Libro de texto en desarrollo

Bioestadística

Una Introducción Conceptual

Iwin Leenen

Universidad Nacional Autónoma de México
Facultad de Medicina
2016

Bioestadística: Una introducción conceptual

Este libro de texto acompaña el curso *Bioestadística Básica* (con clave 67168) del Programa de Posgrado *Maestría y Doctorado en Ciencias Médicas, Odontológicas y de la Salud* de la Universidad Nacional Autónoma de México. El curso se imparte durante el 2do semestre del año académico 2015–2016.

Iwin Leenen
Universidad Nacional Autónoma de México

Organización del curso

Organización práctica

- Lugar de las clases: Aula A-002 de Formación Docente, Facultad de Medicina, UNAM.
- Horario: Los jueves del segundo semestre del año académico 2015–2016 (04/02 – 26/05) de 17h00 a 20h00.
- El profesor se pone a la disposición de los alumnos para aclaraciones adicionales a las clases (por correo electrónico o en su oficina). Sin embargo, se estimula la búsqueda de soluciones por otras fuentes (libros, Internet, etc.) y la colaboración entre alumnos.
- Se distribuirán las presentaciones electrónicas y este texto de apuntes.

Evaluación

La evaluación del curso se hace a través de:

1. *Participación en clase* (resolución de problemas) [30 % de la nota final]
   La última parte de cada clase, se dedica a la resolución de los problemas correspondientes al tema tratado. Los estudiantes preparan estos ejercicios en casa y presentan su solución en clase.
2. *Dos tareas* que implican un análisis de datos con diferentes métodos [20 % de la nota final]
   Los detalles de cada trabajo se anunciarán al menos 15 días antes de su fecha límite de entrega.
3. *Un examen parcial* [20 % de la nota final]
   El examen parcial incluye la materia de los temas 1 a 8 (es decir, la primera parte: estadística descriptiva).
4. *Un examen final* [30 % de la nota final]
   El examen final incluye la materia de los 12 temas cubiertos en el curso (con el énfasis en los temas de la segunda parte: estadística inferencial).

Nótense:

- Las preguntas de los exámenes parcial y final son similares en grado de dificultad y forma de resolución a los problemas que acompañan a cada tema y que se discuten en clase.
- El estudiante prepara tanto las dos tareas como los dos exámenes en casa y envía sus respuestas al profesor por vía electrónica antes de la fecha límite. Trabajos o exámenes que no lleguen a tiempo reciben una nota de cero.
- Al elaborar las respuestas para las tareas y los exámenes, el estudiante puede consultar el material que desee (libro de apuntes, las presentaciones, otros libros, fuentes en Internet, etc.). Para *las tareas*, además se permite que los estudiantes discutan el problema planteado entre sí (únicamente dentro del grupo de estudiantes: no deben consultar a un experto en estadística). Por otro lado, *los exámenes* parcial y final son individuales y los estudiantes no deben consultar a otras personas (aunque puedan consultar el material que quieran).

- Después de ambos exámenes, cada estudiante tendrá una sesión con el profesor en el cual éste indagará más en algunos temas y realizará preguntas adicionales (relacionadas con las preguntas del examen y una de las tareas; la primera tarea se discute junto con el examen parcial, la segunda tarea junto con el examen final). En esta sesión, se determina la nota del estudiante en la tarea y el examen correspondientes.

- Después del examen parcial (pero antes del examen final), cada estudiante puede decidir descartar su nota en este examen. En este caso, se aplica otra versión del examen final (que da un peso igual a los temas de la primera y los de la segunda parte) y el examen final contribuirá el 50 % de la nota final de la asignatura.

Bibliografía

Aunque las presentaciones y los apuntos que se distribuyen, conforman el material para el examen, el estudiante interesado puede consultar los siguientes libros que incluyen una discusión de (algunos de) los temas de este curso, posiblemente desde otra perspectiva:

♣ Dawson B. & Trapp, R. G. (2005). *Bioestadística médica* (4a. Ed., M. R. Carsolio Pacheco, Trad.). Ciudad de México: Manual Moderno.

♣ García Nogales, A. (2004). *Bioestadística básica*. Badajoz, España: Abecedario.

♣ Norman, G. R. & Streiner, D. L. (1994). *Bioestadística* (J. Tarres, Trad.). Madrid, España: Harcourt.

♣ Pagano, M. & Gauvreau, K. (2000). *Principles of Biostatistics* (2a. Ed.). Pacific Groove, CA, EE.UU.: Duxbury.

Índice general

Tema 1
Introducción

Índice

Para este primer tema, en la Sección 1.1 se presentan brevemente los dos objetivos o tareas principales de la estadística. En la Sección 1.2, se resumen algunas nociones importantes de las matemáticas, a las que apelamos con frecuencia en este curso. La Sección 1.3 introduce algunas definiciones básicas necesarias para abordar los siguientes temas. Por último, en la Sección 1.4 se discuten las escalas de medición y los tipos de variables que se derivan de ellas.

1.1 Objetivos de la estadística

El área de la estadística se puede dividir según dos tareas principales*: (a) desarrollar métodos para describir (algunos aspectos de) los datos (Subsección 1.1.1) y (b) desarrollar métodos para inferir información más general a partir de los datos (Subsección 1.1.2).

1.1.1 Describir los datos

En muchas ocasiones, los datos que se han coleccionado y que se desean meter a análisis estadístico inicialmente no están en formato electrónico, lo cual es prácticamente necesario para llevar a cabo el análisis. Tener los datos en formato electrónico muchas veces implica un proceso de conversión que puede consistir en, por ejemplo, la captura manual de los datos o la lectura de hojas especialmente diseñadas por un lector óptico. En consideración de que fácilmente ocurren errores durante dicha conversión, previo a cualquier análisis (descriptivo o inferencial) de datos, es necesario *verificar la validez* de los mismos. Además, aún cuando se recopilan los datos directamente a través de una computadora (como, por ejemplo, el examen profesional teórico en la Facultad de Medicina o muchas encuestas que se aplican por Internet), es primordial investigar la validez de la información que se ha recibido, dado que también la persona que provee los datos puede haberse equivocado al introducir algún elemento en la computadora (por ejemplo, año de nacimiento, sexo, etc). Obviamente, si la información es errónea y no confiable en varias observaciones, los análisis estadísticos tampoco llevarán a conclusiones válidas.

En estudios clínicos—y estudios en las ciencias sociales y de la salud en general—, es común que los datos sean incompletos en el sentido que para una porción significativa de las observaciones el dato en uno o más campos falta y, por lo tanto, este tipo de estudios debe lidiar con el problema de los valores faltantes. La cuestión de cómo tratar la información perdida para que se llegue todavía a conclusiones precisas, conforma un área de investigación en si misma dentro de la estadística. Como parte del proceso de evaluación de la validez de los datos, el investigador examinará los valores perdidos y evaluará hasta qué grado le pueden afectar sus resultados.

Inicialmente los datos recopilados, sobre todo cuando son muchos, constituyen una multitud de información desordenada y exhíben una estructura compleja. La persona responsable del análisis de datos en un proyecto de investigación buscará cómo *ordenar la información*, de tal forma que se pueden acceder eficientemente durante su análisis.

Después de la limpieza y la ordenación de los datos, el análisis descriptivo se enfoca en *representar y resumir* los datos metódicamente, con el objetivo de extraer la información más relevante y comunicarla con colegas u otros investigadores. En un análisis descriptivo, contrario al análisis inferencial, el objetivo es describir la información en los datos disponibles (en la muestra bajo consideración), *sin llegar a conclusiones que transciendan a la muestra*. Una herramienta importante para el resumen y la representación de los datos son las tablas y, más importante aún, las gráficas. Muchas veces, hay verdad en el dicho "una gráfica vale más que mil palabras" (por lo menos, si es una buena gráfica...).

1.1.2 Inducir información más general

Los datos con los que el investigador trabaja y sobre los cuales saca conclusiones siempre son específicos.

*A veces, se distingue una tercera tarea de la estadística, que tiene que ver con el desarrollo de diseños y métodos para la recopilación de datos. Ésta se trata en la asignatura de *Métodos Cuantitativos en Educación Médica*.

Ejemplos

Una investigadora de una empresa farmacéutica quiere estudiar si una nueva medicina es eficaz para retrasar el proceso degenerativo en pacientes con la enfermedad de Alzheimer. Por este fin, diseña un experimento en el cual participan 60 personas a las que se ha diagnosticado Alzheimer en el último mes; a 30 de ellas administra la nueva medicina y a las otras 30 el tratamiento estándar. Un año después observa que los pacientes que recibieron la nueva medicina muestran menos síntomas que los pacientes con tratamiento estándar.

Un psiquiatra quiere iniciar un tratamiento con un niño que tiene problemas de ansiedad. Observa al niño en 10 diferentes situaciones y llega a la conclusión de que el niño desarrolla niveles de angustia más altos en situaciones donde hay adultos presentes.

El alcance de las afirmaciones en los ejemplos anteriores está limitado a los 60 pacientes y a las 10 situaciones exploradas; por otro lado, a los investigadores les interesan conclusiones más generales. En el primer ejemplo, la investigadora quiere encontrar una condición más general que no sólo aplique a los 60 pacientes que incluyó en su muestra. Igualmente, el psiquiatra quiere saber si su paciente *en general* siente más angustia en situaciones donde hay adultos presentes (y no sólo en las 10 situaciones investigadas). Nótese también que, como estos ejemplos muestran, la generalización que se persigue puede ir en diferentes direcciones: La investigadora farmacéutica quiere generalizar a más personas, mientras que el psiquiatra quiere generalizar a más situaciones.

Tales generalizaciones implican exceder los datos; dicho de otra forma, requieren que se haga una *inferencia* o una *inducción*. La estadística inferencial proporciona herramientas para llegar a conclusiones más generales a partir de los datos en una muestra.

Nota sobre la inducción

Se distinguen dos formas de razonamiento: *deducir* e *inducir*. En la deducción se razona desde lo general a lo particular, en la inducción desde lo particular a lo general.

Ejemplo de deducción:

Cada humano es mortal	**general**
Socrates es un humano	
Socrates es mortal	**particular**

Ejemplo de inducción:

Socrates es mortal	**particular**
La princesa Diana es mortal	
Octavio Paz Lozano es mortal	
Mi tatarabuela es mortal	
Cada ser humano es mortal	**general**

Por lo general, las derivaciones deductivas son seguras. Esto no es el caso para las derivaciones inductivas: Por esa razón, en la inducción, el concepto de probabilidad es muy importante. Al generalizar entra una incertidumbre; una probabilidad cuantifica esta incertidumbre. En la estadística inferencial, la noción de probabilidad ocupa un lugar central.

Lo anterior no quiere decir que la estadística inductiva/inferencial en sí misma no sea segura; los teoremas y resultados matemáticos en esta área se comprueban utilizando derivaciones *de*ductivas.

La estructura de este curso refleja la división de la estadística en dos partes como se acaba de describir: La Parte I (con los Temas 2 a 8) corresponde a la estadística descriptiva, mientras la Parte II (con los Temas 9 a 12) presenta temas de la estadística inferencial.

1.2 Algunas nociones de matemáticas

1.2.1 Notación abreviada para sumas y productos

Una suma de términos como $x_3 + x_4 + x_5 + x_6 + x_7$ se abrevia a menudo con el símbolo $\sum\limits_{i=3}^{7} x_i$. El símbolo $\sum$ es la letra mayúscula griega sigma, y en este contexto se interpreta como el *signo sumatorio*. La letra i se llama *índice sumatorio* y el término que sigue al $\sum$ es el *término genérico*. El $i = 3$ abajo del $\sum$ señala que 3 es el *límite inferior* para el índice sumatorio y que el primer término de la suma se obtiene por sustituir i por 3 en el término genérico; el "7" arriba del $\sum$ es el *límite superior* e implica que el término final de la suma se obtiene por sustituir el índice sumatorio por 7 en el término genérico. Los demás terminos se obtienen dando a i los números enteros entre los límites 3 y 7.

Ejemplo

$$\sum_{i=2}^{5} i3^{2i-1} = \underbrace{(2)(3^3)}_{i=2} + \underbrace{(3)(3^5)}_{i=3} + \underbrace{(4)(3^7)}_{i=4} + \underbrace{(5)(3^9)}_{i=5} = 107{,}946$$

A continuación, se enumeran algunas propiedas útiles en expresiones que utilizan el signo sumatorio.

Propiedades de sumas

En las siguientes fórmulas, a y b son números enteros donde $a \leqslant b$; c es cualquier número real.

$$(1) \quad \sum_{i=a}^{b}(x_i + y_i) = \sum_{i=a}^{b} x_i + \sum_{i=a}^{b} y_i$$

$$(2) \quad \sum_{i=a}^{b}(c\, x_i) = c\sum_{i=a}^{b} x_i$$

$$(3) \quad \sum_{i=a}^{b} c = (b - a + 1)\, c$$

Demostración de la tercera propiedad

$$\sum_{i=a}^{b} c = \overbrace{\underbrace{c}_{i=a} + \underbrace{c}_{i=a+1} + \cdots + \underbrace{c}_{i=b-1} + \underbrace{c}_{i=b}}^{b-a+1 \text{ términos}} = (b-a+1)\,c$$

En palabras, la tercera propiedad implica que, cuando el término genérico no contiene el índice sumatorio, siempre es posible usar una expresión más simple que no incluye el signo sumatorio. La comprobación de las primeras dos propiedades se dejan como ejercicio.

Debe ser claro que no importa cual es el índice sumatorio que se usa:

Arbitrariedad del índice sumatorio

$$\sum_{i=a}^{b} x_i = \sum_{j=a}^{b} x_j$$

Casi siempre usaremos las letras i, j y k, como índice sumatorio.

También es posible usar "sumas dobles" (o triples, etc.), como en el siguiente ejemplo:

Ejemplo

$$\begin{aligned}
\sum_{i=2}^{4}\sum_{j=3}^{5} x_{ij} &= \sum_{i=2}^{4}\left(\sum_{j=3}^{5} x_{ij}\right) \\
&= \underbrace{\sum_{j=3}^{5} x_{2j}}_{i=2} + \underbrace{\sum_{j=3}^{5} x_{3j}}_{i=3} + \underbrace{\sum_{j=3}^{5} x_{4j}}_{i=4} \\
&= (x_{23} + x_{24} + x_{25}) + (x_{33} + x_{34} + x_{35}) + (x_{43} + x_{44} + x_{45})
\end{aligned}$$

Obviamente, para que no haya confusión, los índices sumatorios que acompañan ambos signos sumatorios deben ser distintos.

Nótese que, en una suma doble, el límite inferior o superior (o ambos) del segundo signo sumatorio puede contener el índice del primer signo sumatorio. Por ejemplo:

Ejemplo

$$\begin{aligned}
\sum_{i=1}^{4}\sum_{j=1}^{i} x_{ij} &= \sum_{i=1}^{4}\left(\sum_{j=1}^{i} x_{ij}\right) \\
&= \underbrace{\sum_{j=1}^{1} x_{1j}}_{i=1} + \underbrace{\sum_{j=1}^{2} x_{2j}}_{i=2} + \underbrace{\sum_{j=1}^{3} x_{3j}}_{i=3} + \underbrace{\sum_{j=1}^{4} x_{4j}}_{i=4} \\
&= (x_{11}) + (x_{21} + x_{22}) + (x_{31} + x_{32} + x_{33}) + (x_{41} + x_{42} + x_{43} + x_{44})
\end{aligned}$$

Una propiedad importante de las sumas dobles (cuando los límites de un signo sumatorio no contienen el índice del otro) es la siguiente:

Intercambiar el orden de los signos sumatorios en sumas dobles

$$\sum_{i=a}^{b}\sum_{j=c}^{d} x_{ij} = \sum_{j=c}^{d}\sum_{i=a}^{b} x_{ij},$$

donde a, b, c y d son números enteros con $a \leqslant b$ y $c \leqslant d$.

Una notación análoga existe para productos: Para eso, se cambia el $\sum$ por la mayúscula griega pi: $\prod$. Similar al caso de la suma, se obtienen los diferentes factores (elementos en el producto) sustituyendo el índice por los números enteros entre los límites inferiores y superiores. Por ejemplo,

Ejemplo

$$\prod_{i=2}^{6}(i+3) = (2+3)(3+3)(4+3)(5+3)(6+3) = 15{,}120$$

En los ejercicios, se tratan algunas propiedades de productos.

1.2.2 Nociones de la teoría de conjuntos

Conjuntos Un *conjunto* se puede comprender como cualquier colección de objetos bien definida. La palabra *objeto* es un término general y puede referirse a personas, números, escuelas, deportes, etc.. Por lo general, un conjunto se denota con una mayúscula, como A, B, etc., y los objetos se denotan con una minúscula, como a, b, etc.. Para cualquier conjunto A y cualquier objeto a, se puede evaluar si a pertenece o no a A; se escribe $a \in A$ o $a \notin A$. Los objetos que pertenecen a A se llaman los *elementos* de A.

Se dice que dos conjuntos, A y B, son iguales, $A = B$, si cada elemento de A también es elemento de B y cada elemento de B también es elemento de A.

Ejemplos de conjuntos

- El conjunto de las entidades federativas de la República Mexicana.
- El conjunto de los libros escritos por Gabriel García Marquez.
- El conjunto de todas las respuestas que se pueden dar en una pregunta concreta de opción múltiple en un examen.
- Conjuntos de números:
  - $\mathbb{N}$: los números naturales
  - $\mathbb{Z}$: los números enteros
  - $\mathbb{Q}$: los números racionales (recuérdese que un número racional es una fracción a/b donde a y b son números enteros y $b \neq 0$)

  - $\mathbb{R}$: los números reales
  - el intervalo $[0, 1]$
  - ...

- El conjunto que no contiene ningún elemento, llamado el conjunto vacío; se representa por el símbolo $\varnothing$.

Se puede definir un conjunto de dos formas: (a) enumerando sus elementos o (b) especificando una regla que describa las características de los elementos.

Ejemplos

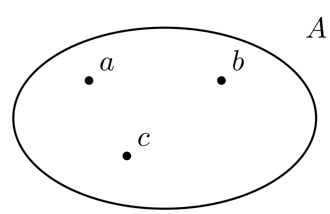
- El conjunto {Adrián, Carlos, Diana, Jesús, Karla, Leonardo, Lizz, Oliver} alternativamente se puede definir como "los estudiantes presentes en la primera clase del curso Bioestadística Básica del año 2015"
- El intervalo $[0, 1]$ se puede definir como $\{x \in \mathbb{R} \mid 0 \leqslant x \leqslant 1\}$.

Si el orden de los elementos en un conjunto tiene importancia, se habla de un conjunto ordenado; un conjunto ordenado se escribe con parentesis redondos como en (primaria, secundaria, preparatoria, licenciatura, posgrado), mientras que si no se toma en cuenta el orden se usan llaves como en {primaria, secundaria, preparatoria, licenciatura, posgrado}. Cuando en este curso se habla de un conjunto sin especificar explícitamente si es ordenado o no, hay que entender que se trata de un conjunto *no* ordenado.

Diagrama de Venn

Una manera común para representar un conjunto es por un *diagrama de Venn*:

$A = \{a, b, c\}$



El número de elementos de un conjunto A se llama *número cardinal* de A y se anota como $\#A$. El número cardinal puede ser finito o infinito. Se dice que el conjunto A es un *subconjunto* del conjunto B si cada elemento de A también es elemento de B. En símbolos se escribe: $A \subseteq B$. Nótese que el conjunto vacío es un subconjunto de cualquier conjunto A. ($\forall A : \varnothing \subseteq A$; el símbolo $\forall$ se lee como "Para cualquier".) La afirmación $A \subseteq B$ no excluye la posibilidad de que $A = B$. De hecho, para cualquier conjunto A, tendremos que $A \subseteq A$. Sin embargo, si $A \subseteq B$ y $A \neq B$, entonces se dice que A es un *subconjunto propio* de B. En estos casos, se puede escribir: $A \subset B$.

El conjunto de todos los subconjuntos de un conjunto A se designa como el *conjunto potencia* y se representa como 2^A.

Ejemplos

- El conjunto de los números naturales es un subconjunto propio del conjunto de los números reales: $\mathbb{N} \subset \mathbb{R}$.
- El intervalo $[0,1]$ no es un subconjunto de $\mathbb{Q}$: $[0,1] \not\subseteq \mathbb{Q}$.
- El conjunto de las personas residentes en el Distrito Federal es un subconjunto del conjunto de las personas residentes en la República Mexicana.
- El conjunto de los diferentes tipos de esquizofrenia es un subconjunto de los transtornos psicóticos.
- Si $A = \{a, b, c\}$, entonces $2^A = \{\varnothing\ , \{a\}\ , \{b\}\ , \{c\}\ , \{a, b\}\ , \{a, c\}\ , \{b, c\}\ , \{a, b, c\}\}$.

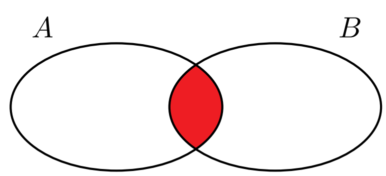
En el último ejemplo, se observa que el número cardinal del conjunto potencia es igual a 2 elevado al número cardinal del conjunto original. Es decir $\#2^A = 2^{\#A} = 8$. Es una regla general; para cualquier A finito, se satisface $\#2^A = 2^{\#A}$.

Tres operaciones asociadas con conjuntos

1. La *intersección* de dos conjuntos, A y B, anotado como $A \cap B$, es el conjunto de los elementos que pertenecen a ambos, a A y a B. En símbolos:

$$A \cap B = \{x \mid x \in A \ \text{ y } \ x \in B\}$$

Gráficamente, la intersección $A \cap B$ es la parte en rojo:

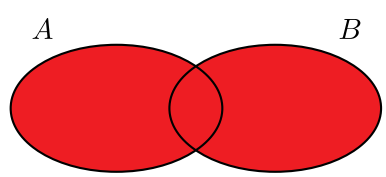


Si la intersección de los conjuntos A y B está vacía, se dice que son conjuntos *disjuntos*.

2. La *unión* de dos conjuntos, A y B, anotado como $A \cup B$, es el conjunto de los elementos que pertenecen, ya sea a A, ya sea a B, ya sea a ambos. En símbolos:

$$A \cup B = \{x \mid x \in A \ \text{ o } \ x \in B\}$$

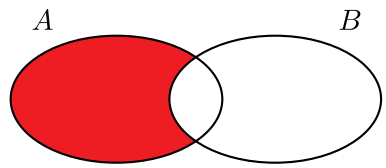
Gráficamente, la unión $A \cup B$ es la parte en rojo:



3. La *diferencia* de dos conjuntos, A menos B, anotado como $A \setminus B$, es el conjunto de los elementos que pertenecen a A y no a B. En símbolos:

$$A \setminus B = \{x \mid x \in A \text{ y } x \notin B\}$$

Gráficamente, la diferencia $A \setminus B$ es la parte en rojo:



Nótese que la intersección y la unión son operaciones conmutativas (es decir, $A \cap B = B \cap A$ y $A \cup B = B \cup A$), mientras que la diferencia no es conmutativa (generalmente, $A \setminus B \neq B \setminus A$).

Ejemplos

Definamos los siguientes conjuntos:

- A: el conjunto de los estados de la República Mexicana cuyo nombre coincide con el nombre de su capital, es decir:

$$A = \{\text{Aguascalientes, Campeche, Chihuahua, Colima, Durango, Guanajuato, Oaxaca, Puebla, Queretaro, San Luis Potosí, Tlaxcala, Zacatecas}\}.$$

- B: el conjunto de todos los estados mexicanos que limitan con un océano, es decir:

$$B = \{\text{Baja California, Baja California Sur, Campeche, Chiapas, Colima, Guerrero, Jalisco, Michoacán, Nayarit, Oaxaca, Quintana Roo, Sinalao, Sonora, Tabasco, Tamaulipas, Veracruz, Yucatán}\}.$$

Entonces,

- La intersección de A y B es el conjunto de todos los estados cuyo nombre coincide con el nombre de su capital y que limitan con un océano, es decir:

$$A \cap B = \{\text{Campeche, Colima, Oaxaca}\}$$

- La unión de A y B es el conjunto de todos los estados cuyo nombre coincide con el nombre de su capital o que limitan con un océano, es decir:

$$A \cup B = \{\text{Aguascalientes, Baja California, Baja California Sur, Campeche, Chiapas, Chihuahua, Colima, Durango, Guanajuato, Guerrero, Jalisco, Michoacán, Nayarit, Oaxaca, Puebla, Queretaro, Quintana oo, San Luis Potosí, Sinalao, Sonora, Tabasco, Tamaulipas, Tlaxcala, Veracruz, Yucatán, Zacatecas}\}.$$

- La diferencia A menos B es el conjunto de todos los estados cuyo nombre coincide con el nombre de su capital y que *no* limitan con un océano, es decir:

$$A \setminus B = \{\text{Aguascalientes, Chihuahua, Durango, Guanajuato, Puebla, Queretaro, San Luis Potosí, Tlaxcala, Zacatecas}\}.$$

Algunas propiedades de operaciones con conjuntos

- $A \cap B = B \cap A$
- $A \cap (B \cap C) = (A \cap B) \cap C$
- $A \cap \varnothing = \varnothing$
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup B = B \cup A$
- $A \cup (B \cup C) = (A \cup B) \cup C$
- $A \cup \varnothing = A$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Es muy instructivo ilustrar las propiedades anteriores con diagramas de Venn. La comprobación se deja como ejercicio.

Producto Cartesiano

El *producto Cartesiano* de dos conjuntos A y B es el conjunto de todos los pares ordenados, donde el primer elemento del par pertenece al conjunto A y el segundo elemento pertenece al conjunto B. Se escribe:

$$A \times B = \{(a, b) \mid a \in A \text{ y } b \in B\}$$

Un caso especial se presenta si los dos conjuntos son iguales. $A \times A$ se escribe también como A^2.

Nótese que $\#(A \times B) = \#A \times \#B$.

Ejemplos

Definamos A como el conjunto de tareas psicólogicas {Hooper, Willeford, Wisconsin} y B como el conjunto de los lóbulos cerebrales {frontal, occipital, parietal, temporal}. (El nombre completo de la prueba de Hooper es la “Prueba de Organización Visual de Hooper”; la prueba de Willeford es la prueba de percepción del habla que alterna rápidamente [Rapidly Alternating Speech Perception]; la prueba de Wisconsin es el “Test de Clasificación de Cartas de Wisconsin”). El producto Cartesiano $A \times B$ es el conjunto de los pares ordenados:

$$A \times B = \{\text{(Hooper, frontal), (Hooper, occipital), (Hooper, parietal), (Hooper, temporal), (Willeford, frontal), (Willeford, occipital), (Willeford, parietal), (Willeford, temporal), (Wisconsin, frontal), (Wisconsin, occipital), (Wisconsin, parietal), (Wisconsin, temporal)}\}$$

El número cardinal del conjunto $A \times B$ es 12; es igual al producto de los números cardinales de A y B.

Siendo A el conjunto de los países {Belice, Canadá, Estados Unidos, Guatemala, México}, A^2 es el conjunto de los pares:

A^2 = {(Belice, Belice), (Belice, Canadá), (Belice, Estados Unidos), (Belice, Guatemala), (Belice, México), (Canadá, Belice), (Canadá, Canadá), (Canadá, Estados Unidos), (Canadá, Guatemala), (Canadá, México), (Estados Unidos, Belice), (Estados Unidos, Canadá), (Estados Unidos, Estados Unidos), (Estados Unidos, Guatemala), (Estados Unidos, México), (Guatemala, Belice), (Guatemala, Canadá), (Guatemala, Estados Unidos), (Guatemala, Guatemala), (Guatemala, México), (México, Belice), (México, Canadá), (México, Estados Unidos), (México, Guatemala), (México, México) }

$$\#(A^2) = (\#A)^2 = 25$$

La definición del producto Cartesiano se puede generalizar a más de dos conjuntos. Se escribe $A \times B \times C$, A^4, etc.. El producto Cartesiano es estonces un conjunto de tripletes, quadrupletes, etc..

Relaciones

Relación

Una *relación R de A a B* es cualquier subconjunto del producto Cartesiano; es decir, si R es una relación de A a B, entonces, $R \subseteq A \times B$.
Un subconjunto de A^2 se llama una *relación en A*.

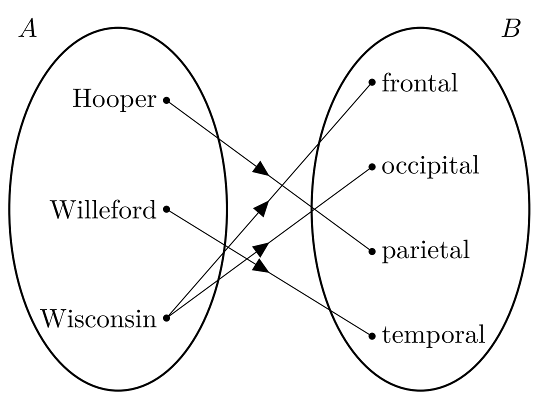
A continuación, se exponen algunos ejemplos de relaciones junto con su representación gráfica.

Ejemplos

Para A = {Hooper, Willeford, Wisconsin} y B = {frontal, occipital, parietal, temporal}, se puede considerar la relación R entre A y B que se lee como "... apela fuertemente al lóbulo cerbral ...":

R = {(Hooper, parietal), (Willeford, temporal), (Wisconsin, frontal), (Wisconsin, occipital)}

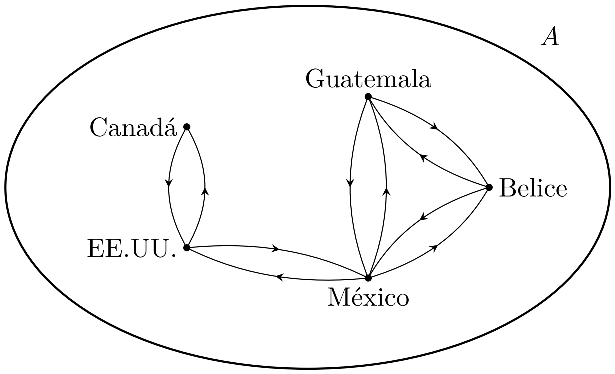
Gráficamente:



Para $A = \{$Belice, Canadá, Estados Unidos, Guatemala, México$\}$ podemos definir la relación R "... colinda con ..." en A:

$R = \{$(Belice, Guatemala), (Belice, México), (Canadá, Estados Unidos), (Estados Unidos, Canadá), (Estados Unidos, México), (Guatemala, Belice), (Guatemala, México), (México, Belice), (México, Estados Unidos), (México, Guatemala)$\}$

Gráficamente:



Nótese que la relación R de este ejemplo es simétrica.

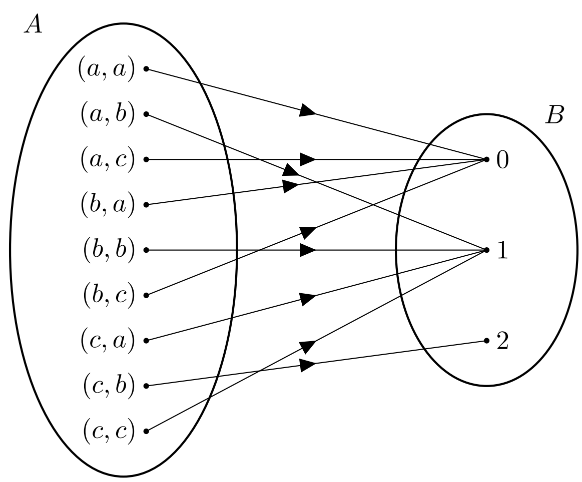
Funciones

Definiciones

- Una *función f de A a B*, anotada como $f : A \to B$, es una relación de A a B donde cada elemento de A se asocia con un único elemento de B.
- El elemento de B con que el elemento a de A se asocia, es la *imagen* de a por la función f y se anota como: $f(a)$. Se escribe: $a \mapsto f(a)$.
- Para $f : A \to B$, el conjunto A se llama el *dominio* de la función (a veces, el *conjunto de partida* o el *conjunto inicial*); el conjunto B es el *codominio* o *contradominio* (a veces, se llama el *conjunto de llegada* o *conjunto final*).
- El conjunto de todos los elementos de B que son la imagen de uno o más elementos de A se conoce como el *rango* o el *recorrido* de la función (a veces, el *conjunto imagen*). Se denota como $f(A)$.
- Si $A = B$, es decir, si el dominio y el codominio son idénticos, se habla a veces de una *transformación*.

Ejemplos

Consideremos, por un lado, el conjunto A de todas los posibles patrones de respuestas en un examen de opción múltiple que consiste en dos preguntas con tres alternativas. Si anotamos las tres posibles respuestas en cada pregunta con a, b, y c, se puede escribir $A = \{(a,a),(a,b),(a,c),(b,a),(b,b),(b,c),(c,a),(c,b),(c,c)\}$. Por otro lado, definamos el conjunto $B = \{0,1,2\}$.
Entonces, la siguiente función $f : A \to B$ asocia con cada patrón de respuestas el número de respuestas correctas en el examen (suponiendo que las respuestas correctas en la primera y segunda pregunta son c y b, respectivamente):

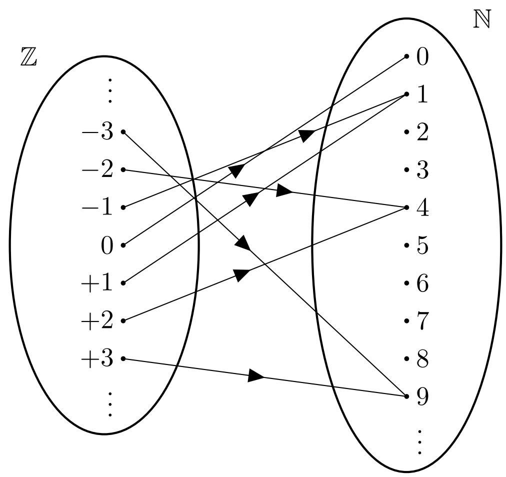


Se trata efectivamente de una función f, puesto que cada elemento de A se asocia con precisamente un elemento de B. (En la representación gráfica, sale una y solo una flecha de cada elemento en A.) El conjunto A con todos los posibles patrones de respuesta es el dominio; el conjunto B con las calificaciones 0, 1, y 2 es el codominio, lo cual en este caso coincide con el rango o el recorrido de la función. La imagen del patrón de respuestas correctas (c,b) por la función f es 2. Se escribe: $f[(c,b)] = 2$.

Como otro ejemplo, considera la función $f : \mathbb{Z} \to \mathbb{N}$ que asocia cada número entero con su cuadrado. Es decir, para cualquier $x \in \mathbb{Z}$, $x \mapsto x^2$. En la parte superior de la siguiente página se encuentra una representación gráfica de (una parte de) esta función.

Desde cada elemento de $\mathbb{Z}$ sale una flecha a su imagen en $\mathbb{N}$, por lo cual f es una función. La definición de f implica que el dominio es $\mathbb{Z}$, el codominio es $\mathbb{N}$ y el rango de la función es el conjunto de números que son cuadrados, es decir, $f(\mathbb{Z}) = \{0,1,4,9,\ldots\}$.

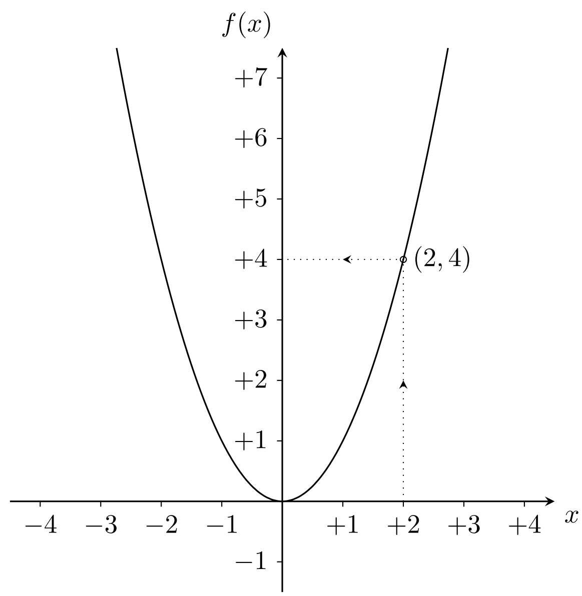
Como ilustran este ejemplo y el anterior, el rango de la función es un subconjunto de (y posiblemente igual a) el codominio.



A veces, se interpreta una función como el procedimiento o algoritmo cuya aplicación en un elemento a de A da como resultado el elemento $f(a)$ de B. Esta interpretación es muy común si se trata de una función entre dos conjuntos de números. En el caso especial donde se trata de una función de $\mathbb{R}$ a $\mathbb{R}$, muchas veces se hace una gráfica en el plano, donde cada par de la función es representado por un punto con coordenadas $(x, f(x))$.

Ejemplo

Consideremos la función $f : \mathbb{R} \to \mathbb{R}$, donde $x \mapsto x^2$. (Nótese que, aunque la regla de esta función, $x \mapsto x^2$, es idéntica que la de la función en el ejemplo anterior, el dominio y el codominio son diferentes.) La representación gráfica en el plano Cartesiano se encuentra en la siguiente figura:



La figura indica explícitamente las coordenadas del par $(2, 4)$, el cual es uno de los elementos de la función. Hay un número infinito de pares en f y los puntos en sus coordenadas forman la parábola representada en la figura.
Nótese que, como el cuadrado siempre es un número positivo, el rango de la función f es $\mathbb{R}^+$ (el conjunto de los números reales positivos), un subconjunto del codominio $\mathbb{R}$.

Entre las funciones de $\mathbb{R}$ a $\mathbb{R}$ se distinguen diferentes familias o tipos de funciones, como las funciones lineales, las cuadráticas, las polinomiales, las exponenciales, las logarítmicas, las goniométricas, las hiperbólicas, etc.. Aquí nos limitamos a una exposición de las funciones lineales, ya que este tipo de funciones se usa con mucha frecuencia en los modelos estadísticos.

Función lineal

Una *función lineal* es una función cuyo dominio y codominio son el conjunto de números reales, es decir $f : \mathbb{R} \to \mathbb{R}$, donde la imagen para cualquier $x \in \mathbb{R}$ se obtiene por

$$f(x) = ax + b,$$

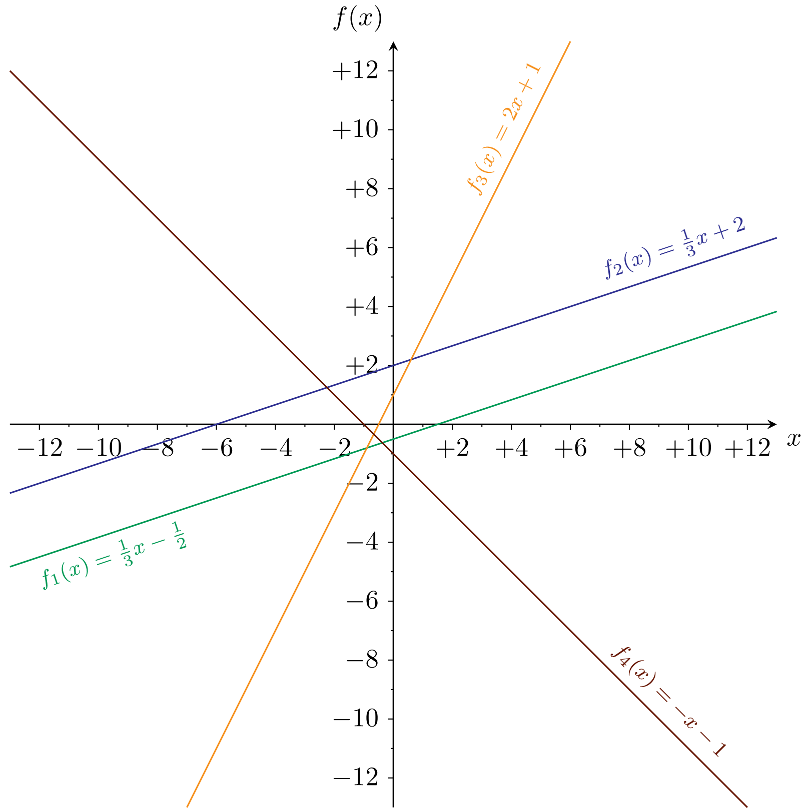
donde a y b son dos números reales fijos, las *constantes* de la función lineal.
Se llama a a la *pendiente* y a b la *intersección* de la función.

Es importante entender bien el concepto de una constante; a y b son constantes porque su valor no cambia al aplicar la regla para conocer los pares $(x, f(x))$ de la función; en otras palabras, al considerar una función lineal concreta se calcula la imagen $f(x) = ax + b$ utilizando siempre los mismos valores para a y b. Una constante es diferente de una variable, cuyo valor sí varía al encontrar los diferentes pares que constituyen la función. Así, x es la variable en la función (se sustituye la x por valores distintos al determinar los pares de la función o al construir la gráfica de la función).

Cabe señalar que, a menudo, se utilizan otros símbolos para a y b en las funciones lineales, por ejemplo $f(x) = mx + q$ o $f(x) = b_1x + b_0$. Debe ser claro, sin embargo, que eso no cambia la esencia de una función lineal. De una forma u otra, la regla de la función lineal indica que se multiplica la variable por una constante y se añade otra constante.

La representación gráfica de una función lineal es una recta. En la siguiente figura se grafican algunas funciones lineales.

Ejemplo



Como se puede apreciar en las funciones representadas en el ejemplo anterior, la constante b (la intersección de la función lineal) es el valor en el eje vertical donde la recta lo corta. Indica qué tan arriba se encuentra en el plano (por ejemplo, con las funciones f_1 y f_2, las cuales difieren únicamente en la constante b, se observa que f_2 está más arriba que f_1 debido a que para f_2 se tiene $b = +2$ y para f_1 se tiene $b = -\frac{1}{2}$.

Por otro lado la pendiente indica cuánto incrementa la imagen $f(x)$ cuando x incrementa una unidad. Por ejemplo, para la función f_3, se puede verificar que, si x incrementa una unidad, entonces $f(x)$ incrementa dos unidades (por ejemplo, si x cambia de 1 a 2 entonces, $f(x)$ incrementa de 3 a 5). En el caso de que la pendiente a sea negativa, entonces la función es decreciente. Se invita al lector que reflexione sobre el caso de una pendiente igual a 0.

Definición

1. Una *función de dos variables* $f : \mathbb{R}^2 \to \mathbb{R}$ es una función que asigna un número real a cada par ordenado de números reales. Se escribe: $(x, y) \mapsto f(x, y)$.

2. Una *función de tres variables* $f : \mathbb{R}^3 \to \mathbb{R}$ es una función que asigna un número real a cada triplete de números reales. Se escribe: $(x, y, z) \mapsto f(x, y, z)$.

3. Una *función de n variables* $f : \mathbb{R}^n \to \mathbb{R}$ es una función que asigna un número real a cada conjunto ordenado de n números reales. Se escribe: $(x_1, x_2, \ldots, x_n) \mapsto f(x_1, x_2, \ldots, x_n)$.

El dominio de las funciones multivariables que se acaban de definir consiste en todos los pares (si es una función de dos variables), todos los tripletes (para funciones de tres variables) o, en el caso general de una función de n variables, en todos los conjuntos ordenados de n números reales. Por otro lado, el codominio de estas funciones multivariables es siempre $\mathbb{R}$.

Ejemplos

- Se puede definir la función en dos variables $f : \mathbb{R}^2 \to \mathbb{R}$, que a dos números reales, (x, y) asigna su diferencia. En símbolos: $f(x, y) = x - y$.
- La nota final para el curso de *Bioestadística Básica* se puede considerar como una función en cuatro variables. Es decir, una función que convierte la participación en clase (x_1), la calificación en las tareas (x_2) y las calificaciones en el exámenes parcial (x_3) y final (x_4), en un solo número. En este caso:

$$f(x_1, x_2, x_3, x_4) = 0.30x_1 + 0.20x_2 + 0.20x_3 + 0.30x_4.$$

- La función $f : \mathbb{R}^3 \to \mathbb{R}$ que calcula para un triplete de valores (x, y, z) la suma de los valores cuadrados, $f(x, y, z) = x^2 + y^2 + z^2$, es un ejemplo de una función de tres variables.
- El porcentaje de respuestas correctas en un examen de opción múltiple de n preguntas es una función de las variables $x_1, x_2, \ldots, x_n$, donde las x_i ($i = 1, \ldots, n$) corresponden con las puntuaciones en las respectivas preguntas (donde $x_i = 1$ si se marcó la respuesta correcta, y $x_i = 0$ si se marco una de las respuestas incorrectas). En este caso,

$$f(x_1, x_2, \ldots, x_n) = 100 \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{1.1}$$

También entre las funciones $f : \mathbb{R}^n \to \mathbb{R}$ se definen funciones lineales, aunque en este caso utilizan más el término combinación lineal.

Combinación lineal

Una *combinación lineal* es una función f de n variables, es decir $f : \mathbb{R}^n \to \mathbb{R}$, donde la imagen se obtiene por

$$f(x_1, x_2, \ldots, x_n) = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

o, utilizando la notación abreviada para sumas

$$f(x_1, x_2, \ldots, x_n) = a_0 + \sum_{i=1}^{n} a_ix_i$$

donde las $a_0, a_1, \ldots, a_n$ son constantes (elementos de $\mathbb{R}$).

Ejemplos

- La función en la Ecuación (1.1) en el ejemplo anterior, que calcula el porcentaje de respuestas correctas en un examen de opción múltiple, es una combinación lineal ya que

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= 100\,\frac{x_1 + x_2 + \cdots + x_n}{n} \\ &= \frac{100}{n}\,x_1 + \frac{100}{n}\,x_2 + \ldots + \frac{100}{n}\,x_n. \end{aligned}$$

  Es decir, es una combinación lineal donde $a_0 = 0$ y $a_1 = a_2 = \cdots = a_n = \frac{100}{n}$. (Nótese que n es el número de preguntas en el examen y por tanto es una constante.)

- La función $f : \mathbb{R}^3 \to \mathbb{R}$, donde $(x, y, z) \mapsto x^2 + y^2 + z^2$ *no* es una combinación lineal.

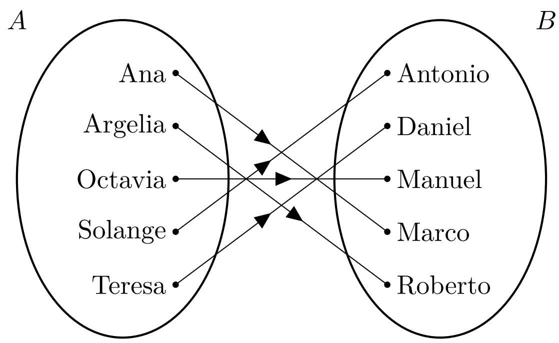
Biyección Una biyección es una función especial:

Biyección

Una función $f : A \to B$ es una *biyección* si y sólo si cada elemento de B es la imagen de un solo elemento de A.
Otro nombre para una biyección es una *relación unívoca*.

Ejemplos

- Sea A el conjunto de mujeres {Ana, Argelia, Octavia, Solange, Teresa} y B el conjunto de los hombres {Antonio, Daniel, Manuel, Marco, Roberto}. Considere la relación “es la esposa de ” entre A y B, representada en el siguiente diagrama:



  Esta relación es una biyección, porque (a) cada elemento de A se asocia con exactamente un elemento de B (lo cual es necesario para que esta relación sea una función) y (b) cada elemento de B es la imagen de exactamente un elemento de A. En la gráfica, se verifica que desde cada elemento de A sale precisamente una flecha y en cada elemento de B llega precisamente una flecha.

- Todas las funciones lineales $f : \mathbb{R} \to \mathbb{R}$ para las cuales $x \mapsto ax + b$ y $a \neq 0$ son biyecciones.

Una nota sobre el número cardinal

Por definición, dos conjuntos tienen el mismo número cardinal si y sólo si se puede establecer una biyección entre ambos. Obviamente, en el caso de conjuntos finitos, se puede verificar la igualdad o desigualdad del número cardinal a través de un simple conteo de los elementos en ambos conjuntos. En este caso, la definición en términos de una biyección no parece ser de mucha utilidad, aunque es trivial que la existencia de una biyección entre ambos conjuntos es equivalente con obtener el mismo resultado en el conteo de los elementos.

Por otro lado, en el caso de conjuntos infinitos, un conteo ya no es posible y la única herramienta para comprobar la igualdad del número cardinal es mostrar que existe una biyección entre los conjuntos. Al respecto, cabe señalar que Georg Canter (1845–1918) comprobó que hay diferentes niveles de infinidad. Es decir que, aunque dos conjuntos A y B son infinitos, posiblemente tienen un número cardinal distinto. Por ejemplo, Canter comprobó que el conjunto $\mathbb{R}$ de los números reales tiene un número cardinal más grande (tiene un nivel de infinidad más alto) que el conjunto $\mathbb{N}$ de los números naturales.

En la estadística inferencial, se distingue entre conjuntos infinitos que son *numerables* y los que son *innumerables*. Un conjunto es numerable infinito si y sólo si tiene el mismo número cardinal que $\mathbb{N}$; si no es numerable, entonces es innumerable. Sinónimo de "numerable" e "innumerable" son "contable" e "incontable". En otras palabras, un conjunto es numerable si se puede indicar una regla de cómo contar todos los elementos del conjunto bajo consideración sin olvidar ninguno.

Estudiemos algunos ejemplos:

- El conjunto $A = \{0, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\}$ es numerable; se puede definir la siguiente función $f : \mathbb{N} \to A$:
$$\begin{cases} x \mapsto x & \text{si } x = 0, \\ x \mapsto \frac{1}{x} & \text{si } x > 0. \end{cases}$$
Es evidente que esta función es una biyección: Cada elemento de $\mathbb{N}$ se asocia con un único elemento de A y cada elemento de A es la imagen de exactamente un elemento de $\mathbb{N}$.

- Mientras que el ejemplo anterior es intuitivamente plausible, en muchas ocasiones se llega a conclusiones contraintuitivas. Por ejemplo, considere la siguiente función de $\mathbb{N}$ a $\mathbb{Z}$:
$$\begin{cases} x \mapsto 2x & \text{si } x \geqslant 0, \\ x \mapsto 2|x| - 1 & \text{si } x < 0, \end{cases}$$
donde $|x|$ denota el valor absoluto de x. También esta función es una biyección, lo que lleva a la conclusión que el conjunto $\mathbb{N}$ de números naturales tiene el mismo número de elementos que el conjunto $\mathbb{Z}$ de números enteros, aunque $\mathbb{N}$ es un subconjunto propio de $\mathbb{Z}$. Efectivamente, se puede contar los elementos en $\mathbb{Z}$ de forma sistemática, garantizando que no se olvide ningún elemento, con la siguiente enumeración: 0, -1, 1, -2, 2, -3, 3, -4, 4,. . . .

- Incluso $\mathbb{Q}$ (el conjunto de todos los números racionales, es decir, el conjunto de todas las fracciones de dos números enteros), y $\mathbb{N}^2$ (el conjunto de todos los pares ordenados de dos números naturales) son numerables y, por lo tanto, tienen el mismo número de elementos que $\mathbb{N}$. Entonces, ¡no hay más fracciones que números naturales!

Para ilustrar lo anterior, consideremos la siguiente manera para ordenar sistemáticamente todos los pares en $\mathbb{N}^2$:

$0 \to (0,0)$	$5 \to (2,0)$	$10 \to (0,4)$	$15 \to (0,5)$	$20 \to (5,0)$
$1 \to (0,1)$	$6 \to (0,3)$	$11 \to (1,3)$	$16 \to (1,4)$	$21 \to (0,6)$
$2 \to (1,0)$	$7 \to (1,2)$	$12 \to (2,2)$	$17 \to (2,3)$	$22 \to (1,5)$
$3 \to (0,2)$	$8 \to (2,1)$	$13 \to (3,1)$	$18 \to (3,2)$	$23 \to (2,4)$
$4 \to (1,1)$	$9 \to (3,0)$	$14 \to (4,0)$	$19 \to (4,1)$	$\vdots$

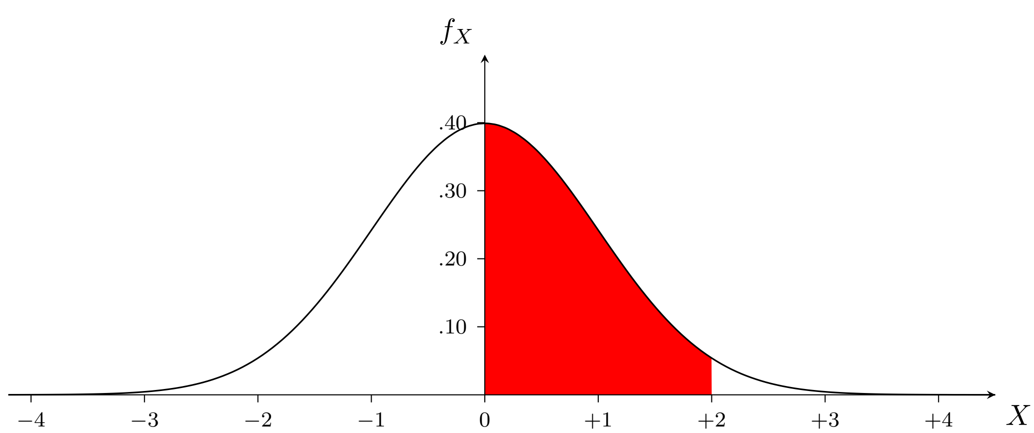
- Asímismo, el conjunto de los números naturales impares $\{1,3,5,7,\dots\}$, el conjunto de los números cuadrados perfectos $\{0,1,4,9,16,25,\dots\}$ y el conjunto de los números primos $\{2,3,5,7,11,13,17\dots\}$ tienen el mismo número cardinal que $\mathbb{N}$.
- El intervalo $[0,1]$, el conjunto potencia de los números naturales ($2^{\mathbb{N}}$, es decir, el conjunto de todos los subconjuntos de $\mathbb{N}$) y $\mathbb{R}$ son ejemplos de conjuntos *in*numerables.

1.2.3 La integral definida

En la estadística inferencial y especialmente en el caso de funciones de densidad continua, el área debajo de la función se interpreta en términos de una probabilidad. Para calcular una área y probabilidad de este tipo se puede recorrer al concepto de la integral definida. Antes de establecer formalmente la relación entre la integral definida y un área debajo de una curva, se ilustra la idea principal mediante un ejemplo.

Ejemplo

Supongamos que nos interesa conocer el área en color rojo en la siguiente figura:



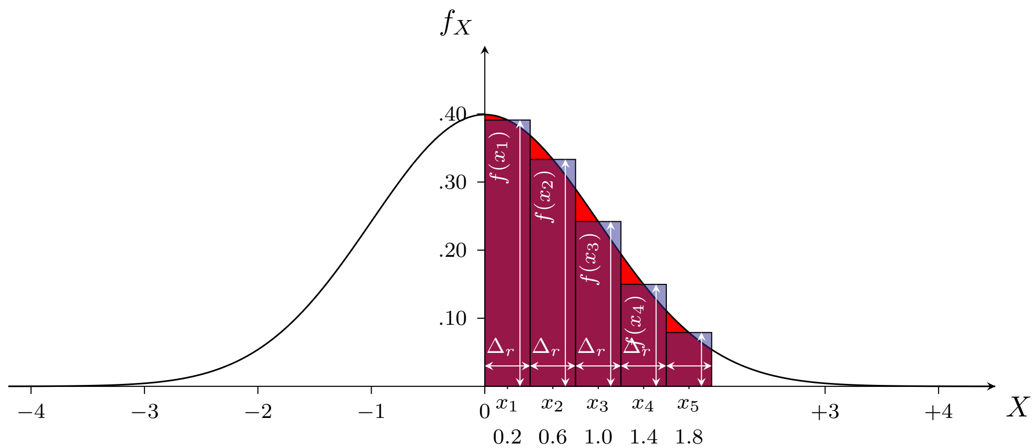
La curva graficada es la función de densidad normal estandarizada, la cual ocupa un lugar central en la estadística inferencial. La regla para calcular la imagen de un valor x por esta función se da por:

$$f(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}, \tag{1.2}$$

donde e es la base de los logaritmos naturales (e ≈ 2.71828).

Es claro que las fórmulas comunes de la geometría son insuficientes para calcular el área en rojo de la figura (es decir, el área debajo de la curva entre los valores 0 y 2, limitada por la abscisa). Sin embargo, se pueden utilizar estas fórmulas para *aproximar* el área.

Por este fin, se procede como sigue: Como primer paso se divide el intervalo entre los límites 0 y 2 en r partes de la misma longitud. La anchura de estos intervales se representa por Δ_r y sus centros por $x_1, x_2, \ldots, x_i, \ldots, x_r$. De lo anterior sigue que la distancia entre cualquier par de centros x_i y x_{i+1} es igual a Δ_r. Como segundo paso, se construye un rectángulo arriba de cada una de las r partes; la base de cada uno de estos r rectángulos es Δ_r, mientras la altura de cada rectángulo depende de la imagen que da la función para el valor x del centro. Es decir, la altura del rectángulo i es igual a $f(x_i)$. La siguiente figura ilustra esta idea para $r = 5$.



La siguiente tabla presenta los cálculos para las áreas de las $r = 5$ rectángulos:

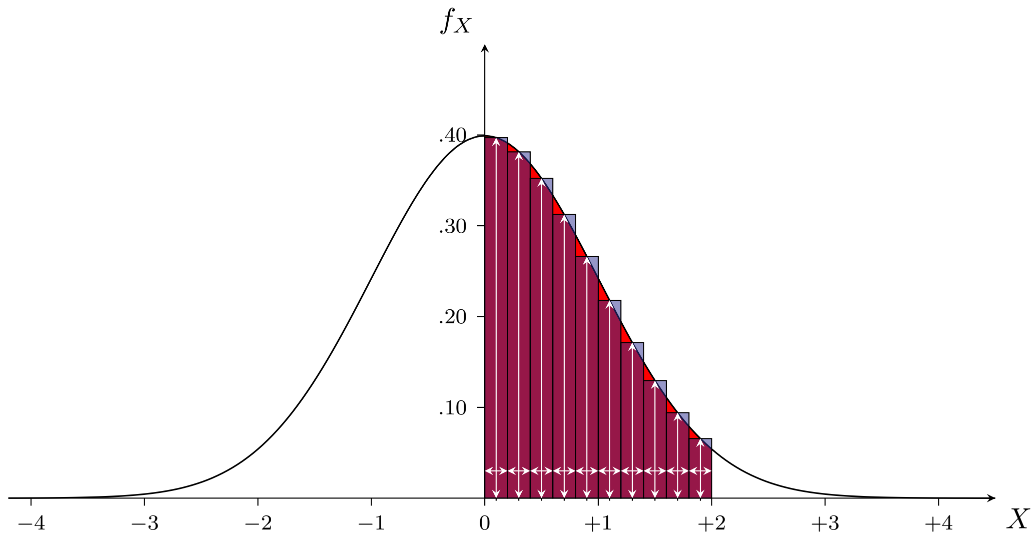
		Base	Altura	Área
i	x_i	Δ_r	$f(x_i)$	$f(x_i)\Delta_r$
1	0.2	0.4	0.3910	0.1564
2	0.6	0.4	0.3332	0.1333
3	1.0	0.4	0.2420	0.0968
4	1.4	0.4	0.1497	0.0599
5	1.8	0.4	0.0790	0.0316
			$\sum_{i=1}^{5} f(x_i)\,\Delta_r =$	0.4780

Por ejemplo, la altura del primer rectángulo se obtiene sustituyendo x en la Ecuación (1.2) de la función de densidad por el valor $x_1 = 0.2$:

$$f(x_1) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(0.2)^2}{2}} = 0.3910.$$

La suma de las áreas de todos los rectángulos (igual a 0.4780, véase la última fila de la tabla anterior) da una aproximación del área debajo de la curva $f(x)$ que interesa conocer (el área en color rojo). Es fácil entender que se trata de una aproximación: Los rectángulos no cubren perfectamente el área debajo de la curva: algunas partes sobran, otras faltan, y el área de las partes que sobran no es exactamente igual al área de las partes que faltan.

Sin embargo, se puede mejorar la precisión de la aproximación si se aumenta el número r de intervalos (así dividiendo el intervalo total de 0 a 2 en más partes). Resulta que la base Δ_r de los rectángulos se reduce, junto con el área de las partes de cada rectángulo que sobran y faltan. La siguiente figura y tabla ilustran el proceso para el caso de $r = 10$.



		Base	Altura	Área
i	x_i	Δ_r	$f(x_i)$	$f(x_i)\Delta_r$
1	0.1	0.2	0.3970	0.0794
2	0.3	0.2	0.3814	0.0763
3	0.5	0.2	0.3521	0.0704
4	0.7	0.2	0.3123	0.0625
5	0.9	0.2	0.2661	0.0532
6	1.1	0.2	0.2179	0.0436
7	1.3	0.2	0.1714	0.0343
8	1.5	0.2	0.1295	0.0259
9	1.7	0.2	0.0940	0.0188
10	1.9	0.2	0.0656	0.0131
		$\sum_{i=1}^{10} f(x_i)\,\Delta_r =$		0.4774

Si se repite este procedimiento incrementando el valor de r, la suma de las áreas de los r rectángulos se aproximará cada vez más al área buscada. De forma más precisa, se define la suma

$$A_r = \sum_{i=1}^{r} f(x_i)\,\Delta_r,$$

la cual corresponde con la aproximación basada en r rectángulos. Si r (el número de rectangulos) aumenta, la aproximación mejorará, hasta que en el límite (es decir, para $r \to \infty$) se obtiene exactamente el área de interés. Se dice que A_r "converge" al área de interés.

La siguiente tabla ilustra esta convergencia; muestra que para valores más grande de r, el área A_r converge al valor exacto.

Número de rectángulos	Área total A_r
r	$\sum_{i=1}^{r} f(x_i)\,\Delta_r$
5	0.477966
10	0.477430
20	0.477295
50	0.477257
100	0.477252
1000	0.477250
$\vdots$	$\vdots$
∞	0.477250

Integral definida

Dado una función $f : \mathbb{R} \to \mathbb{R}$ y un intervalo $[a, b]$, la *integral definida* de $f(x)$ entre los límites a y b se puede interpretar como el área incluida entre la función f, el eje horizontal, y las rectas $x = a$ y $x = b$ y coincide con:

$$\int_a^b f(x)dx = \lim_{r\to\infty} \left[\sum_{i=1}^{r} f(x_i)\,\Delta_r \right],$$

donde $x_1, x_2, \ldots, x_r$ son lo centros de los intervalos obtenidos por dividir el intervalo $[a, b]$ en r partes iguales y $\Delta_r = \frac{b-a}{r}$ es la anchura de cada intervalo.

Los valores a y b se llaman el límite inferior y el límite superior de la integral (e indican el intervalo en el cual se calcula el área debajo de $f(x)$). Nota la similaridad entre la estructura de la notación para la integral y la de la suma de los rectángulos: $f(x)$ es la altura y dx es la base de los rectángulos infinitesimales y el símbolo $\int$ se puede leer como una S alargada que indica (el límite de) la suma de las áreas (base $\times$ altura) de todos los rectángulos. Entonces, geométricamente la integral definida coincide con el área debajo de $f(x)$ entre los valores a y b.

Casi nunca uno mismo debe calcular el valor de una integral definida: Para las funciones más usadas en la estadística existen libros con tablas donde se pueden leer los valores de integrales definidas bajo estas funciones. Además, los paquetes (software) de análisis estadísticos también proporcionan estos valores de manera estándar. Por lo tanto, la discusión anterior no era para enseñar un método para calcular una integral, sino para aclarar la relación entre el concepto de una integral definida y el área debajo de una curva.

1.3 Conceptos básicos

Un paso previo al análisis estadístico es la realización de un *experimento aleatorio*, lo cual nos proporciona los datos.

Experimento aleatorio

Un *experimento* es cualquier procedimiento que se efectúa de acuerdo con reglas fijas y bajo unas circunstancias determinadas y que, en principio, se puede realizar un número infinito de veces.
Si diferentes ejecuciones del experimento pueden llevar a resultados distintos, se habla de un *experimento aleatorio*. En el caso contrario es un *experimento determinístico*.

Ejemplos

Algunos ejemplos de experimentos aleatorios son:

- Echar tres monedas.
- Como en la lotería Melate: sacar, sin reposición, 7 esferas de una urna que contiene 56 esferas numeradas de 1 a 56.
- Presentar a un estudiante un examen de 40 preguntas de opción múltiple, donde cada pregunta tiene 4 opciones de respuesta.
- Elegir al azar una persona de una población determinada de personas y aplicarle una prueba para conocer su grupo sanguíneo.
- Hacer un diagnóstico a un paciente antes y después de la administración de un nuevo fármaco que pretende disminuir la presión arterial.
- Llamar a un número de teléfono en el Distrito Federal, aplicarle a la persona que conteste una encuesta sociodemográfica y preguntarle por cuál candidato votó en las últimas elecciones.

En principio, la ejecución de un experimento aleatorio produce una multitud de información. Sin embargo, sólo parte de esta información se registrará:

Resultados y espacio muestral

La parte de la información generada por un experimento aleatorio que el investigador decide registrar se denomina el *resultado* del experimento.
El conjunto de todos los posibles resultados, se conoce como el *espacio muestral* (o *conjunto de los resultados*), denotado por la mayúscula griega Ω.

Generalmente, son las teorías, las hipótesis y/o los intereses del investigador que definen qué parte de la información se registra para el resultado que se registra.

Ejemplos

- Al echar las tres monedas, sólo se registra, por ejemplo, el número de monedas que tienen "sol" arriba. No se interesa por la posición espacial de las tres monedas o el tiempo que duró hasta que ya no se muevan. En este caso, se puede definir el espacio muestral como:

$$\Omega = \{0, 1, 2, 3\}$$

- Para la lotería Melate, las primeras seis esferas que salgan indican los números principales (los "números naturales" en la terminología de Pronósticos del Estado); la séptima esfera da el número adicional. Para el resultado de la lotería, *no* importa en qué orden hayan salido las primeras seis esferas que salgan.

- En el examen de opción múltiple, se registra únicamente cuáles fueron las respuestas del estudiante en las 40 preguntas. No se registra la velocidad con la que contesta cada pregunta ni la expresión de su cara durante el examen. El espacio muestral Ω consiste de todas las posibles combinaciones de respuestas en las 40 preguntas. (Si hay 4 opciones de respuesta en cada pregunta, entonces el espacio muestral contiene 4^{40} elementos.)

Observaciones y variables

- Diferentes ejecuciones del experimento aleatorio resultan en datos de varios *objetos*, *unidades experimentales* u *observaciones*.
  Es posible que se observe el mismo resultado en diferentes observaciones.

- Para poder sistematizar los datos, se definen *variables*. Una variable asocia un valor a cada resultado en el espacio muestral.

Ejemplos

- El examen de opción múltiple se aplica a 800 estudiantes. En este caso, se considera a cada estudiante como una observación.
  Aplicando la clave de respuestas correctas, se define una variable X que asocia con cada estudiante/observación el número de preguntas en las que dio la respuesta correcta.

- Se administra el nuevo fármaco a un grupo de 60 pacientes (lo cual conlleva a 60 observaciones) y de los diagnósticos que se realizan antes y después de la administración se define una variable X que corresponde con la diferencia entre la presión arterial en ambos momentos para cada paciente.

- El sociólogo que investiga el comportamiento electoral llama a 500 personas y asigna a cada persona valores en diferentes variables: (a) el sexo del respondiente, (b) su religión, (c) su grado de estudios y (d) el nombre del partido por el cual votó en las últimas elecciones. Es decir, tendrá una base de datos con 500 observaciones y 4 variables.

En muchas ocasiones, los valores de una variable serán numéricos, aunque no siempre es así. Por ejemplo, los valores en la variable *Sexo* son, en principio, no numéricos. Por otro lado, siempre es posible convertir valores no númericos a valores númericos (por ejemplo, “masculino” $\mapsto$ 1, “femenino” $\mapsto$ 2).

Notación

Es fundamental distinguir entre la variable, por un lado, y los valores en la variable, por otro lado. Así, se puede decir que la variable *Sexo* asume dos posibles valores: masculino o femenino; o el *Resultado en el examen* asume 41 posibles valores (del conjunto $\{0, 1, \ldots, 40\}$).

Para que no se pierda de vista esta distinción, las variables siempre las anotaremos con letras mayúsculas (por ejemplo, X, Y, Z, X_1, X_2, etc.); un valor de una variable se anotará con una minúscula (como x, y, z, x_1, etc.).

En la segunda parte de este curso, cuando se inicie el estudio de la estadística inferencial, se indagará más en los conceptos mencionados anteriormente. Respecto a los conceptos que se acaban de introducir en la sección actual, es importante tener una idea, por lo menos intuitiva, de su significado antes de iniciar el estudio de la estadística descriptiva.

1.4 Niveles de medición y tipos de variables

Aunque los valores de una variable sean númericos, no todas las operaciones que se pueden hacer con estos números tienen sentido. Los siguientes ejemplos ilustran el problema del nivel de medición:

Ejemplos

Un sociólogo que investiga el comportamiento electoral de los votantes define la siguiente variable X en el conjunto de los resultados Ω:

- Si la persona votó en las últimas elecciones por el candidato del PRI le asigna un valor de 1
- Si la persona votó en las últimas elecciones por el candidato del PRD le asigna un valor de 2
- Si la persona votó en las últimas elecciones por el candidato del PAN le asigna un valor de 3
- Si la persona votó en las últimas elecciones por el candidato de la Nueva Alianza le asigna un valor de 4
- Si la persona votó en las últimas elecciones por otro candidato que los anteriormente mencionados le asigna un valor de 5

Las operaciones que se pueden aplicar a números incluyen: comparar la (des)igualdad (¿Qué significa que en una observación X asume el valor de 2 y en otra observación el valor de 1? ¿Se puede concluir algo a partir de que $1 \neq 2$ en este ejemplo?), evaluar el orden (¿Qué significa que $1 < 2$ en este ejemplo?), sumar o restar (¿Tiene sentido calcular la suma de los votos de PRI y PRD? ¿Qué significaría que $1 + 2 = 3$?), entre otros.

Como otro ejemplo, la siguiente anécdota, descrita por uno de los pioneros de la psicometría (F. M. Lord, 1953, "On the statistical treatment of football numbers", *The American Psychologist*, **8**, 181–194) relata sobre una discusión que en esencia va sobre el nivel de medición de una variable:

> *Un profesor ha tenido el privilegio de distribuir los dorsales a los jugadores de un equipo de rugby en su colegio. Sin embargo, los alumnos del primer año se quejan, porque han recibido "muchos números bajos". El profesor se defiende diciendo que los números son meramente etiquetas: se podrían cambiar por letras o símbolos cualesquiera. Los alumnos, que no están satisfechos con esta explicación del profesor, piden a un estadístico que haga una comparación formal; y, de hecho, éste afirma (después de aplicar una prueba t para dos muestras independientes [véase el Tema 12]) que los números de los alumnos del primer año son "significativamente" menores que los números de los alumnos de años mayores. El comentario que los dorsales son meramente etiquetas, ni lo considera "porque los números no saben de dónde vienen".*
>
> *En este ejemplo, observamos que los alumnos interpretan los dorsales como calificaciones: implican un orden entre los jugadores. El profesor, por otro lado, interpreta los dorsales como etiquetas para identificar a los jugadores y no considera las propiedades númericas de los dorsales. La clave, por supuesto, está en el papel del estadístico: ¿Puede calcular la media y la varianza de los dorsales? "Por supuesto," contesta, "acabo de hacerlo, ¿no?".*

Los ejemplos anteriores plantean la admisibilidad de las operaciones aritméticas a los valores númericos que uno asigna a las observaciones. Se habla del problema del *nivel de medición* o el *nivel de la escala*. Una escala se puede entender como un conjunto de números y las relaciones entre estos números que tienen una interpretación empírica. Los valores de una variable son números que pertenecen a una escala.

A continuación, se describen escalas de diferente nivel. El nivel de la escala depende directamente de las transformaciones que se pueden aplicar a los valores *sin que cambie la interpretación empírica*. Cuando hay menos libertad de transformación, más información contienen los números (y más tipos de relaciones y operaciones entre ellos se pueden interpretar de forma sensata).

Aunque el número de diferentes niveles de escala en principio es muy grande, en general se distinguen los cinco niveles que presentamos a continuación. Hacemos la presentación de nivel menor (es decir, más libertad de transformación y menos información) a nivel mayor (menos libertad de transformación y más información).

1.4.1 Escala nominal

Escala nominal

En una escala de medición *nominal*, los valores se pueden transformar a otros valores cualesquiera, tomando en cuenta la siguiente restricción:
Observaciones que tienen el mismo número asignado antes de la transformación deben seguir con el mismo número después de la transformación; observaciones con valores distintos deben seguir con valores distintos después de la transformación. Formalmente:

Para cualquier par de valores x y y, transformados a $x^\star$ y $y^\star$, se cumple que

$$x = y \iff x^\star = y^\star$$

En variables de nivel nominal, los números sólo funcionan como etiquetas: las observaciones que el investigador considera iguales reciben el mismo número; las observaciones que considera diferentes tienen números diferentes.

Ejemplo

El ejemplo anterior con los partidos políticos muestra que la variable asociada es de nivel nominal. Pues, no se pierde información al realizar una transformación como la que sigue:

		Valor anterior		Valor nuevo
Votó por:		X		$X^\star$
PRI	$\longrightarrow$	1	$\longrightarrow$	0
PRD	$\longrightarrow$	2	$\longrightarrow$	3.8
PAN	$\longrightarrow$	3	$\longrightarrow$	-1
Nueva Alianza	$\longrightarrow$	4	$\longrightarrow$	$\sqrt{2}$
Otro	$\longrightarrow$	5	$\longrightarrow$	π

La única información en los valores de X es que las personas con el mismo valor votaron por el mismo partido y personas con diferentes valores votaron por diferentes partidos. Esta información se ha preservado en $X^\star$. Nótese que si se hubiese asignado en la nueva variable un valor de 0 (en vez de -1) a los votantes del PAN, sí habría una perdida de información, ya que en este caso la nueva variable no distingue entre las personas que votaron por el PRI y el PAN.

Las variables de escala nominal se llaman variables *cualitativas*.

1.4.2 Escala ordinal

Una variable es de nivel ordinal si sus valores reflejan algún orden en las observaciones:

Escala ordinal

En una escala de medición *ordinal*, los valores se pueden transformar a otros valores cualesquiera, siempre y cuando la transformación no cambia el orden de las observaciones. Formalmente: Para cualquier par de valores x y y, transformados a $x^\star$ y $y^\star$, se cumple que

$$x < y \iff x^\star < y^\star.$$

En otras palabras, la transformación debe respetar el orden original. Matemáticamente, se dice que son permitidas las transformaciones monótonamente crecientes.

El hecho de que una variable es de tipo ordinal solo permite sacar conclusiones sobre el orden de las observaciones. Se puede concluir, por ejemplo, que el valor x es mayor que el valor y, pero no se permite interpretar la magnitud de la diferencia entre ambos valores.

Ejemplo

Una encuesta que sondea en un grupo de altos funcionarios el uso de somníferos pregunta a los participantes con qué frecuencia han utilizado en los últimos seis meses un fármaco con el fin de mejorar el sueño. La pregunta se contesta eligiendo una de las opciones "nunca", "alguna vez", "pocas veces", "frecuentemente", "diariamente".

Se trata de una variable de nivel ordinal. Considere la siguiente definición de las variables X y $X^{\star}$.

		Valor numérico X		Valor nuevo $X^{\star}$
Nunca	$\longrightarrow$	1	$\longrightarrow$	-4
Alguna vez	$\longrightarrow$	2	$\longrightarrow$	-3
Pocas veces	$\longrightarrow$	3	$\longrightarrow$	0
Frecuentemente	$\longrightarrow$	4	$\longrightarrow$	1.2
Diariamente	$\longrightarrow$	5	$\longrightarrow$	5

Puesto que X es una variable ordinal, se permite concluir que, si alguien contesta "frecuentemente" entonces ha utilizado somníferos más frecuentemente que si hubiese contestado "pocas veces". Por otro lado, la información en la escala no permite llegar a conclusiones sobre qué tan grande es la diferencia entre "pocas veces " y "frecuentemente", ni si la diferencia entre estos dos valores es igual a la diferencia entre "nunca" y "alguna vez". Si la escala es meramente ordinal, no tiene sentido interpretar la magnitud de las diferencias entre los valores. Por lo tanto, la variable $X^{\star}$, en la cual las diferencias entre los valores son diferentes pero el orden sigue siendo el mismo, implica una transformación permitida de X.

Las variables de tipo ordinal se llaman alternativamente *variables cuasi-cuantitativas*.

1.4.3 Escala de intervalo

Se dice que una variable tiene el nivel de escala de intervalo, si se concede importancia a las diferencias entre los valores de esta variable. Los valores de la variable no sólo reflejan un orden entre las observaciones, sino también nos informan sobre la magnitud de las diferencias entre ellas. Más preciso, en una variable de tipo intervalo, podemos interpretar de forma sensata las razones entre las diferencias.

Ejemplo

Un ejemplo conocido de valores que forman una escala de medida de intervalo se da por las escalas corrientes de temperatura. En México, se usa normalmente la escala de Celsius expresando la temperatura en grados centígrados. A partir de las propiedades de esta escala, no solo podemos decir que 20°C es menor que 30°C, sino también que la diferencia entre estos dos es igual a la diferencia entre 5°C y 15°C o que la diferencia entre 20°C y 30°C es dos veces la diferencia entre 8°C y 13°C, por ejemplo.

Escala de intervalo

En una escala de medición *intervalo*, únicamente se permiten transformaciones lineales, es decir, transformaciones de la forma

$$x^\star = ax + b$$

con la restricción adicional de que $a > 0$.

Una transformación de los valores en una escala de intervalo debe respetar las propiedades de esta escala: Las razones entre las diferencias deben ser iguales después de una transformación. Lo último implica efectivamente que las transformaciones permitidas son las funciones lineales con la pendiente estrictamente positiva. (Nótese que la restricción de que $a > 0$ es necesaria ya que, si $a < 0$, el orden de los valores se invierte y con $a = 0$ todos los valores se transforman al mismo valor). Cualquiera transformación lineal con $a > 0$ en una escala de intervalo guarantiza que la información relevante en esta escala se mantenga.

Ejemplo

En Estados Unidos, es más común el uso de la escala de Fahrenheit para expresar la temperatura. No obstante, los valores en grados Fahrenheit contienen la misma información que aquellos en grados centígrados. Las dos escalas son transformaciones lineales de la una a la otra. En particular, se dan por las siguientes ecuaciones:

$$F = \frac{9}{5}C + 32$$

$$C = \frac{5}{9}F + \frac{160}{9} = \frac{5}{9}F + 17.78$$

A partir de la siguiente tabla se puede ilustrar lo que quiere decir que “las razones entre las diferencias” son idénticas para ambas escalas de temperatura:

Temperatura en centígrados C		Temperatura en Fahrenheit F
5	$\longleftrightarrow$	41.0
8	$\longleftrightarrow$	46.4
13	$\longleftrightarrow$	55.4
15	$\longleftrightarrow$	59.0
20	$\longleftrightarrow$	68.0
30	$\longleftrightarrow$	86.0

En la escala centígrada, la diferencia $30°\text{C} - 20°\text{C} = 10°\text{C}$ es igual a la diferencia $15°\text{C} - 5°\text{C} = 10°\text{C}$, lo que también observamos en los valores correspondientes en la escala de Fahrenheit: $86°\text{F} - 68°\text{F} = 18°\text{F}$ es igual a $59°\text{F} - 41°\text{F} = 18°\text{F}$. De forma similar, la diferencia $30°\text{C} - 20°\text{C} = 10°\text{C}$ es el doble de la diferencia $13°\text{C} - 8°\text{C} = 5°\text{C}$, tal como $86°\text{F} - 68°\text{F} = 18°\text{F}$ es el doble de $55.4°\text{F} - 46.4°\text{F} = 9°\text{F}$.

Cabe señalar que en variables de tipo intervalo (tal como en variables de un nivel menor), el valor de cero no se puede interpretar como ausencia de algo. Es decir, el origen de la escala no refleja un cero absoluto (es la razón por la que se puede sumar una constante al transformar la variable).

Ejemplo

En las escalas de temperatura de Celsius y Fahrenheit no se puede interpretar cero grados como ausencia total de temperatura (o calor). En ambas escalas, existen valores negativos de tal forma que cero grados indica más calor que por ejemplo -5 grados. Además, el punto cero en ambas escalas difiere y se ha elegido de forma arbitraria (Celsius eligió para 0°C la temperatura de congelación de agua mientras Fahrenheit fijó el punto cero de su escala en la temperatura de congelación del cloruro amónico en agua). Nota que la escala de temperatura en Kelvin sí tienen un punto cero absoluto que indica ausencia total de calor, y por lo tanto, es de un nivel de medición mayor que las escalas de Celsius y Fahrenheit.

Una consecuencia de que el origen de la escala es arbitrario es que las razones entre los valores de la escala *no* deben ser interpretados. Por ejemplo, no se puede decir que 30°C es el doble de 15°C (86°F tampoco es el doble de 59°F). Entonces, en una escala de intervalo, las razones que se pueden interpretar son las razones entre diferencias de valores, no las razones entre los valores.

Se llama *variables cuantitativas* a variables que tienen un nivel de medida igual o mayor que intervalo.

1.4.4 Escala de razón

La escala de razón tiene en común con la escala de intervalo que la unidad de la escala se puede elegir de forma arbitraria. (Nótese que en el ejemplo del párrafo anterior, las unidades de ambas escalas de temperatura no son iguales: 1°F corresponde con $\frac{9}{5}$°C.) La diferencia es que en escalas de razón el origen tiene un significado especial: Se da el valor de cero a las observaciones que “no tienen nada de la característica que se está midiendo”. En otras palabras, para escalas de este tipo, sólo podemos elegir libremente la unidad de la escala.

Escala de razón

En una escala de *razón*, únicamente se permiten transformaciones multiplicativas, es decir, transformaciones de la forma:

$$x^\star = ax$$

con la restricción de que $a > 0$.

El valor de 0 no cambia por una transformación multiplicativa. La constante a fija la unidad de la escala. Si una variable tiene el nivel de razón, se permite interpretar las razones entre los valores.

Ejemplo

Un ejemplo de una escala de razón, son los números que indican longitud o distancia. Se puede medir en centimetros, metros, pulgadas o pies, pero en cualquier sistema se asigna el valor de 0 a un objeto que "no tiene longitud". En la escala de longitud, se permite interpretar la razón entre dos números, por ejemplo, se puede afirmar que 200 metros es el doble de 100 metros. Otros ejemplos de variables de nivel de razón incluyen variables que expresan peso (en gramos, kilogramos, onzas), tiempo/edad (en segundos, días, años), velocidad (kilómetros por hora, metros por segundo), fortuna (en pesos mexicanos, euros, dólares, libras) y temperatura en Kelvin.

1.4.5 Escala absoluta

Escala absoluta

Se considera una variable de nivel *absoluto* si no está permitida ninguna transformación.

En general, son absolutas las variables que indican números, como el número de personas que está presente en clase o el número de ovejas que el pastor cuenta en su rebaño.

Comentario final sobre las escalas de medición

Cabe señalar que no existen reglas formales para decidir sobre el nivel de medida de una escala o una variable. Como observamos en la anécdota con que empezamos esta sección, es cuestión de interpretación. Los alumnos consideraron el orden entre los dorsales y por lo tanto, para ellos, es una variable ordinal; por otro lado, según el profesor, no se debe interpretar el orden: los dorsales sólo sirven para distinguir entre los jugadores y, entonces, para él, es una variable nominal. En las ciencias sociales y de la salud, muchas veces se supone que las variables son de un nivel intervalo (básicamente porque el tipo de análisis estadísticos que se aplican a ellas, es para variables cuantitativas), mientras que autores más críticos no están dispuestos a creer que las diferencias entre los valores se pueden interpretar sensatamente.

Por ejemplo, en la medición de la señal BOLD (abreviación de *blood oxygenation level dependent*), la cual comúnmente se utiliza en estudios fMRI, hay consenso que valores más altos indican una actividad cerebral (relativamente) mayor, pero no es claro que estos valores conformen una escala de nivel intervalo. Sin embargo, estrictamente hablando, las técnicas comunes para el análisis de datos fMRI requieren que la variable sea de nivel intervalo.

Presentación de datos ilustrativos

En la siguiente página se presenta un conjunto de datos que forman una muestra de los datos de un estudio en Salud Pública. Se presenta información sobre 50 personas en las siguientes siete variables:

1. *Sexo*: M(asculino) o F(eminino);
2. *Edad*: Edad en años;
3. *Educ*: Nivel educativo de la persona, en cuatro categorías: (a) Primaria, (b) Secundaria, (c) Bachillerato (incluye Preparatoria y Estudios Técnicos) y (d) Universitario;
4. *IMC*: Índice de masa corporal (en kg/m^2);
5. *GS*: Grupo sanguíneo en cuatro categorías (según la clasificación ABO): (a) Grupo O, (b) Grupo A, (c) Grupo B, (d) Grupo AB;
6. *PA*: Presión arterial, donde los dos números indican la presión arterial (en mmHg) sistólica y diastólica, respectivamente;
7. *Fumador*: Esta variable clasifica cada persona según el número de cigarros que fuma por día (al promedio): (a) No fuma, (b) Fumador bajo (entre 1 y 10 cigarros/día) y (c) Fumador intensivo (más de 10 cigarros/día).

Este conjunto de datos se utilizará en los capítulos siguientes para ilustrar los conceptos introducidos y los diferentes tipos de análisis.

No.	*Sexo*	*Edad*	*Educ*	*IMC*	*GS*	*PA*	*Fumador*	No.	*Sexo*	*Edad*	*Educ*	*IMC*	*GS*	*PA*	*Fumador*
1	F	58	Secundaria	21	O	109/82	No fuma	26	M	55	Secundaria	22	A	116/75	No fuma
2	F	55	Bachillerato	25	O	129/77	No fuma	27	F	55	Secundaria	28	A	138/93	Intensivo
3	F	38	Bachillerato	27	A	128/86	Bajo	28	F	54	Secundaria	25	B	127/84	No fuma
4	F	57	Secundaria	20	A	112/95	No fuma	29	M	47	Universitario	22	A	144/88	No fuma
5	M	48	Bachillerato	33	A	129/86	Bajo	30	F	43	Bachillerato	24	O	145/103	Intensivo
6	M	53	Secundaria	23	O	121/90	Intensivo	31	M	38	Bachillerato	34	A	147/100	Intensivo
7	F	44	Secundaria	24	O	116/73	Bajo	32	F	51	Universitario	29	A	116/76	Bajo
8	M	46	Secundaria	28	A	177/100	No fuma	33	M	35	Universitario	26	AB	144/88	No fuma
9	F	32	Bachillerato	23	A	128/68	No fuma	34	F	44	Universitario	19	A	120/78	No fuma
10	F	35	Universitario	27	O	122/77	No fuma	35	F	47	Bachillerato	24	O	132/94	Intensivo
11	M	34	Bachillerato	22	B	164/86	No fuma	36	M	52	Universitario	24	O	112/73	No fuma
12	F	29	Bachillerato	26	O	114/91	No fuma	37	M	33	Universitario	22	O	127/73	No fuma
13	M	49	Secundaria	29	O	133/80	Intensivo	38	F	41	Universitario	20	A	114/79	No fuma
14	F	41	Bachillerato	24	O	111/99	No fuma	39	F	56	Bachillerato	22	O	138/97	Bajo
15	F	57	Secundaria	22	O	172/99	No fuma	40	M	55	Secundaria	42	B	164/123	Intensivo
16	M	45	Bachillerato	27	B	148/94	No fuma	41	M	59	Bachillerato	23	O	127/88	No fuma
17	M	34	Universitario	20	B	128/85	No fuma	42	F	45	Universitario	22	A	142/91	No fuma
18	F	55	Primaria	25	B	116/79	Bajo	43	M	46	Secundaria	28	A	132/93	No fuma
19	F	46	Bachillerato	23	B	107/68	Bajo	44	F	37	Universitario	19	O	113/83	No fuma
20	F	30	Secundaria	26	B	148/77	Intensivo	45	F	55	Primaria	24	B	126/79	No fuma
21	M	38	Primaria	30	O	114/90	Intensivo	46	M	39	Bachillerato	25	O	112/75	No fuma
22	F	55	Secundaria	24	AB	134/82	No fuma	47	M	58	Secundaria	23	O	129/73	No fuma
23	M	57	Primaria	27	A	144/80	No fuma	48	F	36	Bachillerato	27	A	108/73	Bajo
24	M	58	Secundaria	25	A	159/104	Bajo	49	F	29	Secundaria	26	A	114/75	No fuma
25	M	47	Universitario	23	O	129/79	No fuma	50	F	49	Secundaria	26	AB	121/69	No fuma

Problemas

Algunos de los problemas más difíciles o los que requieren conocimiento especial se han marcado con *.

1. Hallar:

- $\displaystyle\sum_{k=1}^{6} 2k =$
- $\displaystyle\sum_{i=12}^{279} 3^2 =$
- $\displaystyle\sum_{j=1}^{1} 5(3j+9j+5) =$
- $\displaystyle\sum_{j=1}^{2}\sum_{k=3}^{5}(k-j) =$
- $\displaystyle\sum_{i=0}^{4}\sum_{j=i-1}^{i+1} j =$
- $\displaystyle\sum_{i=2}^{6}\sum_{j=1}^{4}\sum_{k=1}^{5}(i+j+k) =$
- $\displaystyle\prod_{i=0}^{9} i =$
- $\displaystyle\prod_{i=2}^{6} i^2 =$

2. Demuestra que las siguientes propiedades para productos son ciertas para cualquier $c \in \mathbb{R}$ y cualquier $n \in \mathbb{N}$ mayor que 0:

- $\displaystyle\prod_{i=1}^{n}(x_i y_i) = \left(\prod_{i=1}^{n} x_i\right)\left(\prod_{i=1}^{n} y_i\right)$
- $\displaystyle\prod_{i=1}^{n} c = c^n$
- $\displaystyle\prod_{i=1}^{n}(c x_i) = c^n \prod_{i=1}^{n} x_i$

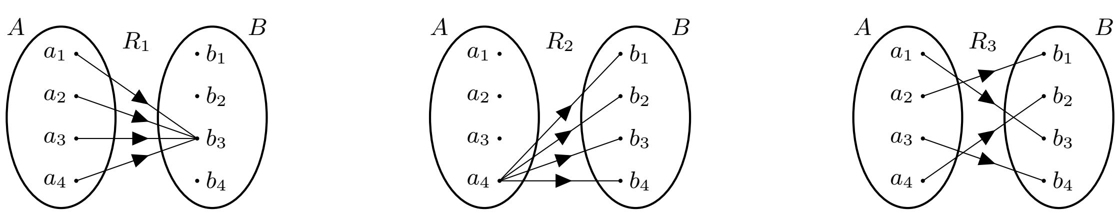
3. Deriva la siguiente igualdad. Para cualquier $n \in \mathbb{N}$:

$$\left(\sum_{i=1}^{n} x_i\right)^2 = \sum_{i=1}^{n} x_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_i x_j)$$

*4. Comprueba la siguiente igualdad. Para cualquier $n \in \mathbb{N}$:

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

5. ¿Cuáles de las siguientes relaciones son funciones? ¿Cuáles son biyecciones?



6. Si existe, da un ejemplo de una...
   - ...función cuyo dominio sea un conjunto infinito y su contradominio un conjunto finito.
   - ...biyección entre un conjunto infinito y un conjunto finito.
   - ...función cuyo dominio sea un conjunto finito y su contradominio el intervalo $[0, 1]$.
   - ...función cuyo dominio sea un conjunto finito y su rango un conjunto infinito.

   Si no existe la función o biyección con las características dadas, argumenta por qué no.

7. La forma general de una función *cuadrática* de $\mathbb{R}$ a $\mathbb{R}$ es:
$$f(x) = ax^2 + bx + c,$$
   donde a, b y c son números reales.
   ¿El conjunto de las funciones lineales es un subconjunto del conjunto de las funciones cuadráticas?

8. ¿Verdadero o falso?
   - Si $C = (A \setminus B)$ y $D = (B \setminus A)$, C y D son disjuntos.
   - Para dos conjuntos A y B siempre se cumple: $A \cap B \;=\; (A \cup B) \;\setminus\; [(A \setminus B) \cup (B \setminus A)]$.
   - El número cardinal de una función es igual al número cardinal de su dominio.
   - Una combinación lineal es una biyección de $\mathbb{R}^n$ a $\mathbb{R}$.

9. ¿Cuál es el valor de las siguientes integrales definidas? (No utilices cálculo diferencial o integral para obtener el resultado, sino utiliza la interpretación de la integral definida en términos de un área debajo de una curva.)

   (a) $\displaystyle\int_0^2 x\,dx =$

   (b) $\displaystyle\int_{10}^{100} 2\,dx =$

   (c) $\displaystyle\int_{-2}^{+2} |x|\,dx =$

   (d) $\displaystyle\int_0^3 (5 - x)\,dx =$

   *(e) $\displaystyle\int_{-1}^{+1} \sqrt{1 - x^2}\,dx =$

Parte I

Estadística descriptiva

Tema 2
Funciones de frecuencia y proporción (una variable)

Índice

En este tema introducimos los conceptos de (la función de) frecuencia y proporción, junto con métodos que permiten "visualizar" los datos, es decir, diferentes métodos para una representación gráfica de estas funciones. Hacemos una distinción entre variables cualitativas (Sección 2.1) y variables cuantitativas (Sección 2.2), ya que, aunque en muchos casos las definiciones y propiedades son similares, en otros es importante distinguir entre los dos tipos de variables, por ejemplo, cuando se introduce la frecuencia y proporción acumulada (solo para variables cuantitativas) y cuando se habla de algunas representaciones gráficas (que son más apropiadas para un tipo de variables que para el otro). En la última sección (Sección 2.3), que sirve como trampolín a los temas siguientes, se discuten las características más relevantes de las funciones de frecuencia y proporción.

2.1 Variables cualitativas

2.1.1 Función de frecuencia

Para cada valor x de una variable X cualitativa, se puede contar en cuántas observaciones o unidades experimentales la variable asume este valor. El resultado de este conteo es la frecuencia de x en esta variable.

Definición

Para (a) una variable X, que asume m distintos valores x_1, x_2, $x_3, \dots,$ x_m
y (b) un conjunto de n observaciones con valores en la variable X,

la *función de frecuencia* asocia cada valor x_j ($j = 1, \dots, m$) con un número natural que indica el número de observaciones que tienen el valor x_j en la variable X.

Notación: $\mathit{frec}(x_j)$ o $\mathit{frec}_X(x_j)$

En la notación $\mathit{frec}(x_j)$ está implícito que se trata de la función de frecuencia asociada con la variable X; cuando se quiere indicar explícitamente la variable, se escribe: $\mathit{frec}_X(x)$.

Ejemplo

Consideremos la variable cualitativa *GS* (*Grupo Sanguíneo*) de los datos de la página 34. La frecuencia del grupo sanguíneo O es 20, lo que se escribe $\mathit{frec}(\mathrm{O}) = 20$ (o $\mathit{frec}_{GS}(\mathrm{O}) = 20$). De forma similar, se obtiene $\mathit{frec}(\mathrm{A}) = 18$, $\mathit{frec}(\mathrm{B}) = 9$ y $\mathit{frec}(\mathrm{AB}) = 3$.

Esta misma información comúnmente se organiza en una tabla que presenta todos los valores de la función de frecuencia:

GS	frec_{GS}
O	20
A	18
B	9
AB	3
Total	50

Algunas propiedades de la función de frecuencia

- $\sum_{j=1}^{m} \mathit{frec}(x_j) = n$
- Para cualquier valor x_j: $0 \leqslant \mathit{frec}(x_j) \leqslant n$

Ejemplo

La variable *GS* (*Grupo Sanguíneo*) en los datos de la página 34 asume $m = 4$ valores diferentes: $x_1 = \mathrm{O}$, $x_2 = \mathrm{A}$, $x_3 = \mathrm{B}$ y $x_4 = \mathrm{AB}$. Se verifica que:

- $\mathit{frec}(x_1) + \mathit{frec}(x_2) + \mathit{frec}(x_3) + \mathit{frec}(x_4) = 20 + 18 + 9 + 3 = 50$,
  el número n de observaciones en el conjunto de datos.
- para cada uno de los valores x_j, $0 \leqslant \mathit{frec}(x_j) \leqslant 50$.

2.1.2 Función de proporción

Además de la frecuencia, se puede calcular la *proporción* de cada valor x de X.

Definición

Para (a) una variable X, que asume m distintos valores x_1, x_2, $x_3, \ldots,$ x_m
y (b) un conjunto de n observaciones con valores en la variable X,

la *función de proporción* asocia cada valor x_j con un número del intervalo $[0, 1]$ que indica la proporción de observaciones que tienen el valor x_j en la variable X.

Notación: $p(x_j)$ o $p_X(x_j)$

Para todos los valores x_j, la proporción se obtiene a partir de la frecuencia:

$$p(x_j) = \frac{\mathit{frec}(x_j)}{n}$$

Algunas propiedades de la función de proporción

De la definición anterior y las propiedades de la función de frecuencia, sigue directamente que:

- $\displaystyle\sum_{j=1}^{m} p(x_j) = 1$
- Para cualquier valor x_j: $0 \leqslant p(x_j) \leqslant 1$

Ejemplo

Se puede extender la tabla de frecuencias para la variable GS, para incluir la función de proporción:

j	x_j	$\mathit{frec}(x_j)$	$p(x_j)$
1	O	20	.40
2	A	18	.36
3	B	9	.18
4	AB	3	.06
Total		50	1.00

Es fácil verificar que, efectivamente,

- $\displaystyle\sum_{j=1}^{m} p(x_j) = p(x_1) + p(x_2) + p(x_3) + p(x_4) = .40 + .36 + .18 + .06 = 1.00$
- para cada uno de los valores x_j, $0 \leqslant p(x_j) \leqslant 1$

Notas

- Frecuentemente, se utiliza porcentajes en vez de proporciones. Por definición, para cualquier número del intervalo $[0, 1]$, se tiene:

$$p \;=\; 100p\,\%$$

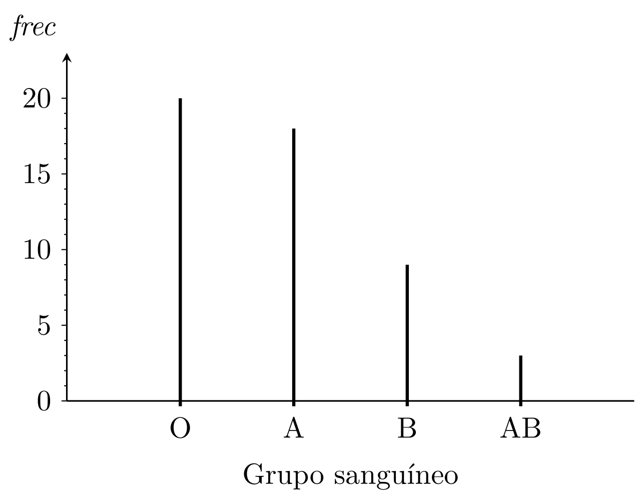
- En vez de hablar de la función de frecuencia o proporción de una variable, se habla de la *distribución de frecuencia o proporción*. Estos términos se utilizan como sinónimos: las funciones de frecuencia y proporción nos informan sobre cómo los valores de la variable "se distribuyen" en la muestra observada.

2.1.3 Representación gráfica de la función de frecuencia o proporción

Existe una amplia gama de herramientas para visualizar la función de frecuencia o proporción de una variable en una gráfica. En esta sección se presentan algunas que son adecuadas para variables cualitativas.

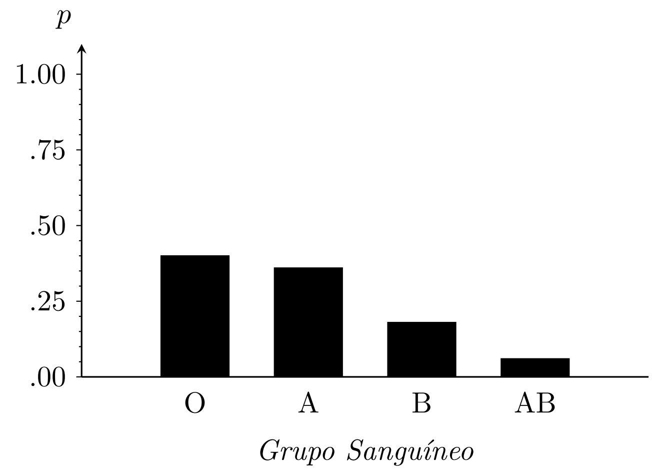
Un *diagrama de barras* tiene en la abscisa (el eje horizontal) los valores de la variable X y en la ordenada (el eje vertical) los valores de la función: $\mathit{frec}(x)$, o bien $p(x)$. Al respecto, nótese que el orden de los valores de X en la abscisa es arbitrario en principio, ya que X es una variable de nivel nominal, lo que implica que el orden de los valores no se puede interpretar sensatamente. La figura abajo presenta un diagrama de barras para la función de frecuencia de la variable GS en los datos de la página 34.

Diagrama de barras para la función de frecuencia de *Grupo Sanguíneo*



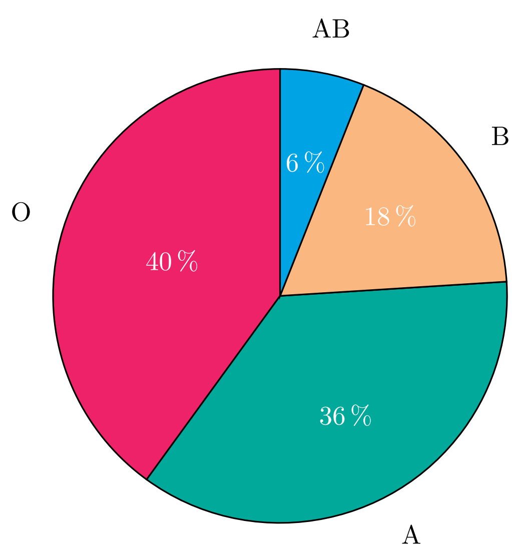
Una representación muy similar es el *diagrama de rectángulos*. En este tipo de gráfica se han sustituido las líneas por rectángulos de igual anchura, de tal forma que la superficie de los rectángulos de los diferentes valores x es directamente proporcional a $\mathit{frec}(x)$ o $p(x)$. Entre los rectángulos se deja un espacio para mostrar que la variable es cualitativa. La siguiente figura muestra un diagrama de rectángulos para la función de proporción de la variable Grupo Sanguíneo.

Diagrama de rectángulos para la función de proporción de *Grupo Sanguíneo*



Por último, se presenta un tipo de diagrama que es muy común para representar la función de proporción de variables cualitativas: el *diagrama de sectores* (también conocido como *gráfica de pastel*). Un diagrama de sectores para nuestro ejemplo se presenta a continuación.

Diagrama de sectores para la función de proporción de *Grupo Sanguíneo*



2.2 Variables cuantitativas

2.2.1 Funciones de frecuencia y proporción

También para variables cuantitativas, se puede considerar el número y la proporción de veces que cualquier valor ocurra y se derivan las funciones de frecuencia y proporción correspondientes. La definición y las propiedades de las funciones de frecuencia y proporción para variables cuantitativas son *idénticas* a las mismas para variables cualitativas.

Función de frecuencia y proporción para la variable *IMC*

La siguiente tabla presenta la función de frecuencia y proporción de la variable *IMC* (*Índice de Masa Corporal*) de los datos de la página 34.

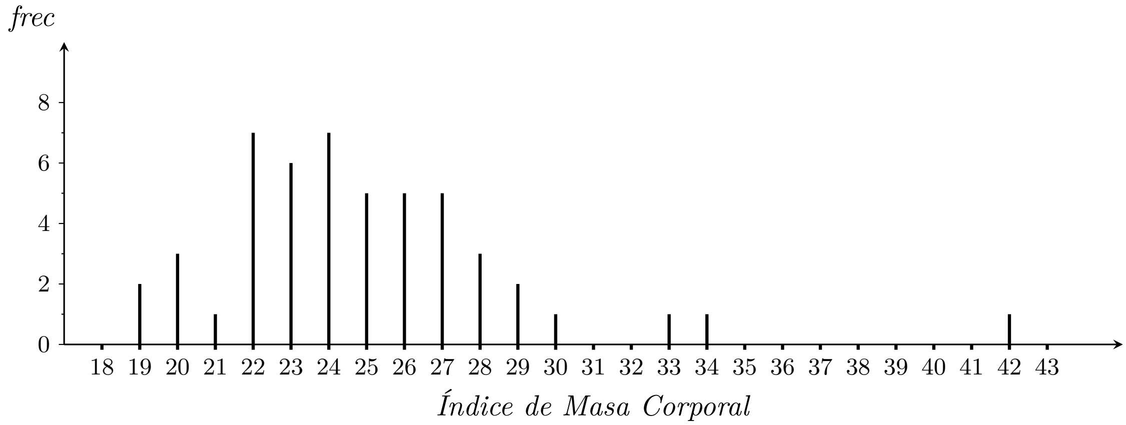
j	x_j	$frec(x_j)$	$p(x_j)$
1	19	2	.04
2	20	3	.06
3	21	1	.02
4	22	7	.14
5	23	6	.12
6	24	7	.14
7	25	5	.10
8	26	5	.10
9	27	5	.10
10	28	3	.06
11	29	2	.04
12	30	1	.02
13	33	1	.02
14	34	1	.02
15	42	1	.02
Total		50	1.00

2.2.2 Representación gráfica de la función de frecuencia o proporción

Algunos de los métodos para representar gráficamente las funciones de frecuencia y proporción de variables cuantitativas son iguales o similares a los mismos para variables cualitativas, pero también existen métodos que son específicos para variables cuantitativas.

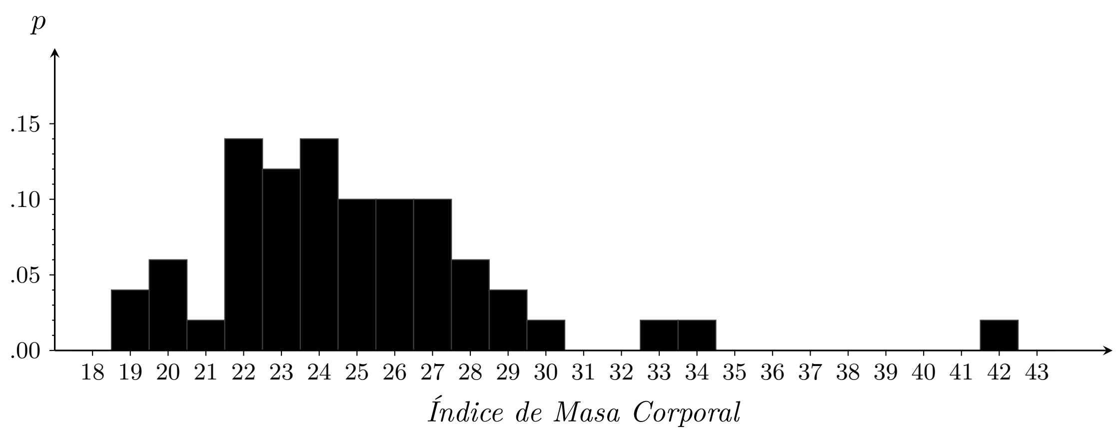
El *diagrama de barras* se construye de forma igual que para variables cualitativas, con el entendimiento de que los valores de la variable cuantitativa sí tienen un orden, el cual se debe respetar al posicionar los distintos valores en la abscisa. A continuación, se presenta el diagrama de barras para la variable *IMC*.

Diagrama de barras para la función de frecuencia de *Índice de Masa Corporal*

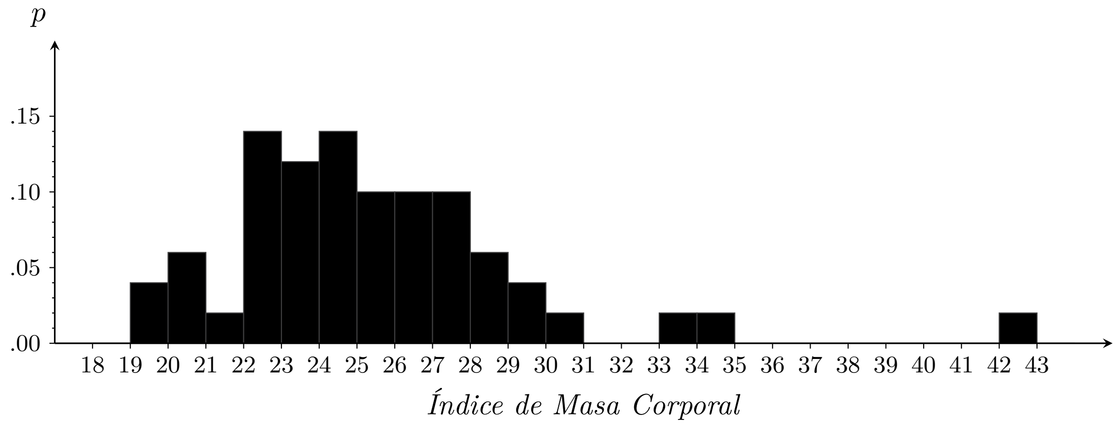


Una representación alternativa es el *histograma*. El histograma es muy similar al diagrama de rectángulos para variables cualitativas. En el histograma, dado que se aplica a variables cuantitativas, no se deja un espacio entre los rectángulos de los valores sucesivos, es decir, es un diagrama con rectángulos conectados. Nótese que el área de cada rectángulo vuelve a ser directamente proporcional a $frec(x)$ resp. $p(x)$. La siguiente figura muestra el histograma de la función de proporción para *Índice de Masa Corporal* de los datos de la página 34.

Histograma para la función de frecuencia de *Índice de Masa Corporal*



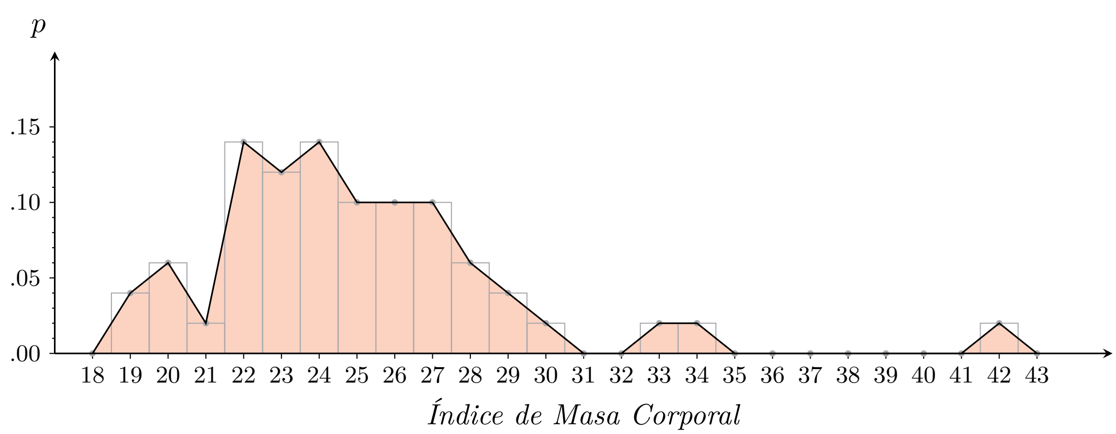
Nótese que el centro de cada rectángulo coincide con la posición exacta del valor correspondiente. Por ejemplo, la base del rectangúlo para el valor 19 inicia en el punto 18.5 y termina en el punto 19.5. Una forma alternativa para posicionar los rectángulos podría ser como en la siguiente gráfica:



La posición correcta de los rectángulos depende de cómo se han obtenido los valores de la variable representada. El índice de masa corporal, cuando se calcula a partir del peso y la talla, típicamente lleva a números decimales. Si estos números se redondean al número entero más cercano (tal que tanto 19.7 como 20.3 se convierten en 20), entonces es más adecuado el posicionamiento de los rectángulos en el primer histograma. Si, por otro lado, los números se redondean, simplemente eliminando la parte decimal (tal que 19.7 se convierte a 19 y 20.3 se convierte a 20), entonces es mejor posicionar los rectángulos como en el segundo histograma.

Otra alternativa para la representación gráfica de una variable cuantitativa es el polígono de frecuencia o proporción. El polígono se obtiene al unir los puntos medios de la parte superior de los rectángulos del histograma. El siguiente ejemplo ilustra la idea.

Polígono de proporción para el *Índice de Masa Corporal*



Se puede interpretar la representación polígona como un histograma que se obtendría si se hubiesen observado los valores de la variable como números decimales sin redondear. En este ejemplo, se ha registrado, como resultado del experimento aleatorio, el *IMC* como número entero, por lo cual en los datos no observamos valores con decimales. Sin embargo, teóricamente, la variable *IMC* puede tener un número infinito de valores entre los números enteros sucesivos. En este caso, el histograma tendría rectángulos infinitamente angostos y la altura de los rectángulos vecinos diferiría muy poco.

Esta situación, en la cual se considera un número infinito de valores para una variable, excede la pura descripción de los datos en la muestra y hace referencia a la estadística inferencial.

Si se ha observado un grán número de valores distintos en una variable, no es muy manejable hacer una tabla de las frecuencias o proporciones tal como se presentó en el ejemplo de la página 44. En este caso, se suele representar los datos por un *diagrama de tallo y hojas*. En dicho diagrama los valores se agrupan según la cifra de las decenas, es decir, se consideran los grupos de valores que pertenecen a la misma decena, como 0–9, 10–19, 20–29, etc. Dichos grupos constituyen el "tallo", mientras que las "hojas" corresponden con las cifras de las unidades de los valores en las diferentes observaciones. El siguiente ejemplo representa la función de frecuencia de la variable *Presión Arterial Sistólica* en los datos de la página 34 a través de un diagrama de tallo y hojas.

Diagrama de tallo y hojas para la *Presión Arterial Sistólica*

Tallo	Hojas
10	789
11	1222344446666
12	011267778889999
13	223488
14	24445788
15	9
16	44
17	27

En esta representación gráfica, se observa de inmediato que la mayor parte de las observaciones tienen un valor para *Presión Arterial Sistólica* entre 100 y 149 mmHg.

Valores atípicos

Volvamos un momento a las representaciones gráficas de la función de frecuencia y proporción para la variable *Índice de Masa Corporal* en los datos de la página 34. El lector habrá notado que estas gráficas muestran que hay una observación con un valor (42) en la variable *IMC* que difiere mucho de los valores de las demás observaciones. A tal observación se la denomina *valor atípico* (en inglés: *outlier*). La detección de valores atípicos es un paso importante en el análisis de los datos. Los valores atípicos pueden suceder por una multitud de factores (un error durante la captura de los datos, un error en la medición, un error del participante al rellenar el cuestionario, un evento raro, etc.) En algunos casos, pueden interferir en el análisis de los datos y llevar a un estadístico poco atento a conclusiones incorrectas.

2.2.3 Funciones de frecuencia y proporción acumulada

Además de la frecuencia y la proporción, también se puede calcular para cada valor de una variable cuantitativa la *frecuencia acumulada* y la *proporción acumulada*.

Definición frecuencia acumulada

Para (a) una variable X, que asume m distintos valores x_1, x_2, $x_3, \ldots,$ x_m
y (b) un conjunto de n observaciones con valores en la variable X

la *función de frecuencia acumulada* asocia cada valor x_j $(j = 1, \ldots, m)$ con un número natural que indica el número de observaciones que tienen *el valor x_j o un valor menor* en la variable X.

Notación: $\mathit{cfrec}(x_j)$ o $\mathit{cfrec}_X(x_j)$

Suponiendo que $x_1 < x_2 < x_3 < \ldots < x_m$, entonces se calcula la frecuencia acumulada para todos los x_j:

$$\mathit{cfrec}(x_j) = \sum_{k=1}^{j} \mathit{frec}(x_k)$$

Definición proporción acumulada

Para (a) una variable X, que asume m distintos valores x_1, x_2, $x_3, \ldots,$ x_m
y (b) un conjunto de n observaciones con valores en la variable X,

la *función de proporción acumulada* asocia cada valor x_j $(j = 1, \ldots, m)$ con un número del intervalo $[0, 1]$ que indica la proporción de observaciones que tienen *el valor x_j o un valor menor* en la variable X.

Notación: $F(x_j)$ o $F_X(x_j)$

La proporción acumulada se puede calcular:

- a partir de las proporciones:
  Suponiendo que $x_1 < x_2 < x_3 < \ldots < x_m$, entonces se cumple para todos los x_j:

$$F(x_j) = \sum_{k=1}^{j} p(x_k)$$

- a partir de la frecuencia acumulada, ya que, para todos los x_j:

$$F(x_j) = \frac{\mathit{cfrec}(x_j)}{n}$$

Función de frecuencia y proporción acumulada para la variable *IMC*

j	x_j	$frec(x_j)$	$p(x_j)$	$cfrec(x_j)$	$F(x_j)$
1	19	2	.04	2	.04
2	20	3	.06	5	.10
3	21	1	.02	6	.12
4	22	7	.14	13	.26
5	23	6	.12	19	.38
6	24	7	.14	26	.52
7	25	5	.10	31	.62
8	26	5	.10	36	.72
9	27	5	.10	41	.82
10	28	3	.06	44	.88
11	29	2	.04	46	.92
12	30	1	.02	47	.94
13	33	1	.02	48	.96
14	34	1	.02	49	.98
15	42	1	.02	50	1.00

Frecuencia y proporción acumulada de valores no observados

Consideremos un valor x que *no* se ha observado en la variable X. Entonces:

- $frec_X(x) = 0$
- $p_X(x) = 0$
- $cfrec_X(x) = cfrec_X(x_j)$
- $F_X(x) = F_X(x_j)$,

donde x_j es el valor *más alto menor que* x que sí se ha observado.

En el ejemplo anterior sobre la variable *IMC*, la proporción acumulada para el valor 32 (no observado) es:

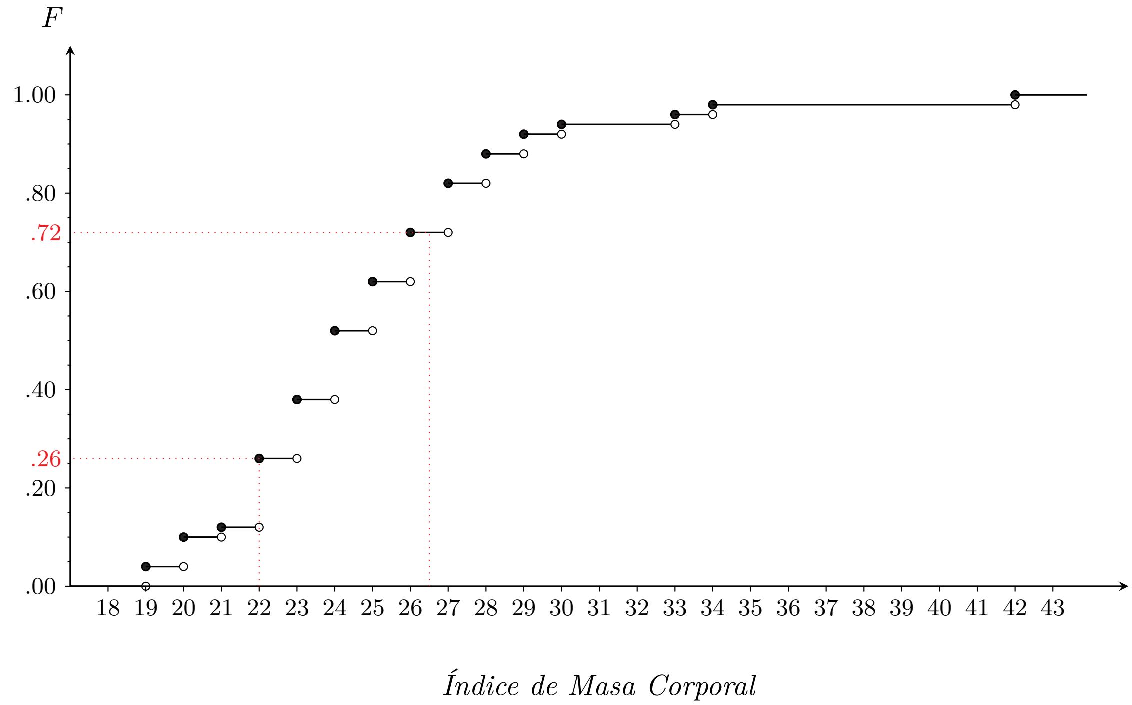
$$F(32) = F(30),$$

ya que 30 es el valor observado que simultáneamente es (a) menor que x y (b) el más alto de todos los valores observados que cumple esta condición.

2.2.4 Representación gráfica de la función de frecuencia o proporción acumulada

El siguiente ejemplo muestra cómo las funciones de frecuencia y proporción acumulada se representan gráficamente.

Representación gráfica de la función de proporción acumulada



La gráfica representa la función F para la variable *IMC* de los datos de la página 34. Cabe señalar que en este tipo de gráficas, se usa un círculo lleno para indicar que el intervalo está cerrado; el valor correspondiente al extremo izquierdo de cada intervalo se incluye en el intervalo. Por otro lado, se usa un círculo abierto para indicar un intervalo abierto; el valor al extremo derecho de cada intervalo no pertenece al intervalo. Así, por ejemplo, se puede afirmar que $F(22) = .26$ (el valor 22 no pertenece al intervalo a la altura de .12, sino es el primer valor del intervalo que está a la altura de .26). Similarmente, se lee que $F(26.5) = 0.72$.

Nótese que F es una función escalonada ascendente que va de 0.00 a 1.00. De forma similar, *cfrec* es una función escalonada ascendente que va de 0 a n.

2.2.5 Cuantiles

La definición de un cuantil se apoya en la función de proporción acumulada F.

Cuantil de orden r

Para cualquier r, $0 < r < 1$, el *cuantil de orden* r se define considerando dos posibles situaciones:

Caso 1: Existe (al menos) un valor x, para el cual $F(x) = r$
Para este caso, *todos* los valores x que cumplen $F(x) = r$ son un cuantil de orden r.

Si se quiere un valor único, se calcula:

$$\frac{x_{\text{min}} + x_{\text{max}}}{2},$$

donde x_{min} es el valor x mínimo que cumple $F(x) = r$ y x_{max} es el valor x mínimo que cumple $F(x) > r$.

Caso 2: No existe un valor x, para el cual $F(x) = r$
En este caso, el cuantil de orden r es el valor x más pequeño que cumple $F(x) > r$.

Notación: x_r

Ejemplo

- Busquemos el cuantil de orden .94 de la variable *IMC* en los datos de la página 34, es decir $x_{.94}$. Al consultar la tabla de la proporción acumulada, se observa que existe un valor x para *IMC*, para el cual $F(x) = .94$, a saber $x = 30$. Por lo tanto, se presenta el caso 1 de la definición anterior. Según la definición todos los valores x para los cuales $F(x) = .94$ se consideran un cuantil $x_{.94}$. Al respecto, recuérdese que no solo $F(30) = .94$ sino que, para cualquier x, que cumple $30 \leqslant x < 33$, $F(x) = .94$.
Si queremos un valor único, hay que hallar x_{min} y x_{max}; $x_{\text{min}} = 30$, ya que 30 es el valor mínimo que cumple $F(x) = .94$ (para $x < 30$, por ejemplo $x = 29.999$, $F(x) < .94$); $x_{\text{max}} = 33$, puesto que 33 es el primer valor x para el cual $F(x) > .94$ (para $x < 33$, por ejemplo $x = 32.999$, $F(x) \leqslant .94$). Entonces el valor único para $x_{.94}$ es:

$$x_{.94} = \frac{30 + 33}{2} = 31.5$$

- Busquemos ahora el cuantil de orden .50 de la misma variable. *No* existe un valor x para el cual $F(x) = .50$ (en la tabla y en la gráfica de la función acumulada, se observa un salto de .38 a .52 en el valor 24). Entonces, se aplica el caso 2 de la definición arriba: el cuantil $x_{.50}$ es igual al primer valor x que satisface $F(x) > .50$. Este valor es 24 ya que para $x = 24$, $F(x) = .52 > .50$, y para cualquier x menor que 24 (por ejemplo $x = 23.999$), $F(x) \leqslant .38 \not> .50$.

Interpretación de un cuantil de orden r

x_r es el valor más bajo de la variable X que por lo menos el $100r\,\%$ de las observaciones en la muestra no supera.

- Por ejemplo, si en el ejemplo anterior se dice que 31.5 es el cuantil de orden .94, se puede decir que el 94 % de las personas en la muestra tiene el *IMC* menor que o igual a 31.5.
- Hallamos que 24 es el cuantil de orden .50 de la variable *IMC* se puede concluir que el 50 % o más de las 50 personas tiene el *IMC* menor que o igual a 24. De hecho, de la función de proporción acumulada se deriva que el 52 % de la muestra no supera el valor 24. (Nótese que no existe ningún valor que exactamente el 50 % de la muestra no supera; 24 es el valor más bajo para el cual "50 % o más de la muestra no lo supera".)

Cuantiles especiales

En la práctica, se usan casi exclusivamente los siguientes cuantiles especiales:

- *Percentiles* El percentil v de una variable X, donde $v \in \{1, 2, \ldots, 99\}$, se define como el cuantil x_r con $r = \frac{v}{100}$. Así, por ejemplo, el percentil 27 corresponde con el cuantil $x_{.27}$. Una manera análoga para entender los percentiles es: Si $100\,r$ es un número entero, entonces el cuantil x_r es el percentil $100\,r$. Percentil v se abrevia como Pc_v.
- *Deciles* El decil v de una variable X, donde $v \in \{1, 2, \ldots, 9\}$, se define como el cuantil x_r con $r = \frac{v}{10}$. El decil 3, por ejemplo, corresponde con el cuantil $x_{.30}$. La abreviación del decil v es D_v.
- *Cuartiles* El cuartil v de una variable X, donde $v \in \{1, 2, 3\}$, se define como el cuantil x_r con $r = \frac{v}{4}$. Hay tres cuartiles, que corresponden con los cuantiles $x_{.25}$, $x_{.50}$ y $x_{.75}$, respectivamente y que se denotan como Q_1, Q_2 y Q_3.

El cuantil $x_{.50}$ (que corresponde con percentil 50, decil 5 y cuartil 2) ocupa un lugar importante en la estadística y se llama la *mediana*.

Para el cálculo práctico de un cuantil (y, entonces, de los percentiles, los deciles, los cuartiles y la mediana), se puede partir de (a) la tabla de la función F de proporción acumulada o de (b) una gráfica de la misma.

Cálculo de cuantiles a través de la tabla de F

Al calcular el cuantil x_r con la tabla de la función F, se verifica si el número r está presente en la columna de $F(x)$ o no.

- Si el número r está presente (Caso 1 de la definición), el valor x correspondiente y todos los valores hasta, pero no incluido, el siguiente valor, son cuantiles x_r. Si se quiere un x_r único, se calcula el medio entre el valor para el cual $F(x) = r$ y el valor siguiente.
- Si el número r *no* se encuentra en la tabla (Caso 2), el cuantil x_r es igual al primer valor x en la tabla para el cual $F(x)$ es mayor que r.

Ejemplo

- Busquemos el Percentil 26 para la variable *IMC*. La siguiente tabla presenta la parte relevante de la tabla de proporción acumulada.

j	x_j	$frec(x_j)$	$p(x_j)$	$cfrec(x_j)$	$F(x_j)$
⋮	⋮	⋮	⋮	⋮	⋮
3	21	1	.02	6	.12
4	22	7	.14	13	.26
5	23	6	.12	19	.38
6	24	7	.14	26	.52
⋮	⋮	⋮	⋮	⋮	⋮

Puesto que el Percentil 26 corresponde con el $x_{.26}$ y que el valor .26 se encuentra en la tabla de la proporción acumulada, se buscan los valores en la misma fila y en la fila siguiente, 22 y 23, y se calcula, como valor único para el Percentil 26

$$\frac{22+23}{2} = 22.5.$$

- Calculemos el Decil 4 (eso es, el cuantil de orden .40) para la misma variable. A continuación, se presenta parte de la tabla de proporción acumulada:

j	x_j	$frec(x_j)$	$p(x_j)$	$cfrec(x_j)$	$F(x_j)$
⋮	⋮	⋮	⋮	⋮	⋮
5	23	6	.12	19	.38
6	24	7	.14	26	.52
7	25	5	.10	31	.62
⋮	⋮	⋮	⋮	⋮	⋮

Puesto que el valor .40 no ocurre en la tabla, se busca en la siguiente fila el valor x para el cual $F(x) > .40$. Se encuentra que el 24 corresponde con el Decil 4.

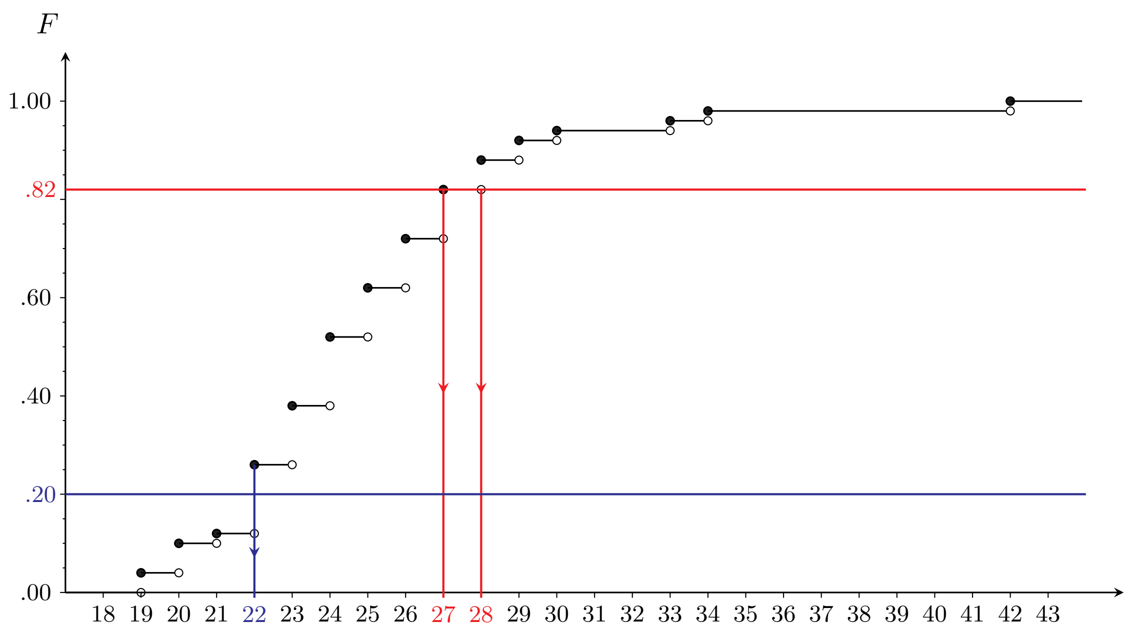
Cálculo de cuantiles a través de la gráfica de F

Para calcular un cuantil a través de una gráfica, se traza una línea horizontal que cruza la ordenada (el eje vertical) en r y se procede como sigue:

- Si esta línea contiene parte de la función $F(x)$, es decir, si coincide con uno de los escalones (se presenta el Caso 1 de la definición), entonces cada valor x cubierto por esta parte es un cuantil x_r.
- Si la línea pasa entre dos escalones de la función (Caso 2), x_r corresponde con el valor x donde la linea horizonal cruza con la linea vertical que conecta los escalones.

Ejemplo

- Busquemos el percentil 82 para la variable IMC, usando la gráfica de la función de proporción acumulada que se presenta a continuación. Pc_{82} corresponde con $x_{.82}$, por lo cual se traza una línea horizontal por el valor .82 de la ordenada (en color rojo). Como esta línea coincide en parte con la función $F(x)$, todos los valores x en la abscisa abajo de esta parte, es decir, todos los x para los cuales $27 \leqslant x < 28$ son un Percentil 82.
- Consideremos ahora el segundo decil, el cual corresponde con $x_{.20}$. Puesto que en la gráfica de F la línea horizontal por el valor .20 (en color azul) no coincide con ninguna parte de la función, el cuantil $x_{.20}$ es igual al valor x donde la linea horizontal cruza la linea que conecta las escalones de la función, en este caso el valor $x = 22$.



Índice de Masa Corporal

2.3 Características de una distribución

Las funciones de frecuencia y proporción (sea acumulada o no) contienen toda la información sobre cómo los valores de la variable se distribuyen en la muestra. Sin embargo, muchas veces uno se satisface con unos pocos números que resumen los datos, dando información sobre las características importantes de la distribución. Estos números se llaman *estadísticos*.

Estadístico

Un *estadístico* es una característica númerica de una distribución.

En el Tema 11 se presentará una definición más formal de un estadístico.

Se puede preguntar: ¿Sobre cuáles aspectos o cuáles características de la distribución deberían informar los estadísticos?
Como respuesta a esta pregunta, se han distinguido cuatro tipos de estadísticos:

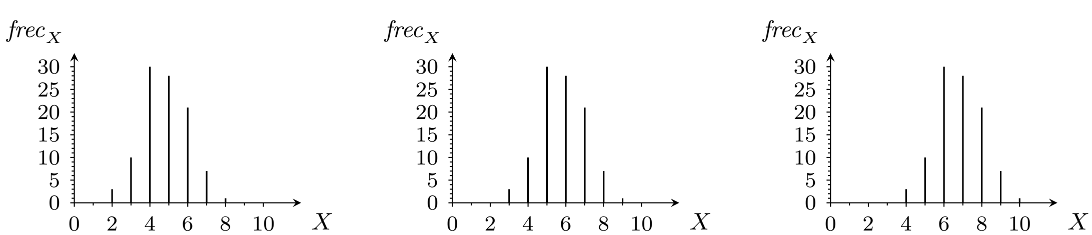
1. Estadísticos de tendencia central
2. Estadísticos de variabilidad (o dispersión)
3. Estadísticos de asimetría
4. Estadísticos de apuntamiento

La tendencia central y la variabilidad son las características más importantes de una distribución.

Tendencia central Los estadísticos de tendencia central informan sobre el "punto medio" de la distribución, es decir, alrededor de qué se centra la distribución. En otras palabras, ¿dónde se encuentra la distribución en la abscisa?; en general, ¿cuál es la magnitud de los números?

Ejemplo de distribuciones que difieren respecto a la tendencia central

Las figuras a continuación presentan el diagrama de barras para una variable X en tres muestras diferentes del mismo tamaño. ¿Cómo difieren estas tres distribuciones?

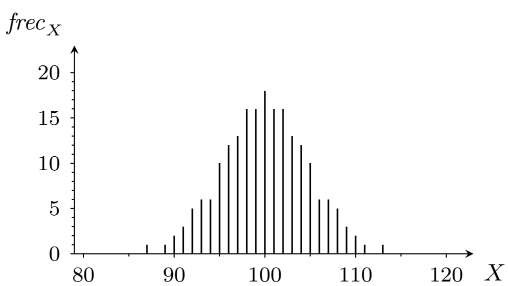


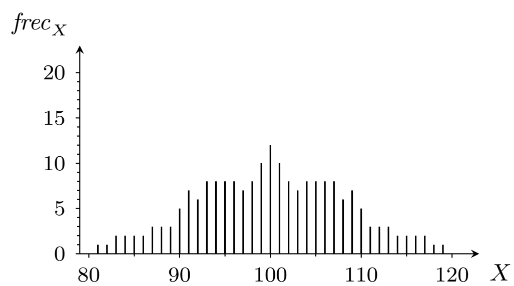
Estas distribuciones difieren respecto a su ubicación en la abscisa. En la primera distribución, la variable X asume valores que generalmente son más bajos en comparación con la segunda distribución. La tercera distribución asume valores más altos en general.

Variabilidad o dispersión Los estadísticos de variabilidad o dispersión informan sobre qué tan alejados están los valores del centro en general o, lo cual es equivalente, qué tanto difieren los valores de la variable. En otras palabras, la variabilidad indica qué tan ancha o qué tan estrecha es la distribución.

Ejemplo de distribuciones que difieren respecto a la variabilidad

Las siguientes dos distribuciones—de la variable X en dos distintas muestras del mismo tamaño—tienen la tendencia central muy similar o idéntica (los estadísticos de tendencia central más importantes dan valores idénticos para ambas distribuciones). Sin embargo, las distribuciones no son idénticas: Difieren respecto de cuánta variabilidad hay entre los valores de la variable.



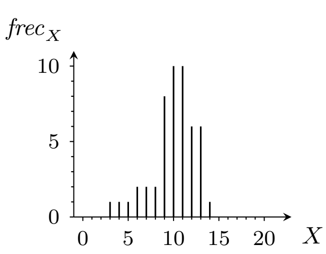


La variabilidad es más grande en la distribución del lado derecho; los valores fluctúan más.

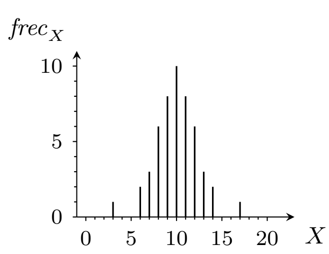
Asimetría Una distribución es simétrica si existe una línea vertical por la cual se puede "espejar" la distribución sin que cambie. En el caso contrario, se dice que la distribución es asimétrica. Cuando es asimétrica, se distingue entre asimetría positiva y negativa.

Ejemplo de distribuciones que difieren respecto a la (a)simetría

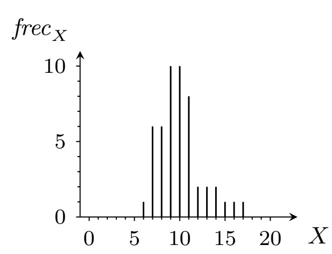
Las siguientes tres distribuciones—de la variable X en tres distintas muestras del mismo tamaño—son muy similares (o idénticas) respecto de la tendencia central y la variabilidad (los estadísticos de tendencia central y variabilidad más importantes dan valores idénticos para ambas distribuciones). Sin embargo, hay otra característica en la cual difieren significativamente: la simetría y el tipo de asimetría.



Asimetría negativa



Simetría



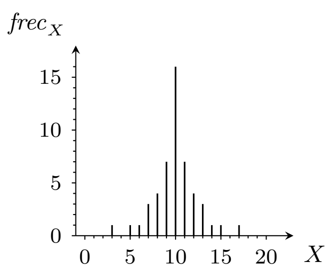
Asimetría positiva

Es obvio que la distribución en medio es simétrica (se puede espejar alrededor de la línea que pasa por $X = 10$ sin que cambie). Las otras dos distribuciones son asimétricas: La distribución al izquierda tiene la cola más larga por el lado negativo; la distribución a la derecha la tiene por el lado positivo. Se dice que estas distribuciones exhíben asimetría negativa y positiva, respectivamente.

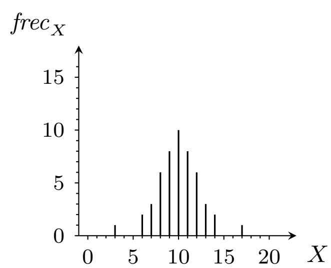
Apuntamiento o curtosis

Ejemplo de distribuciones que difieren respecto al apuntamiento

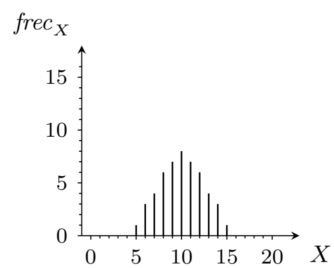
Las tres distribuciones que se grafican a continuación son todas simétricas y tienen también valores idénticos para los estadísticos más importantes de tendencia central y variabilidad. Difieren respecto del apuntamiento (también llamado curtosis).



Leptocúrtica



Mesocúrtica



Platicúrtica

La distribución del lado izquierdo es mucho más puntiaguda cerca del centro de la distribución, mientras la distribución al lado derecho es más plana. Se dice que las distribuciones son leptocúrticas, mesocúrticas y platicúrticas, respectivamente.

Problemas

1. Construir una representación gráfica adecuada de (a) la distribución de proporción y (b) la distribución acumulada de frecuencia (si aplica) para cada variable en los datos de la página 34.

2. La siguiente tabla muestra la distribución de proporción acumulada de la nota final (variable X) de alguna asignatura en un salón de 95 alumnos:

x	$F(x)$	x	$F(x)$
0	0.13	6	0.71
1	0.13	7	0.87
2	0.18	8	0.93
3	0.26	9	0.99
4	0.35	10	1.00
5	0.55		

   (a) A partir de la distribución de proporción acumulada de X, derivar la distribución de frecuencia de X.

   (b) Suponiendo que para aprobar es necesario tener una nota de 6 o mayor, ¿cuál es el porcentaje de personas que aprueban?

   (c) ¿Cuál es la mediana y cómo se interpreta?

3. La siguiente figura representa la distribución de las puntuaciones (número de respuestas correctas) en un examen de opción múltiple que consiste en 100 preguntas de cuatro alternativas cada una, a través de un diagrama de tallo y hojas:

4	58
5	0223556778
6	00555678
7	0002333333466788899
8	024
9	1

   (a) ¿Cuál es el tamaño de la muestra (n)?

   (b) ¿Cuáles son las puntuaciones mínima y máxima que se han observado en esta muestra?

   (c) ¿Cuál es la puntuación con mayor frecuencia (es decir, el valor x para la cual $frec(x)$ es más alto)?

   (d) Hallar y interpretar los tres cuartiles.

Tema 3
Estadísticos de tendencia central y variabilidad

Índice

Al final del tema anterior, se introdujo una tipología de estadísticos que resumen distribuciones univariadas. En particular, esta tipología distinguió entre estadísticos de (a) tendencia central, (b) variabilidad o dispersión, (c) simetría (vs. asimetría) y (d) apuntamiento o curtosis. En el tema actual pasan revista los estadísticos más importantes de las primeras dos categorías; se dejarán a un lado los estadísticos que informan sobre el grado de asimetría y apuntamiento, en consideración de que estos estadísticos se utilizan mucho menos en la práctica y, además, su interpretación suele ser ambigua en varias ocasiones.

En la primera sección, se revisan los estadísticos de tendencia central, incluyendo la moda, la mediana y la media aritmética. Después, se pasa a las medidas que informan sobre la variabilidad en la distribución: la amplitud total, la amplitud intercuartil, la varianza y desviación estándar, y el coeficiente de variación. En la Sección 3.3, se presenta un teorema sobre la distribución de los valores alrededor de la media en términos de la desviación estándar. Concluimos este tema con la introducción de una herramienta para visualizar en una gráfica información de tendencia central y variabilidad simultáneamente (Sección 3.4).

Antes de iniciar la presentación de los distintos estadísticos, dos notas importantes: Primero, no es así que todos los estadísticos aplican para todos los tipos de variables; en otras palabras, algunos estadísticos pueden ser muy apropiados para una variable, mientras que para otras variables simplemente no tiene sentido calcularlos. Si un estadístico es o no es adecuado para un tipo de variable, depende fundamentalmente del nivel de medición de la misma. Como se introdujo en el primer tema (Sección 1.4), los diferentes niveles de medición tienen implicaciones para la validez de ciertas operaciones en la variable. Por ejemplo, para variables de nivel nominal—aunque tengan valores numéricos—no vale interpretar la magnitud de los valores; asimismo, en variables de nivel ordinal, no se justifica la comparación de las diferencias entre valores, etc.. Por otro lado, el cálculo de un estadístico implica ciertas operaciones y, en el caso de que estas operaciones no correspondan con las operaciones permitidas por el tipo de la variable, el estadístico no es apropiado para esta variable. Se deja como ejercicio averiguar cuáles de los estadísticos que se presentan en este tema son adecuados para qué tipo de variables.

Segundo—para los puristas del lenguaje estadístico—, tendencia central y variabilidad son características de una distribución, aunque muchas veces se habla de la tendencia central o variabilidad de una variable. Estrictamente hablando, se debería decir siempre que se trata de la tendencia central o variabilidad de alguna variable *en alguna distribución*. Sin embargo, por conveniencia, no siempre se hace referencia explícita a la distribución, pero en este caso debe ser claro que esta referencia está implícita.

3.1 Estadísticos de tendencia central

Los estadísticos de tendencia central informan sobre “alrededor de qué valor se centra la distribución” o “la magnitud, en general, de los valores”, es decir, sobre cuál es el valor más “típico” o más “común” de la distribución. Existen muchos estadísticos de tendencia central; la diferencias entre ellos se originan de los diferentes significados que se pueden dar a las expresiones como “centrarse alrededor”, “magnitud en general” o “valor típico”.

3.1.1 Moda

Moda

Cada valor x de una variable X para el cual

(a) $\mathit{frec}(x) > 1$

y (b) $\mathit{frec}(x)$ es máxima [$\mathit{frec}(x) \geqslant \mathit{frec}(x^\star)$ para cualquier otro valor $x^\star$]

es una *moda* de X.

Notación: Mo_X

Ejemplos

En la tabla de frecuencia de la variable *Presión Arterial Diastólica* de los datos de la página 34 se observa que el valor 73 tiene la frecuencia más alta. Además, esta frecuencia es mayor que 1. Puesto que no hay otros valores que cumplen esta condición, 73 es la única moda de esta variable.
Se dice que la distribución es *unimodal*.

x_j	$frec(x_j)$	x_j	$frec(x_j)$	x_j	$frec(x_j)$	x_j	$frec(x_j)$
68	2	79	4	88	3	99	2
69	1	80	2	90	2	100	2
73	5	82	2	91	2	103	1
75	3	83	1	93	2	104	1
76	1	84	1	94	2	123	1
77	3	85	1	95	1		
78	1	86	3	97	1		

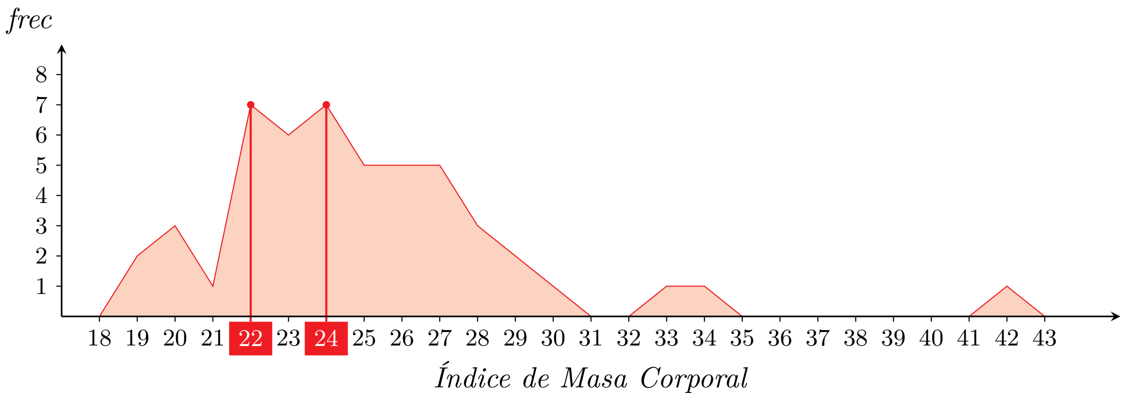
La variable *Presión Arterial Sistólica*, por otro lado, tiene múltiples valores donde la función de frecuencia llega a su máximo: Tanto para 114, como para 116 y 129, hay 4 personas con esta presión arterial sistólica. No hay otros valores con una frecuencia mayor, por lo cual cada uno de estos tres valores es una moda de la distribución. Se dice que esta distribución es *multimodal* (o, más específico, que es *trimodal*).

x_j	$frec(x_j)$	x_j	$frec(x_j)$	x_j	$frec(x_j)$	x_j	$frec(x_j)$
107	1	116	4	129	4	145	1
108	1	120	1	132	2	147	1
109	1	121	2	133	1	148	2
111	1	122	1	134	1	159	1
112	3	126	1	138	2	164	2
113	1	127	3	142	1	172	1
114	4	128	3	144	3	177	1

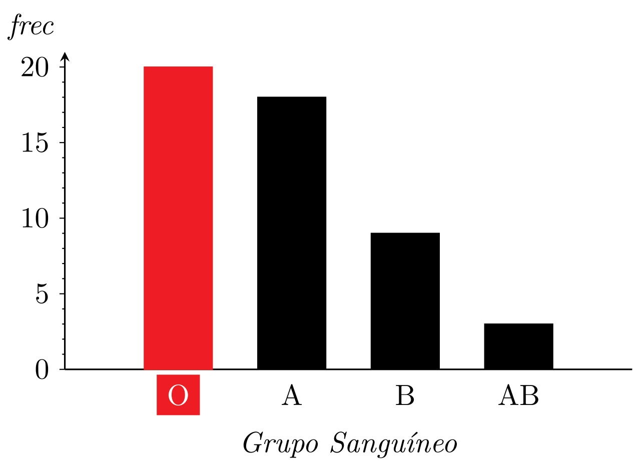
En las representaciones gráficas (como en un diagrama de barras, un diagrama de rectángulos, un histograma o un polígono de frecuencia), la(s) moda(s) corresponde(n) con el (los) punto(s) más alto(s).

Ejemplos

El polígono de frecuencia de la variable *Índice de Masa Corporal* muestra directamente donde están las modas de la distribución. Como la distribución tiene 2 modas, se dice que es *bimodal.*



A partir del diagrama de rectángulos para la variable *Grupo Sanguíneo*, es fácil encontrar que la moda de la distribución es el Grupo O.



Aunque la moda tradicionalmente se conoce como un estadístico de tendencia central, es claro que no necesariamente se encuentra en el centro de la distribución. La moda de *Presión Arterial Diastólica* (73), por ejemplo, está más al lado izquierdo de la distribución. Similarmente, como lo muestra el último ejemplo, la moda puede ser un estadístico apropiado para variables cualitativas (cuyos valores no tienen un orden y para las cuales entonces no tiene sentido hablar del centro de la distribución). Efectivamente, la moda simplemente informa sobre el valor más común en la muestra para la variable bajo consideración y su cálculo no implica ninguna comparación de la magnitud de los valores de la variable. Puesto que un valor extremo puede ser el más frecuente de todos, es posible que la moda se encuentre en uno de los extremos de la distribución. (En este sentido, sería mejor hablar de la moda como un estadística de “tendencia”, quitando el adjectivo “central”). Por lo tanto, rara vez se utiliza la moda como estadístico de tendencia central para variables (cuasi)cuantitativas.

A pesar de que es común hablar de “*la* moda”, los ejemplos anteriores muestran que puede haber más de una moda en una distribución. Al mismo tiempo, también existe el caso especial en el cual una distribución no tenga ninguna moda. Este caso se presenta cuando cada valor ocurre sólo una vez (dado que esto implica que la primera condición en la definición de la moda, $frec(x) > 1$, no se cumple para ningún valor x).

3.1.2 Mediana

Mediana

La *mediana* de una variable X es el valor que corresponde con el cuantil de orden .50 de esta variable.

Notación: Me_X

Al definir los cuantiles (Sección 2.2.5), se consideró la posibilidad de que el valor para un cuantil no era único, sino que correspondía con un intervalo de valores. Así, también puede ocurrir que la mediana de una variable no sea única y que tengamos un intervalo de valores posibles para la mediana. Este caso se presenta si la proporción acumulada para un valor x es exactamente igual a .50. Sin embargo, en aplicaciones prácticas, se prefiere un único valor para la mediana, de tal forma que, cuando el cuantil de orden .50 corresponde con un intervalo de valores, se toma el valor central de este intervalo como la mediana.

Ejemplos

La distribución de proporción acumulada para la variable *Índice de Masa Corporal* de los datos de la página 34 se presenta parcialmente en la siguiente tabla:

	x_j	$frec(x_j)$	$p(x_j)$	$cfrec(x_j)$	$F(x_j)$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	21	1	.02	6	.12
	22	7	.14	13	.26
	23	6	.12	19	.38
$Me_{IMC} = 24$	24	7	.14	26	.52
	25	5	.10	31	.62
	26	5	.10	36	.72
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Como no existe un valor x_j para el cual $F(x_j) = 0.50$, entonces la mediana corresponde con el primer valor x_j que tiene $F(x_j) > 0.50$. Como se lee en la tabla, 24 es la mediana de la variable *Índice de Masa Corporal*.

Para la variable *Presión Arterial Diastólica*, se presenta (parte de) la tabla de frecuencias y proporciones:

x_j	$frec(x_j)$	$p(x_j)$	$cfrec(x_j)$	$F(x_j)$
⋮	⋮	⋮	⋮	⋮
79	4	.08	20	.40
80	2	.04	22	.44
82	2	.04	24	.48
83	1	.02	25	.50
84	1	.02	26	.52
85	1	.02	27	.54
86	3	.06	30	.60
⋮	⋮	⋮	⋮	⋮

$$\begin{aligned} Me_{PAD} &= \frac{83+84}{2} \\ &= 83.5 \end{aligned}$$

Para todos los valores $83 \leqslant x < 84$, $F(x) = 0.50$; es decir, la mediana corresponde con el intervalo de valores [83,84[. El valor único para la mediana entonces es 83.5, el punto medio de este intervalo.

Como se acaba de ilustrar, la definición general de cuantiles de orden r permite calcular la mediana. Sin embargo, el siguiente método es más fácil de aplicar:

Método alternativo para el cálculo de Me_X

El método consiste en dos pasos:

Paso 1: Ordenar las n observaciones según la magnitud de sus valores en la variable X tal que

$$x_1 \leqslant x_2 \leqslant x_3 \leqslant \ldots \leqslant x_n.$$

Paso 2: Identificar el valor central, aplicando la siguiente regla:

$$Me_X = \begin{cases} x_{\frac{n+1}{2}} & \text{si } n \text{ es impar} \\ \dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} & \text{si } n \text{ es par} \end{cases}$$

Ilustración del método alternativo para calcular la mediana

Calculemos, utilizando el nuevo método, la mediana de la variable *Presión Arterial Diastólica* en los datos de la página 34 de forma separada para las submuestras correspondientes con diferentes grupos sanguíneos:

Grupo O (hay 20 personas con grupo sanguíneo O)

Paso 1: Ordenar las 20 observaciones según su valor en la variable *Presión Arterial Diastólica*:

$$\begin{array}{lllllll} x_1=73 & x_4=73 & x_7=77 & x_{10}=82 & x_{13}=90 & x_{16}=94 & x_{19}=99 \\ x_2=73 & x_5=75 & x_8=79 & x_{11}=83 & x_{14}=90 & x_{17}=97 & x_{20}=103 \\ x_3=73 & x_6=77 & x_9=80 & x_{12}=88 & x_{15}=91 & x_{18}=99 & \end{array}$$

Paso 2: Dado que $n = 20$ es un número par, hay que buscar los valores $x_{\frac{20}{2}} = x_{10}$ y $x_{\frac{20}{2}+1} = x_{11}$. Efectivamente, en una muestra de 20 observaciones, los valores en la posición 10 y 11 son los valores centrales. Para obtener la mediana se calcula el promedio de estos dos valores:

$$Me_{PAD} = \frac{x_{10} + x_{11}}{2} = \frac{82 + 83}{2} = 82.5$$

Grupo A (hay 18 personas con grupo sanguíneo A)

Paso 1: Ordenar las 18 observaciones según su valor en la variable *Presión Arterial Diastólica*:

$$\begin{array}{llllll} x_1=68 & x_4=75 & x_7=79 & x_{10}=86 & x_{13}=93 & x_{16}=100 \\ x_2=73 & x_5=76 & x_8=80 & x_{11}=88 & x_{14}=93 & x_{17}=100 \\ x_3=75 & x_6=78 & x_9=86 & x_{12}=91 & x_{15}=95 & x_{18}=104 \end{array}$$

Paso 2: Dado que $n = 18$ es un número par, hay que buscar los valores $x_{\frac{18}{2}} = x_9$ y $x_{\frac{18}{2}+1} = x_{10}$. La mediana se obtiene por el promedio de estos dos valores:

$$Me_{PAD} = \frac{x_9 + x_{10}}{2} = \frac{86 + 86}{2} = 86$$

Grupo B (hay 9 personas con grupo sanguíneo B)

Paso 1: Ordenar las 9 observaciones según su valor en la variable *Presión Arterial Diastólica*:

$$\begin{array}{lllllll} x_1=68 & x_2=77 & x_3=79 & x_4=79 & x_5=84 & x_6=85 & x_7=86 \\ x_8=94 & x_9=123 & & & & & \end{array}$$

Paso 2: Dado que $n = 9$ es un número impar, la mediana corresponde con el valor $x_{\frac{9+1}{2}} = x_5 = 84$.

Grupo AB (hay 3 personas con grupo sanguíneo B)

Paso 1: Ordenar las 3 observaciones según su valor en la variable *Presión Arterial Diastólica*:

$$x_1 = 69 \qquad x_2 = 82 \qquad x_3 = 88$$

Paso 2: Dado que $n = 3$ es un número impar, la mediana corresponde con el valor $x_{\frac{3+1}{2}} = x_2 = 82$. Nótese que, cuando n es impar, el valor central efectivamente está en la posición $\frac{n+1}{2}$; es decir, el quinto valor si $n = 9$ y el segundo valor si $n = 3$.

Equivalencia del método alternativo para calcular la mediana y el método general por cuantiles

Se comprueba que los dos métodos para calcular la mediana siempre llevan al mismo resultado.

Consideremos los dos casos:

Caso 1: n **es impar.** En este caso, no existe ningún valor x con $F(x) = .50$ (ya que, si n es impar y $\frac{cfrec(x)}{n} = \frac{1}{2}$, entonces $cfrec(x)$ no sería un número entero).
Por otro lado, sabemos que $F\left(x_{\frac{n+1}{2}}\right) > .50$, puesto que $F\left(x_{\frac{n+1}{2}}\right) \geqslant \frac{\left(\frac{n+1}{2}\right)}{n}$. Además, para cualquier valor x menor que $x_{\frac{n+1}{2}}$, $F(x) < .50$. Por consecuencia, $x_{\frac{n+1}{2}}$ es el valor más pequeño que cumple $F(x) > .50$ y, entonces, conforme al Caso 2 de la definición en la página 51, corresponde con el cuantil $x_{.50}$.

Caso 2: n **es par.** Sigue que $F\left(x_{\frac{n}{2}}\right) \geqslant .50$. Hay dos posibilidades:

1. $F\left(x_{\frac{n}{2}}\right) = .50$. Ésto implica que el valor siguiente a $x_{\frac{n}{2}}$ es diferente, es decir, $x_{\frac{n}{2}+1} > x_{\frac{n}{2}}$. Por consecuencia, $x_{\frac{n}{2}+1}$ es el primer valor x que cumple $F(x) > .50$. Por lo tanto:

$$Me_X = \frac{x_{\min} + x_{\max}}{2} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

según el Caso 1 de la definición en la página 51.

2. $F\left(x_{\frac{n}{2}}\right) > .50$. Implica que no existe un valor x para el cual $F(x) = .50$. Además, $x_{\frac{n}{2}}$ es el primer valor x que cumple que $F(x) > .50$ y entonces, conforme al Caso 2 de la definición en la página 51, corresponde con el $x_{.50}$.
A continuación, puesto que $F\left(x_{\frac{n}{2}}\right) > .50$, se cumple que $x_{\frac{n}{2}} = x_{\frac{n}{2}+1}$ y entonces:

$$\frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = x_{\frac{n}{2}} = Me_X \qquad \blacksquare$$

Es muy instructivo verificar las diferentes partes de la prueba con los ejemplos anteriores. El cálculo de la mediana de *Presión Arterial Diastólica* en el grupo sanguíneo O ilustra el Caso 2.1 de la prueba (es decir, si n es par y $F(x_{\frac{n}{2}}) = .50$); el cálculo en el grupo sanguínea A permite verificar el Caso 2.2 (cuando n es par y $F(x_{\frac{n}{2}}) > .50$); y con los grupos B y AB, se presenta el Caso 1 de la prueba.

El método alternativo para calcular la mediana que acabamos de presentar y comprobar, más que darnos una nueva herramienta para calcular la mediana—hoy en día el cálculo se realizará casi siempre a través de un software estadístico—, proporciona una justificación de la interpretación de la mediana como "el valor que está en el centro de la distribución" o "el valor que tiene aproximadamente 50 % de las observaciones abajo y 50 % de las observaciones encima".

3.1.3 Media aritmética

3.1.3.1 Definición

Media aritmética

Para n observaciones que tienen valores en una variable X, la *media aritmética* se define como:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3.1}$$

donde x_i es el valor que tiene la observación i en la variable X.

Nota:

Se añade el adjectivo *aritmética* para distinguir este tipo de media de otros tipos, como la media geométrica o la media armónica (que se presentarán en la Sección 3.1.4). No obstante, la media aritmética es la más importante, por lo cual muchas veces se habla de "la media" sin especificar que es la aritmética. También en este curso, cuando se habla sobre "la media", siempre se refiere a la media aritmética.

Ejemplo

Para conocer la media de la variable *Edad* en los datos de la página 34, se calcula primero la suma de los 50 valores observados:

$$\begin{aligned}\sum_{i=1}^{50} x_i &= 58 + 55 + 38 + 57 + 48 + 53 + 44 + 46 + 32 + 35 + 34 + 29 + 49 \\ &\quad + 41 + 57 + 45 + 34 + 55 + 46 + 30 + 38 + 55 + 57 + 58 + 47 + 55 \\ &\quad + 55 + 54 + 47 + 43 + 38 + 51 + 35 + 44 + 47 + 52 + 33 + 41 + 56 \\ &\quad + 55 + 59 + 45 + 46 + 37 + 55 + 39 + 58 + 36 + 29 + 49 \\ &= 2300\end{aligned}$$

y, a continuación, se divide esta suma entre el número de observaciones:

$$\begin{aligned}\overline{x} &= \frac{\sum_{i=1}^{50} x_i}{50} \\ &= \frac{2300}{50} = 46.\end{aligned}$$

De forma equivalente, se puede definir la media aritmética a partir de la función de frecuencia o proporción de la variable:

Definiciones equivalentes de la media aritmética

Para (a) una variable X, que asume m distintos valores $x_1, x_2, x_3, \ldots, x_m$
y (b) un conjunto de n observaciones con valores en la variable X,

la *media aritmética* se define, a través de la función de frecuencia, como:

$$\overline{x} = \frac{\sum_{j=1}^{m} frec(x_j) x_j}{n} \tag{3.2a}$$

o, de forma equivalente, a través de la función de proporción:

$$\overline{x} = \sum_{j=1}^{m} p(x_j) x_j \tag{3.2b}$$

Ejemplo

La siguiente tabla muestra la función de frecuencia y proporción de la variable *Edad* en los datos de la página 34:

j	x_j	$frec(x_j)$	$p(x_j)$	j	x_j	$frec(x_j)$	$p(x_j)$
1	29	2	.04	15	46	3	.06
2	30	1	.02	16	47	3	.06
3	32	1	.02	17	48	1	.02
4	33	1	.02	18	49	2	.04
5	34	2	.04	19	51	1	.02
6	35	2	.04	20	52	1	.02
7	36	1	.02	21	53	1	.02
8	37	1	.02	22	54	1	.02
9	38	3	.06	23	55	7	.14
10	39	1	.02	24	56	1	.02
11	41	2	.04	25	57	3	.06
12	43	1	.02	26	58	3	.06
13	44	2	.04	27	59	1	.02
14	45	2	.04				

A partir de las frecuencias, se obtiene primero:

$$\begin{aligned}\sum_{j=1}^{27} frec(x_j) x_j &= 2 \times 29 + 1 \times 30 + 1 \times 32 + \ldots + 3 \times 58 + 1 \times 59 \\ &= 2300,\end{aligned}$$

lo cual lleva al mismo resultado que con el método de la primera definición:

$$\begin{aligned}\overline{x} &= \frac{\sum_{j=1}^{27} frec(x_j)x_j}{50} \\ &= \frac{2300}{50} = 46.\end{aligned}$$

Similarmente, utilizando la función de proporción lleva a:

$$\begin{aligned}\overline{x} &= \sum_{j=1}^{27} p(x_j)x_j \\ &= .04 \times 29 + .02 \times 30 + .02 \times 32 + \ldots + .06 \times 58 + .02 \times 59 \\ &= 46.\end{aligned}$$

3.1.3.2 Dos propiedades importantes de la media

En esta sección se presentan y comprueban dos propiedades de la media. Sin embargo, primero hay que introducir un nuevo concepto:

Puntuación diferencial

Se llama *puntuación diferencial* de la observación i en la variable X la diferencia entre el valor original x_i y la media del grupo $\overline{x}$:

$$\mathcal{D}(x_i) = x_i - \overline{x}.$$

Propiedad 1 de la media

En cualquier muestra de n observaciones con valores $x_1, x_2, \ldots, x_n$ en una variable X, la suma de las puntuaciones diferenciales en esta variable iguala cero. Es decir,

$$\sum_{i=1}^{n} \mathcal{D}(x_i) = \sum_{i=1}^{n} (x_i - \overline{x}) = 0.$$

Comprobación de la Propiedad 1 de la media

La primera propiedad se comprueba fácilmente usando las propiedades de sumas:

$$\begin{aligned}\sum_{i=1}^{n}(x_i - \overline{x}) &= \left(\sum_{i=1}^{n} x_i\right) - \left(\sum_{i=1}^{n} \overline{x}\right)\\ &= n\,\frac{\sum_{i=1}^{n} x_i}{n} - n\,\overline{x}\end{aligned}$$

ya que, en la segunda suma, $\overline{x}$ es una constante. Simplificando se obtiene:

$$\begin{aligned}&= n\,\overline{x} - n\,\overline{x}\\ &= 0\end{aligned}$$

■

Nótese que la primera propiedad implica que la media de las puntuaciones diferenciales es igual a 0.

Ilustración de la Propiedad 1 de la media

Calcular la puntuación diferencial de cada observación en la variable *Edad* y después sumar todas las puntuaciones diferenciales da efectivamente un resultado de 0:

i	x_i	$x_i - \overline{x}$	$= \mathcal{D}(x_i)$
1	58	$58 - 46$	$= +12$
2	55	$55 - 46$	$= +9$
3	38	$38 - 46$	$= -8$
4	57	$57 - 46$	$= +11$
5	48	$48 - 46$	$= +2$
6	53	$53 - 46$	$= +7$
7	44	$44 - 46$	$= -2$
8	46	$46 - 46$	$= 0$
⋮	⋮	⋮	⋮
48	36	$36 - 46$	$= -10$
49	29	$29 - 46$	$= -17$
50	49	$49 - 46$	$= +3$
		$\sum_{i=1}^{50}(x_i - \overline{x}) =$	0

Esta propiedad ayuda para entender mejor la media aritmética. En particular, ha dado lugar a la interpretación de la media aritmética como el *punto de equilibrio* (o el centro de gravedad) de una distribución. A modo de ilustración, considérese la Figura 3.1 de la siguiente página: Cada observación en nuestra muestra de ejemplo se ha representado por un bloque de madera y se ha apilado, junto con las otras observaciones, en una palanca sin peso en la posición que corresponde con su valor en la variable *Edad*. Según el principio de la palanca (de Arquímedes), la palanca está en equilibrio cuando el torque (momento de fuerza) total hacia la izquierda es igual al torque total hacia la derecha. Recuérdese que el torque de un elemento corresponde con el producto de su peso y su

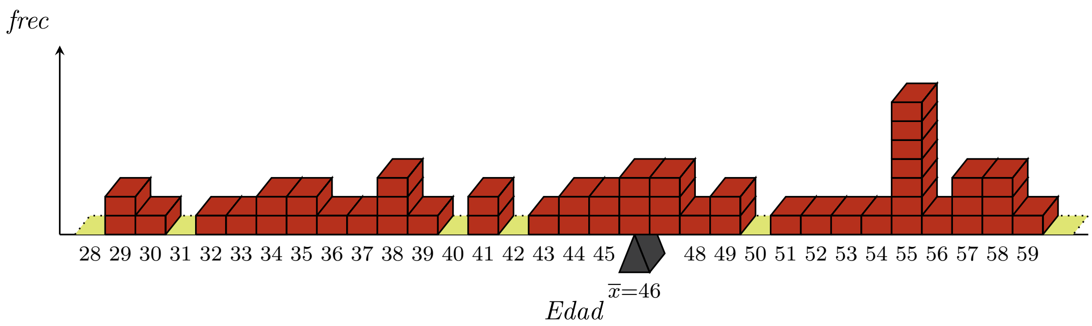


Figura 3.1: La media aritmética se puede comprender como el punto de equilibrio de la distribución.

distancia hacia el fulcro y que el torque total es la suma de los torques asociados con cada elemento. Dado que los pesos de los 50 bloques son iguales, el torque total a la izquierda es proporcional a la suma de las distancias de todos los elementos a la izquierda del fulcro; similarmente, el torque total a la derecha es proporcional a la suma de todas las distancias a la derecha del fulcro. Las distancias son nada más las puntuaciones diferenciales, así que se puede concluir que la palanca está en equilibrio si las puntuaciones diferenciales negativas neutralizan las puntuaciones diferenciales positivas, o bien, cuando el fulcro se encuentra donde la suma de las puntuaciones diferenciales sea cero.

Propiedad 2 de la media

En cualquier muestra de n observaciones con valores $x_1, x_2, \ldots, x_n$ en una variable X, la media aritmética es la constante hacia la cual la suma de diferencias cuadráticas es mínima.
En símbolos: Siempre se cumple que

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 < \sum_{i=1}^{n}(x_i - c)^2,$$

donde c es cualquier constante diferente de $\overline{x}$ (es decir, $c \in \mathbb{R}$ y $c \neq \overline{x}$).

En otras palabras, la Propiedad 2 dice que la suma de las diferencias cuadráticas hacia la media es menor que la suma de las diferencias cuadráticas hacia cualquier otra constante. Antes de demostrar esta propiedad, la ilustramos con un ejemplo.

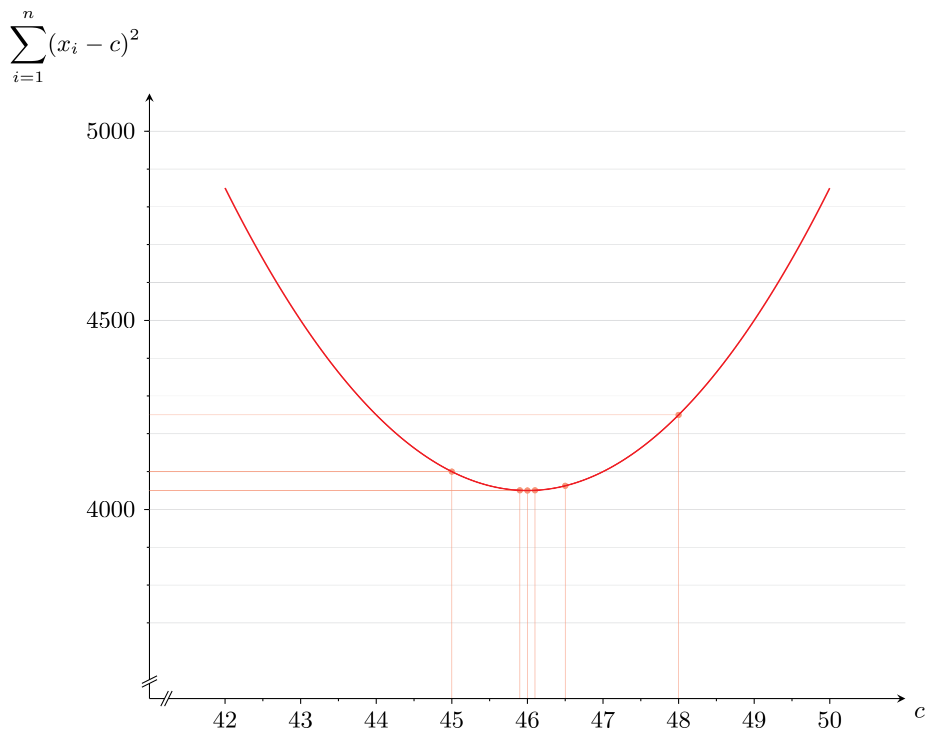
Ejemplo

En los ejemplos anterios ya se halló que la media aritmética de la variable *Edad* en los datos de la página 34 es 46. En la siguiente tabla se muestra el resultado de la suma $\sum_{i=1}^{n}(x_i - c)^2$ para $c = \overline{x} = 46$ y cinco otros valores para c:

i	x_i	$(x_i - 46)^2$	$(x_i - 48)^2$	$(x_i - 45)^2$
1	58	$(58-46)^2 = 144$	$(58-48)^2 = 100$	$(58-45)^2 = 169$
2	55	$(55-46)^2 = 81$	$(55-48)^2 = 49$	$(55-45)^2 = 100$
3	38	$(38-46)^2 = 64$	$(38-48)^2 = 100$	$(38-45)^2 = 49$
4	57	$(57-46)^2 = 121$	$(57-48)^2 = 81$	$(57-45)^2 = 144$
5	48	$(48-46)^2 = 4$	$(48-48)^2 = 0$	$(48-45)^2 = 9$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
48	36	$(36-46)^2 = 100$	$(36-48)^2 = 144$	$(36-45)^2 = 81$
49	29	$(29-46)^2 = 289$	$(29-48)^2 = 361$	$(29-45)^2 = 256$
50	49	$(49-46)^2 = 9$	$(49-48)^2 = 1$	$(49-45)^2 = 16$
		$\sum_{i=1}^{50}(x_i - 46)^2 = 4050$	$\sum_{i=1}^{50}(x_i - 48)^2 = 4250$	$\sum_{i=1}^{50}(x_i - 45)^2 = 4100$

i	x_i	$(x_i - 46.5)^2$	$(x_i - 46.1)^2$	$(x_i - 45.9)^2$
1	58	$(58-46.5)^2 = 132.25$	$(58-46.1)^2 = 141.61$	$(58-45.9)^2 = 146.41$
2	55	$(55-46.5)^2 = 72.25$	$(55-46.1)^2 = 79.21$	$(55-45.9)^2 = 82.81$
3	38	$(38-46.5)^2 = 72.25$	$(38-46.1)^2 = 65.61$	$(38-45.9)^2 = 62.41$
4	57	$(57-46.5)^2 = 110.25$	$(57-46.1)^2 = 118.81$	$(57-45.9)^2 = 123.21$
5	48	$(48-46.5)^2 = 2.25$	$(48-46.1)^2 = 3.61$	$(48-45.9)^2 = 4.41$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
48	36	$(36-46.5)^2 = 110.25$	$(36-46.1)^2 = 102.01$	$(36-45.9)^2 = 98.01$
49	29	$(29-46.5)^2 = 306.25$	$(29-46.1)^2 = 292.41$	$(29-45.9)^2 = 285.61$
50	49	$(49-46.5)^2 = 6.25$	$(49-46.1)^2 = 8.41$	$(49-45.9)^2 = 9.61$
		$\sum_{i=1}^{50}(x_i - 46.5)^2 = 4062.50$	$\sum_{i=1}^{50}(x_i - 46.1)^2 = 4050.50$	$\sum_{i=1}^{50}(x_i - 45.9)^2 = 4050.50$

Se observa efectivamente que la suma de las diferencias cuadráticas hacia la media aritmética siempre es más baja en comparación con la suma de las diferencias cuadráticas hacia otra constante. La tabla anterior muestra el resultado de la suma para 6 valores de c. En principio, se puede calcular el resultado para todos las posibles constantes c y construir una gráfica con, en la abscisa, los diferentes valores de c y en la ordenada el resultado de la suma $\sum_{i=1}^{n}(x_i - c)^2$. Desarrollando esta idea para nuestro ejemplo, lleva a la figura en la siguiente página. La figura muestra también los resultados de los cálculos en la tabla anterior. Nótese que la gráfica llega a su mínimo cuando $c = \overline{x} = 46$, tal como se especifica en la segunda propiedad.



La Propiedad 2 es una consecuencia directa de la Regla de Steiner.

Regla de Steiner

En cualquier muestra de n observaciones con valores $x_1, x_2, \ldots, x_n$ en una variable X, se cumple la siguiente igualdad:

$$\sum_{i=1}^{n}(x_i - c)^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2 + n\,(\overline{x} - c)^2$$

para cualquier $c \in \mathbb{R}$.

La Propiedad 2 de la media aritmética sigue directamente de la Regla de Steiner, ya que, si $c \neq \overline{x}$, entonces $n\,(\overline{x} - c)^2$ es estrictamente mayor de 0 y, por lo tanto, $\sum_{i=1}^{n}(x_i - c)^2 > \sum_{i=1}^{n}(x_i - \overline{x})^2$. Solo falta demostrar que la Regla de Steiner es correcta.

Comprobación de la Regla de Steiner

La primera propiedad se comprueba fácilmente usando las propiedades de sumas:

$$\begin{aligned}
\sum_{i=1}^{n}(x_i-c)^2 &= \sum_{i=1}^{n}\left[(x_i-\overline{x})+(\overline{x}-c)\right]^2 \\
&= \sum_{i=1}^{n}\left[(x_i-\overline{x})^2+2(x_i-\overline{x})(\overline{x}-c)+(\overline{x}-c)^2\right] \\
&= \sum_{i=1}^{n}(x_i-\overline{x})^2 \;+\; \sum_{i=1}^{n}2(x_i-\overline{x})(\overline{x}-c) \;+\; \sum_{i=1}^{n}(\overline{x}-c)^2 \\
&= \sum_{i=1}^{n}(x_i-\overline{x})^2+2(\overline{x}-c)\underbrace{\sum_{i=1}^{n}(x_i-\overline{x})}_{\substack{=0\\ \text{(prop. 1)}}}+n\,(\overline{x}-c)^2 \\
&= \sum_{i=1}^{n}(x_i-\overline{x})^2+0+n\,(\overline{x}-c)^2 \qquad \blacksquare
\end{aligned}$$

Ejemplo

Se puede verificar la Regla de Steiner con los datos del ejemplo anterior. La siguiente tabla recopila la información sobre las sumas cuadráticas de la variable *Edad*:

c	$\sum_{i=1}^{n}(x_i-c)^2$	$=$	$\sum_{i=1}^{n}(x_i-\overline{x})^2$	$+$	$n\,(\overline{x}-c)^2$
46	4050	$=$	4050	$+$	$50\times(46-46)^2$
48	4250	$=$	4050	$+$	$50\times(46-48)^2$
45	4100	$=$	4050	$+$	$50\times(46-45)^2$
46.5	4062.5	$=$	4050	$+$	$50\times(46-46.5)^2$
46.1	4050.5	$=$	4050	$+$	$50\times(46-46.1)^2$
45.9	4050.5	$=$	4050	$+$	$50\times(46-45.9)^2$

3.1.3.3 Comparación de la media aritmética con la mediana

Aunque tanto la media como la mediana informan sobre la tendencia central, arrojan luz sobre diferentes aspectos de la distribución. La media y la mediana utilizan de forma distinta la información numérica de la variable, lo cual lleva a las siguientes dos consecuencias:

La media y la mediana requieren diferentes niveles de medición de las variables En la Sección 3.1.2, se indicó que para el cálculo de la mediana se puede utilizar tanto el procedimiento general para hallar cuantiles como un procedimiento específico para la mediana. Ambos métodos utilizan únicamente la información ordinal de los valores en la variable. El procedimiento general para cuantiles se apoya en la función de proporción acumulada; para conocer la proporción acumulada asociado con un valor x de X, solo se necesita saber cuántas observaciones tienen un valor menor o igual a x, es decir, únicamente se comparan los valores en términos de mayor y menor. También el método alternativo solo utiliza información ordinal: Hay que (a) ordenar los valores de menor a mayor y (b) determinar el valor que está en el centro. Para la media aritmética, por otro lado, no solo importa el orden de los valores, sino también su magnitud y las diferencias entre ellos. Esto se aclara al considerar que el cálculo de la media implica una suma de todos los valores.

Ejemplo

Consideremos la variable *Nivel Educativo* de los datos de la página 34, la cual asume cuatro valores ordenados en los datos. Para la siguiente tabla, se ha construido una variable X que representa los valores numéricos de la variable *Educ* original. La tabla presenta también la proporción y proporción acumulada de X, lo cual nos facilita calcular la media y la mediana:

Nivel Educativo	X	p	F
Primaria	1	.08	.08
Secundaria	2	.36	.44
Bachillerato	3	.32	.76
Universitario	4	.24	1.00

A partir de esta tabla, se obtiene directamente la mediana y la media:

$$\begin{aligned} Me_X &= 3 \\ \overline{x} &= .08 \times 1 + .36 \times 2 + .32 \times 3 + .24 \times 4 = 2.72 \end{aligned}$$

Para interpretar estos resultados, vale volver a la variable original: La mediana $Me_X = 3$ corresponde con “Bachillerato”, mientras que la media de $\overline{x} = 2.72$ corresponde con “un nivel intermedio de Secundaria y Bachillerato”.

La conversión de la variable *Educ* original a la variable númerica X ha sido algo arbitraria, ya que, en principio, es admisible cualquier valoración que respeta el orden entre los cuatro niveles de educación. (Recuérdese que, al considerar la variable *Nivel Educativo* como una variable de nivel ordinal, una transformación que respeta el orden original entre los valores no implica una pérdida de información.) Es decir, la definición de otra variable $X^\star$ con una valoración alternativa de los niveles de educación es igualmente válida:

Nivel Educativo	$X^\star$	p	F
Primaria	0	.08	.08
Secundaria	1	.36	.44
Bachillerato	4	.32	.76
Universitario	10	.24	1.00

Similar a los cálculos de la variable X, se puede calcular la mediana y la media de $X^\star$:

$$\begin{aligned} Me_{X^\star} &= 4 \\ \overline{x}^\star &= .08 \times 0 + .36 \times 1 + .32 \times 4 + .24 \times 10 = 4.04 \end{aligned}$$

Llegamos a las siguientes dos conclusiones importantes:

(a) la interpretación del valor obtenido para la mediana *no* ha cambiado: $Me_{X^\star} = 4$ corresponde con el nivel de "Bachillerato", tanto como $Me_X = 3$;

(b) la interpretación del resultado para la media *sí* ha cambiado: $\overline{x}^\star = 4.04$ equivale a un nivel (ligeramente) mayor que "Bachillerato" mientras que $\overline{x} = 2.72$ equivalió a "un nivel intermedio de Secundaria y Bachillerato".

El ejemplo ilustra que la interpretación del resultado para una mediana fundamentalmente no cambiará cuando se transforme una variable respetando el orden entre los valores. Por otro lado, el valor obtenido para la media sí tendrá una interpretación distinta. En otras palabras, la mediana solo utiliza información ordinal de la variable; la media utiliza más que esta información (también importan las diferencias entre los valores), por lo cual una media, estrictamente hablando, requiere variables de nivel intervalo.

La mediana es menos sensible a valores extremos que la media En el cálculo de la media, cada observación contribuye con el mismo peso, dicho de otra forma, la media trata igual a cada valor observado: El valor de cada observación en la variable simplemente se añade a la suma. Por otro lado, la mediana es más afectada por valores en el centro de la distribución; valores extremos muchas veces se pueden cambiar sin que se afecte la mediana. Se dice que la mediana, en comparación con la media, es *más robusta* contra valores extremos o valores atípicos.

Ejemplo

En el Tema 2 (página 47), ya se señaló que la observación $i = 40$ en los datos de la página 34 tiene un valor extremo para *Índice de Masa Corporal*. Se puede verificar que la mediana y la media de esta variable, incluyendo las 50 observaciones, es igual a:

$$\begin{aligned} Me_{IMC} &= 24 \\ \overline{IMC} &= 25 \end{aligned}$$

Sin embargo, si se quita esta observación extrema de los datos y se calcula la mediana y la media para las 49 observaciones restantes, se obtiene:

$$\begin{aligned} Me_{IMC} &= 24 \\ \overline{IMC} &= 24.65 \end{aligned}$$

En otras palabras, al eliminar el valor extremo, la media bajó con .35, mientras que la mediana no resultó afectada.

Considerando de nuevo las 50 observaciones en la variable *Índice de Masa Corporal*, ¿qué pasaría con la media y la mediana si la persona que capturó los datos se equivocó al introducir el valor de la última observación en esta variable: en vez de capturar 26, capturó 76? Se puede verificar que la mediana seguiría siendo 24, mientras que la media incrementaría un punto completo a 26.

Consideremos una pequeña asociación de médicos y paramédicos que dan consultas a pacientes con varios tipos de problemas. En la asociación trabajan 8 personas. Sigue una lista con la función de cada quien y su sueldo mensual:

secretaria/administradora	\$ 9,000	médico familiar	\$ 15,000
enfermera	\$ 11,000	dentista	\$ 19,000
pedagoga	\$ 12,000	oftalmólogo	\$ 19,000
psicoterapeuta	\$ 14,000	pediatra	\$ 85,000

El pediatra es el presidente de la asociación y, además, es director de una clínica privada. La media de los sueldos es

$$\frac{9,000 + 11,000 + 12,000 + 14,000 + 15,000 + 19,000 + 19,000 + 85,000}{8} = 23,000$$

pesos. El sueldo mediano de 14500 pesos, sin embargo, parece resumir mejor el sueldo típico en esta asociación. En ocasiones como esta puede ser preferible calcular la mediana en vez de la media.

La conclusión de los ejemplos anteriores no debe ser que, en caso de la presencia de valores extremos, siempre sea preferible la mediana a la media. Es necesario conocer bien las características de ambas y tener claro cuál es el aspecto de la distribución que más interesa. El estadístico relevante se debe eligir con fundamento en estas consideraciones. En cierto modo, se recomienda calcular ambas, la media y la mediana, ya que cada una expresa un aspecto diferente de la tendencia central de la distribución.

3.1.4 Otros estadísticos de tendencia central

3.1.4.1 Media geométrica

Media geométrica

Para una variable X, cuyos valores $x_1, x_2, \ldots, x_n$ en n observaciones son estrictamente mayores que 0, la *media geométrica*, representada por g_X, se define como:

$$g_X = \left[\prod_{i=1}^{n} x_i\right]^{\frac{1}{n}}, \tag{3.3a}$$

o de forma equivalente,

$$g_X = \operatorname{antilog}\left(\frac{\sum_{i=1}^{n} \log x_i}{n}\right). \tag{3.3b}$$

Si existe alguna observación i, para la cual $x_i \leqslant 0$, entonces la media geométrica es indefinida.

Mientras que el cálculo de la media aritmética implica sumar todos los valores y dividir la suma entre n, la media geométrica implica multiplicar todos los valores y sacar la raiz n-ésima del producto. La fórmula en la Ecuación (3.3b) da el mismo resultado que la Ecuación (3.3a) (una propiedad que no comprobaremos, pero que sigue directamente de las propiedades de los logaritmos). Nótese que la fórmula no especifica la base de los logaritmos; no es necesario ya que se puede utilizar cualquier base (positiva), siempre y cuando el log y antilog utilizan la misma base. Si, por ejemplo, se utiliza $\log_{10} x_i$ entonces el antilog x debe corresponder con 10^x.

Ejemplo

Supongamos que hemos observado los siguientes seis valores en una variable X:

$$x_1 = 1 \qquad x_2 = 2 \qquad x_3 = 4 \qquad x_4 = 16 \qquad x_5 = 32 \qquad x_6 = 64.$$

La media geométrica de estos seis valores es:

$$\begin{aligned} g_X &= (x_1 \times x_2 \times x_3 \times x_4 \times x_5 \times x_6)^{\frac{1}{6}} \\ &= (1 \times 2 \times 4 \times 16 \times 32 \times 64)^{\frac{1}{6}} \\ &= \sqrt[6]{262144} \\ &= 8. \end{aligned}$$

Para fines de comparación, la media aritmética es:

$$\begin{aligned} \overline{x} &= \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{6} \\ &= \frac{1 + 2 + 4 + 16 + 32 + 64}{6} \\ &= \frac{119}{6} \\ &= 19.83. \end{aligned}$$

A menos de que el número de observaciones sea pequeño (como en el ejemplo anterior), la Fórmula (3.3a) es difícil para aplicar, dado que el producto de los valores crece muy rápido. En este caso es útil la Fórmula (3.3b). Calculemos la media geométrica del *Índice de Masa Corporal* en los datos de la página 34: Primero calculamos la suma de los logaritmos de los valores observados:

$$\begin{aligned} \sum_{i=1}^{50} \log x_i &= \log x_1 + \log x_2 + \log x_3 + \cdots + \log x_{49} + \log x_{50} \\ &= \log 21 + \log 25 + \log 27 + \cdots + \log 26 + \log 26, \end{aligned}$$

lo cual, utilizando el logaritmo con base 10 ($\log_{10}$), nos da

$$
\begin{aligned}
&= 1.322 + 1.398 + 1.431 + \cdots + 1.415 + 1.415 \\
&= 69.650.
\end{aligned}
$$

Como siguiente paso, se divide esta suma entre 50, el número de observaciones:

$$\frac{\sum_{i=1}^{50} \log x_i}{50} = \frac{69.650}{50} = 1.393$$

Finalmente, se toma el antilogaritmo de ambos lados:

$$
\begin{aligned}
g_X = \text{antilog}\left(\frac{\sum_{i=1}^{50} \log x_i}{50}\right) &= \text{antilog } 1.393 \\
&= 10^{1.393} \\
&= 24.717
\end{aligned}
$$

En los dos ejemplos anteriores se observa que la media geométrica es menor que la media aritmética. Esto refleja una propiedad más general: Para cualquier conjunto de n valores $x_1, x_2, \ldots, x_n$, la media geométrica (si es definida) siempre es menor que o igual a la media aritmética. El caso de igualdad sólo se cumple si todos los valores son iguales, es decir, si $x_1 = x_2 = \cdots = x_n$.

La media geométrica, en comparación con la media aritmética, da menos importancia a los valores más altos (lo cual lleva la consecuencia mencionada que, en general, es menor que la media aritmética). Por ejemplo, la media geométrica de los dos valores 10 y 1000 es 100, es decir, si se calcula la media geométrica de estos valores, se considera que la tendencia central es 100. En otra palabras, la media geométrica considera la distancia entre 10 y 100 la misma que la distancia entre 100 y 1000.

En consecuencia, el uso de la media geométrica implica que el nivel de la medición de la variable no se considera de intervalo, ya que, en variables de este tipo, las diferencias entre los valores significan lo mismo en cualquier parte de la escala. En variables de tipo intervalo, la distancia entre 10 y 100 es (diez veces) menor que la distancia entre 100 y 1000. De hecho, la media geométrica se usa cuando se considera que la escala de medición es logarítmica. Para interpretar las distancias entre los valores en una escala logarítmica, hay que mirar a las diferencias entre los *logaritmos* de los valores. Así, la distancia entre 10 y 100 se considera igual a la distancia entre 100 y 1000, ya que $\log 100 - \log 10 = \log 1000 - \log 100$. Es justamente esta relación con la escala logarítmica que se manifiesta en la Ecuación (3.3b). Esta ecuación permite interpretar *el logaritmo de la media geométrica* como la media aritmética de los valores después de que esos se han transformado a una escala logarítmica. Para una discussión sobre transformaciones, véase el Tema 4.

Ejemplo

Muchas veces es conveniente considerar variables que son tasas o razones como variables de escala logarítmica. El siguiente ejemplo justifica la preferencia de la media geométrica sobre la media aritmética en el caso de una variable que se defina como una razón.

En un exámen clínico de la sangre, uno de los parámetros que comúnmente se calcula es la relación entre la albúmina y globulina. Supongamos que se tiene esta información de 9 pacientes:

	Albúmina (g/dL)	Globulina (g/dL)	Rel. A/G	Rel. G/A
i	x_i	y_i	z_i	z'_i
1	5.05	2.71	1.86	0.54
2	3.00	2.66	1.12	0.89
3	5.09	2.79	1.83	0.55
4	3.58	3.27	1.10	0.91
5	4.59	2.68	1.71	0.58
6	3.95	3.10	1.27	0.79
7	3.46	2.49	1.39	0.72
8	3.42	2.96	1.15	0.87
9	3.80	2.54	1.50	0.67
Media aritmética	3.99	2.80	1.44	0.72
Media geométrica	3.93	2.79	1.41	0.71

Comparando la media aritmética y la media geométrica de las variables, se observa que hay pocas diferencias. Sin embargo, dado que la variable Z es la razón de las variables X y Y, es teóricamente mejor presentar la media geométrica para informar sobre la tendencia central de Z. El ejemplo ilustra claramente que, por un lado,

$$g_Z = \frac{g_X}{g_Y} \qquad \Big(1.41 = \frac{3.93}{2.79}\Big),$$

mientras que

$$\overline{z} \neq \frac{\overline{x}}{\overline{y}} \qquad \Big(1.44 \neq \frac{3.99}{2.80}\Big).$$

Es decir, la media geométrica de la razón entre albúmina y globulina es la razón de las medias geométricas de estas dos variables. Con la media aritmética, una relación similar no existe.

Una forma similar para entender lo mismo es comparar las variables Z y Z', las cuales son recíprocas, es decir:

$$Z = \frac{1}{Z'}$$

Comparando las medias geométricas, se observa que:

$$g_Z = \frac{1}{g_{Z'}} \qquad \Big(1.41 = \frac{1}{0.71}\Big),$$

mientras que la relación análoga para las medias aritméticas no se cumple:

$$\overline{z} \neq \frac{1}{\overline{z'}} \qquad \Big(1.44 \neq \frac{1}{0.72}\Big).$$

3.1.4.2 Media armónica

Media armónica

Para una variable X, cuyos valores $x_1, x_2, \ldots, x_n$ en n observaciones son estrictamente mayores que 0, la *media armónica*, representada por h_X, se define como:

$$h_X = \frac{n}{\displaystyle\sum_{i=1}^{n} \frac{1}{x_i}}. \tag{3.4}$$

Si existe alguna observación i, para la cual $x_i \leqslant 0$, entonces la media armónica es indefinida.

En palabras, la definición dice que la media armónica es la inversa de la media aritmética de los valores inversos. La media armónica, tal como la media geométrica, da relativamente menos peso a los valores altos. Es decir, las observaciones que tienen un valor alto en comparación con las otras observaciones afectan menos el valor de la media armónica.

Ejemplo

Calculemos la media armónica de la variable *Índice de Masa Corporal* de los datos de la página 34. Primero calculamos el valor del denominador de la Fórmula (3.4):

$$\begin{aligned}
\sum_{i=1}^{50} \frac{1}{x_i} &= \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \cdots + \frac{1}{x_{50}} \\
&= \frac{1}{21} + \frac{1}{25} + \frac{1}{27} + \cdots + \frac{1}{26} \\
&= 0.0476 + 0.0400 + 0.0370 + \cdots + 0.0385 \\
&= 2.0441
\end{aligned}$$

lo cual lleva a la media armónica:

$$\begin{aligned}
h_X = \frac{50}{\displaystyle\sum_{i=1}^{50} \frac{1}{x_i}} &= \frac{50}{2.0441} \\
&= 24.46
\end{aligned}$$

3.1.4.3 Estadísticos robustos para la tendencia central

Como vimos anteriormente (página 76), la media aritmética es relativamente poco robusta para valores atípicos y cambios en los valores, aunque pequeños, directamente cambian el resultado para la media. La mediana, por otro lado, es bastante más robusta que la media, pero tiene la desventaja que usa relativamente poca información presente en los datos. Como compromiso, se han desarrollado algunos estadísticos de tendencia central que son más robustos que la media aritmética, pero que recogen más de la información presente en los datos que la mediana.

En la literatura se ha presentado una amplia gama de estadísticos robustos para la tendencia central. Aquí se presentan brevemente tres de estos estadísticos: la trimedia, la media recortada y la centrimedia (que resulta ser un caso especial de la media recortada).

Trimedia

Trimedia

Para una variable X, con valores $x_1, x_2, \ldots, x_n$ en n observaciones, la *trimedia* se define a partir de los tres cuartiles como:

$$\frac{Q_1 + 2Q_2 + Q_3}{4}.$$

Ejemplo

A partir de la función de proporción acumulada para el *Índice de Masa Corporal*, presentada en la tabla de la página 49, se puede calcular los tres cuartiles de esta variable:

$$\begin{aligned} Q_1 &= 22 \\ Q_2 &= 24 \;\; (= Me_X) \\ Q_3 &= 27. \end{aligned}$$

Por lo tanto, la trimedia de *IMC* es:

$$\frac{22 \,+\, 2 \times 24 \,+\, 27}{4} = 24.25.$$

Media recortada

Media recortada

Para una variable X, con valores $x_1, x_2, \ldots, x_n$ en n observaciones, la *media recortada al* $c\,\%$ es la media aritmética que se obtiene después de haber quitado el $c\,\%$ de valores más altos y el $c\,\%$ de valores más bajos.

El *corte* c es un número entre 0 y 50.

Existen diferentes medias recortadas, dependiendo del valor del corte c. Para calcular la media recortada correspondiente del corte c, se quitan el $c\,\%$ de la distribución al lado izquierdo y al lado derecho y, a continuación, se calcula la media aritmética de los valores restantes.

Ejemplo

La media recortada a 10 % para la variable *Índice de Masa Corporal* se obtiene como sigue: Primero, se ordenan las 50 observaciones según su valor en la variable *IMC*, es decir, tal que $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_{50}$. Considerando que el 10 % de 50 es 5, se eliminan las primeras 5 y las últimas 5 observaciones de la lista ordenada. De los 40 valores que quedan, se calcula la media aritmética:

$\cancel{x_1 = 19}$	$x_{11} = 22$	$x_{21} = 24$	$x_{31} = 25$	$x_{41} = 27$
$\cancel{x_2 = 19}$	$x_{12} = 22$	$x_{22} = 24$	$x_{32} = 26$	$x_{42} = 28$
$\cancel{x_3 = 20}$	$x_{13} = 22$	$x_{23} = 24$	$x_{33} = 26$	$x_{43} = 28$
$\cancel{x_4 = 20}$	$x_{14} = 23$	$x_{24} = 24$	$x_{34} = 26$	$x_{44} = 28$
$\cancel{x_5 = 20}$	$x_{15} = 23$	$x_{25} = 24$	$x_{35} = 26$	$x_{45} = 29$
$x_6 = 21$	$x_{16} = 23$	$x_{26} = 24$	$x_{36} = 26$	$\cancel{x_{46} = 29}$
$x_7 = 22$	$x_{17} = 23$	$x_{27} = 25$	$x_{37} = 27$	$\cancel{x_{47} = 30}$
$x_8 = 22$	$x_{18} = 23$	$x_{28} = 25$	$x_{38} = 27$	$\cancel{x_{48} = 33}$
$x_9 = 22$	$x_{19} = 23$	$x_{29} = 25$	$x_{39} = 27$	$\cancel{x_{49} = 34}$
$x_{10} = 22$	$x_{20} = 24$	$x_{30} = 25$	$x_{40} = 27$	$\cancel{x_{50} = 42}$

Entonces, la media recortada al 10 % es igual a:

$$\begin{aligned}
\frac{\sum_{i=6}^{45} x_i}{40} &= \frac{x_6 + x_7 + x_8 + x_9 + x_{10} + \cdots + x_{43} + x_{44} + x_{45}}{40} \\
&= \frac{21 + 22 + 22 + 22 + 22 + \cdots + 28 + 28 + 29}{40} \\
&= \frac{984}{40} = 24.6
\end{aligned}$$

Centrimedia

Centrimedia

La *centrimedia* se define como la media recortada a 25 %.

Se puede interpretar la centrimedia como la media aritmética de los 50 % de valores que se encuentran en el centro de la distribución.

Ejemplo

Para obtener la centrimedia de la variable *Índice de Masa Corporal*, hay que descartar los 13 (aproximadamente 25 % de 50) valores más bajos y los 13 valores más altos y calcular la media aritmética sobre los 24 valores que sobran. Nótese que, si el 25 % de n no es un número entero, es necesario redondear; sin embargo, siempre se quitan el mismo número de observaciones a ambos lados de la distribución.

$\cancel{x_1 = 19}$	$\cancel{x_{11} = 22}$	$x_{21} = 24$	$x_{31} = 25$	$\cancel{x_{41} = 27}$
$\cancel{x_2 = 19}$	$\cancel{x_{12} = 22}$	$x_{22} = 24$	$x_{32} = 26$	$\cancel{x_{42} = 28}$
$\cancel{x_3 = 20}$	$\cancel{x_{13} = 22}$	$x_{23} = 24$	$x_{33} = 26$	$\cancel{x_{43} = 28}$
$\cancel{x_4 = 20}$	$x_{14} = 23$	$x_{24} = 24$	$x_{34} = 26$	$\cancel{x_{44} = 28}$
$\cancel{x_5 = 20}$	$x_{15} = 23$	$x_{25} = 24$	$x_{35} = 26$	$\cancel{x_{45} = 29}$
$\cancel{x_6 = 21}$	$x_{16} = 23$	$x_{26} = 24$	$x_{36} = 26$	$\cancel{x_{46} = 29}$
$\cancel{x_7 = 22}$	$x_{17} = 23$	$x_{27} = 25$	$x_{37} = 27$	$\cancel{x_{47} = 30}$
$\cancel{x_8 = 22}$	$x_{18} = 23$	$x_{28} = 25$	$\cancel{x_{38} = 27}$	$\cancel{x_{48} = 33}$
$\cancel{x_9 = 22}$	$x_{19} = 23$	$x_{29} = 25$	$\cancel{x_{39} = 27}$	$\cancel{x_{49} = 34}$
$\cancel{x_{10} = 22}$	$x_{20} = 24$	$x_{30} = 25$	$\cancel{x_{40} = 27}$	$\cancel{x_{50} = 42}$

Entonces, la centrimedia es:

$$\begin{aligned}\frac{\sum_{i=14}^{37} x_i}{24} &= \frac{x_{14} + x_{15} + \cdots + x_{36} + x_{37}}{24} \\ &= \frac{23 + 23 + \cdots + 26 + 27}{24} \\ &= \frac{588}{24} = 24.5\end{aligned}$$

3.2 Estadísticos de variabilidad o dispersión

3.2.1 Amplitud total

Definición

Para n observaciones que tienen valores $x_1, x_2, \ldots, x_n$ en una variable X, la *amplitud total* se define como la diferencia entre el valor máximo y el valor mínimo que se ha observado en esta variable. En símbolos:

$$\text{Amplitud total} = \max_i x_i - \min_i x_i,$$

donde $\max_i x_i$ y $\min_i x_i$ son el valor más alto y el valor más bajo, respectivamente, de todos los valores x_i observados en la variable X.

Ejemplo

Para calcular la amplitud total de la variable *Presión Arterial Sistólica* en los datos de la página 34, se buscan primero las observaciones con los valores extremos en esta variable; encontramos que son la persona 19, con $PAS = 107$, y la persona 8, con $PAS = 177$. Por lo tanto,

$$\begin{aligned}\text{Amplitud total} &= \max_i x_i - \min_i x_i \\ &= 177 - 107 = 70.\end{aligned}$$

Es evidente que la amplitud total es una medida de variabilidad: Si la amplitud es 0, entonces hay ausencia total de variabilidad (no hay diferencia entre el valor máximo y el valor mínimo lo cual implica que todas las observaciones tienen el mismo valor en esta variable) y cuanto mayor es, hay más diferencia entre el valor máximo y el valor mínimo. Al mismo tiempo, es importante darse cuenta de que la amplitud total:

(a) utiliza muy poca información presente en los datos: Únicamente utiliza la diferencia entre el valor más grande y el valor más pequeño; todas las demás diferencias entre pares de valores que se pueden considerar ni se consideran al calcular la amplitud total; y

(b) es muy sensible a valores atípicos, dado que se calcula precisamente a partir de los valores más extremos en la muestra.

3.2.2 Amplitud intercuartil

Definición

La *amplitud intercuartil* de una variable X es la diferencia entre el tercer cuartil y el primer cuartil en esta variable. En símbolos:

$$\text{Amplitud intercuartil} = Q_3 - Q_1.$$

En el caso de que, al calcular los cuartiles, se encuentre frente a la situación de que Q_1 o Q_3 (o ambos) corresponde con un *intervalo de valores*, se procede de la misma forma como con la mediana: Es decir, siempre se calcula un valor único para estos estadísticos (el punto medio del intervalo de valores que cumplen la definición del cuantil buscado).

Ejemplo

A partir de la tabla de frecuencias y proporciones, se hallan el primer y el tercer cuartil de la variable *Presión Arterial Sistólica* en los datos de la página 34:

	x_j	$frec(x_j)$	$p(x_j)$	$cfrec(x_j)$	$F(x_j)$
	107	1	.02	1	.02
	108	1	.02	2	.04
	109	1	.02	3	.06
	111	1	.02	4	.08
	112	3	.06	7	.14
	113	1	.02	8	.16
	114	4	.08	12	.24
$Q_1 = 116$	116	4	.08	16	.32
	120	1	.02	17	.34
	121	2	.04	19	.38
	122	1	.02	20	.40
	126	1	.02	21	.42
	127	3	.06	24	.48
	128	3	.06	27	.54
	129	4	.08	31	.62
	132	2	.04	33	.66
	133	1	.02	34	.68
	134	1	.02	35	.70
	138	2	.04	37	.74
$Q_3 = 142$	142	1	.02	38	.76
	144	3	.06	41	.82
	145	1	.02	42	.84
	147	1	.02	43	.86
	148	2	.04	45	.90
	159	1	.02	46	.92
	164	2	.04	48	.96
	172	1	.02	49	.98
	177	1	.02	50	1.00

Esto lleva a:

$$\begin{aligned}\text{Amplitud intercuartil} &= Q_3 - Q_1 \\ &= 142 - 116 \\ &= 26\end{aligned}$$

La amplitud intercuartil se puede interpretar como la amplitud que se obtendría considerando sólo el 50 % central de las observaciones (análogo a la interpretación de la centrimedia, p. 83). Debe ser claro que la amplitud intercuartil es menos sensible a valores atípicos que la amplitud total.

Por otro lado, nótese que, contrario a la amplitud total (y los demás estadísticos de variabilidad presentados en este tema), un valor de 0 para la amplitud intercuartil *no* implica que todos los valores son iguales.

3.2.3 Varianza y desviación estándar

3.2.3.1 La varianza: definición, cálculo e interpretación

En la búsqueda de estadísticos de variabilidad se han considerado varias medidas que resumen, de una forma más directa, la dispersión de los valores observados alrededor de la media aritmética $\overline{x}$. En particular, se utiliza la desviación entre cada valor y la media: la puntuación diferencial. Un asunto importante al respecto es cómo tratar a estas puntuaciones diferenciales. Consideremos las siguientes posibilidades:

1. Calcular la media de las puntuaciones diferenciales:

$$\frac{\sum_{i=1}^{n}(x_i - \overline{x})}{n}$$

   Aunque, a primera vista, es la opción más natural, esta forma para cuantificar las diferencias hacia la media no servirá: La primera propiedad de la media implica que la suma (y por lo tanto también la media) de estas desviaciones siempre es 0. En otras palabras, esta primera posibilidad no resulta útil para construir un estadístico de variabilidad.

2. Calcular la media del *valor absoluto* de las puntuaciones diferenciales:

$$\frac{\sum_{i=1}^{n}|x_i - \overline{x}|}{n}.$$

   Este estadístico se conoce como la *desviación media* de la variable X. No obstante, la desviación media es utilizada muy pocas veces en la estadística. La razón principal *no* es que sea inválido como estadístico de variabilidad, sino que matemáticamente funciones con valores absolutos involucrados conllevan propiedades menos elegantes (por ejemplo, no son diferenciables para todos los valores del dominio).

3. Calcular la media de las puntuaciones diferenciales *elevadas al cuadrado*, lo cual nos lleva al siguiente estadístico muy importante:

Varianza

Para n observaciones que tienen valores en una variable X, la *varianza* se define como:

$$s_X^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} \tag{3.5}$$

donde x_i es el valor que tiene la observación i en la variable X.

Ejemplo

Calculemos la varianza de la variable *Edad* de los datos de la página 34. Anteriormente se encontró que la media de esta variable era 46. Entonces, el numerador en la fórmula para la varianza es igual a:

$$\begin{aligned}\sum_{i=1}^{50}(x_i-\overline{x})^2 &= \sum_{i=1}^{50}(x_i-46)^2 \\ &= (58-46)^2+(55-46)^2+(38-46)^2+\cdots+(29-46)^2+(49-46)^2 \\ &= 4050.\end{aligned}$$

Nótese que esta misma suma ya se calculó en la página 72, cuando se ilustró la segunda propiedad de la media.

Para obtener la varianza, se divide el numerador entre el número total de observaciones:

$$\begin{aligned}s_X^2 &= \frac{\sum_{i=1}^{50}(x_i-\overline{x})^2}{50} \\ &= \frac{4050}{50} = 81.\end{aligned}$$

Similar a las definiciones equivalentes de la media aritmética, se puede definir la varianza a partir de las funciones de frecuencia o proporción:

Definiciones equivalentes de la varianza

Para (a) una variable X, que asume m distintos valores $x_1, x_2, x_3, \ldots, x_m$
y (b) un conjunto de n observaciones con valores en la variable X,

la *varianza* se define, a través de la función de frecuencia, como:

$$s_X^2 = \frac{\sum_{j=1}^{m} \mathit{frec}(x_j)\,(x_j-\overline{x})^2}{n} \tag{3.6a}$$

o, de forma equivalente, a través de la función de proporción:

$$s_X^2 = \sum_{j=1}^{m} p(x_j)\,(x_j-\overline{x})^2 \tag{3.6b}$$

Ejemplo

Utilizando la función de frecuencia de la variable Edad (representada en el ejemplo de la página 68), se puede calcular la varianza como sigue:

$$
\begin{aligned}
s_X^2 &= \frac{\sum_{j=1}^{27} \mathit{frec}(x_j)\,(x_j - 46)^2}{50} \\
&= \frac{2 \times (29-46)^2 + 1 \times (30-46)^2 + \ldots + 3 \times (58-46)^2 + 1 \times (59-46)^2}{50} \\
&= \frac{4050}{50} = 81
\end{aligned}
$$

o a través de la función de proporción:

$$
\begin{aligned}
s_X^2 &= \sum_{j=1}^{27} p(x_j)\,(x_j - 46)^2 \\
&= .04 \times (29-46)^2 + .02 \times (30-46)^2 + \ldots + .06 \times (58-46)^2 + .02 \times (59-46)^2 \\
&= 81
\end{aligned}
$$

La varianza, al igual que la media aritmética, es un estadístico muy importante. No obstante, hay que atribuir la importancia de la varianza más a su papel dentro de la estadística inferencial (como en el análisis de varianza, por ejemplo, el cual es un método que implica una comparación de las varianzas de diferentes grupos) que a su significado para la estadística descriptiva. La razón por la cual se usa pocas veces la varianza en la estadística descriptiva es que es difícil interpretar su valor. En particular, las unidades en las que la varianza expresa la variabilidad son las unidades originales *cuadráticas*, lo que impide una interpretación sencilla.

Ejemplo

Por ejemplo, encontramos que la varianza de la variable *Edad* era igual a 81. Sin embargo, no es claro qué significa. La media es igual a 46 y es correcto decir que la media aritmética de la edad es 46 *años*. De forma similar, se puede decir que la amplitud total y la amplitud intercuartil son de 30 *años* y 17 *años*, respectivamente. Sin embargo, respecto de la varianza, no se debe decir que iguala 81 *años*, sino 81 *años cuadrados*.

Por estas consideraciones, en la estadística descriptiva se extrae muchas veces la raíz cuadrada de la varianza, lo que nos lleva al estadístico de variabilidad que se introduce a continuación.

3.2.3.2 La desviación estándar: definición, cálculo e interpretación

Definición

La *desviación estándar* se define como la raíz cuadrada de la varianza:

$$\begin{aligned} s_X &= \sqrt{s_X^2} \\ &= \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}}. \end{aligned}$$

Ejemplo

Para conocer la desviación estándar de la variable *Edad* en los datos de la página 34, utilizamos el resultado para la varianza que ya se obtuvo anteriormente y sacamos la raíz cuadrada:

$$\begin{aligned} s_X &= \sqrt{81} \\ &= 9. \end{aligned}$$

Como la desviación estándar se expresa en las mismas unidades que los valores de la variable, es más fácil su interpretación. Así, se puede decir, en el ejemplo anterior, que la desviación estándar es 9 *años*. Se puede interpretar aproximadamente como "la distancia promedio entre los valores y la media".

Nota: El denominador en las fórmulas de la varianza y la desviación estándar

Muchos libros y casi todos los paquetes estadísticos suelen utilizar una definición ligeramente distinta para la varianza y la desviación estándar en comparación con las definiciones en esta sección. En particular, utilizan las definiciones:

$$\tilde{s}_X^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} \qquad \text{y} \qquad \tilde{s}_X = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}},$$

respectivamente. Es decir, en vez de dividir la suma de las puntuaciones diferenciales cuadráticas entre n, se divide entre $n-1$. Esta diferencia se relaciona con las propiedades de la varianza y la desviación estándar cuando se utilizan en la estadística inferencial (por ejemplo, para estimar la varianza y desviación estándar de una población a partir de los datos de una muestra de esta población). En este curso, se añadirá una tilde a la s de la varianza y de la desviación estándar si se hace la división entre $n-1$. A veces, $\tilde{s}_X^2$ se denomina *cuasivarianza*. En aplicaciones siempre hay que tener cuidado y dedicar atención para saber cuál es la definición que se ha utilizado en un caso particular. Por otro lado, si el número de observaciones n es grande, la diferencia entre s_X^2 y $\tilde{s}_X^2$ (y entre s_X y $\tilde{s}_X$) es despreciable.

3.2.3.3 Dos propiedades importantes de la varianza y la desviación estándar

Propiedad 1 de la varianza y desviación estándar

$$s_X^2 \geqslant 0$$

$$s_X \geqslant 0$$

La primera propiedad dice que la varianza y la desviación estándar *no pueden ser negativas*.

Comprobación de la Propiedad 1 de la varianza y desviación estándar

La propiedad sigue directamente de la definición de la varianza y la desviación estándar. La varianza como se encuentra definida en la página 87 es la razón de dos números positivos: En el nominador hay una suma que consiste en términos positivos (como cada término es un cuadrado, entonces es igual a o mayor que 0); el denominador es el número de observaciones, lo cual siempre es (estrictamente) mayor que 0. Por lo tanto, la varianza siempre es mayor o igual a 0.

La desviación estándar, por definición, es la raiz cuadrada de la varianza. La raiz cuadrada siempre es un número positivo, y por lo tanto, la desviación estándar siempre es mayor que o igual a 0. ■

De la prueba anterior (es decir, analizando las definiciones), es directamente claro que la varianza y la desviación estándar asumen un valor de 0 si y solo si todas las observaciones tienen el mismo valor. (Todos los valores deben ser iguales a la media para que todos los cuadrados que componen la suma en el numerador sean iguales a 0.)

Propiedad 2 de la varianza

Para n observaciones que tienen valores $x_1, x_2, \ldots, x_n$ en una variable X, siempre se cumple que:

$$2\, s_X^2 = \frac{\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2}{n^2} \tag{3.7}$$

Comprobación de la Propiedad 2 de la varianza

Consideremos la doble suma en el numerador al lado derecho de la igualdad. Como se explicó en el primer tema (p. 5), en una expresión con doble signo sumatorio se puede intercambiar los dos signos sumatorios (si, como es el caso aquí, los límites de la segunda suma no incluyen el índice de la primera suma):

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2 \;=\; \sum_{j=1}^{n}\left[\sum_{i=1}^{n}(x_i - x_j)^2\right] \tag{3.7a}$$

Como primer paso, desarrollamos la suma entre los corchetes:

$$\begin{aligned}\sum_{i=1}^{n}(x_i - x_j)^2 \;&=\; \sum_{i=1}^{n}(x_i - \overline{x})^2 \;+\; n(\overline{x} - x_j)^2 \\ &=\; n\,\underbrace{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}}_{=s_X^2} + n(\overline{x} - x_j)^2 \\ &=\; ns_X^2 \;+\; n(\overline{x} - x_j)^2\end{aligned} \tag{3.7b}$$

La igualdad en la Ecuación (3.7b) resulta de la aplicación de la Regla de Steiner (p. 73). Pues, x_j en la suma del lado izquierdo es una constante (es decir, x_j representa en cada término de la suma el mismo número, ya que el índice sumatorio es i). Es decir, se aplica la Regla de Steiner con $c = x_j$.

Como siguiente paso, sustituimos el resultado obtenido en la Ecuación (3.7a):

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2 \;&=\; \sum_{j=1}^{n}\left[ns_X^2 + n(\overline{x} - x_j)^2\right] \\ &=\; \sum_{j=1}^{n}(ns_X^2) \;+\; \sum_{j=1}^{n}\left[n(\overline{x} - x_j)^2\right] \\ &=\; n^2 s_X^2 \;+\; n\sum_{j=1}^{n}\left[(x_j - \overline{x})^2\right] \\ &=\; n^2 s_X^2 \;+\; n^2\,\frac{\sum_{j=1}^{n}(x_j - \overline{x})^2}{n} \\ &=\; n^2 s_X^2 \;+\; n^2 s_X^2 \\ &=\; 2n^2 s_X^2,\end{aligned}$$

de lo cual inmediatamente sigue la Propiedad 2. ■

Los términos de la suma doble en el numerador de la fracción al lado derecho de la Ecuación (3.7) son las diferencias cuadráticas entre pares de valores en la variable X. El número de términos en la suma es $n \times n = n^2$. Si no se usara símbolos sumatorios, la suma en el numerador se escribiría como:

$$\begin{array}{rcccccccccc}
\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2 & = & (x_1-x_1)^2 & + & (x_1-x_2)^2 & + & (x_1-x_3)^2 & + & \dots & + & (x_1-x_n)^2 \\
 & + & (x_2-x_1)^2 & + & (x_2-x_2)^2 & + & (x_2-x_3)^2 & + & \dots & + & (x_2-x_n)^2 \\
 & + & (x_3-x_1)^2 & + & (x_3-x_2)^2 & + & (x_3-x_3)^2 & + & \dots & + & (x_3-x_n)^2 \\
 & & \vdots & & \vdots & & \vdots & & \ddots & & \vdots \\
 & + & (x_n-x_1)^2 & + & (x_n-x_2)^2 & + & (x_n-x_3)^2 & + & \dots & + & (x_n-x_n)^2.
\end{array}$$

Por lo tanto, esta suma dividida entre n^2 se puede interpretar como la media de las diferencias cuadráticas entre los pares de observaciones (recuérdese que una media siempre es "una suma dividida entre el número de términos en la suma"; en este caso, el "número de términos" es n^2). Lo que dice la Propiedad 2, entonces, es que la media de las diferencias cuadráticas entre los pares de observaciones es directamente proporcional a—y más precisamente, siempre es el doble de—la varianza.

Este resultado nos permite interpretar la varianza (y la desviación estándar) como una medida que expresa cuántas diferencias hay *entre (pares de) observaciones*, aunque su definición común es en términos de diferencias *entre cada valor y la media*.

Ejemplo

Consideremos la variable *Presión Arterial Sistólica* en los datos de la página 34. Por fines de ilustración (y para no tener un número demasiado de términos), ilustraremos la Propiedad 2 únicamente utilizando los datos de las tres personas con grupo sanguíneo AB. Es decir, consideramos una variable X, *Presión Arterial Sistólica*, con tres observaciones:

$$x_1 = 134 \qquad x_2 = 144 \qquad x_3 = 121.$$

Primero calculamos la varianza de esta variable con tres observaciones, utilizando la definición tradicional de la varianza en página 87. Para este cálculo, se requiere la media:

$$\overline{x} = \frac{134 + 144 + 121}{3} = 133$$

Sustituyendo la media en la fórmula para la varianza, se obtiene:

$$s_X^2 = \frac{(134-133)^2 + (144-133)^2 + (121-133)^2}{3} = \frac{266}{3} = 88.67$$

De forma alternativa, se puede calcular la media de las diferencias cuadráticas entre todos los pares de valores, como especifica la Propiedad 2. Calculamos primero la suma correspondiente (es decir, el numerador de la media):

$$\begin{aligned}\sum_{i=1}^{3}\sum_{j=1}^{3}(x_i - x_j)^2 &= (134-134)^2 + (134-144)^2 + (134-121)^2 \\ &\quad + (144-134)^2 + (144-144)^2 + (144-121)^2 \\ &\quad + (121-134)^2 + (121-144)^2 + (121-121)^2 \\ &= 1596.\end{aligned}$$

Dividir entre el número de términos ($3^2 = 9$) lleva a:

$$\frac{\sum_{i=1}^{3}\sum_{j=1}^{3}(x_i - x_j)^2}{3^2} = \frac{1596}{9} = 177.33.$$

Se verifica que, efectivamente:

$$\frac{\sum_{i=1}^{3}\sum_{j=1}^{3}(x_i - x_j)^2}{3^2} = 2s_X^2$$

$$(177.33 = 2 \times 88.67)$$

3.2.4 Coeficiente de variación

En algunas circunstancias, diferencias grandes entre las medias de dos distribuciones dificultan la comparación de la variabilidad de estas distribuciones. Consideremos el siguiente ejemplo:

Ejemplo

A 40 adolescentes de 18 años se les pregunta cuál fue su peso al nacer (variable X) y cuál es su peso actual (variable Y). Se calcula la media aritmética y la desviación estándar de ambas variables y se obtiene el siguiente resultado:

$$\overline{x} = 3.20 \text{ kg.} \qquad \overline{y} = 60 \text{ kg.}$$
$$s_x = 0.48 \text{ kg.} \qquad s_Y = 9 \text{ kg.}$$

¿Sería correcto concluir que estas personas difieren más entre sí ahora que son adolescentes que cuando recién nacieron?

Haciendo la comparación de forma *absoluta*, esta conclusión sí sería correcta: La desviación estándar en la distribución de los pesos actuales es mucho mayor que la de los pesos al nacer (y también lo serían la amplitud total y la amplitud intercuartil). Dicho de otra forma, ahora no es nada extraordinario que en este grupo de adolescentes uno difiera por 9 kg. de otro, mientras que entre recién nacidos es simplemente imposible observar tal diferencia. En este sentido, se puede decir que estas personas difieren más ahora que antes.

No obstante, el peso medio actual es muy distinto al peso medio al nacer y, según el sentido común, una diferencia de, por ejemplo, 2 kgs. entre dos bebés se considera grande, mientras que la misma diferencia entre dos adultos se considera pequeña. Por lo tanto, puede ser más informativo hacer la comparación de las dos distribuciones de forma *relativa*, considerando la variabilidad de los pesos en relación con la magnitud "común" —la tendencia central— del peso en los respectivos momentos. Por ejemplo, se puede considerar la desviación estándar en relación con la media aritmética, calculando la razón entre ambos:

$$\frac{s_X}{\overline{x}} = \frac{0.48}{3.2} = 0.15 \qquad \text{para el peso al nacer}$$

$$\text{y } \frac{s_Y}{\overline{y}} = \frac{9}{60} = 0.15 \qquad \text{para el peso actual.}$$

Observamos que la variabilidad *relativa* a la media es igual para el peso al nacer y el peso actual. A los 18 años, las diferencias entre estas personas respecto del peso son relativamente iguales a las diferencias a nacer.

Este ejemplo nos lleva a la siguiente definición:

Definición

Para una variable X, que asume valores positivos para todas las observaciones en una muestra, se define el *coeficiente de variación*, representado por CV_X, como:

$$CV_X = \frac{s_X}{\overline{x}}$$

Nótese que la definición excluye la posibilidad de que X asuma valores negativos. Muchas veces se multiplica el coeficiente de variación por 100 para tener un porcentaje. En el ejemplo anterior, el coeficiente de variación de ambas variables (peso al nacer y peso actual) es 15 %.

El coeficiente de variación es un estadístico "sin dimensión" (es decir, no tiene alguna unidad asociada; en el ejemplo anterior, tanto la media como la desviación estándar tienen como unidad kilogramos, pero esto no es el caso para el coeficiente de variación). Precisamente por eso, el coeficiente de variación permite la comparación de variables muy distintas, es decir, variables que no se han medido en la misma escala.

Comparando la desviación estándar y el coeficiente de variación, llegamos a la conclusión de que la desviación estándar se modifica cuando se cambia la unidad de la escala. Por ejemplo, la desviación estándar de una variable que es "peso en kilogramos" es diferente de la desviación estándar de la misma variable transformada "peso en gramos". El coeficiente de variación, por otro lado, sigue siendo el mismo si se cambia la unidad de la escala. Sin embargo, el coeficiente de variación es variable ante cambios del origen (o el punto cero) de la escala, mientras que un cambio del origen de la escala no afectará a la desviación estándar. En el siguiente tema, se investigará cómo transformaciones de variables afectan las medidas de tendencia central y variabilidad.

3.3 Relacionando la media aritmética y la desviación estándar: La desigualdad de Chebyshev

Desigualdad de Chebyshev

Para cualquier variable X que tiene valores en n observaciones de una muestra, y cualquier $k \in \mathbb{R}$ mayor que 0, se cumple que:

$$p(|X - \overline{x}| \geqslant ks_X) \leqslant \frac{1}{k^2} \tag{3.8a}$$

La prueba de la desigualdad de Chebyshev se dará al final de esta sección. Primero, analicemos su significado y aplicación.

El lado izquierdo de la desigualdad corresponde con: la proporción de observaciones cuya puntuación diferencial absoluta es mayor que o igual a k desviaciones estándares; en otras palabras: la proporción de observaciones que *no* están a menos de k desviaciones estándares de la media. La desigualdad de Chebyshev dice sobre esta proporción que no puede ser mayor que $1/k^2$.

En la estadística, la diferencia entre un valor y la media de una variable (es decir, la puntuación diferencial) muchas veces se interpreta en términos de desviaciones estándares. En el siguiente tema se introducirá el concepto de una variable estandarizada cuyo valor indica cuántas desviaciones estándares difiere de la media. Para entender mejor la desigualdad de Chebyshev es útil la siguiente formulación equivalente:

Desigualdad de Chebyshev: Formulación equivalente

Para cualquier variable X que tiene valores en n observaciones de una muestra, y cualquier $k \in \mathbb{R}$ mayor que 0, se cumple que:

$$p(\overline{x} - ks_X \;<\; X \;<\; \overline{x} + ks_X) \geqslant 1 - \frac{1}{k^2}. \tag{3.8b}$$

Equivalencia de las dos formulaciones de la desigualdad de Chebyshev

Efectivamente, la Ecuación (3.8b) es equivalente con la Ecuación (3.8a), como lo muestra la siguiente derivación:

$$\begin{aligned}
& p(|X - \overline{x}| \geqslant ks_X) \leqslant \frac{1}{k^2} && (3.8a)\\
\iff\quad & 1 - p(|X - \overline{x}| \geqslant ks_X) \geqslant 1 - \frac{1}{k^2} \\
\iff\quad & p(|X - \overline{x}| < ks_X) \geqslant 1 - \frac{1}{k^2} \\
\iff\quad & p(\overline{x} - ks_X \;<\; X < \; \overline{x} + ks_X) \geqslant 1 - \frac{1}{k^2} && (3.8b)
\end{aligned}$$

■

En palabras, la formulación de la desigualdad de Chebyshev en la Ecuación (3.8b) dice que la proporción de observaciones cuyo valor se encuentra a menos de k desviaciones estándares de la media (es decir, cuyo valor es mayor que la media menos k desviaciones pero menor que la media más k desviaciones) siempre es mayor que o igual a $1 - 1/k^2$.

Ilustración de la desigualdad de Chebyshev

Ilustramos la desigualdad de Chebyshev aplicando el teorema a la variable *Índice de Masa Corporal* de los datos de la página 34. Se puede verificar que calculando la media y la desviación estándar de esta variable se obtiene el siguiente resultado:

$$\overline{x} = 25 \qquad\qquad s_X = 4$$

La desigualdad de Chebyshev es válida para cualquier $k > 0$. Elijamos, como ejemplo, $k = 1.5$. Sustituyendo los resultados para la media y desviación estándar y nuestro valor de k en la Ecuación (3.8b) da:

$$p(25 - 1.5 \times 4 \ < \ X \ < \ 25 + 1.5 \times 4) \geqslant 1 - \frac{1}{1.5^2}$$

o bien:

$$\iff \qquad\qquad p(19 \ < \ X \ < \ 31) \geqslant .56.$$

Es decir, el 56 % o más de los valores en la variable *IMC* se encuentran entre 19 y 31 (estos límites *no* incluidos). Podemos verificar que esta afirmación es correcta: Nuestra muestra tiene 50 observaciones y contando el número de observaciones con un *IMC* mayor que 19 y menor que 31 (véanse los datos de la página 34), se obtiene que 45 observaciones, o el 90 %, cumplen esta condición, y dado que 90 % > 56 %, se verifica que la afirmación derivada de la desigualdad de Chebyshev es correcta.

Verifiquemos que utilizando la primera formulación (en la Ecuación (3.8a)) llegamos al mismo resultado (que resulta ser equivalente con el anterior):

$$p(|X - 25| \geqslant 1.5 \times 4) \ \leqslant \ \frac{1}{1.5^2}$$

lo cual resulta en

$$p(|X - 25| \geqslant 6) \ \leqslant \ .44$$

La proporción de observaciones para las cuales la diferencia absoluta con la media de 25 es mayor que o igual a 6 no es mayor que 44 %. Una observación tiene una diferencia absoluta de 6 o más con 25 si y solo si es igual o menor que 19 o igual o mayor que 31. En nuestro ejemplo, solo hay 5 observaciones en esta situación (a saber, las personas con no. 5, 31, 34, 40, 44), o bien el 10 % de la muestra, lo cual efectivamente no supera el 44 % como dice la desigualdad.

Nótese la equivalencia de las dos formulaciones de la desigualdad de Chebyshev: Decir que 56 % o más de las observaciones tienen un valor entre 19 y 31 (no incluidos) es precisamente lo mismo que decir que no más de 44 % de las observaciones tienen un valor menor que o igual a 19 o mayor que o igual a 31.

Se puede elegir otros valores para k y cada vez verificar que la proporción exacta efectivamente excede el límite inferior proporcionado por la Ecuación (3.8b). La siguiente tabla presenta los resultados para k igual a 1.25, 1.4, 2, 4 y 5 (siempre para la variable *Índice de*

Masa Corporal con media $\overline{x} = 25$ y desviación estándar $s_X = 4$):

k	$p(\overline{x} - ks_X < X < \overline{x} + ks_X)$	límite inferior por la Ec. (3.8b)	proporción exacta
1.25	$p(\ 20 < X < 30\ )$	$1 - \frac{1}{1.25^2} = .36$	$\frac{41}{50} = .82$
1.4	$p(\ 19.4 < X < 30.6\ )$	$1 - \frac{1}{1.4^2} = .49$	$\frac{45}{50} = .90$
2	$p(\ 17 < X < 33\ )$	$1 - \frac{1}{2^2} = .75$	$\frac{48}{50} = .96$
4	$p(\ 9 < X < 41\ )$	$1 - \frac{1}{4^2} = .94$	$\frac{49}{50} = .98$
5	$p(\ 5 < X < 45\ )$	$1 - \frac{1}{5^2} = .96$	$\frac{50}{50} = 1.00$

La desigualdad de Chebyshev se utiliza de varias maneras. Por ejemplo, la demostración de la ley de los grandes números, uno de los teoremas más importantes en la estadística, se apoya en la desigualdad de Chebyshev. Como se acaba de mostrar en el ejemplo anterior, la desigualdad sirve también para obtener un límite para el número o la proporción de observaciones que tienen el valor en la variable X dentro de algún intervalo de interés. La importancia de la desigualdad de Chebyshev es que proporciona este límite sin tener conocimiento completo de la distribución de la variable, únicamente a partir de la media y la desviación estándar. Normalmente se necesita conocer la función de frecuencia o de proporción para poder hacer afirmaciones sobre la proporción de valores en una variable que se encuentran dentro de algún intervalo.

Como veremos en la parte sobre estadística inferencial, es muy común derivar una estimación de la proporción de observaciones en un intervalo de interés a partir de las propiedades de la distribución normal (una distribución teórica en la cual, por ejemplo, alrededor de 95 % de los valores se encuentran a menos de dos desviaciones estándares de la media). La aplicación de la desigualdad de Chebyshev difiere de este método en que *no* hace supuestos adicionales sobre la distribución; es decir, la conclusión que se saca de la desigualdad de Chebyshev es válida *cualquiera que sea la distribución* de la variable considerada.

La demostración matemática de la desigualdad de Chebyshev se realizará en dos pasos. Primero, comprobaremos el siguiente teorema:

Teorema de Markov

Si en una variable X que tiene valores $x_1, x_2, \ldots, x_n$, en n observaciones, todos los valores satisfacen $x_i \geqslant 0$ $(i = 1, \ldots, n)$, entonces para cualquier número $c > 0$ se cumple la siguiente desigualdad:

$$p(X \geqslant c) \ \leqslant\ \frac{\overline{x}}{c}$$

Prueba del teorema de Markov

Reordenamos los n valores observados de tal forma que $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ y buscamos el índice j que cumple que $x_j < c \leqslant x_{j+1}$. Es decir, buscamos en la secuencia ordenada de valores $x_1, x_2, \ldots, x_n$, entre cuáles dos valores se encuentra c y asignamos a j el índice correspondiente al valor más bajo de estos dos. En el caso de que c sea menor o igual al valor mínimo observado (es decir, $c \leqslant x_1$), ponemos j igual a 0; similarmente, ponemos j igual a n en el caso de que c sea mayor que el valor máximo observado (es decir, cuando $x_n < c$). Ahora podemos derivar:

$$\begin{aligned}
\overline{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\
&= \frac{1}{n}\underbrace{\sum_{i=1}^{j} x_i}_{\geqslant 0} + \frac{1}{n}\underbrace{\sum_{i=j+1}^{n} x_i}_{\geqslant (n-j)\,c} \\
&\geqslant \quad 0 \quad + \frac{n-j}{n}\,c \\
&\geqslant p(X \geqslant c)\,c
\end{aligned} \tag{3.9}$$

De ahí, inmediatemente sigue el teorema de Markov. La suma $\sum_{i=1}^{j} x_i$ en la Ecuación (3.9) es la suma de los j valores más bajos (si $j = 0$, definimos que esta suma es 0). Dado que cada x_i es positivo, una suma de los x_i también es positiva. Por otro lado, $\sum_{i=j+1}^{n} x_i$ es la suma de los $n - j$ valores restantes (los más altos entonces) (también aquí definimos que si la suma no tiene ningún término (porque $j = n$), entonces es igual a 0). Dado que cada uno de estos $(n - j)$ valores es mayor que o igual a c, la suma es mayor que o igual a $(n - j)\,c$. ■

El teorema de Markov nos sirve para comprobar la desigualdad de Chebyshev:

Prueba de la desigualdad de Chebyshev

Primero, construimos una nueva variable $Y = (X - \overline{x})^2$, es decir, para cada observación asignamos el valor y_i en la variable Y aplicando la fórmula:

$$y_i = (x_i - \overline{x})^2.$$

Nótese que todos los valores en la variable Y son iguales a o mayores que 0. Por lo tanto, se cumple la condición para poder aplicar el teorema de Markov a la variable Y. Para el valor de c elegimos $k^2 s_X^2$. Así obtenemos:

$$p(Y \geqslant k^2 s_X^2) \leqslant \frac{\overline{y}}{k^2 s_X^2}. \tag{3.10a}$$

Elaboramos el término de la izquierda, recambiando Y por $(X - \overline{x})^2$:

$$\begin{aligned} p(Y \geqslant k^2 s_X^2) &= p\left[(X - \overline{x})^2 \geqslant k^2 s_X^2\right] \\ &= p\big(|X - \overline{x}| \geqslant k s_X\big). \end{aligned} \tag{3.10b}$$

De manera similar, elaboramos el término derecho de la Ec. (3.10a):

$$\begin{aligned} \frac{\overline{y}}{k^2 s_X^2} &= \frac{1}{k^2 s_X^2} \frac{\sum_{i=1}^{n} y_i}{n} \\ &= \frac{1}{k^2 s_X^2} \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n} \\ &= \frac{1}{k^2 s_X^2} \qquad s_X^2 \\ &= \frac{1}{k^2}. \end{aligned} \tag{3.10c}$$

Cambiar el término izquierdo y derecho en la Ec. (3.10a) por los resultados en las Ecuaciones (3.10b) y (3.10c), respectivamente, nos lleva directamente a la desigualdad de Chebyshev en la Ecuación (3.8a):

$$p\big(|X - \overline{x}| \geqslant k s_X\big) \leqslant \frac{1}{k^2} \qquad \blacksquare$$

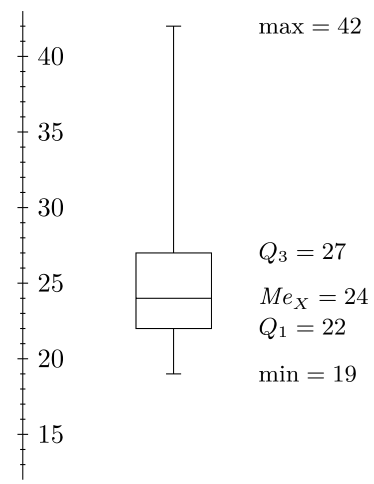
3.4 Representación gráfica de la información de tendencia central y variabilidad

La información sobre la tendencia central y la variabilidad de alguna distribución se puede representar en un *diagrama de caja*, lo cual se conoce también como *diagrama de caja y bigotes*. Aunque existen muchas variantes de este tipo de diagrama, en esta sección se presentan solo dos, como ejemplo para representar información sobre la variable *Índice de Masa Corporal*de los datos de la página 34. Debe ser claro que, al construir un diagrama de caja y bigotes, el investigador se puede sentir libre para añadir información adicional que considere importante.

Ejemplo

En la primera variante se representan (a) el valor mínimo, (b) el cuartil Q_1, (c) la mediana, (d) el cuartil Q_3 y (e) el valor máximo, tal como en la siguiente figura.

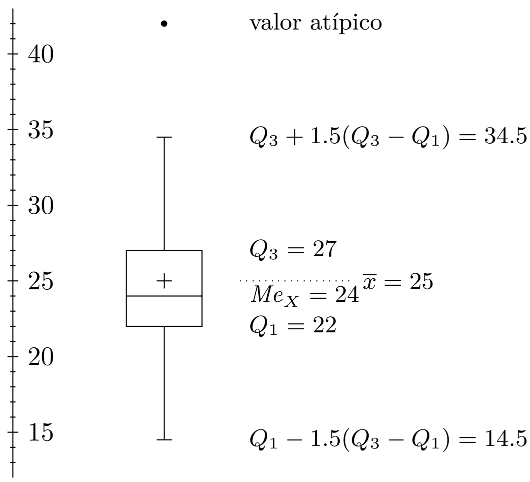
Diagrama de caja y bigotes (variante 1) para la variable *IMC*



De esta representación gráfica podemos leer, por ejemplo, que (a) (aproximadamente) el 50 % de las observaciones tienen un valor entre 22 y 27, (b) la mediana es igual a 24 y (c) hay observaciones que tienen el valor mínimo de 19 y el valor máximo de 42.

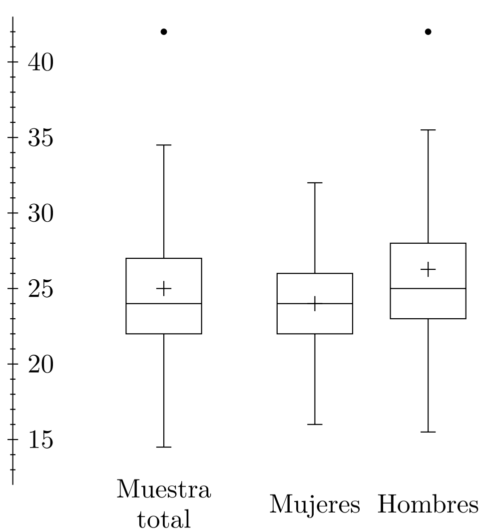
En la segunda variante, se incluye también la media aritmética en la gráfica y, en vez de representar el valor mínimo y máximo, se ponen puntos para los valores atípicos. En general, cuando uno desea representar los valores atípicos, se plantea el problema de cómo decidir cuáles valores son atípicos. Una regla común es considerar a todos los valores que están por debajo de $Q_1 - 1.5(Q_3 - Q_1)$ o por encima de $Q_3 + 1.5(Q_3 - Q_1)$ como atípicos. La siguiente figura ilustra la segunda variante del diagrama de caja y bigotes.

Diagrama de caja y bigotes (variante 2) para la variable *IMC*



Un diagrama de caja y bigotes se utiliza muchas veces para comparar distribuciones de una variable en diferentes muestras. Por ejemplo, en la siguiente figura se presenta el mismo diagrama que en la figura anterior, para la muestra total, y para las mujeres y los hombres, por separado.

Diagrama de caja y bigotes (variante 2) para la variable *IMC*
(muestra total, mujeres y hombres)



Problemas

1. Al inicio del año académico en la Facultad de Medicina, se les aplica un cuestionario a los estudiantes de nuevo ingreso, que sondea su grado de familiaridad con la computadora y otras herramientas tecnológicas (para darles un curso remedial a los que tengan escasas habilidades tecnológicas). El cuestionario consiste en 80 preguntas; cada una presenta una tarea de cómputo concreta que el estudiante contesta utilizando una escala tipo Likert con cuatro opciones de respuesta: (i) "No estoy *nada familiarizado/a* con esta tarea", (ii) "*Poco familiarizado/a*", (iii) "*Familiarizado/a*", (iv) "*Muy familiarizado/a*". La tabla que se presenta a continuación da, para cinco de los estudiantes, el número de veces que han contestado en cada categoría:

	Frecuencias de respuestas en cada categoría				
Categorías de Respuesta	Alumno A	Alumno B	Alumno C	Alumno D	Alumno E
Nada familiarizado/a	16	9	27	0	26
Poco familiarizado/a	25	4	24	3	15
Familiarizado/a	26	18	11	16	9
Muy familiarizado/a	13	49	18	61	30

   (a) ¿Cuál es la moda, la mediana y la media para cada estudiante?

   (b) ¿Cuáles de los estadísticos anteriores consideras adecuados para resumir la distribución de cada estudiante? Argumenta tu respuesta y aclara los supuestos que haces.

   (c) Si estos cinco alumnos formaran la muestra total, ¿cómo resumirías los datos disponibles para hacer una afirmación sobre el nivel global de las habilidades de cómputo en esta muestra?

2. A continuación, se presentan las calificaciones de 12 estudiantes de maestría en la asignatura de "Diseños de investigación":

	x_i		x_i		x_i
Alma	9	Erika	10	Ilona	8
Beatriz	8	Francesca	6	Julián	9
Carlos	9	Gerardo	6	Kime	6
David	7	Heidi	7	Leo	5

   (a) Representa la distribución de las calificaciones gráficamente a través de un histograma o un polígono de frecuencia.

   (b) Calcula la media y la mediana de la variable X. Utiliza para la mediana el método alternativo (p. 64).

   (c) ¿Cuál es la relación entre la media y la mediana? A partir de la representación gráfica, ¿esperaste esta relación?

3. La Propiedad 2 de la media dice que la media es la constante hacia la cual la suma de diferencias cuadráticas es mínima. Existe una propiedad similar de la mediana, según la cual en cualquier muestra se cumple que:

$$\sum_{i=1}^{n}|x_i - Me_X| < \sum_{i=1}^{n}|x_i - c|,$$

donde c es cualquier constante diferente de Me_X.

   (a) ¿Cómo formularías esta propiedad en palabras?

   (b) Verifica esta propiedad de la mediana utilizando la variable *Índice de Masa Corporal* de los datos de la página 34 en la submuestra de las 9 personas con grupo sanguíneo B. En particular, calcula

   $$\sum_{i=1}^{n}|x_i - c|$$

   para tres diferentes valores de c: la mediana, la media y un valor que se encuentra entre la mediana y la media.

4. Para examinar la esperanza de vida de personas que han recibido un trasplante de corazón, se han seguido 45 pacientes desde el momento del trasplante y se ha registrado (a) si todavía están vivos (variable X, 0 = fallecido, 1 = vivo) y (b) el número de meses que han vivido con el nuevo corazón (variable Y, para los que todavía viven se presenta el número de meses pasados desde la operación). Los datos[*] son los siguientes:

i	x_i	y_i	i	x_i	y_i	i	x_i	y_i	i	x_i	y_i	i	x_i	y_i
1	0	1	10	0	35	19	1	124	28	0	34	37	0	3
2	0	1	11	0	1	20	0	67	29	0	2	38	1	100
3	0	2	12	0	2	21	0	0	30	0	50	39	0	91
4	0	21	13	0	25	22	0	2	31	0	9	40	0	19
5	0	5	14	0	66	23	0	2	32	0	30	41	0	3
6	0	3	15	0	5	24	0	2	33	0	5	42	0	3
7	0	45	16	0	1	25	0	96	34	0	2	43	0	8
8	0	1	17	0	28	26	1	114	35	0	11	44	1	94
9	0	10	18	0	2	27	0	2	36	1	101	45	0	1

   (a) Elige un estadístico de tendencia central y otro de variabilidad para resumir la distribución de la variable "número de meses que un paciente vive después de un trasplante de corazón (hasta fallecer)" y argumenta tu elección.

   (b) Calcula el valor en estos estadísticos.

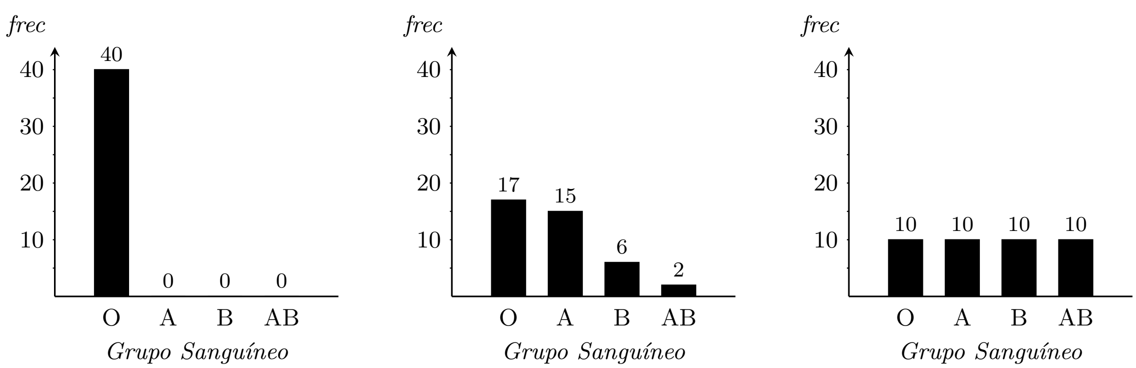
5. Evalúa cada estadístico de variabilidad definido en este tema respecto de (a) el nivel de medición que requiere para poder aplicarse de forma sensata y (b) su sensibilidad para valores atípicos. Incluye también la desviación media (introducida en la p. 87) en esta evaluación.

[*] Datos obtenidos de Miller, R. (1980). *Survival analysis*. New York: Wiley.

6. Si es posible, da un ejemplo de una distribución para la cual...
   - ... la amplitud total es igual a la amplitud intercuartil
   - ... la varianza es menor que la desviación típica
   - ... la amplitud intercuartil es mayor que la amplitud total
   - ... la amplitud total es menor que el coeficiente de variabilidad

   Si no es posible, argumenta por qué no.

*7. No son muy comunes los estadísticos para resumir la variabilidad en una variable cualitativa (de nivel nominal). Sin embargo, comparando las siguientes distribuciones de la variable *Grupo Sanguíneo* en tres muestras distintas de 40 personas, se podría concluir que las distribuciones difieren respecto de la variabilidad: la distribución del lado izquierdo exhíbe nula variabilidad (no hay diferencias entre las personas en esta muestra), mientras que la distribución al lado derecho exhíbe máxima variabilidad (las personas están distribuidas equitativamente entre los cuatro valores de la variable). Construye un estadístico de variabilidad cuyo cálculo implica únicamente operaciones que se permiten para variables cualitativas y calcula su valor para las tres distribuciones graficadas.



8. El examen profesional teórico de la Facultad de Medicina del año 2013 consistió en 330 preguntas de cuatro opciones de respuesta, de las cuales una y solo una era correcta. La calificación total de cada alumno es el número de respuestas correctas que ha dado en las 330 preguntas. En este examen participaron 798 alumnos; la media aritmética y varianza de la calificación total X resultó en $\overline{x} = 220$ y $s_X^2 = 625$. Únicamente utilizando la información dada, deriva un límite superior para el número de estudiantes con una calificación menor de 165 (lo cual corresponde con el 50 % de la calificación máxima).

9. Construye un diagrama de caja y bigotes (variante 2) para la variable *Índice de Masa Corporal* de los datos de la página 34 para diferentes grupos de fumadores.

Tema 4
Transformaciones de variables

Índice

Al tratar las distribuciones de frecuencia y proporción y los estadísticos de tendencia central y dispersión en los temas anteriores, trabajamos con variables que se definieron directamente en el espacio muestral del experimental aleatorio. Sin embargo, en algunas ocasiones se considera *transformar* estas variables originales; es decir, se define una nueva variable, cuyos valores en las distintas observaciones se obtienen aplicando una regla específica (es decir, aplicando una función) a los valores de la variable original. En la primera sección de este tema se presentarán varios ejemplos de casos concretos en los cuales el investigador consideró transformar la variable que observó en la muestra; a partir de estos ejemplos, se disciernen las razones más comunes por las cuales se transforman variables. A continuación, se distinguen diferentes tipos de transformaciones (Sección 4.2) y se presentan algunos teoremas sobre cómo se puede hallar (a) la función de frecuencia y proporción y (b) los valores en los estadísticos de tendencia central y variabilidad de una variable transformada a partir de la variable original (Sección 4.3).

4.1 Introducción: Razones para transformar variables

Caso A: Compartiendo datos de la temperatura corporal

Un médico mexicano, especialista en virología, sospecha que uno de sus pacientes es portador de una nueva variante de un virus y, por lo tanto, lo mantiene en observación durante una semana. Durante este periodo le aplica varios tratamientos y evalúa como el paciente reacciona. Uno de los parámetros que registra (entre muchos otros) es la temperatura corporal, la cual mide en grados centígrados. El médico quiere compartir y discutir los datos con sus colegas virólogos en Estados Unidos, que han tenido más experiencia con este nuevo virus. No obstante, sus colegas estadounidenses suelen expresar la temperatura en Fahrenheit.

Por lo tanto, considera cambiar (transformar) los valores de la temperatura corporal, como los observó en grados centígrados, a valores en la escala de temperatura de Fahrenheit. En particular, obtiene las temperaturas en Fahrenheit, aplicando la siguiente función:

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto \frac{9}{5}x + 32. \end{aligned} \tag{4.1}$$

En palabras: f es una transformación dentro de los números reales (es una función con dominio y codominio $\mathbb{R}$), que asocia cada valor x con su imagen por la operación $\frac{9}{5}x + 32$. Esta función efectivamente convierte temperaturas en centígrados a temperaturas en Fahrenheit: Por ejemplo, 40°C corresponde con $\frac{9}{5}40 + 32 = 104$°F.

El ejemplo anterior describe una razón muy común para transformar. En la Sección 1.4 se presentaron distintos niveles de medición asociados con las variables. Dichos niveles de medición implican que se pueden aplicar determinadas transformaciones a las variables, sin pérdida de información. A una variable de nivel nominal, por ejemplo, se pueden asignar nuevos valores siempre y cuando las observaciones que tienen valores distintos en la variable original también lo tienen en la nueva variable y observaciones con valores idénticos en la variable original siguen teniendo valores idénticos en la variable transformada. Efectivamente, en variables nominales, tal transformación no implicaría ninguna pérdida de información. Similarmente, no se pierde información si una variable de nivel intervalo se transforma a través de una función lineal. En el ejemplo anterior, la variable original X, temperatura en grados centígrados, se mide con nivel de medición intervalo, por lo cual la nueva variable Y que se obtiene de la variable X por la función lineal en la Ecuación (4.1) contendría la misma información que la variable original: Efectivamente, los valores de X en grados centígrados nos cuentan lo mismo sobre la temperatura corporal del paciente que los valores de Y en grados Fahrenheit.

Lo anterior ilustra que las transformaciones son posibles y, a veces, necesarias, porque los valores de una variable (que se han asignado a cada resultado en el espacio muestral) son, hasta cierto punto, arbitrarios. La única excepción son las variables de nivel absoluto, ya que en estas variables no se permite ninguna transformación.

Conceptualizar una transformación

Una variable asocia cada resultado en el espacio muestral Ω con algún valor (véase la definición de *variable* en la p. 25). En el ejemplo del Caso A, la variable X formalmente es una función, definida por:

$$\begin{aligned} X : \Omega &\to \mathbb{R}, \\ \omega &\mapsto X(\omega). \end{aligned}$$

Es decir, la variable X asocia cada resultado ω (alguna temperatura corporal del paciente) en el espacio muestral Ω con un valor numérico, representado por $X(\omega)$, que indica la temperatura en centígrados de este resultado.

De manera similar, se puede definir la variable Y por:

$$\begin{aligned} Y : \Omega &\to \mathbb{R}, \\ \omega &\mapsto Y(\omega), \end{aligned}$$

que asocia cada resultado ω del espacio muestra Ω a un (otro) valor numérico $Y(\omega)$, que indica la temperatura en Fahrenheit.

La función f, como se define en la Ecuación (4.1), transforma (los valores de) la variable X en (valores de) la variable Y sin pasar por el espacio muestral Ω, como lo representa la siguiente gráfica:

$$\begin{tikzcd} & \mathbb{R} \arrow[dd, "f"] \\ \Omega \arrow[ur, "X"] \arrow[dr, "Y"'] & \\ & \mathbb{R} \end{tikzcd}$$

Caso B: Investigando el efecto de la temperatura ambiental

Un profesor sospecha que la atención que le prestan sus alumnos está relacionada con la temperatura en clase. Por eso, registra la temperatura (en centígrados) en el aula cada vez que da clase durante un año. Denomina a la variable que resulta X.

La hipótesis del profesor, sin embargo, es que la atención disminuye a medida que la temperatura *se aleja de la temperatura ideal*. Supone que la temperatura ideal es de 23°C y que desviaciones de esta temperatura, independientemente de si es por arriba o por abajo, afectan de forma similar la atención de los alumnos. Por ejemplo, una temperatura de 28°C, que corresponde con “5 grados demasiado calor” y una temperatura de 18°C, que corresponde con “5 grados demasiado frío” tienen, según su hipótesis, el mismo efecto en sus alumnos.

En otras palabras, la hipótesis del profesor implica saber para cada temperatura registrada X, cuántos grados difiere de la temperatura ideal. Se aplica la siguiente transformación f a la variable X:

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto |x - 23|. \end{aligned} \tag{4.2}$$

Contrario al Caso A, donde la transformación era una consecuencia de que el investigador podía eligir de forma arbitraria la unidad de las medidas, el profesor del Caso B aplica la transformación porque es necesaria para investigar su hipótesis de la investigación. El caso siguiente es otro ejemplo donde la investigación impone una transformación de la variable original.

Caso C: Evaluando el efecto de la intensidad percibida de sonidos

Un psiquiatra quiere tratar el trastorno de sueño en uno de sus pacientes. El paciente vive cerca del aeropuerto de la Ciudad de México y el trastorno parece estar relacionado con los aviones sobrevolando la zona. El psiquiatra decide investigar esta hipótesis y registra la intensidad sonora X (en decibelios) de los aviones sobrevolando durante una semana.

Sin embargo, el psiquiatra consulta a un colega psicólogo que le informa que estudios en el área de la psicofísica han mostrado que la intensidad *percibida* de un sonido *no* es directamente proporcional con la intensidad *física* del mismo; sino que la relación entre la intensidad percibida y la física es logarítmica. Por lo tanto, el psiquiatra construye una nueva variable Y, intensidad sonora percibida, aplicando la siguiente transformación f a la variable X, intensidad sonora física:

$$\begin{aligned} f : \mathbb{R}_0^+ &\to \mathbb{R}, \\ x &\mapsto \log_{10} x. \end{aligned}$$

Nota: $\mathbb{R}_0^+$ denota el conjunto de números reales estrictamente positivos (es decir, mayor que 0). Recuérdese que el logaritmo es definido únicamente para valores estrictamente positivos.

Es claro que, si el objetivo de la investigación es explorar cómo el ruido de los aviones afecta el sueño del paciente, el elemento crucial es cómo éste percibe el ruido (y no tanto, los aspectos físicos del ruido). Se puede concluir que la pregunta de la investigación requiere la transformación logaritmica de la variable original.

Caso D: Construyendo las notas en los exámenes departamentales

En los primeros años de la carrera de Médico Cirujano en la Facultad de Medicina, la evaluación de los estudiantes en las distintas asignaturas se realiza a través de exámenes de opción múltiple. En cada examen, se calcula para cada estudiante el número de preguntas en las que dio una respuesta correcta. Denominemos la variable así definida como X.

Al interpretar la puntuación X, hay que conocer el número de preguntas en el examen. Sin esta información no es posible saber si el resultado obtenido, por ejemplo $X = 38$, corresponde con una calificación aprobatoria o reprobatoria.

Por esta razón, se ha establecido dentro del sistema universitario, un esquema que facilita la interpretación de las calificaciones. En particular, las “puntuaciones crudas” X (número de respuestas correctas en el examen) se transforma en la calificación Y, que es la nota en el examen en una escala de 0 a 10. Utilizando esta escala, se aplican criterios uniformes, por ejemplo, para decidir quienes aprueban y reprueban una materia, o quienes, a partir de los exámenes departamentales, quedan exentos para el examen final, etc.

El cálculo de la nota sobre 10 en un examen implica aplicar una transformación a las puntuaciones originales:

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto f(x), \end{aligned}$$

donde $f(x)$ depende del número de preguntas en el examen y, posiblemente, de otros factores como procedimientos para redondear y correcciones por el grado de dificultad del examen.

La transformación (a notas sobre 10) en el ejemplo anterior se realiza para facilitar la interpretación de los valores en la variable. En la educación (y en otras ciencias sociales) la puntuación original obtenida por algún instrumento no proporciona mucha información en sí misma; para poder interpretar estas puntuaciones, se requiere compararla con algún estándar o con otras puntuaciones obtenidas. El siguiente caso presenta una transformación muy particular cuya función principal también es la interpretación clara y directa de la puntuación obtenida.

Caso E: Construyendo normas de referencia para un nuevo instrumento

Para su proyecto de investigación, un estudiante de doctorado ha desarrollado un nuevo instrumento para evaluar el grado de ansiedad en pacientes que sufren de varios tipos de fobias. El instrumento consiste en 75 preguntas que se contestan a través de una escala tipo Likert de cinco puntos. Para calcular la puntuación global X en el instrumento, se convierten las respuestas en la escala Likert en puntuaciones de 1 a 5 y se calcula la suma de todas las 75 puntuaciones.

Para evaluar la confiabilidad y validez de su instrumento, el investigador diseña un estudio en el cual aplica el instrumento a una muestra aleatoria de 500 personas normales. Después de una serie de análisis psicométricos concluye que su instrumento cumple con los estándares mínimos de confiabilidad y validez.

Sin embargo, la puntuación global X sufre de un problema importante: Aunque se mostró confiable y válido, es complicado interpretar el valor en X. En otras palabras, es difícil saber si una puntuación corresponde o no con un nivel de ansiedad normal. Para remediar este problema, el investigador propone la siguiente transformación:

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto \frac{x - \overline{x}}{s_X}, \end{aligned}$$

donde $\overline{x}$ y s_X son la media y desviación estándar, respectivamente, de las puntuaciones en la muestra de 500 personas normales.

Esta transformación indica qué tan lejos se encuentra la puntuación x de la media, en términos de desviaciones estándares. Por ejemplo, si el valor en la variable transformada es $+2$, entonces sabe que el nivel de ansiedad de la persona se encuentra dos desviaciones estándares encima de la media del grupo normal.

La transformación en el ejemplo anterior ocupa un lugar muy importante en la estadística y ha recibido el nombre de una *transformación* $\mathcal{Z}$ o *estandarización*.

Transformación $\mathcal{Z}$

Para una variable X, cuyos valores en una muestra de n observaciones no son todos idénticos, la transformación $\mathcal{Z}$ de X se define por:

$$\begin{aligned}\mathcal{Z} : \mathbb{R} &\to \mathbb{R},\\ x &\mapsto \frac{x-\overline{x}}{s_X}.\end{aligned}$$

Sinónimos: Estandarización, tipificación

Representaremos las variables estandarizadas (es decir, la variables obtenidas por una transformación $\mathcal{Z}$) casi siempre con la mayúscula Z. Se escribe, por ejemplo, $Z = \mathcal{Z}(X)$ para decir que la variable Z se obtuvo mediante una estandarización de la variable X.

Interpretación de los valores de una variable estandarizada

Un valor en una variable estandarizada tiene una interpretación clara y directa que indica cómo se posiciona en comparación con los otros valores observados en la variable. En otras palabras, una variable estandarizada proporciona directamente información sobre dónde cada valor se ubica en la distribución. Concretamente, la transformación $\mathcal{Z}$ implica una comparación con la media $\overline{x}$ que permite concluir que:

$$\begin{aligned}\mathcal{Z}(x) < 0 &\iff x < \overline{x}\\ \mathcal{Z}(x) = 0 &\iff x = \overline{x}\\ \mathcal{Z}(x) > 0 &\iff x > \overline{x}.\end{aligned}$$

Es decir, el signo de los valores en una variable estandarizada nos informa si se encuentran debajo o encima de la media. Por otra parte, la magnitud absoluta de $\mathcal{Z}(x)$ nos informa qué tan lejos, en términos de desviaciones estándares, el valor X se encuentra de la media. Si, por ejemplo, x se encuentra dos desviaciones típicas por abajo de la media (es decir, $x = \overline{x} - 2s_X$), entonces $\mathcal{Z}(x) = -2$.

Ejemplo

Definamos la variable Z, que se obtiene por una transformación $\mathcal{Z}$ de la variable X, *Índice de Masa Corporal*, de los datos de la página 34 y calculemos el valor de Z para cada observación.

En el tema anterior (véase, por ejemplo, la página 97), calculamos la media y la desviación estándar de esta variable:

$$\overline{x} = 25 \qquad\qquad s_X = 4.$$

Calculando el valor en la variable Z para las primeras tres observaciones, se obtiene:

$$z_1 = \frac{x_1 - \overline{x}}{s_X} = \frac{21-25}{4} = -1$$

$$z_2 = \frac{x_2 - \overline{x}}{s_X} = \frac{25 - 25}{4} = 0$$

$$z_3 = \frac{x_3 - \overline{x}}{s_X} = \frac{27 - 25}{4} = +0.5$$

Efectivamente, la primera observación tiene un valor (21) que se encuentra una desviación estándar (4) debajo de la media (25). Por lo tanto, su valor en la variable Z es -1. De la misma forma interpretamos el valor de $Z = 0$ para la segunda observación, ya que su valor de 25 coincide exactamente con la media. Por otro lado, el valor de la tercera observación está dos unidades encima de la media; puesto que dos puntos es media desviación estándar, el valor estandarizado corresponde con $+0.5$.

Los valores en la variable Z para cada una de las 50 observaciones se encuentran en la siguiente tabla:

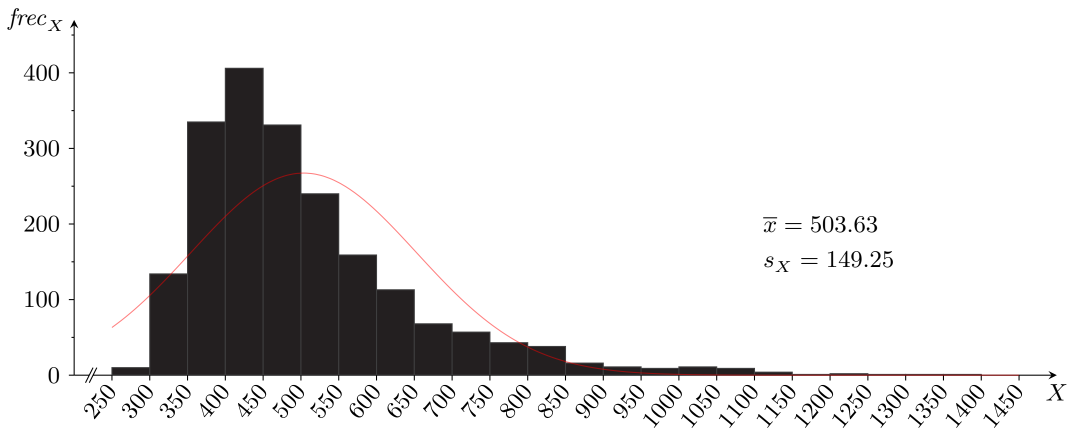
i	x_i	z_i	i	x_i	z_i	i	x_i	z_i	i	x_i	z_i	i	x_i	z_i
1	21	−1.00	11	22	−0.75	21	30	1.25	31	34	2.25	41	23	−0.50
2	25	0.00	12	26	0.25	22	24	−0.25	32	29	1.00	42	22	−0.75
3	27	0.50	13	29	1.00	23	27	0.50	33	26	0.25	43	28	0.75
4	20	−1.25	14	24	−0.25	24	25	0.00	34	19	−1.50	44	19	−1.50
5	33	2.00	15	22	−0.75	25	23	−0.50	35	24	−0.25	45	24	−0.25
6	23	−0.50	16	27	0.50	26	22	−0.75	36	24	−0.25	46	25	0.00
7	24	−0.25	17	20	−1.25	27	28	0.75	37	22	−0.75	47	23	−0.50
8	28	0.75	18	25	0.00	28	25	0.00	38	20	−1.25	48	27	0.50
9	23	−0.50	19	23	−0.50	29	22	−0.75	39	22	−0.75	49	26	0.25
10	27	0.50	20	26	0.25	30	24	−0.25	40	42	4.25	50	26	0.25

Por último, se presenta un sexto caso, que ilustra otra razón por la cual puede ser necesario transformar variables:

Caso F: Analizando distribuciones asimétricas y con valores atípicos

Un investigador del Centro Nacional de Prevención de Accidentes realiza un estudio para investigar el efecto de niveles bajos de alcohol en la conducción de un coche. Realiza un experimento en el cual cada una de 50 personas conduce en un simulador bajo 4 condiciones: sin haber tomado alcohol, después de haber tomado 1 cerveza, después de 2 cervezas y después de 3 cervezas. Durante la simulación, se presentan 10 situaciones que requieren una reacción instantánea para evitar un accidente. El investigador registra el tiempo de reacción (entre el momento que apareció la situación peligrosa y el momento de presionar el freno), denotado X, en las 10 situaciones.

Como primer paso en el análisis de los datos, construye un histograma de X (con los valores agrupados en intervalos de 50 milisegundos).



Tiempo de reacción en milisegundos

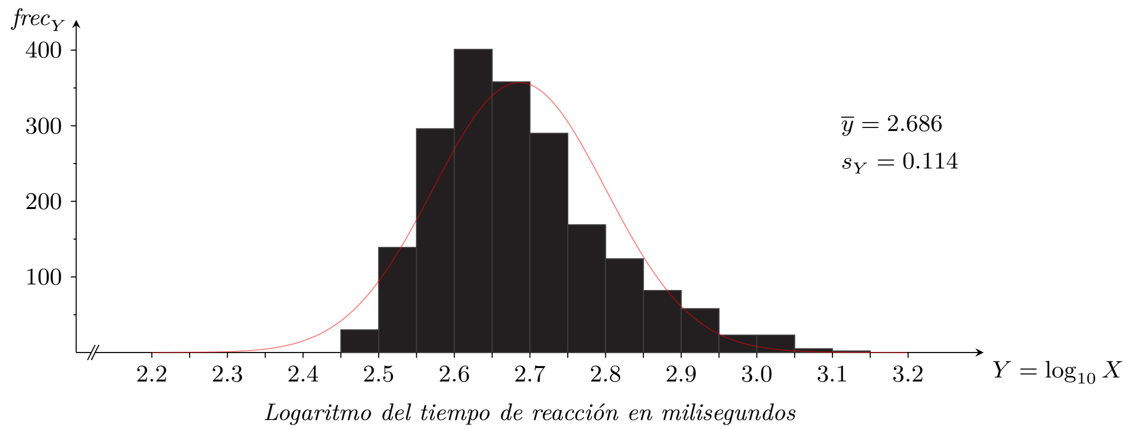
Esta representación gráfica, junto con la tabla de frecuencia (que no se muestra aquí), revela dos problemas que pueden interferir con el análisis subsecuente de los datos. Primero, se observan unos valores atípicos (de más de 1000 milisegundos). Los valores atípicos, aunque son muy comunes en datos de tiempo de reacción, frecuentemente causan problemas en los análisis inferenciales porque tienden a tener un impacto disproporcional en los resultados obtenidos (similar a que los valores atípicos pueden tener un impacto disproporcional al calcular la media). Segundo, el análisis que el investigador planea hacer (conocido como análisis de varianza) incluye entre sus supuestos que la distribución de la variable dependiente (el tiempo de reacción) en las diferentes condiciones sea simétrica (y más preciso, que siga una distribución normal, véase el Tema 10 para una introducción a la distribución normal). El histograma claramente muestra que la distribución *no* cumple este supuesto. (Compárese visualmente la distribución con la curva en color rojo que representa la distribución normal que mejor se ajusta a estos datos; existen pruebas formales para evaluar la bondad de ajuste de una distribución observada a una distribución teórica como la normal.)

La solución más común a estos problemas consiste en aplicar una transformación logarítmica a los tiempos de reacción. Después de construir una variable Y, que resulta de aplicar la transformación

$$\begin{aligned} f : \mathbb{R}_0^+ &\to \mathbb{R}, \\ x &\mapsto \log_{10} x \end{aligned} \tag{4.3}$$

a la variable original X, el investigador obtiene el histograma de la siguiente página (con los valores de Y agrupados en intervalos de 0.05).

Se observa un efecto doble de la transformación: (a) los valores atípicos ya no son tan extremos en la distribución de Y (por ejemplo, el valor máximo en la variable original X es 1380 milisegundos y se encuentra 5.9 desviaciones estándares arriba de la media, mientras que este mismo valor da un máximo de $\log_{10} 1380 = 3.14$ en la distribución de Y, el cual está 4.0 desviaciones estándares de la media) y (b) la distribución de Y resulta menos asimétrica que la distribución de la variable original (aunque todavía hay discrepancias con la distribución teórica, estas discrepancias son menores).



Resumen: ¿Por qué transformar?

En grandes líneas, como ilustran los casos anteriores, se pueden distinguir cuatro razones de por qué hacer una transformación de variables:

(a) Los valores en las variables originales son, hasta cierto punto, arbitrarios (Caso A);

(b) La pregunta de investigación precisa una transformación de la variable orginal (Casos B y C);

(c) Es difícil o imposible dar una interpretación a los valores originales (Casos D y E);

(d) Para algunos métodos estadísticos, los supuestos subyacentes no se cumplen para la variable original; sin embargo, se puede conseguir que se cumplan después de una transformación (Caso F).

4.2 Tipos de transformaciones

En esta sección, se presenta una clasificación de las transformaciones en cuatro tipos: transformaciones generales, transformaciones monótonas, transformaciones lineales y transformaciones $\mathcal{Z}$ o estandarizaciones. Es importante mencionar que los cuatro tipos están ordenados en el sentido de que un tipo previo incluye los tipos siguientes como caso especial. Por ejemplo, las transformaciones $\mathcal{Z}$ son también transformaciones lineales, y estos últimos cumplen también la definición de las transformaciones monótonas. La clasificación propuesta es, hasta cierto punto, arbitraria, pero nos servirá en la siguiente sección donde se investiga el efecto de las transformaciones en las distribuciones de frecuencia y proporción y en los estadísticos de tendencia central y variabilidad asociados. En particular, el efecto de transformaciones más específicas (como las lineales o las estandarizaciones) típicamente resulta en expresiones más simples y fáciles de aplicar.

4.2.1 Transformaciones generales

Transformación general

Un procedimiento que asigna un (único) valor y de la variable Y a cada valor x de una variable X, se llama una *transformación*.

Formalmente, una transformación es una función de $\mathbb{R}$ a $\mathbb{R}$:

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto f(x) = y \end{aligned} \tag{4.4}$$

Cabe señalar que el dominio de la función en la Ecuación (4.4), en vez de ser el conjunto completo de números reales $\mathbb{R}$, posiblemente es un subconjunto de $\mathbb{R}$, como en el caso de una transformación logarítmica (véanse los Casos C y F de la sección anterior). Otro detalle es que, aunque al definir el concepto de una variable en el tema introductorio se permitió que la variable asumiera valores no numéricos (véase, en particular, el comentario en la página 25), en este tema se supone para todas las variables que sus valores son números reales.

4.2.2 Transformaciones monótonas

Transformación estrictamente monótona

La transformación $f : \mathbb{R} \to \mathbb{R}$ es *estrictamente monótona* si satisface una de las siguientes dos condiciones:

(1) Para cualquier par de valores x_i y x_j se cumple que:

$$x_i > x_j \iff f(x_i) > f(x_j) \qquad (\textit{transformación estrictamente creciente}) \tag{4.5a}$$

(2) Para cualquier par de valores x_i y x_j se cumple que:

$$x_i > x_j \iff f(x_i) < f(x_j) \qquad (\textit{transformación estrictamente decreciente}) \tag{4.5b}$$

De la definición anterior, debe ser claro que una transformación que es estrictamente monótona es una de dos: estrictamente creciente o estrictamente decreciente.

La definición implica que si la variable Y se obtiene por una transformación *estrictamente creciente* de la variable X, entonces el orden de las observaciones en la variable Y es el mismo que su orden en la variable X. En otras palabras, si la observación i tiene un valor mayor que la observación j en la variable X (es decir, $x_i > x_j$), entonces también en la variable Y el valor de i es mayor que el de j ($y_i > y_j$). Por otro lado, si la transformación es *estrictamente decreciente*, entonces el orden de las observaciones según su valor en X es inverso al orden según su valor en Y: Si la observación i tiene un valor mayor que la observación j en la variable X (es decir, $x_i > x_j$), entonces en la variable Y el valor de i es menor que el de j ($y_i < y_j$).

Ejemplo

Todas las transformaciones que se introdujeron en los casos de la Sección 4.1, excepto la del Caso B son transformaciones *estrictamente crecientes*. Por ejemplo, la transformación que convierte temperaturas de centigrados a Fahrenheit (Caso A) guarda el orden entre las observaciones; es decir, si el paciente tuvo en el momento i una temperatura en centigrados mayor que en el momento j, entonces también su temperatura en Fahrenheit correspondiente del momento i sería mayor que la del momento j. Asimismo, la transformación logarítmica de los Casos C y F guarda el orden. Ya se ilustró, por ejemplo, que la observación que tiene el tiempo de reacción (en milisegundos) máximo también es la observación que tiene el valor máximo en la variable transformada.

Por otro lado, la transformación que realizó el profesor del Caso B *no* es una transformación (estrictamente) monótona. La primera condición en la definición anterior no se cumple ya que, si $x_i = 24$ y $x_j = 21$, entonces:

$$\begin{aligned} f(x_i) &= |x_i - 23| = |24 - 23| = 1 \\ \text{y } f(x_j) &= |x_j - 23| = |21 - 23| = 2. \end{aligned}$$

Es decir, existe un par de valores x_i y x_j, para el cual $x_i > x_j$ y $f(x_i) < f(x_j)$, contradiciendo la primera condición. Además, tampoco se cumple la segunda condición ya que, para $x_i = 24$ y $x_j = 23$ tenemos

$$\begin{aligned} f(x_i) &= |x_i - 23| = |24 - 23| = 1 \\ \text{y } f(x_j) &= |x_j - 23| = |23 - 23| = 0. \end{aligned}$$

lo cual demuestra que existe un par de valores para el cual $x_i > x_j$ y $f(x_i) > f(x_j)$.

La Sección 4.1 no incluye casos de transformaciones estrictamente *decrecientes*. Como ejemplo de este tipo de transformaciones, considérese la siguiente:

$$\begin{aligned} f : \mathbb{R}_0^+ &\to \mathbb{R} \\ x &\mapsto \frac{1}{x}. \end{aligned}$$

Efectivamente, es una transformación monótona decreciente, que invierte el orden entre las observaciones. En el ejemplo de la página 79, la variable Z' (razón Globulina/Albúmina) se obtiene precisamente por aplicar esta transformación a la variable Z (razón Albúmina/Globulina). Como se puede observar en la tabla de este ejemplo, el orden en la variable Z' es totalmente inverso al orden en Z: Siempre se cumple que si $z_i > z_j$, entonces $z'_i < z'_j$ (y viceversa).

Las transformaciones estrictamente monótonas son biyecciones (para la definición de *biyección*, véase la página 18). Una transformación es biyectiva si a cada valor de la variable original X se asigna otro valor de Y; es decir, no ocurre que dos valores diferentes de X se transforman en el mismo valor de Y. La representación gráfica de una transformación biyectiva (véase, por ejemplo, el diagrama en la página 18) se caracteriza no solo por que sale una flecha de cada elemento del dominio (los valores de la variable X), sino también que en cada elemento del codominio (los valores de la variable Y) llega una y solo una flecha.

Por lo tanto, si una transformación es biyectiva (como lo son las transformaciones estrictamente monótonas), se puede definir la transformación inversa f^{-1}.

Transformación inversa

Si $f : \mathbb{R} \to \mathbb{R}$ es una transformación biyectiva que asigna un valor y de la variable Y a cada valor x de la variable X, la *transformación inversa*, denotada f^{-1}, es la función definida por:

$$f(x) = y \quad \Longleftrightarrow \quad f^{-1}(y) = x$$

En palabras, f^{-1} es la función que realiza el "camino de vuelta". La función inversa tiene como dominio los valores de la variable Y y como codominio los valores de la variable X (es decir, en la representación gráfica las flechas salen de los valores de Y y llegan a los valores de X); $f^{-1}(y)$ devuelve aquel valor de X que por la transformación f resultó en el valor y de Y.

Ejemplos

Consideremos para cada transformación introducida en los Casos A a F de la section anterior, la transformación inversa.

- *Caso A*

  La transformación en la Ecuación (4.1) convierte temperaturas en grados centígrados a temperaturas en grados Fahrenheit. Una manipulación algebraica de la Ecuación (4.1) inmediatamente lleva a la función inversa:

  $$\begin{aligned} f^{-1} &: \mathbb{R} \to \mathbb{R}, \\ y &\mapsto \frac{5}{9}y - 17.78 \end{aligned} \tag{4.6}$$

  En el ejemplo, se calculó—aplicando la transformación f a 40°C—que 40°C corresponde con 104°F. Ahora aplicando la transformación inversa a 104°F, devuelve el valor original en centígrados:

  $$f^{-1}(104) = \frac{5}{9}104 - 17.78 = 40.$$

- *Caso B*

  Como se comentó al inicio de esta sección, la transformación en el Caso B *no* es monótona, ni biyectiva. (Aunque en teoría una transformación puede ser no monótona pero sí biyectiva, muy rara vez se presentan transformaciones de este tipo en la estadística.) Por lo tanto, la transformación inversa no existe.

  Al examinar en detalle la transformación en la Ecuación (4.2), se aclara porque la transformación inversa no existe: Todos los valores, excepto 0, en la variable transformada Y son la imagen de dos valores distintos de la variable original X (es decir, a cada valor de Y llegan dos flechas, de dos valores distintos de X). Por consiguiente, no se puede definir la transformación inversa ya que esto implicaría que a cada y se asociasen dos valores distintos de X (lo cual contradice la definición general de una transformación, que exige que a cada valor del dominio se asocia un *único* valor del codiminio).

- *Casos C y F*

  En estos dos casos, la transformación es logarítmica. El logaritmo con base 10 aplicado a un valor x (donde x es estrictamente mayor que 0) corresponde con el exponente al cual hay que elevar 10 para obtener x. Por ejemplo, $\log_{10} 1000 = 3$, porque $10^3 = 1000$. Esto quiere decir que la transformación inversa de la transformación en los Casos C y F, es la siguiente:

$$\begin{aligned} f^{-1} &: \mathbb{R} \to \mathbb{R}, \\ y &\mapsto 10^y. \end{aligned}$$

Si, por ejemplo, el logaritmo del tiempo de reacción más alto en el Caso F es igual a 3.14, entonces, se puede aplicar la función inversa para conocer el valor original de

$$f^{-1}(3.14) = 10^{3.14} = 1380.$$

- *Caso D*
  Para este ejemplo, no se especificó la transformación en detalle. Sin embargo, supongamos que para conocer la nota sobre 10 (variable Y) a partir del número de respuestas correctas (variable X) en un examen de P preguntas de opción múltiple, simplemente se aplica la regla:

$$\begin{aligned} f &: \mathbb{R} \to \mathbb{R}, \\ x &\mapsto 10\,\frac{x}{P}. \end{aligned} \tag{4.7a}$$

  Esto significa, por ejemplo, que un estudiante que acierta 68 de 80 preguntas, verá en su expediente una nota de

$$f(68) = 10 \times \frac{68}{80} = 8.5.$$

  Ahora, para conocer el número de respuestas correctas en el examen a partir de la nota sobre 10, hay que aplicar la transformación inversa:

$$\begin{aligned} f^{-1} &: \mathbb{R} \to \mathbb{R}, \\ y &\mapsto \frac{P\,y}{10}. \end{aligned} \tag{4.7b}$$

  Efectivamente, el estudiante con la nota de 8.5, tuvo

$$f^{-1}(8.5) = \frac{80 \times 8.5}{10} = 68$$

  de las 80 preguntas correctas.

- *Caso E*
  Como la transformación $\mathcal{Z}$ también es estrictamente monótona (siempre creciente), se puede considerar la transformación inversa que "transforma de vuelta" un valor estandarizado z al valor original x correspondiente:

$$\begin{aligned} \mathcal{Z}^{-1} &: \mathbb{R} \to \mathbb{R}, \\ z &\mapsto s_X\, z + \overline{x}. \end{aligned} \tag{4.8}$$

En el ejemplo de la página 112 se calculó para la primera observación un valor estandarizado $z = -1$ en la variable *Índice de Masa Corporal*. Aplicando la transformación $\mathcal{Z}$ inversa, se obtiene su valor original para IMC:

$$\mathcal{Z}^{-1}(-1) = -1 \times 4 + 25 = 21.$$

Transformaciones monótonas (sin ser *estrictamente* monótonas)

En las matemáticas, se consideran también *transformaciones monótonas* en general, sin añadir el adverbio *estrictamente*. En la definición de estas transformaciones se han sustituido las condiciones de la Ecuación (4.5) por una variante más débil. En particular, una transformación monótona debe satisfacer una de las siguientes dos condiciones:

(1) Para cualquier par de valores x_i y x_j se cumple que:

$$x_i \geqslant x_j \quad \Longleftrightarrow \quad f(x_i) \geqslant f(x_j)$$

(2) Para cualquier par de valores x_i y x_j se cumple que:

$$x_i \geqslant x_j \quad \Longleftrightarrow \quad f(x_i) \leqslant f(x_j)$$

Esta definición implica que puede haber valores x_i y x_j, para el cual $x_i > x_j$ y, sin embargo, $f(x_i) = f(x_j)$, un caso que no se permite con transformaciones estrictamente monótonas.

Un ejemplo de una transformación monótona que no lo sea estrictamente, se presentaría si en el Caso D se aplicara una transformación que implica redondear las notas sobre 10 a un número entero. Supongamos, por ejemplo, que un profesor calcula las notas Y para su asignatura a partir de la siguiente transformación del número de respuestas correctas X en su examen de 30 preguntas: "Divide x entre 3 y si resulta un número con una parte decimal, simplemente redondea al número entero más bajo". Así, por ejemplo, los tres valores 21, 22 y 23 se transforman a una nota de 7 sobre 10 para esta asignatura.

Este ejemplo muestra que, si no se añade la cualidad de "estrictamente" a las transformaciones monótonas, existe la posibilidad que no sean biyectivas. (En el ejemplo pueden llegar hasta tres flechas a cada valor de Y.) Sin embargo, es precisamente la característica de ser biyectiva que conlleva propiedades interesantes, por lo cual esta sección empezó con la definición de transformaciones *estrictamente* monótonas.

4.2.3 Transformaciones lineales

Transformación lineal

La transformaciones f es una *transformación lineal* si es del siguiente tipo:

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto ax + b, \end{aligned}$$

donde a y b son dos constantes cualesquiera.

Transformaciones y funciones lineales son muy frecuentes en la estadística. Nótese que las transformaciones lineales son un caso especial de las transformaciones monótonas y que, si se añade la restricción de que la constante multiplicativa $a \neq 0$, la transformación lineal es una transformación *estrictamente* monótona. Es fácil comprobar que si $a > 0$, entonces se trata de una transformación estrictamente creciente y si $a < 0$, entonces es estrictamente decreciente.

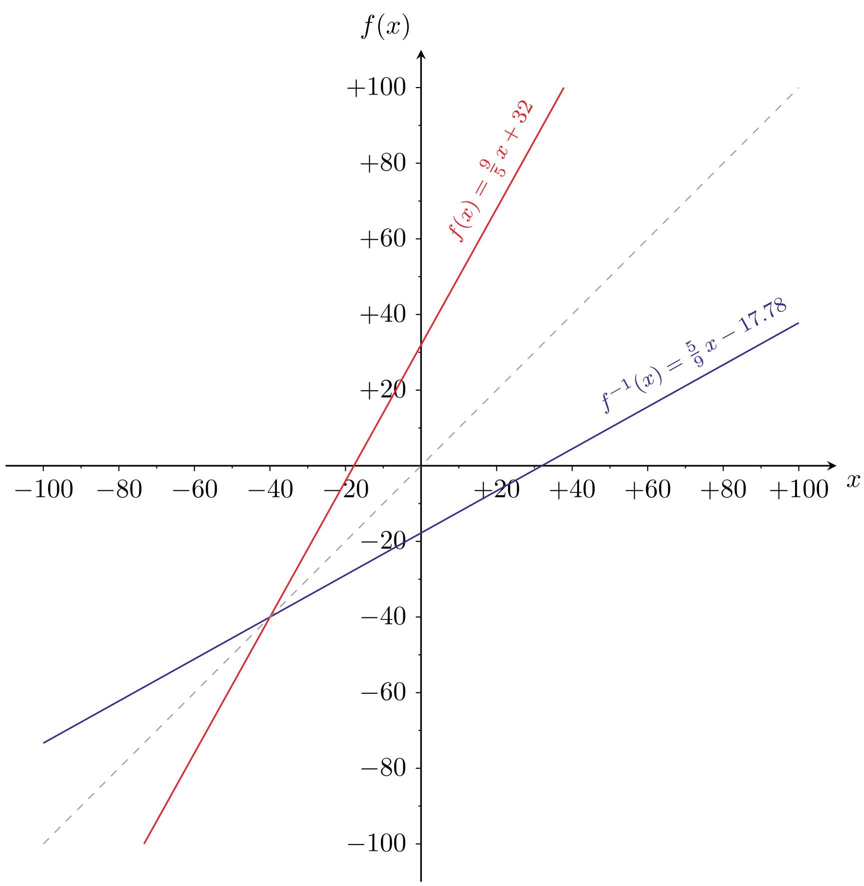
Ejemplo

Las transformaciones introducidas en los Casos A y D (si para este último se utiliza la transformación en la Ecuación (4.7a)) son lineales. También la transformación $\mathcal{Z}$ del Caso E es lineal, lo cual se demostrará en la siguiente sección. Las transformaciones logarítmicas de los Casos C y F no cumplen con la forma de una transformación lineal.

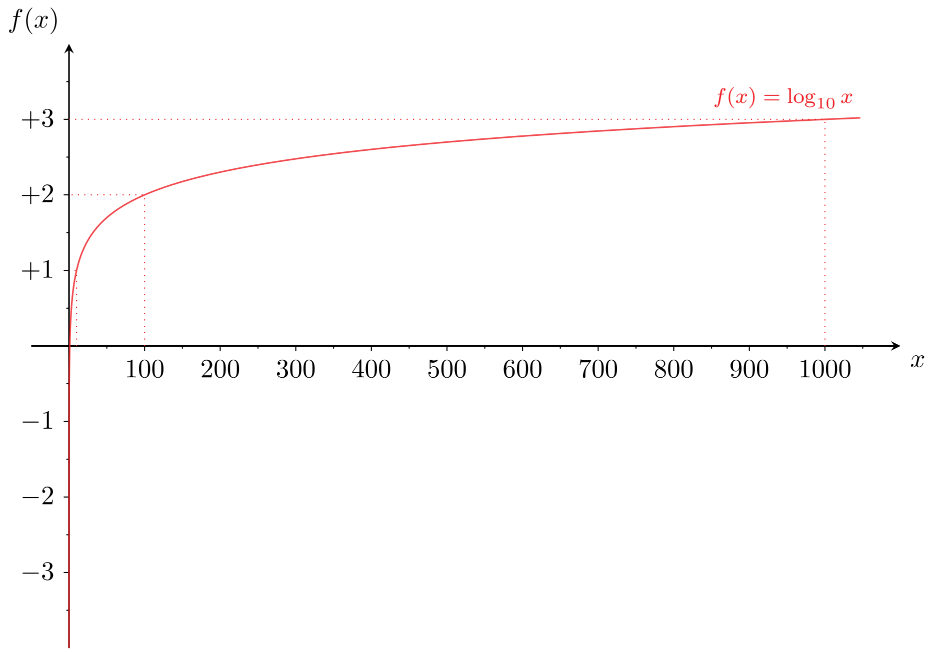
Nótese que la transformación inversa de una transformación lineal con $a \neq 0$ también es una transformación lineal. Véanse, por ejemplo, las funciones inversas en las Ecuaciones (4.6), (4.7b) y (4.8), para los Casos A, D y E, respectivamente.

Representación gráfica en el plano de una transformación lineal

Tal como se mencionó también en el ejemplo de la página 15, la representación gráfica de una transformación lineal es una recta. La siguiente figura, por ejemplo, representa la transformación lineal en la Ecuación (4.1) del Caso A. Nótese que la transformación inversa (en la Ecuación (4.6)) es la imagen especular que se obtiene reflejando la recta de la transformación original por la primera bisectriz.



Por otro lado, la representación gráfica de la transformación logarítmica de los Casos C y F *no* tiene forma de una recta.



4.2.4 Estandarizaciones

La definición y la interpretación de una transformación $\mathcal{Z}$ o estandarización se introdujeron en la página 112. Es fácil demostrar que cumple con la definición de una transformación lineal.

La transformación $\mathcal{Z}$ es una transformación lineal

Una transformación $\mathcal{Z}$ se define como:

$$\mathcal{Z}(x) = \frac{x - \overline{x}}{s_X},$$

Separando la diferencia en el nominador lleva a:

$$\begin{aligned}\mathcal{Z}(x) &= \frac{x}{s_X} - \frac{\overline{x}}{s_X}\\ &= \underbrace{\frac{1}{s_X}}_{a}\, x + \underbrace{\left(-\frac{\overline{x}}{s_X}\right)}_{b}.\end{aligned}$$

Entonces, $\mathcal{Z}(x)$ es una transformación lineal $ax + b$, con $a = \dfrac{1}{s_X}$ y $b = -\dfrac{\overline{x}}{s_X}$. ■

Nótese que la transformación $\mathcal{Z}$ es una transformación estrictamente *creciente*, ya que la constante multiplicativa $\frac{1}{s_X}$ siempre es mayor que 0 (por la primera propiedad de varianzas y desviaciones estándares).

4.3 Efectos de una transformación

4.3.1 Efectos en la función de frecuencia y proporción

La pregunta que se plantea en esta sección es la siguiente: Supongamos que se conoce la función de frecuencia de una variable X y se define una transformación f en esta variable, ¿cómo será la función de frecuencia de la variable transformada $Y = f(X)$? Consideremos el siguiente ejemplo:

Ejemplo

Volvamos al Caso B del profesor que quiere examinar el efecto de la temperatura en clase en la atención de sus alumnos. El profesor recopiló datos durante un año académico en 40 ocasiones; nos comparte el resultado de sus cálculos para obtener las funciones de frecuencia y proporción para la variable X, temperatura en clase (véase la tabla del lado izquierdo):

j	x_j	$frec_X(x_j)$	$p_X(x_j)$
1	18	2	.050
2	19	3	.075
3	20	4	.100
4	21	4	.100
5	22	7	.175
6	23	6	.150
7	24	8	.200
8	25	3	.075
9	26	2	.050
10	27	1	.025

k	y_k	$frec_Y(y_k)$	$p_Y(y_k)$
1	0	6	.150
2	1	15	.375
3	2	7	.175
4	3	6	.150
5	4	4	.100
6	5	2	.050

Recuérdese que, para investigar su hipótesis, el profesor construyó una variable Y que se definió a partir de la variable original X aplicándole la transformación f: $Y = f(X) = |X - 23|$. La distribución de Y se encuentra en la tabla a la derecha, pero ¿cómo se obtuvo de la tabla de frecuencia de X?

Para conocer la frecuencia del valor 0 en la variable Y, hay que reconocer que este valor coincide con la temperatura ideal; efectivamente, para que Y asuma el valor 0, la temperatura original en X debe ser 23. Por lo tanto, el número de observaciones cuya temperatura coincide con la temperatura ideal (es decir, para las cuales $Y = 0$) es igual al número de observaciones con un valor de 23 en la variable original X; o bien:

$$\begin{aligned} frec_Y(0) &= frec_X(23) \\ &= 6. \end{aligned}$$

En las tablas anteriores, se ha dado el mismo color a las filas correspondientes; en este caso, las filas con $x_6 = 23$ y $y_1 = 0$.

Para conocer $\mathit{frec}_Y(1)$, se buscan entre los valores de la variable X cuáles temperaturas corresponden con un grado de diferencia de la temperatura ideal, es decir, se buscan todos los valores x cuya imagen $f(x)$ es igual a 1. Se encontrará que hay dos vías para llegar a un valor de 1 en Y: Tanto el valor 22 como el valor 24 en la variable X corresponden con una diferencia de 1 con la temperatura ideal. Esto implica que:

$$\begin{aligned} \mathit{frec}_Y(1) &= \mathit{frec}_X(22) + \mathit{frec}_X(24) \\ &= 7 + 8 = 15 \end{aligned}$$

Para indicar que la frecuencia de $Y = 1$ se obtiene de las frecuencias $X = 22$ y $X = 24$, las filas correspondientes en las tablas anteriores tienen el mismo color. La frecuencia de los demás valores en Y se obtienen de la misma forma.

La derivación de la función de proporción p_Y se deriva de forma análoga de la función de proporción p_X. Por ejemplo, se tiene:

$$\begin{aligned} p_Y(4) &= p_X(19) + p_X(27) \\ &= .075 + .025 = .100, \end{aligned}$$

puesto que el resultado de la transformación es igual a 4 para los valores 19 y 27 (ambos difieren 4 grados de la temperatura ideal). Similarmente,

$$p_Y(5) = p_X(28)$$

ya que 28 es la única temperatura observada que tiene una diferencia de 5 con la temperatura ideal.

El siguiente teorema formaliza las ideas expuestas en el ejemplo anterior:

Función de frecuencia y proporción de una variable transformada

Si la variable Y se ha obtenido por una transformación f de la variable X, es decir, $Y = f(X)$, entonces:

$$\mathit{frec}_Y(y) = \sum_{\substack{x \\ f(x)=y}} \mathit{frec}_X(x) \tag{4.9a}$$

y

$$p_Y(y) = \sum_{\substack{x \\ f(x)=y}} p_X(x) \tag{4.9b}$$

Nota:

El uso del signo sumatorio en este teorema es ligeramente distinto a la forma que se ha utilizado previamente. Como se ha ilustrado en el ejemplo anterior, calcular la suma

$$\sum_{\substack{x \\ f(x)=y}} \mathit{frec}_X(x)$$

implica los siguientes tres pasos:

Paso 1: Encontrar todos los valores x cuya imagen $f(x)$ es igual al valor y. Esto es precisamente lo que indica la doble expresión debajo del signo sumatorio: El conjunto de los valores x para los cuales $f(x) = y$.

Paso 2: Obtener la frecuencia de cada uno de los valores (en la función de frecuencia de la variable X) del paso 1.

Paso 3: Sumar las frecuencias del paso 2.

El teorema anterior es válido para cualquier transformación (una *transformación general*, como se definió en la página 116). En el caso de una transformación estrictamente monótona—como es biyectiva—existe para cada valor y únicamente un valor x cuya imagen es igual a y. Esto implica que, para este caso, la suma en las Ecuaciones (4.9) consiste en solo un término.

Como se discutió anteriormente, para transformaciones biyectivas f, el valor x para el cual $f(x) = y$ se encuentra aplicando la transformación inversa al valor de y, es decir:

$$x = f^{-1}(y).$$

Estas consideraciones llevan al siguiente teorema:

Funciones de frecuencia y proporción de una variable obtenida por una transformación estrictamente monótona

Si la variable Y se obtiene por una transformación f estrictamente monótona de la variable X, entonces:

$$\mathit{frec}_Y(y) = \mathit{frec}_X\left[f^{-1}(y)\right] \tag{4.10a}$$

y

$$p_Y(y) = p_X\left[f^{-1}(y)\right] \tag{4.10b}$$

Ejemplo

Si el colega estadounidense del investigador del Caso A quiere saber $p_Y(104)$ (es decir, la proporción de registros en los cuales el paciente tuvo una temperatura corporal de 104ºF) y solo tiene acceso a la tabla de p_X (con las proporciones de las temperaturas corporales en centígrados), entonces primero calcula

$$\begin{aligned} x &= f^{-1}(104) \\ &= \frac{5}{9} \times 104 \ - \ 17.78 \qquad \text{(véase la Ec. (4.6))} \\ &= 40 \end{aligned}$$

y a continuación consulta la tabla de la función de proporción de la variable X, para encontrar:

$$p_Y(104) = p_X(40).$$

Este cálculo dice nada más que la proporción con que ocurre el valor de 104ºF es la misma que la proporción con que ocurre el valor de 40ºC.

4.3.2 Efectos en la media aritmética

En vez de derivar las frecuencias y proporciones de una variable transformada a partir de las mismas de la variable original, se puede enfocar únicamente en uno o más estadísticos que resumen la distribución de la variable transformada. En esta sección se trata la pregunta de cómo una transformación afecta a los estadísticos de tendencia central y, en particular, la media aritmética. (En principio, se puede investigar el efecto de diferentes tipos de transformaciones en cualquier estadístico de tendencia central; sin embargo, en esta sección nos limitamos a la media aritmética. Al final de este tema, se incluyen algunos problemas que investigan el efecto de una transformación en otros estadísticos de tendencia central, como la moda y la mediana.)

El primer teorema en esta sección considera el caso de una transformación general y cómo calcular la media aritmética en una variable transformada a partir de la variable original.

Media aritmética de una variable transformada

Si la variable Y se obtiene por una transformación f de la variable X, es decir $Y = f(X)$, entonces se puede calcular $\overline{y}$

(a) a partir de la función de frecuencia de X como sigue:

$$\overline{y} \ = \ \frac{\displaystyle\sum_{j=1}^{m} \mathit{frec}_X(x_j) f(x_j)}{n} \tag{4.11a}$$

(b) a partir de la función de proporción de X como sigue:

$$\overline{y} \ = \ \sum_{j=1}^{m} p_X(x_j) f(x_j). \tag{4.11b}$$

En ambas ecuaciones, $x_1, x_2, \ldots, x_m$ son los m valores distintos observados en la variable X.

Comprobación de las Ecuaciones (4.11)

Sólo se demuestra la Ecuación (4.11a). La demostración de (4.11b) se desarrolla de forma análoga. Según la Ecuación (3.2a), se puede calcular la media de Y como:

$$\overline{y} = \frac{\sum_{k=1}^{r} \mathit{frec}_Y(y_k)\, y_k}{n},$$

donde $y_1, y_2, \ldots, y_r$ son los r valores distintos observados en Y. Nótese que $r \leqslant m$ (el número de valores distintos en Y es menor que o igual al número de valores distintos en X, ya que con cada valor x se asocia sólo un valor y de la variable Y, mientras que, en el caso de que f sea una transformación general, el mismo valor y puede ser la imagen de diferentes valores de X).

Desarrollando el numerador de la expresión del lado derecho resulta en:

$$\begin{aligned}\sum_{k=1}^{r} \mathit{frec}_Y(y_k)\, y_k &= \sum_{k=1}^{r} \left[\left(\sum_{\substack{x \\ f(x)=y_k}} \mathit{frec}_X(x) \right) y_k \right] && \text{(por la Ecuación (4.9a))} \\ &= \sum_{k=1}^{r} \sum_{\substack{x \\ f(x)=y_k}} \left(\mathit{frec}_X(x) f(x) \right).\end{aligned}$$

Los términos de la doble suma en la última expresión corresponden con los m valores distintos de la variable X (el segundo signo sumatorio incluye los valores de x que tienen como imagen y_k, mientras que el primer signo sumatorio pasa por todos los valores y_k de Y; por lo tanto, la combinación de ambos pasa por todos los valores "que tiene como imagen algún valor de Y", es decir, pasa por todos los valores de X). Es decir, se acaba de comprobar la siguiente igualdad:

$$\sum_{k=1}^{r} \mathit{frec}_Y(y_k)\, y_k = \sum_{j=1}^{m} \mathit{frec}_X(x_j) f(x_j),$$

lo cual lleva directamente a la Ecuación (4.11a). ■

Ejemplo

Ilustramos el teorema anterior utilizando los datos del profesor del Caso B sobre la temperatura observada (X) y cuánto desvía, en términos absolutos, de la temperatura ideal (Y); la función de frecuencia y proporción de X y Y se presentaron en la tabla del ejemplo en la página 123. Apliquemos la Ecuación (4.11b) para conocer la media de Y:

$$\begin{aligned}\overline{y} &= \sum_{j=1}^{10} p_X(x_j) f(x_j) \\ &= p_X(18)f(18) + p_X(19)f(19) + p_X(20)f(20) + p_X(21)f(21) + p_X(22)f(22) \\ &\qquad + p_X(23)f(23) + p_X(24)f(24) + p_X(25)f(25) + p_X(26)f(26) + p_X(27)f(27) \\ &= .050 \times 5 + .075 \times 4 + .100 \times 3 + .100 \times 2 + .175 \times 1 \\ &\qquad + .150 \times 0 + .200 \times 1 + .075 \times 2 + .050 \times 3 + .025 \times 4 \qquad (\star) \\ &= 1.825\end{aligned}$$

Al verificar este resultado calculando la media aritmética directamente a partir de la función de proporción de la variable Y (es decir, utilizando la tabla del lado derecho en el ejemplo en la página 123 y la fórmula (3.2b)), se obtiene

$$\begin{aligned}
\overline{y} &= \sum_{k=1}^{6} p_Y(y_k) y_k \\
&= p_Y(0) \times 0 + p_Y(1) \times 1 + p_Y(2) \times 2 + p_Y(3) \times 3 + p_Y(4) \times 4 + p_Y(5) \times 5 \\
&= .150 \times 0 + .375 \times 1 + .175 \times 2 + .150 \times 3 + .100 \times 4 + .050 \times 5 \qquad (\star\star) \\
&= 1.825,
\end{aligned}$$

el mismo resultado que el anterior. Nótese también por qué los dos resultados coinciden: Reordenar los términos en la suma $(\star)$, como

$$\underbrace{.150 \times 0}_{\sum_{\substack{x \\ f(x)=0}} p_X(x) f(x)} + \underbrace{.175 \times 1 + .200 \times 1}_{\sum_{\substack{x \\ f(x)=1}} p_X(x) f(x)} + \underbrace{.100 \times 2 + .075 \times 2}_{\sum_{\substack{x \\ f(x)=2}} p_X(x) f(x)}$$

$$+ \underbrace{.050 \times 3 + .100 \times 3}_{\sum_{\substack{x \\ f(x)=3}} p_X(x) f(x)} + \underbrace{.075 \times 4 + .025 \times 4}_{\sum_{\substack{x \\ f(x)=4}} p_X(x) f(x)} + \underbrace{.050 \times 5,}_{\sum_{\substack{x \\ f(x)=5}} p_X(x) f(x)}$$

muestra que cada término en $(\star\star)$ corresponde con una suma de uno o dos términos en $(\star)$.

En el caso especial de una transformación lineal, se puede obtener la media aritmética de la variable transformada directamente a partir de la media de la variable original, a través del siguiente teorema:

Media aritmética de una variable obtenida por una transformación lineal

Si la variable Y se obtiene por una transformación lineal de la variable X, es decir $Y = aX + b$, con constantes cualesquiera a y b, entonces:

$$\overline{y} = a\overline{x} + b. \tag{4.12}$$

En palabras, esta propiedad dice que, para obtener la media aritmética de una variable que se obtuvo por una transformación lineal, se puede aplicar la misma transformación lineal a la media de la variable original.

Prueba del teorema sobre la media de una variable linealmente transformada

Este teorema se puede demostrar de varias maneras. Aquí se presenta una demostración que parte de la definición de la media aritmética de Y:

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Si Y se obtiene por una transformación lineal de X, entonces para cada observación i, se cumple que $y_i = ax_i + b$. Por consiguiente:

$$\begin{aligned}\overline{y} &= \frac{\sum_{i=1}^{n}(ax_i + b)}{n}\\ &= \frac{\sum_{i=1}^{n}(ax_i)}{n} + \frac{\sum_{i=1}^{n} b}{n}\end{aligned}$$

Como a y b son constantes en los n términos de las respectivas sumas, se tiene

$$\begin{aligned}&= \frac{a\sum_{i=1}^{n} x_i}{n} + \frac{nb}{n}\\ &= a\,\frac{\sum_{i=1}^{n} x_i}{n} + b\end{aligned}$$

Dándose cuenta de que $\frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}$ se obtiene directamente:

$$\overline{y} = a\overline{x} + b. \qquad \blacksquare$$

Ejemplo

Tomemos como ejemplo los datos de un profesor que da clase de Bioestadística; para evaluar la asignatura con sus 12 alumnos les aplicó un examen de 12 preguntas (abiertas). Cada pregunta valía entre 2 y 6 puntos, tal que la puntuación máxima (denotada X) en el examen es 50. Para los expedientes de los alumnos, necesita transformar las puntuaciones en una nota sobre 10 (variable Y).

Después de haber aplicado el examen, el profesor se dio cuenta que el examen en general era bastante difícil y que exploró conocimientos más avanzados de la materia en comparación de lo que había planeado al inicio del curso (los alumnos mostraron mucho interés en la materia y en varias ocasiones pidieron al profesor que extendiera algún tema). Por esta razón decide sumar 6 puntos a la puntuación X de cada alumno, antes de dividirla entre 5 para obtener la nota sobre 10. Es decir, define la siguiente regla para transformar la puntuación X en la nota Y:

$$Y = \frac{X + 6}{5}$$

La siguiente tabla presenta el resultado de cada alumno, en ambas variables:

	x_i	$\mapsto$	y_i		x_i	$\mapsto$	y_i		x_i	$\mapsto$	y_i
Adrián	42	$\mapsto$	9.6	Enrique	38	$\mapsto$	8.8	Ingrid	38	$\mapsto$	8.8
Bruno	37	$\mapsto$	8.6	Fernanda	44	$\mapsto$	10.0	Jacinto	38	$\mapsto$	8.8
Cristina	39	$\mapsto$	9.0	Gisela	35	$\mapsto$	8.2	Kurt	34	$\mapsto$	8.0
Diana	30	$\mapsto$	7.2	Héctor	43	$\mapsto$	9.8	Lucía	26	$\mapsto$	6.4

Es fácil verificar que la media de X y Y son 37 y 8.6, respectivamente. Reconociendo que la transformación de la cual se obtiene Y a partir de X es lineal con

$$\begin{aligned} a &= \frac{1}{5} \\ \text{y} \quad b &= \frac{6}{5}, \end{aligned}$$

se puede aplicar el teorema y verificar que efectivamente:

$$\begin{aligned} \overline{y} &= a\,\overline{x} + b \\ 8.6 &\overset{!}{=} \frac{1}{5} \times 37 + \frac{6}{5}. \end{aligned}$$

Como la transformación $\mathcal{Z}$ es un caso especial de una transformación lineal, se puede aplicar el teorema anterior para comprobar que:

Media aritmética de una variable estandarizada

Si la variable Z se obtuvo por la estandarización de la variable X, es decir, si $Z = \mathcal{Z}(X)$, entonces:

$$\overline{z} = 0.$$

Prueba

Como la estandarización es una transformación lineal con $a = \frac{1}{s_X}$ y $b = -\frac{\overline{x}}{s_X}$, se aplica el teorema anterior para obtener:

$$\begin{aligned} \overline{z} &= a\overline{x} + b \\ &= \frac{1}{s_X}\overline{x} - \frac{\overline{x}}{s_X} \\ &= \frac{s_X}{\overline{x}} - \frac{\overline{x}}{s_X} \\ &= 0. \end{aligned}$$

■

Este resultado no sorprende: Una variable estandarizada informa sobre cuántas desviaciones estándares las observaciones se encuentran debajo o encima de la media aritmética. En promedio, la desviación hacia la media es 0 (véase también la Propiedad 1 de la media).

Ejemplo

En la tabla del ejemplo en la página 113, se presentaron los valores en la variable Z que se obtuvo por una estandarización de la variable *Índice de Masa Corporal* de los datos de la página 34. Calculando la media aritmética de estos 50 valores efectivamente resulta en:

$$\begin{aligned}\overline{z} &= \frac{-1.00 + 0.00 + 0.50 + (-1.25) + 2.00 + (-0.50) + \cdots + 0.25 + 0.25}{50} \\ &= 0.\end{aligned}$$

4.3.3 Efectos en la varianza y desviación estándar

En esta sección se examina el efecto de una trasformación en los estadísticos de varianza y desviación estándar. En los problemas, al final de este tema, se incluyen también algunos ejercicios que implican explorar cómo una transformación afecta a los otros estadísticos de dispersión.

De nuevo, el primer teorema considera el caso de una transformación f *general* de una variable X:

Varianza de una variable transformada

Si la variable Y se obtiene por una transformación f de la variable X, es decir $Y = f(X)$, entonces se puede calcular s_Y^2

(a) a partir de la función de frecuencia de X como sigue:

$$s_Y^2 = \frac{\sum_{j=1}^{m} frec_X(x_j)\left[f(x_j) - \overline{f(x)}\right]^2}{n} \tag{4.13a}$$

(b) a partir de la función de proporción de X como sigue:

$$s_y^2 = \sum_{j=1}^{m} p_X(x_j)\left[f(x_j) - \overline{f(x)}\right]^2. \tag{4.13b}$$

En ambas ecuaciones, $x_1, x_2, \ldots, x_m$ son los m valores distintos observados en la variable X. Nótese que $\overline{f(x)}$ corresponde con $\overline{y}$.

Comprobación de las Ecuaciones (4.13)

Sólo se demuestra la Ecuación (4.13b). La demostración de (4.13a) se desarrolla de forma análoga. Según la Ecuación (3.6b), se puede calcular la varianza de Y como:

$$s_Y^2 = \sum_{k=1}^{r} p_Y(y_k)\,(y_k - \overline{y})^2\,,$$

donde $y_1, y_2, \ldots, y_r$ son los r valores distintos observados en Y.

Desarrollando la expresión del lado derecho resulta en:

$$\begin{aligned}\sum_{k=1}^{r} p_Y(y_k)\,(y_k-\overline{y})^2 &= \sum_{k=1}^{r}\left[\left(\sum_{\substack{x\\ f(x)=y_k}} p_X(x)\right)(y_k-\overline{y})^2\right] \qquad \text{(por la Ecuación (4.9b))}\\ &= \sum_{k=1}^{r}\sum_{\substack{x\\ f(x)=y_k}}\left[p_X(x)\left(f(x)-\overline{f(x)}\right)^2\right].\end{aligned}$$

Dado que los términos de la doble suma en la última expresión corresponden con los m valores distintos de la variable X (véase un argumento similar en la prueba del teorema de la media aritmética de una variable transformada), se obtiene la igualdad:

$$s_Y^2 = \sum_{k=1}^{r} p_Y(y_k)\,(y_k-\overline{y})^2 = \sum_{j=1}^{m} p_X(x)\left[f(x_j)-\overline{f(x)}\right]^2 \qquad \blacksquare$$

Ejemplo

Calculemos la varianza de la variable Y en el ejemplo del profesor del Caso B, utilizando los datos en la tabla de frecuencia de X de la página 123.

$$\begin{aligned} s_Y^2 &= \frac{\displaystyle\sum_{j=1}^{10} \mathit{frec}_X(x_j)\left[f(x_j)-\overline{f(x)}\right]^2}{40}\\ &= \frac{\mathit{frec}_X(x_1)\left[f(x_1)-\overline{f(x)}\right]^2 + \mathit{frec}_X(x_2)\left[f(x_2)-\overline{f(x)}\right]^2 + \cdots + \mathit{frec}_X(x_{10})\left[f(x_{10})-\overline{f(x)}\right]^2}{40}\end{aligned}$$

La media $\overline{y} = \overline{f(x)} = 1.825$, tal como se calculó en el ejemplo de la página 127. Por lo tanto,

$$\begin{aligned} s_Y^2 &= \frac{\mathit{frec}_X(18)\,[f(18)-1.825]^2 + \mathit{frec}_X(19)\,[f(19)-1.825]^2 + \cdots + \mathit{frec}_X(27)\,[f(27)-1.825]^2}{40}\\ &= \frac{2\,(|18-23|-1.825)^2 + 3\,(|19-23|-1.825)^2 + \cdots + 1\,(|27-23|-1.825)^2}{40}\\ &= \frac{77.775}{40}\\ &= 1.9444\end{aligned}$$

La varianza calculada directamente a través de la fórmula (3.6a) (utilizando la función de frecuencia de la variable Y en la página 123) da el mismo resultado:

$$
\begin{aligned}
s_Y^2 &= \frac{\displaystyle\sum_{k=1}^{6} \mathit{frec}_Y(y_k)(y_k - \overline{y})^2}{n} \\
&= \frac{\mathit{frec}_Y(0)(0-1.825)^2 + \mathit{frec}_Y(1)(1-1.825)^2 + \ldots + \mathit{frec}_Y(5)(5-1.825)^2}{40} \\
&= \frac{6(0-1.825)^2 + 15(1-1.825)^2 + \ldots + 2(5-1.825)^2}{40} \\
&= \frac{77.775}{40} \\
&= 1.9444
\end{aligned}
$$

El siguiente teorema permite obtener la varianza y desviación estándar de una variable *linealmente transformada* directamente a partir de la varianza de la variable original:

Varianza y desviación estándar de una variable obtenida por una transformación lineal

Si la variable Y se obtiene por una transformación lineal de la variable X, es decir $Y = aX + b$, con constantes cualesquiera a y b, entonces:

$$s_Y^2 = a^2\, s_X^2 \tag{4.14a}$$

$$\text{y} \quad s_Y = |a|\, s_X \tag{4.14b}$$

Prueba del teorema sobre la varianza y desviación estándar de una variable linealmente transformada

El resultado para la desviación estándar sigue directamente de su definición como la raíz cuadrada de la varianza. Es decir:

$$
\begin{aligned}
s_Y &= \sqrt{s_Y^2} \\
&= \sqrt{a^2\, s_X^2} \\
&= |a|\, s_X .
\end{aligned}
$$

Nótese que es necesario tomar el valor absoluto de la constante multiplicativa a, puesto que el resultado de una raíz cuadrada ($\sqrt{a^2}$) siempre es positivo.

Falta demostrar la Ecuación (4.14a). Partimos de la definición de la varianza de Y:

$$s_Y^2 = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n}$$

Si Y se obtiene por una transformación lineal de X, entonces para cada observación i, se cumple que $y_i = ax_i + b$. Además, como se comprobó en la página 128, $\overline{y} = a\overline{x} + b$. Por lo tanto:

$$\begin{aligned} s_Y^2 &= \frac{\sum_{i=1}^{n}\left[(ax_i + b) - (a\overline{x} + b)\right]^2}{n} \\ &= \frac{\sum_{i=1}^{n}\left[(ax_i - a\overline{x} + b - b)\right]^2}{n} \\ &= \frac{\sum_{i=1}^{n}\left[a\,(x_i - \overline{x})\right]^2}{n} \\ &= \frac{\sum_{i=1}^{n} a^2(x_i - \overline{x})^2}{n}. \end{aligned}$$

Como a es una constante, se puede mover delante del signo sumatorio:

$$\begin{aligned} s_Y^2 &= a^2 \underbrace{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}}_{=s_X^2} \\ &= a^2\, s_X^2 \end{aligned}$$

■

Ejemplo

Calculemos la varianza de las puntuaciones originales X y de las notas Y sobre 10 de los 12 alumnos en el ejemplo del profesor de Bioestadístico, utilizando para ambas la fórmula de la definición de varianza:

$$s_X^2 = \frac{\sum_{i=1}^{12}(x_i - \overline{x})^2}{12}.$$

Para la media $\overline{x}$ se encontró en la página 130 un valor de 37. Por lo tanto:

$$\begin{aligned} s_X^2 &= \frac{(42-37)^2 + (37-37)^2 + (39-37)^2 + \ldots + (26-37)^2}{12} \\ &= \frac{300}{12} \\ &= 25. \end{aligned}$$

Similarmente, para Y se obtiene:

$$s_Y^2 = \frac{\sum_{i=1}^{12}(y_i - \overline{y})^2}{12}.$$

La media $\overline{y}$ resultó ser 8.6, lo cual si se sustituye en la fórmula de la varianza da:

$$\begin{aligned} s_Y^2 &= \frac{(9.6-8.6)^2 + (8.6-8.6)^2 + (9.0-8.6)^2 + \ldots + (6.4-8.6)^2}{12} \\ &= \frac{12}{12} \\ &= 1. \end{aligned}$$

Sin embargo, el último resultado se podría haber calculado de forma más simple, cuando se da cuenta de que $Y = aX + b$, con $a = \frac{1}{5}$ y $b = \frac{6}{5}$:

$$\begin{aligned} s_Y^2 &= a^2\, s_X^2 \\ &= \left(\frac{1}{5}\right)^2 25 \\ &= 1. \end{aligned}$$

A primera vista, puede asombrar el resultado de que la constante aditiva b de una transformación lineal no afecta la varianza (ni la desviación estándar). Sin embargo, sumar la misma constante a todos los valores de una variable, no afecta las diferencias entre los valores. En el ejemplo anterior, la operación que el profesor aplicó para corregir por la dificultad del examen implica sumar 1.2 puntos a cada nota sobre 10; es claro que las diferencias entre (cada par de) alumnos no cambien por esta operación: La distribución se desplaza simplemente a la derecha, pero las diferencias siguen siendo las mismas.

Similarmente, multiplicar una variable por -1 tampoco afecta el tamaño de las diferencias entre los valores. Únicamente cambia la dirección (multiplicar por -1 implica una transformación estrictamente decreciente, es decir el orden entre las observaciones se invierte); sin embargo, el tamaño de las diferencias sigue siendo el mismo y, por lo tanto, la varianza y desviación estándar también.

Por último, se concluye este tema aplicando el teorema anterior a variables estandarizadas:

Varianza y desviación estándar de una variable estandarizada

Si la variable Z se obtuvo por la estandarización de la variable X, es decir, si $Z = \mathcal{Z}(X)$, entonces:

$$\begin{aligned} & s_Z^2 = 1 \\ \text{y} \quad & s_Z = 1 \end{aligned}$$

Prueba

Estandarizar una variable X implica aplicar la transformación lineal

$$\mathcal{Z}(X) = \frac{1}{s_X} X - \frac{\overline{x}}{s_X},$$

así que, aplicando el teorema anterior. se obtiene:

$$\begin{aligned} s_Z^2 &= \left(\frac{1}{s_X}\right)^2 s_X^2 \\ &= 1. \end{aligned}$$

Similarmente,

$$\begin{aligned} s_Z &= \left|\frac{1}{s_X}\right| s_X \\ &= \frac{1}{s_X} s_X \\ &= 1. \end{aligned}$$

■

Ejemplo

Calculemos la varianza de los 50 valores de la variable Z en la tabla en la página 113, que se obtuvo tras una estandarización de la variable *Índice de Masa Corporal* de los datos de la página 34:

$$s_Z^2 = \frac{\sum_{i=1}^{50} (z_i - \overline{z})^2}{50}$$

lo cual, puesto que $\overline{z} = 0$, se simplifica a:

$$\begin{aligned} s_Z^2 &= \frac{\sum_{i=1}^{50} z_i^2}{50} \\ &= \frac{(-1.00)^2 + (0.00)^2 + (0.50)^2 + (-1.25)^2 + \cdots + (0.25)^2 + (0.25)^2}{50} \\ &= 1, \end{aligned}$$

lo cual coincide con el teorema anterior.

Problemas

1. Supongamos que la variable Y se ha obtenido por una transformación estrictamente monótona f de la variable X. Las Ecuaciones (4.10) presentan las fórmulas para obtener la función de frecuencia y proporción de Y a partir de las mismas para la variable X. De forma alternativa, se pueden considerar las funciones de frecuencia y proporción *acumulada*:

   (a) Desarrolla las fórmulas para derivar $cfrec_Y$ y F_Y a partir de $cfrec_X$ y F_X.
   Pista: Conviene separar los casos de transformaciones estrictamente crecientes y decrecientes.

   (b) En el Caso F, se aplicó una transformación logarítmica a los tiempos de reacción. La figura en la Ecuación (4.1) menciona que la media aritmética de la variable transformada es 2.686. Utilizando las fórmulas anteriores, encuentra una expresión para la proporción acumulada $p_Y(2.686)$.

   (c) En el ejemplo de la página 79, se definieron las variables Z (razón Albúmina/Globulina) y Z' (razón Globulina/Albúmina) entre las cuales existe la siguiente relación:

   $$Z' = \frac{1}{Z}.$$

   Utilizando las fórmulas derivas en el inciso (a),

   - Si la proporción acumulada del valor 1 en la variable Z es igual a .0012, entonces ¿cuál es la proporción acumulada del valor 1 en la variable Z'?
   - ¿Qué puedes derivar de la función de proporción acumulada de la variable Z' a partir de la información $F_Z(2) = .967$?
   - Si la mediana de la variable Z es igual a 1.39, ¿cuál es la mediana de Z'?

2. Si la variable Y se obtiene por una transformación f de la variable X (es decir, $Y = f(x)$), ¿de qué tipo debe ser f para que la siguiente relación entre los cuantiles de orden r de ambas variables

   $$y_r = f(x_r)$$

   se cumpla para cualquier r (entre 0 y 1)?

3. Considera una variable Y que se define a partir de una variable X por la siguiente transformación:

   $$Y = \log X$$

   (donde la base del logaritmo es arbitraria).
   Si se conocen las modas de X, ¿qué se puede concluir sobre las modas de Y?

4. ¿Cuál es la relación entre las medianas de las variables X y Y, si Y se obtuvo por

   (a) una transformación lineal de X?

   (b) una transformación estrictamente creciente de X (no necesariamente lineal)?

5. La amplitud total de la variable X se definió como la diferencia entre el valor máximo y el valor mínimo observado:

$$\text{Amplitud total de } X = \max_i x_i - \min_i x_i.$$

Si $Y = f(X)$, ¿de qué tipo debe ser la transformación f para que la amplitud total de la variable Y sea:

$$\text{Amplitud total de } Y = f\left(\max_i x_i\right) - f\left(\min_i x_i\right) \text{ ?}$$

6. Considera las variables X y Y, entre las cuales existe la siguiente relación: $Y = aX + b$, donde a y b son dos constantes cualesquiera. Si se definen las variables Z_1 y Z_2 por la estandarización de X y Y, respectivamente (es decir, $Z_1 = \mathcal{Z}(X)$ y $Z_2 = \mathcal{Z}(Y)$), ¿cuál es la relación entre Z_1 y Z_2?

7. Si Z representa la variable estandarizada correspondiente a la variable *Presión Arterial Sistólica* de los datos de la página 34, ¿cuál es el valor en la variable original de una persona que tiene el valor z_i igual a

   - -1 ?
   - 0 ?
   - $+1$?

   Recuérdese que la media y la varianza de la variable *Presión Arterial Sistólica* son 130 y 289, respectivamente.

Tema 5
Distribuciones bivariadas y multivariadas

Índice

En los temas anteriores se introdujeron una serie de métodos que describen y/o resumen la distribución de una variable; sin embargo, estos métodos trataron las variables por separado, es decir, el análisis no tomé en cuenta otras variables.[*] Las distribuciones que se refieren a una única variable se llaman *distribuciones univariadas* y los estadísticos asociados con estas distribuciones (p.e., la media, la mediana, la varianza, etc.) se llaman *estadísticos univariados*.

En general, las ciencias se dedican a descubrir relaciones entre variables; esto, obviamente, no solo es verdad para las ciencias médicas y de la salud, sino también para las ciencias naturales o exactas (como la física, química, biología, etc.) y las ciencias sociales (psicología, educación, economía, etc.). Por ejemplo, los físicos descubrieron leyes que relacionan fuerza, masa y aceleración ($F = ma$) o energía y materia ($E = mc^2$); los epidemiólogos comprobaron que existe una relación entre los hábitos de fumar y el desarrollo del cáncer de pulmón; según los psicoanalistas, existe una relación entre los traumas en la infancia y las neurosis que se desarrollan en la edad adulta; los sociólogos investigan la relación entre la cifra de desempleo y el índice de suicidio en un país; etc. Cada uno de estos ejemplos implica la relación entre dos o más variables.

[*] Sin embargo, en algunos ejemplos anteriores consideramos la distribución de una variable, tomando en cuenta únicamente aquellas observaciones que tienen un valor específico en otra variable. Como se verá en la Sección 5.4 del tema actual, esta idea corresponde con la de una distribución condicional y examinar la distribución condicional es una forma para estudiar la relación entre dos o más variables.

En los temas que faltan de la primera parte de este libro se introducirán métodos para analizar simultáneamente dos o más variables. Es importante reconocer la similitud del desarrollo de los temas 2, 3 y 4 por un lado y los temas 5, 6 y 7, por otro lado. La siguiente tabla resume dicha correspondencia:

	Estadística descriptiva		
	para una variable		para dos variables
Distribuciones	univariadas	⟷	bivariadas o multivariadas
	Tema 2: Funciones de frecuencia y proporción		Tema 5: Distribuciones bivariadas y multivariadas
Estadísticos	univariados	⟷	bivariados
	Tema 3: Estadísticos de tendencia central y variabilidad		Tema 6: Estadísticos de covariación
Construir nuevas variables	a partir de una variable	⟷	a partir de dos o más variables
	Tema 4: Transformaciones de variables		Tema 7: Sumas de variables

En el tema actual se introducen los conceptos básicos relacionados con *distribuciones bivariadas y multivariadas* y, por lo tanto, es homólogo al Tema 2 donde se introdujeron los conceptos básicos de distribuciones *univariadas*. El siguiente tema presenta unos *estadísticos bivariados* (es decir, estadísticos que resumen, a través de un número, una distribución bivariada), tal como el Tema 3 presentó algunos estadísticos *univariados*. En el Tema 4, se explicó que a veces uno desea construir una nueva variable a partir de otra variable (original), aplicándole una transformación, y se investigó cómo se puede obtener información sobre la distribución de la nueva variable (como su función de frecuencia, la media y la varianza) a partir de la información correspondiente de la variable original; similarmente, el Tema 7 considerará el caso de construir una nueva variable, esta vez *a partir de dos o más variables* originales, y cómo la distribución de esta nueva variable depende de las variables originales. El último capítulo de la primera parte trata un tema particular de distribuciones bivariadas, a saber, la descripción de la relación lineal entre dos variables y cómo utilizar esta relación para predecir valores en una de las variables a partir de la otra.

5.1 Datos bivariados y su representación

Considérese el caso donde se dispone de datos bivariados, es decir, se dispone de los valores de n observaciones en dos variables. Estas variables se denotarán muchas veces X y Y. Respecto del tipo de las variables, cualquier combinación es posible: puede ser que ambas sean cualitativas, ambas cuantitativas, o bien que una sea cualitativa y la otra cuantitativa. Representando los datos bivariados, lo más habitual es utilizar una tabla en la cual hay una columna para cada variable y donde las n filas corresponden con las observaciones y contienen los valores x_i en la variable X y y_i en la variable Y de cada observación i. La siguiente tabla muestra, de forma abstracta, cómo se organiza la información:

	X	Y
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3
$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n

Esta forma de representar datos bivariados (y multivariados) en una tabla es la más común y también es el método que utilizan los programas para análisis estadístico y las hojas de cálculo. Para representar los datos de la página 34, se utilizó el mismo principio: las columnas corresponden con las variables, las filas con las observaciones.

En el caso de que ambas variables sean cuantitativas (o cuasicuantitativas), a menudo se utiliza un *diagrama de dispersión* para representar gráficamente los datos bivariados. En este tipo de diagrama, cada una de las n observaciones se representa por un punto en el plano, donde las dos dimensiones del plano se refieren a las dos variables. Las coordenadas de cada punto son los valores de la observación correspondiente en las dos variables.

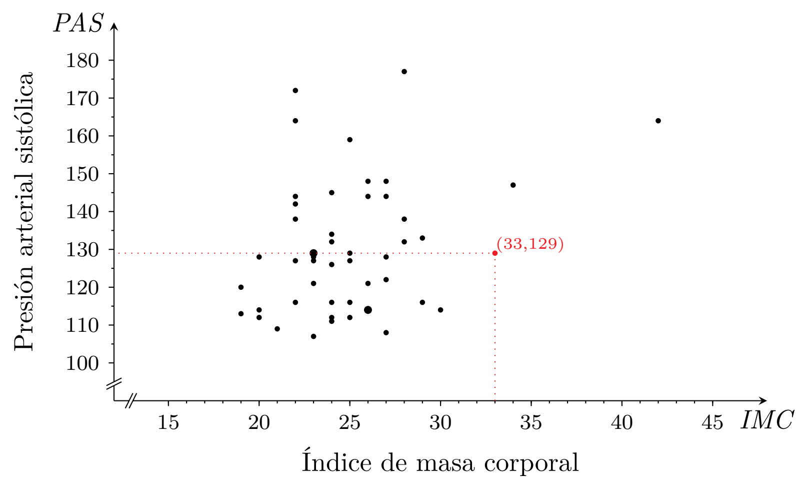
Ejemplo

A continuación, se presentan los datos de las variables *Índice de Masa Corporal* y *Presión Arterial Sistólica* de los datos de la página 34, en forma tabular y gráficamente, mediante un diagrama de dispersión.

Datos bivariados

i	IMC	PAS
1	21	109
2	25	129
3	27	128
4	20	112
5	33	129
6	23	121
7	24	116
8	28	177
9	23	128
10	27	122
11	22	164
12	26	114
13	29	133
$\vdots$	$\vdots$	$\vdots$
49	26	114
50	26	121

Diagrama de dispersión



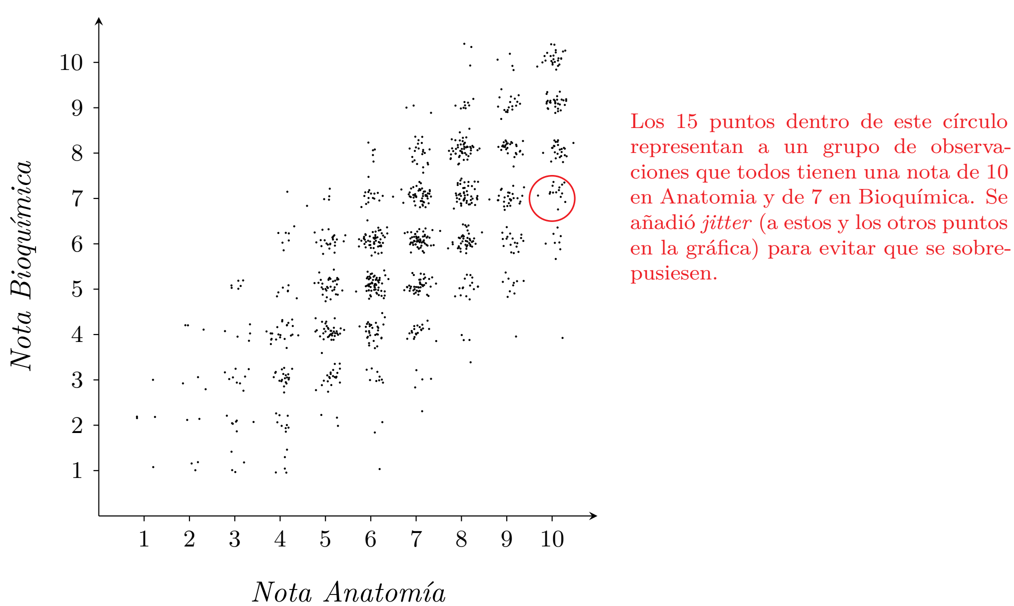
La quinta observación se representa por el punto en color rojo que tiene como coordenadas: la posición 33 en el eje horizontal y la posición 129 en el eje vertical, por los valores que tiene en las variables IMC y PAS, respectivamente.

Superposición de puntos en un diagrama de dispersión

El lector atento habrá notado que en el diagrama de dispersión del ejemplo anterior, había dos puntos con un grosor diferente a los demás puntos. Esta diferencia se origina en que cada uno de estos puntos representa dos observaciones: Las observaciones 12 y 49 del conjunto de datos tienen precisamente los mismos valores en ambas variables; lo mismo es cierto para las observaciones 25 y 47. Para que un diagrama de dispersión no genere una imagen distorcionada de la relación bivariada representada, es importante aclararlo cuando los puntos de diferentes observaciones se sobreponen. Una forma para hacer esta aclaración es—como se hizo en la gráfica anterior—utilizar un grosor diferente para estos puntos; otra posibilidad es representarlos en un color distinto o con otro símbolo (por ejemplo, un rumbo en vez de un círculo), entre otros.

En el caso de que las variables tengan relativamente pocos valores distintos y para cada combinación de valores en ambas variables existan múltiples observaciones, una posible técnica para resolver el problema de la superposición de puntos en el diagrama de dispersión, se conoce como "*jittering*" (un término en inglés que significa "añadir una pequeña variación"). Esto significa que cada punto se mueve ligeramente de su posición para evitar que se sobreponga con otros puntos.

Para ilustrar esta técnica, supongamos que los profesores responsables de las asignaturas de Anatomía y Bioquímica en el primer año de la carrera de Médico Cirujano de la UNAM desean construir un diagrama de dispersión de las notas finales (sobre 10) que los estudiantes obtuvieron. Como las notas finales se presentan como números enteros, sin decimales, y el número de estudiantes es muy alto, habrá múltiples estudiantes que tienen la misma nota en ambas asignaturas, por ejemplo, una nota de 10 en Anatomía y 7 en Bioquímica. En la siguiente gráfica se ha aplicado *jittering* para resolver el problema de que los puntos de estos estudiantes se sobrepongan.



Está fuera del alcance de este texto describir en detalle el procedimiento para aplicar *jitter* a un diagrama de dispersión.

5.2 Función de frecuencia y proporción para dos variables

5.2.1 Definición, interpretación y propiedades

Para un conjunto de datos bivariados se puede considerar para cada par de valores (x, y) la frecuencia con que ocurre este par en los datos, es decir, el número de observaciones que adoptan simultáneamente el valor x en la variable X y el valor y en la variable Y.

Frecuencia y proporción bivariada

Para (a) una variable X que asume m distintos valores x_1, $x_2, \ldots,$ x_m,
(b) una variable Y que asume r distintos valores y_1, $y_2, \ldots,$ y_r,
y (c) un conjunto de n observaciones con valores en ambas variables,

la *función de frecuencia bivariada* asocia cada par de valores (x_j, y_k) con un número natural que indica el número de observaciones que tienen el valor x_j en la variable X y el valor y_k en la variable Y.

Notación: $frec(x_j, y_k)$ o $frec_{XY}(x_j, y_k)$

Similarmente, la *función de proporción bivariada* asocia cada par de valores (x_j, y_k) con un número del intervalo $[0, 1]$ que indica la proporción de observaciones que tienen el valor x_j en la variable X y el valor y_k en la variable Y.

Notación: $p(x_j, y_k)$ o $p_{XY}(x_j, y_k)$

Por definición, la proporción bivariada de cualquier par de valores (x_j, y_k) se obtiene de la frecuencia correspondiente como:

$$p(x_j, y_k) = \frac{frec(x_j, y_k)}{n}$$

Ejemplo

Consideremos las variables *Nivel Educativo* y *Fumador* en los datos de la página 34. Se observa, por ejemplo, que dos personas tienen el valor “Primaria” para *Nivel Educativo* y el valor “No fuma” para *Fumador*. Por lo tanto, la función de frecuencia bivariada de las variables X y Y asocia el valor de 2 con el par (Primaria, No fuma); la función de proporción bivariada asocia el valor de $\frac{2}{50} = .04$ al mismo par.

La siguiente tabla da la frecuencia y proporción bivariada para todos los pares que resultan como combinaciones de valores en las variables *Nivel Educativo* y *Fumador*.

j	k	x_j	y_k	$frec_{XY}(x_j, y_k)$	$p_{XY}(x_j, y_k)$
1	1	Primaria	No fuma	2	.04
1	2	Primaria	Bajo	1	.02
1	3	Primaria	Intensivo	1	.02
2	1	Secundaria	No fuma	11	.22
2	2	Secundaria	Bajo	2	.04
2	3	Secundaria	Intensivo	5	.10
3	1	Bachillerato	No fuma	8	.16
3	2	Bachillerato	Bajo	5	.10
3	3	Bachillerato	Intensivo	3	.06
4	1	Universitario	No fuma	11	.22
4	2	Universitario	Bajo	1	.02
4	3	Universitario	Intensivo	0	.00
				50	1.00

Propiedades de frecuencias y proporciones bivariadas

Para cualquier par de valores (x_j, y_k), se cumple:

- $0 \leqslant frec_{XY}(x_j, y_k) \leqslant n$ • $0 \leqslant p_{XY}(x_j, y_k) \leqslant 1$ (5.1a)

En general, se cumple que:

- $\sum_{j=1}^{m}\sum_{k=1}^{r} frec_{XY}(x_j, y_k) = n$ • $\sum_{j=1}^{m}\sum_{k=1}^{r} p_{XY}(x_j, y_k) = 1$ (5.1b)

Para cada valor x_j de la variable X, se cumple:

- $\sum_{k=1}^{r} frec_{XY}(x_j, y_k) = frec_X(x_j)$ • $\sum_{k=1}^{r} p_{XY}(x_j, y_k) = p_X(x_j)$ (5.1c)

Para cada valor y_k de la variable Y, se cumple:

- $\sum_{j=1}^{m} frec_{XY}(x_j, y_k) = frec_Y(y_k)$ • $\sum_{j=1}^{m} p_{XY}(x_j, y_k) = p_Y(y_k)$ (5.1d)

Estas propiedades siguen directamente de la definición de las funciones bivariadas de frecuencia y proporción. El siguiente ejemplo aclara e ilustra su interpretación.

Ejemplo

Las fórmulas (5.1a) indican que la frecuencia bivariada siempre se encuentra entre 0 y n, el número de observaciones, y que la proporción bivariada es un número entre 0 y 1. Por otro lado, las dos ecuaciones en (5.1b) quieren decir que, al considerar todos los posibles pares de valores en las dos variables involucradas, las frecuencias se suman a n, mientras que las proporciones se suman a 1. Es importante tener claro que estas propiedades son las homólogas de las propiedades de frecuencias y proporciones univariadas que se presentaron en las páginas 40 y 41.

El doble signo sumatorio en las dos fórmulas de (5.1b) implica una suma de $m \times r$ términos. Efectivamente, el número de pares de valores que se pueden componer a partir de m distintos valores en X y r distintos valores en Y es igual al producto de m y r. Por ejemplo, en la tabla del ejemplo anterior, que presenta las frecuencias y proporciones bivariadas para todos los pares de valores en las variables *Nivel Educativo* y *Fumador*, se observan 12 pares (x_j, y_k), que son el resultado de combinar cada uno de los 4 valores de *Nivel Educativo* con los 3 valores en *Fumador*. La tabla también muestra que las 12 frecuencias correspondientes suman a 50 (el número de observaciones), y las 12 proporciones a 1, tal como especifican las ecuaciones de (5.1b).

Las ecuaciones en (5.1c) consideran un valor x_j particular de la variable X y dicen que sumar todas las frecuencias o proporciones bivariadas que incluyen x_j de X da como resultado la frecuencia univariada de este valor en la variable X. Por ejemplo, volvamos a considerar las mismas frecuencias bivariadas del ejemplo anterior, de las variables *Nivel Educativo* y *Fumador*, y apliquemos la formúla izquierda en (5.1c) al valor x_1 ("Primaria") de la variable X (*Nivel Educativo*). Efectivamente,

$$\begin{aligned}\sum_{k=1}^{3} \mathit{frec}_{XY}(x_1, y_k) &= \mathit{frec}_{XY}(x_1, y_1) + \mathit{frec}_{XY}(x_1, y_2) + \mathit{frec}_{XY}(x_1, y_3) \\ &= \mathit{frec}_{XY}(\text{“Primaria”}, \text{“No fuma”}) + \mathit{frec}_{XY}(\text{“Primaria”}, \text{“Bajo”}) \\ &\qquad + \mathit{frec}_{XY}(\text{“Primaria”}, \text{“Intensivo”}) \\ &= 2 + 1 + 1 \\ &= 4 \;=\; \mathit{frec}_{X}(x_1),\end{aligned}$$

es decir, la frecuencia de "Primaria" para la variable *Nivel Educativo*. En palabras, la frecuencia (univariada) del valor "Primaria" coincide con la suma de las frecuencias de todos los pares que se obtienen combinando este valor con cada uno de los valores en la variable *Fumador*. Finalmente no sorprende este resultado: El número de personas con nivel primaria en nuestro conjunto de datos es igual a (a) el número de personas con nivel primaria que no fuman, más (b) el número de personas con nivel primaria que son fumadores de baja intensidad más (c) el número de personas con nivel primaria que son fumadores intensivos.

Finalmente, las últimas ecuaciones reconsideran la propiedad anterior, que se fija en un valor de X, desde la perspectiva de la variable Y. Dicen que la frecuencia o proporción univariada de un valor y_k en la variable Y coincide con la suma de las frecuencias o proporciones bivariadas de todos los pares de valores que involucran este valor de Y. Por ejemplo, si queremos conocer la proporción de personas que son fumadores intensivos (es decir, $p_Y(y_3)$), se suma:

$$\begin{aligned}\sum_{j=1}^{4} p_{XY}(x_j, y_3) &= p_{XY}(x_1, y_3) + p_{XY}(x_2, y_3) + p_{XY}(x_3, y_3) + p_{XY}(x_4, y_3) \\ &= p_{XY}(\text{“Primaria”}, \text{“Intensivo”}) + p_{XY}(\text{“Secundaria”}, \text{“Intensivo”}) \\ &\qquad + p_{XY}(\text{“Bachillerato”}, \text{“Intensivo”}) + p_{XY}(\text{“Universitario”}, \text{“Intensivo”}) \\ &= .02 \ + \ .10 \ + \ .06 \ + \ .00 \\ &= .18 \ = \ p_Y(y_3).\end{aligned}$$

5.2.2 Tabla de contingencia y representación gráfica

Aunque la tabla de frecuencias y proporciones en la página 144 es una herramienta válida y correcta para representar la distribución bivariada de las variables *Nivel Educativo* y *Fumador*, existe una alternativa, a saber, la *tabla de contingencia*, que tiene la ventaja de presentar la misma información de una forma más clara. Una tabla de contingencia es una tabla de doble entrada (filas y columnas), en la cual cada fila se refiere a un valor de la primera variable y cada columna a un valor de la otra variable y donde cada celda muestra la frecuencia o proporción de observaciones que tienen la combinación de valores correspondientes a la fila y la columna. El siguiente ejemplo aclara la idea.

Ejemplo

La siguiente tabla de contingencia muestra las funciones de frecuencia (en color azul) y proporción (en color rojo) bivariada para las variables *Nivel Educativo* y *Fumador* de los datos de la página 34:

	Nivel Educativo				
Fumador	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
	.04	.22	.16	.22	.64
Bajo	1	2	5	1	9
	.02	.04	.10	.02	.18
Intensivo	1	5	3	0	9
	.02	.10	.06	.00	.18
Total	4	18	16	12	50
	.08	.36	.32	.24	1.00

En una tabla de contingencia habitualmente se añade una fila por el lado inferior y una columna al lado derecho que contienen las frecuencias y proporciones univariadas de las dos variables ($frec_X(x_j)$ y $frec_Y(y_k)$). Dichas frecuencias y proporciones se llaman *frecuencias marginales* o *proporciones marginales*. En este caso, el *gran total* (es decir, el número total de observaciones) está en la celda inferior derecha.

Tal como existen varias formas para representar distribuciones univariadas gráficamente, también se desarrollaron gráficas para representar las distribuciones bivariadas. A continuación se presentan algunos ejemplos de diagramas que representan la distribución bivariada (de frecuencia o proporción) de las variables *Nivel Educativo* y *Fumador* de los datos de la página 34.

Diagrama de rectángulos conjuntos para *Nivel Educativo* y *Fumador*

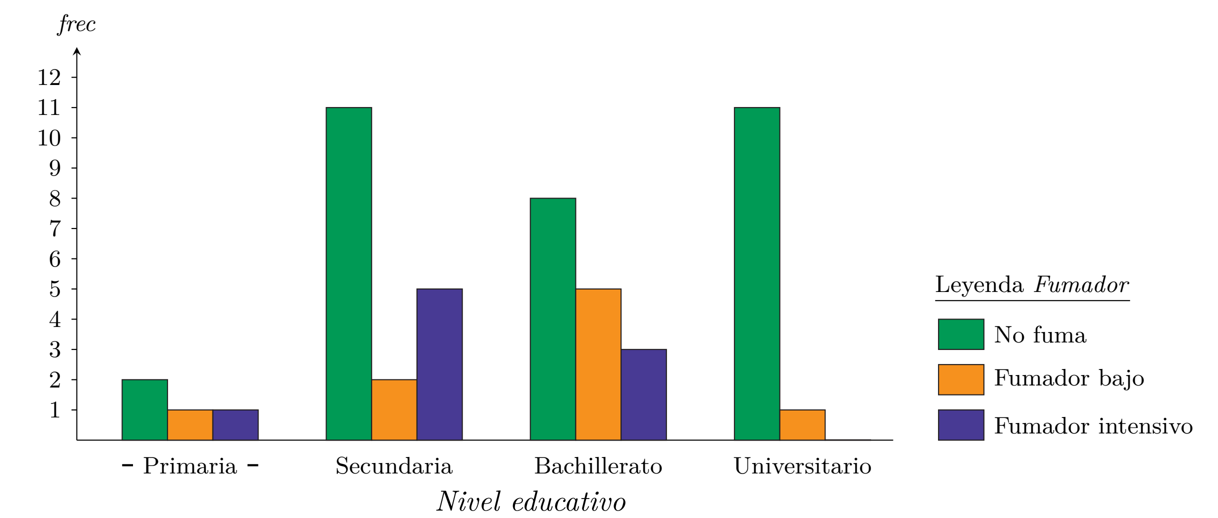


Diagrama de rectángulos partidos para *Nivel Educativo* y *Fumador*

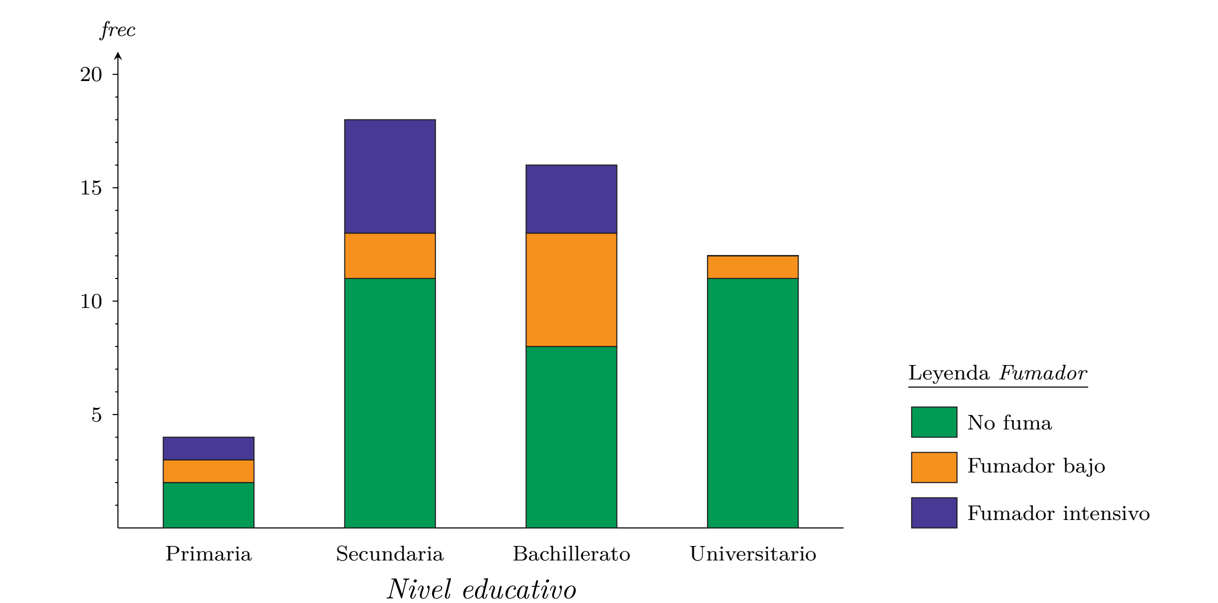
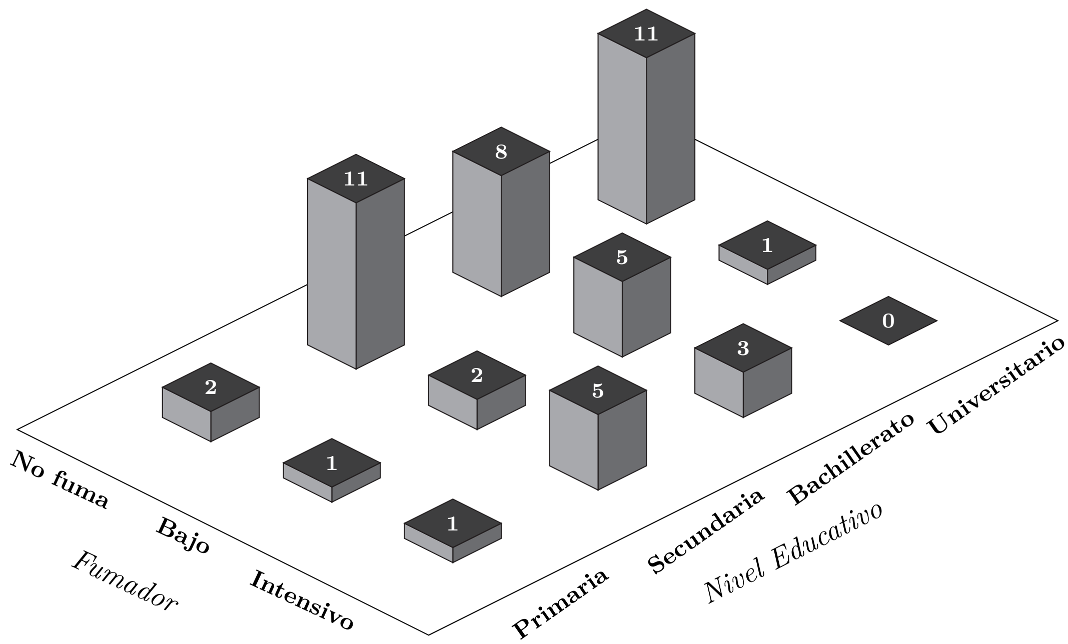


Diagrama tridimensional de barras para *Nivel Educativo* y *Fumador*



5.3 Función de frecuencia y proporción acumulada para dos variables

Para dos variables (cuasi)cuantitativas, se puede considerar la función de frecuencia *acumulada* de estas variables, de forma muy similar a cómo se definió la función de frecuencia acumulada para una variable cuantitativa (véase la Sección 2.2.3). En particular, la frecuencia acumulada para el par de valores (x, y) en las variables X y Y se define como el número de observaciones que tienen un valor menor que o igual a x en la variable X y un valor menor que o igual a y en la variable Y. La siguiente definición formaliza esta idea:

Función bivariada de frecuencia acumulada

Para (a) una variable X que asume m distintos valores $x_1, x_2, \ldots, x_m$,
(b) una variable Y que asume r distintos valores $y_1, y_2, \ldots, y_r$,
y (c) un conjunto de n observaciones con valores en ambas variables,

la *función bivariada de frecuencia acumulada* asocia cada par de valores (x_j, y_k) con un número natural que indica el número de observaciones que tienen

el valor x_j o un valor menor en la variable X
y
el valor y_k o un valor menor en la variable Y

Notación: $cfrec(x_j, y_k)$ o $cfrec_{XY}(x_j, y_k)$

Suponiendo que $x_1 < x_2 < x_3 < \ldots < x_m$ y $y_1 < y_2 < y_3 < \ldots < y_r$, entonces se calcula la frecuencia acumulada bivariada para cada par (x_j, y_k):

$$cfrec(x_j, y_k) = \sum_{h=1}^{j}\sum_{l=1}^{k} frec(x_h, y_l) \tag{5.2}$$

Ejemplo

Evaluemos $cfrec_{XY}(x_j, y_k)$, donde X y Y son las variables *Nivel Educativo* y *Fumador*, respectivamente, y los valores x_j y y_k son respectivamente el tercero y segundo valor en las variables anteriores, es decir, $x_j = x_3 =$ "Bachillerato" y $y_k = y_2 =$ "Fumador bajo". Los valores que preceden "Bachillerato" en X son "Primaria" y "Secundaria"; en la variable Y, el valor "No fuma" es el que precede "Fumador bajo". Por lo tanto, para conocer la frecuencia acumulada, se realizan las siguientes operaciones especificadas por la Ecuación (5.2):

$$\begin{aligned} cfrec(x_3, y_2) &= \sum_{h=1}^{3}\sum_{l=1}^{2} frec(x_h, y_l) \\ &= frec(x_1, y_1) + frec(x_1, y_2) + frec(x_2, y_1) + frec(x_2, y_2) + frec(x_3, y_1) + frec(x_3, y_2) \\ &= frec(\text{“Primaria”}, \text{“No fuma”}) + frec(\text{“Primaria”}, \text{“Bajo”}) \\ &\quad + frec(\text{“Secundaria”}, \text{“No fuma”}) + frec(\text{“Secundaria”}, \text{“Bajo”}) \\ &\quad + frec(\text{“Bachillerato”}, \text{“No fuma”}) + frec(\text{“Bachillerato”}, \text{“Bajo”}) \\ &= 2 + 1 + 11 + 2 + 8 + 5 \ = \ 29. \end{aligned}$$

Se puede utilizar el mismo formato que una tabla de contingencia para resumir la función bivariada de frecuencia acumulada. Las siguientes tablas ilustran el proceso.

Frecuencias

	Nivel Educativo			
Fumador	Primaria	Secundaria	Bachillerato	Universitario
No fuma	2	11	8	11
Bajo	1	2	5	1
Intensivo	1	5	3	0

Frecuencias *acumuladas*

	Nivel Educativo			
Fumador	Primaria	Secundaria	Bachillerato	Universitario
No fuma	2	13	21	32
Bajo	3	16	29	$\boxed{41}$
Intensivo	4	$\boxed{22}$	38	50

La tabla superior muestra las frecuencias bivariadas, mientras que la tabla inferior muestra las mismas, pero acumuladas. Si las filas y columnas están ordenadas de menor a mayor valor en las variables X y Y, entonces la Ecuación (5.2) implica que la frecuencia acumulada $cfrec_{XY}(x, y)$ corresponde con la suma de todas las frecuencias de las celdas que se encuentran en las filas superiores y las columnas izquierdas de la celda que corresponde con el par (x, y). Las tablas anteriores ilustran esta idea para *cfrec*("Universitario", "Fumador bajo"), cuyo valor de 41 corresponde con la suma de las frecuencias en el rectángulo rojo de la tabla superior. Similarmente, *cfrec*("Secundaria", "Fumador intensivo") = 22, que se obtiene por la suma de las frecuencias en el rectángulo azul de la misma tabla.

El cuadro que sigue formaliza el concepto análogo de una *proporción* bivariada acumulada.

Función bivariada de proporción acumulada

Para (a) una variable X que asume m distintos valores $x_1, x_2, \ldots, x_m$,
(b) una variable Y que asume r distintos valores $y_1, y_2, \ldots, y_r$,
y (c) un conjunto de n observaciones con valores en ambas variables,

la *función bivariada de proporción acumulada* asocia cada par de valores (x_j, y_k) con un número del intervalo $[0, 1]$ que indica la proporción de observaciones que tienen

el valor x_j o un valor menor en la variable X
y
el valor y_k o un valor menor en la variable Y

Notación: $F(x_j, y_k)$ o $F_{XY}(x_j, y_k)$

Las siguientes relaciones siempre se cumplen, para cualquier par (x_j, y_k):

- $F(x_j, y_k) = \dfrac{cfrec(x_j, y_k)}{n}$;
- Suponiendo que $x_1 < x_2 < x_3 < \ldots < x_m$ y $y_1 < y_2 < y_3 < \ldots < y_r$, entonces:

$$F(x_j, y_k) = \sum_{h=1}^{j} \sum_{l=1}^{k} p(x_h, y_l).$$

Ejemplo

La tabla que se presenta a continuación muestra las proporciones acumuladas de las variables *Nivel Educativo* y *Fumador* del ejemplo anterior.

	Nivel Educativo			
Fumador	Primaria	Secundaria	Bachillerato	Universitario
No fuma	.04	.26	.42	.64
Bajo	.06	.32	.58	.82
Intensivo	.08	.44	.76	1.00

5.4 Distribuciones condicionales

Una forma común para estudiar la relación entre dos variables consiste en explorar la distribución de una de las variables, no en la muestra total, sino en una submuestra de observaciones que se define a partir de sus valores en la otra variable. Esta idea ya se aplicó en la página 65, donde se consideró la distribución en la variable *Presión Arterial Diastólica* únicamente dentro de las submuestras de observaciones con un valor específico para *Grupo Sanguíneo*. En particular, se calculó la mediana de *Presión Arterial Diastólica* de forma separada para las personas con la sangre tipo O, las personas con tipo A, aquellas con tipo B y, finalmente las del grupo sanguíneo AB. La distribución de una variable únicamente considerando una submuestra de observaciones con un valor específico en otra variable se denomina *distribución condicional*. A partir de la distribución condicional, se definen la *función de proporción condicional* y *estadísticos condicionales* (es decir, la media, mediana, varianza, etc., de la distribución condicional).

Función de proporción condicional

Función de proporción condicional

Para (a) una variable X que asume m distintos valores x_1, $x_2, \ldots,$ x_m,
(b) una variable Y que asume r distintos valores y_1, $y_2, \ldots,$ y_r,
y (c) un conjunto de n observaciones con valores en ambas variables,

la *función de proporción condicional*, $p_{X|Y}(x_j|y_k)$, asocia cada valor x_j con un número del intervalo $[0,1]$ que indica la proporción de observaciones que tienen el valor x_j en la variable X, *dentro de la submuestra de observaciones con el valor y_k en la variable Y*. En particular, se define para todos los x_j y y_k:

$$p_{X|Y}(x_j|y_k) = \frac{frec_{XY}(x_j, y_k)}{frec_Y(y_k)}$$

Ejemplo

Consideremos la distribución de la variable *Nivel Educativo* (X), condicional al valor “No fuma” de la variable *Fumador* (Y) en los datos de la página 34. La función de frecuencia y proporción asociada con esta distribución condicional se presenta en la siguiente tabla:

j	x_j	$frec_{X\|Y}(x_j\|\text{“No fuma”})$	$p_{X\|Y}(x_j\|\text{“No fuma”})$
1	Primaria	2	$\frac{2}{32} = .062$
2	Secundaria	11	$\frac{11}{32} = .344$
3	Bachillerato	8	$\frac{8}{32} = .250$
4	Universitario	11	$\frac{11}{32} = .344$
Total		32	$\frac{32}{32} = 1.000$

En la muestra hay 32 personas que no fuman. A partir de esta submuestra de 32 personas se define la distribución de *Nivel Educativo* condicional al valor “No fuma” de la variable *Fumador*.

Se observa, por ejemplo, que la proporción condicional p(“Bachillerato”|“No fuma”) = .25, es decir, un cuarto de las personas que no fuman tienen Bachillerato como *Nivel Educativo*.

De la misma forma, se pueden considerar la distribuciones de *Nivel Educativo*, condicional a los otros valores de *Fumador*. A continuación, se muestran las funciones de frecuencia y proporción condicionales a “fumador bajo” y “fumador intensivo”, respectivamente:

j	x_j	$frec_{X\|Y}(x_j\|\text{“Bajo”})$	$p_{X\|Y}(x_j\|\text{“Bajo”})$
1	Primaria	1	$\frac{1}{9} = .111$
2	Secundaria	2	$\frac{2}{9} = .222$
3	Bachillerato	5	$\frac{5}{9} = .556$
4	Universitario	1	$\frac{1}{9} = .111$
Total		9	$\frac{9}{9} = 1.000$

j	x_j	$frec_{X\|Y}(x_j\|\text{“Intensivo”})$	$p_{X\|Y}(x_j\|\text{“Intensivo”})$
1	Primaria	1	$\frac{1}{9} = .111$
2	Secundaria	5	$\frac{5}{9} = .556$
3	Bachillerato	3	$\frac{3}{9} = .333$
4	Universitario	0	$\frac{0}{9} = .000$
Total		9	$\frac{9}{9} = 1.000$

Aunque se puede hablar de (y mostrar, como se acaba de ejemplificar en las tablas anteriores) la función de frecuencia condicional, es poco común utilizar este concepto; en distribuciones condicionales es más habitual considerar la función de proporción. Esto, sin duda, tiene que ver con que la función de frecuencia condicional coincide con la función de frecuencia bivariada. Obsérvese, por ejemplo, que la frecuencia condicional $frec_{X|Y}$(“Bachillerato”|“Intensivo”) = 3 en la tabla anterior es precisamente la frecuencia bivariada $frec_{XY}$(“Bachillerato”, “Intensivo”) = 3 que se encuentra en la tabla de la página 144.

Obviamente, se puede invertir el papel de las variables X y Y al estudiar distribuciones condicionales. Al respecto, es importante hacer constar que la proporción condicional $p_{X|Y}(x, y)$ generalmente difiere de la proporción condicional $p_{Y|X}(y, x)$. Es decir, la función de proporción no es simétrica en sus argumentos. (La proporción bivariada, por otro lado, sí lo es: Para cualquier par de valores x y y se cumple que $p_{XY}(x, y) = p_{YX}(y, x)$.)

Ejemplo

Es importante tener muy claro que en las distribuciones condicionales, las dos variables juegan un papel diferente y que, por lo tanto, intercambiarlas lleva a resultados distintos. Ilustrémoslo siguiendo con el mismo ejemplo de las variables *Nivel Educativo* y *Fumador* de los datos de la página 34.

En el ejemplo anterior se calculó la función de proporción de *Nivel Educativo* condicional a distintos valores de *Fumador*. El cuadro siguiente organiza los resultados que se obtuvieron en una tabla:

Distribución condicional $p_{X|Y}$ (con *Nivel Educativo* X y *Fumador* Y)

	Nivel Educativo				
Fumador	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	.062	.344	.250	.344	1.000
Bajo	.111	.222	.556	.111	1.000
Intensivo	.111	.556	.333	.000	1.000
p_X	.080	.360	.320	.240	1.000

Sin embargo, intercambiando el papel de las dos variables, es decir, calculando ahora las proporciones en la variable *Fumador* condicionales a distintos valores de *Nivel Educativo*, lleva a un resultado distinto:

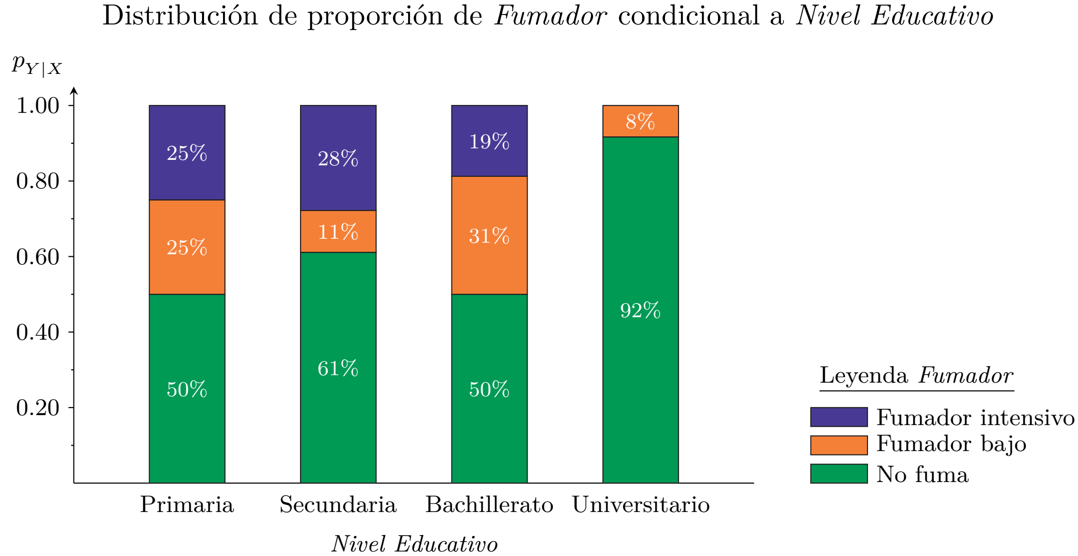
Distribución condicional $p_{Y|X}$ (con *Nivel Educativo* X y *Fumador* Y)

	Nivel Educativo				
Fumador	Primaria	Secundaria	Bachillerato	Universitario	p_Y
No fuma	.500	.611	.500	.917	.640
Bajo	.250	.111	.313	.083	.180
Intensivo	.250	.278	.187	.000	.180
Total	1.000	1.000	1.000	1.000	1.000

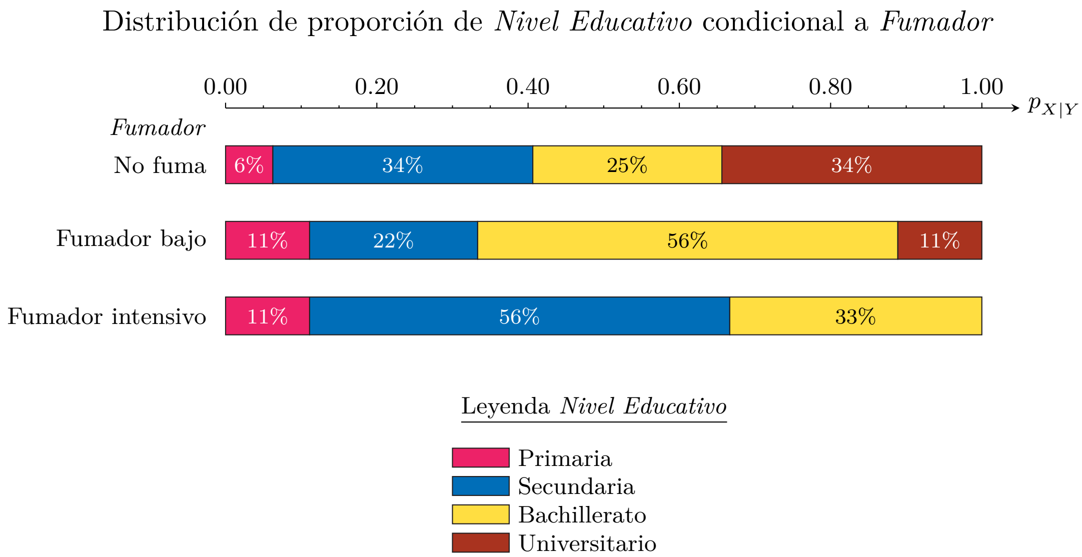
Se observa, por ejemplo que $p_{X|Y}$(“Primaria”|“No fuma”) $\neq p_{Y|X}$(“No fuma”|“Primaria”); al final de cuentas, la proporción de personas con nivel primaria dentro del grupo de no fumadores tiene otro concepto que la proporción de personas que no fuman dentro del grupo de personas con nivel primaria.

Una herramienta muy útil para representar gráficamente la distribución de X condicional a diferentes valores de Y es el diagrama de rectángulos partidos, el cual tiene una variante con rectángulos verticales y otra con rectángulos horizontales.

Diagramas de rectángulos partidos para *Nivel Educativo* (X) y *Fumador* (Y)



Esta gráfica muestra claramente que dentro del grupo de universitarios, predominan los no fumadores. Similarmente, en el grupo de personas con nivel de bachillerato, los fumadores bajos tienen una representación relativamente mayor, en comparación con su representación en los grupos de otro nivel educativo.



Se observa que en el grupo de no fumadores, la proporción de personas con nivel educativo de secundaria es igual a la proporción de universitarios. Sin embargo, en la submuestra de fumadores intensivos, las personas con nivel secundaria representa a más de la mitad de la personas, mientras que los universitarios ni están presentes en este grupo.

Estadísticos condicionales En el Tema 3 se introdujo una serie de estadísticos para resumir la distribución de una variable X. Los mismos estadísticos se pueden utilizar para resumir distribuciones condicionales. En este caso, se habla de la *media aritmética condicional*, la *amplitud total condicional*, la *varianza condicional*, etc.

Ejemplo

En la página 65, se calculó la mediana de la variable *Presión Arterial Diastólica* (X), condicional a diferentes valores de la variable *Grupo Sanguíneo* (Y). Se encontró que:

- $Me_X \mid (Y = \text{"O"}) = 82.5$
- $Me_X \mid (Y = \text{"A"}) = 86$
- $Me_X \mid (Y = \text{"B"}) = 84$
- $Me_X \mid (Y = \text{"AB"}) = 82$

Para construir el diagrama de caja y bigotes en la página 102, se consideraron las distribuciones de *Índice de Masa Corporal* (X) condicionales a *Sexo* (Y). Este diagrama requirió calcular, entre otros estadísticos, la media aritmética y los tres cuartiles de *IMC*, condicionales a (*Sexo* = "Femenino") y (*Sexo* = "Masculino"). Se encontró que:

$$
\begin{aligned}
\overline{x} \mid (Y = \text{"Femenino"}) &= 24 & \qquad \overline{x} \mid (Y = \text{"Masculino"}) &= 26.27 \\
Q_1 \mid (Y = \text{"Femenino"}) &= 22 & \qquad Q_1 \mid (Y = \text{"Masculino"}) &= 23 \\
Me_X \mid (Y = \text{"Femenino"}) &= 24 & \qquad Me_X \mid (Y = \text{"Masculino"}) &= 25 \\
Q_3 \mid (Y = \text{"Femenino"}) &= 26 & \qquad Q_3 \mid (Y = \text{"Masculino"}) &= 28
\end{aligned}
$$

Se puede calcular que la desviación estándar de la *Presión Arterial Diastólica* (X_1) y *Sistólica* (X_2), condicional al *Sexo* (Y) es igual a:

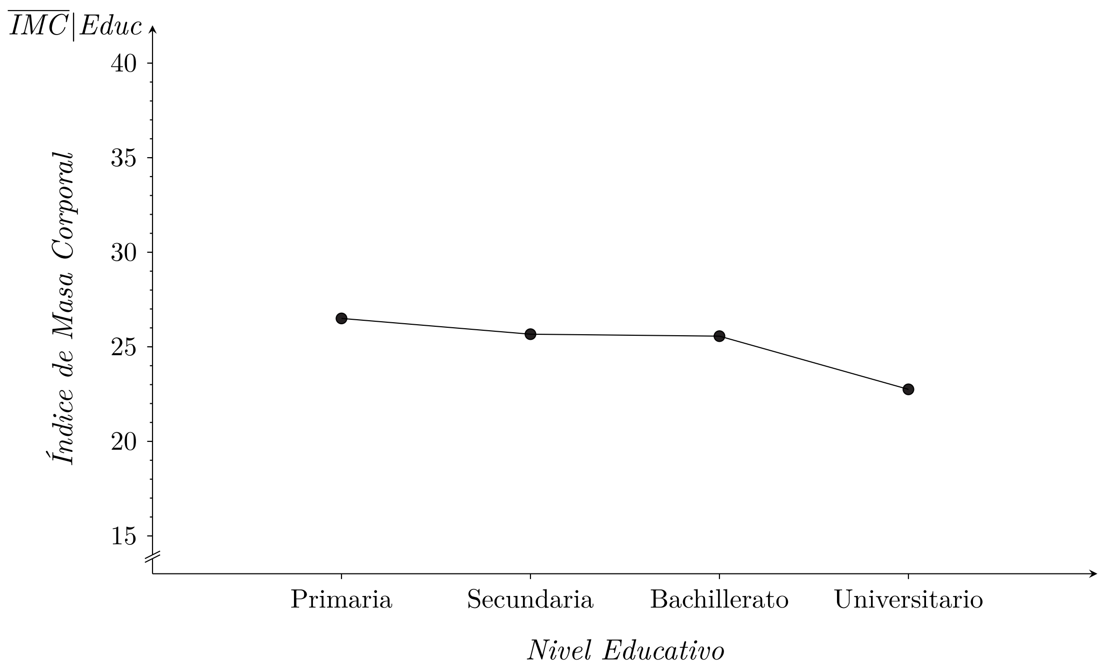
$$
\begin{aligned}
s_{X_1} \mid (Y = \text{"Femenino"}) &= 9.92 & \qquad s_{X_2} \mid (Y = \text{"Femenino"}) &= 11.81 \\
s_{X_1} \mid (Y = \text{"Masculino"}) &= 14.60 & \qquad s_{X_2} \mid (Y = \text{"Masculino"}) &= 17.70
\end{aligned}
$$

Se descubre que, tanto para *Presión Arterial Diastólica* como para *Presión Arterial Sistólica*, la desviación estándar en el grupo de los hombres es mayor que para las mujeres. En esta muestra, hay más variabilidad respecto de la presión arterial entre hombres que entre mujeres.

Las medias aritméticas de una variable condicionales a valores de otra variable suelen representarse en un diagrama de líneas. Por ejemplo, al calcular la media aritmética de *Índice de Masa Corporal* (X) condicional para cada nivel de la variable *Nivel Educativo* (Y) en los datos de la página 34, se obtiene:

- $\overline{x}\,|\,(Y = \text{“Primaria”}) = 26.50$
- $\overline{x}\,|\,(Y = \text{“Secundaria”}) = 25.67$
- $\overline{x}\,|\,(Y = \text{“Bachillerato”}) = 25.56$
- $\overline{x}\,|\,(Y = \text{“Universitario”}) = 22.75$,

lo cual se puede representar en un diagrama de líneas como sigue:



A veces, se añade también información sobre la variabilidad condicional, usando “bigotes” (como en un diagrama de caja y bigotes) para representar el primer y el tercer cuartil, o los valores correspondientes a $\overline{x} - s_X$ y $\overline{x} + s_X$.

5.5 Una característica de distribuciones bivariadas: Covariación

Al estudiar distribuciones bivariadas, el interés principal suele enfocarse en la relación entre las dos variables. En la estadística, se utiliza el término *covariación* (a veces, *correlación*) para referirse a la relación entre variables. La covariación es una característica de distribuciones bivariadas tal como la tendencia central y la variabilidad son características de distribuciones univariadas. El término covariación origina de las palabras “variación conjunta” (co-variación), es decir, la variación entre los valores en una variable y cómo depende de la variación entre los valores en la otra variable.

Una forma para estudiar la covariación en una distribución bivariada, consiste en examinar la distribución de una variable condicional a la otra, a través de las proporciones condicionales o los estadísticos condicionales, tal como se explicó en la sección anterior. Por ejemplo, el diagrama de líneas a finales del ejemplo anterior permite concluir que, en promedio, el índice de masa corporal tiende a disminuir conforme el nivel educativo esté más elevado.

Sin embargo, a través de las distribuciones condicionales es difícil obtener una idea general de la covariación entre las variabels. Lo que se desea típicamente es poder expresar la covariación en un solo número, es decir, a través del valor en un estadístico de covariación, tal como se resume la tendencia central a través de la media aritmética o la mediana de una variable. En el siguiente tema se presentarán varios estadísticos para cuantificar la covariación en una distribución bivariada. No obstante, antes de introducir estos índices, la primera subsección de la sección actual define el concepto de *independencia* entre las dos variables de una distribución bivariada, el cual se puede interpretar como "ausencia de covariación". A continuación, se presentan varios ejemplos que ilustrar los aspectos generales de la covariación entre dos variables.

5.5.1 Independencia de variables

Para introducir independencia entre variables en una distribución bivariada, consideremos el siguiente ejemplo.

Ejemplo: independencia de variables

Supongamos que se recopilaron datos en una muestra de 120 personas respecto de las variables *Grupo Sanguíneo* y *Fumador*. Para simplificar la exposición, estas dos variables se definen de la misma forma que las variables con el mismo nombre en los datos de la página 34 (es decir, se observaron los mismos cuatro valores para *Grupo Sanguíneo* y los mismos tres valores para *Fumador*). Sin embargo, la muestra, tanto el tamaño como la distribución de frecuencia y proporción de estas variables, es distinta. En particular, la tabla de contingencia correspondiente a estas variables en la muestra de 120 personas se presenta a continuación:

Tabla de contingencia

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Fumador bajo	10	8	4	2	24
Fumador intensivo	5	4	2	1	12
Total	50	40	20	10	120

A partir de la tabla de contingencia, se pueden calcular las proporciones marginales de ambas variables, asi como la función de proporción de una variable condicional a los valores de la otra variable. La siguiente tabla muestra la distribución de proporción de *Grupo Sanguíneo* condicional a los valores de *Fumador* y, en la última fila, las proporciones marginales de *Grupo Sanguíneo*.

Tabla de la distribución condicional $p_{Y|X}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	$\frac{35}{84} = .417$	$\frac{28}{84} = .333$	$\frac{14}{84} = .167$	$\frac{7}{84} = .083$	$\frac{84}{84} = 1.000$
Fumador bajo	$\frac{10}{24} = .417$	$\frac{8}{24} = .333$	$\frac{4}{24} = .167$	$\frac{2}{24} = .083$	$\frac{24}{24} = 1.000$
Fumador intensivo	$\frac{5}{12} = .417$	$\frac{4}{12} = .333$	$\frac{2}{12} = .167$	$\frac{1}{12} = .083$	$\frac{12}{12} = 1.000$
p_Y	$\frac{50}{120} = .417$	$\frac{40}{120} = .333$	$\frac{20}{120} = .167$	$\frac{10}{120} = .083$	$\frac{120}{120} = 1.000$

Se observa un patrón muy particular: Las proporciones con que ocurren los cuatro grupos sanguíneos son idénticos para cada nivel de fumador: Tanto en la submuestra de los no fumadores, como en las submuestras de fumadores bajos y fumadores intensivos, el 41.7 % de la muestra tiene grupo O, el 33.3 % tiene el grupo A, el 16.7 % tiene el grupo B y el 8.3 % tiene el grupo AB. En otras palabras, la distribución de los grupos sanguíneos es idéntica *independientemente* del valor en la variable *Fumador*. Por consiguiente, estas distribuciones condicionales coinciden también perfectamente con la distribución marginal de la variable *Grupo Sanguíneo*.

¿Qué pasaría si se intercambiase el papel de las dos variables, es decir, si se calculase la distribución marginal de *Fumador* y la distribución de esta variable condicional a *Grupo Sanguíneo*? La siguiente tabla proporciona esta información.

Tabla de la distribución condicional $p_{X|Y}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	p_X
No fuma	$\frac{35}{50} = .700$	$\frac{28}{40} = .700$	$\frac{14}{20} = .700$	$\frac{7}{10} = .700$	$\frac{84}{120} = .700$
Fumador bajo	$\frac{10}{50} = .200$	$\frac{8}{40} = .200$	$\frac{4}{20} = .200$	$\frac{2}{10} = .200$	$\frac{24}{120} = .200$
Fumador intensivo	$\frac{5}{50} = .100$	$\frac{4}{40} = .100$	$\frac{2}{20} = .100$	$\frac{1}{10} = .100$	$\frac{12}{120} = .100$
Total	$\frac{50}{50} = 1.000$	$\frac{40}{40} = 1.000$	$\frac{20}{20} = 1.000$	$\frac{10}{10} = 1.000$	$\frac{120}{120} = 1.000$

Volvemos a observar el mismo patrón: La forma de cómo se distribuyen los tres niveles de *Fumador* es idéntica en las cuatro submuestras definidas por el grupo sanguíneo. En particular, tanto en el grupo O, como en el grupo A, el grupo B y el grupo AB, 70 % de las personas no fuman, el 20 % son fumadores bajos y el 10 % son fumadores intensivos, y estas proporciones condicionales coinciden con las proporciones marginales de *Fumador* en la última columna de la tabla.

Si la función de proporción para una variable X, condicional a la variable Y, es idéntica para cualquier valor y de Y, entonces, las dos variables son independientes. Al evaluar independencia, no importa el papel de las dos variables; es decir, evaluar si la función de proporción condicional de X sea idéntica para distintos valores de Y, o bien, evaluar si la función de proporción condicional de Y sea idéntica para distintos valores de X llevaría a la misma conclusión.

Además existe una tercera forma para evaluar independencia. Para ilustrarla, se presentan en la siguiente tabla las proporciones bivariadas de las variables *Nivel Educativo* y *Fumador* en la muestra de 120 personas:

Tabla de la distribución bivariada p_{XY}

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	p_X
No fuma	$\frac{35}{120} = .292$	$\frac{28}{120} = .233$	$\frac{14}{120} = .117$	$\frac{7}{120} = .058$	$\frac{84}{120} = .700$
Fumador bajo	$\frac{10}{120} = .083$	$\frac{8}{120} = .067$	$\frac{4}{120} = .033$	$\frac{2}{120} = .017$	$\frac{24}{120} = .200$
Fumador intensivo	$\frac{5}{120} = .042$	$\frac{4}{120} = .033$	$\frac{2}{120} = .017$	$\frac{1}{120} = .008$	$\frac{12}{120} = .100$
p_Y	$\frac{50}{120} = .417$	$\frac{40}{120} = .333$	$\frac{20}{120} = .167$	$\frac{10}{120} = .083$	$\frac{120}{120} = 1.000$

Se puede verificar que para cada par de valores (x, y) en las variables X y Y (es decir, para cada una de las 12 combinaciones que se obtienen por los cuatro grupos sanguíneos y los tres niveles de fumador), se cumple la siguiente condición:

$$p_{XY}(x, y) \;=\; p_X(x)\, p_Y(y).$$

En particular, se verifica que:

$$\begin{aligned}
p_{XY}(\text{“No fuma”}, \text{“O”}) \;&=\; p_X(\text{“No fuma”})\, p_Y(\text{“O”}) \;=\; .700 \times .417 = .292\\
p_{XY}(\text{“No fuma”}, \text{“A”}) \;&=\; p_X(\text{“No fuma”})\, p_Y(\text{“A”}) \;=\; .700 \times .333 = .233\\
&\vdots\\
p_{XY}(\text{“Intensivo”}, \text{“AB”}) \;&=\; p_X(\text{“Intensivo”})\, p_Y(\text{“AB”}) \;=\; .100 \times .083 = .008
\end{aligned}$$

Si para cada celda la proporción bivariada es igual al producto de las dos proporciones marginales correspondientes, entonces las dos variables de la tabla son independientes.

En el contexto de evaluar independencia en la tabla de proporciones bivariadas, se suele hablar de la *proporcionalidad* de filas y columnas. Esto quiere decir que, en caso de independencia, los valores de dos filas cualesquiera (o dos columnas cualesquiera) mantienen una razón constante; por ejemplo, en la tabla anterior, se observa que las proporciones en la segunda fila son el doble de las proporciones en la tercera fila, y las de la primera fila son 3.5 veces las proporciones correspondientes de la segunda fila. Asímismo, aplica el principio de proporcionalidad entre las columnas: por ejemplo, las proporciones en la cuarta columna son la mitad de las proporciones en la tercera columna y la primera columna es 1.25 veces la segunda columna.

Definición: Independencia de dos variables

Dos variables X y Y son *independientes* si, y sólo si, se cumple cualquiera de las siguientes tres condiciones:

$$\text{(a) } p_{X|Y}(x|y) = p_X(x) \text{ para todos los pares } (x, y) \text{ con } \mathit{frec}_Y(y) > 0. \tag{5.3a}$$

$$\text{(b) } p_{Y|X}(y|x) = p_Y(y) \text{ para todos los pares } (x, y) \text{ con } \mathit{frec}_X(x) > 0. \tag{5.3b}$$

$$\text{(c) } p_{XY}(x, y) = p_X(x)\, p_Y(y) \text{ para todos los pares } (x, y). \tag{5.3c}$$

Equivalencia de las tres condiciones para independencia

Demostramos la equivalencia de las tres condiciones. Para cualquier valor x de la variable X y cualquier valor y en la variable Y, se pueden establecer las siguientes equivalencias:

$$
\begin{aligned}
& p_{X|Y}(x|y) = p_X(x) && (5.3a) \\
\iff\quad & \frac{p_{XY}(x,y)}{p_Y(y)} = p_X(x) && \text{(a)} \Leftrightarrow \text{(c)} \\
\iff\quad & p_{XY}(x,y) = p_X(x)\, p_Y(y) && (5.3c) \qquad \text{(a)} \Leftrightarrow \text{(b)} \\
\iff\quad & \frac{p_{XY}(x,y)}{p_X(x)} = p_Y(y) && \text{(b)} \Leftrightarrow \text{(c)} \\
\iff\quad & p_{Y|X}(y|x) = p_Y(y) && (5.3b)
\end{aligned}
$$

En las primeras dos ecuaciones, hay que añadir que $p_Y(y) > 0$ o $\mathit{frec}_Y(y) > 0$ ya que la proporción condicional o la división entre $p_Y(y)$ es indefinida si $p_Y(y) = 0$. Por razones similares, hay que añadir que $p_X(x) > 0$ o $\mathit{frec}_X(x) > 0$ en las últimas dos ecuaciones. ■

5.5.2 Formas de covariación

Si las dos variables no son independientes, se puede distinguir varias formas de cómo la una depende de la otra. Los diagramas de dispersión en la siguiente página dan algunos ejemplos de diferentes tipos de covariación en distribuciones bivariadas. En todos los diagramas, excepto (h), las variables son dependientes, pero la covariación entre ellas toma formas diferentes. En los diagramas (a) y (b), todas las observaciones (es decir, todos los puntos) se encuentran en una línea, y por lo tanto, la covariación entre X y Y en estos diagramas es *lineal*. Además, en el diagrama (a), observaciones que tienen (en comparación con otras observaciones) valores altos en X, también tienen valores altos en Y; y observaciones con valores bajos en X tienen valores bajos en Y también. Entonces, se trata de una relación *lineal directa* entre ambas variables. Por otro lado, la covariación en el diagrama (b) es *lineal inversa*: A medida que la observación tiene un valor alto en X, tiene un valor bajo en Y, y viceversa, un valor bajo en X va a la mano de un valor alto en Y.

En los diagramas (c) y (d), la relación entre X y Y es *curvilínea*. Existen varios tipos de relaciones curvilíneas. Como ejemplo, se muestran una relación *exponencial* en el diagrama (c) y una relación *polinómica* (del segundo grado, o bien, cuadrática) en el diagrama (d).

Diagramas de dispersión ilustrando diferentes tipos de covariación

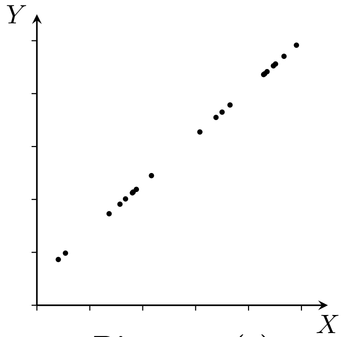


Diagrama (a)

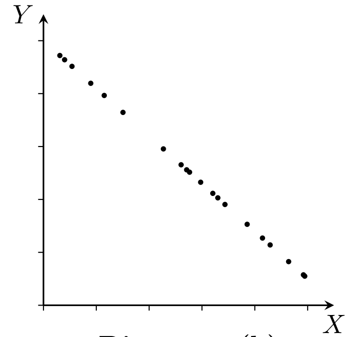


Diagrama (b)

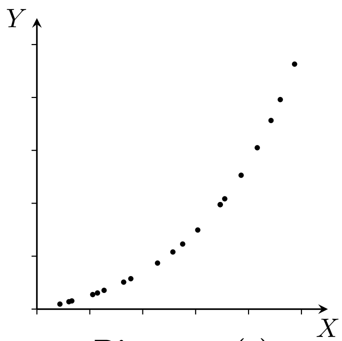


Diagrama (c)

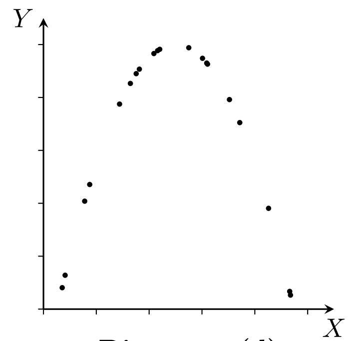


Diagrama (d)

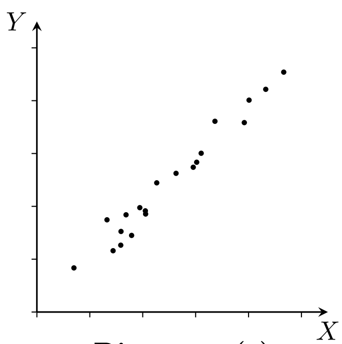


Diagrama (e)

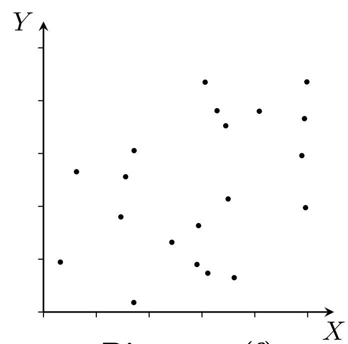


Diagrama (f)

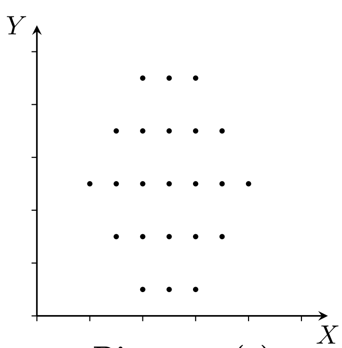


Diagrama (g)

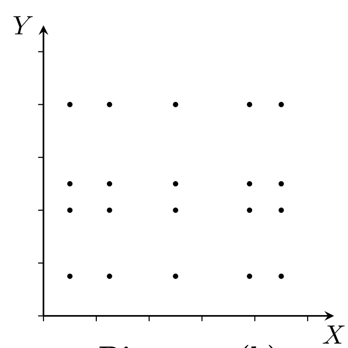


Diagrama (h)

Además del tipo de la función que describe mejor la relación entre las dos variables (como lineal o curvilínea), también se puede evaluar la fuerza de la covariación. En los diagramas de (a) a (d), la covariación es perfecta: Todos los puntos se encuentran en la recta o en la curva exponencial o polinómica. En los diagramas (e) y (f), se aprecia una relación lineal directa: Valores altos en X *generalmente* van acompañados con valores altos en Y y valores bajos en X van acompañados con valores bajos en Y. Sin embargo, la relación *no es perfectamente* lineal: Hay desviaciones de la recta. En el diagrama (e), la relación lineal, aunque no es perfecta, todavía es bastante fuerte (las desviaciones son menores), mientras que en el diagrama (f) la relación lineal es más débil.

Por último, examinemos los diagramas (g) y (h). En ninguno de ellos se puede descubrir un patrón específico para describir la covariación entre X y Y. (Nótese que las medias condicionales de Y dado un valor de X son todas iguales, y viceversa.) El diagrama (h) es un ejemplo de dos variables independientes y, como se explicó en la subsección anterior, independencia implica "ausencia de covariación". El diagrama (g) es un ejemplo de que las dos variables aparentemente no muestran covariación a pesar de que las dos variables no son independientes (nótese que, en este diagrama, las varianzas condicionales de Y dado los valores de X no son todas iguales —entonces, las distribuciones condicionales son diferentes— y, por lo tanto X y Y, son dependientes).

Como conclusión de esta sección, recuérdese que la covariación de dos variables se puede describir al respecto de:

1. El *tipo* de covariación: lineal (directa o inversa) o curvilínea (exponencial, polinómica, etc.)
2. La *magnitud* o la *fuerza* de la covariación: perfecta, fuerte, débil o ausente.

5.6 Generalización a distribuciones multivariadas

Las ideas expuestas en este tema para el caso de dos variables se pueden extender directamente a casos con tres o más variables. Así, por ejemplo, $\mathit{frec}_{XYZ}(x, y, z)$ es una frecuencia trivariada; más específico, es el número de observaciones que adoptan simultáneamente el valor x en la variable X, el valor y en la variable Y y el valor z en la variable Z. En el caso de que X, Y y Z sean variables cuantitativas, se puede considerar la proporción acumulada $F_{XYZ}(x, y, z)$, la cual corresponde con la proporción de observaciones que tienen el valor x o un valor menor en la variable X, el valor y o un valor menor en la variable Y y el valor z o un valor menor en la variable Z. La representación mediante una tabla de contingencia o una gráfica de las funciones de frecuencia o proporción trivariada suele ser más complicada. Una solución existe en presentar una serie de tablas de contingencia bivariadas para diferentes pares de variables. Otra solución, adecuada para diagramas de dispersión, es representar los valores en una tercera variable (posiblemente cualitativa) mediante el uso de diferentes colores o diferentes símbolos. Por ejemplo, para los puntos que representan a los hombres se utiliza color azul; los puntos que representan a las mujeres se pueden colorear en rojo.

También es posible considerar distribuciones condicionales. Por ejemplo, para tres variables X, Y y Z, se pueden considerar las proporciones condicionales en la variable X, dado el valor y en Y y el valor z en Z; asimismo, se puede estudiar la función de proporción bivariada de X y Y, condicional a diferentes valores de Z. Como ejemplo concreto, para los datos de la página 34, se puede calcular la proporción de fumadores intensivos dentro de la submuestra de mujeres con nivel educativo de secundaria. Puesto que dicha submuestra incluye a 10 personas, de las cuales 2 son fumadores intensivas, la proporción condicional buscada es $\frac{2}{10}$ o 20 %.

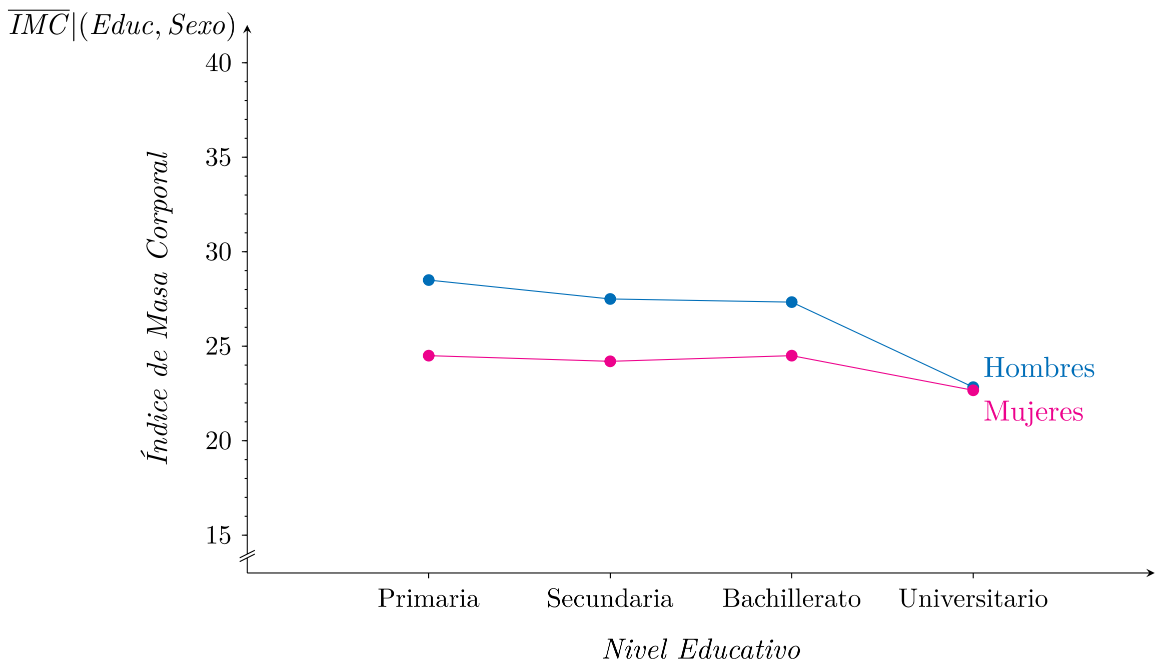
Obviamente, de estas distribuciones condicionales se puede calcular la media, la desviación típica o cualquier otro estadístico. Por ejemplo, calculando la media condicional en la variable

Índice de Masa Corporal, dentro del grupo de diez mujeres con nivel educativo de secundaria, se obtiene 24.2. Por otro lado, es posible examinar la covariación entre dos variables (y calcular uno de los estadísticos que se introducirán en el siguiente tema), condicional a una tercera variable. Por ejemplo, puede ser interesante estudiar la relación entre el *Índice de Masa Corporal* y la *Presión Arterial Diastólica*, tanto en la submuestra de hombres y la submuestra de mujeres.

Un ejemplo de un análisis que estudia la relación entre tres variables simultáneamente es el análisis de varianza bifactorial. Para representar e interpretar los resultados de este análisis es muy común utilizar una extensión del tipo de diagrama que se presentó en la página 156. En particular, se presenta la media de una variable, condicional a la combinación de valores en dos otras variables. El siguiente ejemplo ilustra la idea.

Ejemplo

Al estudiar cómo el *Índice de Masa Corporal* depende de las variables *Nivel Educativo* y *Sexo*, conviene utilizar un diagrama de líneas, que representa la media de *Índice de Masa Corporal*, para cada combinación de *Nivel Educativo* y *Sexo*. Es decir, se calcula la media de *IMC* dentro del grupo de mujeres de nivel primaria, el grupo de mujeres de nivel secundaria, etc.. A continuación se representan dichas medias en un diagrama de líneas donde se utilizan líneas de diferentes colores para representar las medias correspondientes en las submuestras de hombres y en las submuestras de mujeres.



Esta gráfica visualiza claramente la diferencia que existe entre hombres y mujeres respecto del *Índice de Masa Corporal*: la diferencia entre ambos sexos se manifiesta únicamente en los niveles educativos de primaria, secundaria y bachillerato; en la submuestra de universitarios, las medias del *IMC* para hombres y mujeres coinciden.

Por último, el concepto de *independencia* también se generaliza a más de dos variables. La definición formal se presenta en el siguiente cuadro.

Definición: Independencia de k variables

Las variables $X_1, X_2, X_3, \ldots X_k$ son *independientes* si, y sólo si, se cumple la siguiente condición para cualquier combinación de valores x_1 de X_1, x_2 de X_2, …, x_k de X_k:

$$p_{X_1 X_2 \ldots X_k}(x_1, x_2, \ldots, x_k) \;=\; p_{X_1}(x_1)\, p_{X_2}(x_2)\; \ldots\; p_{X_k}(x_k)$$

En palabras, k variables son conjuntamente independientes si y solo si la proporción conjunta de cualquier combinación de valores en las k variables es igual al producto de las k proporciones marginales correspondientes.

Problemas

1. A todos los estudiantes de nuevo ingreso a la Facultad de Medicina de la Universidad Nacional Autónoma de México UNAM) se les aplica un examen diagnóstico en el primer semestre de la carrera. Dicho examen se divide en ocho áreas: matemáticas, física, química, biología, historia universal, historia de México, literatura y geografía. Un investigador del centro de servicios escolares de la misma facultad dispone de las calificaciones de cada estudiante en las distintas áreas del examen diagnóstico. Además, tiene las respuestas de cada participante en unas preguntas socioeconómicas, que incluyen, entre otras, la escuela de procedencia del estudiante. Las escuelas de procedencia se clasifican en tres categorías: las Escuelas Nacionales Preparatorias (ENP), los Colegios de Ciencias y Humanidades (CCH) y otras escuelas, que no forman parte de los sistemas de educación media superior que ofrece la UNAM.

   En este ejercicio, investigamos la distribución bivariada de las calificaciones en el área de geografía (X) y la escuela de procedencia (Y). La calificación en geografía es una nota sobre 10 (el número de respuestas correctas en 10 preguntas de opción múltiple). La siguiente tabla muestra las frecuencias bivariadas en una tabla de contingencia.

	Escuela de procedencia (Y)		
Nota geografía (X)	ENP	CCH	No UNAM
0	2	0	1
1	6	23	6
2	36	44	11
3	70	68	16
4	90	69	20
5	93	52	34
6	69	35	41
7	47	12	52
8	23	11	45
9	12	6	32
10	5	0	16

   (a) Completa la tabla de contingencia con las frecuencias marginales.

   (b) Construye una tabla similar que representa

      i. las proporciones bivariadas p_{XY};

      ii. las proporciones condicionales $p_{X|Y}$;

      iii. las proporciones condicionales $p_{Y|X}$;

   (c) Representa la proporción condicional $p_{X|Y}$ gráficamente por un diagrama de rectángulos partidos.

   (d) ¿Son independientes las variables X y Y en esta muestra? Evalúa la independencia a tráves de cada una de las tres condiciones equivalentes de la definición en la página 160.

   (e) Calcula la media y la desviación estándar de la nota en geografía, condicional a los distintos valores de la escuela de procedencia. Interpreta los resultados y construye una hipótesis que los explique.

2. Los siguientes datos salieron de una investigación en la que se estudia la relación entre la conducta fumadora de futuros padres durante el embarazo y el peso del niño a nacer. Concretamente, se registró para cada una de 1,000 parejas que acababan de tener un hijo, si la madre fumaba durante el embarazo (variable X), si el padre fumaba durante el embarazo (Y), y el peso al nacer del niño en gramos (Z). Los datos se resumieron en la siguiente tabla:

Conducta de la madre x	Conducta del padre y	Proporción bivariada $p_{XY}(x, y)$	Media condicional $\overline{z}\vert(X = x, Y = y)$
"No fuma"	"No fuma"	0.60	3,500
"No fuma"	"Sí fuma"	0.15	3,400
"Sí fuma"	"No fuma"	0.15	3,200
"Sí fuma"	"Sí fuma"	0.10	3,000

(a) Representa mediante una gráfica adecuada el efecto de la conducta fumadora de la madre y del padre en el peso al nacer del bebé. ¿Cómo resumirías el resultado?

(b) ¿Cuál es la media aritmética del peso del bebé dentro de la submuestra de madres fumadores? ¿y dentro de la submuestra de madres que no fuman?

(c) Para este conjunto de datos, ¿el hecho de que la madre sea fumadora es independiente de que el padre sea fumador?

3. El grupo sanguíneo en el sistema ABO está controlado por un solo gen con tres alelos que se nombran: alelo A, alelo B y alelo O. El alelo O es recesivo respecto a los alelos A y B, mientras que A y B son codominantes; esto quiere decir que la relación entre el genotipo (la combinación de (a) el alelo de la copia materna del gen y (b) el alelo de la copia paterna) y el fenotipo (uno de los cuatro grupos sanguíneos) es la siguiente:

		Alelo materno		
		A	B	O
Alelo paterno	A	**A**	**AB**	**A**
	B	**AB**	**B**	**B**
	O	**A**	**B**	**O**

Consideremos ahora un grupo de 10,000 personas y deja que X represente el alelo materno y Y el alelo paterno de cada individuo. Entonces, estas dos variables asumen tres valores: A, B y O. Supongamos que tienen la siguiente distribución de proporción:

$$\begin{aligned} p_X(\text{"A"}) = p_Y(\text{"A"}) &= .27, \\ p_X(\text{"B"}) = p_Y(\text{"B"}) &= .06, \\ p_X(\text{"O"}) = p_Y(\text{"O"}) &= .67. \end{aligned}$$

Si la variable Z representa el fenotipo del grupo sanguíneo cuyos cuatro valores se obtienen de las variables X y Y según el esquema de la tabla arriba, ¿cuál será la función de frecuencia $frec_Z$ si las variables X y Y son independientes?

4. En un estudio se investigó la relación entre la hipertensión arterial (HTA) y la obesidad. En específico, se observó en una muestra de 240 personas las siguientes dos variables:

   - *Presión arterial* en cuatro niveles: (i) Normal ($PAS < 140$ y $PAD < 90$, donde PAS y PAD representan la presión arterial sistólica y diastólica, respectivamente, en mmHg), (ii) HTA estadio 1 ($140 \leqslant PAS < 160$ o $90 \leqslant PAD < 100$), (iii) HTA estadio 2 ($160 \leqslant PAS < 180$ o $100 \leqslant PAD < 110$) y (iv) HTA grave (($PAS \geqslant 180$ o $PAD \geqslant 110$);
   - *Peso* con tres niveles: (i) Normal ($IMC \leqslant 25$, donde IMC es el índice de masa corporal en kg/m^2), (ii) Sobrepeso ($25 < IMC \leqslant 30$), (iii) Obesidad ($IMC > 30$).

   La función de proporción acumulada de las dos variables se presenta en la siguiente tabla:

	Presión arterial			
Peso	Normal	HTA estadio 1	HTA estadio 2	HTA grave
Peso normal	.305	.380	.400	.400
Sobrepeso	.500	.680	.750	.750
Obesidad	.560	.870	.980	1.000

   (a) Construye una tabla de contingencia para las frecuencias bivariadas.

   (b) La hipótesis de los investigadores es que la obesidad puede tener un efecto perjudicial en la hipertensión. En esta muestra, ¿son dependientes las dos variables?

   (c) Construye una gráfica que represente cómo la presión arterial depende del peso. Argumenta la selección del estadístico que has elegido y comenta los resultados.

Tema 6
Estadísticos de covariación

Índice

En el tema anterior se identificó la covariación como una característica de distribuciones bivariadas, similar a que la tendencia central es una característica de distribuciones univariadas. En el tema actual se introducen diferentes estadísticos que cuantifican la covariación entre dos variables. Al respecto, es importante distinguir entre estadísticos de covariación para variables cualitativas (Sección 6.1) y para variables cuantitativas (Sección 6.2).

6.1 Covariación entre variables cualitativas

6.1.1 La proporción de acuerdo

En varios contextos, es muy común pedir el juicio de uno o más expertos sobre una persona o un objeto. Por ejemplo, para obtener una medida de la calidad de una tesis doctoral, se pide el juicio de un tribunal de investigadores expertos en el área de la tesis. En los juegos olímpicos, la actuación de los atletas en las disciplinas deportivas de gimnasia, patinaje artístico, natación sincronizada o saltos es apreciada por un grupo de jueces con mucha experiencia al respecto. Otro ejemplo son los concursos musicales, como el *Concurso Internacional Reina Elisabeth* que se organiza anualmente en Bélgica y donde un jurado compuesto de músicos de fama internacional y profesores destinguidos eligen al laureado.

En los ejemplos anteriores, los juicios tienen consecuencias altas para las personas evaluadas: la evaluación de la tesis doctoral afectará en gran medida la carrera académica y/o las oportunidades que se le ofrecen al candidato en su vida futura. Asimismo, la carrera de un músico joven puede recibir un empujón significativo después de ganar el *Concurso Internacional Reina Elisabeth* y regresar de los juegos olímpicos con una medalla en la maleta hace un mundo de diferencia para los atletas.

Por lo tanto, es muy importante, en cada uno de los contextos anteriormente descritos, que haya un alto grado de concordancia entre los juicios de los expertos: mucha gente estaría de acuerdo que sería injusto que, en los juegos olímpicos la distribución de las medallas en gimnasia cambiaría si otros jueces hubiesen evaluado a los atletas. Asímismo, se espera que el resultado en la defensa de la tesis doctoral o la decisión sobre el laureado en el concurso musical no cambien si otro tribunal o jurado hubiese sido electo para evaluar la calidad de los candidatos.

También en el contexto médico, son muy comunes los juicios de personas y objetos y es de igual importancia que los juicios sean (relativamente) constantes entre diferentes jueces. Por ejemplo, si dos médicos opinan sobre el mismo caso respecto de la necesidad de una operación, es de esperar que lleguen a la misma conclusión. En caso contrario, puede ser que se someta el paciente a una operación innecesaria, o aún peor, que le nieguen una operación que le salve la vida. Para otro ejemplo, en el ámbito de la educación médica, considérese el examen clínico objetivo estructurado (ECOE) en la Facultad de Medicina, el cual cada estudiante debe pasar para su titulación. Este examen práctico consiste en una serie de estaciones, en las cuales diferentes profesores evalúan el desempeño de los estudiantes. Obviamente, si diferentes profesores llegasen a calificaciones (muy) distintas para el mismo desempeño, habría suficiente fundamento para la apertura de procedimientos legales con las instancias superiores.

En todos los casos anteriores, aunque es normal—hasta cierto punto—que haya diferencias entre los juicios de diferentes evaluadores, es igualmente evidente que, en caso de diferencias grandes entre los juicios sobre una misma persona o un mismo objeto, la validez de los juicios obtenidos es cuestionable. Los estadísticos que se proponen en esta y la siguiente subsección son útiles para cuantificar el grado de acuerdo entre dos jueces. En este sentido, son estadísticos particulares de covariación entre dos variables.

Ejemplo

Para un ejemplo más formal, considérese el siguiente escenario: Un psiquiatra y un psicólogo trabajan en el mismo equipo multidisplinario de un hospital psiquiátrico. Cada vez que un nuevo paciente llega al centro, la enfermera psiquiátrica del mismo equipo le realiza un diagnóstico y escribe un reporte detallado en el cual resume la información principal de una entrevista con el paciente y de los resultados en diversas pruebas psicológicas, incluyendo, entre otros, el test de Rorschach, el Inventario Multifásico de Personalidad de Minnesotta (MMPI) y el Test de Apercepción Temática. Con base en el reporte de su asistente, el psiquiatra y el psicólogo toman una decisión sobre el tratamiento, es decir, la terapia más adecuada para el paciente. Se consideran cuatro tipos de terapias: (a) terapia farmacológica, (b) terapia psicoanalítica,

(c) terapia conductual y (d) terapia cognitiva.

El psiquiatra y el psicólogo quieren evaluar hasta que grado coinciden sus ideas sobre cuál es la terapia más adecuada para cada paciente. Por este fin, deciden que ambos analicen, de forma separada, cada nuevo paciente que ingrese a su centro; es decir, tanto el psiquiatra como el psicólogo llegan, de forma independiente, a una propuesta sobre la terapia para los pacientes. La siguiente tabla resume los datos obtenidos después de un mes. La primera columna son las iniciales del paciente; las siguientes dos indican la propuesta para la terapia del psiquiatra y el psicólogo, respectivamente.

Paciente	Juicio Psiquiatra	Juicio Psicólogo	Paciente	Juicio Psiquiatra	Juicio Psicólogo
L.H.E.	Farmacológica	Conductual	B.F.I.	Cognitiva	Conductual
S.T.A.	Cognitiva	Cognitiva	T.R.O.	Farmacológica	Farmacológica
B.O.J.	Psicoanalítica	Conductual	B.O.K.	Farmacológica	Farmacológica
F.M.R.	Psicoanalítica	Psicoanalítica	P.G.G.	Farmacológica	Farmacológica
H.F.B.	Psicoanalítica	Conductual	P.P.D.	Farmacológica	Farmacológica
O.R.S.	Psicoanalítica	Psicoanalítica	B.J.L.	Conductual	Conductual
O.G.B.	Psicoanalítica	Conductual	A.R.Y.	Farmacológica	Farmacológica
C.S.A.	Farmacológica	Farmacológica	M.A.M.	Farmacológica	Farmacológica
W.L.L.	Psicoanalítica	Conductual	H.P.G.	Cognitiva	Psicoanalítica
E.R.K.	Cognitiva	Cognitiva	V.A.D.	Psicoanalítica	Psicoanalítica
E.G.R.	Farmacológica	Conductual	C.G.E.	Psicoanalítica	Cognitiva
V.B.S.	Farmacológica	Farmacológica	G.R.G.	Cognitiva	Cognitiva
V.M.R.	Conductual	Conductual	G.M.S.	Farmacológica	Psicoanalítica
R.B.C.	Cognitiva	Cognitiva	G.S.E.	Farmacológica	Farmacológica
I.E.S.	Psicoanalítica	Psicoanalítica	D.P.R.	Conductual	Conductual
S.A.J.	Farmacológica	Farmacológica	B.G.A.	Psicoanalítica	Cognitiva
R.S.L.	Farmacológica	Conductual	R.T.I.	Farmacológica	Farmacológica
T.C.J.	Farmacológica	Conductual	L.P.S.	Farmacológica	Farmacológica
G.M.A.	Farmacológica	Farmacológica	R.M.E.	Farmacológica	Farmacológica
U.J.P.	Farmacológica	Conductual	R.T.A.	Conductual	Conductual
D.R.C.	Farmacológica	Farmacológica	R.C.C.	Farmacológica	Cognitiva
S.J.G.	Conductual	Conductual	F.M.T.	Psicoanalítica	Psicoanalítica
L.S.A.	Farmacológica	Farmacológica	G.P.A.	Farmacológica	Conductual
P.R.M.	Conductual	Cognitiva	S.S.E.	Psicoanalítica	Psicoanalítica

Observamos, por ejemplo, que para el primer paciente (L.H.E.) el psiquiatra y el psicólogo llegaron a una propuesta distinta: el psiquiatra propone una terapia farmacológica; el psicólogo una terapia conductual. Por otro lado, los dos llegaron a la misma conclusión para el segundo paciente (S.T.A.): una terapia cognitiva es la mejor respuesta para sus problemas. Nótese que la tabla incluye información de 48 pacientes.

Debe ser claro que los datos en la tabla presentan una distribución bivariada: las 48 observaciones tienen un valor en la variable X (juicio del psiquiatra) y otro valor en la variable

Y (juicio del psicólogo). Como se explicó en el tema anterior, se pueden resumir las frecuencias y proporciones bivariadas en una tabla de contingencia:

	Juicio del psicólogo (X)				
Juicio del psiquiatra (Y)	Farmaco-lógico	Psico-analítico	Conductual	Cognitiva	**Total**
Farmacológica	16 .333	1 .021	6 .125	1 .021	24 .500
Psicoanalítica	0 .000	6 .125	4 .083	2 .042	12 .025
Conductual	0 .000	0 .000	5 .104	1 .021	6 .125
Cognitiva	0 .000	1 .021	1 .021	4 .083	6 .125
Total	16 .333	8 .167	16 .333	8 .167	48 1.000

A partir de la información en esta tabla, es fácil corroborar que el psicólogo y el psiquiatra llegan al mismo juicio para 31 de los 48 pacientes (los pacientes correspondientes a las frecuencias en la diagonal principal de la tabla). En específico, ambos proponen una terapia farmacológica para 16 pacientes; para 6, los dos creen que una terapia psicoanalítica es la mejor; ambos creen que 5 pacientes serían mejor ayudados en una terapia conductual; y, por último, hay 4 pacientes sobre los cuales el psiquiatra y el psicólogo coinciden que la terapia cognitiva es la más adecuada. Estos 31 pacientes corresponden con una propoción de $\frac{31}{48} = .646$, o bien, en alrededor de 65 % de los casos los dos expertos concuerdan sobre la terapia más apropiada. Se dice que *la proporción de acuerdo* o *la proporción de concordancia* es .646.

Proporción de acuerdo

Para (a) dos variables, X y Y, que pueden asumir los mismos m valores (categorías o juicios de valor), $v_1, \ldots, v_m$,

y (b) un conjunto de n observaciones con valores en ambas variables,

la *proporción de acuerdo* se define como:

$$p_c = \sum_{j=1}^{m} p_{XY}(v_j, v_j).$$

Sinónimo: *proporción de concordancia*

Ejemplo

En el ejemplo anterior, las variables X (juicio de psicólogo) y Y (juicio del psiquiatra) efectivamente asumen los mismos $m = 4$ valores: v_1 = farmacológica, v_2 = psicoanalítica, v_3 = conductual, v_4 = cognitiva y el conjunto de datos contiene $n = 48$ observaciones (pacientes) que fueron juzgados en ambas variables. La proporción de acuerdo es la que se obtuvo en el ejemplo anterior:

$$\begin{aligned} p_c &= \sum_{j=1}^{4} p_{XY}(v_j, v_j) \\ &= p_{XY}(v_1, v_1) + p_{XY}(v_2, v_2) + p_{XY}(v_3, v_3) + p_{XY}(v_4, v_4) \\ &= p_{XY}(\text{“farmacológica”}, \text{“farmacológica”}) + p_{XY}(\text{“psicoanalítica”}, \text{“psicoanalítica”}) \\ &\quad + p_{XY}(\text{“conductual”}, \text{“conductual”}) + p_{XY}(\text{“cognitiva”}, \text{“cognitiva”}) \\ &= .333 + .125 + .104 + .083 = .646 \end{aligned}$$

Nótese que las proporciones que entran en el cálculo de la *proporción de acuerdo* se encuentran en la diagonal principal de la tabla de contingencia (puesto que las filas y las columnas de la tabla ordenan los valores de la misma forma).

6.1.2 El coeficiente kappa

A pesar de su sencillez, tanto respecto del cálculo como la interpretación, la proporción de acuerdo sufre de una desventaja importante. Ilustremos esta desventaja a partir del ejemplo del psicólogo y el psiquiatra que asignan los pacientes a diferentes tipos de terapia.

Ejemplo

Supongamos, como ejemplo muy artificial, que tanto el psicólogo como el psiquiatra hubiesen asignado los pacientes a los diferentes tipos de terapia, aplicando el siguiente procedimiento: cada uno tira un dado (independientemente del otro) y decide sobre la terapia para el paciente a partir del resultado obtenido:

Psicólogo		Psiquiatra	
Resultado	Juicio (X)	Resultado	Juicio (Y)
1,2	Farmacológica	1,2,3	Farmacológica
3	Psicoanalítica	4	Psicoanalítica
4,5	Conductual	5	Conductual
6	Cognitiva	6	Cognitiva

Es decir, para cada paciente, ambos tiran un dado. Si el dado del psicólogo muestra 1 ó 2, entonces asigna el paciente a la terapia farmacológica; si sale 3, le asigna a la terapia psicoanalítica, etc. El psiquiatra, por su cuenta, aplica otro esquema: si sale 1, 2 ó 3 en su dado, asigna el paciente a la terapia farmacológica; si sale 4, le asigna a la terapia psicoanalítica, etc.

El punto que este ejemplo artificial quiere ilustrar es que, aún en el caso tan extremo (y artificial) de que los dos clínicos decidan sobre la terapia de sus pacientes después de tirar un dado, habrá pacientes para los cuales el psicólogo y el psiquiatra proponen el mismo tipo de terapia. (Aunque el concepto de probabilidad no se defina hasta la segunda parte de este curso, el lector podrá verificar, por ejemplo, que se espera que para alrededor de un sexto de los pacientes, ambos propongan una terapia farmacológica.) En otras palabras, habrá concordancia por azar; la proporción de acuerdo p_c sería mayor que 0, aunque los dos juicios fuesen independientes.

En general, al construir un estadístico de covariación, se prefiere—con el objetivo de facilitar la interpretación—que tome un valor de 0 en el caso de independencia (que implica ausencia de covariación). La proporción de acuerdo no cumple esta preferencia.

Para remediar esta inconveniencia de la proporción de acuerdo, Cohen propuso el coeficiente κ (pronunciado “kappa”, la letra griega). Este coeficiente corrige la proporción de acuerdo por los efectos del azar. Para este fin, compara la distribución bivariada observada (las proporciones en la tabla de contingencia) con la distribución que se obtendría en caso de que las dos variables fuesen independientes. Es decir, se construye una tabla de contingencia *hipotética*—bajo el supuesto de que las variables son independientes—y se compara esta tabla de contingencia hipotética con la tabla de contingencia observada.

Antes de poder aplicar este procedimiento, hay que contestar la pregunta: ¿cómo se construye dicha tabla hipotética? Se pueden construir muchas tablas de contingencia, donde las dos variables son independientes. Por lo tanto, se añade una restricción: Se construye una tabla de contingencia bajo el supuesto de independencia *respetando las frecuencias y proporciones marginales*. Ilustrémoslo con un ejemplo.

Ejemplo

En el caso de nuestro ejemplo en el que un psiquiatra y un psicólogo eligen entre cuatro tipos de terapia para 48 pacientes, efectivamente existen muchas maneras en que la distribución bivariada implica independencia entre los dos juicios. Por ejemplo, si la tabla de contingencia presentara las siguientes frecuencias, las dos variables X y Y serían independientes.

	Juicio del psicólogo (X)				
Juicio del psiquiatra (Y)	Farmaco-lógica	Psico-analítica	Conductual	Cognitiva	**Total**
Farmacológica	3	3	3	3	12
Psicoanalítica	3	3	3	3	12
Conductual	3	3	3	3	12
Cognitiva	3	3	3	3	12
Total	12	12	12	12	48

Sin embargo, esta tabla de contingencia no respeta las frecuencias marginales que se encuentran en la tabla de la página 172 (que presenta la función de frecuencia bivariada para los datos observados). ¿Cómo se construye una tabla de contingencia que sí las respeta?

La definición de independencia (véase la Sección 5.5.1 y, en particular, la Ecuación (5.3c)) prescribe que independencia implica que, para todas las celdas se cumple que la proporción bivariada es el producto de las proporciones marginales; es decir, para todos los pares de valores (x, y), se cumple que

$$p_{XY}(x, y) = p_X(x)p_Y(y).$$

Esta condición para independencia permite calcular la proporción en cada celda a partir de las proporciones marginales; en otras palabras, permite la construcción de la tabla de contingencia *bajo el supuesto de independencia* y *respetando las frecuencias y proporciones marginales*. Aplicando esta idea a nuestro ejemplo, se obtiene como tabla de contingencia hipotética:

	Juicio del psicólogo (X)				
Juicio del psiquiatra (Y)	Farmaco-lógica	Psico-analítica	Conductual	Cognitiva	**Total**
Farmacológica	8 .167	4 .083	8 .167	4 .083	24 .500
Psicoanalítica	4 .083	2 .042	4 .083	2 .042	12 .250
Conductual	2 .042	1 .021	2 .042	1 .021	6 .125
Cognitiva	2 .042	1 .021	2 .042	1 .021	6 .125
Total	16 .333	8 .167	16 .333	8 .167	48 1.000

Se observa que, efectivamente, las frecuencias y proporciones marginales en esta tabla hipotética son idénticas a las de la tabla observada en la página 172. Además, las dos variables son independientes (lo cual también se manifiesta a través del principio de proporcionalidad de filas y columnas).

Como siguiente paso, se calcula la proporción de acuerdo en esta tabla de contingencia que se construyó bajo la hipótesis de independencia. Dicha proporción es igual a:

$$p_a \;=\; .167 + .042 + .042 + .021 \;=\; .271$$

Nótese que esta proporción de acuerdo se ha denotado p_a, ya que es la *proporción de acuerdo debido al azar* (o, más preciso, la proporción de acuerdo bajo el supuesto de independencia).

El coeficiente κ compara p_a con el p_c (proporción de acuerdo observada, que para estos datos resultó en .646, véase la página 173), a través de la fórmula en la siguiente definición.

Coeficiente kappa de Cohen

Para (a) dos variables, X y Y, que pueden asumir los mismos m valores (categorías o juicios de valor), $v_1, \ldots, v_m$,

y (b) un conjunto de n observaciones con valores en ambas variables,

el coeficiente kappa se define como:

$$\kappa = \frac{p_c - p_a}{1 - p_a} \tag{6.1a}$$

donde p_c es la proporción de acuerdo observada y p_a es la proporción de acuerdo debido al azar, la cual se define como:

$$p_a = \sum_{j=1}^{m} p_X(v_j)\, p_Y(v_j) \tag{6.1b}$$

Nótese que la suma en la Fórmula (6.1b) es efectivamente la suma de las proporciones bajo el supuesto de independencia en las celdas que se encuentran en la diagonal principal de la tabla de contingencia. El producto $p_X(v_j)\, p_Y(v_j)$ es precisamente la proporción de observaciones para las cuales los dos jueces llegarían al juicio v_j en caso de independencia (y respetando las proporciones marginales).

Nótese también que el cálculo del coeficiente κ presupone que los dos jueces conjuntamente han utilizado dos o más categorías. (El caso contrario quiere decir que se ha utilizado solo una categoría para todos los juicios y, entonces, no habría diferencias entre los juicios. En otras palabras, no habría variación en las variables X y Y, lo cual conlleva que $p_c = p_a = 1$ y el coeficiente kappa no es definido ya que su nominador y denominador serán 0. Varios estadísticos de covariación—aunque no todos—requieren variación en los valores de las variables para que queden definidos.)

Ejemplo

En las páginas anteriores, ya se calcularon la proporción de acuerdo p_c y la proporción de acuerdo debido al azar p_a para el ejemplo del psiquiatra y el psicólogo:

$$p_c = .646 \quad \text{y} \quad p_a = .271.$$

Por lo tanto, el coeficiente kappa que expresa el grado de concordancia entre ambos clínicos es:

$$\begin{aligned} \kappa &= \frac{p_c - p_a}{1 - p_a} \\ &= \frac{.646 - .271}{1 - .271} \\ &= .514. \end{aligned}$$

Interpretación del coeficiente kappa

Revisemos los posibles valores para κ:

- $\kappa = 1$
  De la Ecuación (6.1a) sigue
  $$\kappa = 1 \implies p_c = 1$$
  Por lo tanto, si la concordancia entre juicios llega a $\kappa = 1$, entonces los dos jueces concuerdan en todos los casos evaluados.

- $\kappa = 0$
  También se puede derivar de la Ecuación (6.1a) que
  $$\kappa = 0 \implies p_c = p_a.$$
  Es decir, la concordancia entre los jueces es precisamente igual a la concordancia que se esperaría en caso de que sus juicios fuesen independientes. En otras palabras, no hay más concordancia que la que se espera por coincidencia accidental.

- $0 < \kappa < 1$
  En este caso, el grado de concordancia se encuentra entre concordancia perfecta y concordancia por coincidencia. A medida de que κ se acerque a 1, la concordancia entre los jueces es más pronunciada.

- $\kappa < 0$
  Teóricamente es posible que kappa adopte un valor negativo; quiere decir que los jueces exhiben menos acuerdo que el que se espera por azar. Si en una aplicación práctica se observa una kappa negativa, es recomendable revisar el cálculo del coeficiente y/o la conveniencia del uso de este estadístico para el problema bajo consideración.

6.1.3 El estadístico χ^2_{P} de Pearson

Los dos estadísticos anteriores cuantifican el acuerdo entre jueces que han evaluado n personas u objetos utilizando las mismas m categorías. Como caso más general, se puede investigar también la covariación entre dos variables cualitativas X y Y aunque los valores de X *no* sean los mismos que los valores de Y. Por ejemplo, las variables *Nivel Educativo* y *Fumador* en los datos de la página 34 asumen valores totalmente distintos (además, *Nivel Educativo* asume cuatro valores, mientras que *Fumador* asume tres); sin embargo, sería válido investigar la relación entre ambas variables, por ejemplo, examinar si son independientes y en caso de que no, se podría cuantificar el grado de dependencia.

El estadístico χ^2_{P} (pronunciado como “ji-cuadrado”), el cual se introduce a continuación, permite cuantificar la dependencia entre dos variables, aunque los valores en estas variables no coinciden. En este sentido difiere del coeficiente kappa cuya definición incluye el requisito que X y Y asuman los mismos m valores. Por otro lado, es importante reconocer una similitud muy particular entre estos dos estadísticos: Ambos implican una comparación entre, por un lado, la tabla de contingencia observada y, por otro lado, una tabla de contingencia hipotética que se construye bajo el supuesto de independencia de las dos variables representadas. El siguiente ejemplo explica cómo se realiza dicha comparación en el caso del estadístico χ^2_{P}.

Ejemplo

Consideremos la distribución bivariada definida por las variables *Nivel Educativo* y *Fumador* en los datos de la página 34. Nótese que es válido utilizar un estadístico para variables cualitativas para expresar el grado de asociación entre *Nivel Educativo* y *Fumador*, aunque las dos son de nivel ordinal (y, por lo tanto, son cuasicuantitativas en vez de cualitativas). Debe ser claro que, por ser un estadístico para variables de nivel nominal, el χ^2_{P} no utiliza toda la información relevante en estos datos bivariados (ya que descarta la información ordinal en las variables). Como alternativa, se podría considerar estadísticos que sí toman en cuenta la información ordinal presente en *Nivel Educativo* y *Fumador* (véase, por ejemplo, el estadístico introducido en la Sección 6.2.3).

Los números en color azul en la siguiente tabla de contingencia representa la función de frecuencia bivariada. Estas frecuencias son las que se calcularon y presentaron en el mismo color en la tabla de la página 146.

	Nivel Educativo (X)				
Fumador (Y)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2 2.56 $\frac{(2-2.56)^2}{2.56}$ 0.12	11 11.52 $\frac{(11-11.52)^2}{11.52}$ 0.02	8 10.24 $\frac{(8-10.24)^2}{10.24}$ 0.49	11 7.68 $\frac{(11-7.68)^2}{7.68}$ 1.44	32 32
Bajo	1 0.72 $\frac{(1-0.72)^2}{0.72}$ 0.11	2 3.24 $\frac{(2-3.24)^2}{3.24}$ 0.47	5 2.88 $\frac{(5-2.88)^2}{2.88}$ 1.56	1 2.16 $\frac{(1-2.16)^2}{2.16}$ 0.62	9 9
Intensivo	1 0.72 $\frac{(1-0.72)^2}{0.72}$ 0.11	5 3.24 $\frac{(5-3.24)^2}{3.24}$ 0.96	3 2.88 $\frac{(3-2.88)^2}{2.88}$ 0.01	0 2.16 $\frac{(0-2.16)^2}{2.16}$ 2.16	9 9
Total	4 4	18 18	16 16	12 12	50 50

Esto quiere decir que los números en color azul representan las frecuencias *observadas*. Por otro lado, los números en color rojo representan las frecuencias *esperadas*; estas son las frecuencias hipotéticas, bajo el supuesto de independencia de las variables. Se obtienen dichas frecuencias esperadas de una forma análoga a cómo se calcularon las frecuencias y proporciones en la tabla de contingencia de la página 175. Es decir, siguiendo la definición de independencia en la Ecuación (5.3c), la proporción esperada bajo el supuesto de independencia es igual al producto de las dos proporciones marginales. Puesto que el estadístico χ^2_{P} requiere *frecuencias* (en vez de proporciones), se multiplican las proporciones esperadas por n, el número de observaciones. Así, por ejemplo, se obtiene para la frecuencia esperada en la primera celda, correspondiente a ("Primaria", "No fuma"):

$$e_{11} \;=\; n\, p_X(x_1)\, p_Y(y_1) \;=\; 50\, p_X(\text{“Primaria”})\, p_Y(\text{“No fuma”}) \;=\; 50\, \frac{4}{50}\, \frac{32}{50} = 2.56.$$

Similarmente, para la celda correspondiente al par de valores ("Secundaria", "No fuma"), la frecuencia esperada es:

$$e_{21} \;=\; n\, p_X(x_2)\, p_Y(y_1) \;=\; 50\, p_X(\text{“Secundaria”})\, p_Y(\text{“No fuma”}) \;=\; 50\, \frac{18}{50}\, \frac{32}{50} = 11.52.$$

Así se calcula la frecuencia esperada en cada celda (es decir, para cada combinación de un valor x_j en X y y_k en Y).

Como siguiente paso, el estadístico χ^2_{P} compara la frecuencia observada (o_{jk}) y la frecuencia esperada (e_{jk}) para cada par de valores (x_j, y_k). En específico, calcula para cada celda

$$\frac{(o_{jk} - e_{jk})^2}{e_{jk}}.$$

Debe ser claro que esta fracción es una medida de la discrepancia entre la frecuencia observada y la esperada para la celda correspondiente a la combinación de valores (x_j, y_k). En el caso particular de que $o_{jk} = e_{jk}$, es decir, si la frecuencia observada en una celda coincide perfectamente con lo que se espera en caso de independencia, entonces esta medida de discrepancia resulta 0. Conforme la diferencia sea más grande, el valor en la fracción anterior aumenta. (Nótese que la fracción siempre da valores iguales a o mayores que cero.) Para la primera celda ("Primaria", "No fuma"), la comparación resulta en:

$$\frac{(o_{11} - e_{11})^2}{e_{11}} = \frac{(2 - 2.56)^2}{2.56} = 0.12.$$

Para la celda ("Secundaria", "No fuma"), se tiene:

$$\frac{(o_{21} - e_{21})^2}{e_{21}} = \frac{(11 - 11.52)^2}{11.52} = 0.02.$$

Las fracciones en color verde en la tabla anterior presentan estas comparaciones para cada celda y los números en color naranja corresponden con el resultado del cálculo. El último paso para conocer el valor del χ^2_{P} implica sumar dichos números naranja de las $4 \times 3 = 12$ celdas. Así, se obtiene:

$$\begin{aligned}\chi^2_{\text{P}} &= 0.12 + 0.02 + 0.49 + 1.44 + 0.11 + 0.47 + 1.56 + 0.62 \\ &\quad + 0.11 + 0.96 + 0.01 + 2.16 \\ &= 8.07.\end{aligned}$$

Esta introducción lleva a la definición formal del estadístico χ^2_{P} de Pearson:

El estadístico χ^2_{P} de Pearson

Para (a) una variable X que asume m distintos valores $x_1, x_2, \ldots, x_m$,
(b) una variable Y que asume r distintos valores $y_1, y_2, \ldots, y_r$,
y (c) un conjunto de n observaciones con valores en ambas variables,

el *estadístico* χ^2_P *de asociación* se define por:

$$\chi^2_{\mathrm{P}} = \sum_{j=1}^{m} \sum_{k=1}^{r} \frac{(o_{jk} - e_{jk})^2}{e_{jk}}, \tag{6.2}$$

donde o_{jk} y e_{jk} denotan las *frecuencias observadas* y *esperadas*, respectivamente, definidas como:

$$o_{jk} = \mathit{frec}_{XY}(x_j, y_k) \quad \text{y} \quad e_{jk} = n\, p_X(x_j)\, p_Y(y_k).$$

Del ejemplo anterior que introdujo el estadístico de asociación de Pearson y la definición en el cuadro precedente, es claro que (a) χ^2_{P} siempre es positivo y (b) $\chi^2_{\mathrm{P}} = 0$ si y solo si las dos variables son independientes. Efectivamente, para que $\chi^2_{\mathrm{P}} = 0$, la diferencia entre la frecuencia observada y la frecuencia bajo el supuesto de independencia debe ser 0 *en cada celda*; es decir, la tabla de contingencia cumple con la definición de independencia entre ambas variables.

Entre mayor sea χ^2_{P}, más difieren las frecuencias observadas de las frecuencias que suponen independencia, lo cual se puede interpretar como un mayor grado de asociación entre las variables. No obstante, si χ^2_{P} es diferente de 0, es difícil interpretar su valor, ya que el resultado en el estadístico se ve fuertemente afectado por (a) el número de observaciones (n) y (b) el número de categorías en las variables (m y r). Por ejemplo, aunque las proporciones bivariadas $p_{XY}(x, y)$ en una muestra de 100 personas son idénticas a las proporciones correspondientes en una muestra de 1000 personas, entonces, χ^2_{P} será 10 veces mayor en la muestra más grande. Por lo tanto, el resultado $\chi^2_{\mathrm{P}} = 8.07$ en el ejemplo anterior, únicamente informa que en la muestra las dos variables no son independientes; no ayuda mucho a saber si la asociación es fuerte o débil. Por esta razón, el estadístico no es muy utilizado en la estadística descriptiva. Sin embargo, se introdujo aquí porque (a) este estadístico es de gran importancia en la estadística inferencial y (b) constituye la base para el coeficiente C de contingencia, que se presenta en la siguiente sección.

Nota: El término "ji-cuadrado" en la estadística

En esta sección se introdujo "el estadístico ji-cuadrado de asociación de Pearson". Además del estadístico, también se habla de la "distribución ji-cuadrado" y la "prueba ji-cuadrado". Aunque los tres conceptos están intimamente relacionados, es importante distinguir entre ellos. Hasta el momento basta con tener claro que en esta sección se introdujo el estadístico. Recuérdese que un estadístico es una característica númerica de una muestra. La distribución y la prueba ji-cuadrado son conceptos de la estadística inferencial.

6.1.4 El coeficiente C de contingencia

Una vez que se conoce el valor en el estadístico χ^2_{P}, es fácil calcular el coeficiente C de contingencia.

El coeficiente C de contingencia

Para (a) una variable X que asume m distintos valores $x_1, x_2, \ldots, x_m$,
(b) una variable Y que asume r distintos valores $y_1, y_2, \ldots, y_r$,
y (c) un conjunto de n observaciones con valores en ambas variables,

el *coeficiente C de contingencia* se define como:

$$C = \sqrt{\frac{\chi^2_{\mathrm{P}}}{\chi^2_{\mathrm{P}} + n}},$$

donde χ^2_{P} se define como en la Ecuación (6.2).

Ejemplo

Se obtuvo, en el ejemplo anterior, para el estadístico χ^2_{P} para la asociación entre las variables *Nivel Educativo* y *Fumador*:

$$\chi^2_{\mathrm{P}} = 8.07.$$

Por lo tanto, el coeficiente C de contingencia asume el siguiente valor:

$$\begin{aligned} C &= \sqrt{\frac{\chi^2_{\mathrm{P}}}{\chi^2_{\mathrm{P}} + n}} \\ &= \sqrt{\frac{8.07}{8.07 + 50}} \\ &= .373. \end{aligned}$$

Interpretación del coeficiente C de contingencia

Al analizar la definición del coeficiente C, se verifica que siempre satisface:

$$0 \leqslant C < 1.$$

- Además, se puede derivar que

$$C = 0 \iff \chi^2_{\mathrm{P}} = 0,$$

lo cual implica que:

$$C = 0 \iff \text{independencia}$$

En palabras, el coeficiente C adopta un valor de 0, si y solo si las dos variables son independientes.

- A medida que C se acerca a 1, más fuerte es la asociación entre las variables.

El valor obtenido de $C = 0.373$ en el ejemplo anterior indica una asocación débil entre *Nivel Educativo* y *Fumador*.

Un inconveniente del coeficiente C es que nunca llega a asumir el valor de 1, ni siquiera en el caso de asociación perfecta. La razón se encuentra en que n, el número de observaciones, es mayor que 0 de tal forma que el numerador y el denominador en la fórmula de la definición de C nunca pueden ser iguales. Existen variantes que remedian este inconveniente; sin embargo, una discusión de estas variantes excede el alcance de este curso.

6.2 Covariación lineal entre variables cuantitativas

En esta sección, se introducen dos estadísticos importantes para cuantificar la covariación entre dos variables cuantitativas. Es importante destacar desde un inicio que se tratan de estadísticos de covariación *lineal*. Es decir, su valor da una indicación de la fuerza de relación lineal entre las dos variables de una distribución bivariada, o dicho de otra forma, qué tan bueno se puede describir la relación entre las variables por una recta. Sin duda, la covariación lineal es el tipo de covariación más estudiado en las ciencias médicas y de la salud (y en muchas otras ramas de la ciencia). Sin embargo, la popularidad de las relaciones lineales en la ciencia no se debe a que siempre sean las más adecuadas, sino a su sencillez de interpretación.

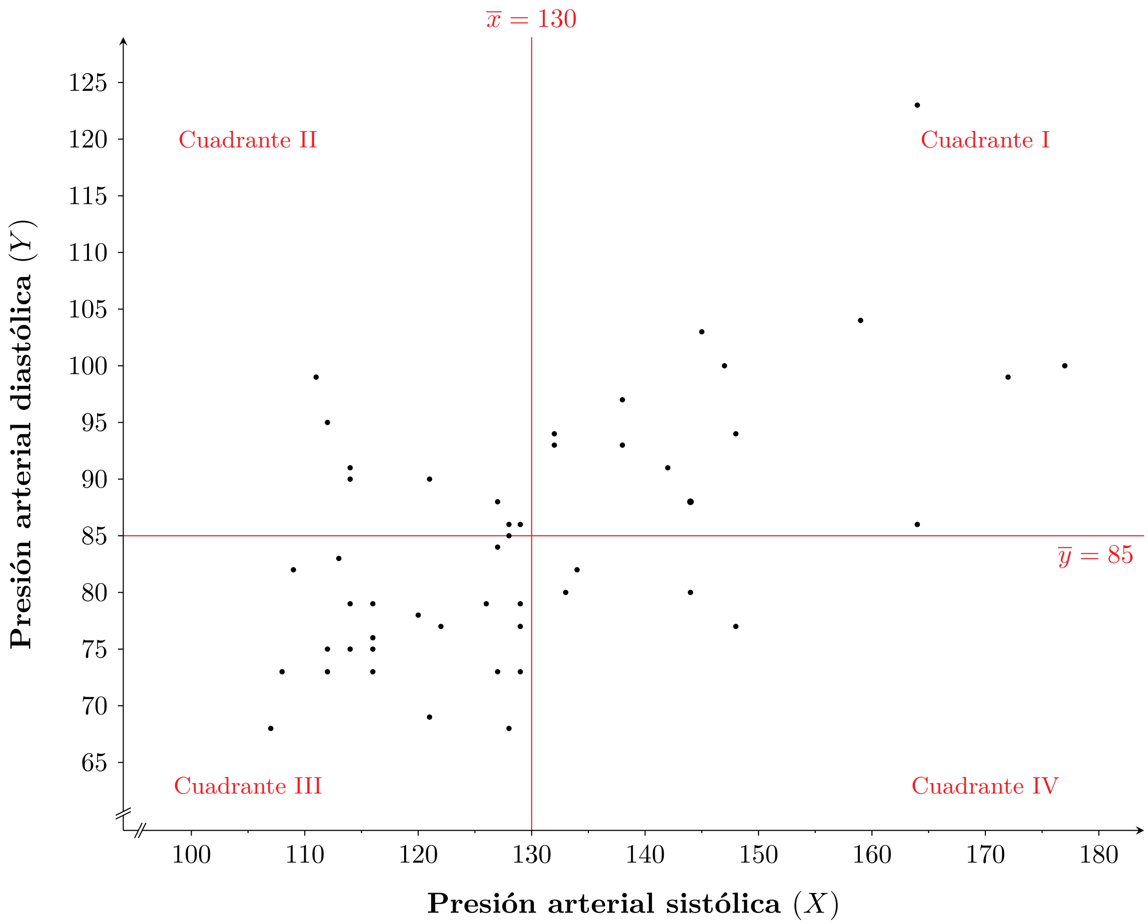
6.2.1 La covarianza

6.2.1.1 Principios y definición

Introduzcamos el estadístico de covarianza a través de un ejemplo.

Ejemplo

Al estudiar la relación entre dos variables, el primer paso comúnmente es representar los datos disponibles en un diagrama de dispersión. En este ejemplo se estudiará la relación entre la *Presión Arterial Sistólica* (variable X) y la *Presión Arterial Diastólica* (variable Y) a través de los datos de la página 34. Un diagrama de dispersión para estas variables se presenta a continuación.



En este diagrama se observa que las observaciones con un valor alto en la variable X tienden a tener un valor alto en la variable Y; similarmente, las que tienen un valor bajo en una de las variables tienden a tener un valor bajo en la otra variable también. Es decir, hay una tendencia para una relación lineal *directa*. Para una comprehensión correcta de esta tendencia, es crucial interpretar correctamente el concepto de "valor alto" y "valor bajo" en una variable. Al estudiar la covariación (lineal) entre dos variables, se entiende valor alto y valor bajo siempre *de forma relativa*, es decir, un valor es alto o bajo porque es alto o bajo *en comparación con las otras observaciones* o, aún más preciso, *en comparación con la media de la variable* en la muestra estudiada. Por ejemplo, en un estudio que involucra exclusivamente a pacientes que sufren de hipertensión arterial grave, una presión arterial sistólica de 148 mmHg puede ser un "valor bajo" (si la media de la variable en este estudio es, digamos, 170 mmHg), mientras que, en el ejemplo de los datos de la página 34, la misma presión arterial sistólica de 148 mmHg es un "valor alto" ya que la media de la variable en esta muestra es 130 mmHg.

"Valor alto" y "valor bajo" deben entenderse entonces en relación con la media de la variable. Por eso, se trazaron en el diagrama de dispersión dos líneas en color rojo que indican la media aritmética de ambas variables. Estas líneas permiten identificar de forma más fácil cuáles son las observaciones con valores altos y bajos en las variables: las observaciones que se encuentran al lado derecho de la línea roja vertical tienen un valor alto en *Presión Arterial Sistólica*; las que están a la izquierda de esta línea tienen un valor bajo. Similarmente, los puntos arriba de la línea roja horizontal representan a personas con un valor alto en *Presión Arterial Diastólica*, mientras que los puntos que están por debajo de esta línea son de personas con valores bajos en esta variable.

Ahora, se entiende de forma más precisa cuando se interpreta una relación lineal directa diciendo "valores altos en una variable van de la mano de valores altos en la otra variable" y "valores bajos en una variable van de la mano de valores bajos en la otra variable". Significa en el ejemplo del diagrama de dispersión arriba que hay más observaciones en los cuadrantes I y III (definidos por las líneas rojas) que en los cuadrantes II y IV. Similarmente, una relación lineal inversa implicaría que la mayoría de las observaciones está en los cuadrantes II y IV; en este caso, hay una tendencia de que "valores altos en una variable van de la mano de valores bajos en la otra".

Para llegar a una cuantificación global de la relación lineal directa o inversa, la covarianza calcula para cada observación las puntuaciones diferenciales en ambas variables y las multiplica. Recuérdese que las puntuaciones diferenciales se definen como la diferencia (positiva o negativa) entre el valor de la observación y la media de la variable, es decir, corresponden en el diagrama de dispersión con la distancia hacia las medias (las líneas en color rojo). Para las observaciones en los cuadrantes I y III, el producto de las puntuaciones diferenciales es positivo (en el cuadrante I, las observaciones tienen puntuaciones diferenciales positivas para ambas variables; en el cuadrante III, las puntuaciones diferenciales son negativas en ambas variables). En los cuadrantes II y IV, este producto es negativo (una puntuación diferencial es positiva, la otra es negativa). A continuación, se suman estos productos, lo cual da una impresión global de la relación entre las dos variables. La siguiente tabla elabora esta idea e ilustra el cálculo.

i	x_i	y_i	$x_i - \overline{x} = \mathcal{D}(x_i)$	$y_i - \overline{y} = \mathcal{D}(y_i)$	$\mathcal{D}(x_i) \times \mathcal{D}(y_i)$
1	109	82	$109 - 130 = -21$	$82 - 85 = -3$	$-21 \times -3 = 63$
2	129	77	$129 - 130 = -1$	$77 - 85 = -8$	$-1 \times -8 = 8$
3	128	86	$128 - 130 = -2$	$86 - 85 = 1$	$-2 \times 1 = -2$
4	112	95	$112 - 130 = -18$	$95 - 85 = 10$	$-18 \times 10 = -180$
5	129	86	$129 - 130 = -1$	$86 - 85 = 1$	$-1 \times 1 = -1$
6	121	90	$121 - 130 = -9$	$90 - 85 = 5$	$-9 \times 5 = -45$
7	116	73	$116 - 130 = -14$	$73 - 85 = -12$	$-14 \times -12 = 168$
8	177	100	$177 - 130 = 47$	$100 - 85 = 15$	$47 \times 15 = 705$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
17	128	85	$128 - 130 = -2$	$85 - 85 = 0$	$-2 \times 0 = 0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
48	108	73	$108 - 130 = -22$	$73 - 85 = -12$	$-22 \times -12 = 264$
49	114	75	$114 - 130 = -16$	$75 - 85 = -10$	$-16 \times -10 = 160$
50	121	69	$121 - 130 = -9$	$69 - 85 = -16$	$-9 \times -16 = 144$
					$\sum_{i=1}^{50} \mathcal{D}(x_i)\mathcal{D}(y_i) = 5610$

Se invita al lector a verificar que las observaciones que contribuyen con un producto positivo a la suma (es decir, aquellas con un resultado positivo en la última columna) efectivamente se sitúan en el cuadrante I o III del diagrama de dispersión; los que contribuyen con un producto negativo están en uno de los cuadrantes II y IV. Además de esto, la tabla muestra que las observaciones que se encuentran muy alejadas del centro de la distribución (es decir, el punto que corresponde con las medias de ambas variables, o bien, donde las dos líneas se cruzan) aportan más a la suma que los puntos que están más cerca. Por ejemplo, las observaciones 4 y 5 pertenecen ambas al cuadrante II (y por lo tanto su contribución a la suma es negativa); sin embargo, la observación 5 se encuentra muy cerca a ambas medias (es el punto que está más cerca del cruce de las líneas rojas en el diagrama de dispersión) y contribuye solo -1 a la suma, mientras que la observación 4 está relativamente más lejos y contribuye -180. Esto quiere decir que el resultado que se obtiene para la suma, no solo depende de cuántas observaciones hay en cada cuadrante, sino también de qué tan alejadas se encuentran estas observaciones del centro. Nótese también la observación 17, cuyo valor en la variable Y coincide con la media, por lo cual su puntuación diferencial en esta variable es 0; por consiguiente, no aporta nada a la suma de productos de puntuaciones diferenciales. En general, observaciones cuyos valores en una o ambas variables coinciden con la media no aportan nada (ya que un valor que coincide con la media no es "alto" ni "bajo").

Finalmente, para la covarianza, se divide la suma de productos de puntuaciones diferenciales por el número de observaciones. En este caso, se obtiene:

$$s_{XY} = \frac{5610}{50} = 112.2.$$

El valor positivo que se obtiene para la covarianza confirma lo que ya hemos constatado por inspección visual del diagrama de dispersión: La relación entre las variables *Presión Arterial Sistólica* y *Presión Arterial Diastólica* se deja describir, hasta cierto punto, como una relación lineal *directa*. Como se aclarará en la siguiente sección, es más difícil sacar conclusiones a partir del valor obtenido respecto de la fuerza de la relación lineal (es decir, ¿qué tan buena es la descripción de la relación entre X y Y como relación lineal directa?).

Llegamos a la siguiente definición:

Covarianza

Para dos variables X y Y, que tienen valores en n observaciones, la *covarianza*, representada por s_{XY}, se define como:

$$s_{XY} = \frac{\sum_{i=1}^{n} \mathcal{D}(x_i)\mathcal{D}(y_i)}{n}, \tag{6.3a}$$

lo cual es equivalente a:

$$s_{XY} = \frac{\sum_{i=1}^{n} [(x_i - \overline{x})(y_i - \overline{y})]}{n}. \tag{6.3b}$$

La definición anterior implica que se puede interpretar la covarianza como la media del producto de las puntuaciones diferenciales en ambas variables.

Cuando se definieron la media y la varianza, se presentaron fórmulas alternativas (véanse las Ecuaciones (3.2) y (3.6)) que permiten calcular estos estadísticos a partir de la función de frecuencia o proporción de la variable bajo consideración. También para la covarianza existen fórmulas alternativas que incorporan la función de frecuencia o proporción bivariada de X y Y.

Definiciones equivalentes de la covarianza

Para (a) una variable X, que asume m distintos valores $x_1, x_2, x_3, \ldots, x_m$,
(b) una variable Y, que asume r distintos valores $y_1, y_2, y_3, \ldots, y_r$
y (c) un conjunto de n observaciones con valores en ambas variables,

la *covarianza* se define, a través de la función de frecuencia, como:

$$s_{XY} = \frac{\sum_{j=1}^{m}\sum_{k=1}^{r} \mathit{frec}_{XY}(x_j, y_k)\, \mathcal{D}(x_j)\mathcal{D}(y_k)}{n} \tag{6.3c}$$

o, de forma equivalente, a través de la función de proporción:

$$s_{XY} = \sum_{j=1}^{m}\sum_{k=1}^{r} p_{XY}(x_j, y_k)\, \mathcal{D}(x_j)\mathcal{D}(y_k) \tag{6.3d}$$

Ejemplo

Las Formúlas (6.3c) y (6.3d) son especialmente útiles cuando las variables X y Y asumen pocos valores (es decir, m y r son valores pequeños) y n, el tamaño de la muestra es grande. Para ilustrar las fórmulas anteriores, consideremos un ejemplo de la psicometría (la cual estudia las propiedades de pruebas y cuestionarios). En la psicometría, es común calcular la covarianza entre las puntuaciones de dos ítems en un test. Por ejemplo, supongamos que se analiza psicométricamente un examen de preguntas de opción múltiple donde la puntuación para cada pregunta es 1 (si el sustentante lo contestó correcto) o 0 (en el caso contrario) y que para las dos primeras preguntas del examen la distribución bivariada de las puntuaciones es la siguiente:

	Puntuación ítem 1 (X)		
Puntuación item 2 (Y)	0	1	**Total**
0	135 .18	240 .32	375 .50
1	15 .02	360 .48	375 .50
Total	150 .20	600 .80	750 1.00

donde, como antes, los números en color azul corresponden con las frecuencias bivariadas y los rojos con las proporciones. La tabla de contingencia muestra que hay 750 sustentantes y que el ítem 1, con 80 % que lo contesta correctamente, parece ser un ítem más fácil que el ítem 2, que tiene solo 50 % de respuestas correctas. Nótese que las medias aritméticas de las variables X y Y correspondientes son .80 y .50, respectivamente. (El lector puede verificar que, como regla general, la media aritmética de una variable que asume únicamente los valores 0 y 1 es igual a la proporción de observaciones con el valor de 1.)

Por lo tanto, aplicando la Ecuación (6.3c), se obtiene para la covarianza:

$$\begin{aligned}
s_{XY} &= \frac{\sum_{j=1}^{m}\sum_{k=1}^{r} \mathit{frec}_{XY}(x_j, y_k)\,\mathcal{D}(x_j)\mathcal{D}(y_k)}{n} \\
&= \frac{\sum_{j=1}^{2}\sum_{k=1}^{2} \mathit{frec}_{XY}(x_j, y_k)\,(x_j - \overline{x})(y_k - \overline{y})}{750} \\
&= \frac{135(0-0.80)(0-0.50) + 15(0-0.80)(1-0.50) + 240(1-0.80)(0-0.50)}{750} \\
&\quad + \frac{360(1-0.80)(1-0.50)}{750} \\
&= \frac{60}{750} = 0.08.
\end{aligned}$$

Interpretando este resultado, se puede concluir que existe una tendencia de relación lineal en esta muestra: los sustentantes que contestan correctamente el primer ítem tienden a contestar también correctamente el segundo ítem. Sin embargo, a partir del resultado obtenido es difícil sacar conclusiones sobre la fuerza de la relación.

Nota: El denominador en la fórmula de la covarianza

Cuando se introdujeron la varianza y desviación estándar se mencionó que muchos libros y software para el análisis estadístico suelen definir estos estadístocos de una forma (ligeramente) distinta (véase la nota en la página 90), en particular, que dividen la suma del numerador entre $n-1$ en vez de n. Para la covarianza se aplica la misma lógica y se divide la suma en las Ecuaciones (6.3) entre $n-1$. Conforme al acuerdo de anotar estas variantes añadiendo una tilde, se define:

$$\tilde{s}_{XY} = \frac{\sum_{i=1}^{n} \left[(x_i - \overline{x})(y_i - \overline{y})\right]}{n-1}.$$

Para referirse al estadístico $\tilde{s}_{XY}$, se habla a veces de la *cuasicovarianza*.

6.2.1.2 Propiedades de la covarianza

Propiedad 1: Conmutatividad de la covarianza

Para variables X y Y cualesquiera, se cumple que:

$$s_{XY} = s_{YX}.$$

Comprobación de la Propiedad 1 de la covarianza

Esta propiedad sigue directamente de la conmutatividad del producto:

$$\begin{aligned} s_{XY} &= \frac{\sum_{i=1}^{n} \mathcal{D}(x_i)\,\mathcal{D}(y_i)}{n} \\ &= \frac{\sum_{i=1}^{n} \mathcal{D}(y_i)\,\mathcal{D}(x_i)}{n} \\ &= s_{YX}. \end{aligned}$$

■

Propiedad 2: Covarianza de una variable consigo misma

Para cualquier variable X, se cumple que:

$$s_{XX} = s_X^2.$$

Comprobación de la Propiedad 2 de la covarianza

Aplicando la Ecuación (6.3b) al caso de $Y = X$, se obtiene:

$$\begin{aligned} s_{XX} &= \frac{\sum_{i=1}^{n} [(x_i - \overline{x})(x_i - \overline{x})]}{n} \\ &= \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n} \\ &= s_X^2. \end{aligned}$$

■

Propiedad 3: Covarianza entre dos variables independientes

Si X y Y son dos variables independientes, entonces:

$$s_{XY} = 0.$$

Hacemos hincapié en que la relación inversa no necesariamente es verdad. Es decir, la Propiedad 3 dice que:

$$\text{independencia} \quad \Longrightarrow \quad \text{covarianza nula.}$$

Sin embargo, la relación inversa:

$$\text{covarianza nula} \quad \Longrightarrow \quad \text{independencia}$$

no es correcta en general. En otras palabras, no es válido concluir que dos variables son independientes simplemente por haber observado que su covarianza es cero.

Se entenderá la relación entre independencia y covarianza nula mejor, al darse cuenta de que una covarianza nula significa independencia *lineal*. Así, es claro que *independencia (general)* implica *independencia lineal*—siempre se puede hacer una deducción de lo general a un caso particular—, pero que la implicación inversa no es siempre verdad.

Comprobación de la Propiedad 3 de la covarianza

La forma más fácil para demostrar la Propiedad 3 es partiendo de la variante de la definición de covarianza en la Fórmula (6.3d):

$$s_{XY} = \sum_{j=1}^{m} \sum_{k=1}^{r} p_{XY}(x_j, y_k)\, \mathcal{D}(x_j)\mathcal{D}(y_k).$$

Si X y Y son independientes, la definición de independencia en la Ecuación (5.3c) dice que $p_{XY}(x_j, y_k) = p_X(x_j)\, p_Y(y_k)$, por lo cual la ecuación anterior se convierte en:

$$\begin{aligned} s_{XY} &= \sum_{j=1}^{m} \sum_{k=1}^{r} p_X(x_j) p_Y(y_k)\, \mathcal{D}(x_j)\mathcal{D}(y_k) \\ &= \sum_{j=1}^{m} \sum_{k=1}^{r} p_X(x_j)\mathcal{D}(x_j)\, p_Y(y_k)\mathcal{D}(y_k). \end{aligned}$$

Puesto que $p_X(x_j)\mathcal{D}(x_j)$ es una constante en la suma que tiene como índice k, se puede mover esta expresión adelante del signo sumatorio, lo cual lleva a:

$$\begin{aligned} &= \sum_{j=1}^{m} \left[p_X(x_j)\mathcal{D}(x_j) \underbrace{\sum_{k=1}^{r} p_Y(y_k)\mathcal{D}(y_k)}_{=\,0} \right], \\ &= 0. \end{aligned}$$

La suma marcada en la penúltima expresión es 0, ya que corresponde con la media de las puntuaciones diferenciales en Y, la cual siempre es 0 según la Propiedad 1 de la media (p. 69). ■

Ejemplos

El Comité Ético de un hospital grande quiere conocer la opinión de su personal de salud respecto de las prácticas de la interrupción del embarazo (aborto) y la terminación de la vida a petición propia (eutanasia). Por este fin, pide a cada miembro del personal que exprese su grado de acuerdo con las siguientes dos afirmaciones:

- La ley debe permitir la interrupción del embarazo hasta 12 semanas de gestación en cualquier circunstancia y a cualquier mujer que lo desee libremente.
- La ley debe permitir la terminación activa de la vida a pacientes que se encuentran actualmente en estado de coma irreversible y que expresaron formalmente su acuerdo con la eutanasia antes de entrar en este estado.

Ambas preguntas se responden en una escala de cuatro opciones de respuesta: "Totalmente en desacuerdo", "En desacuerdo", "De acuerdo", "Totalmente de acuerdo". A partir de las respuestas obtenidas se definen las variables X (actitud hacia el aborto) y Y (actitud hacia la eutanasia), con valores numéricos de 1 a 4, que corresponden con las respectivas opciones de respuesta. La distribución de proporción bivariada de X y Y en el grupo de 1000 profesionales de salud asociados con el hospital se muestra en la siguiente tabla de contingencia:

Actitud hacia la eutanasia (Y)	*Actitud hacia el aborto* (X) 1	2	3	4	**Total**
1	.091	.021	.049	.119	.280
2	.078	.018	.042	.102	.240
3	.039	.009	.021	.051	.120
4	.117	.027	.063	.153	.360
Total	.325	.075	.175	.425	1.000

Se puede evaluar la definición de independencia y verificar que en este conjunto de datos las variables X y Y son efectivamente independientes. Para cada celda, se cumple que la proporción observada coincide con el producto de las proporciones marginales. En concreto,

$$\begin{aligned} .091 &= .325 \times .280 \\ .021 &= .075 \times .280 \\ &\vdots \\ .153 &= .425 \times .360. \end{aligned}$$

Para conocer la covarianza entre X y Y, se necesita primero calcular la media de ambas variables. Aplicando la fórmula en la Ecuación (3.2b), resulta en:

$$\begin{aligned} \overline{x} &= \sum_{j=1}^{4} p(x_j)x_j \\ &= .325 \times 1 + .075 \times 2 + .175 \times 3 + .425 \times 4 = 2.70 \end{aligned}$$

y

$$
\begin{aligned}
\overline{y} &= \sum_{k=1}^{4} p(y_k) y_k \\
&= .280 \times 1 \,+\, .240 \times 2 \,+\, .120 \times 3 \,+\, .360 \times 4 \,=\, 2.56.
\end{aligned}
$$

Con estos resultados se puede aplicar la fórmula en la Ecuación (6.3d) para la covarianza:

$$
\begin{aligned}
s_{XY} &= \sum_{j=1}^{4} \sum_{k=1}^{4} p_{XY}(x_j, y_k)\, \mathcal{D}(x_j) \mathcal{D}(y_k) \\
&= .091\,(1-2.7)\,(1-2.56) \,+\, .078\,(1-2.7)\,(2-2.56) \,+\, .039\,(1-2.7)\,(3-2.56) \\
&\quad +\, .117\,(1-2.7)\,(4-2.56) \,+\, .021\,(2-2.7)\,(1-2.56) \,+\, .018\,(2-2.7)\,(2-2.56) \\
&\quad +\, .009\,(2-2.7)\,(3-2.56) \,+\, .027\,(2-2.7)\,(4-2.56) \,+\, .049\,(3-2.7)\,(1-2.56) \\
&\quad +\, .042\,(3-2.7)\,(2-2.56) \,+\, .021\,(3-2.7)\,(3-2.56) \,+\, .063\,(3-2.7)\,(4-2.56) \\
&\quad +\, .119\,(4-2.7)\,(1-2.56) \,+\, .102\,(4-2.7)\,(2-2.56) \,+\, .051\,(4-2.7)\,(3-2.56) \\
&\quad +\, .153\,(4-2.7)\,(4-2.56) \\
&= 0.00.
\end{aligned}
$$

Se confirma que la covarianza adopta un valor de 0 cuando las dos variables son independientes.

Consideremos ahora los siguientes datos:

i	x_i	y_i	i	x_i	y_i	i	x_i	y_i	i	x_i	y_i	i	x_i	y_i
1	9	8.2	7	16	5.0	13	9	7.3	19	10	7.9	25	13	8.9
2	8	6.5	8	16	6.4	14	11	8.7	20	8	6.9	26	9	7.1
3	17	4.9	9	12	10.0	15	15	7.0	21	13	8.7	27	15	8.2
4	8	7.5	10	6	4.2	16	9	6.3	22	16	6.1	28	6	4.0
5	18	4.0	11	9	6.0	17	14	8.4	23	14	8.1	29	13	9.1
6	16	6.8	12	11	9.3	18	17	5.2	24	15	7.6	30	7	5.7

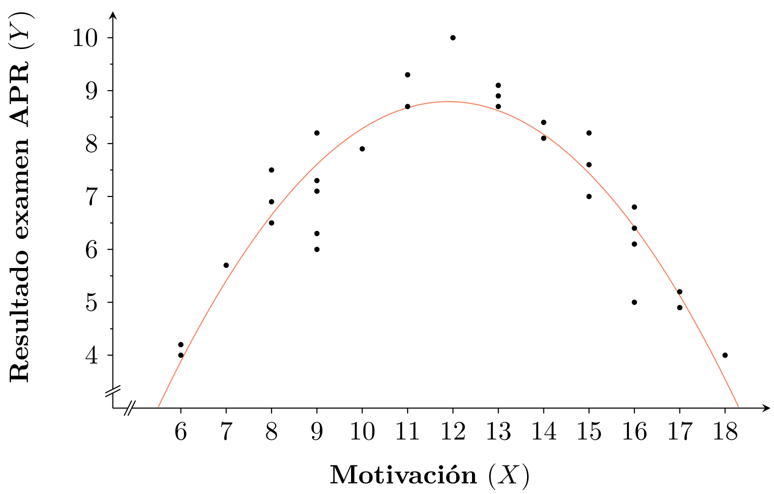
Son datos de un investigador del departamento de Educación Médica que estudia la relación entre, por un lado, el grado de motivación que tiene un estudiante para titularse (variable X) y, por otro lado, el resultado (variable Y) que obtiene en el examen ante paciente real (APR), el cual forma parte del examen profesional en la Facultad de Medicina. La variable X es el resultado en un cuestionario desarrollado por psicólogos de la universidad de Baja California que mide la motivación de estudiantes en medicina, con puntuaciones que varían de 1 (falta completa de motivación) a 20 (motivación máxima). Después de haber averiguado que las medias aritméticas de las variables X y Y son 12 y 7, respectivamente, se obtiene para la covarianza, aplicando la Fórmula (6.3b):

$$\begin{aligned} s_{XY} &= \frac{\sum_{i=1}^{30}[(x_i-\overline{x})(y_i-\overline{y})]}{30} \\ &= \frac{(9-12)(8.2-7)+(8-12)(6.5-7)+(17-12)(4.9-7)+\cdots+(7-12)(5.7-7)}{30} \\ &= 0. \end{aligned}$$

¿Cómo se puede interpretar este resultado?
Es correcto concluir que hay independencia *lineal* entre la motivación del estudiante y el resultado en el examen APR, o que no es adecuado describir la relación entre ambas variables por una recta. Sin embargo, la covarianza nula *no* significa que las dos variables son independientes en general. Las proporciones bivariadas *no* son iguales a los productos de las proporciones marginales; por ejemplo:

$$p_{XY}(6,4) = \frac{1}{30} = .033 \neq p_X(6)p_Y(4) = \frac{2}{30} \times \frac{2}{30} = .0044.$$

Además, un diagrama de dispersión muestra efectivamente que la relación entre las dos variables no es lineal, sino que es posible aproximar la relación por un polinomio cuadrático:



La relación curvilínea entre motivación y rendemiento ha sido confirmada múltiples veces en estudios psicológicos y en el área de educación. Se entiende que un rendimiento óptimo requiere niveles intermedios de motivación: demasiado poca motivación obviamente conlleva un rendimiento bajo, pero también niveles muy altos de motivación afectan negativamente el rendimiento, probablemente porque están asociados con estrés inapropiado. Por lo tanto, un investigador que a partir de un resultado de 0 en un estadístico de covariación lineal infiere que no están relacionados motivación y rendimiento, ha llegado a una conclusión errónea.

Propiedad 4: Covarianza con una variable obtenida por una transformación lineal

Si la variable T es una transformación lineal de la variable X, es decir, si

$$T = aX + b$$

para constantes a y b cualesquiera,

entonces la covarianza de T con cualquier otra variable Y se obtiene por

$$s_{TY} = a\, s_{XY}.$$

Comprobación de la Propiedad 4 de la covarianza

La definición en la Ecuación (6.3b) implica para la covarianza entre las variables T y Y que:

$$s_{TY} = \frac{\sum_{i=1}^{n} \left[(t_i - \overline{t})(y_i - \overline{y})\right]}{n}.$$

$T = aX + b$ quiere decir que $t_i = ax_i + b$ para cualquier observación i; además, por la Ecuación (4.12), se tiene $\overline{t} = a\overline{x} + b$, lo cual resulta en:

$$\begin{aligned} s_{TY} &= \frac{\sum_{i=1}^{n} \left[\left((ax_i + b) - (a\overline{x} + b)\right)(y_i - \overline{y})\right]}{n} \\ &= \frac{\sum_{i=1}^{n} \left[(ax_i + b - a\overline{x} - b)\,(y_i - \overline{y})\right]}{n} \\ &= \frac{\sum_{i=1}^{n} \left[a\,(x_i - \overline{x})\,(y_i - \overline{y})\right]}{n} \\ &= a\,\frac{\sum_{i=1}^{n} \left[(x_i - \overline{x})(y_i - \overline{y})\right]}{n} \\ &= a\, s_{XY}. \end{aligned}$$

■

Examinemos por separado las dos operaciones implicadas en una transformación lineal y su efecto en la covarianza. Por un lado, la Propiedad 4 implica que *sumar* una constante a todos los valores de una variable no cambia la covarianza de esta variable con otras variables. Pensando en la lógica subyacente de la covarianza, es claro por qué sumar una constante no afecta la covarianza. Pues, al sumar una constante a todos los valores observados, la media se transforma sumando la misma constante y, por lo tanto, las puntuaciones diferenciales no cambian y la covarianza sigue siendo la misma. Por otro lado, *multiplicar* todos los valores observados por una constante conlleva que las puntuaciones diferenciales se multiplican por esta constante. Por consiguiente, la covarianza de esta variable con cualquier otra variable se multiplica también por la misma constante.

Ejemplo

El médico virólogo del Caso A del Tema 4 (véase p. 108) realiza un ensayo clínico para evaluar la eficacia de un fármaco para un nuevo tipo de infección, que típicamente se asocia con niveles altos de fiebre. Tiene una muestra de 16 pacientes, infectados con el virus, a los cuales aplica, en una prueba doble ciego, diferentes dosis de la sustancia activa del fármaco. En el diagnóstico posterior mide varios parámetros, entre los cuales la temperatura corporal media hora después de que el paciente ingirió el medicamento. La variable X del estudio es la temperatura corporal del paciente en grados centígrados; la variable Y es la dosis (en mg.) de la sustancia activa del fármaco administrado. Para poder compartir los resultados del estudio con sus colegas estadounidenses, define también la variable T, temperatura corporal en grados Fahrenheit, que es una transformación lineal de la variable observada X con $T = \frac{9}{5}X + 32$.

Los datos se resumen en la siguiente tabla:

i	y_i	x_i	t_i	i	y_i	x_i	t_i
1	0	40.2	104.36	9	2	39.2	102.56
2	0	40.5	104.90	10	2	36.3	97.34
3	0	41.4	106.52	11	2	40.0	104.00
4	0	40.6	105.08	12	2	38.2	100.76
5	1	40.1	104.18	13	3	37.1	98.78
6	1	38.3	100.94	14	3	36.1	96.98
7	1	39.5	103.10	15	3	36.5	97.70
8	1	42.0	107.60	16	3	38.0	100.40

Calcula la media de las variables X y Y:

$$\begin{aligned}
\overline{x} &= \frac{\sum_{i=1}^{16} x_i}{16} \\
&= \frac{40.2 + 40.5 + 41.4 + 40.6 + 40.1 + 38.3 + \cdots + 36.1 + 36.5 + 38.0}{16} \\
&= \frac{624}{16} = 39,
\end{aligned}$$

$$\begin{aligned}
\overline{y} &= \frac{\sum_{i=1}^{16} y_i}{16} \\
&= \frac{0+0+0+0+1+1+1+1+2+2+2+2+3+3+3+3}{16} \\
&= \frac{24}{16} = 1.5,
\end{aligned}$$

y obtiene para la covarianza entre estas variables:

$$\begin{aligned} s_{XY} &= \frac{\sum_{i=1}^{16}[(x_i-\overline{x})(y_i-\overline{y})]}{16}. \\ &= \frac{(40.2-39)(0-1.5)\ +\ (40.5-39)(0-1.5)\ +\cdots+\ (38.0-39)(3-1.5)}{16} \\ &= \frac{-25.6}{16}\ =\ -1.6. \end{aligned}$$

A partir de la covarianza negativa, el investigador concluye que hay una relación inversa entre la dosis del fármaco y la temperatura corporal: Aporta evidencia de que dosis más altas hacen bajar más la fiebre. Quiere compartir el resultado que halló con sus colegas en Estados Unidos y calcula la covarianza entre la variable transformada T, temperatura en Fahrenheit, y la dosis Y. Primero, calcula la media de T y obtiene

$$\begin{aligned} \bar{t} &= \frac{\sum_{i=1}^{16} t_i}{16} \\ &= \frac{104.36+104.90+106.52+105.08+\cdots+96.98+97.70+100.40}{16} \\ &= \frac{1635.20}{16}\ =\ 102.2, \end{aligned}$$

un resultado que, tras una reflexión, esperó, ya que aplicar la propiedad en la Ecuación (4.12) lleva más fácilmente a $\bar{t}=\frac{9}{5}\overline{x}+32=\frac{9}{5}\,39+32=102.2$. A continuación, calcula la covarianza entre T y Y:

$$\begin{aligned} s_{TY} &= \frac{\sum_{i=1}^{16}[(t_i-\bar{t})(y_i-\overline{y})]}{16} \\ &= \frac{(104.36-102.2)(0-1.5)+(104.90-102.2)(0-1.5)+\cdots+(100.4-102.2)(3-1.5)}{16} \\ &= \frac{-46.08}{16}\ =\ -2.88. \end{aligned}$$

Vuelve a encontrar un resultado negativo para la covarianza, lo cual le permite confirmar que existe una relación inversa entre ambas variables. Sin embargo, constata que la covarianza entre la dosis y la temperatura corporal en Fahrenheit no es la misma que la covarianza entre la dosis y la temperatura en centígrados. Efectivamente, a partir de la Propiedad 4 de la covarianza, habría encontrado directamente

$$\begin{aligned} s_{TY} &= \tfrac{9}{5}\,s_{XY} \\ &= \tfrac{9}{5}\,(-1.6)=-2.88. \end{aligned}$$

Como conclusión más importante de esta sección, hay que recordar que el signo de la covarianza informa sobre el tipo de covariación lineal entre dos variables: Una covarianza positiva señala una relación lineal directa, una covarianza negativa apunta a una relación lineal inversa y una covarianza nula significa independencia lineal. No obstante, es difícil inferir la fuerza de la relación a partir de la covarianza. Es verdad que una covarianza mayor (en valor absoluto) generalmente significa una relación más fuerte; seguramente, la covariación entre la dosis del fármaco (Y) y la temperatura corporal en centígrados (X) de -1.6 indica una relación más fuerte que una covarianza de, digamos, -1.2, entre las mismas variables. Sin embargo, es difícil saber si la covarianza de -1.6 indica una relación inversa fuerte o débil.

El último ejemplo además ilustra claramente que la covarianza es sensible a la unidad de las mediciones utilizadas para las variables: Covarianzas con temperaturas en Fahrenheit siempre serán mayores en valor absoluto que temperaturas en centígrados (a menos que, obviamente, la covarianza sea nula). Por lo tanto, sería poco informativo decir "la covarianza entre temperatura corporal y dosis del medicamento es igual a -1.6"; lo anterior muestra que es necesario espeficar la unidad de la escala que se utilizó para la temperatura corporal (y también la unidad en que se expresó la dosis). Sería más preciso decir: "la covarianza entre temperatura corporal en centígrados y dosis del medicamento en miligramos es igual a -1.6". Y aún con esta precisión sigue siendo difícil interpretar la convarianza.

Estas reflexiones han causado que la covarianza se utilice con poca frecuencia en la estadística descriptiva. Como alternativa, es mucho más popular el coeficiente de correlación que se introduce a continuación. Como se aclarará en la siguiente sección, este estadístico no es sensible a la unidad de la escala en que se miden las variables y es más fácil de interpretar.

6.2.2 El coeficiente de correlación de Pearson

6.2.2.1 Principios y definición

El ejemplo de la página 182, donde se introdujo el principio subyacente de la covarianza, ilustró que el núcleo de la fórmula para la covarianza es el producto de las puntuaciones diferenciales en ambas variables. Sin embargo, las puntuaciones diferenciales (y, por lo tanto, también su producto y la covarianza en su totalidad) son sensibles a la unidad de la escala, lo cual—según el argumento al final de la sección anterior—resulta en que la covarianza generalmente es difícil de interpretar.

Para remediar esta inconveniencia de la covarianza, Pearson propuso utilizar las *puntuaciones estandarizadas* en vez de las puntuaciones diferenciales. Como se explicó en la página 112, la puntuación estandarizada (también llamada puntuación $\mathcal{Z}$) se define como la puntuación diferencial dividida entre la desviación estándar de la variable. En símbolos:

$$\mathcal{Z}(x_i) = \frac{\mathcal{D}(x_i)}{s_X} = \frac{x_i - \overline{x}}{s_X}.$$

Gracias a que la puntuación $\mathcal{Z}$ hace abstracción de la unidad de la escala, el coeficiente de correlación, que se obtiene reemplazando las puntuaciones diferenciales por puntuaciones estandarizadas en la Ecuación (6.3a) de la definición de covarianza, resulta tener una interpretación más sencilla y directa.

Ejemplo

Debe ser claro que la puntuación estandarizada efectivamente hace abstracción de la unidad de la escala. Para entender bien esta afirmación, vuélvase a considerar el ejemplo al final de la sección anterior, donde se estudió el efecto de la transformación de la temperatura corporal a grados Fahrenheit. La media y desviación estándar de la variable original X de temperatura en centígrados son

$$\overline{x} = 39 \qquad \text{y} \qquad s_X = \sqrt{3.2} = 1.789.$$

Es importante reconocer que estos dos estadísticos se expresan en unidades "centígrados". Es decir, sería más preciso escribir:

$$\overline{x} = 39°\text{C} \qquad \text{y} \qquad s_X = 1.789°\text{C}.$$

A partir de la media y la desviación estándar, se puede calcular la puntuación diferencial y la puntuación estandarizada, por ejemplo, para la primera observación que tiene un valor $x_1 = 40.2°\text{C}$:

$$\mathcal{D}(40.2°\text{C}) = 40.2°\text{C} - 39°\text{C} = 1.2°\text{C}$$

y

$$\mathcal{Z}(40.2°\text{C}) = \frac{40.2°\text{C} - 39°\text{C}}{1.789°\text{C}} = \frac{1.2°\text{C}}{1.789°\text{C}} = .671.$$

Estas fórmulas explicitan las unidades asociadas. Mientras que la puntuación diferencial para esta observación es 1.2 grados centígrados, la puntuación estandarizada es un resultado *sin dimensiones*. Su significado es que el valor de la primera observación se encuentra .671 *desviaciones estándares encima de la media*.

Realizando las misma operaciones para la variable transformada T, se obtiene

$$\bar{t} = 102.2°\text{F} \qquad \text{y} \qquad s_T = 3.220°\text{F},$$

y para la puntuación diferencial y la puntuación estandarizada de la primera observación, cuya temperatura en Fahrenheit es $t_1 = 104.36°\text{C}$:

$$\mathcal{D}(104.36°\text{F}) = 104.36°\text{F} - 102.2°\text{F} = 2.16°\text{F}$$

y

$$\mathcal{Z}(104.36°\text{F}) = \frac{104.36°\text{F} - 102.2°\text{F}}{3.220°\text{F}} = \frac{2.16°\text{F}}{3.220°\text{F}} = .671.$$

Mientras que $102.2°\text{F} = 39°\text{C}$ (para la media de temperatura corporal), $3.220°\text{F} = 1.789°\text{C}$ (para la desviación estándar), $104.36°\text{F} = 40.2°\text{C}$ (para la temperatura corporal del primer paciente) y $2.16°\text{F} = 1.2°\text{C}$ (para su puntuación diferencial), los valores en las variables X y T son distintos (y ¡son estos valores que entran en las fórmulas!) Por otro lado, la puntuación estandarizada, que en el proceso de su cálculo quita la unidad de la escala, resulta idéntica para las dos variables X y T.

Coeficiente de correlación

Para dos variables X y Y, que tienen valores en n observaciones, el *coeficiente de correlación*, representada por r_{XY}, se define como:

$$r_{XY} = \frac{\sum_{i=1}^{n} \mathcal{Z}(x_i)\, \mathcal{Z}(y_i)}{n}. \tag{6.4a}$$

En algunas ocasiones, se utiliza el término "correlación" también como sinónimo de "covariación". Entonces, "correlación" se puede referir a una característica de una distribución bivariada, o bien, a un estadístico que cuantifica esta propiedad. Para aclarar que se refiere al estadístico, se dice explícitamente "*coeficiente de* correlación". Sin embargo, si desde el contexto es claro qué se está refiriendo al estadístico, se dice simplemente "correlación".

Se puede interpretar el coeficiente de correlación como la media del producto de las puntuaciones estandarizadas en ambas variables. Como ya se mencionó en la parte introductoria de esta sección, la diferencia entre la covarianza y la correlación es que el primero utiliza las puntuaciones diferenciales y el segundo las puntuaciones estandarizadas.

A pesar de esta diferencia, es claro que la misma lógica que subyace la covarianza, con la división del plano en diferentes cuadrantes que identifican las observaciones con una contribución positiva y negativa, también aplica a la correlación. Debido a que las puntuaciones diferenciales y las estandarizadas siempre tienen el mismo signo, las observaciones en los cuadrantes I y III aportan términos positivos a la suma de la correlación, mientras que aquellas en los cuadrantes II y IV contribuyen hacia una correlación negativa.

Ejemplo

Retomemos el ejemplo con el que se introdujo el concepto de la covarianza en la página 182 y que estudió la relación entre *Presión Arterial Sistólica* y *Presión Arterial Diastólica*. En vez de calcular para cada observación la puntuación diferencial en ambas variables, construimos una tabla con las puntuaciones estandarizadas. Tomando en cuenta la media y desviación estándar de las dos variables, los valores en *Presión Arterial Sistólica* se transforman por:

$$\mathcal{Z}(x_i) = \frac{x_i - \overline{x}}{s_X} = \frac{x_i - 130}{17}$$

mientras que los valores en la variable *Presión Arterial Diastólica* se estandarizan aplicando:

$$\mathcal{Z}(y_i) = \frac{y_i - \overline{y}}{s_Y} = \frac{y_i - 85}{11}.$$

De esta forma, se construye una tabla análoga a la tabla que se encuentra en la página 184:

i	x_i	y_i	$\dfrac{x_i - \overline{x}}{s_X} = \mathcal{Z}(x_i)$	$\dfrac{y_i - \overline{y}}{s_Y} = \mathcal{Z}(y_i)$	$\mathcal{Z}(x_i) \times \mathcal{Z}(y_i)$
1	109	82	$\dfrac{109-130}{17} = -1.24$	$\dfrac{82-85}{11} = -0.27$	$-1.24 \times -0.27 = 0.34$
2	129	77	$\dfrac{129-130}{17} = -0.06$	$\dfrac{77-85}{11} = -0.73$	$-0.06 \times -0.73 = 0.04$
3	128	86	$\dfrac{128-130}{17} = -0.12$	$\dfrac{86-85}{11} = 0.09$	$-0.12 \times 0.09 = -0.01$
4	112	95	$\dfrac{112-130}{17} = -1.06$	$\dfrac{95-85}{11} = 0.91$	$-1.06 \times 0.91 = -0.96$
5	129	86	$\dfrac{129-130}{17} = -0.06$	$\dfrac{86-85}{11} = 0.09$	$-0.06 \times 0.09 = -0.01$
6	121	90	$\dfrac{121-130}{17} = -0.53$	$\dfrac{90-85}{11} = 0.45$	$-0.53 \times 0.45 = -0.24$
7	116	73	$\dfrac{116-130}{17} = -0.82$	$\dfrac{73-85}{11} = -1.09$	$-0.82 \times -1.09 = 0.90$
8	177	100	$\dfrac{177-130}{17} = 2.76$	$\dfrac{100-85}{11} = 1.36$	$2.76 \times 1.36 = 3.77$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
48	108	73	$\dfrac{108-130}{17} = -1.29$	$\dfrac{73-85}{11} = -1.09$	$-1.29 \times -1.09 = 1.41$
49	114	75	$\dfrac{114-130}{17} = -0.94$	$\dfrac{75-85}{11} = -0.91$	$-0.94 \times -0.91 = 0.86$
50	121	69	$\dfrac{121-130}{17} = -0.53$	$\dfrac{69-85}{11} = -1.45$	$-0.53 \times -1.45 = 0.77$
					$\displaystyle\sum_{i=1}^{50} \mathcal{Z}(x_i)\mathcal{Z}(y_i) = 30.00$

Los productos de puntuaciones estandarizadas suman a 30. Por lo tanto, el coeficiente de correlación entre *Presión Arterial Sistólica* y *Presión Arterial Diastólica* se da por:

$$r_{XY} = \frac{\sum_{i=1}^{50} \mathcal{Z}(x_i)\,\mathcal{Z}(y_i)}{50} = \frac{30}{50} = 0.60.$$

La Propiedad 5 de la correlación (p. 205) proporcionará las herramientas para interpretar correctamente este valor obtenido para la correlación.

La definición de la correlación que se presentó en la Ecuación (6.4a) no es la que se encuentra comúnmente en los libros de estadística o las páginas web sobre el tema. Es más común definir la correlación como se hace en el siguiente cuadro:

Definición equivalente de la correlación: Relación con la covarianza

Para dos variables X y Y, el *coeficiente de correlación* se puede definir como:

$$r_{XY} = \frac{s_{XY}}{s_X\, s_Y}. \tag{6.4b}$$

Esta definición relaciona explicítamente la correlación con la covarianza. Es fácil demostrar que la definición en la Ecuación (6.4b) es equivalente con la que se presentó inicialmente en la Ecuación (6.4a).

Comprobación de la equivalencia de las definiciones (6.4a) y (6.4b)

Se parte de la definición en (6.4a):

$$\begin{aligned} r_{XY} &= \frac{\sum_{i=1}^{n} \mathcal{Z}(x_i)\mathcal{Z}(y_i)}{n} \\ &= \frac{\sum_{i=1}^{n} \frac{\mathcal{D}(x_i)}{s_X}\frac{\mathcal{D}(y_i)}{s_Y}}{n}. \end{aligned}$$

Reconociendo que s_X y s_Y son constantes en la suma, se pueden poner delante del signo sumatorio:

$$\begin{aligned} &= \frac{1}{s_X\, s_Y}\frac{\sum_{i=1}^{n} \mathcal{D}(x_i)\mathcal{D}(y_i)}{n} \\ &= \frac{1}{s_X s_Y} s_{XY} \\ &= \frac{s_{XY}}{s_X s_Y} \qquad \blacksquare \end{aligned}$$

Ejemplo

A partir de la covarianza entre *Presión Arterial Sistólica* y *Presión Arterial Diastólica* que se halló en la página 185 y conociendo las desviaciones estándares de ambas variables, se puede calcular la correlación aplicando la Ecuación (6.4b):

$$r_{XY} = \frac{s_{XY}}{s_X\, s_Y} = \frac{112.2}{17 \times 11} = 0.60,$$

el mismo resultado obtenido por el cálculo elaborado a partir de la Ecuación (6.4a) (p. 199).

El médico virólogo del ejemplo en la página 194 obtuvo para la covarianza un resultado de -1.60 entre la temperatura corporal X y la dosis administrada del fármaco Y. Después de calcular las desviaciones estándares de las dos variables y obteniendo

$$s_X \;=\; \sqrt{3.2} \;=\; 1.789 \qquad \text{y} \qquad s_Y \;=\; \sqrt{1.25} \;=\; 1.118$$

halló la correlación entre estas variables aplicando la Ecuación (6.4b):

$$r_{XY} \;=\; \frac{s_{XY}}{s_X\, s_Y} \;=\; \frac{-1.60}{1.789 \,\times\, 1.118} \;=\; -.80.$$

Cabe señalar que la correlación no es definida en el caso de que una de las variables (o ambas) tenga variación nula. Pues, en este caso, la desviación estándar es igual a 0 y no se puede llevar a cabo la estandarización implicada en la definición (6.4a). El problema se manifiesta también al calcular la correlación por la Ecuación (6.4b) ya que en caso de que $s_X = 0$ o $s_Y = 0$, el denominador (tanto como la covarianza en el numerador) queda en cero.

6.2.2.2 Propiedades del coeficiente de correlación

Propiedad 1: Conmutatividad del coeficiente de correlación

Para variables X y Y cualesquiera, se cumple que:

$$r_{XY} = r_{YX}.$$

Comprobación de la Propiedad 1 del coeficiente de correlación

Esta propiedad sigue directamente de la definición de la correlación en la Ecuación (6.4b) y la conmutatividad de la covarianza (en el nominador) y la conmutatividad del producto (en el denumerador de esta ecuación). ■

Propiedad 2: Correlación de una variable consigo misma

Para cualquier variable X para la cual $s_X > 0$, se cumple que:

$$r_{XX} = 1.$$

Comprobación de la Propiedad 2 del coeficiente de correlación

Aplicando la definición en la Ecuación (6.4b) al caso de $Y = X$, se obtiene:

$$r_{XX} = \frac{s_{XX}}{s_X\, s_X}.$$

Según la segunda propiedad de la covarianza, el numerador es igual a la varianza de X. De ahí sigue:

$$r_{XX} = \frac{s_X^2}{s_X^2}$$

lo cual, en el caso de que $s_X^2 > 0$ (o $s_X > 0$), implica que

$$r_{XX} = 1.$$

(Nótese que la fracción queda indefinida si $s_X = 0$.) ■

Propiedad 3: Correlación entre dos variables independientes

Si X y Y son dos variables independientes y $s_X > 0$ y $s_Y > 0$, entonces:

$$r_{XY} = 0.$$

Igual a que fue importante entender bien la dirección de la implicación en la Propiedad 3 de la covarianza, también hay que entender bien esta propiedad de la correlación: puesto que la correlación es un estadístico de covariación *lineal*, $r_{XY} = 0$ implica independencia *lineal* entre las dos variables, pero una correlación nula no quiere decir que las dos variables son independientes en general.

Comprobación de la Propiedad 3 del coeficiente de correlación

Según la Propiedad 3 de la covarianza, independencia entre X y Y implica una covarianza nula; por otro lado, de la Ecuación (6.4b) se deriva que una covarianza nula implica una correlación nula siempre y cuando $s_X > 0$ y $s_Y > 0$. ■

Propiedad 4: Correlación con una variable obtenida por una transformación lineal

Si la variable T es una transformación lineal de la variable X, es decir, si

$$T = aX + b,$$

donde a es una constante diferente de 0 y b una constante cualesquiera,
entonces la correlación de T con cualquier otra variable Y se da por

$$r_{TY} = \begin{cases} r_{XY} & \text{si } a > 0, \\ -r_{XY} & \text{si } a < 0. \end{cases}$$

Comprobación de la Propiedad 4 del coeficiente de correlación

Según la propiedad en la página 133 (y la Ecuación (4.14b) en particular), si $T = aX + b$, entonces la desviación estándar de T mantiene la siguiente relación con la desviación estándar de X:

$$s_T \;=\; |a|\, s_X.$$

Además, la Propiedad 4 de la covarianza implica que la covarianza de T con cualquier otra variable Y obedece

$$s_{TY} \;=\; a\, s_{XY}.$$

Por otro lado, la Ecuación (6.4b) define la correlación entre T y Y por:

$$r_{TY} \;=\; \frac{s_{TY}}{s_T\, s_Y},$$

lo cual, si se reemplazan los dos resultados anteriores para s_T y s_{TY}, lleva a:

$$\begin{aligned} r_{TY} &= \frac{a\, s_{XY}}{|a| s_X\, s_Y} \\ &= \frac{a}{|a|}\,\frac{s_{XY}}{s_X\, s_Y} \\ &= \frac{a}{|a|}\, r_{XY} \\ &= \begin{cases} r_{XY} & \text{si } a > 0, \\ -r_{XY} & \text{si } a < 0. \end{cases} \end{aligned}$$

Nótese que si $a = 0$, entonces tanto $s_{TY} = 0$ como $s_T = 0$, así que r_{TY} es indefinida en este caso. ■

Esta propiedad importante implica que transformar linealmente a una de las variables (o ambas) involucrada en la correlación, nunca cambiará la magnitud de la correlación. Si la transformación lineal implica que se invierte el orden (porque la constante multiplicativa es negativa), entonces la correlación cambia de signo. Ésto último quiere decir que si una variable se multiplica por una constante negativa, entonces cambia el tipo de la relación lineal con otra variable de directa a inversa o, viceversa, de inversa a directa.

Ejemplos

En el ejemplo de la página 201, el virólogo que investigó el efecto de la dosis de un medicamento en la temperatura corporal de sus pacientes encontró una correlación de $-.80$ entre Y, la dosis, y X, la temperatura en grados centígrados.

Calculemos ahora la correlación entre la misma variable Y y la variable T, temperatura en *Fahrenheit*. A partir de los datos en la tabla de la página 194, se encontró que:

$$\begin{aligned} s_Y &= 1.118 && \text{(véase la página 201)},\\ s_T &= 3.220 && \text{(véase la página 197)},\\ \text{y} \quad s_{TY} &= -2.88 && \text{(véase la página 195)}, \end{aligned}$$

así que la correlación resulta en:

$$\begin{aligned} r_{TY} &= \frac{s_{TY}}{s_T\, s_Y}\\ &= \frac{-2.88}{3.220 \times 1.118}\\ &= -.80, \end{aligned}$$

la misma correlación que la correlación entre la dosis Y y la temperatura en *centígrados* X. Es decir, se verifica la Propiedad 4 del coeficiente de correlación, ya que T es una transformación lineal de X con una constante multiplicativa positiva ($T = \frac{9}{5}X + 32$), por lo cual $r_{TY} = r_{XY}$.

En el examen profesional que organizó la Facultad de Medicina en enero 2013, 702 estudiantes participaron tanto en el examen teórico como en el examen clínico objetivo estructurado (el ECOE, es decir la parte práctica). Denominemos X la calificación en el teórico, la cual es nada más el número de preguntas que el estudiante contestó correctamente del total de 330 preguntas en el examen; y Y la calificación en el ECOE, la cual es el promedio de las calificaciones otorgadas por diferentes profesores en las 18 estaciones del examen. Se encontró que la correlación entre X y Y era .32.

Ahora, definamos la variable T como el número de preguntas con respuesta *in*correcta en el examen teórico. Es decir, si alguien contesta correctamente todas las preguntas, entonces su valor en T es 0; si contestase todas las preguntas incorrectas, entonces el valor en T sería 330. Es decir, la relación entre T y X es la siguiente:

$$T = 330 - X.$$

Puesto que es una transformación lineal con una constante multiplicativa negativa ($a = -1$), la correlación entre T y Y se da por:

$$\begin{aligned} r_{TY} &= -r_{XY}\\ &= -.32. \end{aligned}$$

Nota: La independencia de la correlación de la unidad de la escala

La propiedad anterior permite entender de forma alternativa y más clara de por qué el coeficiente de correlación es independiente de la unidad de la escala. Puesto que las escalas de temperatura en centígrados y Fahrenheit están relacionadas por una transformación lineal (con una constante multiplicativa positiva), la correlación entre temperatura y otra variable, no depende de qué escala se utilice para temperatura. Asimismo, las escalas para expresar longitud—en centímetros, metros, kilómetros, pies, pulgadas, varas o milas— están relacionadas por una transformación lineal y mantienen correlaciones idénticas con otras variables; lo mismo es verdad de las escalas de peso—gramas, kilogramos, onzas, libras, etc.

Sin embargo, si las escalas *no* están relacionadas por una transformación lineal, la correlación con otras variables generalmente cambia. Por ejemplo, la intensidad *física* de un estímulo auditivo, que se suele expresar en dB, y la intensidad *percibida* del estímulo, que se sitúa en una escala de $\log \mathrm{dB}$ se relacionan por una función *logarítmica* (véase, por una explicación más detalla, el Caso C en el tema sobre transformaciones) y por lo tanto las correlaciones con una u otra variable pueden ser diferentes. (Nótese que es posible, considerando una muestra de estímulos auditivos, que el estímulo i se encuentre encima de la media en la escala de intensidad física y debajo de la media en la escala de intensidad percibida, lo cual afecta su contribución a la suma en (6.4a) que define su correlación con otra variable.)

Propiedad 5: Valor máximo y mínimo de una correlación

Para variables X y Y cualesquiera (con $s_X > 0$ y $s_Y > 0$), se cumple que:

$$-1 \leqslant r_{XY} \leqslant +1.$$

Comprobación de la Propiedad 5 del coeficiente de correlación

Primero demostramos que $r_{XY} \leqslant 1$. Para esto partimos de la diferencia entre las puntuaciones estandarizadas en X y Y de cada observación i (que son definidas si $s_X > 0$ y $s_Y > 0$). Evidentemente, esta diferencia llevada al cuadrado, siempre es positiva:

$$0 \leqslant \left[\mathcal{Z}(x_i) - \mathcal{Z}(y_i)\right]^2 .$$

También la media de esta expresión entre todas las observaciones será positiva. Es decir:

$$\Longrightarrow \qquad 0 \leqslant \frac{\sum_{i=1}^{n} \left[\mathcal{Z}(x_i) - \mathcal{Z}(y_i)\right]^2}{n}.$$

Desarrollando la diferencia cuadrática, nos lleva a:

$$\Longleftrightarrow \qquad 0 \leqslant \frac{\sum_{i=1}^{n} \left[\left(\mathcal{Z}(x_i)\right)^2 + \left(\mathcal{Z}(y_i)\right)^2 - 2\mathcal{Z}(x_i)\mathcal{Z}(y_i)\right]}{n}.$$

$$\iff \quad 0 \leqslant \frac{\sum_{i=1}^{n} \left(\mathcal{Z}(x_i)\right)^2}{n} + \frac{\sum_{i=1}^{n} \left(\mathcal{Z}(y_i)\right)^2}{n} - 2\,\frac{\sum_{i=1}^{n} \mathcal{Z}(x_i)\mathcal{Z}(y_i)}{n}.$$

Dado que la media de $\mathcal{Z}(X)$ es cero, el primer termino de la suma del lado derecho de la desigualdad representa la varianza de $\mathcal{Z}(X)$. De forma similar, el segundo término es la varianza de $\mathcal{Z}(Y)$. El quebrado en el tercer término coincide precisamente con la Ecuación (6.4a) de la correlación entre X y Y. Por lo tanto,

$$\iff \quad 0 \leqslant s^2_{\mathcal{Z}(X)} + s^2_{\mathcal{Z}(Y)} - 2r_{XY}.$$

De ahí sigue (ya que la varianza de una variable estandarizada es 1):

$$\begin{aligned} \iff &\quad 0 \leqslant 1 + 1 - 2r_{XY} \\ \iff &\quad 2r_{XY} \leqslant 2 \\ \iff &\quad r_{XY} \leqslant 1. \end{aligned}$$

La demostración de que $r_{XY} \geqslant -1$ sigue, de modo análogo, a partir de $0 \leqslant [\mathcal{Z}(x_i) + \mathcal{Z}(y_i)]^2$. ■

La Propiedad 5 ofrece un punto de referencia para la interpretación de un coeficiente de correlación en aplicaciones concretas. Conforme una correlación se acerca a $+1$ o -1, mejor se describe la relación entre las dos variables por una función lineal; cuando llega a uno de estos límites, entonces la relación es perfectamente lineal (es decir, todos los puntos del diagrama de dispersión se encuentran en una recta).

Como regla general—y aunque depende mucho del contexto—, correlaciones cuyo valor absoluto llega a .70 o mayor se interpretan como indicadores de relaciones lineales fuertes; correlaciones de alrededor de .40 o .50 se consideran moderadas y de .20 a .30 débiles. La correlación de .60 entre *Presión Arterial Sistólica* y *Presión Arterial Diastólica* en los datos de la página 34 es moderada-alta; la correlación de $-.80$ que encontró el virólogo del ejemplo en la página 201 entre la dosis de un fármaco y la temperatura corporal apunta a una relación fuerte entre ambas variables.

En la Sección 6.2.2.3 se presentarán algunos diagramas de dispersión típicos que visualizan la fuerza de correlación entre dos variables.

Cabe mencionar que de la propiedad de la correlación se desprende una propiedad homóloga de la covarianza:

Corolario: Valor máximo y mínimo de la covarianza

Para variables X y Y cualesquiera, siempre se cumple:

$$-s_X\, s_Y \;\leqslant\; s_{XY} \;\leqslant\; +s_X\, s_Y.$$

Propiedad 6: Correlaciones de -1 y $+1$

Para variables X y Y cualesquiera, las siguientes equivalencias se cumplen siempre:

$$\begin{aligned}
&\text{(a)}\ r_{XY} = +1 \iff \forall i : \mathcal{Z}(x_i) = \mathcal{Z}(y_i)\\
&\text{(b)}\ r_{XY} = -1 \iff \forall i : \mathcal{Z}(x_i) = -\mathcal{Z}(y_i)
\end{aligned}$$

En palabras, la equivalencia en (a) dice que si la correlación entre dos variables es máxima, entonces las puntuaciones estandarizadas en ambas variables son iguales para todas las observaciones; y viceversa, si las puntuaciones estandarizadas en ambas variables son iguales para todas las observaciones, entonces la correlación es máxima. De modo similar, (b) quiere decir que si el coeficiente de correlación entre dos variables es igual a -1, entonces las puntuaciones estandarizadas en ambas variables son opuestas para todas las observaciones; y viceversa, si las puntuaciones estandarizadas en ambas variables son opuestas para todas las observaciones, entonces la correlación es igual a -1.

Comprobación de la Propiedad 6 del coeficiente de correlación

- Primero se demuestra la implicación: $r_{XY} = +1 \Rightarrow \forall i : \mathcal{Z}(x_i) = \mathcal{Z}(y_i)$.
  A partir de la media de las diferencias entre las puntuaciones tipificadas

$$\frac{\sum_{i=1}^{n} [\mathcal{Z}(x_i) - \mathcal{Z}(y_i)]^2}{n}$$

  se obtiene que (véase la la comprobación de la Propiedad 5):

$$\frac{\sum_{i=1}^{n} [\mathcal{Z}(x_i) - \mathcal{Z}(y_i)]^2}{n} = 1 + 1 - 2r_{XY},$$

  lo cual, si $r_{XY} = 1$, implica que

$$\frac{\sum_{i=1}^{n} [\mathcal{Z}(x_i) - \mathcal{Z}(y_i)]^2}{n} = 1 + 1 - 2 = 0.$$

  La expresión al lado izquierdo es igual a 0 si y solo si el numerador es igual a 0. Por otro lado, el numerador es una suma de la cual cada término es igual a o mayor que 0 (ya que cada término es el cuadrado de un número real). Por lo tanto, el numerador es 0 si y solo si cada término de la suma es igual a 0. Es decir,

$$\begin{aligned}
\frac{\sum_{i=1}^{n} [\mathcal{Z}(x_i) - \mathcal{Z}(y_i)]^2}{n} = 0 &\iff \forall i : \quad [\mathcal{Z}(x_i) - \mathcal{Z}(y_i)]^2 = 0\\
&\iff \forall i : \quad \mathcal{Z}(x_i) = \mathcal{Z}(y_i).
\end{aligned}$$

- A continuación, se demuestra la implicación inversa: $\forall i : \mathcal{Z}(x_i) = \mathcal{Z}(y_i) \Rightarrow r_{XY} = +1$. A partir de la definición del coeficiente de correlación:

$$r_{XY} = \frac{\sum_{i=1}^{n} \mathcal{Z}(x_i)\mathcal{Z}(y_i)}{n}$$

se deriva que, cuando $\mathcal{Z}(x_i) = \mathcal{Z}(y_i)$:

$$\begin{aligned} r_{XY} &= \frac{\sum_{i=1}^{n} \mathcal{Z}(y_i)\mathcal{Z}(y_i)}{n} \\ &= \frac{\sum_{i=1}^{n} \left(\mathcal{Z}(y_i)\right)^2}{n}, \end{aligned}$$

lo cual, como vimos anteriormente, corresponde con la varianza de la variable estandarizada $\mathcal{Z}(Y)$. Esta varianza es igual a 1. Entonces,

$$r_{XY} = 1.$$

La equivalencia (b) se demuestra de forma análoga:

- Para la implicación $r_{XY} = -1 \Rightarrow \forall i : \mathcal{Z}(x_i) = -\mathcal{Z}(y_i)$, se desarrolla

$$\frac{\sum_{i=1}^{n} \left[\mathcal{Z}(x_i) + \mathcal{Z}(y_i)\right]^2}{n}$$

y se obtiene, cuando $r_{XY} = -1$,

$$\frac{\sum_{i=1}^{n} \left[\mathcal{Z}(x_i) + \mathcal{Z}(y_i)\right]^2}{n} = 0,$$

lo cual, a su vez, conlleva que

$$\forall i : \quad \mathcal{Z}(x_i) = -\mathcal{Z}(y_i).$$

- Por último, la implicación inversa, $\forall i : \mathcal{Z}(x_i) = -\mathcal{Z}(y_i) \Rightarrow r_{XY} = -1$, sigue directamente de

$$\begin{aligned} r_{XY} &= \frac{\sum_{i=1}^{n} \mathcal{Z}(x_i)\mathcal{Z}(y_i)}{n} \\ &= \frac{\sum_{i=1}^{n} -\mathcal{Z}(y_i)\mathcal{Z}(y_i)}{n} \\ &= \frac{\sum_{i=1}^{n} -\left(\mathcal{Z}(x_i)\right)^2}{n} \\ &= -1. \end{aligned}$$

■

Es importante darse cuenta de que $\mathcal{Z}(y_i) = \mathcal{Z}(x_i)$ conlleva:

$$\frac{y_i - \overline{y}}{s_Y} = \frac{x_i - \overline{x}}{s_X}$$

y, de ahí:

$$\begin{aligned} y_i &= \frac{x_i - \overline{x}}{s_X} s_Y + \overline{y} \\ \Longleftrightarrow \quad y_i &= \underbrace{\left(\frac{s_Y}{s_X}\right)}_{a} x_i + \underbrace{\left(\overline{y} - \frac{s_Y}{s_X}\,\overline{x}\right)}_{b}. \end{aligned}$$

Este resultado quiere decir que si la correlación entre dos variables es $+1$, entonces la relación entre ambas es lineal directa y se describe perfectamente por una recta con la pendiente a y la intersección b como indicado en la última ecuación. Del mismo modo, si la correlación entre X y Y es -1, entonces según la Propiedad 6 tenemos para todas las observaciones $\mathcal{Z}(y_i) = -\mathcal{Z}(x_i)$, por lo cual

$$\begin{aligned} \frac{y_i - \overline{y}}{s_Y} &= -\frac{x_i - \overline{x}}{s_X} \\ \Longleftrightarrow \quad y_i &= -\frac{x_i - \overline{x}}{s_X} s_Y + \overline{y} \\ \Longleftrightarrow \quad y_i &= \underbrace{\left(-\frac{s_Y}{s_X}\right)}_{a} x_i + \underbrace{\left(\overline{y} + \frac{s_Y}{s_X}\,\overline{x}\right)}_{b}. \end{aligned}$$

Es decir, la relación entre X y Y es lineal inversa y se describe perfectamente por la función lineal en la última ecuación.

Ejemplo

La siguiente tabla muestra los valores estandarizados en las variables X y T (temperatura en centígrados y en Fahrenheit, respectivamente) para las observaciones en la tabla de la página 194 (véase también el ejemplo en la página 197):

i	x_i	t_i	$\frac{x_i - \overline{x}}{s_X} = \mathcal{Z}(x_i)$	$\frac{t_i - \overline{t}}{s_T} = \mathcal{Z}(t_i)$	$\mathcal{Z}(x_i) \times \mathcal{Z}(t_i)$
1	40.2	104.36	$\frac{40.2-39}{1.789} = 0.67$	$\frac{104.36-102.2}{3.220} = 0.67$	$0.67 \times 0.67 = 0.45$
2	40.5	104.90	$\frac{40.5-39}{1.789} = 0.84$	$\frac{104.90-102.2}{3.220} = 0.84$	$0.84 \times 0.84 = 0.70$
3	41.4	106.52	$\frac{41.4-39}{1.789} = 1.34$	$\frac{106.52-102.2}{3.220} = 1.34$	$1.34 \times 1.34 = 1.80$
4	40.6	105.08	$\frac{40.6-39}{1.789} = 0.89$	$\frac{105.08-102.2}{3.220} = 0.89$	$0.89 \times 0.89 = 0.80$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
14	36.1	96.98	$\frac{36.1-39}{1.789} = -1.62$	$\frac{96.98-102.2}{3.220} = -1.62$	$-1.62 \times -1.62 = 2.63$
15	36.5	97.70	$\frac{36.5-39}{1.789} = -1.40$	$\frac{97.70-102.2}{3.220} = -1.40$	$-1.40 \times -1.40 = 1.95$
16	38.0	100.40	$\frac{38.0-39}{1.789} = -0.56$	$\frac{100.40-102.2}{3.220} = -0.56$	$-0.56 \times -0.56 = 0.31$
					$\sum_{i=1}^{16} \mathcal{Z}(x_i)\mathcal{Z}(t_i) = 16$

La tabla ilustra (a) que las puntuaciones $\mathcal{Z}$ en ambas variables son idénticas para cada observación y (b) la correlación entre X y T es igual a:

$$r_{XT} = \frac{\sum_{i=1}^{16} \mathcal{Z}(x_i)\mathcal{Z}(t_i)}{16} = \frac{16}{16} = 1,$$

conforme lo comprobado en la Propiedad 6.

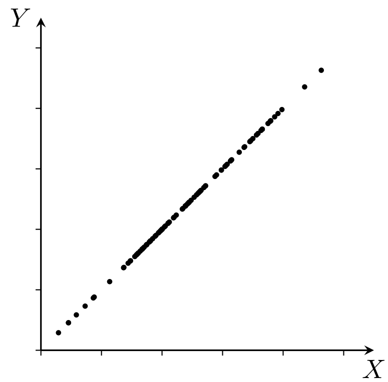
Además, según la derivación en la página anterior, la relación lineal entre T y X se describe por la siguiente recta:

$$\begin{aligned} T &= \left(\frac{s_T}{s_X}\right) X + \left(\overline{t} - \frac{s_T}{s_X}\,\overline{x}\right) \\ &= \left(\frac{3.220}{1.789}\right) X + \left(102.2 - \frac{3.220}{1.789}\,39\right) \\ &= \frac{9}{5}X + 32. \end{aligned}$$

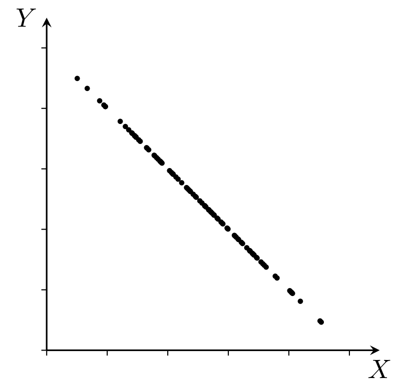
6.2.2.3 Visualizar correlaciones

Concluimos la sección sobre el coeficiente de correlación con algunos diagramas de dispersión que ilustran la fuerza de covariación lineal implicada por diferentes valores típicos del coeficiente de correlación. Cada diagrama representa los datos de 100 observaciones.

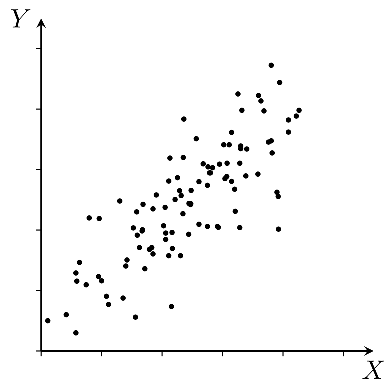
Fuerza de covariación lineal para algunos valores de r_{XY}



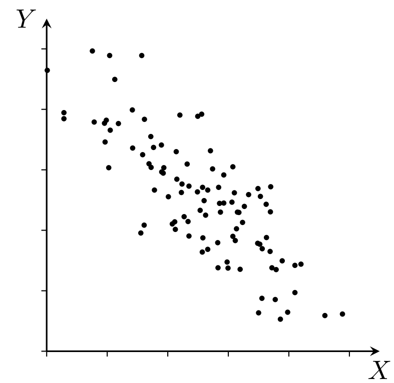
$r_{XY} = +1.00$



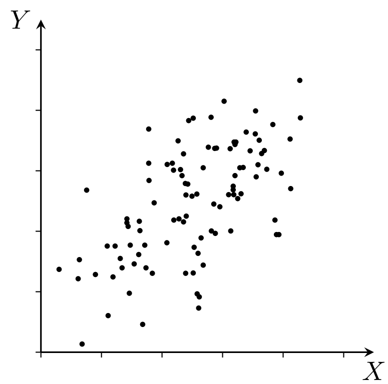
$r_{XY} = -1.00$



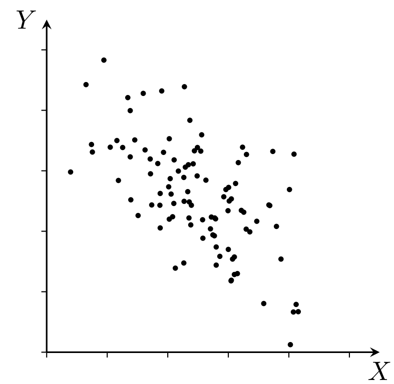
$r_{XY} = +0.80$



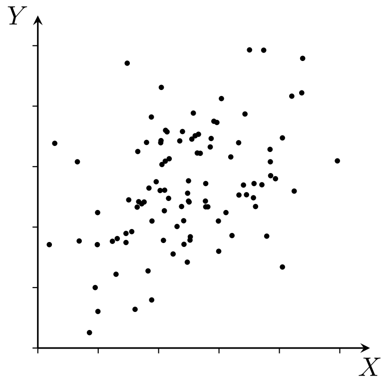
$r_{XY} = -0.80$



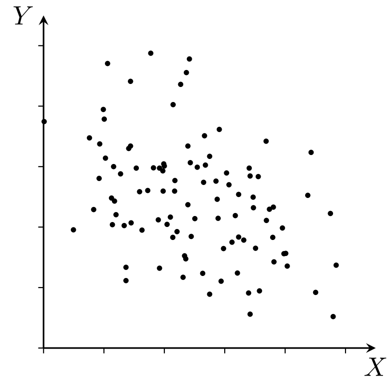
$r_{XY} = +0.60$



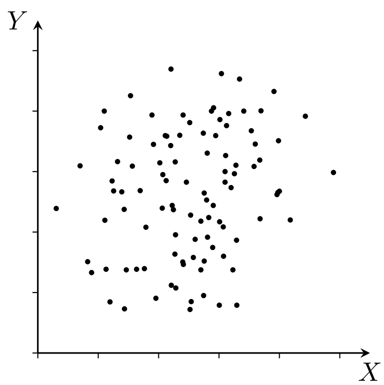
$r_{XY} = -0.60$



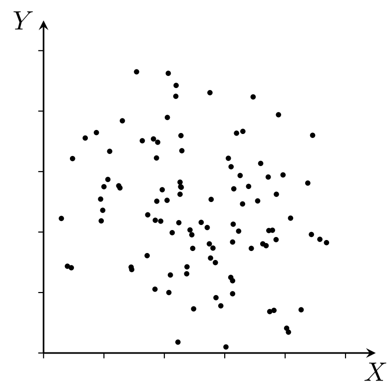
$r_{XY} = +0.40$



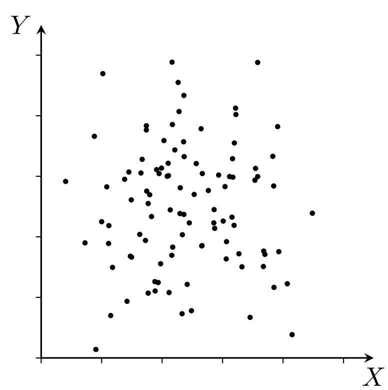
$r_{XY} = -0.40$



$r_{XY} = +0.20$



$r_{XY} = -0.20$



$r_{XY} = 0.00$

Los diagramas anteriores son únicamente ilustraciones de relaciones típicas entre variables X y Y para determinados valores para el coeficiente de correlación. Debe ser claro que la covariación para un valor específico de correlación puede ser distinta a la graficada. Por ejemplo, el diagrama de dispersión en la página 192 ilustra que una correlación de 0 puede esconder una clara covariación curvilínea.

6.2.3 Nota: La correlación de Spearman para variables ordinales

De la definición del coeficiente de correlación sigue que se requiere que las dos variables involucradas tengan el nivel de medición de intervalo: la Ecuación (6.4a) involucra a operaciones (como restar y dividir para calcular la estandarización y multiplicar y sumar las puntuaciones estandarizadas) que sólo tienen sentido con variables de este nivel de medición. Estrictamente hablando, entonces, la correlación no se debería calcular para variables de nivel ordinal. Sin embargo, puede haber interés en cuantificar la relación, directa o inversa, entre dos variables ordinales. Por ejemplo, podríamos tener la hipótesis de que las variables *Nivel Educativo* y *Fumador* en los datos de la página 34 tengan una relación inversa: a mayor nivel de estudios, menos cigarros se fuman.

Un estadístico para cuantificar la covariación entre dos variables de nivel ordinal es el *coeficiente de correlación propuesto por Spearman*. Este estadístico es en esencia la correlación de Pearson introducida en la sección anterior aplicada al número de orden que las observaciones ocupan en las dos variables. El siguiente ejemplo ilustra los detalles del cálculo del coeficiente de Spearman.

Ejemplo

La siguiente tabla muestra, para los 15 profesores de la asignatura de Informática Biomédica, los resultados que obtuvieron en la evaluación de su docencia y la proporción de estudiantes de su grupo que gracias a sus resultados en los exámenes departamentales (un promedio superior a 8.5) quedaron exentos del exámen final.

	Variables originales		Número de orden	
Profesor	Evaluación docencia (X)	Prop. alumnos exentos (Y)	Evaluación docencia ($X^\star$)	Prop. alumnos exentos ($Y^\star$)
J.A.Z.	8.0	.13	12	9
R.A.F.	5.9	.05	2	2
O.P.L.	6.8	.29	4	14
R.C.F.	6.9	.02	5	1
J.S.G.	7.2	.11	7	7
F.G.I.	7.7	.26	10	13
J.E.P.	7.5	.12	9	8
J.P.M.	7.1	.23	6	12
A.R.O.	8.2	.18	13	11
F.P.M.	7.3	.10	8	6
E.G.M.	8.4	.15	14	10
R.L.V.	8.7	.38	15	15
C.F.M.	2.7	.08	1	5
M.M.R.	7.8	.07	11	4
I.C.A.	6.4	.06	3	3

Las dos últimas columnas contienen las variables con los números de orden derivadas de las variables originales. El profesor con la evaluación más baja en X (C.F.M.) ocupa la posición 1 en la variable $X^\star$ y el que tiene la evaluación más alta (R.L.V.) tiene el valor de 15 en $X^\star$. De forma análoga, se define la variable $Y^\star$.

La correlación de Spearman entre las variables X y Y es nada más la correlación de Pearson entre las variables $X^\star$ y $Y^\star$. Gracias a las propiedades de los números de orden, la fórmula para el cálculo de la correlación de Spearman simplifica a:

$$r_s = 1 - \frac{6\sum_{i=1}^{n}(x_i^\star - y_i^\star)^2}{n^3 - n}. \tag{6.5}$$

Calcular la suma en el numerador para los datos de la tabla resulta en:

$$\begin{aligned}\sum_{i=1}^{15}(x_i^\star - y_i^\star)^2 &= (12-9)^2 + (2-2)^2 + (4-14)^2 + (5-1)^2 + (1-5)^2 \\ &\quad + (7-7)^2 + (10-13)^2 + (9-8)^2 + (6-12)^2 + (13-11)^2 \\ &\quad + (8-6)^2 + (14-10)^2 + (15-15)^2 + (11-4)^2 + (3-3)^2 \\ &= 260,\end{aligned}$$

tal que

$$\begin{aligned}r_s &= 1 - \frac{6 \times 260}{15^3 - 15} \\ &= 0.54.\end{aligned}$$

La correlación de Spearman para estos datos indica que hay una relación directa moderada entre el desempeño docente de los profesores y el porcentaje de sus alumnos que aprueban los exámenes departamentales con una nota promedio de 8.5 a más.

Cuando dos o más observaciones empatan en una de las variables, habitualmente se asigna la posición media. Por ejemplo, supongamos que los profesores O.P.L. y R.C.F. tuviesen la misma calificación para su desempeño docente X (por ejemplo, supongamos que O.P.L. tuviese la calificación 6.9 en vez de 6.8). Puesto que O.P.L. y R.C.F. ocupan las posiciones 4 y 5 y los datos no permiten distinguir entre ambos, el valor en la variable $X^\star$ sería 4.5. Del mismo modo, si los profesores C.F.M., M.M.R., y I.C.A. tuviesen el mismo porcentaje de alumnos exentos (por ejemplo, .07), entonces tendrían el valor de 4 en la variable $Y^\star$, ya que 4 es el valor medio de las posiciones—3, 4 y 5—que ocupan. Cabe señalar que la Fórmula (6.5) ya no vale para calcular la correlación de Spearman en caso de que haya empates. En este caso, hay que volver a las Fórmulas (6.4) anteriormente presentadas para calcular la correlación de Pearson y aplicarlas a las variables de orden.

A pesar de su amplio uso, la correlación de Spearman, como estadístico descriptivo para distribuciones de variables ordinales, carece de un fundamento teórico firme. Como se acaba de explicar, la correlación de Spearman es la correlación de Pearson calculada entre dos variables cuyos valores son los números ordinales en las respectivas variables originales. En principio, estas dos variables (las variables $X^\star$ y $Y^\star$ en el ejemplo anterior) deberían cumplir los requisitos de la correlación de Pearson. Es decir, deberían formar una escala de intervalo (lo cual significa, por ejemplo, que la diferencia entre los profesores en las posiciones 1 y 2 respecto de su desempeño docente debería ser la misma que entre los profesores en las posiciones 3 y 4, tal como la diferencia entre 1 metro y 2 metros es la misma que entre 3 metros y 4 metros). No obstante, casi nunca se puede justificar este supuesto.

Por otro lado, la correlación de Spearman, en comparación con la correlación de Pearson (entre las variables originales), tiene como ventaja práctica que es menos sensible a valores atípicos. Además, dentro de la estadística inferencial, tiene un papel importante (y teóricamente justificado), porque la técnica basada en ella permite sacar conclusiones generales (es decir, hacer inferencias) sobre relaciones directas o inversas con supuestos menos exigentes que las técnicas habituales.

Problemas

1. Un sociólogo hace investigación con parejas casadas y recopiló los siguientes datos sobre la relación entre el nivel de estudios alcanzados de ambos esposos en 400 parejas mexicanas:

Nivel educativo del hombre (Y)	*Nivel educativo de la mujer* (X)				
	Primaria	Secundaria	Bachillerato	Universitario	**Total**
Primaria	30	19	5	0	54
Secundaria	13	118	44	5	180
Bachillerato	3	24	49	46	122
Universitario	0	0	8	36	44
Total	46	161	106	87	400

   (a) Selecciona un estadístico para describir la relación de los estudios alcanzados entre los dos miembros de una pareja. Justifica tu respuesta.

   (b) Calcula el valor en este estadístico e interpreta el resultado.

2. Un investigador del departamento de Servicios Escolares de la Facultad de Medicina recopiló la siguiente información sobre los 1102 estudiantes de nuevo ingreso a la facultad de Medicina:

   - *Tipo de bachillerato de procedencia* (variable X), con tres posibles valores: Escuela Nacional Preparatoria (ENP), Colegio de Ciencias y Humanidades (CCH) y un grupo restante de escuelas que no forman parte del sistema de educación superior que ofrece la UNAM (No UNAM).
   - *Sexo* (variable Y) con los valores: femenino y masculino.

   Resumiendo la distribución bivariada de las dos variables en una tabla de contingencia obtiene las siguiente frecuencias:

	Tipo de bachillerato (X)			
Sexo (Y)	ENP	CCH	No UNAM	**Total**
Femenino	301	274	120	695
Masculino	179	102	126	407
Total	480	376	246	1102

   (a) Elige un estadístico para cuantificar la covariación entre *Tipo de bachillerato* y *Sexo*. ¿Cómo justificas la elección?

   (b) Calcula el valor en el estadístico seleccionado con este conjunto de datos;

   (c) Interpreta el valor obtenido.

3. A dos médicos, un especialista en nutrición y un cirujano plástico, se les pide recomendar sobre la necesidad de una operación de reducción de estómago a 20 pacientes con obesidad mórbida. Para cada paciente indican sí o no consideran pertinente la operación. Las opiniones de los dos médicos se presentan en la siguiente tabla:

Número Paciente	Opinión Nutriólogo	Opinión Cirujano plástico	Número Paciente	Opinión Nutriólogo	Opinión Cirujano plástico
1	No	Sí	11	Sí	No
2	No	Sí	12	Sí	No
3	No	Sí	13	No	Sí
4	Sí	No	14	Sí	No
5	No	Sí	15	Sí	No
6	Sí	No	16	No	Sí
7	No	Sí	17	No	Sí
8	No	Sí	18	No	Sí
9	No	Sí	19	No	Sí
10	No	Sí	20	No	Sí

(a) Calcula el valor en el coeficiente κ para cuantificar el grado de acuerdo entre ambos médicos;

(b) Calcula el valor en el coeficiente C de contingencia para los mismos datos;

(c) Interpreta los resultados y justifica tus conclusiones.

4. (a) ¿Qué estadístico utilizarías para cuantificar la covariación entre las variables *Nivel Educativo* y *Fumador* en los datos de la página 34?

(b) Calcula el valor en este estadístico para las 50 personas observadas;

(c) Interpreta el resultado.

5. (a) Calcula la correlación entre *Índice de Masa Corporal* y *Presión Arterial Sistólica* para los datos de la página 34. Interpreta el resultado obtenido.

(b) Anteriormente, ya se señaló que la observación 40 tiene un valor extremo o atípico para *IMC* (de 42 kg/m^2). Elimina esta observación del conjunto de datos y vuelve a calcular la correlación entre *Índice de Masa Corporal* y *Presión Arterial Sistólica.*

(c) También hay dos personas (observaciones 8 y 15) con valores extremos en *Presión Arterial Sistólica*, superando 170 mmHg. Elimina estas dos observaciones y calcula la correlación entre *Índice de Masa Corporal* y *Presión Arterial Sistólica* dentro del conjunto de 48 observaciones.

(d) Formula una regla general de cómo los valores extremos afectan el coeficiente de correlación.

6. Si la variable T se ha obtenido por una transformación lineal de X con $T = aX + b$ y la variable V por una transformación lineal de Y donde $V = cY + d$, entonces

   (a) ¿Cuál es la relación entre s_{TV} y s_{XY}?

   (b) ¿Cuál es la relación entre r_{TV} y r_{XY}?

7. A 100 mujeres, se les midió la estatura en centímetros y en pulgadas. Denominemos la variable X estatura en centímetros y la variable T estatura en pulgadas. Se encontró que:

$$\overline{x} = 165.1 \qquad s_X = 7.62$$
$$\bar{t} = 65.0 \qquad s_T = 3.00$$
$$s_{XT} = 22.86$$

   (a) ¿Cuál es el valor en el coeficiente de correlación r_{XT}?

   (b) ¿Cuál es la función para transformar pulgadas a centímetros? ¿y la función inversa para transformar centímetros a pulgadas?
   Pista: Utiliza el resultado de la derivación en la página 209.

8. En el Caso F del Tema 4 (p. 113), se introdujo un investigador del Centro Nacional de Prevención de Accidentes que realizó un estudio para investigar la relación entre la dosis de alcohol ingerido por un conductor (variable X, medida en mg.) y el tiempo de reacción del mismo en una situación vial simulada (Y, en milisegundos). Supongamos que encontró una correlación de Spearman entre X y Y de .50.

   (a) Interpreta el resultado obtenido para la correlación.

   *(b) Para que la distribución de los tiempos de reacción fuese más simétrica, aplicó una transformación logarítmica a la variable Y como descrita por la Ecuación (4.3), es decir, construyó una nueva variable $Y^\star = \log_{10} Y$. ¿Cómo afecta esta transformación la correlación? En otras palabras, ¿cómo se relacionan, por un lado, la correlación de Spearman entre X y Y y, por otro lado, la correlación de Spearman entre X y $Y^\star$?

Tema 7
Sumas de variables

Índice

En el Tema 4 se introdujo el concepto de la transformación de una variable. En específico, se explicó que una transformación implica la aplicación de una función a una variable original y que resulta en la creación de una nueva variable. Dependiendo del tipo de la función, diferentes tipos de transformaciones fueron definidas: transformaciones en general, transformaciones monótonas y lineales, y estandarizaciones. En la sección subsecuente, se propusieron algunos teoremas que permiten evaluar las características de la distribución de una variable transformada a partir de las características de la distribución de la variable original. Por ejemplo, se presentaron teoremas de la cuales se desprende que la media aritmética de una variable linealmente transformada se calcula fácilmente aplicando la misma transformación lineal a la media de la variable original o que una transformación lineal afecta la varianza únicamente por la constante multiplicativa.

El tema actual generaliza la idea de una transformación: en vez de crear una nueva variable a partir de una única variable original, se considera el caso de crear nuevas variables a partir de dos o más variables existentes. La definición de la nueva variable requiere la especificación de una regla—o una función multivariable (véase la página 16)—que explique cómo combinar las variables originales. Un ejemplo de una función multivariable es la ecuación que permite calcular el índice de masa corporal (variable T) a partir de las variables peso en kilogramos (variable X) y talla en metros (variable Y):

$$T = f(X, Y) = \frac{X}{Y^2}.$$

En palabras, esta función multivariable especifica la regla:

> "Primero, eleva la estatura en metros al cuadrado y, a continuación, divide el peso en kilogramos entre el resultado de la primera operación."

En principio, se podrían considerar diferentes tipos de funciones multivariables, tal como se consideraron en el Tema 4 diferentes tipos de transformaciones. Sin embargo, dentro del alcance de este curso, nos limitamos a funciones multivariables que son relativamente sencillas, a saber, funciones que implican únicamente sumas (o sumas ponderadas) de las variables originales. Aunque el ejemplo del índice de masa corporal no es una función de este tipo, sumas de variables son muy comunes en el área de la salud (tanto como en las otras ramas de la ciencia). Por ejemplo, si una base de datos epidemiológicos incluye una variable para cada entidad federativa mexicana con el número anual de nuevos casos de cáncer de pulmón desde 1950, se puede construir una nueva variable que indica el número de nuevos casos para toda la República Mexicana por una simple suma de las 32 variables correspondientes de las entidades federativas. Otro ejemplo es el índice desarrollado por la Organización Mundial de la Salud para cuantificar el impacto de un problema de salud (como enfermedades, suicidios, accidentes de tráfico, exposición a la contaminación, etc.) y que se llama *Años de Vida Ajustados por Discapacidad*, abreviado como DALY (por sus siglas en inglés: *Disability-Adjusted Life Years*). Este índice se ha definido como la suma de dos otros índices: YLL (*Years of Life Lost*), el número de años perdidos debido a muerte temprana (es decir, el número de años que la persona habría vivido si hubiese cumplido con la esperanza de vida), y YLD (*Years Lived with Disability*), el número de años perdidos por discapacidad. Es decir, DALY $=$ YLL $+$ YLD.

En el área de educación médica y la evaluación educativa, también son muy comunes las sumas de variables. El ejemplo típico son las calificaciones finales para las asignaturas que casi siempre se calculan a través de una suma (simple o ponderada) de partes más elementales. Por ejemplo, la nota final para la asignatura de *Bioestadística Básica* se define a partir de la suma de las siguientes cuatro variables: Las calificaciones en dos tareas, la calificación en el examen teórico y la calificación en el examen de problemas prácticas.

También diferencias se pueden conceptualizar como sumas, o más precisamente como una suma ponderada que incluye coeficientes negativos. Considérese un estudio que evalúa el impacto de una intervención, por ejemplo, una nueva dieta, para pacientes con sobrepeso. En este tipo de estudios, es común realizar una medición previa a la intervención (antes de iniciar la dieta, variable X) y otra posterior a la intervención (después de haber seguido la dieta por un tiempo determinado, variable Y). Entonces, el impacto de la dieta se expresa como la diferencia entre ambas mediciones; es decir, se crea una nueva variable T que se define por

$$T = Y - X.$$

Debe ser claro que la diferencia entre las dos variables se puede entender como una suma de las mismas variables, pero después de haber multiplicado la variable X por la constante -1. En otras palabras, aunque este tema se fija en sumas, las herramientas presentadas también servirán para conocer las propiedades de variables que se obtienen restando dos variables. La última sección de este tema considera explícitamente el caso de sumas ponderadas, que, como se acaba de explicar, incluye el caso especial de diferencias.

Se inicia el tema con el caso más sencillo de una variable T que es la suma simple de dos variables componentes X y Y (Sección 7.1). En particular, se investiga cómo la media y la varianza de T, y la covarianza de T con otras variables, dependen de los estadísticos correspondientes asociados con X y Y. La investigación de los estadísticos de T se centra en la media, la varianza, y la covarianza porque (a) estos son de los estadísticos más importantes en la estadística y (b) mantienen relaciones relativamente simples con los estadísticos de las variables originales. En la Sección 7.2 se generalizan los resultados anteriores al caso de sumas de m variables, donde m es cualquier número natural mayor que a igual a 2. Finalmente, en la última sección se considera el caso más general de sumas ponderadas (o, más preciso, combinaciones lineales) de m variables y cómo los estadísticos de la media, varianza y covarianza asociados con la nueva variable se derivan de los estadísticos correspondientes de las variables originales.

7.1 Sumas de dos variables

Ejemplo

Como ejemplo guía en esta sección, considérese el examen final de la asignatura de *Bioestadística Básica* que consiste en una parte teórica y otra parte de resolver problemas. Cada parte recibe una calificación por separado. Para una muestra de 12 estudiantes que participaron en el examen, denominemos X la calificación que obtuvieron en la parte teórica y Y la calificación en la parte práctica. La calificación total es nada más la suma de las calificaciones en ambas partes. En símbolos, si T representa la calificación total, entonces:

$$T = X + Y.$$

Supongamos que, además de las calificaciones en el examen de *Bioestadística Básica*, se dispone de las notas finales en la asignatura *Metodología de la investigación cualitativa* de los mismos estudiantes. Con la variable V representando dichas notas, la siguiente tabla muestra este conjunto de datos:

	Bioestadística básica			*Investigación cualitativa*
Iniciales estudiante	Teoría X	Problemas Y	Total T	V
J.M.S.	28	43	71	8
Y.C.S.	26	32	58	9
A.M.M.	29	37	66	9
E.F.U.	18	29	47	8
H.H.K.	22	40	62	7
S.D.M.	26	39	65	7
E.M.G.	21	28	49	8
M.C.C.	22	31	53	8
J.M.C.	25	42	67	8
C.P.E.	25	41	66	10
A.P.S.	22	37	59	8
R.C.L.	24	33	57	6

A través de las fórmulas que definen la media aritmética y la varianza (véase las Ecuaciones (3.1) y (3.5)), se pueden calcular estos estadísticos para las cuatro variables. También se pueden calcular las covarianzas entre cualquier par de variables (a través de la Ec. (6.3b)). La siguiente tabla resume los resultados:

	X	Y	T	V
Media aritmética	24	36	60	8
Varianza	9	25	52	1
Covarianzas				
X		9.00	18.00	0.75
Y			34.00	0.50
T				1.25

La última parte de la tabla presenta en cada celda la covarianza entre las variables correspondientes de la fila y de la columna. Se lee, por ejemplo, que la covarianza entre X y T es 18.00 y entre T y V es 1.25. (En la nota de la página 231 se explicará de forma más detallada esta manera para representar las covarianzas entre variables.)

Las preguntas que se contestan en esta sección son las siguientes:

(1) ¿Cuál es la relación entre la media aritmética de la nueva variable T y las medias de las variables originales X y Y?

(2) ¿Cuál es la relación entre la varianza de T y la varianza de las variables originales X y Y?

(3) Si se conocen las covarianzas de las variables originales con otra variable, es decir, las covarianzas entre X y V y entre Y y V, ¿es posible conocer la covarianza entre la nueva variable T y V?

7.1.1 Media aritmética

Media aritmética de una suma de dos variables

Si la variable T se obtiene por la suma de las variables X y Y, es decir $T = X + Y$, entonces:

$$\bar{t} = \overline{x} + \overline{y}. \tag{7.1}$$

Demostración del teorema sobre la media aritmética de una suma de dos variables

De la definición de la media aritmética de T

$$\bar{t} = \frac{\sum_{i=1}^{n} t_i}{n},$$

y puesto que $T = X + Y$ quiere decir que $t_i = x_i + y_i$ para cualquier observación i, sigue directamente:

$$\begin{aligned}
\bar{t} &= \frac{\sum_{i=1}^{n}(x_i + y_i)}{n} \\
&= \frac{\left(\sum_{i=1}^{n} x_i\right) + \left(\sum_{i=1}^{n} y_i\right)}{n} \\
&= \frac{\sum_{i=1}^{n} x_i}{n} + \frac{\sum_{i=1}^{n} y_i}{n} \\
&= \overline{x} + \overline{y}.
\end{aligned}$$

■

Ejemplo

Con los datos del ejemplo en la página 221, aplicando la Ecuación (7.1) a las medias de las calificaciones de las partes teóricas y de problemas del examen, se verifica directamente que

$$\begin{aligned}
\bar{t} &= \overline{x} + \overline{y} \\
&= 24 + 36 \\
&= 60,
\end{aligned}$$

un resultado que efectivamente coincide con lo anteriormente calculado (que se presenta en la tabla de estadísticos de este ejemplo).

7.1.2 Varianza

Varianza de una suma de dos variables

Si la variable T se obtiene por la suma de las variables X y Y, es decir $T = X + Y$, entonces:

$$s_T^2 = s_X^2 + s_Y^2 + 2\, s_{XY}. \tag{7.2}$$

Demostración del teorema sobre la varianza de una suma de dos variables

La varianza de T se define como:

$$s_T^2 \;=\; \frac{\sum\limits_{i=1}^{n}(t_i-\overline{t})^2}{n},$$

Reemplazando t_i por x_i+y_i, y $\overline{t}$ por $\overline{x}+\overline{y}$ (gracias al resultado anterior), se obtiene:

$$\begin{aligned}
s_T^2 \;&=\; \frac{\sum\limits_{i=1}^{n}\left[(x_i+y_i)-(\overline{x}+\overline{y})\right]^2}{n} \\
&=\; \frac{\sum\limits_{i=1}^{n}\left[(x_i-\overline{x})+(y_i-\overline{y})\right]^2}{n} \\
&=\; \frac{\sum\limits_{i=1}^{n}\left[(x_i-\overline{x})^2+(y_i-\overline{y})^2+2(x_i-\overline{x})(y_i-\overline{y})\right]}{n} \\
&=\; \frac{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}{n} \;+\; \frac{\sum\limits_{i=1}^{n}(y_i-\overline{y})^2}{n} \;+\; 2\,\frac{\sum\limits_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{n} \\
&=\; s_X^2 \;+\; s_Y^2 \;+\; 2\,s_{XY}.
\end{aligned}$$

■

El teorema anterior implica que, para conocer la varianza de la suma de dos variables, *no* basta tener información sobre las varianzas de estas variables, sino que también se requiere la covarianza entre ambas. Solo en el caso especial de que dos variables sean linealmente independientes, la varianza de su suma coincide con la suma de sus varianzas.

Como ayuda mnemotécnica, conviene ver la analogía entre la estructura en el teorema anterior y la estructura de una igualdad algebraica bien conocida. Al desarrollar el cuadrado de una suma, se obtiene:

$$(a+b)^2 \;=\; a^2 \;+\; b^2 \;+\; 2ab.$$

En palabras, el *cuadrado* de una suma es igual a la suma del *cuadrado* del primer *término*, más el *cuadrado* del segundo *término* más el doble del *producto* de ambos *términos*. Si en esta frase se cambia “cuadrado” por “varianza”, “término” por “variable” y “producto” por “covarianza”, se obtiene la formulación en palabras de la Ecuación (7.2). En la estadística, calcular la varianza efectivamente significa calcular (la suma de) cuadrados (de puntuaciones diferenciales) y la covarianza significa calcular (la suma de) productos (de puntuaciones diferenciales). También la demostración del teorema (véase la elaboración de la expresión en la tercera línea) aclara la correspondencia entre las estructuras de la Ecuación (7.2) y la ecuación algebraica de $(a+b)^2$.

Ejemplo

Verifiquemos que la varianza de la variable T en los datos del ejemplo en la página 221 (la cual se presenta en la tabla de la página 222) efectivamente coincide con el resultado que se obtiene tras aplicar la Ecuación (7.2):

$$\begin{aligned} s_T^2 &= s_X^2 + s_Y^2 + 2s_{XY} \\ &= 9 + 25 + 2 \times 9 \\ &= 52. \end{aligned}$$

7.1.3 Covarianza

Propiedad distributiva de la covarianza respecto de la suma (Versión 1)

Si la variable T se obtiene por la suma de las variables X y Y, es decir $T = X + Y$, entonces:

$$s_{TV} = s_{XV} + s_{YV}, \tag{7.3}$$

donde V es cualquier variable definida sobre las mismas observaciones.

Demostración de la propiedad distributiva de la covarianza (Versión 1)

La covarianza entre T y V se define como:

$$s_{TV} = \frac{\sum_{i=1}^{n}(t_i - \overline{t})(v_i - \overline{v})}{n},$$

Utilizando las identidades $t_i = x_i + y_i$ (para cada i) y $\overline{t} = \overline{x} + \overline{y}$, se obtiene:

$$\begin{aligned} s_{TV} &= \frac{\sum_{i=1}^{n}\Big[(x_i + y_i) - (\overline{x} + \overline{y})\Big](v_i - \overline{v})}{n} \\ &= \frac{\sum_{i=1}^{n}\Big[(x_i - \overline{x}) + (y_i - \overline{y})\Big](v_i - \overline{v})}{n} \\ &= \frac{\sum_{i=1}^{n}\Big[(x_i - \overline{x})(v_i - \overline{v}) + (y_i - \overline{y})(v_i - \overline{v})\Big]}{n} \\ &= \frac{\sum_{i=1}^{n}(x_i - \overline{x})(v_i - \overline{v})}{n} + \frac{\sum_{i=1}^{n}(y_i - \overline{y})(v_i - \overline{v})}{n} \\ &= s_{XV} + s_{YV}. \end{aligned}$$

■

También esta propiedad tiene una estructura similar a la de una propiedad algebraica mejor conocida. También en la algebra se conocen propiedades distributivas, como la distributividad del producto respecto de la suma, la cual se expresa en la siguiente fórmula:

$$(a + b) \times c = (a \times c) + (b \times c).$$

Nótese que la demostración del teorema (en la transición de la tercera a la cuarta línea) utiliza la identidad expresada en esta fórmula.

Ejemplo

Utilizar la Ecuación (7.3) para derivar la covarianza entre T y V en los datos del ejemplo en la página 221 resulta en:

$$\begin{aligned} s_{TV} &= s_{XV} + s_{YV} \\ &= 0.75 + 0.50 \\ &= 1.25. \end{aligned}$$

Este resultado coincide con la covarianza que se calculó por la fórmula habitual de la covarianza, presentada en la tabla de la página 222.

7.2 Sumas de m variables

En esta sección se extienden las propiedades derivadas en la sección anterior al caso más general de una suma de m variables, donde m es un número natural mayor que o igual a 2. Mientras que, con dos variables originales, se utilizaban los símbolos X y Y para representarlas, ahora para el caso de m variables, se utilizará $X_1, X_2, \ldots, X_m$. Por lo tanto, se define la nueva variable T como:

$$T = X_1 + X_2 + X_3 + \cdots + X_{m-1} + X_m,$$

o, utilizando el signo sumatorio:

$$T = \sum_{j=1}^{m} X_j.$$

Es importante enfatizar que $X_1, X_2, \ldots, X_m$ son *variables* (contrario a $x_1, x_2, \ldots, x_m$, que hemos utilizado con frecuencia para anotar los distintos *valores* en una variable X, como, por ejemplo, en la definición de la función de frecuencia en la página 40).* Como se aclaró en la nota de la página 26, es fundamental distinguir entre variables y los valores en las variables. Con la importancia de esta distinción entre variables y sus valores en mente, seguimos representando variables con mayúsculas y valores (o números) por minúsculas.

*Aunque en los temas anteriores el número m indicaba el número de valores distintos en la variable X, aquí se utiliza el mismo símbolo para referir al número de variables en la suma que define la variable T. En general, será claro del contexto a qué refiere m. En este tema, m se utiliza exclusivamente para indicar el número de variables en una suma.

Cada una de las variables $X_1, X_2, \ldots, X_m$ tienen valores en n observaciones. (Recuérdese que en la representación habitual de un conjunto de datos multivariados, las filas representan las observaciones y las columnas las variables.) Para referir al valor de una observación en una de las variables, se necesitan entonces dos subíndices: Se escribe x_{ij} para referirse al valor que tiene la observación i en la variable X_j. Es decir, el primer subíndice se refiere a la observación, el segundo a la variable.

Ejemplo

Para el ejemplo guía de esta sección, se presentarán unos datos del examen diagnóstico que se aplica a todos los estudiantes de nuevo ingreso a la Facultad de Medicina, al inicio del primer semestre. Este examen diagnóstico consiste en un total de 120 preguntas de opción múltiple, divididas en 8 áreas: matemáticas (32 preguntas), física (16 preguntas), química (16 preguntas), biología (16 preguntas), historia universal (10 preguntas), historia de México (10 preguntas), literatura (10 preguntas) y geografía (10 preguntas). Para cada estudiante, se calcula su calificación (el número de preguntas que contestó correctamente) para las 8 áreas. Se definen las siguientes variables:

X_1 : Calificación en matemáticas
X_2 : Calificación en física
X_3 : Calificación en química
X_4 : Calificación en biología
X_5 : Calificación en historia universal
X_6 : Calificación en historia de México
X_7 : Calificación en literatura
X_8 : Calificación en geografía

La calificación total (T) en el examen diagnóstico iguala la suma de las calificaciones parciales:

$$T = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8.$$

Consideremos adicionalmente otra variable definida sobre la misma muestra de estudiantes: su nota promedio de las asignaturas del primer año, representada por V. Así, los valores de la variable T se obtuvieron al inicio del año académico; la información de la variable V se añadió al final del mismo.

La siguiente tabla muestra la estructura de los datos de 910 estudiantes de nuevo ingreso en las variables anteriores:

	Examen diagnóstico									*Nota promedio*
No. estudiante	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	T	V
1	7	6	8	9	6	2	4	3	45	5.8
2	18	8	6	8	4	6	8	7	65	7.8
3	24	12	13	14	10	10	9	10	102	8.5
4	17	8	11	10	5	3	8	6	68	8.3
5	10	7	7	5	4	3	6	3	45	6.7
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
908	19	12	12	11	8	7	6	5	80	8.3
909	23	6	10	12	6	9	6	5	77	8.2
910	8	10	9	10	6	5	9	6	63	7.7

Un investigador del Centro de Servicios Escolares de la Facultad nos comparte los siguientes estadísticos calculados sobre estos datos:

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	V
Media aritmética	17.28	8.88	9.89	10.41	6.69	5.86	6.65	6.25	7.48
Varianza	40.76	8.23	8.38	7.60	4.67	5.98	3.52	4.43	1.08
Covarianza con V	3.52	1.54	1.63	1.50	0.96	1.08	0.79	1.00	—

Las preguntas que se contestan en esta sección son las siguientes:

(1) ¿Cuál es la relación entre la media aritmética de la nueva variable T (calificación total en el examen diagnóstico) y las medias de las variables originales $X_1, X_2, \ldots, X_8$ (calificaciones por área)?

(2) ¿Cuál es la relación entre la varianza de T y las varianzas de las variables originales $X_1, X_2, \ldots, X_8$?

(3) Si se conoce la covarianza de cada variable original con la variable V (nota promedio), ¿es posible conocer la covarianza entre la nueva variable T y V?

7.2.1 Media aritmética

Media aritmética de una suma de m variables

Si la variable T se obtiene por la suma de las variables $X_1, X_2, \ldots, X_m$, es decir $T = \sum_{j=1}^{m} X_j$, entonces:

$$\bar{t} = \sum_{j=1}^{m} \overline{x_j}. \tag{7.4}$$

Las demostraciones de este teorema y las que siguen en el resto de este tema se dejan como ejercicios; son generalizaciones directas de las demostraciones de los teoremas correspondientes presentados en la Sección 7.1. Por otro lado, sí se discute su interpretación y se ilustran los teoremas con ejemplos.

Ejemplo

El teorema anterior permite conocer la media de la calificación total en el examen diagnóstico a partir de las medias de las ocho calificaciones parciales para los datos del ejemplo de la página 227. Simplemente corresponde con la suma de dichas medias:

$$\bar{t} = \overline{x_1} + \overline{x_2} + \overline{x_3} + \overline{x_4} + \overline{x_5} + \overline{x_6} + \overline{x_7} + \overline{x_8},$$

las cuales se resumieron en la última tabla del ejemplo. Sustituirlas por los valores correctos resulta en:

$$\begin{aligned}\bar{t} &= 17.28 + 8.88 + 9.89 + 10.41 + 6.69 + 5.86 + 6.65 + 6.25\\ &= 71.90.\end{aligned}$$

Es decir, en promedio los estudiantes de nuevo ingreso dan la respuesta correcta a alrededor de 72 de las 120 preguntas del examen diagnóstico.

7.2.2 Varianza

Varianza de una suma de m variables

Si la variable T se obtiene por la suma de las variables $X_1, X_2, \ldots, X_m$, es decir $T = \sum_{j=1}^{m} X_j$, entonces:

$$s_T^2 = \sum_{j=1}^{m} s_{X_j}^2 + 2 \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} s_{X_j X_k}. \tag{7.5}$$

A pesar de que parece más complejo la expresión de la varianza en este teorema en comparación con la Ecuación (7.2), es una generalización directa de la misma. La Ecuación (7.2) dice: “la varianza de la suma de dos variables se obtiene por la suma de las varianzas de cada una de las dos variables, más el doble de la covarianza entre las dos variables”. El teorema para la suma de m variables se lee como: “la varianza de la suma de m variables se obtiene por la suma de las varianzas de cada una de las m variables, más el doble de las covarianzas entre *pares de dos variables* que se pueden formar a partir de las m variables”. Ilustrémoslo con un ejemplo.

Ejemplo

Apliquemos el teorema anterior al ejemplo de la página 227 para conocer la varianza de la calificación total en el examen diagnóstico. Entonces, se desarrollan las sumas en la Ecuación (7.5). La primera suma es simplemente:

$$\sum_{j=1}^{m} s_{X_j}^2 = s_{X_1}^2 + s_{X_2}^2 + s_{X_3}^2 + s_{X_4}^2 + s_{X_5}^2 + s_{X_6}^2 + s_{X_7}^2 + s_{X_8}^2.$$

Estas varianzas se encuentran en la última tabla del ejemplo. Sustituir da:

$$\begin{aligned}\sum_{j=1}^{m} s_{X_j}^2 &= 40.76 + 8.23 + 8.38 + 7.60 + 4.67 + 5.98 + 3.52 + 4.43\\ &= 83.57.\end{aligned}$$

Esto es un resultado intermedio. Para conocer la varianza de T, es necesario desarrollar también la segunda suma. Hagámoslo en dos pasos; primero la suma que corresponde al primer signo sumatorio. Este signo sumatorio especifica que el índice j varía de 1 a $m-1$; dado que $m = 8$ en nuestro ejemplo, entonces j varía de 1 a 7:

$$\sum_{j=1}^{7}\sum_{k=j+1}^{8} s_{X_jX_k} = \underbrace{\sum_{k=1+1}^{8} s_{X_1X_k}}_{j=1} + \underbrace{\sum_{k=2+1}^{8} s_{X_2X_k}}_{j=2} + \underbrace{\sum_{k=3+1}^{8} s_{X_3X_k}}_{j=3} + \underbrace{\sum_{k=4+1}^{8} s_{X_4X_k}}_{j=4}$$
$$+ \underbrace{\sum_{k=5+1}^{8} s_{X_5X_k}}_{j=5} + \underbrace{\sum_{k=6+1}^{8} s_{X_6X_k}}_{j=6} + \underbrace{\sum_{k=7+1}^{8} s_{X_7X_k}}_{j=7}.$$

Como siguiente paso, se elabora cada término (cada una de las siete sumas del lado derecho de la ecuación):

$$\sum_{k=1+1}^{8} s_{X_1X_k} = \sum_{k=2}^{8} s_{X_1X_k} = \underbrace{s_{X_1X_2}}_{k=2} + \underbrace{s_{X_1X_3}}_{k=3} + \underbrace{s_{X_1X_4}}_{k=4} + \underbrace{s_{X_1X_5}}_{k=5} + \underbrace{s_{X_1X_6}}_{k=6} + \underbrace{s_{X_1X_7}}_{k=7} + \underbrace{s_{X_1X_8}}_{k=8},$$

lo cual es la suma de todas las covarianzas entre pares de variables que incluyen X_1. El siguiente término corresponde con la suma de las covarianzas que incluyen X_2, y no X_1 (ya que éstas ya se incluyeron en la suma anterior):

$$\sum_{k=2+1}^{8} s_{X_2X_k} = \sum_{k=3}^{8} s_{X_2X_k} = \underbrace{s_{X_2X_3}}_{k=3} + \underbrace{s_{X_2X_4}}_{k=4} + \underbrace{s_{X_2X_5}}_{k=5} + \underbrace{s_{X_2X_6}}_{k=6} + \underbrace{s_{X_2X_7}}_{k=7} + \underbrace{s_{X_2X_8}}_{k=8}.$$

El desarrollo de los otros términos (y su interpretación) es similar:

$$\sum_{k=3+1}^{8} s_{X_3X_k} = \sum_{k=4}^{8} s_{X_3X_k} = \underbrace{s_{X_3X_4}}_{k=4} + \underbrace{s_{X_3X_5}}_{k=5} + \underbrace{s_{X_3X_6}}_{k=6} + \underbrace{s_{X_3X_7}}_{k=7} + \underbrace{s_{X_3X_8}}_{k=8}$$
$$\sum_{k=4+1}^{8} s_{X_4X_k} = \sum_{k=5}^{8} s_{X_4X_k} = \underbrace{s_{X_4X_5}}_{k=5} + \underbrace{s_{X_4X_6}}_{k=6} + \underbrace{s_{X_4X_7}}_{k=7} + \underbrace{s_{X_4X_8}}_{k=8}$$
$$\sum_{k=5+1}^{8} s_{X_5X_k} = \sum_{k=6}^{8} s_{X_5X_k} = \underbrace{s_{X_5X_6}}_{k=6} + \underbrace{s_{X_5X_7}}_{k=7} + \underbrace{s_{X_5X_8}}_{k=8}$$
$$\sum_{k=6+1}^{8} s_{X_6X_k} = \sum_{k=7}^{8} s_{X_6X_k} = \underbrace{s_{X_6X_7}}_{k=7} + \underbrace{s_{X_6X_8}}_{k=8}$$
$$\sum_{k=7+1}^{8} s_{X_7X_k} = \sum_{k=8}^{8} s_{X_7X_k} = \underbrace{s_{X_7X_8}}_{k=8}.$$

Entonces, la doble suma en la Ecuación (7.5) corresponde con la suma de las siete ecuaciones anteriores (incluye 28 covarianzas). Antes de seguir con el cálculo de la varianza de T, se introduce un nuevo concepto.

(Continuará...)

La elaboración anterior muestra que la doble suma en la Ecuación (7.5) efectivamente es la suma de las covarianzas entre todos los posibles pares que se pueden ensamblar de $X_1, X_2, \ldots, X_m$. Una herramienta que se utiliza con frecuencia para representar la información que se requiere para la Ecuación (7.5) es la *matriz de varianzas y covarianzas*.

Matriz de varianzas y covarianzas

La *matriz de varianzas y covarianzas* para variables $X_1, X_2, \ldots, X_m$ es una representación tabular con m filas y m columnas. La primera fila y la primera columna refieren a X_1, la segunda fila y la segunda columna a X_2, etc., hasta la última fila y la última columna que refieren a X_m. Las celdas de la matriz contienen la covarianza entre la variable de la fila y la variable de la columna.

$$\begin{array}{c c}
 & \begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & \cdots & X_m \end{array} \\
\begin{array}{c} X_1 \\ X_2 \\ X_3 \\ X_4 \\ \vdots \\ X_m \end{array} &
\begin{pmatrix}
s_{X_1X_1} & s_{X_1X_2} & s_{X_1X_3} & s_{X_1X_4} & \cdots & s_{X_1X_m} \\
s_{X_2X_1} & s_{X_2X_2} & s_{X_2X_3} & s_{X_2X_4} & \cdots & s_{X_2X_m} \\
s_{X_3X_1} & s_{X_3X_2} & s_{X_3X_3} & s_{X_3X_4} & \cdots & s_{X_3X_m} \\
s_{X_4X_1} & s_{X_4X_2} & s_{X_4X_3} & s_{X_4X_4} & \cdots & s_{X_4X_m} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
s_{X_mX_1} & s_{X_mX_2} & s_{X_mX_3} & s_{X_mX_4} & \cdots & s_{X_mX_m}
\end{pmatrix}
\end{array}$$

Conviene resaltar las siguientes dos propiedades de esta matriz:

- Cada celda en la diagonal principal de la matriz contiene una covarianza de una de las variables X_j consigo misma: $s_{X_1X_1}, s_{X_2X_2}, \ldots, s_{X_mX_m}$. Por la segunda propiedad de covarianzas (véase, p. 188), estas covarianzas coinciden con las varianzas de las variables: $s_{X_1X_1} = s^2_{X_1}, \ldots, s_{X_mX_m} = s^2_{X_m}$.
- Por la primera propiedad de covarianzas (véase, p. 188), la covarianza en la celda correspondiente de la fila 1 y la columna 2 es la misma que la covarianza en la celda de la fila 2 y la columna 1 ($s_{X_1X_2} = s_{X_2X_1}$). En general, cada elemento arriba de la diagonal principal corresponde con otro elemento debajo de esta diagonal ($s_{X_jX_k} = s_{X_kX_j}$). Es decir, la matriz de varianzas y covarianzas es simétrica.

Estas propiedades permiten escribir la matriz de varianzas y covarianzas como sigue:

$$\begin{array}{c c}
 & \begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & \cdots & X_m \end{array} \\
\begin{array}{c} X_1 \\ X_2 \\ X_3 \\ X_4 \\ \vdots \\ X_m \end{array} &
\begin{pmatrix}
s^2_{X_1} & s_{X_1X_2} & s_{X_1X_3} & s_{X_1X_4} & \cdots & s_{X_1X_m} \\
s_{X_1X_2} & s^2_{X_2} & s_{X_2X_3} & s_{X_2X_4} & \cdots & s_{X_2X_m} \\
s_{X_1X_3} & s_{X_2X_3} & s^2_{X_3} & s_{X_3X_4} & \cdots & s_{X_3X_m} \\
s_{X_1X_4} & s_{X_2X_4} & s_{X_3X_4} & s^2_{X_4} & \cdots & s_{X_4X_m} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
s_{X_1X_m} & s_{X_2X_m} & s_{X_3X_m} & s_{X_4X_m} & \cdots & s^2_{X_m}
\end{pmatrix}
\end{array}$$

Ejemplo

Continuemos con la derivación de la varianza de la calificación total en el examen diagnóstico. La matriz de varianzas y covarianzas de las calificaciones parciales no se incluyó en la tabla de los estadísticos en la página 228. Se presenta a continuación:

$$\begin{array}{cc} & \begin{array}{cccccccc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 & X_8 \end{array} \\ \begin{array}{c} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \\ X_8 \end{array} & \begin{pmatrix} 40.76 & 12.03 & 11.12 & 9.64 & 7.18 & 7.80 & 5.35 & 6.94 \\ 12.03 & 8.23 & 4.94 & 4.48 & 3.30 & 3.81 & 2.51 & 3.19 \\ 11.12 & 4.94 & 8.38 & 4.76 & 3.12 & 3.81 & 2.43 & 3.08 \\ 9.64 & 4.48 & 4.76 & 7.60 & 3.10 & 3.63 & 2.36 & 2.80 \\ 7.18 & 3.30 & 3.12 & 3.10 & 4.67 & 3.27 & 1.90 & 2.62 \\ 7.80 & 3.81 & 3.81 & 3.63 & 3.27 & 5.98 & 2.15 & 2.94 \\ 5.35 & 2.51 & 2.43 & 2.36 & 1.90 & 2.15 & 3.52 & 1.67 \\ 6.94 & 3.19 & 3.08 & 2.80 & 2.62 & 2.94 & 1.67 & 4.43 \end{pmatrix} \end{array}$$

Las varianzas (que sí se presentaron en la tabla de los estadísticos en la página 228) se encuentran en la diagonal principal (en color azul) de la matriz. Además, se observa que la matriz es simétrica respecto a la diagonal principal: por ejemplo, los valores en las celdas de la primera fila coinciden con los de las celdas de la primera columna.

Retomando el cálculo de la varianza de la variable $T = \sum_{j=1}^{8} X_j$ a través de la Ecuación (7.5), se reconoce que las dos sumas de la fórmula coninciden con diferentes partes de la matriz de varianzas y covarianzas:

$$\sum_{j=1}^{8} s_{X_j}^2 \quad \longrightarrow \quad \text{la suma de los elementos en la diagonal principal (azul)}$$

$$\sum_{j=1}^{8-1} \sum_{k=j+1}^{m} s_{X_j X_k} \quad \longrightarrow \quad \text{la suma de los elementos arriba de la diagonal principal (rojo)}$$

Además, como la suma de los elementos arriba de la diagonal principal es igual a la suma de los elementos debajo de esa diagonal, es claro que el resultado total de la Ecuación (7.5) es precisamente la suma de todos los elementos (tanto los en la diagonal principal, como los que se encuentran arriba y debajo de la diagonal) de la matriz de varianzas y covarianzas.

La derivación anterior muestra que la Ecuación (7.5) se puede escribir de forma alternativa y más breve como:

$$s_T^2 = \sum_{j=1}^{m} \sum_{k=1}^{m} s_{X_j X_k}.$$

Sin embargo, esta fórmula no explicita que la suma incluye covarianzas de variables consigo mismas y covarianzas que se repiten.

Concluyendo este ejemplo, se deriva que la varianza de la calificación total a partir de la información en la matriz de varianzas y covarianzas de las calificaciones parciales es igual a:

$$\begin{aligned} s_T^2 = \sum_{j=1}^{8} s_{X_j}^2 \; + \; 2\sum_{j=1}^{8-1}\sum_{k=j+1}^{m} s_{X_jX_k} &= 40.76 + 8.23 + 8.38 + \cdots + 3.52 + 4.43 \\ &\quad + 2\,(12.03 + 11.12 + 9.64 + \cdots + 2.94 + 1.67) \\ &= 335.42. \end{aligned}$$

La varianza de T entonces es 335.42; la desviación estándar es $\sqrt{335.42} = 18.31$.

7.2.3 Covarianza

Propiedad distributiva de la covarianza respecto de la suma (Versión 2)

Si la variable T se obtiene por la suma de las variables $X_1, X_2, \ldots, X_m$, es decir $T = \sum_{j=1}^{m} X_j$, entonces:

$$s_{TV} \; = \; \sum_{j=1}^{m} s_{X_jV}, \tag{7.6}$$

donde V es cualquier variable definida sobre las mismas observaciones.

Ejemplo

El teorema anterior permite calcular la covarianza entre T, la calificación total en el examen diagnóstico, y V, la nota promedio en las asignaturas del primer año. La tabla con los estadísticos del ejemplo de la página 227 incluye, en la última fila, las covarianzas de cada calificación parcial (X_j) con la nota promedio V. Por lo tanto, directamente se obtiene:

$$\begin{aligned} s_{TV} &= s_{X_1V} + s_{X_2V} + s_{X_3V} + s_{X_4V} + s_{X_5V} + s_{X_6V} + s_{X_7V} + s_{X_8V} \\ &= 3.52 + 1.54 + 1.63 + 1.50 + 0.96 + 1.08 + 0.79 + 1.00 \\ &= 12.02. \end{aligned}$$

Nótese que, a partir de este resultado, y la información sobre la varianza de T (véase el final de la sección anterior) y V (en la tabla con los estadísticos del ejemplo), se puede derivar la correlación entre ambas variables (aplicando la Ecuación (6.4b)):

$$\begin{aligned} r_{TV} &= \frac{s_{TV}}{s_T\, s_V} \\ &= \frac{12.02}{\sqrt{335.42}\,\sqrt{1.08}} \\ &= .63. \end{aligned}$$

Es decir, existe una relación lineal directa relativamente fuerte entre la calificación total en el examen diagnóstico y la nota promedio de las asignaturas del primera año de la carrera.

En el teorema anterior se mencionó que la variable V de la Ecuación (7.6) puede ser cualquier variable. Es posible, por ejemplo, que también V, tal como T, sea una variable que resulta de una suma de otras variables. Por ejemplo, considerando la nota promedio de las asignaturas del primer año—es decir, la variable V en nuestro ejemplo guía—, es claro que esta variable también es compuesta de otras variables más elementales, a saber las notas finales de las asignaturas del primer año.

El siguiente teorema, que generaliza el teorema anterior, expresa la covarianza entre dos variables T y V, donde ambas resultaron de una suma de variables más elementales; el teorema expresa esta varianza en términos de la suma de las covarianzas entre pares de variables componentes:

Propiedad distributiva de la covarianza respecto de la suma (Versión 3)

Si (a) la variable T se obtiene por la suma de las variables $X_1, X_2, \ldots, X_m$, es decir $T = \sum_{j=1}^{m} X_j$, y (b) la variable V se obtiene por la suma de las variables $Y_1, Y_2, \ldots, Y_r$, es decir $V = \sum_{k=1}^{r} Y_k$, entonces:

$$s_{TV} = \sum_{j=1}^{m} \sum_{k=1}^{r} s_{X_j Y_k}. \tag{7.7}$$

En palabras, si T y V son sumas de variables más elementales, entonces su covarianza se obtiene por la suma de todas las distintas covarianzas que involucran un componente de T y un componente de V. Las covarianzas involucradas en la suma de la Ecuación (7.7) se pueden organizar en una *matriz de covarianzas.*

Matriz de covarianzas entre dos conjuntos de variables

En la nota de la página 231, se introdujo el concepto de la matriz de varianzas y covarianzas para un conjunto de variables $X_1, X_2, \ldots, X_m$. Esta matriz era una matriz cuadrada (es decir, con el número de filas igual al número de columnas); en efecto, las filas y las columnas de la matriz referían a las mismas variables. Además, esta matriz era simétrica respecto de la diagonal principal. En cada celda de la matriz se encontraba la covarianza entre la variable de la fila y la variable de la columna (la cual en la diagonal principal coincidía con una varianza).

Un concepto muy relacionado es la *matriz de covarianzas* entre dos conjuntos de variables. Supongamos que el primer conjunto de variables es $X_1, X_2, \ldots, X_m$ y el segundo conjunto consiste en las variables $Y_1, Y_2, \ldots, Y_r$. En este caso, la matriz de covarianzas tiene m filas, que corresponden con las respectivas variables $X_1, X_2, \ldots, X_m$, y r columnas, que corresponden con las respectivas variables $Y_1, Y_2, \ldots, Y_r$. La celda en el cruce de la fila j y la columna k presenta la covarianza $s_{X_j Y_k}$, tal como se muestra a continuación:

$$
\begin{array}{c}
 \\ X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_m
\end{array}
\begin{array}{c}
\begin{array}{cccccc} Y_1 & Y_2 & Y_3 & Y_4 & \cdots & Y_r \end{array} \\
\begin{pmatrix}
s_{X_1Y_1} & s_{X_1Y_2} & s_{X_1Y_3} & s_{X_1Y_4} & \cdots & s_{X_1Y_r} \\
s_{X_2Y_1} & s_{X_2Y_2} & s_{X_2Y_3} & s_{X_2Y_4} & \cdots & s_{X_2Y_r} \\
s_{X_3Y_1} & s_{X_3Y_2} & s_{X_3Y_3} & s_{X_3Y_4} & \cdots & s_{X_3Y_r} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
s_{X_mY_1} & s_{X_mY_2} & s_{X_mY_3} & s_{X_mY_4} & \cdots & s_{X_mY_r}
\end{pmatrix}
\end{array}
$$

Ejemplo

En el ejemplo guía de esta sección (p. 227), se introdujeron las ocho calificaciones parciales en el examen diagnóstico aplicado a los estudiantes de nuevo ingreso a la Facultad de Medicina: $X_1, X_2, \ldots, X_8$. Consideremos ahora, para los mismos estudiantes, sus notas finales en las seis asignaturas del primer año de la carrera:

Y_1 : Nota final en *Anatomía*
Y_2 : Nota final en *Biología del Desarrollo*
Y_3 : Nota final en *Bioquímica y Biología Molecular*
Y_4 : Nota final en *Biología Celular y Tisular*
Y_5 : Nota final en *Salud Pública I*
Y_6 : Nota final en *Psicología Médica I*

A continuación, se presenta la matriz de covarianzas entre, por un lado, las calificaciones parciales del examen diagnóstico $(X_1, X_2, \ldots, X_8)$ y, por otro lado, las notas finales de las asignaturas $(Y_1, Y_2, \ldots, Y_6)$

$$
\begin{array}{c}
 \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \\ X_8
\end{array}
\begin{array}{c}
\begin{array}{cccccc} Y_1 & Y_2 & Y_3 & Y_4 & Y_5 & Y_6 \end{array} \\
\begin{pmatrix}
4.89 & 4.50 & 4.14 & 4.61 & 1.32 & 1.67 \\
2.06 & 2.04 & 1.78 & 1.99 & 0.62 & 0.73 \\
2.02 & 2.11 & 1.95 & 2.11 & 0.71 & 0.86 \\
1.83 & 1.99 & 1.70 & 1.95 & 0.67 & 0.88 \\
1.28 & 1.17 & 1.12 & 1.21 & 0.44 & 0.52 \\
1.41 & 1.42 & 1.19 & 1.44 & 0.47 & 0.56 \\
1.00 & 1.01 & 0.88 & 1.05 & 0.35 & 0.46 \\
1.33 & 1.31 & 1.15 & 1.33 & 0.39 & 0.49
\end{pmatrix}
\end{array}
$$

Nótese que esta matriz (contrario a la matriz de varianzas y covarianzas), no es simétrica (ni siquiera es cuadrada).

Si la suma de las notas finales se denota por la variable $V^\star$, es decir,

$$V^\star = \sum_{k=1}^{6} Y_k,$$

entonces la Ecuación (7.7) permite calcular la covarianza entre T (la suma de las ocho variables $X_1, X_2, \ldots, X_8$) y $V^\star$:

$$\begin{aligned}
s_{TV^\star} &= \sum_{j=1}^{m}\sum_{k=1}^{r} s_{X_jY_k}\\
&= s_{X_1Y_1} + s_{X_1Y_2} + s_{X_1Y_3} + s_{X_1Y_4} + s_{X_1Y_5} + s_{X_1Y_6}\\
&\quad s_{X_2Y_1} + s_{X_2Y_2} + s_{X_2Y_3} + s_{X_2Y_4} + s_{X_2Y_5} + s_{X_2Y_6}\\
&\quad s_{X_3Y_1} + s_{X_3Y_2} + s_{X_3Y_3} + s_{X_3Y_4} + s_{X_3Y_5} + s_{X_3Y_6}\\
&\quad s_{X_4Y_1} + s_{X_4Y_2} + s_{X_4Y_3} + s_{X_4Y_4} + s_{X_4Y_5} + s_{X_4Y_6}\\
&\quad s_{X_5Y_1} + s_{X_5Y_2} + s_{X_5Y_3} + s_{X_5Y_4} + s_{X_5Y_5} + s_{X_5Y_6}\\
&\quad s_{X_6Y_1} + s_{X_6Y_2} + s_{X_6Y_3} + s_{X_6Y_4} + s_{X_6Y_5} + s_{X_6Y_6}\\
&\quad s_{X_7Y_1} + s_{X_7Y_2} + s_{X_7Y_3} + s_{X_7Y_4} + s_{X_7Y_5} + s_{X_7Y_6}\\
&\quad s_{X_8Y_1} + s_{X_8Y_2} + s_{X_8Y_3} + s_{X_8Y_4} + s_{X_8Y_5} + s_{X_8Y_6},
\end{aligned}$$

lo cual es precisamente la suma de todos los elementos de la matriz arriba:

$$\begin{aligned}
s_{TV^\star} &= 4.89 + 4.50 + 4.14 + 4.61 + \cdots + 1.33 + 0.39 + 0.49\\
&= 72.09.
\end{aligned}$$

La nota *promedio* V que se introdujo al inicio de esta sección no es la misma variable que $V^\star$, la cual es la *suma* de las notas finales. Sin embargo, ambas están relacionadas por la siguiente ecuación:

$$V = \frac{V^\star}{6}.$$

Es decir, V es una transformación lineal de $V^\star$ (con constante multiplicativa $\frac{1}{6}$ y constante aditiva 0). Por la Propiedad 4 de la covarianza (p. 192), las covarianzas s_{TV} y $s_{TV^\star}$ están relacionadas por:

$$\begin{aligned}
s_{TV} &= \frac{s_{TV^\star}}{6}\\
&= \frac{72.09}{6}\\
&= 12.02,
\end{aligned}$$

el mismo resultado al que llegamos por otra vía en la página 233.

7.3 Combinaciones lineales de m variables

En la última sección de este tema, se considera el caso más general de una combinación lineal de m variables. Recuérdese que una combinación lineal de las variables $X_1, X_2, \ldots, X_m$ tiene la siguiente forma:

$$T \;=\; b_0 \;+\; b_1\,X_1 \;+\; b_2\,X_2 \;+\; b_3\,X_3 \;+\; \cdots \;+\; b_{m-1}\,X_{m-1} \;+\; b_m\,X_m,$$

o, de forma equivalente,

$$T \;=\; b_0 \;+\; \sum_{j=1}^{m} b_j\,X_j,$$

donde $b_0, b_1, \ldots, b_m$ son constantes (elementos de $\mathbb{R}$). La diferencia entre una suma simple de m variables (que se introdujo en la sección anterior) y una combinación lineal de m variables es que, en el último caso, (a) cada variable tiene *un peso* por el cual se multiplica (el peso de la variable X_j es b_j) y (b) a la suma se añade una constante aditiva b_0. (Se puede entender una combinación lineal como una suma ponderada a la cual se ha añadido una constante b_0.) A veces, los pesos $b_1, \ldots, b_m$ se llaman alternativamente las *pendientes* de las variables, o las *constantes multiplicativas*. Debe ser claro que, si $b_0 = 0$ y $b_1 = b_2 = \cdots = b_m = 1$, entonces la combinación lineal se reduce a una suma simple de las m variables. Nótese también que una combinación lineal con $m = 1$ (es decir, con solo una variable), es una transformación lineal de esta variable. En este sentido, los teoremas que se presentan en esta sección generalizan varios de los teoremas y propiedades presentados en este tema y en temas anteriores.

Ejemplo

Para ilustrar los teoremas de esta sección, se utilizarán los mismos datos que de la sección anterior, a saber, las variables $X_1, X_2, \ldots, X_8$, que representan las ocho calificaciones parciales en el examen diagnóstico de la Facultad de Medicina. Sin embargo, en vez de definir T como la suma simple de estas variables (como se hizo en la sección anterior), se redefine la variable T de tal forma que cada parte del examen tenga el mismo peso. (Nótese que, como el examen cuenta con 32 preguntas de matemáticas, el peso de matemáticas en la calificación total, si éste se obtiene por una simple suma, es más grande que el peso de las otras partes, que tienen menos preguntas.) En particular, se define T en esta sección como:

$$T \;=\; \frac{100}{32}X_1 \;+\; \frac{100}{16}X_2 \;+\; \frac{100}{16}X_3 \;+\; \frac{100}{16}X_4 \;+\; \frac{100}{10}X_5 \;+\; \frac{100}{10}X_6 \;+\; \frac{100}{10}X_7 \;+\; \frac{100}{10}X_8.$$

De esta forma, cada una de las ocho partes contribuye de forma igual a la variable T (ya que el peso de cada parte implica dividir el número de respuestas correctas entre el número total de preguntas y multiplicar por 100, así que el peso convierte el número de respuestas correctas a un porcentaje).

Las preguntas que se contestan en esta sección son básicamente las mismas que las que se especificaron al final de la parte introductoria de la Sección 7.2, bien entendido que la variable T ahora es una combinación lineal en vez de una suma simple de las ocho calificaciones parciales.

7.3.1 Media aritmética

Media aritmética de una combinación lineal de m variables

Si la variable T se obtiene por una combinación lineal de las variables $X_1, X_2, \ldots, X_m$, es decir $T = b_0 + \sum_{j=1}^{m} b_j X_j$ para constantes $b_0, b_1, \ldots, b_m$ cualesquiera, entonces:

$$\bar{t} \;=\; b_0 \;+\; \sum_{j=1}^{m} b_j \overline{x_j}.$$

En palabras: La media aritmética de una combinación lineal de variables es igual a la combinación lineal aplicada a las medias de las respectivas variables. Este teorema es una generalización, tanto de la Ecuación (4.12) (sobre el efecto de una transformación lineal en la media aritmética) como de la Ecuación (7.4) en la sección anterior.

Ejemplo

Para conocer la media aritmética de la variable T, definida por

$$T \;=\; \frac{100}{32}X_1 + \frac{100}{16}X_2 + \frac{100}{16}X_3 + \frac{100}{16}X_4 + \frac{100}{10}X_5 + \frac{100}{10}X_6 + \frac{100}{10}X_7 + \frac{100}{10}X_8,$$

se aplica la misma combinación lineal a las medias de las variables $X_1, \ldots, X_8$:

$$\bar{t} \;=\; \frac{100}{32}\,\overline{x_1} + \frac{100}{16}\,\overline{x_2} + \frac{100}{16}\,\overline{x_3} + \frac{100}{16}\,\overline{x_4} + \frac{100}{10}\,\overline{x_5} + \frac{100}{10}\,\overline{x_6} + \frac{100}{10}\,\overline{x_7} + \frac{100}{10}\,\overline{x_8}.$$

Dichas medias se han resumido en la tabla de la página 228. Sistituirlas resulta en:

$$\begin{aligned}\bar{t} \;&=\; \frac{100}{32}\,17.28 + \frac{100}{16}\,8.88 + \frac{100}{16}\,9.89 + \frac{100}{16}\,10.41 + \frac{100}{10}\,6.69 + \frac{100}{10}\,5.86 \\ &\quad + \frac{100}{10}\,6.65 + \frac{100}{10}\,6.25 \\ &= 490.84.\end{aligned}$$

Para la interpretación de este valor, conviene recordar que T es la suma de las ocho calificaciones convertidas a porcentajes. Entonces, se puede decir que la media de la calificación final del examen diagnóstico, definida de tal forma que cada parte tenga el mismo peso, es 490.84 sobre un máximo de 800, o bien, un poco más que 61 %.

7.3.2 Varianza

Varianza de una combinación lineal de m variables

Si la variable T se obtiene por una combinación lineal de las variables $X_1, X_2, \ldots, X_m$, es decir $T = b_0 + \sum_{j=1}^{m} b_j X_j$ para constantes $b_0, b_1, \ldots, b_m$ cualesquiera, entonces:

$$s_T^2 \;=\; \sum_{j=1}^{m} b_j^2 s_{X_j}^2 \;+\; 2 \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} b_j b_k s_{X_j X_k}. \tag{7.8}$$

Obsérvese la similitud entre esta ecuación y la Ecuación (7.5). Como se explicó en la Sección 7.2.2, la primera suma incluye las varianzas de las variables originales. Cuando se trata de una combinación lineal, donde cada variable original se multiplica por su peso b_j, se multiplica cada varianza por el cuadrado del peso asociado. Esto es una consecuencia directa de la Ecuación (4.14a), sobre el efecto de una transformación lineal (y, específicamente, el efecto de la multiplicación de una variable por una constante) en la varianza. Similarmente, las covarianzas en la segunda parte de la fórmula se multiplican por los pesos asociados con cada variable involucrada en la covarianza. Esto tiene relación directa con la cuarta propiedad de la covarianza, sobre el efecto de una transformación lineal (y la multiplicación por una constante) en la covarianza.

Ejemplo

Calculemos la varianza de la variable T, definida como una combinación lineal de las variables $X_1, X_2, \ldots, X_8$ en el ejemplo de la página 237. Como primer paso, desarrollemos la suma

$$\sum_{j=1}^{8} b_j^2 s_{X_j}^2,$$

la cual, como se explicó en el ejemplo de la página 232, incluye las varianzas en la diagonal principal de la matriz de varianzas y covarianzas. La diferencia con el caso de una simple suma es que cada varianza se multiplica por el cuadrado del peso asociado con la variable:

$$\begin{aligned}
\sum_{j=1}^{8} b_j^2 s_{X_j}^2 \;&=\; b_1^2 s_{X_1}^2 \;+\; b_2^2 s_{X_2}^2 \;+\; b_3^2 s_{X_3}^2 \;+\; b_4^2 s_{X_4}^2 \;+\; b_5^2 s_{X_5}^2 \;+\; b_6^2 s_{X_6}^2 \;+\; b_7^2 s_{X_7}^2 \;+\; b_8^2 s_{X_8}^2 \\
&=\; \left(\frac{100}{32}\right)^2 40.76 \;+\; \left(\frac{100}{16}\right)^2 8.23 \;+\; \left(\frac{100}{16}\right)^2 8.38 \;+\; \cdots \;+\; \left(\frac{100}{10}\right)^2 4.43 \\
&=\; 3203.50.
\end{aligned}$$

Como segundo paso, hay que calcular la suma

$$\sum_{j=1}^{7} \sum_{k=j+1}^{8} b_j b_k s_{X_j X_k},$$

la cual involucra las covarianzas por encima de la diagonal principal. Tal como en el ejemplo de la página 232, se suman todos los elementos de la parte superior a la diagonal principal, pero en el caso de una combinación lineal se multiplica cada covarianza primero por los pesos asociados con las dos variables involucradas. Es decir,

$$\begin{aligned}\sum_{j=1}^{7}\sum_{k=j+1}^{8} b_j b_k s_{X_jX_k} &= b_1\,b_2\,s_{X_1X_2} + b_1\,b_3\,s_{X_1X_3} + b_1\,b_4\,s_{X_1X_4}\\ &\quad + b_1\,b_5\,s_{X_1X_5} + \cdots + b_6\,b_8\,s_{X_6X_8} + b_7\,b_8\,s_{X_7X_8}\\ &= \frac{100}{32}\,\frac{100}{16}\,12.03 + \frac{100}{32}\,\frac{100}{16}\,11.12 + \frac{100}{32}\,\frac{100}{16}\,9.64\\ &\quad + \frac{100}{32}\,\frac{100}{10}\,7.18 + \cdots + \frac{100}{10}\,\frac{100}{10}\,2.94 + \frac{100}{10}\,\frac{100}{10}\,1.67\\ &= 5822.34.\end{aligned}$$

Juntando los dos resultados anteriores como prescribe la Ecuación (7.8), resulta en:

$$\begin{aligned}s_T^2 &= \sum_{j=1}^{8} b_j^2 s_{X_j}^2 + 2\sum_{j=1}^{7}\sum_{k=j+1}^{8} b_j b_k s_{X_jX_k}.\\ &= 3203.50 + 2 \times 5822.34\\ &= 14848.19.\end{aligned}$$

Para una interpretación más directa, sacamos la raiz cuadrada:

$$s_T = \sqrt{14848.19} = 121.85.$$

Es decir, la calificación total en el examen diagnóstico (definida como en la página 237, con un máximo de 800) tiene una desviación estándar de aproximadamente 122.

7.3.3 Covarianza

Propiedad distributiva de la covarianza respecto de la suma (Versión 4)

Si (a) la variable T se obtiene por una combinación lineal de las variables $X_1, X_2, \ldots, X_m$, es decir $T = b_0 + \sum_{j=1}^{m} b_j X_j$,

y (b) la variable V se obtiene por una combinación lineal de las variables $Y_1, Y_2, \ldots, Y_r$, es decir $V = a_0 + \sum_{k=1}^{r} a_k Y_k$,

donde $b_0, b_1, \ldots, b_m$ y $a_0, a_1, \ldots, a_r$ son constantes cualesquiera, entonces:

$$s_{TV} = \sum_{j=1}^{m}\sum_{k=1}^{r} b_j\,a_k\,s_{X_jY_k}. \tag{7.9}$$

Este teorema extiende la Versión 3 de la propiedad distributiva de la covarianza en la Ecuación (7.7). La diferencia entre las dos versiones son las constantes multiplicativas que se incluyen en la definición de las variables T y V y que afectan la covarianza a través de la multiplicación de las covarianzas entre las variables componentes por sus respectivos pesos. El efecto de las constantes multiplicativas sigue directamente de la Propiedad 4 de covarianzas, la cual describe el efecto de una transformación lineal de una variable (multiplicar por una constante y sumar otra constante) en la covarianza de esta variable con otras variables.

Ejemplo

Ilustremos el teorema anterior calculando la covarianza entre la variable T, definida como en el ejemplo de la página 237, y la variable V, que se defina como la nota promedio de las asignaturas en el primer año. Dichas asignaturas y las respectivas variables $Y_1, Y_2, \ldots, Y_6$ asociadas que representan la nota final se introdujeron en el ejemplo de la página 235. La nota promedio V se calcula a partir de $Y_1, Y_2, \ldots, Y_6$ como sigue:

$$\begin{aligned} V &= \frac{Y_1 + Y_2 + Y_3 + Y_4 + Y_5 + Y_6}{6} \\ &= \frac{1}{6} Y_1 + \frac{1}{6} Y_2 + \frac{1}{6} Y_3 + \frac{1}{6} Y_4 + \frac{1}{6} Y_5 + \frac{1}{6} Y_6, \end{aligned}$$

es decir, V es una combinación lineal de las variables $Y_1, Y_2, \ldots, Y_6$, donde todas las constantes multiplicativas (es decir, los pesos) son iguales a $\frac{1}{6}$ y la constante aditiva es igual a 0.

Para conocer la covarianza entre T y V, se puede aplicar la Ecuación (7.9) en el teorema anterior. Las covarianzas que entran en esta fórmula se presentaron en la matriz de covarianzas en la página 235. La Ecuación (7.9) prescribe que estas covarianzas se deben multiplicar por los respectivos pesos de las variables en las combinaciones lineales que definen T y V:

$$\begin{aligned} s_{TV} = \sum_{j=1}^{8} \sum_{k=1}^{6} b_j\, a_k\, s_{X_j Y_k} &= b_1\, a_1\, s_{X_1 Y_1} + b_1\, a_2\, s_{X_1 Y_2} + b_1\, a_3\, s_{X_1 Y_3} + \cdots + b_1\, a_6\, s_{X_1 Y_6} \\ &\quad + b_2\, a_1\, s_{X_2 Y_1} + b_2\, a_2\, s_{X_2 Y_2} + \cdots + b_2\, a_6\, s_{X_2 Y_6} \\ &\quad + b_3\, a_1\, s_{X_3 Y_1} + b_3\, a_2\, s_{X_3 Y_2} + \cdots + b_3\, a_6\, s_{X_3 Y_6} \\ &\quad + \cdots \\ &\quad + b_8\, a_1\, s_{X_8 Y_1} + b_8\, a_2\, s_{X_8 Y_2} + \cdots + b_8\, a_6\, s_{X_8 Y_6} \\ &= \frac{100}{32}\frac{1}{6} 4.89 + \frac{100}{32}\frac{1}{6} 4.50 + \frac{100}{32}\frac{1}{6} 4.14 + \cdots + \frac{100}{32}\frac{1}{6} 1.67 \\ &\quad + \frac{100}{16}\frac{1}{6} 2.06 + \frac{100}{16}\frac{1}{6} 2.04 + \cdots + \frac{100}{16}\frac{1}{6} 0.73 \\ &\quad + \frac{100}{16}\frac{1}{6} 2.02 + \frac{100}{16}\frac{1}{6} 2.11 + \cdots + \frac{100}{16}\frac{1}{6} 0.86 \\ &\quad + \cdots \\ &\quad + \frac{100}{10}\frac{1}{6} 1.33 + \frac{100}{10}\frac{1}{6} 1.31 + \cdots + \frac{100}{10}\frac{1}{6} 0.49 \\ &= 78.44. \end{aligned}$$

A partir de este resultado y las varianzas de T (en la página 240) y V (en la página 228), se puede calcular la correlación entre ambas variables:

$$\begin{aligned} r_{TV} &= \frac{s_{TV}}{s_T\, s_V} \\ &= \frac{78.44}{\sqrt{14848.19}\,\sqrt{1.08}} \\ &= .62. \end{aligned}$$

Problemas

1. Para una variable T que es la diferencia entre las variables X y Y, es decir, para

$$T = X - Y,$$

deriva fórmulas en función de los estadísticos definidos sobre la distribución de X y Y, para:

(a) la media de T

(b) la varianza de T

(c) la covarianza de T con cualquier otra variable V.

2. Demuestra las propiedades presentadas en las Secciones 7.2 y 7.3.

3. En el contexto de la lucha contra el SIDA, una organización no gubernamental en Sudáfrica hace una intervención en una escuela secundaria de 500 alumnos: durante 1 semana, los alumnos reciben un curso intensivo con información sobre la enfermedad (vías de transmisión, medidas para la prevención, etc.). Para evaluar el impacto de la intervención, se ha construido una prueba que mide conocimientos sobre el SIDA y que se aplica a cada alumno al principio y al final de (la semana de) la intervención. Se definen variables *Pre* y *Post* para representar las calificaciones en esta prueba. Las medias y desviaciones estándares en estas variables, tal como la correlación entre ellas, se presentan a continuación:

	Pre	*Post*
Media aritmética	8.5	13.1
Desviación estándar	2.9	3.5
Correlación entre *Pre* y *Post*: 0.70		

A partir de las variables *Pre* y *Post*, se define una nueva variable *Impacto*:

$$\textit{Impacto} \;=\; \textit{Post} - \textit{Pre}.$$

A partir de los datos anteriores:

(a) Halla la media y la desviación estándar de *Impacto*.

(b) Halla el coeficiente de correlación entre (i) *Impacto* y *Pre* y entre (ii) *Impacto* y *Post*.

(c) Interpreta los resultados para los estadísticos hallados

4. Según el "modelo clásico" de la psicometría, la puntuación que se obtiene al aplicar un test (por ejemplo, un examen o un cuestionario) a una persona i (representada por x_i) es una suma de dos componentes: (a) la puntuación verdadera de esta persona en el test (v_i) y (b) un error aleatorio (e_i) (el cual explica por qué el test no es perfectamente confiable, es decir, por qué la puntuación verdadera v_i generalmente no coincide perfectamente con la puntuación observada x_i). En símbolos, el modelo supone para cualquier persona i al que se administra el test:

$$x_i = v_i + e_i.$$

Supongamos que se aplica el test a 1000 personas y que el modelo se cumple para cada una. Además, supongamos que, en esta muestra, (i) la media aritmética de E es igual a 0 y (ii) las variables V y E son independientes (los cuales son dos supuestos adicionales comunes en el modelo clásico de la psicometría). Entonces,

(a) ¿Cuál es la relación entre las medias $\overline{x}$ y $\overline{v}$?

(b) ¿Cuál es la varianza de X en función de las varianzas de V y E?

(c) Demuestra que, bajo los supuestos mencionados,

$$r_{XV} = \frac{s_V}{s_X}.$$

Como paso siguiente, consideremos el caso de que se apliquen *dos tests* a la misma muestra de personas y supongamos que el modelo clásico de la psicometría se cumple para ambos. Esto quiere decir que, para las variables X_1 y X_2 que se definen a partir de las puntuaciones observadas en los tests respectivos, se cumplen las siguiente ecuaciones:

$$\begin{aligned} x_{i1} &= v_{i1} + e_{i1} \\ x_{i2} &= v_{i2} + e_{i2} \end{aligned}$$

Un concepto importante en la psicometría clásica es el de *tests paralelos*. Dos tests son paralelos si (i) para cada persona del grupo, las puntuaciones verdaderas en los tests son iguales (es decir, $\forall i : v_{i1} = v_{i2}$) y (ii) si las varianzas del error de ambos tests son iguales (es decir, $s^2_{E_1} = s^2_{E_2}$).

(d) Suponiendo que las puntuaciones verdaderas (V_1 y V_2) son independientes de los errores (E_1 y E_2) y que los errores son independientes entre sí (es decir, E_1 es independiente de E_2), demuestra que, si X_1 y X_2 son tests paralelos:

$$r_{X_1X_2} = \frac{s^2_{V_1}}{s^2_{X_1}} = \frac{s^2_{V_2}}{s^2_{X_2}}.$$

Tema 8
Predicción óptima y regresión lineal (descriptiva)

Índice

En los temas anteriores se estudiaron distribuciones bivariadas y, en específico, la característica de covariación entre dos variables. El concepto de covariación está muy relacionado con la problemática que se aborda en el tema actual: predicción o regresión. Si existe una asociación entre dos variables, es natural preguntarse, como paso siguiente, si se puede utilizar el conocimiento del valor que tiene una observación en una de las variables para predecir su valor en la otra.

Se destacaron diferentes formas de covariación entre dos variables (véase, por ejemplo, los diagramas de dispersión en la página 161). A la luz del tipo de covariación, se puede decidir sobre el tipo de predicción. En este tema se discuten dos tipos de predicción: (a) Predicción a partir de la asociación general que existe entre dos variables y (b) predicción lineal, la cual—ciertamente—explota la relación lineal entre las variables involucradas.

Los métodos descritos en este tema generalmente consideran el caso de variables cuantitativas, aunque el método presentado en la primera sección serviría también para predecir a partir de una variable cualitativa. Ejemplos típicos en este tema incluyen la predicción de la presión arterial con base en el índice de masa corporal y la predicción de la nota final promedio en las asignaturas del primer año de la carrera de Médico Cirujano con base en las calificaciones en una prueba diagnóstica que se aplica al inicio del primer semestre. Generalmente, se representa la variable *predictora* (cuyos valores se *utilizan* para predecir) por X y la variable *criterio* (cuyos valores se *estiman*) por Y.

Cabe reconocer que, en un contexto de predicción, como en el ejemplo de la predicción del promedio de la notas finales, comúnmente el objetivo final es realizar predicciones para observaciones *nuevas* (es decir, para estudiantes nuevos, por ejemplo, de la siguiente generación), para las cuales se conoce únicamente la puntuación en la variable predictora. Tales predicciones podrían motivar decisiones sobre la admisión o el rechazo a la carrera o sobre la organización de cursos remediales para estudiantes que mostraron falta de conocimientos básicos en el examen diagnóstico. Sin embargo, tal objetivo implica hacer inferencias sobre la variable criterio que superan los datos bivariados originalmente observados; por consiguiente, esta problemática pertenece al área de la estadística inferencial. En este tema, se aproxima el problema de predicción dentro del marco de la estadística descriptiva; es decir, los métodos abordados sirven para describir la asociación (general o lineal) que existe (o no) entre las variables observadas. También es posible considerar la problemática de predicción dentro del marco de la estadística inferencial. Sin embargo, la elaboración de la regresión lineal inferencial se encuentra fuera del alcance de este curso.

La consideración anterior acarrea un uso del término "predicción" algo peculiar, ya que en este tema se "predicen" valores en la variable criterio ya observados y, por lo tanto, conocidos. De forma alternativa se podría hablar de "la asociación de Y con X" (aunque esta expresión no es muy precisa sobre cuál variable es el criterio y cuál la predictora). En la estadística se utiliza a menudo la expresión "regresión de Y sobre X". A pesar de la peculiaridad de hablar sobre "predicciones en una variable cuyos valores ya se observaron", se seguirá utilizando el término "predicción" en este tema; es importante tener claro que predicción en este caso es nada más que la descripción (a través de una función matemática, véase el inicio de la primera sección) de la relación que existe entre las variables estudiadas.

El tema consiste en tres secciones. La primera inicia con una explicación de qué es predecir matemáticamente y distingue entre los casos para los cuales una predicción exacta (o perfecta) es posible y los casos en los que es imposible. Para los casos de que no sea posible una predicción exacta, se derivará un criterio que permita comparar la calidad de diferentes predicciones y se introducirá una metodología general para realizar predicciones óptimas bajo este criterio. Desde la segunda sección se añade un requisito adicional a la predicción: En particular, se exige que la predicción—o regresión—sea lineal. Se explicará cómo conseguir predicciones óptimas que respetan esta restricción y, a continuación, se discute la relación entre regresión lineal y el coeficiente de correlación, el cual, como se explicó cuando se introducía en la Sección 6.2.2, cuantifica que tan bueno se puede describir la relación entre dos variables por una función lineal. La última sección da una introducción al método de *regresión lineal múltiple*, el cual generaliza las ideas de la Sección 8.2 para predecir una variable criterio a partir de una *combinación lineal* de dos o más variables. Se explican también algunos conceptos definidos dentro de este marco, como la correlación múltiple, la correlación parcial y semiparcial, y el problema de multicolinealidad en regresión múltiple.

8.1 Predicción óptima: Conceptos básicos

8.1.1 Predicción perfecta

Consideremos un conjunto de datos bivariados (X, Y) en n observaciones y supongamos que se desea predecir la variable Y a partir de la variable X, por ejemplo, la predicción del promedio de la nota final en las asignaturas del primer año de la carrera a partir de la calificación global en el examen diagnóstico aplicado al inicio del primer semestre. Esto quiere decir que se asigna, a través de alguna regla o algún procedimiento, a cada valor de la variable X un valor pronosticado en la variable Y.

Se anotará el valor pronosticado para la observación i como $\hat{y}_i$ y la variable correspondiente $\hat{Y}$. Es importante distinguir entre Y (la variable con valores observados) y $\hat{Y}$ (la variable con los valores pronosticados).

"Asignar a cada valor de una variable X un valor en otra variable $\hat{Y}$ a través de alguna regla" se traduce en lenguaje matemático a "aplicar una función f que tiene como dominio los valores de X y codominio los valores de $\hat{Y}$".

Predicción de Y a partir de X

Para una variable X que asume valores $x_1, x_2, \ldots, x_m$ y otra variable Y cuyos valores de Y son elementos de $\mathbb{R}$, la *predicción de Y a partir de X* se define como la aplicación de una función f que asigna a cada valor de X un valor en la variable $\hat{Y}$:

$$\begin{aligned} f : V &\to \mathbb{R} \\ x_j &\mapsto f(x_j) = \hat{y}_j, \end{aligned}$$

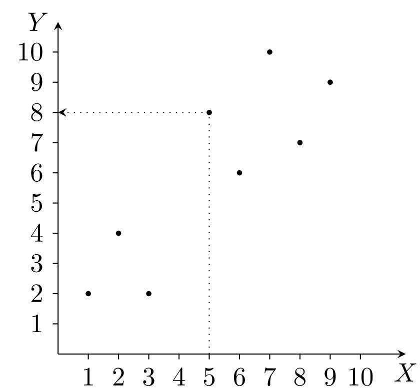
donde $V = \{x_1, x_2, \ldots, x_m\}$.

Sinónimo: *Regresión de Y sobre X*

Ejemplo

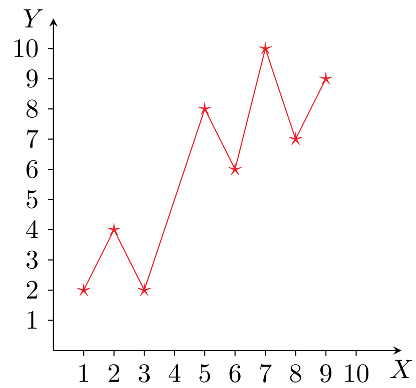
Para ocho estudiantes de la carrera de Médico Cirujano se dispone de sus notas en el primer examen departamental (variable X) y en el segundo examen departamental (variable Y) de la asignatura de *Bioquímica*. Los datos son los siguientes:

Iniciales estudiante	X	Y
M.R.E.	7	10
L.S.S.	2	4
M.A.M.	8	7
A.R.A.	5	8
P.E.S.	9	9
A.O.L.	1	2
T.O.G.	6	6
G.P.S.	3	2



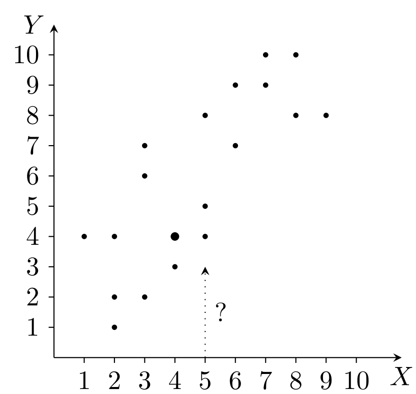
A partir de estos datos, si uno recibe la información de que un estudiante tiene en el primer departamental una nota de 5, ¿qué valor le asignaría para el segundo departamental? Puesto que para el valor de 5 en X solo hay un valor correspondiente para Y en estos datos, no hay incertidumbre sobre el valor en la variable Y que tiene una persona i con $x_i = 5$. También para los otros valores en X, se sabe con certeza cuál sería el valor en Y. En otras palabras, es posible realizar una *predicción perfecta* en este ejemplo. La siguiente tabla representa la función f de predicción que asigna a cada uno de los $m = 8$ valores de x_j el valor pronosticado $\hat{y}_j$; la figura da una representación gráfica de la misma función (las estrellas en la gráfica representan los valores pronósticados para cada valor de X):

j	X	$\hat{Y}$
1	1	2
2	2	4
3	3	2
4	5	8
5	6	6
6	7	10
7	8	7
8	9	9



Consideremos ahora las mismas variables en una muestra más amplia de 20 estudiantes, con los siguientes datos:

Iniciales estudiante	X	Y	Iniciales estudiante	X	Y
N.H.V.	4	3	J.S.G.	3	6
F.P.A.	3	7	M.G.L.	6	7
I.S.B.	5	8	R.G.R.	7	9
M.P.R.	8	10	O.S.B.	9	8
C.B.B.	2	4	S.J.M.	4	4
V.E.G.	2	2	M.T.R.	8	8
B.G.M.	1	4	B.S.R.	4	4
J.M.C.	5	5	C.G.G.	7	10
J.P.H.	3	2	C.M.C.	5	4
J.Y.S.	6	9	A.G.E.	2	1



Ahora, considerando la misma pregunta que en el ejemplo anterior "¿qué valor se le asignaría para el segundo departamental (variable Y) a una persona de la cual se sabe que en el primer departamental tiene una nota de 5?", es claro que, cualquiera que sea el valor que se prediga, dicha predicción será acompañada con un grado de incertidumbre. Como la asociación entre ambas variables en esta muestra no es perfecta, ya no hay certeza de que la predicción coincida siempre con el valor real observado; es decir, en este ejemplo una predicción perfecta (para todos los valores en X) no es posible.

El ejemplo anterior muestra que una predicción perfecta es posible si, y solo si, en los datos hay un único valor de la variable criterio asociada con cada valor de la variable predictora. La pregunta que surge en casos como el segundo ejemplo es: ¿Cómo predecir cuando una predicción perfecta no es posible?

8.1.2 Bondad de predicción

Se observó en el ejemplo anterior que, para la mayoría de los valores observados en la variable predictora, resulta imposible garantizar una predicción exacta. Esto no quiere decir que todas las predicciones son igualmente buenas. Supuestamente, todo el mundo estará de acuerdo que, para alguien con una nota de $x_i = 5$ en el primer departamental, pronosticar un valor de $\hat{y}_i = 6$ es "mejor" que predecir $\hat{y}_i = 0$. Por otro lado, ¿$\hat{y}_i = 6$ sería mejor que $\hat{y}_i = 5$ en este caso? La conclusión de estas reflexiones es que hace falta un criterio claro y objetivo para evaluar la bondad de diferentes predicciones. En esta sección se desarrolla tal criterio.

Ejemplo

Considérese el siguiente escenario: Ocho alumnos en el último año del bachillerato de la misma escuela deciden iniciar los estudios de la carrera de Médico Cirujano en la Facultad de Medicina. Tres de sus maestros, Nemo (su maestro de matemáticas), Paco (su maestro de física) y Pepe (su maestro de historia) organizan entre sí una apuesta que consiste en que realizan un pronóstico para cada uno de los ocho estudiantes sobre la nota final promedio que obtengan en las asignaturas del primer año de la carrera. Los pronósticos de los tres maestros se representan en las variables $\hat{Y}_{\text{Nemo}}$, $\hat{Y}_{\text{Paco}}$ y $\hat{Y}_{\text{Pepe}}$, respectivamente. Al final del año, los tres maestros recopilan las notas finales de cada estudiante y cálculan el promedio que los alumnos realmente obtuvieron (Y). Los datos se presentan en la siguiente tabla:

		Predicciones		
	Nota observada Y	Maestro Nemo $\hat{Y}_{\text{Nemo}}$	Maestro Paco $\hat{Y}_{\text{Paco}}$	Maestro Pepe $\hat{Y}_{\text{Pepe}}$
Helena	7.6	7	8	8
Juan Carlos	3.4	5	5	4
María	8.5	10	8	8
Paulina	6.1	6	5	9
Fernando	9.4	7	6	6
Luis	7.3	9	7	6
Álvaro	6.4	8	7	7
Carmen	7.3	8	6	8

A partir de estos datos, ¿cómo se puede decidir quién de los tres, Nemo, Paco o Pepe, realizó la mejor predicción?

Para contestar esta pregunta, es de sentido común calcular, para cada alumno, la diferencia entre la nota pronosticada y la nota efectivamente obtenida (separadamente para cada maestro). La idea es que, si la nota pronósticada se diferiese mucho de la nota final observada, menos adecuada sería la predicción. En otras palabras, para evaluar la bondad de predicción, el punto de partida es el *error* que se ha producido en cada observación:

$$\text{error}_i = \hat{y}_i - y_i.$$

En la siguiente tabla (columnas error_i) se presentan los errores que cometieron los tres maestros al predecir la nota final para cada estudiante.

	Bondad de predicción para Maestro Nemo			Bondad de predicción para Maestro Paco			Bondad de predicción para Maestro Pepe		
	error_i	$\lvert\text{error}_i\rvert$	$(\text{error}_i)^2$	error_i	$\lvert\text{error}_i\rvert$	$(\text{error}_i)^2$	error_i	$\lvert\text{error}_i\rvert$	$(\text{error}_i)^2$
Helena	−0.6	0.6	0.36	+0.4	0.4	0.16	+0.4	0.4	0.16
Juan Carlos	+1.6	1.6	2.56	+1.6	1.6	2.56	+0.6	0.6	0.36
María	+1.5	1.5	2.25	−0.5	0.5	0.25	−0.5	0.5	0.25
Paulina	−0.1	0.1	0.01	−1.1	1.1	1.21	+3.9	3.9	15.21
Fernando	−2.4	2.4	5.76	−3.4	3.4	11.56	−4.4	4.4	19.36
Luis	+1.7	1.7	2.89	−0.3	0.3	0.09	−1.3	1.3	1.69
Álvaro	+1.6	1.6	2.56	+0.6	0.6	0.36	+0.6	0.6	0.36
Carmen	+0.7	0.7	0.49	−1.3	1.3	1.69	+0.7	0.7	0.49
Suma:	+4.0	10.2	16.88	−4.0	9.2	17.88	0.0	12.4	37.88

La última fila de la tabla presenta la suma de los errores. Para Maestro Nemo la suma de errores es $+4.0$, para Maestro Paco es -4.0, y para Maestro Pepe es 0.0. Es decir, la suma de errores es más baja para Maestro Paco, pero ¿significa que realizó las mejores predicciones? Debe ser claro que, al interpretar los errores de predicción, una subestimación (cuando $\text{error}_i < 0$) se debe "contar" de la misma forma que una sobreestimación ($\text{error}_i > 0$); para alguien que obtiene una nota efectiva de 7, una predicción de 8 y otra de 6 implican el mismo tamaño del error. Es decir, al evaluar la bondad de predicción, hay que deshacerse del signo del error de predicción.

Mirándolo desde otra perspectiva: Los errores del Maestro Pepe suman a 0.0. ¿Quiere decir que sus predicciones son perfectas o que realizó mejores predicciones que los dos otros maestros? La respuesta a ambas preguntas es obviamente *no*: La suma de errores puede ser 0, aun cuando hay errores grandes, puesto que en la suma las sobreestimaciones se compensan por las subestimaciones. Respecto del Maestro Pepe, por ejemplo, se observa que los errores de predicción para Paulina y Fernando son relativamente grandes; sin embargo, en la suma, los dos errores prácticamente se neutralizan ya que sobreestimó a Paulina y subestimó a Fernando.

Las consideraciones anteriores implican que la operación "deshacerse del signo del error" (es decir, "tomar el valor absoluto del error") se debe realizar de forma separada para cada estudiante. Es decir, un aceptable índice para evaluar la bondad de predicción en n observaciones puede ser

$$\sum_{i=1}^{n} |\hat{y}_i - y_i|. \tag{8.1}$$

Un valor de 0 en este índice reflejaría un caso de predicción perfecta: El índice solo puede ser 0, si para cada observación i, se tiene $\hat{y}_i = y_i$. En general, valores más bajos en el índice de la Ecuación (8.1) apuntan a mejores predicciones. En el ejemplo de los tres maestros, el Maestro Paco realizó la mejor predicción bajo este criterio.

No obstante, Maestro Nemo, quien en su carrera como matemático tuvo un curso de estadística, objeta que no es muy común tomar el valor absoluto de los errores, sino que normalmente se llevan los errores al cuadrado para deshacerse de su signo. Argumenta que el índice

$$\sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \tag{8.2}$$

cuantificaría de mejor forma la calidad de las predicciones. Igual al índice en la Ecuación (8.1), esta alternativa tiene un valor de 0 en caso de predicción perfecta y valores más bajos generalmente apuntan a una mejor predicción. Sin embargo, los maestros, al calcular el valor en este índice, se dan cuenta de que, bajo el criterio en la Ecuación (8.2), ya no es el Maestro Paco quien realizó la mejor predicción, sino que este criterio apunta al Maestro Nemo como mejor pronosticador.

El ejemplo anterior enseña que el criterio para evaluar la bondad de predicción es, hasta cierto punto, arbitrario: A pesar de que tanto el índice (8.1) como el índice (8.2) son criterios válidos para evaluar la bondad de predicción, no necesariamente llevan a la misma conclusión. Por otro lado, hay que admitir que el ejemplo anterior fue especialmente construido para que estos dos índices resultasen en conclusiones opuestas. En aplicaciones prácticas, ambos índices casi siempre sí llevan a la misma conclusión. En la estadística, lo más común para comparar predicciones es el índice de la Ecuación (8.2). Cuando se divide por el número de observaciones, se conoce como el *error cuadrático medio*:

Error cuadrático medio

Si n entidades (personas, objetos, etc.)

(a) tienen valores observados $y_1, y_2, \ldots, y_n$ en alguna variable Y

y (b) recibieron, a través de algún método de predicción, valores pronosticados $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$, respectivamente,

entonces el *error cuadrático medio* se define como

$$\mathrm{ECM}(Y, \hat{Y}) = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}.$$

Como resultado de un acuerdo, se evaluará la bondad de predicción en el resto de este tema utilizando el criterio ECM. En la literatura, este criterio se conoce también como el *principio* o el *criterio de mínimos cuadrados*.

8.1.3 Predicción óptima

Con el criterio de bondad de predicción que se acaba de definir en la sección anterior, el objetivo es hacer la predicción de tal forma que el error cuadrático medio sea mínimo. Cuando este objetivo se realiza para un contexto de predicción concreto, se dice que la predicción es *óptima*. En esta sección se da respuesta a la pregunta: ¿Cuál procedimiento se debe seguir para que la predicción en la variable Y a partir de la variable X sea óptima?

Al predecir Y a partir de X, se representa el error cuadrático medio con el símbolo $s^2_{Y.X}$. Es decir,

$$s^2_{Y.X} = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}, \tag{8.3a}$$

lo cual, a la luz de la definición en la página 247, implica que:

$$s^2_{Y.X} = \frac{\sum_{i=1}^{n} [f(x_i) - y_i]^2}{n}. \tag{8.3b}$$

Ejemplo

Al final de ejemplo de la página 248 se concluyó que una predicción perfecta no era posible para los datos presentados. Por eso, se cambia de objetivo y se busca una predicción óptima. Debe ser claro que minimizar el error cuadrático medio es equivalente a minimizar la expresión en la Ecuación (8.2); dividir esta suma entre la constante n, no cambiará las conclusiones respecto de la bondad de predicción.

Para minimizar la suma de error cuadráticos

$$\sum_{i=1}^{20} [f(x_i) - y_i]^2 \tag{8.4}$$

para los datos de la página 248, conviene reordenar las 20 observaciones según su valor en X, en consideración de que las personas con el mismo valor x_j en X, recibirán la misma predicción $\hat{y}_j = f(x_j)$ en $\hat{Y}$. (Recuérdese que predecir implica especificar una función f, por lo cual se debe asociar una sola predicción al valor x_j.) En la siguiente tabla, se presentan los datos reorganizados (aún no se considere la última columna):

Iniciales estudiante	X	Y	$\hat{Y} = f(X)$
B.G.M.	1	4	4.00
A.G.E.		1	
V.E.G.	2	2	$\frac{1+2+4}{3} = 2.33$
C.B.B.		4	
J.P.H.		2	
J.S.G.	3	6	$\frac{2+6+7}{3} = 5.00$
F.P.A.		7	
N.H.V.		3	
S.J.M.	4	4	$\frac{3+4+4}{3} = 3.67$
B.S.R.		4	
C.M.C.		4	
J.M.C.	5	5	$\frac{4+5+8}{3} = 5.67$
I.S.B.		8	
M.G.L.	6	7	$\frac{7+9}{2} = 8.00$
J.Y.S.		9	
R.G.R.	7	9	$\frac{9+10}{2} = 9.50$
C.G.G.		10	
M.T.R.	8	8	$\frac{8+10}{2} = 9.00$
M.P.R.		10	
O.S.B.	9	8	8.00

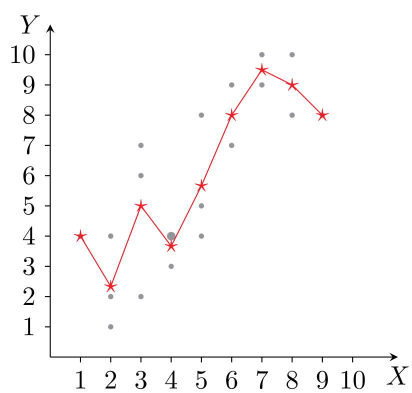
Si se desarrolla la suma en la Ecuación (8.4), se obtiene:

$$
\begin{aligned}
\sum_{i=1}^{20} [f(x_i) - y_i]^2 \;=\;& [f(1) - 4]^2 \\
&+ [f(2) - 1]^2 + [f(2) - 2]^2 + [f(2) - 4]^2 \\
&+ [f(3) - 2]^2 + [f(3) - 6]^2 + [f(3) - 7]^2 \\
&+ [f(4) - 3]^2 + [f(4) - 4]^2 + [f(4) - 4]^2 \\
&+ [f(5) - 4]^2 + [f(5) - 5]^2 + [f(5) - 8]^2 \\
&+ [f(6) - 7]^2 + [f(6) - 9]^2 \\
&+ [f(7) - 9]^2 + [f(7) - 10]^2 \\
&+ [f(8) - 8]^2 + [f(8) - 10]^2 \\
&+ [f(9) - 8]^2
\end{aligned}
$$

Se logra minimizar esta suma, si se minimiza *por separado* la suma en cada ringlón de la expresión anterior, eligiendo el valor adecuado para $f(1), f(2), \ldots, f(9)$, respectivamente. Esto significa, por ejemplo, que se debe elegir el valor $f(2)$ de tal forma que minimice la suma

$$[f(2) - 1]^2 + [f(2) - 2]^2 + [f(2) - 4]^2 .$$

Debe ser claro que $f(2)$ en esta expresión es una constante. Gracias a la segunda propiedad de la media (p. 71), se sabe que la constante que minimiza una suma de cuadrados como en la expresión anterior, es la media aritmética. Es decir, la expresión se minimiza eligiendo para $f(2)$ la media aritmética de los valores 1, 2 y 4. Como regla general, una predicción es óptima cuando se asocia a cada valor x_j la media aritmética de los valores en Y correspondientes a las observaciones cuyo valor en X es x_j. Como se explicó en la Sección 5.4, p. 155, tal media es una media condicional. La última columna de la tabla anterior, muestra las predicciones óptimas para el ejemplo. En la siguiente figura se representa la función de predicción gráficamente en color rojo.



X	$\hat{Y}$
1	4
2	2.33
3	5
4	3.67
5	5.67
6	8
7	9.5
8	9
9	8

Cabe mencionar que el error cuadrático medio asociado con esta predicción es igual a:

$$\begin{aligned} s^2_{Y.X} &= \frac{\sum_{i=1}^{20} (\hat{y}_i - y_i)^2}{20} \\ &= \frac{(4-4)^2 + (2.33-1)^2 + (2.33-2)^2 + \ldots + (9-10)^2 + (8-8)^2}{20} \\ &= \frac{32.5}{20} \\ &= 1.625. \end{aligned}$$

Por ser el error cuadrático medio asociado con la predicción óptima, es el mínimo posible.

El ejemplo anterior lleva directamente a la siguiente propiedad.

Función de predicción óptima

Si a partir de un conjunto de observaciones que tienen valores en las variables X y Y se realiza una predicción de Y sobre X, es decir, con

$$\hat{y}_i = f(x_i)$$

para todas las observaciones, entonces la función f de *predicción óptima*—que minimiza $s^2_{Y.X}$ en la Ecuación (8.3)—se da por:

$$f(x_i) = \overline{y} \,|\, (X = x_i).$$

Prueba

Como se argumentó en el ejemplo anterior, esta propiedad es una consecuencia directa de la segunda propiedad de la media (p. 71).

8.2 Regresión lineal simple

8.2.1 Predicción lineal óptima

Como se puede observar en las figuras de los ejemplos anteriores (páginas 248 y 253), la función de predicción f que minimiza el error cuadrático medio, en general tiene una forma bastante irregular. Sin embargo, muchas veces se prefieren funciones simples, por la facilidad de la interpretación. Es decir, se añade un requisito adicional al problema de predicción: Además de que la función f provea una buena predicción de Y a partir de X, se exige que f tenga una forma especial. La forma de función que más se elige para estos fines es la lineal.

Esto implica que la función f en

$$\hat{y}_i = f(x_i)$$

sea una función lineal, lo que quiere decir que existan constantes a y b de tal forma que:

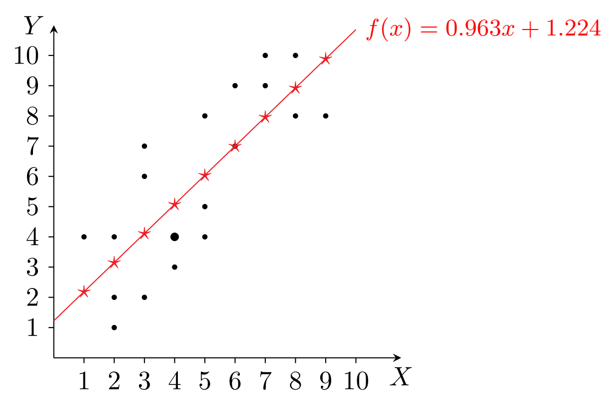
$$\hat{y}_i = ax_i + b.$$

La última ecuación se denomina *ecuación de regresión lineal de Y sobre X*; a y b son las *constantes* o los *coeficientes* de regresión correspondientes. La recta definida por a y b es la *recta de regresión*. El objetivo es eligir entre todos los valores posibles para a y b (es decir, entre todos los números reales), aquellos que minimicen el error cuadrático medio. En este caso, la predicción es *lineal óptima*.

Como ya se mencionó, la restricción adicional facilita interpretar la función de predicción. Sin embargo, esta ventaja tiene un precio: Generalmente, el error cuadrático medio en el caso de una predicción lineal óptima es mayor en comparación con el error cuadrático medio asociado con la predicción óptima. Pues, en el primer caso la restricción de linealidad exige que la función f tenga una forma muy particular, mientras que, en el caso de predicción óptima, f puede ser cualquier función (incluyendo, si acaso, una función lineal). Por lo tanto, el error cuadrático medio asociado con predicción lineal óptima nunca puede ser menor que el mismo en caso de predicción óptima sin esta restricción. (En realidad, ya se sabe de la exposición en la sección anterior que cualquier función de predicción f tendrá el error cuadrático medio mayor que o igual al error cuadrático medio de predicción óptima, ya que éste es el más bajo posible.)

Ejemplo

Volvamos a considerar el ejemplo de la página 248, donde se presentaron las notas en el primer (X) y segundo (Y) departamental de 20 estudiantes de Bioquímica. La siguiente figura representa la recta de regresión lineal óptima para la predicción de Y sobre X.



X	$\hat{Y} = 0.963X + 1.224$
1	2.187
2	3.150
3	4.113
4	5.076
5	6.039
6	7.002
7	7.965
8	8.928
9	9.891

Al concluir este ejemplo, se explicará como se obtuvieron los valores para las constantes $a = 0.963$ y $b = 1.224$. Por el momento, fíjese en los siguientes dos resultados:

1. La predicción es lineal. Todas las predicciones (representadas por las estrellas en la figura anterior) se encuentran en una recta. Nótese que es más fácil interpretar esta función de predicción (en comparación con la de la página 253), gracias a que se trata de una relación lineal directa entre ambas variables: Si el valor en X es bajo/alto, se predice un valor bajo/alto en Y. En específico, para cada punto adicional en la calificación del primer departamental, se suma 0.963 puntos en la predicción para el segundo departamental.

2. Por otro lado, el error cuadrático medio de esta predicción es igual a:

$$
\begin{aligned}
s^2_{Y.X} &= \frac{\sum_{i=1}^{20} (\hat{y}_i - y_i)^2}{20} \\
&= \frac{(2.187-4)^2 \,+\, (3.150-1)^2 \,+\, (3.150-2)^2 \,+\, \ldots \,+\, (8.928-10)^2 \,+\, (9.891-8)^2}{20} \\
&= \frac{56.813}{20} \\
&= 2.841.
\end{aligned}
$$

Compárese este valor con el error cuadrático medio asociado con predicción óptima sin restricciones (en la página 254). Por la restricción adicional, la interpretación volvió más sencilla a costo de que el error de predicción en esta muestra subió ligeramente.

La pregunta clave en el ejemplo anterior es cómo se han hallado los valores $a = 0.963$ y $b = 1.224$. En general: a partir de datos bivariados de n observaciones en las variables X y Y, ¿cuáles son los valores que se deben asignar a las constantes de la regresión para que la predicción lineal de Y sobre X sea óptima? El siguiente teorema nos da la respuesta:

Ecuación de regresión para predicción lineal óptima

Si a partir de un conjunto de observaciones que tienen valores en las variables X y Y, se realiza una predicción lineal de Y sobre X, es decir, donde

$$\hat{y}_i \;= a\,x_i \,+\, b$$

para todas las observaciones, los valores de a y b que llevan a *predicción lineal óptima* (aquellos que minimizan el error cuadrático medio) se dan por:

$$a \;= \frac{s_{XY}}{s^2_X}. \tag{8.5a}$$

$$b \;= \overline{y} \,-\, a\,\overline{x} \tag{8.5b}$$

Antes de ilustrarlo con unos ejemplos, se presenta la demostración del teorema anterior.

Dereviación de las constantes de regresión en predicción lineal óptima

Los valores óptimos para a y b son aquellos que minimizan el error cuadrático medio en la Ecuación (8.3) o, de forma equivalente, el numerador de esta expresión. Dado que $\hat{y}_i = ax_i + b$ para todas las observaciones, el númerador toma la forma:

$$\sum_{i=1}^{n} [(ax_i + b) - y_i]^2 .$$

Como paso siguiente, considérese la expresión anterior como una función de a y b; es decir, se define la función h

$$h(a, b) \; = \; \sum_{i=1}^{n} [(ax_i + b) - y_i]^2 .$$

A continuación, se busca el mínimo de esta función. Es decir, se evalúa la función h para diferentes valores de a y b y se busca cuál par de valores (a, b) minimiza la función.

El cálculo diferencial ofrece las herramientas para encontrar el mínimo de una función. El procedimiento consiste en tres pasos:

(a) calcular las derivadas parciales de h, tanto respecto de a como de b

(b) igualar ambas derivadas parciales a cero

(c) resolver el sistema de ecuaciones resultante para a y b.

Aplicándo este procedimiento a la función h anteriormente definida, se obtiene

(a) Las derivadas parciales respecto de a y b tienen la siguiente forma:

$$\begin{cases} \dfrac{\partial h(a,b)}{\partial a} & = \; 2a\displaystyle\sum_{i=1}^{n} x_i^2 + 2b\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} x_i y_i \\ \dfrac{\partial h(a,b)}{\partial b} & = \; 2nb + 2a\displaystyle\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} y_i \end{cases}$$

(b) Igualar las derivadas a cero y reorganizar algebraicamente resulta en el siguiente sistema de ecuaciones:

$$\begin{cases} a\displaystyle\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \\ nb + a\displaystyle\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \end{cases}$$

(c) Los valores a y b que satisfacen este sistema de ecuaciones se obtienen más fácilmente primero despejando b en la segunda ecuación:

$$b \; = \; \frac{\sum_{i=1}^{n} y_i}{n} \; - \; a\,\frac{\sum_{i=1}^{n} x_i}{n} \tag{8.6a}$$

y, a continuación, en la primera ecuación reemplazando b por esta última expresión:

$$a\sum_{i=1}^{n} x_i^2 \; + \; \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n} \; - \; a\,\frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} \; = \; \sum_{i=1}^{n} x_i y_i .$$

Resolviendo para a resulta en:

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{8.6b}$$

Lo anterior implica que, si se asignan a b y a los valores obtenidos por las Ecuaciones (8.6a) y (8.6b) respectivamente, entonces la función h—o bien, el error cuadrático medio—es mínimo. (El cálculo diferencial estipula un requisito adicional para esta conclusión, a saber, las segundas derivadas de $h(a, b)$ respecto de a y b deben ser positivas. Se puede verificar que esto es efectivamente el caso.)

Falta demostrar que las Ecuaciones (8.6a) y (8.6b) son equivalentes a las Ecuaciones (8.5b) y (8.5a) mencionadas en el teorema. Para la constante b, la equivalencia de (8.5b) y (8.6a) sigue directamente de la definición de la media aritmética de X y Y:

$$\begin{aligned} b &= \frac{\sum_{i=1}^{n} y_i}{n} - a\frac{\sum_{i=1}^{n} x_i}{n} \\ &= \overline{y} - a\overline{x}. \end{aligned}$$

La equivalencia de las expresiones (8.5a) y (8.6b) para la pendiente a se demuestra como sigue:

$$\begin{aligned} a &= \frac{s_{XY}}{s_X^2} \\ &= \frac{\frac{1}{n}\sum_{i=1}^{n}[(x_i - \overline{x})(y_i - \overline{y})]}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \\ &= \frac{\sum_{i=1}^{n}(x_i y_i - x_i\overline{y} - \overline{x}y_i + \overline{x}\,\overline{y})}{\sum_{i=1}^{n}(x_i^2 + \overline{x}^2 - 2x_i\overline{x})} \\ &= \frac{\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i\overline{y} - \overline{x}\sum_{i=1}^{n} y_i + n\,\overline{x}\,\overline{y}}{\sum_{i=1}^{n} x_i^2 + n\overline{x}^2 - 2\overline{x}\sum_{i=1}^{n} x_i}, \end{aligned}$$

lo cual, después de multiplicar cada término del numerador y el denominador por n, lleva a:

$$\begin{aligned} a &= \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 + \left(\sum_{i=1}^{n} x_i\right)^2 - 2\left(\sum_{i=1}^{n} x_i\right)^2} \\ &= \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}. \end{aligned}$$

■

Ejemplos

En el ejemplo de la página 255, se indicó que los valores óptimos para las constantes de regresión eran:

$$a = 0.963$$
$$b = 1.224.$$

A partir de los valores que tienen las 20 observaciones en las variables estudiadas X y Y (véase la tabla de la página 248), se puede verificar los siguientes estadísticos para ambas variables:

$$\overline{x} = 4.70 \qquad \overline{y} = 5.75$$
$$s_X^2 = 5.01 \qquad s_Y^2 = 7.49$$
$$s_{XY} = 4.83$$

Aplicando las fórmulas en la Ecuación (8.5), se verifica que

$$a = \frac{s_{XY}}{s_X^2} = \frac{4.83}{5.01} = 0.963$$
$$b = \overline{y} - a\,\overline{x} = 5.75 - 0.963 \times 4.70 = 1.224.$$

Para el ejemplo de la página 198 se calcularon varios estadísticos para las variables *Presión Arterial Sistólica* (X) y *Presión Arterial Diastólica* (Y) de los datos de la página 34. En particular, las medias y desviaciones estándares de ambas variables y su correlación eran:

$$\overline{x} = 130 \qquad \overline{y} = 85$$
$$s_X = 17 \qquad s_Y = 11$$
$$r_{XY} = .60$$

A partir de esta información, se puede calcular la recta de regresión de Y sobre X:

$$a = \frac{s_{XY}}{s_X^2}.$$

Considerando la Ecuación (6.4b), que relaciona la correlación con la covarianza, la expresión anterior equivale a

$$a = \frac{r_{XY}\, s_X\, s_Y}{s_X^2}$$
$$= r_{XY}\,\frac{s_Y}{s_X}.$$

Reemplazando los estadísticos en la última ecuación por los valores mencionados arriba, da como valor para a:

$$\begin{aligned} a &= .60\,\frac{11}{17} \\ &= 0.388. \end{aligned}$$

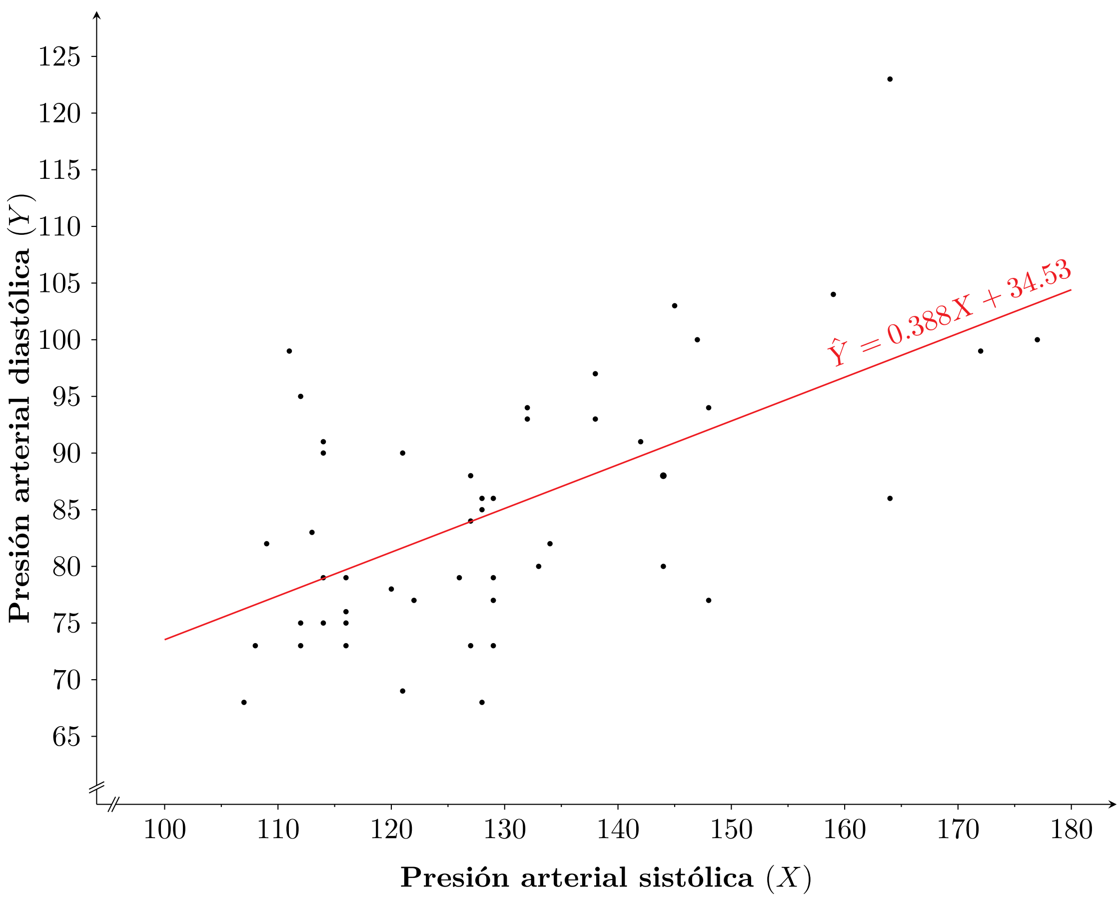
Para la constante b, se obtiene

$$\begin{aligned} b &= \overline{y} - a\,\overline{x} \\ &= 85 - 0.388 \times 130 = 34.53. \end{aligned}$$

Por lo tanto, la ecuación de regresión lineal es:

$$\hat{Y} = 0.388\,X + 34.53$$

La representación gráfica de la ecuación de regresión en el diagrama de dispersión de ambas variables se encuentra en la siguiente figura:



8.2.2 Regresión lineal simple y el coeficiente de correlación

Cuando se realiza una predicción por regresión lineal simple, se aprovecha la relación lineal que existe entre la variable predictora y la variable criterio. En la Sección 6.2.2—y especialmente en la Subsección 6.2.2.3—, se afirmó que el coeficiente de correlación de Pearson cuantifica la fuerza o la magnitud de la covariación lineal entre dos variables. Por lo tanto, no sorprende que los conceptos de correlación y predicción lineal óptima están estrechamente relacionados. En esta sección se introducen tres propiedades asociadas con la predicción lineal óptima, que aclaran el papel de la correlación de Pearson dentro del contexto de regresión lineal.

Propiedad 1: La correlación como enlace entre las puntuaciones estandarizadas

Si la predicción en la variable Y a partir de la variable X es lineal óptima, entonces, para todas las observaciones se cumple que:

$$\mathcal{Z}(\hat{y}_i) \;=\; r_{XY}\, \mathcal{Z}(x_i),$$

donde $\mathcal{Z}(\hat{y}_i)$ y $\mathcal{Z}(x_i)$ son las puntuaciones estandarizadas de $\hat{y}_i$ y x_i, definidas por:

$$\mathcal{Z}(\hat{y}_i) \;=\; \frac{\hat{y}_i - \overline{y}}{s_Y} \qquad \text{y} \qquad \mathcal{Z}(x_i) \;=\; \frac{x_i - \overline{x}}{s_X}.$$

En palabras: La correlación es el enlace entre, por un lado, el valor pronosticado en la variable criterio estandarizada y, por otro lado, el valor estandarizado en la variable predictora. En otras palabras, con base en la puntuación estandarizada que tiene una observación en la variable X se puede predecir su puntuación estandarizada en la variable Y, simplemente multiplicando la puntuación estandarizada en X por la correlación r_{XY}.

Demostración de la Propiedad 1

Se reemplazan las constantes a y b de la ecuación de regresión lineal por sus respectivos valores óptimos (en la Ecuación (8.5)):

$$\hat{y}_i \;=\; a\, x_i \;+\; b$$

Como primer paso, se reemplaza b por su valor óptimo (8.5b)

$$\begin{aligned} \hat{y}_i \;&=\; a\, x_i \;+\; \overline{y} \;-\; a\,\overline{x} \\ &=\; \overline{y} \;+\; a\,(x_i - \overline{x}) \end{aligned}$$

Como segundo paso, se reemplaza a por su valor óptimo (8.5a)

$$\hat{y}_i \;=\; \overline{y} \;+\; \frac{s_{XY}}{s_X^2}\,(x_i - \overline{x}).$$

La Ecuación (6.4b) relaciona la covarianza y la correlación e implica $s_{XY} = r_{XY} s_X s_Y$. Por lo tanto:

$$\begin{aligned}\hat{y}_i &= \overline{y} + \frac{r_{XY}\, s_X\, s_Y}{s_X^2}(x_i - \overline{x})\\ &= \overline{y} + r_{XY}\frac{s_Y}{s_X}(x_i - \overline{x}).\end{aligned}$$

De la última ecuación, se desprende que:

$$\frac{\hat{y}_i - \overline{y}}{s_Y} = r_{XY}\frac{x_i - \overline{x}}{s_X}. \qquad \blacksquare$$

Ejemplo

Ilustremos la Propiedad 1 utilizando los resultados de la regresión lineal de la calificación en el segundo departamental (Y) sobre la calificación en el primer departamental (X) de Bioquímica, presentados en el ejemplo en la página 255. En particular, calculemos para cada valor x_j en X y la predicción $\hat{y}_j$ asociada, la puntuación estandarizada. Para este cálculo se requiere la media y desviación estándar de ambas variables, las cuales se pueden derivar de la información en la página 259. Así, las fórmulas para calcular las puntuaciones estandarizadas son:

$$\mathcal{Z}(\hat{y}_j) = \frac{\hat{y}_j - 5.75}{2.74} \qquad\qquad \mathcal{Z}(x_j) = \frac{x_j - 4.70}{2.24}.$$

$\hat{y}_j$	$\xrightarrow{\mathcal{Z}}$	$\mathcal{Z}(\hat{y}_j)$				$\mathcal{Z}(x_j)$	$\xleftarrow{\mathcal{Z}}$	x_j
2.187	$\xrightarrow{\mathcal{Z}}$	−1.302	=	.788	×	−1.653	$\xleftarrow{\mathcal{Z}}$	1
3.150	$\xrightarrow{\mathcal{Z}}$	−0.950	=	.788	×	−1.206	$\xleftarrow{\mathcal{Z}}$	2
4.113	$\xrightarrow{\mathcal{Z}}$	−0.598	=	.788	×	−0.760	$\xleftarrow{\mathcal{Z}}$	3
5.076	$\xrightarrow{\mathcal{Z}}$	−0.246	=	.788	×	−0.313	$\xleftarrow{\mathcal{Z}}$	4
6.039	$\xrightarrow{\mathcal{Z}}$	+0.106	=	.788	×	+0.134	$\xleftarrow{\mathcal{Z}}$	5
7.002	$\xrightarrow{\mathcal{Z}}$	+0.458	=	.788	×	+0.581	$\xleftarrow{\mathcal{Z}}$	6
7.965	$\xrightarrow{\mathcal{Z}}$	+0.810	=	.788	×	+1.028	$\xleftarrow{\mathcal{Z}}$	7
8.928	$\xrightarrow{\mathcal{Z}}$	+1.161	=	.788	×	+1.474	$\xleftarrow{\mathcal{Z}}$	8
9.891	$\xrightarrow{\mathcal{Z}}$	+1.513	=	.788	×	+1.921	$\xleftarrow{\mathcal{Z}}$	9

La correlación entre ambas variables se calcula a partir de la covarianza y las desviaciones estándares de ambas variables (que se calcularon en la página 259) y aplicando la Ecuación (6.4b):

$$r_{XY} = \frac{s_{XY}}{s_X\, s_Y} = \frac{4.83}{\sqrt{5.01}\,\sqrt{7.49}} = .788.$$

Se observa que efectivamente las puntuaciones estandarizadas de la predicción y el valor en la variable predictora están relacionadas por la correlación, tal como se muestra en color rojo en la tabla.

Es importante interpretar correctamente la propiedad 1. Dice que si una observación tiene un valor en la variable predictora que se encuentra z desviaciones estándares arriba de la media de esta variable, entonces se predice que esta observación tiene un valor en la variable criterio que se encuentra $r \times z$ desviaciones estándares arriba de la media de la variable criterio (donde r es la correlación entre la predictora y el criterio). Concretamente, en el ejemplo de la presión arterial, se puede concluir que, si la presión arterial sistólica de una persona i se encuentra dos desviaciones encima de la media del grupo, entonces se predice que su presión arterial diastólica estará $0.60 \times 2 = 1.2$ desviaciones estándares encima de la media de la presión diastólica del grupo. Efectivamente, si la presión sistólica es 164 mmHg (es decir, 34 mmHg o dos veces la desviación estándar de 17 mmHg encima de la media de 130 mmHg), se predice según la ecuación derivada en la página 260 para la presión diastólica un valor de 98.2 mmHg (el cual se encuentra 13.2 mmHg o 1.2 desviaciones estándares de 11 mmHg encima de 85 mmHg).

Propiedad 2: Descomposición de la varianza total

Si la predicción en la variable Y a partir de la variable X es lineal óptima, entonces:

$$\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} + \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{n} \tag{8.7}$$

La fracción del lado izquierdo de la Ecuación (8.7) es la varianza (total) de la variable criterio Y. Es un índice de las diferencias globales que existen dentro de la muestra en la variable Y. La Propiedad 2 dice que las diferencias globales se dividen (en términos matemáticas: se descomponen) en dos partes:

(1) La primera parte es el error cuadrático medio asociado con la predicción lineal óptima de Y a partir de X. El nucleo del error cuadrático medio, como explicamos anteriormente, es el error para cada observación i entre el valor y_i real, observado y la predicción $\hat{y}_i$ con base en la variable X. Por lo tanto, el primer término del lado derecho de la igualdad es un índice de las diferencias en la variable Y que *no* se pueden predecir a partir de X. Se llama también la *varianza error* o la *varianza residual*.

(2) La segunda parte es la varianza de las predicciones. (Nótese que la media de las predicciones es igual a la media observada: $\overline{\hat{y}} = \overline{y}$, lo cual sigue directamente de la definición de $\hat{Y} = aX + b$ y los valores óptimos para a y b.) Para entender bien su significado, conviene recordar que $\overline{y}$ sería la predicción en caso de que no se despusiese de información adicional (como la variable X) para predecir Y (ya que, según el mismo argumento que lo utilizado en la Sección 8.1.3, se asigna la misma predicción a todas las observaciones y, en este caso, la media es la constante que minimiza los errores cuadráticos).
Es decir, el nucleo de la varianza de la segunda parte es la diferencia entre, por un lado, la predicción en caso de sí disponer de la información en X y, por otro lado, la predicción en caso de no tener esta información. En otra palabras, esta varianza expresa la parte de las diferencias que X logra predecir en la variable Y. Por lo tanto, se llama la *varianza explicada*.

Resumiendo: La Propiedad 2 dice que la varianza total de la variable criterio se puede descomponer en dos partes: (1) varianza error o residual y (2) la varianza explicada:

$$s_Y^2 \;=\; s_{Y.X}^2 \;+\; s_{expl}^2.$$

Demostración de la Propiedad 2

La siguiente igualdad se cumple para cada observación i:

$$(y_i - \overline{y}) \;=\; (y_i - \hat{y}_i) \;+\; (\hat{y}_i - \overline{y}).$$

Elevar ambos lados al cuadrado, resulta en:

$$(y_i - \overline{y})^2 \;=\; (y_i - \hat{y}_i)^2 \;+\; (\hat{y}_i - \overline{y})^2 \;+\; 2\,(y_i - \hat{y}_i)\,(\hat{y}_i - \overline{y}).$$

Puesto que esta igualdad se cumple para cada observación, también se cumple la igualdad que se obtiene sumando tanto el lado izquierdo como el lado derecho entre todas las observaciones:

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 \;=\; \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \;+\; \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 \;+\; 2\sum_{i=1}^{n}[(y_i - \hat{y}_i)\,(\hat{y}_i - \overline{y})].$$

Dividir ambos lados entre n nos da:

$$\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n} \;=\; \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \;+\; \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{n} \;+\; 2\,\frac{\sum_{i=1}^{n}[(y_i - \hat{y}_i)\,(\hat{y}_i - \overline{y})]}{n}.$$

Este resultado es igual al resultado en la Ecuación (8.7) ya que el último quebrado es igual a 0 según:

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \overline{y}) &= \sum_{i=1}^{n}(y_i - ax_i - b)(ax_i + b - \overline{y}) \\
&= \sum_{i=1}^{n}(y_i - ax_i - \overline{y} + a\overline{x})(ax_i + \overline{y} - a\overline{x} - \overline{y}) \\
&= \sum_{i=1}^{n}[(y_i - \overline{y}) - (ax_i - a\overline{x})](ax_i - a\overline{x}) \\
&= a\sum_{i=1}^{n}[(y_i - \overline{y})(x_i - \overline{x}) - a(x_i - \overline{x})^2] \\
&= a\,n\left(\frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{n} - a\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2]}{n}\right) \\
&= a\,n\,\left(s_{XY} - a s_X^2\right) \\
&= a\,n\left(s_{XY} - \frac{s_{XY}}{s_X^2}\,s_X^2\right) \\
&= 0.
\end{aligned}$$

■

Ejemplo

Verifiquemos la Ecuación (8.7) con las variables *Presión Arterial Diastólica* (X) e *Índice de Masa Corporal* (Y) de los datos de la página 34. Los estadísticos uni y bivariados asociados con la distribución (X, Y) son:

$$\begin{aligned} \overline{x} &= 85 \\ s_X &= 11 \end{aligned} \qquad r_{XY} = .40 \qquad \begin{aligned} \overline{y} &= 25 \\ s_Y &= 4, \end{aligned}$$

por lo cual los coeficientes de regresión de Y sobre X se dan por:

$$\begin{aligned} a &= r_{XY} \frac{s_Y}{s_X} \qquad \text{(véase el ejemplo en la página 259)} \\ &= .40 \, \frac{4}{11} = 0.145 \\ \text{y} \qquad b &= \overline{y} - a\,\overline{x} \\ &= 25 - 0.145 \times 85 = 12.636. \end{aligned}$$

Resumiendo:

$$\hat{y}_i = 0.145\, x_i + 12.636.$$

La siguiente tabla presenta, para cada observación, (a) la predicción $\hat{y}_i$ a partir de esta ecuación de regresión, (b) las diferencias $y_i - \overline{y}$, $y_i - \hat{y}_i$ y $\hat{y}_i - \overline{y}$ (las diferencias que forman el nucleo de las varianzas en la Ecuación (8.7)), y (c) estas diferencias al cuadrado:

i	x_i	y_i	$\hat{y}_i$	$y_i - \overline{y}$	$y_i - \hat{y}_i$	$\hat{y}_i - \overline{y}$	$(y_i - \overline{y})^2$	$(y_i - \hat{y}_i)^2$	$(\hat{y}_i - \overline{y})^2$
1	82	21	24.564	-4	-3.564	-0.436	16	12.700	0.190
2	77	25	23.836	0	1.164	-1.164	0	1.354	1.354
3	86	27	25.145	2	1.855	0.145	4	3.439	0.021
4	95	20	26.455	-5	-6.455	1.455	25	41.661	2.116
5	86	33	25.145	8	7.855	0.145	64	61.694	0.021
6	90	23	25.727	-2	-2.727	0.727	4	7.438	0.529
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
49	75	26	23.545	1	2.455	-1.455	1	6.025	2.116
50	69	26	22.673	1	3.327	-2.327	1	11.071	5.416
			Suma por columna:	0	0.000	0.000	800	672.000	128.000

Se verifica que la igualdad en la Ecuación (8.7) se cumple:

$$\begin{aligned} \frac{\sum_{i=1}^{50} (y_i - \overline{y})^2}{50} &= \frac{\sum_{i=1}^{50} (y_i - \hat{y}_i)^2}{50} + \frac{\sum_{i=1}^{50} (\hat{y}_i - \overline{y})^2}{50} \\ \iff \frac{800}{50} &\overset{!}{=} \frac{672}{50} + \frac{128}{50}. \end{aligned}$$

Propiedad 3: Coeficiente de determinación

Si la predicción en la variable Y a partir de la variable X es lineal óptima, entonces:

$$r_{XY}^2 = \frac{s_{expl}^2}{s_Y^2} \tag{8.8}$$

donde

$$s_{expl}^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2}{n}$$

La correlación cuadrática se denomina *coeficiente de determinación*.

Según la Ecuación (8.8), la correlación elevada al cuadrado es igual a la proporción de la varianza de Y que queda explicada por una predicción lineal óptima a partir de X. A veces se habla de la proporción de la varianza en la variable criterio que está *determinada* por X y se llama la correlación cuadrática "coeficiente de determinación". Por otro lado, aunque el término "determinación" tiene una conotación de causalidad, es importante tener claro que la correlación entre dos variables no necesariamente implica un efecto causal de una de las variables a la otra (tal como se argumentará en la página 270).

Demostración de la Propiedad 3

Se elabora la varianza explicada:

$$\begin{aligned}
s_{expl}^2 &= \frac{\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2}{n} \\
&= \frac{\sum_{i=1}^{n} (ax_i + b - \overline{y})^2}{n} \\
&= \frac{\sum_{i=1}^{n} [ax_i + (\overline{y} - a\overline{x}) - \overline{y}]^2}{n} \\
&= \frac{\sum_{i=1}^{n} (ax_i - a\overline{x})^2}{n} \\
&= a^2 \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n} \\
&= a^2 s_X^2 \\
&= r_{XY}^2 \frac{s_Y^2}{s_X^2} s_X^2 \\
&= r_{XY}^2 s_Y^2.
\end{aligned}$$

De la última expresión sigue inmediatamente la Ecuación (8.8). ■

Ejemplos

En el ejemplo anterior, se mencionó que el coeficiente de correlación entre *Presión Arterial Diastólica* (X) e *Índice de Masa Corporal* (Y) era igual a .40. Por otro lado, en el mismo ejemplo se encontró que:

$$\begin{aligned} s^2_{expl} &= \frac{\sum_{i=1}^{50}(\hat{y}_i - \overline{y})^2}{50} \\ &= \frac{128}{50} = 2.56 \end{aligned} \qquad \text{y} \qquad \begin{aligned} s^2_Y &= \frac{\sum_{i=1}^{50}(y_i - \overline{y})^2}{50} \\ &= \frac{800}{50} = 16. \end{aligned}$$

La razón de estas varianzas es:

$$\frac{s^2_{expl}}{s^2_Y} = \frac{2.56}{16} = .16,$$

lo cual efectivamente coincide con la correlación al cuadrado:

$$r^2_{XY} = (.40)^2 = .16.$$

El resultado anterior se interpreta como sigue: el 16 % de las diferencias entre las 50 personas de la muestra respecto de su índice de masa corporal se explican (en una regresión lineal) por diferencias respecto de su presión arterial diastólica.

Cambiemos en el ejemplo anterior la variable predictora: en lugar de utilizar *Presión Arterial Diastólica* se predice *Índice de Masa Corporal* a partir de la variable *Edad*. Dejando que X representa la *Edad* y Y el *Índice de Masa Corporal*, se resumen los estadísticos de esta distribución bivariada como sigue:

$$\begin{aligned} \overline{x} &= 46 \\ s^2_X &= 81 \end{aligned} \qquad s_{XY} = 0.00 \qquad \begin{aligned} \overline{y} &= 25 \\ s^2_Y &= 16. \end{aligned}$$

A partir de esta información, se obtiene para las constantes a y b de la regresión de Y sobre X:

$$\begin{aligned} a &= \frac{s_{XY}}{s^2_X} = \frac{0.00}{81} = 0 \\ b &= \overline{y} - a\,\overline{x} = 25 - 0 \times 46 = 25, \end{aligned}$$

lo cual implica que $\hat{y}_i = 0\,x_i + 25 = 25$ para todas las personas en la muestra.

La siguiente tabla es la análoga de la tabla en el ejemplo de la página 265.

i	x_i	y_i	$\hat{y}_i$	$y_i - \overline{y}$	$y_i - \hat{y}_i$	$\hat{y}_i - \overline{y}$	$(y_i - \overline{y})^2$	$(y_i - \hat{y}_i)^2$	$(\hat{y}_i - \overline{y})^2$
1	82	21	25	−4	−4	0	16	16	0
2	77	25	25	0	0	0	0	0	0
3	86	27	25	2	2	0	4	4	0
4	95	20	25	−5	−5	0	25	25	0
5	86	33	25	8	8	0	64	64	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
49	75	26	25	1	1	0	1	1	0
50	69	26	25	1	1	0	1	1	0
	Suma por columna:			0	0	0	800	800	0

Este ejemplo arroja luz a las propiedades de regresión lineal en caso de que la variable predictora y el criterio sean linealmente independientes (es decir, si la covarianza es 0):

(1) En caso de independencia lineal, la predicción no depende de X ya que la pendiente de la regresión es igual a 0. Por consiguiente, se predice para todas las observaciones la media de la variable Y (debido a que—como ya se ha mencionado varias veces en este tema y temas anteriores—es la constante que minimiza la suma de errores cuadráticos). En el ejmplo se ilustra con que la cuarta columna en la tabla es una constante: $\overline{y}$.
Este resultado también se puede interpretar en términos de la primera propiedad en regresión lineal. Si hay independencia lineal (y las dos variables involucradas en la regresión tienen una varianza mayor que 0), entonces la correlación es 0, y el valor estandarizado $\mathcal{Z}(\hat{y}_i) = 0$, para cualquier observación. Un valor $\mathcal{Z}(\hat{y}_i)$ de 0 implica que $\hat{y}_i = \overline{y}$, es decir, se vuelve a confirmar que para todas las observaciones se predice la media en Y.

(2) La igualdad en la Ecuación (8.7) toma una forma especial en el caso de independencia lineal. En el ejemplo anterior, aplicar la fórmula lleva a:

$$\begin{aligned} \frac{\sum_{i=1}^{50}(y_i - \overline{y})^2}{50} &= \frac{\sum_{i=1}^{50}(y_i - \hat{y}_i)^2}{50} + \frac{\sum_{i=1}^{50}(\hat{y}_i - \overline{y})^2}{50} \\ \iff \quad \frac{800}{50} &= \frac{800}{50} + \frac{0}{50}. \end{aligned}$$

Es decir, la varianza total es igual a la varianza residual (o la varianza error); o de forma equivalente, la varianza explicada es 0: la variable predictora (*Edad* en este ejemplo) no logra explicar nada de las diferencias en la variable criterio (*Índice de Masa Corporal*).

(3) Lo anterior se refleja directamente en el coeficiente de determinación: la correlación es igual a 0 (y entonces también el coeficiente de correlación), lo cual implica que nada de las diferencias entre personas respecto de su *Índice de Masa Corporal* se explica por diferencias respecto de *Edad*. Sin embargo, hay que ser prudente con esta afirmación: es necesario tener claro que *Edad* no explica nada de *Índice de Masa Corporal* en una regresión (o predicción) *lineal*.

Las tres propiedades anteriores acarrean que la correlación juega un doble papel en la regresión lineal. El siguiente cuadro lo resume:

El doble papel de la correlación en regresión lineal:

En el caso de predicción lineal óptima de Y a partir de X, la correlación r_{XY}

(1) es el enlace entre los valores estandarizados en X y las predicciones estandarizadas en Y;

(2) llevada al cuadrado indica la proporción de la varianza en Y que queda explicada por X.

La Propiedad 5 de la correlación (p. 205) especificó que r_{XY} siempre se encuentra entre -1 y 1. Ahora pasemos por las diferentes posibilidades para el valor de la correlación y su significado en el contexto de la regresión lineal de Y sobre X:

$\boxed{r_{XY} = 1}$ En este caso, la Propiedad 1 de regresión lineal implica que:

$$\forall i : \mathcal{Z}(\hat{y}_i) = \mathcal{Z}(x_i).$$

Además, en este caso, $r^2_{XY} = 1$, lo cual implica que las predicciones son perfectas. Es decir, para todas las observaciones se cumple $\hat{y}_i = y_i$; por lo tanto

$$\forall i : \mathcal{Z}(y_i) = \mathcal{Z}(x_i).$$

Nótese que ésta es la misma conclusión a la que llegamos—por otras vías—en la Propiedad 6a de la correlación (p. 207).

$\boxed{r_{XY} = -1}$ Este caso es similar al anterior, excepto que la relación entre las variables es inversa:

$$\forall i : \mathcal{Z}(\hat{y}_i) = -\mathcal{Z}(x_i).$$

Además, el coeficiente de determinación vuelve a ser igual a 1, por lo cual $\hat{y}_i = y_i$ para todas las observaciones. Sigue que:

$$\forall i : \mathcal{Z}(y_i) = -\mathcal{Z}(x_i),$$

lo cual, a su vez, corresponde con la Propiedad 6b de la correlación (p. 207).

$\boxed{0 < r_{XY} < 1}$ En este caso, hay una relación lineal directa no perfecta entre X y Y. La Propiedad 1 de regresión lineal implica que en este caso $\mathcal{Z}(\hat{y}_i)$ se encuentra más cerca a 0 que $\mathcal{Z}(x_i)$ (ya que multiplicar por un número cuyo valor absoluto es menor que 1 siempre resulta en un número más cercano a 0). Esto quiere decir que la puntuación pronosticada en la variable criterio, $\hat{y}_i$, se encontrará relativamente más cercana a su media $\overline{y}$, en comparación con el valor en la variable predictora, x_i, respecto a su media $\overline{x}$. (Se añade "relativamente" porque se debe interpretar la diferencia con la media en términos de desviaciones estándares.) Este fenómeno se conoce en la literatura como la *"regresión a la media"*.

Para entender bien el significado de regresión a la media, sirve mucho un ejemplo de Sir Francis Galton, quien era el primero que describió este fenómeno en el contexto de sus estudios de herencia. Supongamos que se conocen, para 1000 familias, la altura del padre (X) y la altura de la hija mayor adulta (Y) y que la correlación r_{XY} es igual a 0.40. La Propiedad 1 implica que, si el padre es muy alto (por ejemplo, de tres desviaciones estándares encima de la media de los padres) entonces se predecirá que su hija también será alta con una altura que está por encima de la media de las hijas, pero ella no tendrá una altura tan extrema que su padre (sino se estima que tenga una altura que corresponde con 1.2 desviaciones estándares encima de la media de las hijas). En lo anterior es importante contemplar que la altura del padre (x_i) se compara con la media de los *padres* ($\overline{x}$), y la diferencia ($x_i - \overline{x}$) se interpreta en términos de la desviación estándar de los padres (s_X); por otro lado, la altura pronosticada de la hija ($\hat{y}_i$) se compara con la media de las *hijas* ($\overline{y}$) —la cual, en principio, podría ser mayor o menor que la media de los padres— y la diferencia ($\hat{y}_i - \overline{y}$) se interpreta en términos de la desviación típica de las *hijas* (s_Y).

$\boxed{-1 < r_{XY} < 0}$ Implica que existe una relación entre X y Y que es lineal inversa y no perfecta. También en este caso se produce una regresión a la media, incluyendo un cambio de sentido. Es decir, $\mathcal{Z}(\hat{y}_i)$ se encuentra más cerca a 0 que $\mathcal{Z}(x_i)$ y además tendrá un signo opuesto. Por ejemplo, si $r_{XY} = -0.50$ y la puntuación de alguna observación en X es dos desviaciones *encima* de la media $\overline{x}$, entonces se predice que esta observación tendrá un valor en Y que se encuentra una desviación típica *debajo* de la media $\overline{y}$ ($\mathcal{Z}(\hat{y}_i) = -0.50 \times 2 = -1$).

$\boxed{r_{XY} = 0}$ Como se ilustró en el último ejemplo, este caso implica que $\mathcal{Z}(\hat{y}_i) = 0$ o equivalentemente $\hat{y}_i = \overline{y}$ y que X no contiene nada de información para predecir Y en una regresión lineal ($r^2_{XY} = 0$). Se puede interpretar esta situación como un caso extremo de regresión a la media.

Concluyendo esta sección, se fija la atención en los siguientes comentarios:

(1) Se ha enfatizado de sobra que la correlación r_{XY} establece una relación entre puntuaciones *estandarizadas*. También la definición del coeficiente de correlación (véase la Ecuación (6.4a)) incluye únicamente información estandarizada. Debe ser claro que las variables estandarizadas no contienen información sobre la media ni la varianza (o desviación estándar) de las variables originales y, por consiguiente, no se debe sacar conclusiones a partir de una correlación sobre la media y la desviación estándar de X y Y.

(2) El coeficiente de correlación sólo informa sobre la relación *lineal* entre X y Y. Un coeficiente de determinación $r^2_{XY} \neq 1$ *no* quiere decir que una predicción perfecta de Y a partir de X sea imposible. Lo que *sí* quiere decir es que una predicción *lineal* perfecta de Y a partir de X es imposible.

(3) En este tema se revisó el papel del coeficiente de correlación a la luz de una problemática formal-matemática: la predicción de Y a partir de X. Sin embargo, en esta revisión no se habló de una interpretación de la correlación a la luz del contexto en el cual se han definido las variables X y Y. Tal interpretación generalmente es bastante compleja y requiere un análisis adicional del diseño de la recopilación de datos. Es general, diferentes tipos de explicaciones son posibles al buscar una interpretación por qué dos variables mantienen una correlación:

(a) Una primera posibilidad es la de una relación causal directa entre X y Y. Por ejemplo, se puede observar en un grupo de alcohólicos que existe una relación lineal *inversa* entre el grado de adicción (X) y el tamaño de la red social (es decir, la extensión del círculo de amistades/amigos, Y). Una explicación de esta correlación puede ser la de una relación causal entre ambas variables:

$X \longrightarrow Y$ El grado de adicción y el consumo de alcohol causa que poca gente se quiera relacionar con la persona.

$Y \longrightarrow X$ El control social por los amigos disminuye el consumo de alcohol.

(b) Otra posibilidad es que exista una relación causal *in*directa entre X y Y, es decir, la causalidad pasa por una variable intermedia Z:

$X \longrightarrow Z \longrightarrow Y$ El abuso de alcohol causa un trastorno de ansiedad (variable Z) y esta ansiedad causa que la persona tenga dificultades para relacionarse con otras personas.

$Y \longrightarrow Z \longrightarrow X$ El tamaño reducido del círculo social causa un estado de depresión (variable Z) y el estado de depresión causa el abuso de alcohol.

(c) Una tercera posibilidad es que una variable Z causa (directa o indirectamente) tanto X como Y:

$$\begin{tikzcd} & X \\ Z \arrow[ur] \arrow[dr] & \\ & Y \end{tikzcd}$$

Por un problema no resuelto de la infancia, la persona tiene una personalidad dependiente (Z) la cual, por un lado, causa que otra gente se sienta incómoda y no quiera relacionarse con ella (llevando a un círculo reducido de amigos) y, por otro lado, provoca el abuso de alcohol en esta persona.

(d) Por último, puede ser que no subyaga *ningún* patrón real a la correlación. En este caso se habla de una *correlación espuria.*

8.3 Introducción a la regresión lineal múltiple

Sección por desarrollar...

En muchas ocasiones se dispone de información en más de una variable para realizar una predicción. Por ejemplo, para predecir el desempeño (Y) de un estudiante en la maestría en Ciencias Médicas, Odontológicas y de la Salud, se puede evaluar no solo el promedio de sus notas en las asignaturas durante la licenciatura (X_1), sino también se le puede aplicar una prueba de personalidad para evaluar si dispone de las actitudes apropiadas (X_2) y/o una entrevista para conocer sus motivaciones (X_3). Para este tipo de situaciones, se extendieron los métodos de predicción presentados en las dos secciones anteriores para acomodar más variables predictoras.

En la página 247, se conceptualizó la predicción a partir de una variable predictora X como la aplicación de una función f a X. En el caso de m variables predictoras, la función de predicción f es una función multivariable:

$$\hat{y}_i \;=\; f(x_{i1}, x_{i2}, \ldots, x_{im}).$$

En palabras: para obtener la predicción $\hat{y}_i$ para una observación i, se aplica la función f a los valores $x_{i1}, x_{i2}, \ldots, x_{im}$ que se han observado en las respectivas variables predictoras.

Al introducir regresión lineal simple en la Sección 8.2.1, la función f se restringió para que tuviese una forma particular, la de una función lineal. En el caso de regresión lineal múltiple, se restringe la función multivariable f en la ecuación anterior a que tenga una forma particular, específicamente, una combinación lineal de las m variables predictoras. Esto quiere decir que los valores $x_{i1}, x_{i2}, \ldots, x_{im}$ que se observaron para i en las m variables predictoras $X_1, X_2, \ldots, X_m$ se combinan de la siguiente forma para obtener el valor pronosticado:

$$\hat{y}_i \;=\; b_0 \;+\; b_1\, x_{i1} \;+\; b_2\, x_{i2} \;+\; \ldots \;+\; b_m\, x_{im}.$$

En la expresión anterior, $b_0, b_1, \ldots, b_m$ son las constantes o los coeficientes de la regresión lineal de Y sobre $X_1, X_2, \ldots, X_m$ (similares a las constantes a y b en regresión lineal simple).

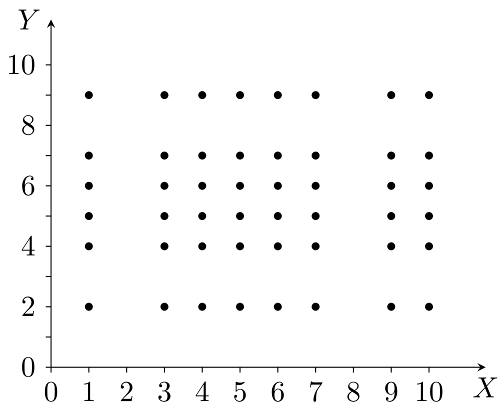
También en el caso de regresión lineal múltiple, el criterio para evaluar la bondad de predicción es el error cuadrático medio. Esto quiere decir que el objetivo es determinar los coeficientes $b_0, b_1, \ldots, b_m$ de tal forma que la suma de errores de las predicciones en la muestra de n observaciones,

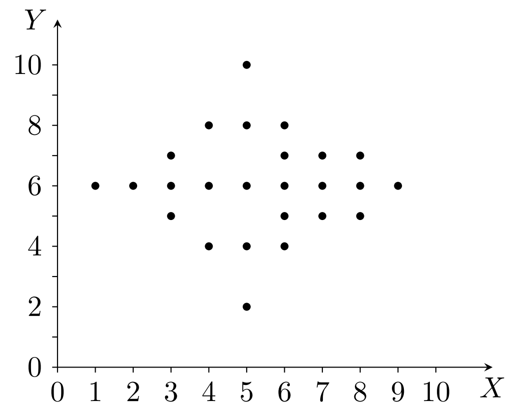
$$\sum_{i=1}^{n} (\hat{y}_i - y_i)^2,$$

sea mínima. El teorema en la página 256 presentó las fórmulas para las constantes de regresión a y b en el caso de regresión lineal simple; como se detalló en la demostración de este teorema, dichas fórmulas se derivaron considerando la suma de errores cuadráticos como una función de las constantes a y b, y a continuación aplicando cálculo diferencial para encontrar las expresiones para a y b que minimizan esta función. Para regresión lineal múltiple, el procedimiento es similar, aunque es más complejo debido a que se analiza la suma de errores cuadráticos en función de $m+1$ constantes (en lugar de solo 2). Es decir, el cálculo diferencial permite buscar el mínimo del error cuadrático medio en función de las $m+1$ constantes $b_0, b_1, \ldots, b_m$. Para calcular y representar la solución se suele utilizar algebra matricial. A pesar de que las matemáticas son más complicadas y que pueden surgir problemas específicos por las relaciones entre las variables predictoras, la interpretación del resultado obtenido de una regresión lineal múltiple es similar a la interpretación en el caso de regresión lineal simple.

Problemas

1. Considérense los dos conjuntos de datos bivariados representados en los siguientes diagramas de dispersión:





   (a) Para ambos diagramas, ¿cuál es el pronóstico que se asigna a cada valor observado en la variable X, si se realiza una predicción óptima de Y a partir de X?

   (b) ¿Y si se realiza una predicción lineal óptima?

   (c) Compara los resultados obtenidos y saca una conclusión general.

2. ¿Qué valor tiene el error medio, el cual se define como

$$\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)}{n},$$

   (a) en caso de predicción óptima de Y a partir de X?

   (b) en caso de predicción lineal óptima de Y a partir de X?

3. (a) Deriva una fórmula que permite calcular el error cuadrático medio ($s^2_{Y.X}$) en caso de regresión lineal de Y sobre X a partir de la varianza de la variable criterio (s^2_Y) y la correlación entre la variable predictora y el criterio (r_{XY}).

   (b) Verifica la fórmula derivada hallando el error cuadrático medio al predecir *Índice de Masa Corporal* a partir *Presión Arterial Diastólica* en los datos de la página 34 por una regresión lineal.
   (El ejemplo de la página 265 incluye las tres partes de información que se relacionan en la fórmula derivada.)

4. La siguiente tabla presenta, para 12 parejas casadas, la altura (en cm) de la mujer y del hombre.

Mujer	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer	Hombre
156	172	171	186	153	164	166	156
171	180	161	155	157	154	165	175
151	174	170	167	162	165	173	180

(a) Realiza una regresión lineal con la altura de la mujer como variable criterio y la altura del hombre como variable predictora. En específico, halla las constantes de la ecuación de regresión y calcula también la varianza residual, la varianza explicada y el coeficiente de determinación.

(b) Repite el ejercicio realizando una regresión lineal donde se intercambia el papel de ambas variables. Es decir, realiza una predicción lineal óptima para la altura del hombre a partir de la altura de la mujer.

(c) En general, ¿cuáles de los estadísticos anteriores mantienen su valor (y cuáles no) al intercambiar el papel de la variable predictora y la variable criterio en una regresión lineal simple?

5. Supongamos que se realiza tanto una predicción óptima como una predicción lineal óptima de la variable Y a partir de X y que, para las predicciones obtenidas, se calcula el error cuadrático medio. El siguiente cuadro considera diferentes combinaciones para el valor del error cuadrático medio en ambos casos. Para cada una de las nueve combinaciones, indica si puede ocurrir o no.

	Predicción lineal óptima		
Predicción óptima	$s^2_{Y.X} = 0$	$0 < s^2_{Y.X} < s^2_Y$	$s^2_{Y.X} = s^2_Y$
$s^2_{Y.X} = 0$			
$0 < s^2_{Y.X} < s^2_Y$			
$s^2_{Y.X} = s^2_Y$			

Si es imposible que ocurra, argumenta por qué; en caso contrario, construye un diagrama de dispersión de X e Y que representa unos datos bivariados que exhiben los errores cuadráticos medios especificados para ambos tipos de predicciones.

*6. La siguiente tabla presenta la evolución de la cifra de población en México desde el inicio del Siglo XX:

Año del censo	*Cifra poblacional*
1900	13,607,259
1910	15,160,369
1921	14,334,780
1930	16,552,722
1940	19,653,552
1950	25,791,017
1960	34,923,129
1970	48,225,238
1980	66,846,833
1990	81,249,645
1995	91,158,290
2000	97,483,412
2005	103,263,388
2010	112,336,538

Fuente: INEGI

(a) Presenta los datos gráficamente en un diagrama de dispersión con el *Año del censo* (X) en la abscisa y la *Cifra poblacional* (Y) en la ordenada.

(b) Para cada uno de los valores observados en *Año del censo*, calcula el pronóstico en *Cifra poblacional* para el caso de una predicción lineal óptima. Presenta los pronósticos en el diagrama de dispersión (mediante una estrella).

(c) Aplica a la cifra poblacional una transformación logarítmica, $T = \log_{10} Y$, y cálcula la ecuación de regresión para predecir T linealmente a partir de X.

(d) Aplica la transformación inversa de la logarítmica para hallar el valor pronosticado en la variable Y original a partir del pronóstico obtenido para la variable T. Representa los pronósticos en Y en el diagrama de dispersión (mediante un cuadradito).

(e) ¿Cuál de los dos métodos ha generado la mejor predicción de Y?

Parte II

Estadística inferencial

Tema 9
Marco conceptual

Índice

En la primera parte de este curso, las afirmaciones y las conclusiones que se sacaron de los análisis nunca sobrepasaron el alcance de los datos observados. Por ejemplo, se encontró en *la muestra de 50 personas* cuyos datos se presentaron en la página 34, que en el grupo de universitarios había menos fumadores que en los otros grupos (véase el diagrama de rectángulos en la página 154) o que *en esta muestra* la correlación entre *Presión Arterial Diastólica* y *Presión Arterial Sistólica* era .60 (véase el ejemplo en la página 199). De forma similar, se concluyó que el psiquiatra y el psicólogo del ejemplo en la página 170 coincidieron sobre la mejor terapia para el 65 % de *los 48 pacientes que se presentaron a su centro en el último mes*. En muchas ocasiones, sin embargo, el objetivo final de un investigador no es concluir que en este grupo de 50 personas hay más o menos fumadores entre las personas con niveles educativos más altos o que en este grupo la correlación es .60, sino quiere llegar a conclusiones similares sobre una población más grande y afirmar, por ejemplo, que los mexicanos con un nivel educativo alto fuman menos en comparación con los que tienen el nivel educativo bajo. Asimismo, el psiquiatra y el psicólogo quieren saber hasta qué grado coinciden sus juicios para todos los pacientes que lleguen a su centro. En esta segunda parte del curso, se introducen las herramientas para realizar este tipo de generalizaciones. El tema actual inicia esta parte con una revisión de los conceptos básicos.

9.1 Población y muestra

Población

La *población* es el conjunto de todos los objetos o unidades experimentales (p.e. pacientes, estudiantes) sobre los cuales se quiere llegar a conclusiones.

La población puede ser finita, infinita numerable o (infinita) innumerable (para una explicación de estos términos, véase la nota en la página 19). Si es finita, se indica el número de elementos por el símbolo N.

Ejemplos

Un jugador en el casino quiere conocer el porcentaje de veces que la bola se cae en el número 7 considerando la *población* de todos los lanzamientos de la bola.

Una empresa farmacéutica que ha desarrollado un nuevo fármaco contra la enfermedad de Parkinson, quiere conocer la efectividad del medicamento en la *población* de todos los pacientes que sufren de esta enfermedad.

Un político, candidato en las elecciones presidenciales, quiere saber con cuánto apoyo cuenta en la *población* de todas las personas que tienen derecho de votar.

Un médico chino quiere conocer cuál es el efecto perjudical de la contaminación ambiental en la esperanza de vida dentro de la *población* de personas que viven en Beijing.

Un atleta está interesado en conocer qué tan lejos puede echar la jabalina considerando la *población* de los posibles lanzamientos (por ejemplo, en diversas circunstancias climatológicas).

Un investigador del Centro Nacional de Prevención de Accidentes quiere conocer dentro de la *población* de todos los taxis que circulan en el Distrito Federal, cuántos tienen cinturones de seguridad instalados y funcionando.

Un psiquiatra, que tiene en tratamiento un paciente con un trastorno esquizofrénico, quiere conocer cómo ciertas características situacionales desencadenan un episodio psicótico en este paciente considerando la *población* de todas las situaciones en las que se pueda encontrar.

El primer caso y los ultimos tres ejemplifican que la población no siempre consiste en personas, sino que, dependiendo de la pregunta de investigación, pueden ser también situaciones, objetos (como coches), países, etc.

Estos ejemplos también muestran que en muchas ocasionas el tamaño de la población es (muy) grande (en el primer ejemplo, incluso es infinita, hay un número infinito de lanzamientos que se pueden considerar). El ser demasiado grande es una de las razones principales por las que generalmente no se estudia la población entera; por razones de economía, se suele estudiar solo una *muestra* de objetos a partir de la cual se espera llegar a conclusiones sobre la población. Otras razones pueden ser de naturaleza ética. Por ejemplo, algunos estudios de cerebro con primates sacrifican a los animales que participaron en el experimento para examinar su cerebro a un análisis *post mortem*. Evidentemente, es deseable minimizar el número de animales sacrificados en este tipo de estudios.

Muestra

Una *muestra* es un subconjunto ordenado de elementos de la población.

Se dice que una muestra *se extrae* de la población. El tamaño de la muestra se representa por n.

Ejemplos

La empresa farmacéutica que desarrolló un nuevo fármaco contra el Parkinson realiza un estudio con 20 pacientes con esta enfermedad.

Para el político que quiere saber con cuánto apoyo cuenta, una empresa consultora experta en estudios de opinión llega a la conclusión de que el candidato puede contar con aproximadamente el 40 % de los votos, con base en una *muestra* de 1000 personas.

El investigador del Centro Nacional de Prevención de Accidentes detiene (con la ayuda de la policía federal) una *muestra* de 500 taxis circulando por las vías públicas del Distrito Federal y revisa los cinturones de seguridad.

Unas notas respecto de la definición de muestras

Nota 1: La definición no excluye la posibilidad de que el mismo elemento se encuentre más de una vez en la muestra. En la siguiente sección se introducirá una distinción relevante al respecto, a saber la diferencia entre muestras extraídas *con reposición* y *sin reposición*.

Nota 2: La definición de una muestra especifica que es un subconjunto *ordenado*. El orden de los elementos de la muestra se define por el orden en el cual se extraen de la población. En principio, los elementos de la muestra se extraen uno por uno. Si no es el caso, se define un orden a partir de otro criterio que la secuencia temporal; por ejemplo, si se aplica un examen a una muestra de cien personas *a la vez*, se ordenan los elementos posteriormente, por ejemplo, basado en el orden en que aparezcan sus cuestionarios en la pila.

Nota 3: No siempre es importante el orden de los elementos en la muestra. Muchas veces no lo es, como en el caso de la aplicación de un examen. Por otro lado, hay casos en que el orden es de suma importancia, por ejemplo en una muestra de medidas repetidas de parámetros de salud de un paciente donde se quiere investigar la evolución de una enfermedad.

Nota 4: En algunos libros de estadística se consideran también subconjuntos de la población no ordenados como muestras. En este curso, sin embargo, si no se especifica lo contrario, "muestras" siempre refieren a conjuntos *ordenados*.

La estadística inferencial se dedica a derivar procedimientos para llegar a conclusiones sobre la población a partir de información de una muestra obtenida de esta población. Para entender estas derivaciones es crucial analizar en detalle el proceso de la recopilación de los datos. Dicho proceso se puede dividir en tres pasos:

Paso 1: *Realizar un experimento aleatorio*
Esto quiere decir que se ejecuta un procedimiento de acuerdo con reglas fijas, lo cual implica que se extrae una muestra de una población en la que se aplicarán algunas operaciones que, al final, nos proporcionan los datos.

Paso 2: *Determinar el resultado*
De toda la información que nos proporciona el experimento aleatorio, se registra sólo una parte reducida: el resultado.

Paso 3: *Determinar el valor en una o más variables aleatorias*
Para cada resultado se determinará el valor numérico en una(s) variable(s) de interés.

Ejemplo

Consideremos el proceso de la recopilación de datos para el caso de la empresa farmacéutica que desea investigar la eficacia de un nuevo fármaco para la enfermedad de Parkinson:

Paso 1: Respecto de la realización del experimento aleatorio se decide extraer, siguiendo reglas específicas, 20 pacientes de la población de pacientes con Parkinson y aplicar a cada uno de estos pacientes una prueba de psicomotricidad antes y después de la ingestión de la nueva medicina.

Paso 2: Como el resultado del experimento aleatorio, se decide registrar el número de errores en la prueba de psicomotricidad (antes y después de la ingestión de la medicina). Otra información del experimento aleatorio (por ejemplo, la velocidad con la que el paciente realiza las pruebas, su mímica durante la ingestión de la medicina, etc.) no se toma en cuenta para el estudio y, por lo tanto, no forma parte del resultado.

Paso 3: Se definen dos variables aleatorias X y Y que asignan a los posibles resultados que se podrían haber observado (a) el número de errores en la prueba antes de haber tomado la medicina, (valor en la variable X) y (b) el número de errores después de haber tomado la medicina, (valor en la variable Y).

Las siguientes tres secciones se dedican a un análisis más detallado de estos tres pasos, respectivamente. La Sección 9.2 se trata de experimentos aleatorios (donde también se habla del muestreo, es decir, el procedimiento que proporciona la muestra); la Sección 9.3 es sobre los resultados del experimento aleatorio; y en la Sección 9.4 se introducirá la definición formal de variables aleatorias.

9.2 Experimento aleatorio (incluyendo muestreo)

Experimento aleatorio

Un *experimento aleatorio* es un procedimiento que:

(a) se realiza de acuerdo a reglas fijas

(b) *en principio* se puede repetir un número infinito de veces bajo las mismas condiciones

(c) produce datos que no se pueden predecir con certeza a partir de las reglas bajo las cuales el experimento se realiza.

Ejemplos

Para conocer el comportamiento electoral de los empadronados con derecho a voto en las siguientes elecciones para Jefe del Gobierno del Distrito Federal, se puede realizar el siguiente experimento aleatorio, que consiste en los siguientes tres pasos:

- Seleccionar al azar 50 colonias del Distrito Federal
- En cada una de las colonias seleccionadas, elegir al azar 20 personas del padrón electoral
- Presentar a cada una de las ($50 \times 20 =$) 1,000 personas elegidas una papeleta con una lista de los diferentes candidatos y pedirle que marque el candidato por el que votaría si hubiera elecciones.

El director de la Facultad de Medicina quiere conocer la opinión de los estudiantes sobre la asignatura *Informática Biomédica I* que se introdujo en el Plan de Estudios 2010 y pide a sus colaboradores realizar el siguiente experimento aleatorio:

- Seleccionar al azar 10 grupos, de los cuales uno sea un grupo PAEA (Programa de Alta Exigencia Académica)
- Presentar a todos los estudiantes de los 10 grupos seleccionados, un cuestionario que tiene 80 preguntas con cuatro opciones de respuesta (tipo Likert) sobre diversos aspectos de la docencia de Informática Biomédica I

Para poder planear un tratamiento para una de sus pacientes, que sufre de episodios depresivos que parecen estar relacionados con el ciclo menstrual, su psicólogo le pide que durante un mes, cada día a las 8 de la noche, rellene un breve cuestionario que sondea cuáles han sido los eventos más significativos del día y cómo se sintió al respecto.

Unas notas respecto de la definición de un experimento aleatorio

Nota 1: Aunque el término "experimento" tiene la conotación de una manipulación experimental, como en el contexto de "diseños experimentales" (los cuales se suelen asociar con experimentos controlados), esto no es necesariamente cierto para el concepto de un *experimento aleatorio* en la estadística. Como ilustran los ejemplos anteriores, el experimento aleatorio y las reglas fijas que la definición del mismo incluye, pueden llevarse a cabo aplicando un cuestionario o realizando observaciones de un fenómeno natural. Lo importante es que se realice utilizando un procedimiento estandarizado.

Nota 2: Un experimento aleatorio se puede repetir *en principio* un número infinito de veces. Esto quiere decir que los procedimientos del experimento *permiten* que el experimento se repita infinitamente, entendiendo que estas repeticiones son posibles *conceptualmente* (en teoría) y no necesariamente que se realicen efectivamente. Así, en principio es posible aplicar 1000 veces el procedimiento descrito en el primero de los ejemplos anteriores en el mismo día, mientras que esto no es posible en la práctica.

Nota 3: Debido al tercer criterio, un experimento aleatorio difiere de un experimento *determinístico* cuyo resultado sí se puede predecir exactamente a partir de las condiciones bajo las cuales el experimento se lleva a cabo.

Nota 4: Realizar un experimento aleatorio implica extraer una muestra de la población, de acuerdo con un procedimiento fijo. El tamaño n de la muestra suele especificarse en las reglas del experimento aleatorio; en este caso, diferentes valores de n corresponden con diferentes experimentos aleatorios. Para los métodos estadísticos inferenciales que se introducirán en los siguientes temas, siempre se supone que el tamaño de la muestra es fijo y que se especifica como parte del experimento aleatorio.

Nota 5: Las reglas del experimento aleatorio, por sí mismas, pueden contener ciertas formas de repetición. Por ejemplo, se puede extraer una muestra de 20 pacientes de Parkinson y presentarles *dos veces* la misma tarea de psicomotricidad (antes y después de la administración del nuevo fármaco).
Nótese que la repetición mencionada aquí es otra que las repeticiones implicadas en la nota 2 anterior.

La realización de un experimento aleatorio implica la extracción de una muestra de tamaño n de la población. El proceso de obtener la muestra se denomina *muestreo*. En el resto de esta sección se introducen algunas nociones importantes relacionadas con el muestreo.

Muestreo con y sin reposición

Muestreo con y sin reposición

- Si durante el proceso de extracción de la muestra cada elemento extraído se devuelve antes de extraer el siguiente, el muestreo es *con reposición*.
- Si, al contrario, los elementos que se extraigan de la población no se devuelven a la población durante el muestreo, el muestreo es *sin reposición*.

Comparación entre muestreo con y sin reposición

	Muestreo con reposición	*Muestreo sin reposición*
1.	El mismo elemento de la población puede incluirse múltiples veces en la muestra.	Cada elemento de la población puede incluirse sólo una vez en la muestra.
2.	El tamaño de la muestra puede ser más grande que el tamaño de la población (posiblemente, $n > N$).	El tamaño de la muestra no puede ser más grande que el tamaño de la población (siempre, $n \leqslant N$).
3.	La composición de la población se queda constante durante la extracción de la muestra.	La composición de la población cambia durante la extracción de la muestra.

Ejemplo

Ilustremos las diferencias entre los dos tipos de muestreo a través de un ejemplo. Considérese la siguiente población, con $N = 5$ personas:

$$\{\text{Adrián}, \text{Carlos}, \text{Diana}, \text{Lizz}, \text{Oliver}\}.$$

1. Al extraer una muestra con reposición de tamaño $n = 2$, es posible que se obtenga como primer elemento Carlos. Después de haber regresado a Carlos a la población, es posible que se vuelva a extraer el mismo elemento, tal que la muestra sea:

   $$(\text{Carlos}, \text{Carlos}).$$

   Si se extrae la muestra sin reposición, no se regresa Carlos a la población y no es posible que el segundo elemento extraído vuelva a ser Carlos.

2. El muestreo con reposición permite la extracción de una muestra cuyo tamaño supera el tamaño de la población. Por ejemplo, se podría obtener una muestra de $n = 8$ y se observa que $n > N$:

   $$(\text{Carlos}, \text{Oliver}, \text{Lizz}, \text{Adrían}, \text{Carlos}, \text{Diana}, \text{Carlos}, \text{Diana}).$$

   En el caso de un muestreo sin reposición, después de haber extraído el quinto elemento, ya no quedan elementos en la muestra y por lo tanto, es imposible extraer una muestra de más de $n = 5$ personas.

3. Consideremos la extracción de una muestra de tamaño $n = 2$.
   El primer elemento se saca, para ambos tipos de muestreo, de la población original:

   $$\{\text{Adrián}, \text{Carlos}, \text{Diana}, \text{Lizz}, \text{Oliver}\}.$$

Supongamos que el primer elemento que se saca es Adrián.

- *En el caso de muestreo con reposición*, se regresa a Adrián a la población antes de realizar la segunda extracción. Por lo tanto, el segundo elemento se saca del mismo conjunto que el primer elemento.
- *En el caso de muestreo sin reposición*, Adrián se queda fuera de la población y el segundo elemento se saca del conjunto

$$\{\text{Carlos}, \text{Diana}, \text{Lizz}, \text{Oliver}\}.$$

Es decir, para la extracción del segundo elemento, la composición de la población cambió.

En las ciencias de la salud, el muestreo suele ser sin reposición. Por ejemplo, al evaluar la eficacia de un nuevo fármaco no se permite que la muestra incluya dos veces al mismo paciente. Asimismo, al evaluar el desempeño de los docentes de una asignatura a través de la opinión de estudiantes, la muestra comúnmente no incluye dos veces a un mismo estudiante. Sin embargo, estrictamente hablando, la mayoría de los métodos estadísticos (de los cuales algunos se tratan en los últimos temas) son válidos sólo si la muestra se ha extraído con reposición (o bien, si la población es innumerable). Por otro lado, por fines prácticos, estos métodos llevarán a conclusiones correctas si el tamaño n de la muestra es mucho menor que el tamaño N de la población (en este caso, los cambios en la composición de la población durante la extracción de la muestra serán muy pequeños y prácticamente irrelevantes).

Muestreo aleatorio simple Los métodos estadísticos inferenciales imponen ciertas restricciones sobre el muestreo para asegurar que los métodos estadísticos inferenciales para el análisis de los datos lleven a conclusiones correctas. Además de la distinción entre muestreos sin y con reposición, es importante la siguiente propiedad de muestreos:

Muestreo aleatorio simple

Al extraer una muestra de tamaño n de una población, el muestreo es *aleatorio* si las reglas garantizan que cada subconjunto ordenado de n elementos de la población tiene la misma probabilidad de ser la muestra extraída.

Unas notas respecto de la definición de muestreo aleatorio simple

Nota 1: La definición anterior utiliza el concepto de “probabilidad” todavía de forma intuitiva. En la Sección 9.6.1 se definirá formalmente.

Nota 2: Aunque no todos los muestreos aleatorios son simples, es común omitir el adjetivo “simple”. Al hablar de un “muestreo aleatorio” sin más, se supone que se trata de un muestreo aleatorio simple.

Nota 3: También es común hablar de una "*muestra* aleatoria". No obstante, la definición anterior desprende claramente que la aleatoriedad es una propiedad del muestreo, es decir del *procedimiento* utilizado para extraer la muestra. Observando las características de una muestra, es imposible saber (con certeza) si se ha utilizado o no un procedimiento aleatorio para su extracción. En principio, se debe hablar siempre de "muestras extraídas mediante un proceso aleatorio" en vez de "muestras aleatorias". En la práctica, sin embargo, las dos expresiones se usan como sinónimos.

Ejemplo

Volvamos a considerar la extracción de una muestra de tamaño $n = 2$ de la siguiente población con $N = 5$ elementos:

$$\{\text{Adrián}, \text{Carlos}, \text{Diana}, \text{Lizz}, \text{Oliver}\}\,.$$

- Las posibles muestras de 2 elementos, para el caso de un muestreo con reposición, se enumeran al continuación:

$$\begin{array}{rrrrr}
(\text{Adrián}, \text{Adrián}) & (\text{Adrián}, \text{Carlos}) & (\text{Adrián}, \text{Diana}) & (\text{Adrián}, \text{Lizz}) & (\text{Adrián}, \text{Oliver}) \\
(\text{Carlos}, \text{Adrián}) & (\text{Carlos}, \text{Carlos}) & (\text{Carlos}, \text{Diana}) & (\text{Carlos}, \text{Lizz}) & (\text{Carlos}, \text{Oliver}) \\
(\text{Diana}, \text{Adrián}) & (\text{Diana}, \text{Carlos}) & (\text{Diana}, \text{Diana}) & (\text{Diana}, \text{Lizz}) & (\text{Diana}, \text{Oliver}) \\
(\text{Lizz}, \text{Adrián}) & (\text{Lizz}, \text{Carlos}) & (\text{Lizz}, \text{Diana}) & (\text{Lizz}, \text{Lizz}) & (\text{Lizz}, \text{Oliver}) \\
(\text{Oliver}, \text{Adrián}) & (\text{Oliver}, \text{Carlos}) & (\text{Oliver}, \text{Diana}) & (\text{Oliver}, \text{Lizz}) & (\text{Oliver}, \text{Oliver})
\end{array}$$

  Si el procedimiento utilizado para el muestreo garantiza que estas 25 muestras tienen la misma probabilidad (de $\frac{1}{25}$) de ser la muestra extraída en una ejecución del experimento aleatorio, entonces, por definición, es un muestreo aleatorio simple. En caso contrario, no lo es.

- Para el caso de un muestreo sin reposición, las posibles muestras son 20:

$$\begin{array}{rrrr}
(\text{Adrián}, \text{Carlos}) & (\text{Adrián}, \text{Diana}) & (\text{Adrián}, \text{Lizz}) & (\text{Adrián}, \text{Oliver}) \\
(\text{Carlos}, \text{Adrián}) & (\text{Carlos}, \text{Diana}) & (\text{Carlos}, \text{Lizz}) & (\text{Carlos}, \text{Oliver}) \\
(\text{Diana}, \text{Adrián}) & (\text{Diana}, \text{Carlos}) & (\text{Diana}, \text{Lizz}) & (\text{Diana}, \text{Oliver}) \\
(\text{Lizz}, \text{Adrián}) & (\text{Lizz}, \text{Carlos}) & (\text{Lizz}, \text{Diana}) & (\text{Lizz}, \text{Oliver}) \\
(\text{Oliver}, \text{Adrián}) & (\text{Oliver}, \text{Carlos}) & (\text{Oliver}, \text{Diana}) & (\text{Oliver}, \text{Lizz})
\end{array}$$

  Si y solo si el muestreo garantiza que cada una de estas 20 muestras tienen la misma probabilidad (de $\frac{1}{20}$) de ser la muestra extraída, entonces es aleatorio simple.

Por fines ilustravos, en este ejemplo se utilizaron valores pequeños para N y n. Sin embargo, para poblaciones y muestras más grandes, el principio de un muestreo aleatorio sigue siendo el mismo, aunque el número de posibles muestras que se puedan sacar de la población sea muy grande.

El muestreo aleatorio simple resulta muy importante en la estadística, ya que permite hacer inferencias sobre la población de forma (relativamente) fácil. Una pregunta que probablemente surge después de la definición y el ejemplo anterior es, *cómo* se puede garantizar que cada muestra tenga la misma probabilidad de ser la muestra extraída. Habitualmente se hace uso de tablas con números aleatorios o de algoritmos (implementados en una computadora) que generan números aleatorios. Por ejemplo, para sacar, de forma aleatoria simple y sin reposición, una muestra de 10 elementos de la población de 40 grupos de estudiantes en el primer año de la carrera de medicina, se asigna un número distinto entre 1 y 40 a cada grupo y se utiliza un programa para que genere aleatoriamente 10 números entre 1 y 40 sin repeticiones. Los grupos con los 10 números generados se incluyen en la muestra.

Muestras representativas A veces, un investigador desea que la muestra se parezca a la población en ciertas características (que más le interesan). Es en este contexto que surge el concepto de una muestra representativa.

Muestra representativa

Una muestra es *representativa para alguna característica* de interés, si la diversidad respecto de esta característica en la muestra es la misma que en la población.

Esta definición no es muy precisa en el uso del término "diversidad". Se ilustra el concepto con unos ejemplos.

Ejemplos

Consideramos la siguiente población de seis personas:

$$\{\text{Adrián}, \text{Carlos}, \text{Diana}, \text{Jesús}, \text{Leo}, \text{Lizz}, \text{Oliver}, \text{Olivia}, \text{Uri}\}\,.$$

La muestra

$$(\text{Diana}, \text{Adrián}, \text{Leo})$$

es representativa para la característica *Sexo*, porque la representación de hombres y de mujeres es 67 %-33 %, tanto en la población total como en esta muestra.

Un candidato para la presidencia que quiere conocer el porcentaje de votantes empadronados que votaría por él en las siguientes elecciones saca una muestra de 1000 personas. Sabe que su popularidad es muy distinto en diferentes estratos socioeconómicos de la población: su programa socialista atrae más a los votantes de niveles socioeconómicos bajos. En el sitio web del Instituto Nacional de Estadística y Geografía (INEGI), encontró un estudio que divide la población en cuatro estratos socioeconómicos: muy bajo (30 %), bajo (35 %), medio (25 %) y alto (10 %). Puesto que el nivel socioeconómico es un factor determinante para su popularidad, utiliza un muestreo que garantiza que la muestra obtenida refleje perfectamente la distribución de nivel socioeconómico de la población. Es decir, la muestra tendrá 300 personas de nivel socioeconómico muy bajo, 350 de nivel socioeconómico bajo, 250 de nivel socioeconómico medio, y 100 de nivel socioeconómico alto.

Unas notas respecto de la definición de muestras representativas

Nota 1: Representatividad siempre refiere a una característica específica. Es posible que una muestra sea representativa para una característica y para otra no. En el primer ejemplo del bloque anterior, se concluyó que la muestra era representativa para *Sexo*; sin embargo, probablemente no lo es para la característica *Estado civil*.

Nota 2: Un creencia común *incorrecta* es que un muestreo aleatorio simple lleva siempre a muestras representativas. Recuérdese que la aleatoriedad es una propiedad del *muestreo*, el procedimiento seguido para extraer la muestra; representatividad, al contrario, es una propiedad del producto del muestreo, la *muestra*. (Esta diferencia se refleja en los títulos de los apartados anteriores: *muestreo* aleatorio simple pero *muestra* representativa.) Representatividad y aleatoriedad son conceptos distintos: Una muestra puede ser representativa, aunque no se extraiga de forma aleatoria; y, al revés, es posible que un muestreo aleatorio simple no lleve a una muestra representativa (para una característica de interés).

Nota 3: Si la representatividad de la muestra respecto de una (o más) característica(s) es muy importante, se puede considerar utilizar un *muestreo aleatorio estratificado*. Este tipo de muestreo garantiza que la muestra es perfectamente representativa respecto de una(s) característica(s) eligida(s). El procedimiento se mostró en el último ejemplo: Primero, se realiza una división previa de la población en distintos grupos homogéneos (es decir, estratos) y, segundo, se extrae de cada estrato una muestra aleatoria cuyo tamaño es proporcional a la representación del estrato en la población. En el ejemplo, al obtener una muestra total de 1000 personas, se extrae a través de un muestreo aleatorio simple (a) una muestra de 300 personas de la subpoblación de nivel socioeconómico muy bajo así que la muestra final tenga 30 % de miembros de este estrato, el mismo porcentaje que corresponde a este estrato en la población; según la misma lógica, se extraen (b) una muestra de 350 personas del estrato socioeconómico bajo, (c) una muestra de 250 personas de nivel socioconómico medio y (d) una muestra de 100 personas del estrato socioeconómico alto. Nótese que el muestreo aleatorio estratificado no es aleatorio simple en el sentido que no todas las muestras tienen la misma probabilidad de ser la extraída (para sacar la muestra dentro de cada estrato, el muestreo sí es aleatorio simple, pero la muestra en su totalidad no lo es). Cabe señalar que, si bien es cierto que el muestreo aleatorio estratificado también permite llegar a inferencias válidas sobre la población, los métodos de análisis generalmente se modifican.

La investigación en las ciencias de la salud rara vez utiliza muestreos aleatorios (simples o estratificados). Por ejemplo, la muestra utilizada en las pruebas científicas de nuevos fármacos casi siempre se extrae de alguna subpoblación. Si en México se realiza un estudio para probar un nuevo fármaco contra el Parkinson, se extraerá una muestra de la subpoblación de pacientes con esta enfermedad que se presentan durante algún periodo (por ejemplo, en el transcurso de este año) a una clínica específica (o un grupo de clínicas, por ejemplo, del sistema IMSS) en el Distrito Federal. Sin embargo, el objetivo generalmente es llegar a conclusiones sobre *todos* los pacientes con Parkinson (también los que se presentarán después o los que van a otras clínicas, incluso para pacientes en otros estados y otros países). Asimismo, un equipo de investigadores que quiere comparar la eficacia de clases presenciales versus a distancia, podría utilizar una muestra de los estudiantes que actualmente cursan una asignatura particular en la Facultad de Medicina de la UNAM, a pesar de que el objetivo es llegar a conclusiones que también sean válidas para futuras generaciones, otras asignaturas e incluso para estudiantes de otras carreras u otras universidades.

Lo anterior ilustra la diferencia entre la *población muestreada* (es decir, la población de la cual se extrae la muestra) y la *población objetivo* (sobre la cual se desea hacer inferencias). La pregunta clave es: si la población muestreada no coincide con la población objetivo, ¿serán válidas las inferencias sobre la población objetivo? Para responder esta pregunta valen las siguientes consideraciones:

Primero, es cierto que los métodos estadísticos sólo permiten concluir que las inferencias son válidas para la población muestreada. Por otro lado, esto no quiere decir que sean inválidas en la población objetivo. Es decir, un muestreo aleatorio de la población objetivo es una condición *suficiente* para llegar a conclusiones correctas sobre esta población; sin embargo, no es una condición *necesaria*: resultados obtenidos por otra vía pueden ser correctos también.

Segundo, la respuesta en la pregunta anterior depende de lo que se examina. Por ejemplo, al examinar el efecto de un nuevo fármaco en pacientes de Parkinson, es plausible suponer que los mecanismos subyacentes responsables para la enfermedad sean los mismos para pacientes de la población muestreada que para pacientes de la población objetivo. En este caso es plausible que las conclusiones se generalicen hasta la población objetivo. En el ejemplo de la comparación de clases presenciales y a distancia, la respuesta no es tan obvia. Podría ser que los dos métodos funcionen de forma diferente con otras asignaturas u otras carreras, o que tengan efectos distintos en futuras generaciones de estudiantes (más acostumbrados, por ejemplo, a ciertos avances tecnológicos), etcétera. Es decir, es más probable que la población muestreada difiera en características relevantes de la población objetivo. Concluyendo estas reflexiones, debe ser claro que la cuestión si la población muestreada se parece lo suficiente o no a la población objetivo refleja—en la ausencia de datos adicionales disponibles—una suposición teórica.

9.3 Resultados y el espacio muestral

Resultado

De toda la información que se genera a partir de la ejecución de un experimento aleatorio, el investigador decide retener una parte para su posterior investigación.
Esta parte se denomina el *resultado* del experimento aleatorio.

Un resultado comúnmente se representa por la minúscula griega omega: ω.

Ejemplos

Considérese el experimento aleatorio "echar una moneda". Aunque habitualmente se define la cara que sale como el resultado del experimento, genera una gran cantidad de información de otro tipo, por ejemplo, la trayectoria exacta por la cual pasa la moneda en el aire y el número de revoluciones que da sobre su eje antes de tocar el suelo, el tiempo total desde el lanzamiento hasta que se detiene, etcétera. Si se define como resultado la cara que sale, entonces hay dos posibles resultados $\omega_1 = \text{sol}$, $\omega_2 = \text{águila}$.

La ejecución del experimento aleatorio "lanzar un dado hasta que salga el número seis" generará información de diferente tipo, por ejemplo, todos los números que salieron en los tiros anteriores a tener el seis, el tiempo total (en segundos) que requirió la ejecución del experimento, la posición del dado en la mesa en cada tiro, etcétera. Es posible que únicamente interesa el número total de tiros; en este caso, esto es el resultado del experimento.

La aplicación de un examen de opción múltiple a un estudiante genera un sinfín de información: la respuesta que dio y el tiempo que tardó al contestar cada pregunta, si dejó preguntas sin contestar, su presión arterial durante el examen, si tuvo que interrumpir el examen para ir al baño, si intentó y/o logró copiar algunas respuestas del examen de su compañero, etcétera. En caso de que solo interese el número de preguntas contestadas correctamente, entonces esto es el resultado.

El experimento aleatorio que consiste en la medición de la presión arterial de una persona mediante un esfigmomanómetro digital no solo proporciona información de su presión arterial sistólica y diastólica, sino también la mímica de la persona, si mantiene una conversación durante la medición, etc.. Comúnmente, se define como resultado el par de números que salen en el aparato.

Espacio muestral

El *espacio muestral* es el conjunto de todos los resultados posibles que pueden producirse en un experimento aleatorio.

Notación: Ω
Sinónimo: Conjunto de resultados

Ejemplos

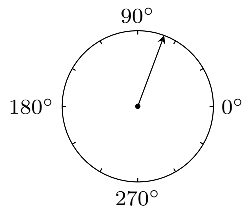
Si en el experimento aleatorio "lanzar una vez una moneda" se define como resultado la cara que sale, entonces:

$$\Omega = \{\text{sol}, \text{águila}\}$$

Si en el experimento aleatorio "lanzar un dado hasta que salga seis" se define como resultado el número total de tiros que se realiza, entonces:

$$\Omega = \{1, 2, 3, 4, \dots\}$$

Considérese la siguiente rueda giratoria



Si se ejecuta el experimento aleatorio "girar la manecilla" y se define como resultado el ángulo bajo el cual se detiene la manecilla, entonces el espacio muestral consiste en todos los números reales entre 0 (incluido) y 360 (no incluido):

$$\Omega \;=\; \{x \in \mathbb{R} \,|\, 0 \leqslant x < 360\}.$$

Para la medición de la presión arterial sistólica por un esfigmomanómetro digital, el conjunto de resultados es (por ejemplo):

$$\Omega \;=\; \{0, 1, 2, 3, 4, \ldots, 299, 300\}$$

Si un examen de opción múltiple consiste en 3 preguntas, con tres opciones de respuesta (A, B y C) cada una, entonces definiendo como resultado el patrón de respuestas de una persona en la aplicación de este examen, se tiene:

$$\begin{aligned}\Omega \;=\; \{&(\mathrm{A},\mathrm{A},\mathrm{A}),\ (\mathrm{A},\mathrm{A},\mathrm{B}),\ (\mathrm{A},\mathrm{A},\mathrm{C}),\ (\mathrm{A},\mathrm{B},\mathrm{A}),\ (\mathrm{A},\mathrm{B},\mathrm{B}),\ (\mathrm{A},\mathrm{B},\mathrm{C}),\ (\mathrm{A},\mathrm{C},\mathrm{A}),\ (\mathrm{A},\mathrm{C},\mathrm{B}),\\ &(\mathrm{A},\mathrm{C},\mathrm{C}),\ (\mathrm{B},\mathrm{A},\mathrm{A}),\ (\mathrm{B},\mathrm{A},\mathrm{B}),\ (\mathrm{B},\mathrm{A},\mathrm{C}),\ (\mathrm{B},\mathrm{B},\mathrm{A}),\ (\mathrm{B},\mathrm{B},\mathrm{B}),\ (\mathrm{B},\mathrm{B},\mathrm{C}),\ (\mathrm{B},\mathrm{C},\mathrm{A}),\\ &(\mathrm{B},\mathrm{C},\mathrm{B}),\ (\mathrm{B},\mathrm{C},\mathrm{C}),\ (\mathrm{C},\mathrm{A},\mathrm{A}),\ (\mathrm{C},\mathrm{A},\mathrm{B}),\ (\mathrm{C},\mathrm{A},\mathrm{C}),\ (\mathrm{C},\mathrm{B},\mathrm{A}),\ (\mathrm{C},\mathrm{B},\mathrm{B}),\ (\mathrm{C},\mathrm{B},\mathrm{C}),\\ &(\mathrm{C},\mathrm{C},\mathrm{A}),\ (\mathrm{C},\mathrm{C},\mathrm{B}),\ (\mathrm{C},\mathrm{C},\mathrm{C})\}.\end{aligned}$$

(Se supuso que no se pueden dejar preguntas sin contestar.)
Obviamente, en el caso de más preguntas y más opciones de respuesta, el número de elementos en Ω sería más grande (pero siempre finito).

Un investigador lleva a cabo el siguiente experimento sobre el efecto del alcohol en el tiempo de reacción de conductores de coches. En un simulador, una persona con un nivel de 0.05g/dL de alcohol en la sangre debe frenar lo más pronto posible a un obstáculo imprevisto. Define como resultado del experimento el tiempo (en segundos) hasta presionar el freno. En este caso:

$$\Omega \;=\; \{x \in \mathbb{R} \,|\, 0 \leqslant x \leqslant 5\} \,\cup\, \{\text{sin frenar}\}.$$

(Si no frena o si el tiempo es mayor de 5 segundos se asigna el resultado "sin frenar".)

Unas notas respecto de la definición de un resultado y el espacio muestral

Nota 1: Los ejemplos anteriores muestran que lo que se considera como el resultado del experimento aleatorio es decisión del investigador. Dos investigadores pueden estar interesados en diferentes aspectos del mismo experimento aleatorio y, por lo tanto, llegan a definir otros resultados. En el ejemplo del exámen de opción múltiple, por ejemplo, de un examen departamental, a los profesores de la asignatura a lo mejor les interesa únicamente la calificación total (número de respuestas correctas), mientras que una investigadora, con el fin de investigar si es más válido un exámen donde las preguntas tienen tres, o bien, cuatro opciones de respuesta, necesita información sobre las respuestas en cada pregunta.

Nota 2: El espacio muestral Ω puede ser finito, infinito numerable, o bien, innumerable. En los ejemplos del bloque anterior, Ω es finito en el primero, el cuarto y el quinto; infinito numerable en el segundo; e innumerable en el tercero y el sexto ejemplo.

Nota 3: Espacios muestrales infinitos (numerables o innumerables) son conceptualizaciones teóricas. En la práctica el conjunto de resultados siempre es finito. Por ejemplo, al ejecutar el experimento aleatorio "tirar un dado hasta que salga seis", si después de 1000 tiros todavía no se observase un seis, probablemente se terminaría el experimento (y obtener un resultado como un millón, aunque teóricamente posible, en la práctica nunca se alcanzará). Asimismo, la precisión de mediciones como peso, tiempo, temperatura e índice de masa corporal, en la práctica es limitada y siempre el número de resultados posibles es finito. Sin embargo, en la estadística resulta para este tipo de resultados más fácil considerar espacios muestrales infinitos.

Nota 4: En una nota a la definición del experimento aleatorio (en la página 284) se aclaró que (casi siempre) el tamaño de la muestra n se considera fijo y es parte del experimento aleatorio. Diferentes valores para n corresponden no sólo con diferentes experimentos aleatorios, sino también con diferentes espacios muestrales. Para explicitar el tamaño de la muestra asociada con el espacio muestral, se puede añadir un subíndice a Ω: por ejemplo, $\Omega_{n=1}$, $\Omega_{n=10}$, etc.

Ejemplo

Para ilustrar los diferentes espacios muestrales asociados con diferentes tamaños de muestras (mencionados en la Nota 4 anterior), reconsideremos el experimento aleatorio del primer ejemplo en la página 291. Para $n = 1$, el espacio muestral es:

$$\Omega_{n=1} = \{\text{sol}, \text{águila}\}.$$

Para el experimento aleatorio análogo con $n = 3$ (lanzar tres veces una moneda), se obtiene:

$$\begin{aligned}\Omega_{n=3} &= \{\text{sol}, \text{águila}\} \times \{\text{sol}, \text{águila}\} \times \{\text{sol}, \text{águila}\} \\ &= \{(\text{sol}, \text{sol}, \text{sol}), (\text{sol}, \text{sol}, \text{águila}), (\text{sol}, \text{águila}, \text{sol}), (\text{sol}, \text{águila}, \text{águila}), \\ &\quad\ (\text{águila}, \text{sol}, \text{sol}), (\text{águila}, \text{sol}, \text{águila}), (\text{águila}, \text{águila}, \text{sol}), (\text{águila}, \text{águila}, \text{águila})\},\end{aligned}$$

donde los tres elementos de cada triplete en $\Omega_{n=3}$ se refieren a la cara que salió en los tres lanzamientos, respectivamente.

9.4 Variables aleatorias

Variable aleatoria

Una *variable aleatoria* X es una función con dominio el espacio muestral de un experimento aleatorio y codominio el conjunto de números reales $\mathbb{R}$:

$$\begin{aligned} X : \Omega &\to \mathbb{R} \\ \omega &\mapsto X(\omega). \end{aligned}$$

Esta definición dice nada más que una variable aleatoria asigna un número a cada resultado que el experimento aleatorio puede producir.

Unas notas respecto de la definición de un resultado y una variable aleatoria

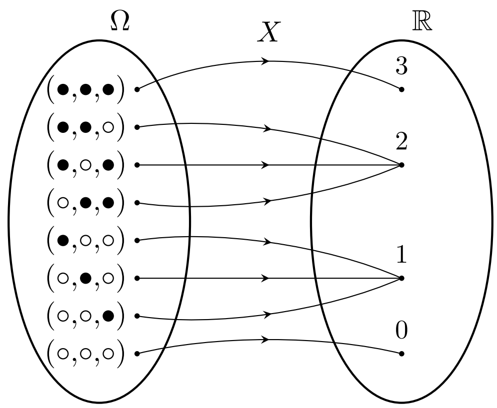
Nota 1: Los números que se asignan a los resultados son los *valores* de la variable y el recorrido de la función (es decir, el conjunto de valores que son la imagen de, al menos, uno de los resultados en Ω, véase la página 12) se llama el *espacio de valores* de X.

Nota 2: Si el espacio de valores de una variable aleatoria es finito o infinito numerable, entonces se trata de una variable *discreta*.
Si el número de posibles valores es innumerable, la variable aleatoria se caracteriza como *continua*.

Nota 3: En la Nota 3 de la página 293, se señaló que espacios muestrales en la práctica siempre son finitos (debido a la limitación por el tiempo y/o porque la precisión de mediciones tiene un límite). Por la misma razón, el espacio de valores de variables aleatorios en la práctica es finito también, así que en principio bastaría definir únicamente variables aleatorias discretas. Sin embargo, la conceptualización teórica en términos de variables continuas simplifica en muchas ocasiones el análisis estadístico.

Ejemplo

Sobre el espacio muestral en el último ejemplo de la sección anterior, que consiste en los resultados del experimento aleatorio de lanzar tres veces una moneda, se puede definir una variable aleatoria X que indica el número de veces que salió águila. Si se representan "águila" y "sol" con bolitas negras y blancas, respectivamente, la variable X se puede visualizar como sigue:



El espacio de valores de X es $\{0, 1, 2, 3\}$; es un conjunto finito, por lo cual esta variable aleatoria es discreta.

Volvamos a considerar el experimento aleatorio que consiste en aplicar un examen de opción múltiple, para el cual se define el espacio muestral como todos los posibles patrones de respuesta (similar al ejemplo en la p. 292). Supongamos que se aplicaron 330 preguntas, con 3 opciones de respuesta cada uno. Una variable aleatoria muy común que se define sobre este espacio muestral es el número de respuestas correctas. Es decir, esta variable asigna a cada posible patrón de respuestas el número de respuestas correctas que contiene (el cual muchas veces corresponde con la calificación en el examen). Claramente, X es una variable discreta ya que el espacio de valores de X, siendo el conjunto de números naturales hasta 330 ($\{0, 1, 2, \ldots, 329, 330\}$), es finito.

Si los resultados en el espacio muestral ya son valores numéricos, la variable aleatoria a veces asigna el mismo número que representa el resultado. Es decir, el espacio de valores de la variable coincide con el espacio muestral. Por ejemplo, en el experimento aleatorio "lanzar un dado hasta que salga seis" con el espacio muestral el número de lanzamientos en total (véase el ejemplo en la página 291), una variable aleatoria puede asignar a cada resultado este mismo número, asi que el espacio de valores de X es el mismo que Ω (lo cual implica que X es discreta, con un número infinito numerable de valores).

En el mismo orden de ideas, se puede definir una variable aleatoria X continua sobre el espacio muestral que corresponde con el experimento aleatorio con la rueda giratoria (véase, la página 292), donde los valores de la variable corresponden precisamente con el ángulo bajo el cual la manecilla se detiene.

9.5 Eventos

Evento

Un *evento* E es un subconjunto del espacio muestral Ω. En símbolos:

$$E \subseteq \Omega.$$

Ejemplo

Reconsidérese el espacio muestral asociado con la aplicación de un examen de opción múltiple de tres preguntas, con tres opciones de respuesta cada una (véase la página 292) y supóngase que las respuestas correctas en las tres preguntas son B, B, y C, respectivamente. Se pueden definir los siguientes eventos (entre otros):

- "Seleccionar la opción de la misma letra en cada pregunta", lo cual corresponde con:

$$E_1 = \{(\mathrm{A},\mathrm{A},\mathrm{A}),\ (\mathrm{B},\mathrm{B},\mathrm{B}),\ (\mathrm{C},\mathrm{C},\mathrm{C})\}.$$

- "Dar la respuesta correcta en las tres preguntas":

$$E_2 = \{(\mathrm{B},\mathrm{B},\mathrm{C})\}.$$

- "No dar ninguna respuesta correcta":

$$\begin{aligned} E_3 = \{&(\mathrm{A},\mathrm{A},\mathrm{A}),\ (\mathrm{A},\mathrm{A},\mathrm{B}),\ (\mathrm{A},\mathrm{C},\mathrm{A}),\ (\mathrm{A},\mathrm{C},\mathrm{B}),\ (\mathrm{C},\mathrm{A},\mathrm{A}),\ (\mathrm{C},\mathrm{A},\mathrm{B}),\ (\mathrm{C},\mathrm{C},\mathrm{A}),\\ &(\mathrm{C},\mathrm{C},\mathrm{B})\}.\end{aligned}$$

- "Aprobar el examen" (donde se aprueba si se tiene más del 60 %, o bien, al menos 2 respuestas correctas):

$$E_4 = \{(\mathrm{A},\mathrm{B},\mathrm{C}),\ (\mathrm{B},\mathrm{A},\mathrm{C}),\ (\mathrm{B},\mathrm{B},\mathrm{A}),\ (\mathrm{B},\mathrm{B},\mathrm{B}),\ (\mathrm{B},\mathrm{B},\mathrm{C}),\ (\mathrm{B},\mathrm{C},\mathrm{C}),\ (\mathrm{C},\mathrm{B},\mathrm{C})\}.$$

Unas notas respecto de la definición de un evento

Nota 1: Se dice que un evento *se realiza* si la ejecución del experimento aleatorio produce un resultado que pertenece al evento.

Nota 2: Algunos eventos importantes recibieron un nombre especial:

- $\varnothing$: el *evento imposible*
- Ω: el *evento seguro* (nótese que $\Omega \subseteq \Omega$ y, entonces, el mismo Ω también es un evento)
- para cualquier $\omega \in \Omega$, se llama el evento $\{\omega\}$ un *evento elemental*. En otras palabras, los eventos elementales son los eventos que contienen solo un resultado.

Nota 3: Como se mostrará en el siguiente tema, resultarán muy importantes los eventos que se derivan de variables aleatorias. Es decir, se considerarán eventos del tipo:

$$E = \{\omega \in \Omega \,|\, X(\omega) = x\},$$

donde x es cualquier valor de X. En palabras, este evento incluye todos los resultados a los cuales la variable X asigna el valor x.

Nota 4: Los eventos son conjuntos y, por lo tanto, se pueden aplicar a los eventos todas las operaciones definidas en conjuntos, como intersección, unión y diferencia. Un evento que se obtiene por la aplicación de alguna(s) operación(es) en uno o más eventos, se llama un *evento compuesto*.

Nota 5: Una operación especial es tomar el *complemento* (o *complementario*) de un evento E, el cual se anota por E^c y se define como:

$$E^c \;=\; \Omega \;\setminus\; E.$$

Es decir, el complemento del evento E contiene todos los resultados en el espacio muestral que no están en E.

Nota 6: Dos eventos cuya intersección es vacía se llaman *exclusivos*. Si considerando tres o más eventos la intersección en cualquier par es vacía, se dice que estos eventos son *mutuamente exclusivos*. Es claro que, si dos eventos son exclusivos, es imposible que en una ejecución del experimento aleatorio ambos se realicen (ya que, por definición, la intersección de dos eventos exclusivos es el evento imposible). Por lo tanto, se habla también de eventos *incompatibles* como sinónimos de eventos exclusivos.

Nota 7: Dos o más eventos cuya unión es igual a Ω son *colectivamente exhaustivos*. En general, $E_1, E_2, \ldots, E_m$ son m eventos colectivamente exhaustivos si:

$$E_1 \,\cup\, E_2 \,\cup \cdots \cup\, E_m \;=\; \Omega.$$

La ejecución del experimento aleatorio siempre produce un resultado que se encuentra en uno (o más) de los eventos que son colectivamente exhaustivos.

Ejemplos

Se ilustran los conceptos en las notas anteriores utilizando los eventos E_1, E_2, E_3 y E_4 del ejemplo anterior:

- Si la ejecución del experimento aleatorio produce como resultado el patrón de respuestas (A,A,A), entonces los eventos E_1 y E_3 se han realizados y los eventos E_2 y E_4 no.
- El evento E_2 en el ejemplo anterior es un evento elemental. Sólo contiene el resultado (B,B,C). Los eventos E_1, E_3, y E_4 no son elementales.

- Considérese la variable aleatoria X que asigna a cada resultado una calificación en términos del número de preguntas contestadas correctamente. Entonces, los eventos E_2 y E_3 se describen en términos de la variable X como sigue:

$$\begin{aligned} E_2 &= \{\omega \in \Omega \mid X(\omega) = 3\} \\ E_3 &= \{\omega \in \Omega \mid X(\omega) = 0\}. \end{aligned}$$

  Corresponden con el conjunto de todos los patrones de respuesta que corresponden con una calificación de 3 y 0, respectivamente. También el evento E_4 se puede definir en términos de la variable X:

$$E_4 = \{\omega \in \Omega \mid X(\omega) \geqslant 2\},$$

  es decir, el conjunto de los patrones de respuesta que corresponden con una calificación de 2 o más.

- La intersección de los eventos E_1 y E_4 se da por:

$$E_1 \cap E_4 = \{(\mathrm{B}, \mathrm{B}, \mathrm{B})\}.$$

  La intersección de estos eventos es un evento elemental; corresponde con “seleccionar la opción de la misma letra en cada pregunta **y** aprobar el examen”.

- La unión de E_1 y E_4 es el siguiente evento:

$$\begin{aligned} E_1 \cup E_4 = \{&(\mathrm{A},\mathrm{A},\mathrm{A}),\ (\mathrm{A},\mathrm{B},\mathrm{C}),\ (\mathrm{B},\mathrm{A},\mathrm{C}),\ (\mathrm{B},\mathrm{B},\mathrm{A}),\ (\mathrm{B},\mathrm{B},\mathrm{B}),\ (\mathrm{B},\mathrm{B},\mathrm{C}), \\ &(\mathrm{B},\mathrm{C},\mathrm{C}),\ (\mathrm{C},\mathrm{B},\mathrm{C}),\ (\mathrm{C},\mathrm{C},\mathrm{C})\}, \end{aligned}$$

  lo cual corresponde con “seleccionar la opción de la misma letra en cada pregunta **o** aprobar el examen”.

- El complemento del evento E_2 es “dar una o más respuestas correctas” y coincide con el conjunto:

$$\begin{aligned} E_2^c = \{&(\mathrm{A},\mathrm{A},\mathrm{C}),\ (\mathrm{A},\mathrm{B},\mathrm{A}),\ (\mathrm{A},\mathrm{B},\mathrm{B}),\ (\mathrm{A},\mathrm{B},\mathrm{C}),\ (\mathrm{A},\mathrm{C},\mathrm{C}),\ (\mathrm{B},\mathrm{A},\mathrm{A}),\ (\mathrm{B},\mathrm{A},\mathrm{B}), \\ &(\mathrm{B},\mathrm{A},\mathrm{C}),\ (\mathrm{B},\mathrm{B},\mathrm{A}),\ (\mathrm{B},\mathrm{B},\mathrm{B}),\ (\mathrm{B},\mathrm{B},\mathrm{C}),\ (\mathrm{B},\mathrm{C},\mathrm{A}),\ (\mathrm{B},\mathrm{C},\mathrm{B}),\ (\mathrm{B},\mathrm{C},\mathrm{C}), \\ &(\mathrm{C},\mathrm{A},\mathrm{C}),\ (\mathrm{C},\mathrm{B},\mathrm{A}),\ (\mathrm{C},\mathrm{B},\mathrm{B}),\ (\mathrm{C},\mathrm{B},\mathrm{C}),\ (\mathrm{C},\mathrm{C},\mathrm{C})\}. \end{aligned}$$

- Los eventos E_2, E_3 y E_4 son mutuamente exclusivos ya que:

$$E_2 \cap E_3 = \varnothing \qquad E_2 \cap E_4 = \varnothing \qquad E_3 \cap E_4 = \varnothing$$

- Nótese que un evento y su complemento siempre son exclusivos y colectivamente exhaustivos ya que

$$E \cap E^c = \varnothing \quad \text{y} \quad E \cup E^c = \Omega.$$

9.6 Probabilidad

El concepto de probabilidad es uno de los más importantes en la estadística inferencial. En esta sección, primero se da una definición formal de probabilidad y se discuten varias formas en que se suele interpretar este concepto (Sección 9.6.1); a continuación se explica, mediante una serie de ejemplos, cómo se pueden calcular probabilidades (Sección 9.6.2); en las Secciones 9.6.3 y 9.6.4 se introducen el concepto de probabilidad condicional y una aplicación importante del mismo en el teorema de Bayes.

9.6.1 Definición e interpretación

Función de probabilidad P

Sea Ω el espacio muestral de un experimento aleatorio, sea $\mathscr{C}$ el conjunto de todos los eventos dentro de Ω, y sea P una función que tiene como dominio $\mathscr{C}$ y como codominio el intervalo [0,1], es decir:

$$\begin{aligned} P : \mathscr{C} &\to [0,1] \\ E &\mapsto P(E). \end{aligned}$$

Entonces, P es una *función de probabilidad* y $P(E)$ es la *probabilidad* del evento $E \in \mathscr{C}$, si y sólo si satisface las siguientes condiciones:

(1) $P(\Omega) = 1$ (9.1)

(2) Para cualquier par de eventos *exclusivos* E_1 y E_2 (elementos de $\mathscr{C}$), se cumple:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2). \tag{9.2a}$$

En lenguaje común, esta definición dice que la función de probabilidad P asigna a cada evento E una probabilidad (que es un número entre 0 y 1). Para que P sea una función de probabilidad válida, debe cumplir dos requisitos: (1) la probabilidad del evento seguro es 1 y (2) si dos eventos son exclusivos, entonces la probabilidad de que se realice uno u otro es igual a la suma de las probabilidades de que se realice cada evento por separado.

Nota: probabilidad de un resultado

La definición de probabilidad es en términos de *eventos*, en el sentido de que la función de probabilidad asigna una probabilidad a cada *evento*. Sin embargo, comúnmente se habla también de la probabilidad de un *resultado*. Lo que se quiere decir en este caso es la probabilidad del evento elemental que consiste en este resultado. Es decir, para un resultado ω, si se escribe

$$P(\omega)$$

se refiere a

$$P(\{\omega\}).$$

Algunas propiedades de la función de probabilidad

De la definición de la función de probabilidad, siguen las siguientes propiedades:

- Si $E_1, E_2, \ldots, E_m$ son eventos *mutuamente exclusivos* del espacio muestral, entonces:

$$P(E_1 \cup E_2 \cup \ldots \cup E_m) = P(E_1) + P(E_2) + \ldots + P(E_m). \tag{9.2b}$$

- Probabilidad de un evento complementario:

$$P(E^c) = 1 - P(E)$$

Nótese que la primera propiedad en este cuadro generaliza la Ecuación (9.2a) a más de dos eventos. Esta propiedad se conoce como la *regla de la suma* en la teoría de la probabilidad. Al respecto, es importante tener claro que esta regla sólo aplica a eventos *mutuamente exclusivos*.

Prueba de la Regla de la suma y la Probabilidad de un evento complementario

- Se demuestra la primera propiedad utilizando inducción matemática. En general, este tipo de demostración consiste en dos pasos. Para el caso actual, son los siguientes:

Paso 1 (base): *La propiedad es cierta para* $m = 2$.
Para $m = 2$, la propiedad es:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2).$$

Es cierto por definición, según la Ecuación (9.2a) en la definición de la función de probabilidad.

Paso 2 (inducción) *Para cualquier* $m \geqslant 3$, si *la propiedad es cierta para* $m - 1$, entonces *también lo es para* m.
Definamos primero el evento A, como la unión de los eventos $E_1, E_2, \ldots, E_{m-1}$:

$$A = E_1 \cup E_2 \cup \ldots \cup E_{m-1},$$

así que:

$$\begin{aligned} P(E_1 \cup E_2 \cup \ldots \cup E_{m-1} \cup E_m) &= P(A \cup E_m). \\ &= P(A) + P(E_m), \end{aligned}$$

puesto que A y E_m son eventos exclusivos en el espacio muestral Ω y tras aplicar la Ecuación (9.2a) a A y E_m.
Si la propiedad es cierta para $m - 1$, entonces:

$$P(A) = P(E_1 \cup E_2 \cup \ldots \cup E_{m-1}) = P(E_1) + P(E_2) + \ldots + P(E_{m-1}),$$

lo cual, combinado con el resultado anterior, lleva a:

$$\begin{aligned} P(E_1 \cup E_2 \cup \ldots \cup E_{m-1} \cup E_m) &= P(A) + P(E_m). \\ &= P(E_1) + P(E_2) + \ldots + P(E_{m-1}) + P(E_m). \quad \blacksquare \end{aligned}$$

- Para la demostración de la segunda propiedad, recuérdese que por la definición de un evento complementario, se tiene que

$$E \cup E^c = \Omega.$$

Dado que E y su complemento E^c son eventos exclusivos, aplica la Ecuación (9.2a), lo cual, en combinación con la Ecuación (9.1) lleva a:

$$\begin{aligned} & P(E \cup E^c) &= P(\Omega) \\ \Longleftrightarrow \quad & P(E) + P(E^c) &= 1 \\ \Longleftrightarrow \quad & P(E^c) &= 1 - P(E). \end{aligned}$$ ∎

Considerando que el evento imposible es el complemento del evento seguro y que, por la Ecuación (9.1), la probabilidad del evento seguro es 1, tenemos directamente el siguiente resultado:

Corolario: Probabilidad del evento imposible

$$P(\varnothing) = 0.$$

Ejemplos

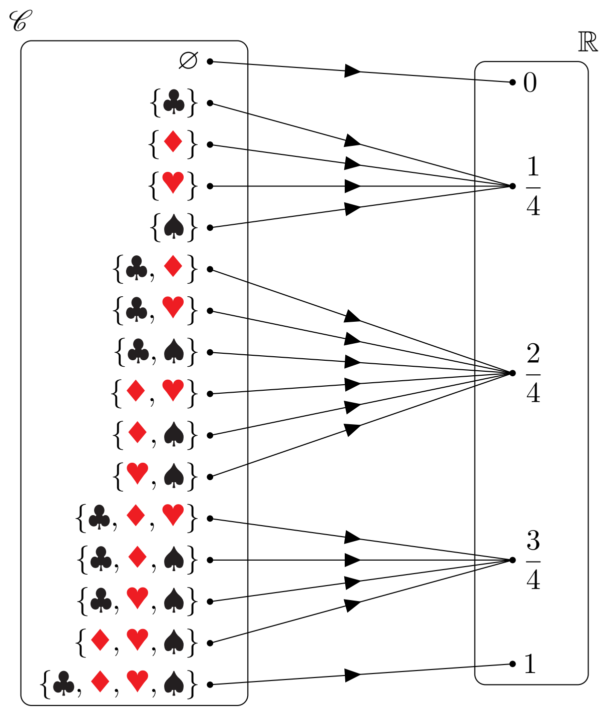
Consideremos el experimento aleatorio que consiste en sacar, de forma aleatoria, una carta de una baraja de 52 cartas y definamos como resultado el palo de la carta extraída. Por lo tanto, el espacio muestral es:

$$\Omega = \{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}.$$

El conjunto $\mathscr{C}$ contiene todos los posibles eventos en el espacio muestral, es decir, contiene todos los subconjuntos posibles de Ω:

$$\begin{aligned} \mathscr{C} = \Big\{ & \varnothing, \{\clubsuit\}, \{\diamondsuit\}, \{\heartsuit\}, \{\spadesuit\}, \{\clubsuit, \diamondsuit\}, \{\clubsuit, \heartsuit\}, \{\clubsuit, \spadesuit\}, \{\diamondsuit, \heartsuit\}, \{\diamondsuit, \spadesuit\}, \{\heartsuit, \spadesuit\}, \\ & \{\clubsuit, \diamondsuit \heartsuit\}, \{\clubsuit, \diamondsuit, \spadesuit\}, \{\clubsuit, \heartsuit, \spadesuit\}, \{\diamondsuit, \heartsuit, \spadesuit\}, \{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\} \Big\} \end{aligned}$$

La función de probabilidad P asigna a cada elemento de $\mathscr{C}$ (es decir, a cada evento en Ω) un número entre 0 y 1 (que se interpreta como la *probabilidad de que este evento se realice* en una ejecución de experimento aleatorio). La siguiente figura muestra una función de probabilidad válida para este ejemplo:



Se trata efectivamente de una función de probabilidad válida ya que las dos condiciones en la definición se cumplen. Por un lado, el evento seguro tiene una probabilidad de 1, ya que $P\left(\{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}\right) = 1$ y por otro lado, se tiene para cualquier par de eventos exclusivos que la probabilidad de que ocurra uno u otro es igual a la probabilidad de cada evento por separado. Por ejemplo:

$$P(\{\clubsuit\} \cup \{\spadesuit\}) \;=\; P(\{\clubsuit, \spadesuit\}) \;=\; \frac{2}{4} \;=\; \underbrace{P(\{\clubsuit\})}_{=\frac{1}{4}} + \underbrace{P(\{\spadesuit\})}_{=\frac{1}{4}}$$

$$P(\{\clubsuit, \diamondsuit\} \cup \{\heartsuit\}) \;=\; P(\{\clubsuit, \diamondsuit, \heartsuit\}) \;=\; \frac{3}{4} \;=\; \underbrace{P(\{\clubsuit, \diamondsuit\})}_{=\frac{2}{4}} + \underbrace{P(\{\heartsuit\})}_{=\frac{1}{4}}.$$

Nótese que también se cumplen las propiedades (p. 300) y el corolario (p. 301) que se derivan de la definición de la función de probabilidad. Por ejemplo, los tres eventos mutuamente exclusivos $\{\clubsuit\}$, $\{\heartsuit, \spadesuit\}$, $\{\diamondsuit\}$ satisfacen:

$$P(\{\clubsuit\} \cup \{\heartsuit, \spadesuit\} \cup \{\diamondsuit\}) \;=\; P(\{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}) \;=\; 1 \;=\; \underbrace{P(\{\clubsuit\})}_{=\frac{1}{4}} + \underbrace{P(\{\heartsuit, \spadesuit\})}_{=\frac{2}{4}} + \underbrace{P(\{\diamondsuit\})}_{=\frac{1}{4}}.$$

y, para el evento $\{\clubsuit, \heartsuit, \spadesuit\}$ y su complemento $\{\diamondsuit\}$, se tiene:

$$P(\{\diamondsuit\}) \;=\; \frac{1}{4} \;=\; 1 - \underbrace{P(\{\clubsuit, \heartsuit, \spadesuit\})}_{=\frac{3}{4}}.$$

Considérese el experimento aleatorio que consiste en aplicar a un alumno un cuestionario de 5 preguntas para evaluar a su docente de estadística y las clases que impartió; se define como resultado el patrón de respuestas en estas cinco preguntas. Si cada pregunta se contesta eligiendo una de las opciones "Totalmente en desacuerdo", "En desacuerdo", "De acuerdo" y "Totalmente de acuerdo" (y no se permite que el alumno dé múltiples respuestas a una pregunta y/o que deje preguntas sin contestar), entonces el espacio muestral contiene un total de $4^5 = 1024$ elementos (es decir, hay 1024 distintos patrones de respuesta, véase la siguiente sección para una justificación de este número).
Si se representan las opciones de respuesta por las siguientes letras:

$$\begin{aligned} \text{A} &\leftarrow \text{“Totalmente en desacuerdo”} \\ \text{B} &\leftarrow \text{“En desacuerdo”} \\ \text{C} &\leftarrow \text{“De acuerdo”} \\ \text{D} &\leftarrow \text{“Totalmente de acuerdo”} \end{aligned}$$

y se representan los patrones de respuesta por las cinco letras ordenadas reflejando la repuesta en cada pregunta, entonces

$$\Omega = \{(\text{AAAAA}), (\text{AAAAB}), (\text{AAAAC}), (\text{AAAAD}), (\text{AAABA}), \ldots, (\text{DDDDC}), (\text{DDDDD})\}.$$

El número de eventos que se pueden derivar a partir de este espacio muestral (es decir, el número de elementos en $\mathscr{C}$) es igual al número astronómico de 2^{1024} (lo cual es mayor que 10^{308} o bien, un 1 seguido por 308 ceros). La función P asigna a cada uno de estos 2^{1024} eventos una probabilidad. Nótese que si se conocieran las probabilidades de los 1024 eventos elementales, se conocería también la probabilidad de cualquier otro evento E en $\mathscr{C}$, ya que (a) E siempre se puede escribir como la unión de eventos elementales y (b) los eventos elementales son mutuamente exclusivos, de tal forma que (c) se puede utilizar la Ecuación (9.2b) para conocer la probabilidad de E.

En el primer ejemplo del cuadro anterior, se asignaron probabilidades concretos a los diferentes eventos en $\mathscr{C}$ (y también en el segundo ejemplo se podría asignar números concretos a cada evento). Sin embargo, no se explicó cómo se llegaron a dichas probabilidades; por ejemplo, ¿de dónde viene el valor $\frac{1}{4}$ para $P(\{\heartsuit\})$? Efectivamente, la definición de la función de probabilidad (en la p. 299) no especifica cómo calcular las probabilidades; únicamente menciona las condiciones que la función de probabilidad debe satisfacer para ser válida. En la literatura existen diferentes enfoques interpretativos sobre el concepto de probabilidad, los cuales conllevan estrategias para asignar probabilidades concretas a los eventos. A continuación, se introducen brevemente tres de ellos:

(a) *El enfoque clásico o el enfoque a priori*
Este enfoque parte de un análisis lógico del experimento aleatorio. Por ejemplo, si se dice que, al lanzar un dado imparcial, la probabilidad de que salga un cinco es $\frac{1}{6}$, se fundamenta en un análisis lógico del experimento aleatorio y del dado utilizado: El dado tiene 6 lados y, si es imparcial, entonces los lados tienen probabilidades iguales de quedar arriba; por lo tanto, la probabilidad de cada lado es $\frac{1}{6}$.

Se le llama a este enfoque *a priori*, porque permite determinar las probabilidades *antes de* que se realice cualquier observación empírica.

(b) El enfoque frecuentista o el enfoque a posteriori
Una desventaja del enfoque clásico es que, para muchas aplicaciones prácticas, es imposible derivar las probabilidades por un análisis lógico del experimento aleatorio. Por ejemplo, ¿cómo se puede determinar la probabilidad de que salga un cinco al lanzar un dado que *no* es imparcial? o ¿cómo se puede determinar la probabilidad de que alguien que obtuvo un resultado positivo en la prueba ELISA, sea portador del VIH?

Para estos casos, el enfoque frecuentista puede facilitar una solución. Según esta aproximación, la probabilidad de un evento E coincide con la proporción de veces que E se realiza en un número infinito de ejecuciones del experimento aleatorio bajo consideración. En otras palabras, el enfoque frecuentista (y de allí se deriva su nombre) se fundamenta en la propiedad del experimento aleatorio de que se puede repetir un número infinito de veces bajo las mismas circunstancias.

En términos matemáticos, dentro del enfoque frecuentista, se tiene:

$$P(E) \;=\; \lim_{n\to\infty} \frac{f_n}{n}, \tag{9.3}$$

donde f_n es el número de veces que el evento E ha ocurrido en una secuencia de n replicaciones del experimento aleatorio.

Este enfoque es *a posteriori* porque se asignan los valores de probabilidad con base en observaciones empíricas, es decir, *después de* haber ejecutado el experimento aleatorio (un gran número de veces) y haber observado el resultado (en cada ejecución).

Ejemplo

Para ilustrar el enfoque frecuentista, considérese el experimento aleatorio que consiste en lanzar un dado imparcial y que se desea atribuir un valor a la probabilidad del evento $E =$ "sale cuatro o cinco". La ecuación anterior dice que se puede aproximar esta probabilidad por el siguiente procedimiento de tres pasos:

Paso 1: Ejecutar el experimento aleatorio n veces; es decir, lanzar el dado n veces.

Paso 2: Contar cuántas veces en estas n ejecuciones se observa uno de los resultados en el evento E; es decir, se cuenta el número de veces que se observa "cuatro o cinco" en las n lanzamientos. Este número de veces que se realiza el evento se denota f_n.

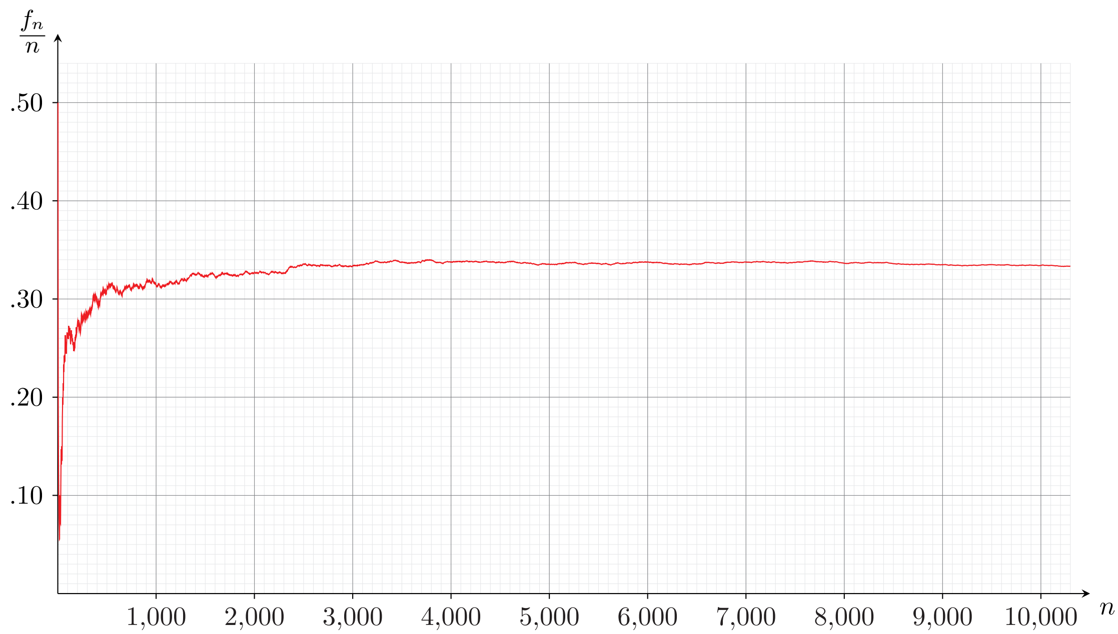
Paso 3: Calcular la proporción de veces que ha ocurrido el evento en estas n ejecuciones y utilizar esta proporción como una aproximación de la probabilidad:

$$P(E) \;\approx\; \frac{f_n}{n}.$$

Realicemos una simulación donde efectivamente se ejecuta este experimento aleatorio un gran número de veces. Después de $n = 100$ ejecuciones, se ha observado $f_{100} = 26$ veces uno de los resultados cuatro o cinco; por lo tanto, $\frac{26}{100} = .26$ es una aproximación de la probabilidad del evento E por el enfoque frecuentista. Nótese que efectivamente es una aproximación, en el sentido que no coincide perfectamente con la probabilidad *a priori* de $\frac{2}{6} = .333$. Por otro lado, la Ecuación (9.3) implica que la fracción $\frac{f_n}{n}$ tiende a ser una mejor aproximación de la probababilidad $P(E)$ conforme n, el número de veces que se ejecuta el experimento aleatorio, aumenta. Por ejemplo, al seguir repetiendo el experimento aleatorio se contó, después de 1000 ejecuciones del experimento, que el evento E se realizó 314 veces, así que:

$$P(E) \approx \frac{f_{1000}}{1000} = \frac{314}{1000} = .314,$$

lo cual es una mejor aproximación en comparación con la que se obtuvo después de 100 ejecuciones. La siguiente gráfica muestra el resultado de la fracción $\frac{f_n}{n}$ en función de n. Se observa que conforme n aumenta, dicha fracción se acerca más a la probabilidad *a priori* de 0.333.



(c) *El enfoque subjetivo*
A veces se utiliza el término *probabilidad* sin referirse a un experimento aleatorio que se puede repetir (ni siquiera en principio). Por ejemplo, se dice que "la probabilidad de que el universo se formó por un *big bang* es 99 %", "la probabilidad de que apruebe el siguiente examen parcial de estadística es $\frac{3}{4}$" o "la probabilidad de que llueva mañana es 20 %". Efectivamente, el experimento aleatorio subyacente a estas afirmaciones es irrepetible: El universo se formó una sola vez, solo habrá un "siguiente examen parcial", y hay solo un día "mañana".

En estas frases, el término "probabilidad" tiene más el sentido de "grado de credibilidad". A pesar de que esta credibilidad puede ser muy personal—por lo cual se suele hablar del enfoque subjetivo de probabilidad—, puede fundamentarse en todo tipo de evidencia científica, por ejemplo, cuando un cosmólogo afirma que hay 99 % de probabilidad de que el universo inició con un *big bang*, o cuando se utilizan modelos numéricos de predicción meteorológica para estimar la probabilidad de precipitación durante el día de mañana. Sin embargo, algunos autores han identificado como problema principal de este enfoque que, en muchas ocasiones, falta un fundamento matemático para asignar valores númericos a la probabilidad/credibilidad.

Esta sección inició con la afirmación de que probabilidad es uno de los conceptos más importantes en la estadística inferencial. El siguiente ejemplo presenta una primera ilustración del papel importante de este concepto para realizar inferencias sobre una población.

Ejemplo

Consideremos el siguiente escenario artificial: Dos urnas contienen cada una 1,000 cuentas, y cada cuenta es de color blanco o negro: mientras que la Urna A contiene 999 cuentas negras y 1 blanca, se desconoce la proporción de cuentas negras y blancas en la Urna B. Se ejecuta el siguiente experimento aleatorio: se pide a alguien externo que seleccione una de las dos urnas y que saque, de forma aleatoria, una cuenta de la urna seleccionada. Esta persona *no* informa sobre la urna de la cuál sacó la cuenta, únicamente revela el color de la cuenta que obtuvo. El objetivo es inferir a partir del dato proporcionado ("cuenta negra" o "cuenta blanca") cuál ha sido la urna de la que sacó la cuenta.

Denominemos la afirmación "la cuenta se sacó de la urna A" como la *Hipótesis A* y, de forma similar, la afirmación "la cuenta se sacó de la urna B" la *Hipótesis B*. Obviamente, cualquiera que sea el resultado obtenido, en ningún caso se puede concluir *con certeza* qué hipótesis es correcta: bajo ambas hipótesis, los dos resultados son posibles. (Nótese que la situación sería diferente, si en vez de una cuenta se extrajeran dos cuentas *sin reposición* de la urna seleccionada; en este caso el dato "dos cuentas blancas" permitiría concluir con certeza que las cuentas se sacaron de la Urna B.) Cuando no es posible concluir con certeza qué hipótesis es la correcta, se puede llegar a afirmaciones *probabilísticas*.

Supongamos que la ejecución del experimento aleatorio da como resultado "cuenta blanca". Si la Hipótesis A es cierta, entonces la probabilidad de obtener este resultado es igual a $\frac{1}{1{,}000} = .001$ (nótese que para llegar a este valor se adoptó el enfoque *a priori*). A partir de esta consideración, son posibles las siguientes dos conclusiones:

- Hipótesis A es verdadera *y se realizó un evento que tuvo una probabilidad* a priori *muy baja*.
- Hipótesis A es falsa.

Generalmente, se eligirá en este caso la segunda conclusión, debido a que se prefiere no aceptar eventos cuya probabilidad de realizarse es muy baja (siempre y cuando existen explicaciones alternativas para el resultado observado).

Aunque el ejemplo anterior parece muy artificial, su estructura representa el molde de un contraste de hipótesis, sin duda la herramienta más utilizada de la estadística inferencial para llegar a conclusiones que generalizan los resultados de una muestra. El Tema 12 tratará en detalle la teoría subyacente a los contrastes de hipótesis y explicará los lazos entre este ejemplo artificial e investigaciones más relevantes en el área de medicina.

9.6.2 Calcular la probabilidad de un evento

Esta sección se dedica a calcular probabilidades de eventos según el enfoque clásico. Es necesario tratar de forma separada espacios muestrales discretos (Sección 9.6.2.1) y continuos (Sección 9.6.2.2).

9.6.2.1 Espacios muestrales discretos (incluyendo combinatoria)

Si Ω es finito o numerable es posible escribir cualquier evento E en Ω como la unión de eventos elementales. De forma abstracta, se escribe:

$$\begin{aligned} \text{si} \quad E &= \{\omega_1, \omega_2, \ldots, \omega_m\}, \\ \text{entonces} \quad E &= \{\omega_1\} \cup \{\omega_2\} \cup \ldots \cup \{\omega_m\}. \end{aligned}$$

Puesto que los eventos elementales son mutuamente exclusivos, se puede aplicar la regla de la suma para conocer $P(E)$, lo cual lleva a:

$$\begin{aligned} P(E) &= P(\{\omega_1\} \cup \{\omega_2\} \cup \ldots \cup \{\omega_m\}) \\ &= P(\omega_1) + P(\omega_2) + \ldots + P(\omega_m). \end{aligned}$$

Ejemplos

Consideremos el caso del experimento aleatorio "lanzar dos dados" y el espacio muestral

$$\begin{aligned} \Omega = \{&(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), \\ &(3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), \\ &(5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\} \end{aligned}$$

donde cada par se refiere al número que salió en el primer y el segundo dado, respectivamente. Definamos el evento E como "la suma de los números en los dos dados es mayor que 9":

$$E = \{(4,6), (5,5), (5,6), (6,4), (6,5), (6,6)\},$$

lo cual es la únion de eventos elementales:

$$E = \{(4,6)\} \cup \{(5,5)\} \cup \{(5,6)\} \cup \{(6,4)\} \cup \{(6,5)\} \cup \{(6,6)\}.$$

Por lo tanto, la probabilidad del evento E coincide con la suma de probabilidades de eventos elementales:

$$P(E) = P((4,6)) + P((5,5)) + P((5,6)) + P((6,4)) + P((6,5)) + P((6,6)).$$

Retomemos el ejemplo del examen de tres preguntas de opción múltiple en la página 292 y los eventos definidos sobre el espacio muestral en la página 296. Según la misma lógica del ejemplo anterior, la probabilidad del evento E_4 "aprobar el examen" se puede escribir como una suma de probabilidades de eventos elementales:

$$\begin{aligned} P(E_4) &= P((\mathrm{A},\mathrm{B},\mathrm{C})) + P((\mathrm{B},\mathrm{A},\mathrm{C})) + P((\mathrm{B},\mathrm{B},\mathrm{A})) + P((\mathrm{B},\mathrm{B},\mathrm{B})) \\ &\quad + P((\mathrm{B},\mathrm{B},\mathrm{C})) + P((\mathrm{B},\mathrm{C},\mathrm{C})) + P((\mathrm{C},\mathrm{B},\mathrm{C})\}). \end{aligned}$$

En los ejemplos anteriores todavía no se obtuvo un valor numérico para la probabilidad. Únicamente se segregó la probabilidad de los eventos en una suma de probabilidades de eventos elementales. Cuando los resultados en el espacio muestral Ω son *equiprobables* (es decir, tienen todos la misma probabilidad), el principio anterior lleva a la *regla de Laplace*, la cual permite asignar probabilidades concretas (es decir, con valores numéricos) a los eventos en Ω.

Regla de Laplace

Si todos los resultados en un espacio muestral finito Ω son equiprobables, entonces para cada evento $E \subseteq \Omega$ se cumple que:

$$P(E) \;=\; \frac{\#E}{\#\Omega}$$

Recuérdese que $\#E$ y $\#\Omega$ son el número cardinal de E y Ω, respectivamente; es decir, el número de elementos en estos conjuntos. La regla de Laplace dice que, *si todos los resultados son igualmente probables*, la probabilidad del evento E corresponde con la proporción de resultados "favorables" (los que implican que el evento se realiza) en comparación con el número total de resultados posibles.

Prueba de la regla de Laplace

La regla de Laplace sigue directamente cuando se escribe la probabilidad de un evento como la suma de las probabilidades de los eventos elementales que lo constituyen. Si Ω es finito y todos los resultados en Ω tienen la misma probabilidad, entonces la probabilidad de un evento elemental $\{\omega\}$ es

$$P(\omega) \;=\; \frac{1}{\#\Omega}.$$

Por lo tanto, para cualquier $E \subseteq \Omega$, que contiene los m resultados $\omega_1, \omega_2, \ldots, \omega_m$, se tiene:

$$\begin{aligned} P(E) &= P(\omega_1) \,+\, P(\omega_2) \,+\, \ldots \,+\, P(\omega_m) \\ &= m\,\frac{1}{\#\Omega} \\ &= \frac{\#E}{\#\Omega}, \end{aligned}$$

puesto que m es el número cardinal de E. ■

Ejemplos

En el ejemplo anterior con el experimento aleatorio de lanzar dos dados, la probabilidad del evento E "la suma de los números es mayor que 9" es igual a:

$$P(E) \;=\; \frac{\#E}{\#\Omega} \;=\; \frac{6}{36} \;=\; .167.$$

Nótese que, si los dos dados son imparciales, los 36 resultados en Ω son equiprobables.

Similarmente, se obtiene en el ejemplo del examen de opción múltiple, suponiendo que todos los resultados son equiprobables, para la probabilidad del evento E_4 "aprobar el examen":

$$P(E) = \frac{\#E}{\#\Omega} = \frac{7}{27} = .259.$$

Se puede verificar en la página 292 que Ω incluye 27 elementos y en la página 296 que el evento E_4 contiene 7 elementos.
Es importante tener claro que esta probabilidad es correcta *sólo bajo el supuesto de equiprobabilidad de los resultados*. En general, *no* es plausible suponer que cualquier patrón de respuesta en un examen de opción múltiple es igualmente probable. Por otro lado, la probabilidad de .259 para el ejemplo puede ser correcto en el caso de que el examinado conteste las preguntas al azar (es cuando, efectivamente, las tres opciones de respuesta en cada pregunta tienen la misma probabilidad de ser elegidas, lo cual conlleva que los 27 patrones de respuesta son equiprobables).

La aplicación de la regla de Laplace requiere un conteo de los elementos en el evento E y en el espacio muestral Ω. Dentro de la *combinatoria* (una rama de las matemáticas), se han propuesto *técnicas de contar* que permiten encontrar el número cardinal de conjuntos que reflejan alguna propiedad o configuración. A continuación, se presentan algunas de las reglas o principios de la combinatoria. La primera regla es fundamental ya que forma la base de las demás reglas que se introducirán en el resto de esta sección.

Regla básica de la combinatoria: la Regla del Producto

Si (a) un experimento aleatorio implica extraer un elemento de cada una de n poblaciones de $N_1, N_2, \ldots, N_n$ elementos, respectivamente, donde los elementos dentro de cada población son diferentes uno de otro

y (b) se considera, como el resultado de este experimento aleatorio, el conjunto ordenado de los n elementos extraídos,

entonces se cumple que:

$$\#\Omega = N_1 \times N_2 \times \ldots \times N_n.$$

Justificación de la Regla del Producto

El principio de la regla del producto se puede entender como sigue: Como para *cada uno* de los N_1 elementos de la primera población hay N_2 elementos de la segunda población con que se puede combinar, hay $N_1 \times N_2$ maneras diferentes en que se pueden obtener los dos primeros elementos. Para *cada una* de estas $N_1 \times N_2$ combinaciones, hay N_3 elementos de la tercera población con que se pueden combinar; entonces hay $N_1 \times N_2 \times N_3$ combinaciones diferentes para los primeros tres elementos. Se sigue aplicando la misma lógica hasta incluir la última población n.

Ejemplos

Tomemos como ejemplo el experimento aleatorio que consiste en lanzar un dado y echar una moneda, definiendo como resultado el número que sale en el dado y la cara que sale en la moneda. Este experimento aleatorio se puede conceptualizar como un procedimiento en el cual se extrae primero un elemento de la población $\{1, 2, 3, 4, 5, 6\}$ (para la cual $N_1 = 6$) y, a continuación, un elemento de la población {águila, sol} (con $N_2 = 2$). El siguiente diagrama visualiza los resultados que tendrá el espacio muestral de este experimento aleatorio:

Dado		Moneda		Resultado
1	→	águila	→	(1, águila)
	→	sol	→	(1, sol)
2	→	águila	→	(2, águila)
	→	sol	→	(2, sol)
3	→	águila	→	(3, águila)
	→	sol	→	(3, sol)
4	→	águila	→	(4, águila)
	→	sol	→	(4, sol)
5	→	águila	→	(5, águila)
	→	sol	→	(5, sol)
6	→	águila	→	(6, águila)
	→	sol	→	(6, sol)

Se observa que cada elemento de la población $\{1, 2, 3, 4, 5, 6\}$ se combina con cada elemento de la población {águila, sol}, y que esto lleva a los 12 resultados en la última columna.

Supongamos que un examen consiste en tres preguntas de opción múltiple, donde la primera pregunta tiene tres alternativas de respuesta (A, B y C), la segunda dos (A y B) y la tercera cuatro alternativas (A, B, C y D). Suponiendo que todos los patrones de tres respuestas son equiprobables, ¿cuál es la probabilidad de contestar correctamente las tres preguntas?
El número total de patrones de respuesta que se pueden observar, es igual a:

$$\#\Omega \ = \ 3 \times 2 \times 4 \ = \ 24.$$

En particular, los posibles patrones son:

(A,A,A)	(A,A,B)	(A,A,C)	(A,A,D)	(A,B,A)	(A,B,B)	(A,B,C)	(A,B,D)
(B,A,A)	(B,A,B)	(B,A,C)	(B,A,D)	(B,B,A)	(B,B,B)	(B,B,C)	(B,B,D)
(C,A,A)	(C,A,B)	(C,A,C)	(C,A,D)	(C,B,A)	(C,B,B)	(C,B,C)	(C,B,D)

Puesto que el evento E "contestar correctamente todas las preguntas" corresponde con un único patrón de respuesta, la probabilidad de obtener este resultado, si todos los 24 patrones son equiprobables, es igual a $\frac{1}{24}$.

Las demás reglas de la combinatoria que se presentan en esta sección se derivan de la regla del producto.

Regla 1: Ordenar N elementos

Si una población contiene N elementos distintos, entonces el número de muestras diferentes de tamaño $n = N$ que se pueden extraer *sin reposición* de esta población es igual a $N!$.

Una forma alternativa y equivalente para expresar esta regla es:
El número de maneras en las que se pueden ordenar N elementos distintos es igual a $N!$.

Factoriales

En el cuadro anterior, $N!$ (léase: N *factorial*), donde N es un número natural, se define como:

$$N! = N \times (N-1) \times (N-2) \times \ldots \times 1.$$

Por ejemplo:

$$\begin{aligned} 3! &= 3 \times 2 \times 1 = 6 \\ 7! &= 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 5,040. \end{aligned}$$

La siguiente tabla da los factoriales hasta 20!:

$1! = 1$	$11! = 39,916,800$
$2! = 2$	$12! = 479,001,600$
$3! = 6$	$13! = 6,227,020,800$
$4! = 24$	$14! = 87,178,291,200$
$5! = 120$	$15! = 1,307,674,368,000$
$6! = 720$	$16! = 20,922,789,888,000$
$7! = 5,040$	$17! = 355,687,428,096,000$
$8! = 40,320$	$18! = 6,402,373,705,728,000$
$9! = 362,880$	$19! = 121,645,100,408,832,000$
$10! = 3,628,800$	$20! = 2,432,902,008,176,640,000$

Se observa que los factoriales crecen muy rápido. Las calculadoras científicas y los programas de cálculo en la computadora generalmente incluyen una función para calcular directamente un factorial.

Cabe señalar que, por definición, se establece:

$$0! = 1.$$

A partir de la Regla 1 se puede calcular el número de muestras—recuérdese que muestras son conjuntos ordenados—de tamaño N que se pueden extraer sin reposición de una población de N elementos. Es decir, el experimento aleatorio consiste en extraer *todos* los elementos de la población uno por uno y el resultado que se considera es el conjunto *ordenado* de estos elementos. Esto explica por qué las dos formas para expresar la Regla 1, que se presentan en el cuadro arriba, son equivalentes.

Justificación de la Regla 1

La Regla 1 es un caso especial de la regla del producto. El experimento aleatorio de extraer sin reposición todos los elementos de una población de N elementos es conceptualmente equivalente al experimento aleatorio de extraer consecutivamente un elemento de N poblaciones, donde la Población 1 es la población original de N elementos, la Población 2 es esta población sin el primer elemento extraído (la cual entonces tiene $N-1$ elementos), la Población 3 es la población original sin los primeros dos elementos que se extrajeron (que tiene por lo tanto $N-2$ elementos), etcétera, hasta la Población N que corresponde con la población original después de haber quitado los $N-1$ elementos extraídos anteriormente (la cual, por lo tanto, contiene sólo un elemento). Por lo tanto, el número de muestras diferentes que se pueden obtener es $N \times (N-1) \times (N-2) \times \ldots \times 1$, o bien, $N!$.

Ejemplos

Considérese la población de las letras $\{a, b, c, d\}$, las cuales podrían ser las letras correspondientes de las cuatro opciones de respuesta de una pregunta de opción múltiple. Se pueden obtener $4! = 4 \times 3 \times 2 \times 1 = 24$ muestras diferentes de tamaño $n = N = 4$ de esta población, si se extrae sin reposición. Es claro que cada una de estas muestras corresponde con un orden de los cuatro elementos:

$$\begin{array}{llllll}
(a,b,c,d) & (a,b,d,c) & (a,c,b,d) & (a,c,d,b) & (a,d,b,c) & (a,d,c,b) \\
(b,a,c,d) & (b,a,d,c) & (b,c,a,d) & (b,c,d,a) & (b,d,a,c) & (b,d,c,a) \\
(c,a,b,d) & (c,a,d,b) & (c,b,a,d) & (c,b,d,a) & (c,d,a,b) & (c,d,b,a) \\
(d,a,b,c) & (d,a,c,b) & (d,b,a,c) & (d,b,c,a) & (d,c,a,b) & (d,c,b,a)
\end{array}$$

En otras palabras, existen 24 maneras distintas para ordenar las cuatro opciones de respuesta en una pregunta de opción múltiple.

Como última fase de la clasificación para la Copa Mundial de Futból de 2014 en Brasil, México entró en un “torneo hexagonal” con otros cinco países de la zona Concacaf (la Confederación de Fútbol de Norte, Centroamérica y el Caribe): Costa Rica, Estados Unidos, Jamaica, Honduras y Panamá. En este torneo cada equipo juega un partido ida y vuelta contra los otros equipos del grupo. Los tres equipos con más puntos al final del torneo se clasifican directamente para la Copa Mundial; el equipo en cuarto lugar juega el partido de repesca frente a Nueva Zelanda y los equipos en quinto y sexto lugar quedan eliminados. Si los seis equipos participantes son igualmente fuertes (de tal suerte que cualquier orden de los seis países después del torneo final

es igualmente probable), ¿cuál es la probabilidad de que México termine en una de las primeras tres posiciones?
Según la Regla 1, existen $6! = 720$ formas diferentes para ordenar los seis equipos. En $5! = 120$, México ocupa la primera posición (nótese que si México ocupa la primera posición, hay $5!$ maneras en que se pueden ordenar los otros cinco equipos); asimismo, hay 120 órdenes donde México ocupa la segunda posición y 120 donde ocupa la tercera posición. Por lo tanto, la probabilidad de que México ocupe una de las primeras tres posiciones es, según la regla de Laplace, igual a:

$$\frac{3 \times 5!}{6!} = \frac{360}{720} = .50,$$

lo cual intuitivamente tiene sentido ya que si hay seis equipos igualmente fuertes, la probabilidad de ocupar cualquier posición es $\frac{1}{6}$ y la probabilidad de terminar en una de las primeras tres posiciones es $\frac{3}{6}$.

Regla 2: Muestras ordenadas de tamaño n extraídas con reposición de una población de N elementos

Si una población contiene N elementos distintos, entonces el número de muestras diferentes de tamaño n que se pueden extraer *con reposición* de esta población es igual a

$$\overbrace{N \times N \times \ldots \times N}^{n \text{ veces}},$$

o bien,

$$N^n.$$

Justificación de la Regla 2

El experimento aleatorio de extraer con reposición n elementos de una población de N elementos es conceptualmente equivalente al experimento aleatorio de extraer consecutivamente un elemento de n poblaciones, donde cada población contiene los mismos N elementos. Recuérdese que, al extraer una muestra *con reposición* de una población, la población *no* cambia durante la extracción (de los diferentes elementos) de la muestra (véase la tabla en la página 285). Por lo tanto los n elementos de la muestra se extraen de la misma población, lo cual implica, después de aplicar la regla del producto, que hay N^n diferentes muestras que se pueden obtener.

Ejemplo

¿Cuántos posibles patrones de respuesta existen en un examen de 10 preguntas de opción múltiple, donde cada pregunta tiene 4 opciones de respuesta?
Se puede conceptualizar este experimento aleatorio como sacar 10 veces una respuesta de la población de opciones/respuestas {A,B,C,D}. Se extrae la muestra ordenada de 10 respuestas *con reposición* ya que, para cada pregunta, la población de respuestas es la misma. Por lo tanto,

la aplicación de la Regla 2 revela que hay

$$4^{10} = 1,048,576$$

distintos patrones de respuesta.

Regla 3: Muestras ordenadas de tamaño n extraídas sin reposición de una población de N elementos

Si una población tiene N elementos distintos, entonces el número de muestras diferentes de tamaño n (para $n \leqslant N$) que se pueden extraer *sin reposición* de esta población es igual a

$$N \times (N-1) \times (N-2) \times \ldots \times (N-n+1),$$

o bien,

$$\frac{N!}{(N-n)!}$$

La Regla 3 es una generalización de la Regla 1 en el sentido que permite calcular el número de muestras de tamaño n que se pueden extraer sin reposición de una población de N elementos, donde $n \leqslant N$ (mientras que la Regla 1 considera el mismo caso pero únicamente para $n = N$). Obsérvese que, si $n = N$, entonces la fórmula de la Regla 3 se reduce a:

$$N \times (N-1) \times (N-2) \times \ldots \times (N-N+1),$$

lo cual es igual a $N!$, tal como especifica la Regla 1.

Permutaciones

En la combinatoria, se le llama *permutación* a un orden específico de n elementos. "Permutar los n elementos" de una muestra quiere decir "variar el orden de los n elementos". La Regla 1 permite calcular el número de permutaciones posibles de N elementos (lo cual es igual a $N!$). Es decir, la Regla 1 considera el caso de permutar o ordenar la población total de N elementos, es decir, el caso de una *permutación total*. La Regla 3, por otro lado, considera el caso de una *permutación parcial* de n elementos de un total de N; en otras palabras: ¿Cuántas permutaciones (o muestras ordenadas) de n elementos se pueden extraer de un total de N?

A veces se escribe el producto $N \times (N-1) \times (N-2) \times \ldots \times (N-n+1)$ de forma abreviada como $(N)_n$, y se llama "el número de permutaciones (parciales) de n elementos tomados de N". Es decir, se define

$$(N)_n = \frac{N!}{(N-n)!}. \tag{9.4}$$

Justificación de la Regla 3

El experimento aleatorio de extraer sin reposición n elementos de una población de N elementos es conceptualmente equivalente al experimento aleatorio de extraer consecutivamente un elemento de n poblaciones, donde la Población 1 contiene los N elementos iniciales, la Población 2 los $N-1$ elementos que se quedan después de haber extraído el primer elemento, la Población 3 los $N-2$ elementos que se quedan después de la extracción de los primeros dos elementos, etcétera, hasta la Población n que contiene los $N-n+1$ elementos después de las $n-1$ extracciones anteriores. Por lo tanto, aplicar la regla del producto lleva a

$$N \times (N-1) \times (N-2) \times \ldots \times (N-n+1)$$

diferentes muestras de tamaño n que se pueden extraer sin reposición de una población de N elementos.

Ejemplo

¿Cuál es la probabilidad de extraer la muestra

$$(\text{Diana}, \text{Carlos}, \text{Lizz})$$

de la población de los estudiantes de la clase de Bioestadística

$$\{\text{Adrián}, \text{Carlos}, \text{Diana}, \text{Lizz}, \text{Oliver}\},$$

si se extrae sin reposición y por un muestreo aleatorio simple?

En este caso, el primer elemento se obtiene de la población completa de 5 elementos, el segundo elemento de una población de 4 elementos, y el tercer elemento de una población de 3 elementos. Por lo tanto, hay $5 \times 4 \times 3 = 60$ muestras diferentes que se pueden obtener sin reposición. Si el muestreo es aleatorio simple, la probabilidad para obtener (Diana, Carlos, Lizz) es $\frac{1}{60}$.

Regla 4: Muestras no ordenadas de tamaño n extraídas sin reposición de una población de N elementos

Si una población tiene N elementos distintos, entonces el número de muestras *no ordenadas* diferentes de tamaño n (para $n \leqslant N$) que se pueden extraer *sin reposición* de esta población es igual a

$$\frac{N!}{n!\,(N-n)!}. \tag{9.5a}$$

Combinaciones y el coeficiente binomial

Contrario a las permutaciones, cuando se consideran *combinaciones* no interesa el orden. Una combinación de n elementos es entonces un (sub)conjunto no ordenado de n elementos. La expresión en la Ecuación (9.5a), que lleva al número de combinaciones de n elementos que se pueden sacar de una población de N elementos, comúnmente se abrevia como

$$\binom{N}{n}$$

y se llama el *coeficiente binomial*.

Nótese que el coeficiente binomial se relaciona con el coeficiente $(N)_n$ (introducido en la página 314) de la siguiente manera:

$$\binom{N}{n} = \frac{(N)_n}{n!}. \tag{9.5b}$$

Justificación de la Regla 4

A partir de la Regla 3, se conoce el número de muestras *ordenadas* de n elementos que se pueden extraer de una población de N elementos:

$$(N)_n = \frac{N!}{(N-n)!}.$$

Para conocer el número de muestras *no ordenadas* de n elementos, se contempla el número de maneras que se pueden ordenar cualquier conjunto de n elements. Según la Regla 1, se pueden ordenar los n elementos de $n!$ maneras posibles. Por lo tanto, para cada combinación (= muestra no ordenada) de n elementos, hay $n!$ permutaciones (= muestras ordenadas) que contienen los mismos n elementos en órdenes diferentes. Esto lleva a la Ecuación (9.5b), que divide el número de permutaciones entre $n!$ para conocer el número de combinaciones.

Ejemplos

Considérese la población de las letras $\{a, b, c, d, e\}$, para la cual $N = 5$. Según la Regla 4, el número de muestras *no ordenadas* de tamaño 3 que se pueden extraer de esta población es igual a:

$$\begin{aligned}\binom{5}{3} = \frac{N!}{n!\,(N-n)!} &= \frac{5!}{3! \times 2!} \\ &= \frac{120}{6 \times 2} \\ &= 10.\end{aligned}$$

Por otro lado, considérese el caso de sacar muestras de la misma población *tomando en cuenta el orden* de los elementos. Para este caso, la Regla 3 indica que el número de muestras posibles es:

$$\begin{aligned}(N)_n &= (5)_3\\ &= 5 \times 4 \times 3\\ &= 60.\end{aligned}$$

La siguiente tabla enumera las 60 muestras ordenadas (entre paréntesis redondos) y al mismo tiempo las 10 muestras no ordenadas (entre llaves). Además, la organización de la tabla aclara que los 3 elementos de cada muestra no ordenada se pueden ordenar de $3! = 6$ maneras. Es decir, hay 6 muestras ordenadas correspondientes de cada muestra no ordenada.

$$\begin{array}{cccccccc}
(a,b,c) & (a,c,b) & (b,a,c) & (b,c,a) & (c,a,b) & (c,b,a) & \longleftrightarrow & \{a,b,c\}\\
(a,b,d) & (a,d,b) & (b,a,d) & (b,d,a) & (d,a,b) & (d,b,a) & \longleftrightarrow & \{a,b,d\}\\
(a,b,e) & (a,e,b) & (b,a,e) & (b,e,a) & (e,a,b) & (e,b,a) & \longleftrightarrow & \{a,b,e\}\\
(a,c,d) & (a,d,c) & (c,a,d) & (c,d,a) & (d,a,c) & (d,c,a) & \longleftrightarrow & \{a,c,d\}\\
(a,c,e) & (a,e,c) & (c,a,e) & (c,e,a) & (e,a,c) & (e,c,a) & \longleftrightarrow & \{a,c,e\}\\
(a,d,e) & (a,e,d) & (d,a,e) & (d,e,a) & (e,a,d) & (e,d,a) & \longleftrightarrow & \{a,d,e\}\\
(b,c,d) & (b,d,c) & (c,b,d) & (c,d,b) & (d,b,c) & (d,c,b) & \longleftrightarrow & \{b,c,d\}\\
(b,c,e) & (b,e,c) & (c,b,e) & (c,e,b) & (e,b,c) & (e,c,b) & \longleftrightarrow & \{b,c,e\}\\
(b,d,e) & (b,e,d) & (d,b,e) & (d,e,b) & (e,b,d) & (e,d,b) & \longleftrightarrow & \{b,d,e\}\\
(c,d,e) & (c,e,d) & (d,c,e) & (d,e,c) & (e,c,d) & (e,d,c) & \longleftrightarrow & \{c,d,e\}
\end{array}$$

De esta forma, la tabla anterior ilustra el principio que subyace la Ecuación (9.5b): Al dividir el número de permutaciones entre $n!$, se obtiene el número de combinaciones.

$$\frac{(5)_3}{3!} = \frac{60}{6} = 10 = \binom{5}{3}.$$

Para conocer las opiniones de estudiantes sobre la educación y el aprendizaje electrónico (*e-learning*), se construye un cuestionario que consiste en 10 afirmaciones que se responden eligiendo uno de las opciones “de acuerdo” y “en desacuerdo”. Suponiendo (a) que una persona está indecisa sobre el aprendizaje electrónico, por lo cual para cada pregunta la probabilidad de que conteste “de acuerdo” es igual a .50, y (b) que sus respuestas en las 10 preguntas son independientes, ¿cuál es la probabilidad de que esta persona exprese estar “de acuerdo” en 8 o más preguntas?

Para responder esta pregunta, se realizan los siguientes cálculos:

(1) Puesto que para cada pregunta se elige una opción entre dos, el número de patrones de respuesta es $2^{10} = 1024$. (Se aplica la Regla 2 ya que el experimento aleatorio implica la extracción de un elemento de la población {“de acuerdo”, “en desacuerdo”} en cada una de las 10 preguntas.)

(2) El número de patrones de respuesta que contienen 8 veces “de acuerdo” y 2 veces “en

desacuerdo” es igual a:

$$\begin{aligned}\binom{10}{8} &= \frac{10!}{8! \times 2!} \\ &= \frac{3{,}628{,}800}{40{,}320 \times 2} \\ &= 45.\end{aligned}$$

La justificación del uso de la Regla 4 en este cálculo es la siguiente: cada patrón de respuesta contiene 10 elementos, correspondientes a las 10 preguntas, y se extrae de esta población de 10 preguntas una muestra de 8 sin reposición y sin que importe el orden. Enumerando las preguntas de 1 a 10 de tal forma que la población de preguntas es {1,2,3,4,5,6,7,8,9,10}, se puede verificar que las combinaciones de 8 elementos de esta población son las siguientes:

$\{1,2,3,4,5,6,7,8\}$	$\{1,2,3,5,6,7,8,9\}$	$\{1,3,4,5,6,7,9,10\}$
$\{1,2,3,4,5,6,7,9\}$	$\{1,2,3,5,6,7,8,10\}$	$\{1,3,4,5,6,8,9,10\}$
$\{1,2,3,4,5,6,7,10\}$	$\{1,2,3,5,6,7,9,10\}$	$\{1,3,4,5,7,8,9,10\}$
$\{1,2,3,4,5,6,8,9\}$	$\{1,2,3,5,6,8,9,10\}$	$\{1,3,4,6,7,8,9,10\}$
$\{1,2,3,4,5,6,8,10\}$	$\{1,2,3,5,7,8,9,10\}$	$\{1,3,5,6,7,8,9,10\}$
$\{1,2,3,4,5,6,9,10\}$	$\{1,2,3,6,7,8,9,10\}$	$\{1,4,5,6,7,8,9,10\}$
$\{1,2,3,4,5,7,8,9\}$	$\{1,2,4,5,6,7,8,9\}$	$\{2,3,4,5,6,7,8,9\}$
$\{1,2,3,4,5,7,8,10\}$	$\{1,2,4,5,6,7,8,10\}$	$\{2,3,4,5,6,7,8,10\}$
$\{1,2,3,4,5,7,9,10\}$	$\{1,2,4,5,6,7,9,10\}$	$\{2,3,4,5,6,7,9,10\}$
$\{1,2,3,4,5,8,9,10\}$	$\{1,2,4,5,6,8,9,10\}$	$\{2,3,4,5,6,8,9,10\}$
$\{1,2,3,4,6,7,8,9\}$	$\{1,2,4,5,7,8,9,10\}$	$\{2,3,4,5,7,8,9,10\}$
$\{1,2,3,4,6,7,8,10\}$	$\{1,2,4,6,7,8,9,10\}$	$\{2,3,4,6,7,8,9,10\}$
$\{1,2,3,4,6,7,9,10\}$	$\{1,2,5,6,7,8,9,10\}$	$\{2,3,5,6,7,8,9,10\}$
$\{1,2,3,4,6,8,9,10\}$	$\{1,3,4,5,6,7,8,9\}$	$\{2,4,5,6,7,8,9,10\}$
$\{1,2,3,4,7,8,9,10\}$	$\{1,3,4,5,6,7,8,10\}$	$\{3,4,5,6,7,8,9,10\}$

Se puede considerar las ocho preguntas de una muestra como las preguntas que se contestan con “de acuerdo”. Así se explica que hay 45 diferentes formas de expresar “de acuerdo” en precisamente 8 preguntas.

(3) Según la misma lógica, se puede argumentar que el número de patrones de respuesta que tienen 9 veces un “de acuerdo” (y una vez “en desacuerdo”) es igual a:

$$\begin{aligned}\binom{10}{9} &= \frac{10!}{9! \times 1!} \\ &= \frac{3{,}628{,}800}{362{,}880 \times 1} \\ &= 10.\end{aligned}$$

(4) Obviamente, solo hay un patrón de respuesta en el cual las 10 preguntas se contestan con “de acuerdo”. Efectivamente, el coeficiente binomial en esta caso es igual a 1:

$$\begin{aligned}\binom{10}{10} &= \frac{10!}{10! \times 0!} \\ &= \frac{3,628,800}{3,628,800 \times 1} \\ &= 1.\end{aligned}$$

(5) Por los tres puntos anteriores, el número de patrones de respuesta que corresponden con el evento “estar de acuerdo en 8 *o más* preguntas” es:

$$\binom{10}{8} + \binom{10}{9} + \binom{10}{10} = 45 + 10 + 1 = 56.$$

(6) Si las probabilidades de contestar “de acuerdo” y “en desacuerdo” son iguales y las respuestas son independientes, entonces los 1024 patrones de respuestas son equiprobables. Por lo tanto, según la regla de Laplace, la probabilidad del evento E “estar de acuerdo en 8 o más preguntas” es la siguiente:

$$\begin{aligned}P(E) &= \frac{\#E}{\#\Omega} \\ &= \frac{56}{1024} \\ &\approx .055.\end{aligned}$$

En el punto (2) del último ejemplo, la Regla 4 sirvió para conocer el número de muestras de tamaño 8 que se pueden sacar de las 10 preguntas. De forma similar, se puede interpretar esta operación como el número de maneras en que se puede ordenar un conjunto de 10 elementos que tiene 8 elementos de un tipo (a saber, preguntas que se contestan con “de acuerdo”) y 2 elementos de otro tipo (preguntas con que se está “en desacuerdo”). Esta consideración lleva a la siguiente variante de la Regla 4.

Regla 4b: Ordenar N elementos de 2 tipos

Si un conjunto consiste en n_1 elementos idénticos de un tipo y n_2 elementos idénticos de otro tipo, entonces el número de maneras en que se pueden ordenar los $N = n_1 + n_2$ elementos del conjunto se da por:

$$\frac{N!}{n_1!\, n_2!},$$

lo cual es igual tanto a $\binom{N}{n_1}$ como a $\binom{N}{n_2}$.

Ejemplo

En el ejemplo anterior, había 45 patrones con 8 respuestas "de acuerdo" y 2 "en desacuerdo". Si se representan las respuestas "de acuerdo" con círculos vacíos y las "en desacuerdo" con círculos rellenos, entonces los 45 patrones de respuesta corresponden con:

$$\begin{array}{ccc}
(\circ,\circ,\circ,\circ,\circ,\circ,\circ,\circ,\bullet,\bullet) & (\circ,\circ,\circ,\bullet,\circ,\circ,\circ,\circ,\circ,\bullet) & (\circ,\bullet,\circ,\circ,\circ,\circ,\circ,\bullet,\circ,\circ) \\
(\circ,\circ,\circ,\circ,\circ,\circ,\circ,\bullet,\circ,\bullet) & (\circ,\circ,\circ,\bullet,\circ,\circ,\circ,\circ,\bullet,\circ) & (\circ,\bullet,\circ,\circ,\circ,\circ,\bullet,\circ,\circ,\circ) \\
(\circ,\circ,\circ,\circ,\circ,\circ,\circ,\bullet,\bullet,\circ) & (\circ,\circ,\circ,\bullet,\circ,\circ,\circ,\bullet,\circ,\circ) & (\circ,\bullet,\circ,\circ,\circ,\bullet,\circ,\circ,\circ,\circ) \\
(\circ,\circ,\circ,\circ,\circ,\circ,\bullet,\circ,\circ,\bullet) & (\circ,\circ,\circ,\bullet,\circ,\circ,\bullet,\circ,\circ,\circ) & (\circ,\bullet,\circ,\circ,\bullet,\circ,\circ,\circ,\circ,\circ) \\
(\circ,\circ,\circ,\circ,\circ,\circ,\bullet,\circ,\bullet,\circ) & (\circ,\circ,\circ,\bullet,\circ,\bullet,\circ,\circ,\circ,\circ) & (\circ,\bullet,\circ,\bullet,\circ,\circ,\circ,\circ,\circ,\circ) \\
(\circ,\circ,\circ,\circ,\circ,\circ,\bullet,\bullet,\circ,\circ) & (\circ,\circ,\circ,\bullet,\bullet,\circ,\circ,\circ,\circ,\circ) & (\circ,\bullet,\bullet,\circ,\circ,\circ,\circ,\circ,\circ,\circ) \\
(\circ,\circ,\circ,\circ,\circ,\bullet,\circ,\circ,\circ,\bullet) & (\circ,\circ,\bullet,\circ,\circ,\circ,\circ,\circ,\circ,\bullet) & (\bullet,\circ,\circ,\circ,\circ,\circ,\circ,\circ,\circ,\bullet) \\
(\circ,\circ,\circ,\circ,\circ,\bullet,\circ,\circ,\bullet,\circ) & (\circ,\circ,\bullet,\circ,\circ,\circ,\circ,\circ,\bullet,\circ) & (\bullet,\circ,\circ,\circ,\circ,\circ,\circ,\circ,\bullet,\circ) \\
(\circ,\circ,\circ,\circ,\circ,\bullet,\circ,\bullet,\circ,\circ) & (\circ,\circ,\bullet,\circ,\circ,\circ,\circ,\bullet,\circ,\circ) & (\bullet,\circ,\circ,\circ,\circ,\circ,\circ,\bullet,\circ,\circ) \\
(\circ,\circ,\circ,\circ,\circ,\bullet,\bullet,\circ,\circ,\circ) & (\circ,\circ,\bullet,\circ,\circ,\circ,\bullet,\circ,\circ,\circ) & (\bullet,\circ,\circ,\circ,\circ,\circ,\bullet,\circ,\circ,\circ) \\
(\circ,\circ,\circ,\circ,\bullet,\circ,\circ,\circ,\circ,\bullet) & (\circ,\circ,\bullet,\circ,\circ,\bullet,\circ,\circ,\circ,\circ) & (\bullet,\circ,\circ,\circ,\circ,\bullet,\circ,\circ,\circ,\circ) \\
(\circ,\circ,\circ,\circ,\bullet,\circ,\circ,\circ,\bullet,\circ) & (\circ,\circ,\bullet,\circ,\bullet,\circ,\circ,\circ,\circ,\circ) & (\bullet,\circ,\circ,\circ,\bullet,\circ,\circ,\circ,\circ,\circ) \\
(\circ,\circ,\circ,\circ,\bullet,\circ,\circ,\bullet,\circ,\circ) & (\circ,\circ,\bullet,\bullet,\circ,\circ,\circ,\circ,\circ,\circ) & (\bullet,\circ,\circ,\bullet,\circ,\circ,\circ,\circ,\circ,\circ) \\
(\circ,\circ,\circ,\circ,\bullet,\circ,\bullet,\circ,\circ,\circ) & (\circ,\bullet,\circ,\circ,\circ,\circ,\circ,\circ,\circ,\bullet) & (\bullet,\circ,\bullet,\circ,\circ,\circ,\circ,\circ,\circ,\circ) \\
(\circ,\circ,\circ,\circ,\bullet,\bullet,\circ,\circ,\circ,\circ) & (\circ,\bullet,\circ,\circ,\circ,\circ,\circ,\circ,\bullet,\circ) & (\bullet,\bullet,\circ,\circ,\circ,\circ,\circ,\circ,\circ,\circ)
\end{array}$$

Obsérvese la correspondencia entre estos 45 conjuntos ordenados y las 45 combinaciones enumeradas en el punto (2) del ejemplo anterior.

Regla 5: Ordenar N elementos de k tipos

Si (a) un conjunto consiste en elementos de k diferentes tipos,
(b) los elementos de cada tipo son idénticos,
y (c) el número de elementos de cada tipo es $n_1, n_2, \ldots, n_k$, respectivamente,

entonces el número de maneras en que se pueden ordenar los $N = n_1 + n_2 + \cdots + n_k$ elementos se da por:

$$\frac{N!}{n_1!\, n_2!\, \ldots\, n_k!}. \tag{9.6}$$

El coeficiente multinomial

Se le llama *coeficiente multinomial* a la expresión en la Ecuación (9.6). Se escribe:

$$\binom{N}{n_1, n_2, \ldots, n_k} = \frac{N!}{n_1!\, n_2!\, \ldots\, n_k!}.$$

Es evidente que la Regla 5 generaliza la Regla 4b y que el coeficiente multinomial es una generalización del coeficiente binomial. En el caso de la Regla 4b, se consideran únicamente dos tipos de elementos en la población, por lo cual el número de maneras de ordenar los elementos se calcula con el coeficiente *bi*nomial; la Regla 5 considera el caso de ordenar una población que consiste en elementos de k diferentes tipos; por lo cual se aplica el coeficiente *multi*nomial.

Justificación de la Regla 5

Según la Regla 1 hay $N!$ maneras de ordenar los elementos de una población de N elementos. Sin embargo, la Regla 1 supone que los N elementos son distintos. Para conocer el número de órdenes cuando algunos de los elementos de la población son idénticos, considérense los siguientes casos:

1. 2 elementos de los N son idénticos y cada uno de los demás $N - 2$ elementos es único (es decir, diferente de cualquier otro elemento). Puesto que los 2 elementos son indistinguibles, intercambiarlos en cualquier orden de los N elementos resulta en el mismo orden. Es decir, para cada uno de los $N!$ órdenes de los N elementos existen dos variantes que son indistinguibles. Por lo tanto, el número de órdenes distinguibles es
$$\frac{N!}{2}.$$
2. n elementos de los N son idénticos y cada uno de los demás $N-n$ elementos es único. Considerando únicamente a los n elementos indistinguibles en un orden específico de los N elementos, es claro que existen $n!$ maneras para ordenar estos n elementos. Por lo tanto, para cada uno de los N órdenes de los N elementos existen $n!$ variantes que son indistinguibles, por lo cual el número de órdenes distinguibles es:
$$\frac{N!}{n!}.$$
3. los N elementos se dividen en k grupos donde los elementos de diferentes grupos son distintos, pero los elementos dentro de cada grupo son idénticos. Según el razonamiento en el punto anterior, si el primer grupo contiene n_1 elementos, entonces para cualquier de los $N!$ órdenes, existen $n_1!$ variantes indistinguibles que corresponden con cambiar el orden de los n_1 elementos. De la misma forma, existen $n_2!$ variantes indistinguibles que corresponden con cambiar el orden de los n_2 elementos del segundo grupo, etcétera, hasta considerar los $n_k!$ variantes correspondientes a combiar el orden de los elementos del último grupo. Por lo tanto, el número de órdenes distinguibles se obtiene dividiendo $N!$ entre los respectivos factoriales del número de elementos de cada grupo:
$$\frac{N!}{n_1!\, n_2!\, \ldots\, n_k!}.$$

Ejemplo

Ilustremos los tres casos de la justificación anterior con un ejemplo artificial:

1. Considérese una población que consiste en 4 cuentas: 2 rojas, 1 azul y 1 verde. Esto efectivamente representa el caso de una población en la cual 2 de los $N = 4$ elementos son idénticos y los otros dos elementos son únicos. ¿De cuántas maneras se pueden ordenar estas cuatro cuentas?
   Si es posible distinguir entre las dos cuentas rojas (por ejemplo, porque traen una etiqueta con su número), entonces en total hay $4! = 24$ diferentes maneras para ordenarlas. Sin embargo, si las dos cuentas rojas son indistinguibles la una de la otra, entonces cada muestra ordenada se duplica y únicamente quedan $\frac{24}{2} = 12$ muestras distinguibles:

   (❶, ❷, 🔵, 🟢) = (❷, ❶, 🔵, 🟢)
   (❶, 🔵, ❷, 🟢) = (❷, 🔵, ❶, 🟢)
   (❶, 🔵, 🟢, ❷) = (❷, 🔵, 🟢, ❶)
   (🔵, ❶, ❷, 🟢) = (🔵, ❷, ❶, 🟢)
   (🔵, ❶, 🟢, ❷) = (🔵, ❷, 🟢, ❶)
   (🔵, 🟢, ❶, ❷) = (🔵, 🟢, ❷, ❶)

   (❶, ❷, 🟢, 🔵) = (❷, ❶, 🟢, 🔵)
   (❶, 🟢, 🔵, ❷) = (❷, 🟢, 🔵, ❶)
   (❶, 🟢, 🔵, ❷) = (❷, 🟢, 🔵, ❶)
   (🟢, ❶, ❷, 🔵) = (🟢, ❷, ❶, 🔵)
   (🟢, ❶, 🔵, ❷) = (🟢, ❷, 🔵, ❶)
   (🟢, 🔵, ❶, ❷) = (🟢, 🔵, ❷, ❶)

2. Similarmente, se calcula el número de maneras en que se pueden ordenar las cuentas de la siguiente población: {🔴,🔴,🔴,🔵,🟢}. Los elementos de esta población, puesto que $N = 5$, se puede ordenar en $5! = 120$ diferentes maneras. No obstante, si las tres cuentas rojas son indistinguibles, cada muestra ordenada se replica 6 veces. Por ejemplo, con la muestra

   (🔴, 🔴, 🔴, 🔵, 🟢)

   corresponden las siguientes muestras indistinguibles:

   (❶, ❷, ❸, 🔵, 🟢) = (❶, ❸, ❷, 🔵, 🟢) = (❷, ❶, ❸, 🔵, 🟢)
   = (❷, ❸, ❶, 🔵, 🟢) = (❸, ❶, ❷, 🔵, 🟢) = (❸, ❷, ❶, 🔵, 🟢).

   Este razonamiento lleva a que el número total de muestras ordenadas *diferentes* (o distinguibles) es $\frac{120}{6} = 20$.

3. En el caso de que no solo haya cuentas rojas indistinguibles, sino también de otros colores, se calcula el número de maneras diferentes en que se pueden ordenar a través del coeficiente multinomial. Por ejemplo, considérese la siguiente población de 7 elementos:

   {🔴,🔴,🔴,🔵,🔵,🟢,🟢}.

Si todas las cuentas fueran distinguibles, entonces habría $7! = 5{,}040$ posibles órdenes. Sin embargo, como las tres cuentas rojas son indistinguibles, tanto como las dos cuentas azules y las dos verdes, este número hay que dividir entre $3! \times 2! \times 2! (= 24)$. Efectivamente, a cada orden subyacen 24 órdenes no distinguibles; por ejemplo, a la muestra

$$(\textcolor{red}{\bullet}, \textcolor{cyan}{\bullet}, \textcolor{green}{\bullet}, \textcolor{red}{\bullet}, \textcolor{red}{\bullet}, \textcolor{green}{\bullet}, \textcolor{cyan}{\bullet})$$

subyacen las muestras indistinguibles:

$$\begin{aligned}
& (\textcolor{red}{\text{❶}}, \textcolor{cyan}{\text{❶}}, \textcolor{green}{\text{❶}}, \textcolor{red}{\text{❷}}, \textcolor{red}{\text{❸}}, \textcolor{green}{\text{❷}}, \textcolor{cyan}{\text{❷}}) = (\textcolor{red}{\text{❶}}, \textcolor{cyan}{\text{❶}}, \textcolor{green}{\text{❷}}, \textcolor{red}{\text{❷}}, \textcolor{red}{\text{❸}}, \textcolor{green}{\text{❶}}, \textcolor{cyan}{\text{❷}}) \\
= & (\textcolor{red}{\text{❶}}, \textcolor{cyan}{\text{❷}}, \textcolor{green}{\text{❶}}, \textcolor{red}{\text{❷}}, \textcolor{red}{\text{❸}}, \textcolor{green}{\text{❷}}, \textcolor{cyan}{\text{❶}}) = (\textcolor{red}{\text{❶}}, \textcolor{cyan}{\text{❷}}, \textcolor{green}{\text{❷}}, \textcolor{red}{\text{❷}}, \textcolor{red}{\text{❸}}, \textcolor{green}{\text{❶}}, \textcolor{cyan}{\text{❶}}) \\
= & (\textcolor{red}{\text{❶}}, \textcolor{cyan}{\text{❶}}, \textcolor{green}{\text{❶}}, \textcolor{red}{\text{❸}}, \textcolor{red}{\text{❷}}, \textcolor{green}{\text{❷}}, \textcolor{cyan}{\text{❷}}) = (\textcolor{red}{\text{❶}}, \textcolor{cyan}{\text{❶}}, \textcolor{green}{\text{❷}}, \textcolor{red}{\text{❸}}, \textcolor{red}{\text{❷}}, \textcolor{green}{\text{❶}}, \textcolor{cyan}{\text{❷}}) \\
= & (\textcolor{red}{\text{❶}}, \textcolor{cyan}{\text{❷}}, \textcolor{green}{\text{❶}}, \textcolor{red}{\text{❸}}, \textcolor{red}{\text{❷}}, \textcolor{green}{\text{❷}}, \textcolor{cyan}{\text{❶}}) = (\textcolor{red}{\text{❶}}, \textcolor{cyan}{\text{❷}}, \textcolor{green}{\text{❷}}, \textcolor{red}{\text{❸}}, \textcolor{red}{\text{❷}}, \textcolor{green}{\text{❶}}, \textcolor{cyan}{\text{❶}}) \\
= & (\textcolor{red}{\text{❷}}, \textcolor{cyan}{\text{❶}}, \textcolor{green}{\text{❶}}, \textcolor{red}{\text{❶}}, \textcolor{red}{\text{❸}}, \textcolor{green}{\text{❷}}, \textcolor{cyan}{\text{❷}}) = (\textcolor{red}{\text{❷}}, \textcolor{cyan}{\text{❶}}, \textcolor{green}{\text{❷}}, \textcolor{red}{\text{❶}}, \textcolor{red}{\text{❸}}, \textcolor{green}{\text{❶}}, \textcolor{cyan}{\text{❷}}) \\
= & (\textcolor{red}{\text{❷}}, \textcolor{cyan}{\text{❷}}, \textcolor{green}{\text{❶}}, \textcolor{red}{\text{❶}}, \textcolor{red}{\text{❸}}, \textcolor{green}{\text{❷}}, \textcolor{cyan}{\text{❶}}) = (\textcolor{red}{\text{❷}}, \textcolor{cyan}{\text{❷}}, \textcolor{green}{\text{❷}}, \textcolor{red}{\text{❶}}, \textcolor{red}{\text{❸}}, \textcolor{green}{\text{❶}}, \textcolor{cyan}{\text{❶}}) \\
= & (\textcolor{red}{\text{❷}}, \textcolor{cyan}{\text{❶}}, \textcolor{green}{\text{❶}}, \textcolor{red}{\text{❸}}, \textcolor{red}{\text{❶}}, \textcolor{green}{\text{❷}}, \textcolor{cyan}{\text{❷}}) = (\textcolor{red}{\text{❷}}, \textcolor{cyan}{\text{❶}}, \textcolor{green}{\text{❷}}, \textcolor{red}{\text{❸}}, \textcolor{red}{\text{❶}}, \textcolor{green}{\text{❶}}, \textcolor{cyan}{\text{❷}}) \\
= & (\textcolor{red}{\text{❷}}, \textcolor{cyan}{\text{❷}}, \textcolor{green}{\text{❶}}, \textcolor{red}{\text{❸}}, \textcolor{red}{\text{❶}}, \textcolor{green}{\text{❷}}, \textcolor{cyan}{\text{❶}}) = (\textcolor{red}{\text{❷}}, \textcolor{cyan}{\text{❷}}, \textcolor{green}{\text{❷}}, \textcolor{red}{\text{❸}}, \textcolor{red}{\text{❶}}, \textcolor{green}{\text{❶}}, \textcolor{cyan}{\text{❶}}) \\
= & (\textcolor{red}{\text{❸}}, \textcolor{cyan}{\text{❶}}, \textcolor{green}{\text{❶}}, \textcolor{red}{\text{❶}}, \textcolor{red}{\text{❷}}, \textcolor{green}{\text{❷}}, \textcolor{cyan}{\text{❷}}) = (\textcolor{red}{\text{❸}}, \textcolor{cyan}{\text{❶}}, \textcolor{green}{\text{❷}}, \textcolor{red}{\text{❶}}, \textcolor{red}{\text{❷}}, \textcolor{green}{\text{❶}}, \textcolor{cyan}{\text{❷}}) \\
= & (\textcolor{red}{\text{❸}}, \textcolor{cyan}{\text{❷}}, \textcolor{green}{\text{❶}}, \textcolor{red}{\text{❶}}, \textcolor{red}{\text{❷}}, \textcolor{green}{\text{❷}}, \textcolor{cyan}{\text{❶}}) = (\textcolor{red}{\text{❸}}, \textcolor{cyan}{\text{❷}}, \textcolor{green}{\text{❷}}, \textcolor{red}{\text{❶}}, \textcolor{red}{\text{❷}}, \textcolor{green}{\text{❶}}, \textcolor{cyan}{\text{❶}}) \\
= & (\textcolor{red}{\text{❸}}, \textcolor{cyan}{\text{❶}}, \textcolor{green}{\text{❶}}, \textcolor{red}{\text{❷}}, \textcolor{red}{\text{❶}}, \textcolor{green}{\text{❷}}, \textcolor{cyan}{\text{❷}}) = (\textcolor{red}{\text{❸}}, \textcolor{cyan}{\text{❶}}, \textcolor{green}{\text{❷}}, \textcolor{red}{\text{❷}}, \textcolor{red}{\text{❶}}, \textcolor{green}{\text{❶}}, \textcolor{cyan}{\text{❷}}) \\
= & (\textcolor{red}{\text{❸}}, \textcolor{cyan}{\text{❷}}, \textcolor{green}{\text{❶}}, \textcolor{red}{\text{❷}}, \textcolor{red}{\text{❶}}, \textcolor{green}{\text{❷}}, \textcolor{cyan}{\text{❶}}) = (\textcolor{red}{\text{❸}}, \textcolor{cyan}{\text{❷}}, \textcolor{green}{\text{❷}}, \textcolor{red}{\text{❷}}, \textcolor{red}{\text{❶}}, \textcolor{green}{\text{❶}}, \textcolor{cyan}{\text{❶}}).
\end{aligned}$$

Por lo tanto el número de maneras en que se pueden ordenar los 7 elementos de la población considerada es igual a:

$$\binom{7}{3,\ 2,\ 2} = \frac{7!}{3! \times 2! \times 2!} = \frac{5040}{24} = 210.$$

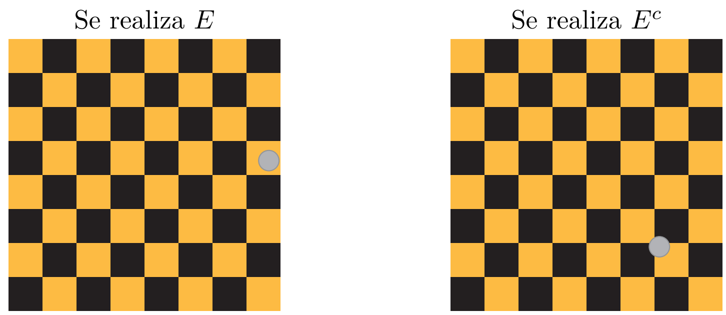
9.6.2.2 Espacios muestrales continuos

Si el número de elementos en el espacio muestral Ω y el evento E es infinito, no se puede aplicar la estrategia explicada al inicio de la sección anterior (ni la Regla de Laplace). En estos casos, es necesario emplear una estrategia diferente para obtener la probabilidad $P(E)$ según el enfoque *a priori*. Ilustramos mediante un ejemplo cómo se puede realizar un análisis lógico para calcular la probabilidad en caso de un espacio muestral infinito e innumerable:

Ejemplo

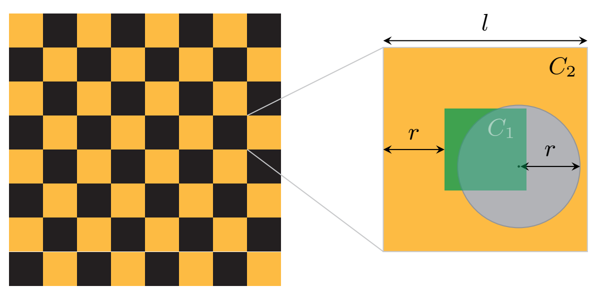
Supongamos que se echa una moneda con radio r sobre un tablero de ajedrez y que se busca la probabilidad del evento E: la moneda se queda perfectamente dentro de uno de los cuadros (con lado igual a l) del tablero. Es decir, el evento E se realiza si la moneda no toca ni se interseca con ningún lado del cuadro.

Por ejemplo, el panel izquierdo de la siguiente figura representa un caso donde el evento E se realiza; en el panel derecho, el evento E no se realiza (y entonces, se realiza su complemento E^c).



Si se consideran como resultados de este experimento aleatorio los puntos del tablero donde se caiga el centro de la moneda, entonces el espacio muestral Ω contiene un número infinito e innumberable de resultados.

Para un análisis lógico del problema, véase la siguiente figura. Se puede derivar que el evento E se realiza si y solo si el centro de la moneda cae dentro del cuadro C_1 (de color verde) de algún cuadro C_2 del tablero, como se ilustra en la siguiente figura:



Considerando el caso de que la moneda se haya caido en el cuadro C_2 del tablero, la probabilidad de que su centro se encuentre dentro del cuadro C_1 (y, por ende, se realice el evento E) se da por:

$$P(E) = \frac{\text{superficie } C_1}{\text{superficie } C_2} = \frac{(l-2r)^2}{l^2}. \tag{9.7}$$

Es importante destacar algunos elementos del análisis anterior:

- Se supone que $l > 2r$ (es decir, que el diámetro de la moneda es menor que el lado de un cuadrado del tablero). Obviamente, en el caso contrario, E representa un evento imposible (es decir, si $l \leqslant 2r$, entonces $P(E) = 0$).
- La Ecuación (9.7) se basa en el supuesto que cada resultado de Ω es *equiprobable*, es decir que el centro de la moneda puede estar en cualquier punto del cuadro C_2 con igual probabilidad.
- Aunque hemos considerado un sólo cuadro del tablero, el resultado es correcto siempre y cuando se supone equiprobabilidad de que se caiga el centro de la moneda en cualquier punto del tablero (puesto que los cuadros del tablero son idénticos y las superficies de C_1 y C_2 son iguales para cualquier cuadro en que se puede caer la moneda.

Nótese que en este ejemplo el cálculo del número de los resultados en el evento E y en Ω (lo que se consideraría en el caso de que estos conjuntos sean finitos) se ha cambiado por el cálculo de las superficies de los cuadros C_1 y C_2.

9.6.3 Probabilidad condicional

Para introducir el concepto de probabilidad condicional, miremos algunos ejemplos.

Ejemplos

Un experimento aleatorio consiste en extraer, de forma aleatoria, una persona de la población mexicana (las que tiene la nacionalidad mexicana); el resultado que se define sobre este experimento aleatorio es el color de los ojos de la persona extraída. Considérese el evento E "la persona extraída tiene ojos azules"; probablemente, la probabilidad de este evento (dentro de la población mexicana) es baja. Por fines de ilustración, supongamos que $P(E) = 0.05$.

Como segundo paso, considérese la subpoblación de mexicanos que tienen cabello rubio. Al extraer aleatoriamente una persona de esta subpoblación y considerando el mismo evento E, es obvio que ya no $P(E) = 0.05$: En la población de mexicanos con cabello rubio, los ojos azules son más recurrentes.

Consideremos el experimento aleatorio de extraer una persona de la población de estudiantes de la Universidad Nacional Autónoma de México y el evento E "el estudiante extraído ha fumado un cigarro, al menos una vez en su vida". Supongamos que la probabilidad del evento es igual a 0.40.

Supongamos que recibimos información adicional sobre el estudiante extraído, a saber: "Alguna vez ha probado la marihuana". ¿Todavía creeríamos que la probabilidad de que haya fumado un cigarro al menos una vez en su vida sea 0.40? Probablemente no, pues, entre personas que han probado marihuana, es más común que también han fumado tabaco. Este ejemplo muestra que la probabilidad de fumar es más alta en la subpoblación de estudiantes que han fumado tabaco, que en la población total de estudiantes.

Similarmente, si se limita la población de todos los estudiantes de la UNAM a los que estudian medicina, posiblemente la probabilidad de fumar tabaco es más baja (si es cierto que los médicos conocen mejor las consecuencias nocivas para la salud y este conocimiento influye en los hábitos de fumar).

Para un ejemplo más elaborado, considérese la población de los 1,384 estudiantes matriculados en el primer año de la carrera de la Facultad de Medicina de la UNAM durante el año académico 2013–2014. Se extrae de forma aleatoria un estudiante y se registran (a) si es un estudiante perteneciente al PAEA (Programa de Alta Exigencia Académica), un estudiante regular, o bien un estudiante del grupo de repetidores y (b) si aprueba o no todas las asignaturas del primer año. La siguiente tabla presenta los resultados en el espacio muestral y el número de estudiantes correspondientes en la población:

Resultados en Ω	Representados por	Resultados en Ω	Representados por
(PAEA, No aprobó)	17 estudiantes	(PAEA, Sí aprobó)	105 estudiantes
(Regular, No aprobó)	562 estudiantes	(Regular, Sí aprobó)	389 estudiantes
(Repetidor, No aprobó)	161 estudiantes	(Repetidor, Sí aprobó)	150 estudiantes

A continuación, considérense los siguientes dos eventos:

1. A: El estudiante aprueba todas las materias
2. B: El estudiante pertenece al PAEA

Si todos los alumnos tienen la misma probabilidad de ser extraídos, entonces se obtiene (mediante la regla de Laplace):

$$P(A) = \frac{105 + 389 + 150}{1384} = \frac{644}{1384} = .465$$

$$P(B) = \frac{17 + 105}{1384} = \frac{122}{1384} = .088$$

A continuación, calculamos la probabilidad de que un estudiante apruebe todas las materias, pero sólo para la subpoblación de estudiantes PAEA. Es decir, consideramos la probabilidad del evento A, condicional al evento B, la cual se escribe como $P(A|B)$. Con los números en la tabla anterior, se obtiene:

$$P(A|B) = \frac{105}{17 + 105} = \frac{105}{122} = .861.$$

Comparando la probabilidad condicional $P(A|B)$ con la probabilidad incondicional $P(A)$ (también llamada la probabilidad *marginal*), se observa que la probabilidad de aprobar todas las materias es considerablemente más alta en la subpoblación de estudiantes PAEA que en la población total.

De forma similar, se puede calcular la probabilidad del evento A, condicional a que el estudiante no pertenece al PAEA:

$$P(A|B^c) = \frac{289 + 150}{562 + 389 + 161 + 150} = \frac{439}{1262} = .348.$$

Probabilidad condicional

La *probabilidad del evento A condicional al evento B*, anotado como $P(A|B)$, se define como:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad \text{si } P(B) \neq 0. \tag{9.8}$$

Si $P(B) = 0$, la probabilidad condicional $P(A|B)$ no está definida.

Al considerar la probabilidad condicional $P(A|B)$, el evento B juega el papel de un *espacio muestral reducido*. Pues, $P(A|B)$ se interpreta como la probabilidad de que se realice el evento A después de haber reducido el espacio muestral a los resultados que pertenecen al evento B. En algunas ocasiones, si es deseable o necesario explicitar que se toma en cuenta el espacio muestral original Ω para la probabilidad de A, se escribe $P(A|\Omega)$. Entonces, $P(A|\Omega)$ es nada más una notación más explícita para $P(A)$.

La definición indica que la probabilidad condicional $P(A|B)$ es la probabilidad de que el evento A se realice, dado que o sabiendo que el evento B se ha realizado. Por lo tanto, en el numerador de la Ecuación (9.8) se consideran todos los resultados en de A que están en B, o bien, los resultados en el evento compuesto $A \cap B$. Para la probabilidad $P(A|B)$ se toma la razón entre $P(A \cap B)$ y $P(B)$.

Es importante enfatizar que, en general, la probabilidad de A condicional a B es diferente de la probabilidad de B condicional a A. Es decir, en general tenemos $P(A|B) \neq P(B|A)$. En la Sección 9.6.4 se introducirá el teorema de Bayes, el cual relaciona $P(A|B)$ y $P(B|A)$.

Ley de la Probabilidad Total

Sean $A_1, A_2, \ldots, A_m$ eventos en el espacio muestral Ω que cumplen las siguientes tres condiciones:

(1) La probabilidad de cualquier de estos eventos es mayor que 0 (es decir, $P(A_j) > 0$ para $j = 1, \ldots, m$);

(2) Estos eventos son mutuamente exclusivos (es decir, $A_j \cap A_k = \varnothing$ para cualquier $j \neq k$ y $j, k = 1, \ldots, m$);

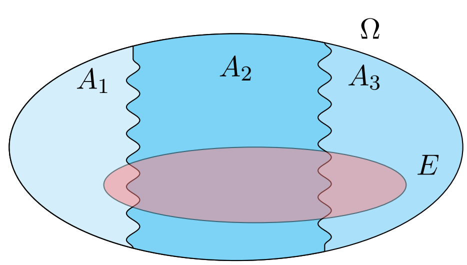
(3) Conjuntamente, forman un evento seguro (es decir, $A_1 \cup A_2 \cup \cdots \cup A_m = \Omega$).

Entonces, para cualquier evento $E \subseteq \Omega$, se satisface:

$$P(E) = P(E|A_1)\,P(A_1) + P(E|A_2)\,P(A_2) + \cdots + P(E|A_m)\,P(A_m).$$

Si los eventos $A_1, A_2, \ldots, A_m$ cumplen los requisitos (1), (2) y (3) del teorema anterior, se dice que forman una *partición* de Ω. Una partición, entonces, parte el espacio muestral en m piezas no vacías que no traslapan y que conjuntamente cubren el espacio muestral completo.

La siguiente gráfica ilustra las condiciones para la Ley de la Probabilidad Total. La elipse grande representa el espacio muestral Ω y está particionado en tres eventos A_1, A_2 y A_3. Efectivamente, los tres eventos forman una partición del espacio muestral, ya que las tres partes no son vacías, no traslapan (es decir, $A_1 \cap A_2 = \varnothing$, $A_1 \cap A_3 = \varnothing$ y $A_2 \cap A_3 = \varnothing$), y conjuntamente cubren la elipse total (es decir, $A_1 \cup A_2 \cup A_3 = \Omega$). Por consiguiente, se cumplen las condiciones para aplicar la Ley de la Probabilidad de Total. Esto quiere decir que para cualquier evento E en el espacio muestral se puede calcular $P(E)$ a partir de las probabilidades condicionales $P(E|A_1)$, $P(E|A_2)$ y $P(E|A_3)$ y las probabilidades incondicionales de los eventos que forman la partición $P(A_1)$, $P(A_2)$ y $P(A_3)$.



Prueba de la Ley de la Probabilidad Total

Se consideran para cada uno de los eventos $A_1, A_2, \dots, A_m$ la parte que tiene en común con el evento E; es decir, se consideran las intersecciones $A_1 \cap E, A_2 \cap E, \dots, A_m \cap E$. Puesto que $A_1, A_2, \dots, A_m$ son mutuamente exclusivos y su unión es Ω, es cierto que:

$$E \;=\; (A_1 \cap E) \,\cup\, (A_2 \cap E) \,\cup\, \dots \,\cup\, (A_m \cap E).$$

(Este resultado se ilustra gráficamente en la figura anterior. Se observa que el evento E (en color rojo) es la unión de las respectivas partes que tiene en común con los eventos A_1, A_2 y A_3.)

Entonces, para la probabilidad del evento E en Ω, se obtiene:

$$\begin{aligned} P(E) \;&=\; P\left[(A_1 \cap E) \,\cup\, (A_2 \cap E) \,\cup\, \dots \,\cup\, (A_m \cap E)\right] \\ &=\; P(A_1 \cap E) \,+\, P(A_2 \cap E) \,+\, \cdots \,+\, P(A_m \cap E) \end{aligned} \tag{9.9}$$

La última ecuación sigue de la propiedad en la Ecuación (9.2b), ya que los eventos compuestos $(A_1 \cap E), (A_2 \cap S), \dots, (A_m \cap E)$ son mutuamente exclusivos (como consecuencia de que $A_1, A_2, \dots, A_m$ son mutuamente exclusivos).

Reconsiderando la definición de probabilidad condicional, es claro que:

$$\begin{aligned} P(A|B) \;&=\; \frac{P(A \cap B)}{P(B)} \\ \Longleftrightarrow \quad P(A \cap B) \;&=\; P(A|B)P(B). \end{aligned}$$

Por lo tanto, la Ecuación (9.9)

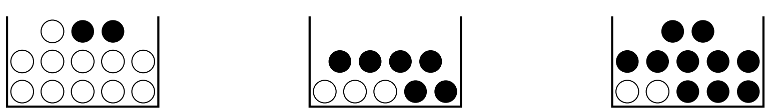
$$P(E) \;=\; \underbrace{P(A_1 \cap E)}_{=P(E|A_1)\,P(A_1)} \;+\; \underbrace{P(A_2 \cap E)}_{=P(E|A_2)\,P(A_2)} \;+\cdots+\; \underbrace{P(A_m \cap E)}_{=P(E|A_m)\,P(A_m)}$$

lleva inmediatemente a la Ley de la Probabilidad Total. ■

Ejemplos

Para ilustrar la Ley de la Probabilidad Total, consideremos el siguiente ejemplo artificial:
Un experimento aleatorio consiste en extraer, de forma aleatoria y sin reposición, cuatro cuentas de una de tres urnas. La primera urna (Urna A) contiene 11 cuentas blancas y 2 negras, la segunda urna (Urna B) contiene 3 cuentas blancas y 6 negras, y la tercera urna (Urna C) contiene 2 cuentas blancas y 10 negras:

Urna A **Urna B** **Urna C**



Para determinar de qué urna se extraen las cuatro cuentas, se lanza un dado: Si sale 1, se extraen de la Urna A; si sale 2 ó 3, se extraen de la Urna B; si sale 4, 5 ó 6, se extraen de la Urna C. Representando los eventos “las cuentas se extraen de la Urna A”, “las cuentas se extraen de la Urna B”, y “las cuentas se extraen de la Urna C” por A, B, y C, respectivamente, tenemos:

$$P(A) \;=\; \tfrac{1}{6} \qquad\qquad P(B) \;=\; \tfrac{1}{3} \qquad\qquad P(C) \;=\; \tfrac{1}{2}.$$

Puesto que (a) los eventos A, B y C son mutuamente exclusivos (ya que no se pueden realizar dos de estos eventos en una ejecución del experimento aleatorio) y (b) conjuntamente forman un evento seguro (ya que siempre se realiza uno de estos tres eventos), se satisfacen las dos condiciones para aplicar la Ley de la Probabilidad Total. Por ejemplo, nos podemos preguntar cuál sería la probabilidad del evento $E_{\{\circ\circ\bullet\bullet\}}$ “la muestra de cuatro cuentas contiene dos blancas y dos negras”. La Ley de la Probabilidad Total, aplicada a este problema, es como sigue:

$$P(E_{\{\circ\circ\bullet\bullet\}}) \;=\; P(E_{\{\circ\circ\bullet\bullet\}}|A)\,P(A) \;+\; P(E_{\{\circ\circ\bullet\bullet\}}|B)\,P(B) \;+\; P(E_{\{\circ\circ\bullet\bullet\}}|C)\,P(C).$$

La ecuación anterior incluye las probabilidades condicionales $P(E_{\{\circ\circ\bullet\bullet\}}|A)$, $P(E_{\{\circ\circ\bullet\bullet\}}|B)$ y $P(E_{\{\circ\circ\bullet\bullet\}}|C)$. Estas probabilidades se pueden calcular utilizando los principios de la combinatoria y la Regla de Laplace (ya que se extraen las cuentas de forma aleatoria, lo cual quiere decir que todas las muestras de cuatro cuentas son equiprobables). Por ejemplo, la probabilidad de que se realice el evento $E_{\{\circ\circ\bullet\bullet\}}$ si la muestra de cuatro cuentas se extrae de la Urna A es igual a:

$$\begin{aligned} P(E_{\{\circ\circ\bullet\bullet\}}|A) \;&=\; \frac{\binom{11}{2}\binom{2}{2}}{\binom{13}{4}} \\ &= \frac{55 \times 1}{715} \\ &= .077. \end{aligned}$$

Efectivamente, la Regla 4 de la combinatoria implica que se pueden sacar $\binom{13}{4} = 715$ muestras de 4 cuentas de una población de 13 cuentas. Para que se realice el evento $E_{\{\circ\circ\bullet\bullet\}}$ se deben extraer (a) dos cuentas blancas de las 11 blancas que hay en la población de la Urna A y (b) las dos cuentas negras que hay en esta población. Puesto que hay $\binom{11}{2} = 55$ maneras para sacar las dos cuentas blancas de las 11 y solo una manera para sacar las dos cuentas negras, el número de muestras que incluyen dos cuentas blancas y dos cuentas negras es 55. Si las 715 muestras son equiprobables, la probabilidad del evento $E_{\{\circ\circ\bullet\bullet\}}$ efectivamente es $\frac{55}{715}$ por la Regla de Laplace.

De la misma forma, se puede derivar la probabilidad del evento $E_{\{\circ\circ\bullet\bullet\}}$ condicional a que se extraen las cuatro cuentas de la Urna B y condicional a que se extraen de la Urna C:

$$\begin{aligned} P(E_{\{\circ\circ\bullet\bullet\}}|B) \;&=\; \frac{\binom{3}{2}\binom{6}{2}}{\binom{9}{4}} \\ &= \frac{3 \times 15}{126} \\ &= .357 \end{aligned}$$

$$\text{y} \qquad P(E_{\{\circ\circ\bullet\bullet\}}|C) = \frac{\binom{2}{2}\binom{10}{2}}{\binom{12}{4}} = \frac{1 \times 45}{495} = .091.$$

Por lo tanto, la probabilidad del evento $E_{\{\circ\circ\bullet\bullet\}}$ según la Ley de la Probabilidad Total es:

$$\begin{aligned} P(E_{\{\circ\circ\bullet\bullet\}}) &= P(E_{\{\circ\circ\bullet\bullet\}}|A)\,P(A) \;+\; P(E_{\{\circ\circ\bullet\bullet\}}|B)\,P(B) \;+\; P(E_{\{\circ\circ\bullet\bullet\}}|C)\,P(C). \\ &= .077 \times \tfrac{1}{6} \;+\; .357 \times \tfrac{1}{3} \;+\; .091 \times \tfrac{1}{2} \\ &= .177. \end{aligned}$$

Es decir, considerando el experimento aleatorio en su totalidad (lo cual implica lanzar el dado para seleccionar la urna y subsecuentemente extraer una muestra de cuatro cuentas de la urna seleccionada), la probabilidad de que la muestra extraída contenga dos cuentas blancas y dos negras es .177.

Para un ejemplo de más relevancia médica, considérese un laboratorio donde se aplica la prueba ELISA a individuos que sospechan que son portadores del VIH. Con base en experiencia previa se sabe que en la población de individuos que acuden al laboratorio, el 1 % de las personas es portador del VIH. Si denominamos el evento “la persona es portador del virus” como A, entonces

$$P(A) = .01 \qquad \text{y} \qquad P(A^c) = .99.$$

Por otro lado, definamos los eventos complementarios E^+, “la prueba ELISA da un resultado positivo”, y E^-, “la prueba ELISA da un resultado negativo”. Si llega al laboratorio una persona para aplicarse la prueba ELISA, ¿cuál es la probabilidad de que se presente un resultado positivo, $P(E^+)$?

Estudios clínicos en la fase de la validación de la prueba ELISA mostraron que su sensibilidad y especificidad son .995 y .992, respectivamente. La sensibilidad de una prueba diagnóstica conceptualmente es la probabilidad de que dé un resultado positivo en casos que son realmente enfermos; aplicado a este ejemplo, corresponde formalmente con la probabilidad condicional $P(E^+|A)$. Similarmente, la especificidad es la probabilidad de obtener un resultado negativo en casos que son realmente sanos; es decir, en nuestro ejemplo corresponde con $P(E^-|A^c)$. Por lo tanto, la información sobre la sensibilidad y especificidad de la prueba ELISA nos indica que:

$$P(E^+|A) = .995 \qquad \text{y} \qquad P(E^-|A^c) = .992$$

y puesto que E^+ y E^- son eventos complementarios:

$$P(E^-|A) = .005 \qquad \text{y} \qquad P(E^+|A^c) = .008.$$

Las últimas dos probabilidades se pueden interpretar como la probabilidad de un falso negativo y un falso positivo, respectivamente.

A partir de esta información es posible calcular $P(E^+)$ aplicando la Ley de la Probabilidad Total:

$$P(E^+) = P(E^+|A)\,P(A) + P(E^+|A^c)\,P(A^c),$$

ya que A y A^c son eventos exclusivos y su unión abarca todas las posibilidades. Por lo tanto, se obtiene que:

$$\begin{aligned} P(E^+) &= .995 \times .01 + .008 \times .99 \\ &= .018. \end{aligned}$$

Es decir, una persona que acude al laboratorio para la prueba ELISA tiene una probabilidad de 1.8 % de que su resultado en la prueba sea positiva.

Es muy instructivo comparar los dos ejemplos anteriores. Formalmente, se puede conceptualizar el ejemplo de la prueba ELISA en términos de un experimento aleatorio que consiste en extraer una cuenta de una de dos cajas que contienen cuentas negras y blancas. Una cuenta negra corresponde con un resultado positivo en la prueba ELISA, una cuenta blanca con un resultado negativo. La primera caja representa el caso de que el paciente es portador del VIH y contiene 995 cuentas negras y 5 blancas. Efectivamente, si el paciente es portador del VIH, entonces la probabilidad de un resultado positivo es .995. Por otro lado, la segunda caja representa el caso de un paciente que no es portador del VIH y contiene 992 cuentas blancas y 8 negras, tal que la probabilidad de un resultado negativo en este caso es .992. La probabilidad de que se extraiga la cuenta de la primera o la segunda caja corresponde con las probabilidades de .01 y .99, respectivamente.

9.6.4 El teorema de Bayes

En muchas ocasiones, se conoce la probabilidad $P(A|B)$ de un evento A condicional a otro evento B, pero se desea conocer la probabilidad condicional inversa, $P(B|A)$. En el ejemplo anterior, los estudios clínicos nos informaron sobre la sensibilidad de la prueba ELISA, la cual nos da la probabilidad de un resultado positivo si el paciente es portador del VIH. Sin embargo, en la práctica, se desea conocer la probabilidad de que un paciente en el cual se obtuvo un resultado positivo en la prueba, efectivamente sea portador del VIH. En la simbología del ejemplo anterior, la sensibilidad corresponde con la probabilidad condicional $P(E^+|A)$, mientras que la probabilidad de interés es la probabilidad $P(A|E^+)$. Como se enfatizó anteriormente, estas probabilidades difieren. Por lo tanto, la pregunta de cómo se relacionan y cómo se puede conocer una a partir de la otra, es de mucha relevancia práctica. El teorema de Bayes da la respuesta.

Teorema de Bayes

El Teorema de Bayes viene en tres variantes:

Variante 1: Sean A y B dos eventos en el espacio muestral Ω, para los cuales $P(A) > 0$ y $P(B) > 0$. Entonces, se satisface:

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}. \tag{9.10a}$$

Variante 2: Sean A y B dos eventos en el espacio muestral Ω, para los cuales $P(A) > 0$ y $0 < P(B) < 1$. Entonces, se satisface:

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A|B)\,P(B) + P(A|B^c)\,P(B^c)}. \tag{9.10b}$$

Variante 3: Sean $B_1, B_2, \ldots, B_m$ eventos que forman una partición del espacio muestral Ω y A cualquier otro evento en Ω. Entonces, para cualquier de los eventos B_j $(j = 1, \ldots, m)$ se satisface:

$$P(B_j|A) = \frac{P(A|B_j)\,P(B_j)}{P(A|B_1)\,P(B_1) + P(A|B_2)\,P(B_2) + \cdots + P(A|B_m)\,P(B_m)}. \tag{9.10c}$$

Prueba del Teorema de Bayes

A partir de la definición de $P(A|B)$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

sigue que

$$P(A \cap B) = P(A|B)\,P(B).$$

Similarmente, se tiene

$$\begin{aligned} P(B|A) &= \frac{P(A \cap B)}{P(A)} \\ \Longleftrightarrow \quad P(A \cap B) &= P(B|A)\,P(A). \end{aligned}$$

Por lo tanto,

$$P(A|B)\,P(B) = P(B|A)\,P(A)$$

lo cual lleva directamente a la Ecuación (9.10a):

$$P(A|B) \;=\; \frac{P(B|A)\,P(A)}{P(B)}.$$

La segunda y tercera variante del Teorema de Bayes se derivan después de aplicar el Teorema de la Probabilidad Total para sustituir $P(A)$ en la Variante 1 por una expresión equivalente. En la Variante 2, B y B^c forman, por definición, una partición de Ω, tal que:

$$P(A) \;= P(A|B)\,P(B) \;+\; P(A|B^c)\,P(B^c)$$

y para la Variante 3, se considera una partición $B_1, B_2, \ldots, B_m$ del espacio muestral, tal que:

$$P(A) \;= P(A|B_1)\,P(B_1) \;+\; P(A|B_2)\,P(B_2) \;+\; \cdots \;+\; P(A|B_m)\,P(B_m).$$

Efectivamente, sustituyendo $P(A)$ en el denominador de la Ecuación (9.10a) por las dos expresiones anteriores lleva directamente a las Ecuaciones (9.10b) y (9.10c), respectivamente. ■

Ejemplos

Retomemos los dos ejemplos que introdujimos para ilustrar la Ley de la Probabilidad Total.

Supongamos que, en el primer ejemplo, se observa como resultado del experimento aleatorio una muestra de dos cuentas blancas y dos negras, es decir, que se ha realizado el evento $E_{\{\circ\circ\bullet\bullet\}}$. Dado esta información, ¿cuál sería la probabilidad de que la muestra se haya extraído de la Caja A?

Aplicando la Variante 3 del Teorema de Bayes, se obtiene:

$$P(A|E_{\{\circ\circ\bullet\bullet\}}) \;=\; \frac{P(E_{\{\circ\circ\bullet\bullet\}}|A)\,P(A)}{P(E_{\{\circ\circ\bullet\bullet\}}|A)\,P(A) \;+\; P(E_{\{\circ\circ\bullet\bullet\}}|B)\,P(B) \;+\; P(E_{\{\circ\circ\bullet\bullet\}}|C)\,P(C)},$$

lo cual con las probabilidades derivadas anteriormente lleva a:

$$P(A|E_{\{\circ\circ\bullet\bullet\}}) \;=\; \frac{.077 \times \frac{1}{6}}{.077 \times \frac{1}{6} + .357 \times \frac{1}{3} + .091 \times \frac{1}{2}} \;=\; \frac{.013}{.177} \;=\; .072.$$

Similarmente, se puede derivar la probabilidad de que se haya extraído la muestra de la Caja B:

$$P(B|E_{\{\circ\circ\bullet\bullet\}}) \;=\; \frac{.357 \times \frac{1}{3}}{.077 \times \frac{1}{6} + .357 \times \frac{1}{3} + .091 \times \frac{1}{2}} \;=\; \frac{.119}{.177} \;=\; .671$$

y la probabilidad de que se haya extraído de la Caja C:

$$P(C|E_{\{\circ\circ\bullet\bullet\}}) \;=\; \frac{.091 \times \frac{1}{2}}{.077 \times \frac{1}{6} + .357 \times \frac{1}{3} + .091 \times \frac{1}{2}} \;=\; \frac{.045}{.177} \;=\; .256.$$

Nótese que el resultado (.177) que se obtiene cada vez en el denominador, es precisamente la probabilidad incondicional del evento $E_{\{\circ\circ\bullet\bullet\}}$, tal y como se calculó en la página 330. Como conclusión de este ejercicio, debe ser claro que la información sobre el evento $E_{\{\circ\circ\bullet\bullet\}}$ cambió nuestra apreciación respecto de las probabilidades de los eventos A, B, y C. Efectivamente, los expertos en estadística Bayesiana suelen interpretar el teorema de Bayes en términos de cómo una *probabilidad previa* (o *a priori*) se transforma, con base en información adicional (en una prueba, un estudio, etc.) a una *probabilidad posterior* (o *a posteriori*). En nuestro ejemplo, sin tener la información que ocurrió $E_{\{\circ\circ\bullet\bullet\}}$, las probabilidades previas de los eventos A, B, y C son $\frac{1}{6}$, $\frac{1}{3}$, y $\frac{1}{2}$, respectivamente; al sí contar con esta información, las probabilidades posteriores de estos eventos son .072, .671 y .256, respectivamente.

En el segundo ejemplo, sobre la prueba ELISA aplicada en un laboratorio médico, una pregunta de mucha relevancia clínica refiere a la probabilidad de que un paciente efectivamente sea portador del VIH posterior a haber observado un resultado positivo en la prueba. Con los datos mencionados previamente, el Teorema de Bayes (Variante 2) nos da la respuesta:

$$\begin{aligned} P(A|E^+) &= \frac{P(E^+|A)\,P(A)}{P(E^+|A)\,P(A) \;+\; P(E^+|A^c)\,P(A^c)} \\ &= \frac{.995 \times .01}{.995 \times .01 + .008 \times .99} \\ &= \frac{.010}{.018} \\ &= .557. \end{aligned}$$

Este resultado implica que en los casos donde la prueba ELISA dé un resultado positivo todavía existe una probabilidad sustancial de 44 % de que la persona *no* es portadora del virus y que, por lo tanto exámenes adicionales son necesarios para llegar a una conclusión definitiva sobre la situación médica del paciente. Por un lado, es sorprendente que, a pesar de la alta sensibilidad y especificidad de la prueba ELISA, un resultado positivo no es resultado concluyente y que se requieren exámenes adicionales. Por otro lado, el resultado positivo en la prueba modificó la muy baja probabilidad previa (de .01) a una probabilidad posterior considerablemente más alta (de .557). Este aumento aclara en qué sentido el resultado de la prueba ELISA es informativa sobre la situación médica del paciente.

9.7 Independencia de eventos

Independencia de dos eventos

Dos eventos A y B en el espacio muestral Ω son *independientes* si y sólo si se cumple cualquiera de las siguientes tres condiciones:

$$\text{(a)}\quad \text{Si } P(B) \neq 0, \ \text{entonces } P(A|B) = P(A) \tag{9.11a}$$

$$\text{(b)}\quad \text{Si } P(A) \neq 0, \ \text{entonces } P(B|A) = P(B) \tag{9.11b}$$

$$\text{(c)}\quad P(A \cap B) = P(A)\,P(B). \tag{9.11c}$$

Similar a las condiciones para independencia de *variables* en una muestra (véase la página 160), es importante tener claro que las tres condiciones en la definición anterior son condiciones equivalentes. Es decir, si una de ellas se cumple, entonces las otras dos también se cumplen (y si una *no* se cumple, tampoco las otras dos se cumplen). Por lo tanto, en determinado contexto, es suficiente verificar la condición que más convenga para llegar a una conclusión sobre la (in)dependencia. A continuación, se presenta una prueba formal de que las tres condiciones siempre llegan a la misma conclusión.

Equivalencia de las tres condiciones para la independencia de dos eventos

Demostramos la equivalencia de las tres condiciones:

$$\begin{array}{rrcll}
 & P(A|B) = & P(A) & (9.11a) & \\
\Longleftrightarrow & \dfrac{P(A\cap B)}{P(B)} = & P(A) & & \text{(a)} \Leftrightarrow \text{(c)} \\
\Longleftrightarrow & P(A\cap B) = & P(A)\,P(B) & (9.11c) & \text{(a)} \Leftrightarrow \text{(b)} \\
\Longleftrightarrow & \dfrac{P(A\cap B)}{P(A)} = & P(B) & & \text{(b)} \Leftrightarrow \text{(c)} \\
\Longleftrightarrow & P(B|A) = & P(B) & (9.11b) &
\end{array}$$

En las primeras dos ecuaciones, hay que añadir que $P(B) > 0$ ya que la probabilidad condicional y la división entre $P(B)$ es indefinida si $P(B) = 0$. Similarmente, hay que añadir que $P(A) > 0$ en las últimas dos ecuaciones. En el caso especial de que $P(A) = 0$ y/o $P(B) = 0$, las tres condiciones son siempre verdaderas y, por lo tanto, equivalentes. ■

Ejemplo

Considérese el experimento aleatorio de lanzar un dado rojo y un dado azul, y como resultado el par de números que están arriba en los dos dados. Definamos los siguientes eventos:

Evento A: "El número en el dado rojo es impar". Corresponde con el siguiente subconjunto de Ω:

$$A = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6)\}.$$

Evento B: "La suma de los números en los dos dados es 5". Este evento contiene los siguientes resultados de Ω:

$$B = \{(1,4), (2,3), (3,2), (4,1)\}.$$

Evento C: "Ambos dados muestran el mismo número", lo cual corresponde con:

$$C = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}.$$

Supongamos que los 36 resultados en Ω son equiprobables, así que podemos calcular las probabilidades dentro de Ω utilizando la Regla de Laplace. Evaluemos la independencia de los eventos A, B, y C:

- Los eventos A y B son independientes ya que $P(A|B) = P(A)$ (véase la Condición (a)):

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{\#(A\cap B)}{\#B} = \frac{2}{4} = \frac{1}{2}$$
$$\shortparallel$$
$$P(A) = \frac{\#A}{\#\Omega} = \frac{18}{36} = \frac{1}{2}.$$

- Los eventos A y C son independientes, ya que $P(A|C) = P(A)$:

$$P(A|C) = \frac{P(A\cap C)}{P(C)} = \frac{\#(A\cap C)}{\#C} = \frac{3}{6} = \frac{1}{2}$$
$$\shortparallel$$
$$P(A) = \frac{1}{2}.$$

- Los eventos B y C no son independientes, ya que $P(B|C) \neq P(B)$:

$$P(B|C) = \frac{P(B\cap C)}{P(C)} = \frac{\#(B\cap C)}{\#C} = \frac{0}{6} = 0$$
$$\nparallel$$
$$P(B) = \frac{\#B}{\#\Omega} = \frac{4}{36} = \frac{1}{9}.$$

En términos Bayesianos, las Ecuaciones (9.11a) y (9.11b) dicen que dos eventos son independientes si la probabilidad *posterior* es la misma que la probabilidad *previa*. Efectivamente, es común interpretar la independencia de dos eventos como si información adicional sobre uno de ellos no cambia nuestra apreciación respecto de la probabilidad de que ocurra el otro. Esta interpretación es evidente en los ejemplos anteriores: Por ejemplo, la probabilidad previa de que ambos dados muestren el mismo número es $\frac{1}{6}$. Si alguien después de haber lanzado los dos dados revela que el dado rojo muestra un número impar, no se modificará la probabilidad de que ambos dados muestren el mismo número (sigue siendo $\frac{1}{6}$); por lo tanto los eventos A y C son independientes. Por otro lado, si supiéramos que la suma de los números en los dos dados es 5, entonces sabríamos con certeza que los dados no muestran el mismo número; por ende, los evento B y C son dependientes.

Es importante no confundir eventos *independientes* con eventos *exclusivos* (lo cual puede resultar de una interpretación vaga de estas nociones como eventos "que no tienen nada que ver el uno con el otro"). No obstante, dos eventos exclusivos A y B *no* son independientes si $P(A) > 0$ y $P(B) > 0$. Pues, en este caso $P(A)\,P(B) > 0$ mientras que, por definición la intersección de dos eventos exclusivos es vacía, por lo cual $P(A \cap B) = 0$. Es decir, si A y B son exclusivos y no representan eventos imposibles , entonces $P(A \cap B) \neq P(A)\,P(B)$, lo cual implica que no son independientes.

La noción de independencia introducida en las páginas anteriores se generaliza también a más de dos eventos. Aunque es posible dar una definición más formal (utilizando expresiones y símbolos matemáticos), aquí se presenta, por conveniencia, solo en palabras:

Independencia de m eventos

Los eventos $A_1, A_2, \ldots, A_m$, definidos en el espacio muestral Ω, son *conjuntamente independientes* si y solo si para *cualquier* subconjunto de estos m eventos la probabilidad de su intersección es igual al producto de las probabilidades de los eventos individuales.

Ejemplos

Se aplica la definición anterior para evaluar la independencia conjunta de tres eventos definidos en el contexto del mismo experimento aleatorio del ejemplo anterior, donde el espacio muestral Ω contiene los 36 resultados equiprobables conformados por los 36 pares de números que se pueden observar en los dados rojo y azul.

Definamos los siguientes eventos:

Evento A: "La suma de los dos números es 7":

$$A = \{(1,6),\ (2,5),\ (3,4),\ (4,3),\ (5,2),\ (6,1)\}.$$

Evento B: "El dado rojo muestra el número 3":

$$B = \{(3,1),\ (3,2),\ (3,3),\ (3,4),\ (3,5),\ (3,6)\}.$$

Evento C: "El dado azul muestra el número 4":

$$C = \{(1,4),\ (2,4),\ (3,4),\ (4,4),\ (5,4),\ (6,4)\}.$$

La definición anterior implica que para concluir que los tres eventos son conjuntamente independientes es necesario que se cumplan las siguientes cuatro condiciones:

(1) $P(A \cap B) = P(A)\,P(B)$.
Aplicando al ejemplo anterior, se verifica ya que, por un lado, $P(A \cap B) = \frac{1}{36}$ (únicamente el resultado $(3, 4)$ se encuentra tanto en A como en B) y, por otro lado, $P(A)\,P(B) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$.

(2) $P(A \cap C) = P(A)\,P(C)$.
Se verifica ya que, por un lado, $P(A \cap C) = \frac{1}{36}$ y, por otro lado, $P(A)\,P(C) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$.

(3) $P(B \cap C) = P(B)\,P(C)$.
Se verifica ya que, por un lado, $P(B \cap C) = \frac{1}{36}$ y, por otro lado, $P(B)\,P(C) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$.

(4) $P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$.
Esta condición *no* se cumple: La intersección $A \cap B \cap C$ contiene un elemento (el resultado $(3, 4)$), por lo cual $P(A \cap B \cap C) = \frac{1}{36}$. Puesto que el producto de las probabilidades de los eventos individuales, dado por $P(A)\,P(B)\,P(C) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{216}$, es diferente de $\frac{1}{36}$, los tres eventos no son conjuntamente independientes.

El último ejemplo ilustró que, en el caso de tres eventos, la independencia por pares (que se consigue cuando se cumplen las condiciones (1) a (3)) *no* es suficiente para la independencia conjunta. Por lo tanto, para llegar a la conclusión de independencia conjunta, hay que evaluar cada una de las cuatro condiciones mencionadas. El siguiente ejemplo muestra que tampoco la cuarta condición ($P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$) es suficiente en sí misma para la independencia conjunta.
Considérense los siguientes eventos:
Evento A: "El número en el dado rojo es menor que o igual a 3":

$$A = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6)\}.$$

Evento B: "El dado rojo muestra el número 3, 4 ó 5":

$$B = \{(3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6)\}.$$

Evento C: "La suma de los números en los dos dados es igual a 9":

$$C = \{(3,6), (4,5), (5,4), (6,3)\}.$$

A pesar de que se cumple que $P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$:

$$P(A \cap B \cap C) = \frac{\#(A \cap B \cap C)}{\#\Omega} = \frac{1}{36}$$

y

$$\begin{aligned} P(A)\,P(B)\,P(C) &= \frac{\#A}{\#\Omega}\,\frac{\#B}{\#\Omega}\,\frac{\#C}{\#\Omega} \\ &= \frac{18}{36}\cdot\frac{18}{36}\cdot\frac{4}{36} \\ &= \frac{1}{36} \end{aligned}$$

los tres eventos no son conjuntamente independientes debido a que no son independientes por pares. Por ejemplo, $P(A\cap B) \neq P(A)\,P(B)$:

$$P(A\cap B) = \frac{\#(A\cap B)}{\#\Omega} = \frac{6}{36} = \frac{1}{6}$$

mientras que

$$\begin{aligned} P(A)\,P(B) &= \frac{\#A}{\#\Omega}\,\frac{\#B}{\#\Omega} \\ &= \frac{18}{36}\cdot\frac{18}{36} \\ &= \frac{1}{4}. \end{aligned}$$

Problemas

1. Se lanzan dos monedas, una de dos pesos y otra de cinco pesos. A partir de este experimento aleatorio, definiendo como resultado la cara que muestra cada moneda después de haberse detenido, define

   (a) el espacio muestral Ω;
   (b) una variable aleatoria X: "Número de monedas en que ha salido cruz";
   (c) la clase de los eventos $\mathscr{C}$;
   (d) la función de probabilidad, suponiendo que las monedas son imparciales;
   (e) la función de probabilidad, suponiendo que para la moneda de dos pesos la probabilidad de que salga águila es 0.40 y para la moneda de cinco pesos esta probabilidad es 0.70.

2. El cajón donde Luis Arturo guarda sus calcetines a lado de su cama, contiene 8 calcetines negros, 6 blancos, 6 azules y 2 verdes. Luis Arturo es bastante flojo con respecto de cómo guarda sus calcetines: No los guarda por pares, sino echa los calcetinos sueltos en el cajón. Una mañana, cuando todavía hay oscuridad total, toma dos calcetines del cajón, pero para no despertar a su hermana quien duerme en la misma habitación, no enciende la luz. Puesto que los calcetines están hechos de la misma tela, no le es posible distinguirlos tactilmente y es plausible suponer que los dos calcetines que ha tomado, forman una muestra aleatoria de los 22 calcetines en el cajón. Bajo este supuesto, calcula las siguientes probabilidades:

   (a) La probabilidad de que los dos calcetines sean verdes.
   (b) La probabilidad de que los dos calcetines sean negros.
   (c) La probabilidad de que los dos calcetines sean del mismo color.

3. Salgo con cinco personas a un bar, donde nos sentamos a una mesa redonda. Estoy desesperadamente enamorado de una de estas personas. Suponiendo que cada uno elige una silla de forma aleatoria, ¿cuál es la probabilidad de que la persona de quien estoy enamorado se siente a mi lado?

4. Si A y B son eventos independientes dentro de un espacio muestral Ω y $P(A) = 0.80$ y $P(B) = 0.50$, ¿cuál es la probabilidad del evento compuesto $A \cup B$?

*5. Para resolver una discusión entre una astróloga que pretende poder predecir las características de personalidad de una persona a partir de su fecha y lugar de nacimiento, y un investigador científico que es escéptico respecto de la astrología, se realiza la siguiente prueba: La astróloga recibe (a) la fecha y lugar de nacimiento de seis personas y (b) una descripción de la personalidad de las mismas (obtenida por una combinación de pruebas psicológicas y entrevistas con amigos y familiares) y le piden que asocie la información sobre la fecha y el lugar de nacimiento de la persona con la información sobre la personalidad correspondiente. Se observa que para tres personas las dos partes de información fueron asociadas correctamente. ¿Cuál es la probabilidad de obtener este resultado o un resultado mejor si la astróloga en realidad forma pares de forma aleatoria?

6. El profesor de la asignatura Métodos Cuantitativos organiza un examen que consiste en 15 preguntas de opción múltiple. Cada pregunta tiene 3 opciones de respuesta (etiquetadas A, B, y C), de las cuales una (y sólo una) es correcta. Los estudiantes dan su respuesta en una hoja de lector óptico, marcando para cada pregunta la letra de la opción de respuesta que, cree, es la correcta.
   Para controlar el fraude (que consistiría en que un estudiante copiase una o más respuestas de otro estudiante), el profesor da a cada uno una copia personalizada. Cada copia contiene las mismas preguntas con las mismas opciones de respuesta, sin embargo se presentan en un orden aleatorizado. Es decir, las diferentes copias presentan en un órden diferente (a) las 15 preguntas y (b) las tres opciones de respuesta asociadas con cada pregunta. El profesor no dice a los estudiantes que reciben copias diferentes; ellos creen que el órden de las preguntas y las opciones de respuesta es el mismo para todos.
   (a) ¿Cuántas variantes (diferentes copias) del examen se pueden construir en total? Dicho de otra forma, ¿de cuántas maneras se pueden ordenar simultáneamente las preguntas y las opciones de respuesta?
   *(b) Al corregir los exámenes, el profesor sospecha que uno de los estudiantes (digamos Alumno X, para no mencionar nombres) copió gran parte de las respuestas de su vecino (Alumno Y, el cual es considerado por los demás como el más listo del grupo). La razón para esta sospecha es que Alumno X marcó en 13 preguntas la misma letra de opción de respuesta que Alumno Y; en particular, el patrón de respuestas (las letras que marcó cada estudiante en su copia, según el orden en el cual se le presentaron las preguntas) es el siguiente para estos estudiantes:

   Alumno X: CBBBCAAABAABABA   Alumno Y: CABBCAAABAABACA

   ¿La similitud en los patrones de respuesta es suficiente para suponer que se cometió algún fraude?

*7. ¿Qué evento tiene la probabilidad más alta: un seis al lanzar seis dados imparciales o dos seises al lanzar doce dados imparciales?

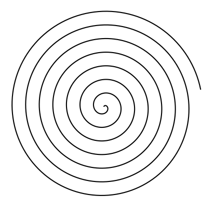
8. Tres síndromes psiquiátricos diferentes A, B y C suelen presentarse con la misma sintomatología E. Según la versión más reciente del *Manual Diagnóstico y Estadístico de los Trastornos Mentales*, la sintomatología E se presenta únicamente en estos tres síndromes; además, especifica que los síndromes no pueden presentarse simultáneamente en un mismo paciente. En un estudio epidemiológico se encontró que:
   (a) la probabilidad de contraer los síndromes A, B y C son de .01, .03 y .005, respectivamente;
   (b) los pacientes que efectivamente tienen el síndrome A, B o C, tienen una probabilidad de desarrollar la sintomatología E de .95, .90 y .70, respectivamente.

   Calcula, a partir de esta información, para cada síndrome la probabilidad de que un paciente que se presenta con la sintomatología E en realidad lo tenga.

9. Con los datos del ejemplo en la página 334, donde se aplicó la prueba ELISA a un paciente que sospechaba haber contraido el VIH, calcula la probabilidad de que el paciente no sea portador del virus cuando la prueba da un resultado negativo.

10. Un estudiante responde una pregunta de un examen de opción múltiple que tiene cuatro opciones de respuesta. Como parte de un modelo psicométrico se supone que la probabilidad de que el estudiante conozca la respuesta es .70 y que en este caso eligirá la opción correcta (es decir, no se considera posible que el estudiante se equivoque al conocer la respuesta). Si el estudiante no conoce la respuesta, el modelo supone que las cuatro opciones de respuesta tienen la misma probabilidad de ser elegidas. Bajo los supuestos de este modelo, ¿cuál es la probabilidad de que el estudiante conozca la respuesta si se ha observado que eligió la opción de respuesta correcta?

11. Para diferenciar si un jóven víctima de un accidente de tránsito ha sufrido o no daño cerebral, un neuropsicólogo usa la prueba de la "espiral de Arquímedes". La prueba consiste en presentar la siguiente espiral



mientras que gira durante 30 segundos en sentido de las manecillas del reloj. Diez segundos después de que el espiral ha dejado de moverse, se le pregunta al paciente cómo la espiral parece moverse. Los pacientes sin daño cerebral generalmente indican que observan la ilusión de que la espiral gira en sentido contrario, mientras que los pacientes con daño cerebral en muchas ocasiones no reportan esta ilusión. Más precisamente, una publicación* estima la probabilidad de que pacientes *sin daño cerebral* observen la ilusión en .97, mientras que en pacientes *con daño cerebral* esta probabilidad es de .46. Con base en su experiencia, el neuropsicólogo cree que el 10 % de las personas que se presentan a su centro después de un accidente de tránsito efectivamente sufra daño cerebral. A partir de esta información,

(a) ¿cuál es la probabilidad de que el paciente sufra daño cerebral si *no* reporta la ilusión?

(b) ¿cuál es la probabilidad de que *no* tenga daño cerebral si reporta la ilusión?

12. Para dos eventos A y B, se sabe que:

$$P(A|B) \;=\; P(A|B^c)$$

(a) Explica, en palabras comunes, qué significa esta igualdad.

(b) Intuitivamente, ¿qué conclusión sacarías de esta igualdad respecto de la dependencia o independencia de A y B?

*(c) Demuestra mediante una prueba matemática esta intuición.

*Gallese, A. J., Jr. (1956). Spiral aftereffect as a test of organic brain damage. *Journal of Clinical Psychology*, *12*, 254–258.

13. Sobre el siguiente problema, el cual se conoce como el *problema de Monty Hall*, existen varias publicaciones en revistas científicas. Para mucha gente, la conclusión a la que se llega a través de un análisis matemático-estadístico es muy contraintuitiva. A continuación, se presenta el problema como aparece en la página correspondiente de Wikipedia:

    Se ofrece un concurso cuya mecánica es la siguiente:

    - Al concursante se le ofrece la posibilidad de escoger entre tres puertas. Tras una de ellas se encuentra un coche, y tras las otras dos hay una cabra. El concursante gana el premio que se oculta detrás de la puerta que escoja.
    - Después de que el concursante escoja una puerta, el presentador abre una de las otras dos puertas, mostrando una cabra. Siempre puede hacerlo ya que incluso si el concursante ha escogido una cabra, queda otra entre las puertas que ha descartado y el presentador conoce lo que hay detrás de cada puerta.
    - Entonces, ofrece al concursante la posibilidad de cambiar su elección inicial y escoger la otra puerta que descartó originalmente, que continúa cerrada.

    La pregunta es: ¿debe cambiar su elección inicial o no?

Tema 10
Funciones de probabilidad y densidad

Índice

En los temas de la primera parte se introdujeron herramientas para organizar y resumir la información en una o más variables para las cuales se han observado valores en una muestra. Por ejemplo, se definieron las funciones de frecuencia y proporción, los cuantiles, estadísticos de tendencia central (como la mediana y la media), estadísticos de variabilidad (como la amplitud intercuartil y la varianza), y estadísticos de covariación (como la covarianza y la correlación). Estos conceptos corresponden con características de una muestra; en el tema actual se consideran características de una población, o más preciso, características de variables aleatorias para el caso de que el experimento aleatorio que subyace los datos se haya realizado un número infinito de veces en la población bajo consideración. Existe una relación estrecha entre los conceptos muestrales y poblacionales—efectivamente, muchos de los conceptos poblacionales tienen su homólogo muestral—, lo cual explica por qué la estructura del tema actual en grandes líneas sigue la estructura de la primera parte (Temas 2 a 7). De cierto modo, el tema actual retomará muchos de los conceptos introducidos en el contexto de la estadística descriptiva y los reconsidera en el contexto de la estadística inferencial.

En la primera sección se definen los conceptos poblacionales que son los homólogos de los conceptos muestrales introducidos en el Tema 2. En particular, se definen la función de probabilidad (Sección 10.1.1) y la función de densidad (Sección 10.1.2) de una variable aleatoria, las cuales se pueden considerar como las homólogas poblacionales de la función de proporción para una muestra. Además, se definen la función de distribución (Sección 10.1.3), homóloga de la función de proporción *acumulada*, y cuantiles poblacionales (Sección 10.1.4) que son homólogos a los cuantiles muestrales.

La función de la Sección 10.2 en la estructura del tema actual es similar a la del Tema 3 en la primera parte de este libro: Introduce algunos *parámetros* que resumen funciones de probabilidad o densidad. Al respecto, es importante entender que un parámetro es el homólogo de un estadístico, ya que éste resume la función de frecuencia o proporción de una muestra. Efectivamente, en la Sección 10.2.1 se introducen parámetros que resumen la función de probabilidad o densidad respecto de su tendencia central y en la Sección 10.2.2 se introducen parámetros de variabilidad. No sorprenderá que los conceptos que se definen en estas secciones traen el nombre de “moda poblacional”, “mediana poblacional”, “media poblacional”, “varianza poblacional”, etc., es decir, son homólogos de los estadísticos que se introdujeron en el Tema 3. Cabe mencionar que, para algunos de estos parámetros se presentan también teoremas sobre el efecto de transformaciones de las variables en estos parámetros y, en este sentido, también existen relaciones entre la Sección 10.2 de este tema y el Tema 4.

La tercera sección es el homólogo del Tema 5: Se generalizan los conceptos de función de probabilidad y densidad a dos o más variables. A continuación, la Sección 10.4 habla sobre características de funciones de probabilidad y densidad para dos variables, en particular, sobre independencia de dos variables aleatorias (Sección 10.4.1) y covariación lineal (Sección 10.4.2). En este sentido, la Sección 10.4 es para la estadística inferencial lo que es el Tema 6 para la estadística descriptiva.

A pesar de la relación estrecha entre ambos, es muy importante diferenciar entre los conceptos poblacionales y muestrales. Los conceptos poblacionales generalmente figuran en un *modelo estadístico*, el cual se puede considerar como una formalización (es decir, una “traducción en términos matemáticos”) de una teoría sobre el origen de los datos y, por lo tanto, provee el fundamento de las características poblacionales. En otras palabras, una función de probabilidad o una función de densidad (y sus características como la media poblacional, independencia poblacional, etc.) son idealizaciones o teorías sobre la realidad. Los conceptos y características muestrales, por otro lado, se pueden conocer a través de los datos observados en la muestra. A fin de cuentas son las características poblacionales sobre las cuales se quieren sacar conclusiones, y en muchas ocasiones se usarán sus homólogas muestrales para estimarlas (lo cual se tratará en más detalle en el siguiente tema).

En la última sección de este tema, se introducen algunas (familias de) funciones de probabilidad y densidad concretas: la distribución de Bernoulli (Sección 10.5.1), la distribución binomial (Sección 10.5.2), la distribución de Poisson (Sección 10.5.3) y la distribución normal (Sección 10.5.4). Resulta que estas distribuciones son parte de los modelos estadísticos que subyacen muchos de los análisis en la estadística aplicada (no solo en las ciencias de la salud, sino en cualquier rama de las ciencias).

10.1 Funciones de probabilidad y densidad para una variable aleatoria

10.1.1 Función de probabilidad para variables aleatorias discretas

Primero se considera el caso de una variable aleatoria X discreta, es decir, una variable aleatoria cuyo espacio de los valores es finito o infinito numerable. Los siguientes ejemplos introducen el concepto de función de probabilidad de una variable aleatoria discreta.

Ejemplo

Retomemos el experimento aleatorio en la página 294, que consiste en lanzar tres veces una moneda y para el cual se definió como resultado la cara que queda arriba en los tres lanzamientos. Además se definió la variable aleatoria X "número de veces que salió aguila" Como mencionamos al desarrollar este ejemplo, el espacio muestral y la variable X son discretos.

El espacio de los valores de X es $\{0, 1, 2, 3\}$; para cada valor x_i de X se puede considerar la probabilidad de que se realice. Se anota esta probabilidad $\pi(x_i)$, y formalmente corresponde con $P(\{\omega \mid X(\omega) = x_i\})$, lo cual, en palabras, es la probabilidad de que se realice el evento que incluye todos los resultados cuya imagen es igual a x_i. Usamos la minúscula griega π para indicar que se trata de una probabilidad teórica; se refiere a la población de todas las secuencias de tres monedas en el experimento aleatorio.

Como siguiente paso, se especifica un modelo estadístico; quiere decir que se postula un valor para los $\pi(x_i)$. En particular, se formula una teoría (es decir, algunos supuestos) acerca del proceso que genera los resultados del experimento aleatorio y se traduce esta teoría a términos matemáticos. En este ejemplo parece plausible suponer que

(a) la moneda es imparcial, que, en términos matemáticos, quiere decir que al lanzarla:

$$P(\circ) = P(\bullet) = \frac{1}{2}.$$

Recuérdese que $\circ$ y $\bullet$ representan "sol" y "aguila", respectivamente.

(b) los tres lanzamientos son independientes, lo cual, en términos matemáticos, quiere decir que la probabilidad de algún triplete es el producto de las probabilidades (véase p. 337) de las caras que se muestran, por ejemplo:

$$P[(\bullet, \circ, \bullet)] \; = \; P(\bullet) \; \times \; P(\circ) \; \times \; P(\bullet),$$

lo cual, bajo el supuesto de imparcialidad, es igual a

$$\begin{aligned} &= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \\ &= \frac{1}{8}. \end{aligned}$$

A partir de estos supuestos, se deriva directamente que:

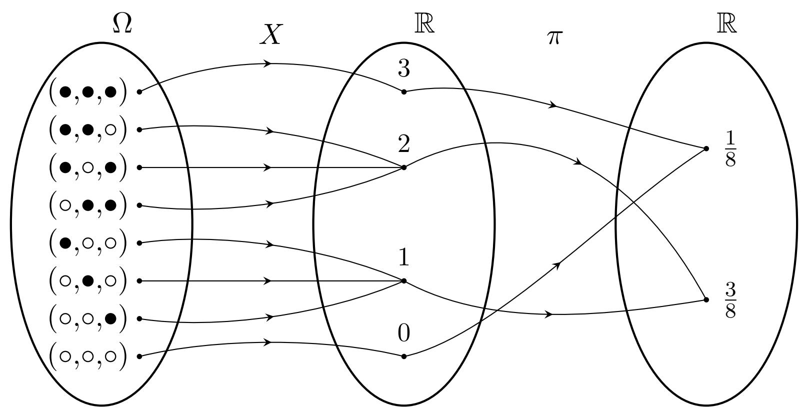
$$\pi(0) \; = P\left[\, \{\omega \mid X(\omega) = 0\} \,\right] \; = \; P[\{(\circ, \circ, \circ)\}] = \frac{1}{8},$$

$$\pi(1) \;= P[\,\{\omega\,|\,X(\omega)=1\}\,] \;=\; P[\{(\bullet,\circ,\circ),(\circ,\bullet,\circ),(\circ,\circ,\bullet)\}] = \frac{3}{8},$$

$$\pi(2) \;= P[\,\{\omega\,|\,X(\omega)=2\}\,] \;=\; P[\{(\bullet,\bullet,\circ),(\bullet,\circ,\bullet),(\circ,\bullet,\bullet)\}] = \frac{3}{8},$$

$$\pi(3) \;= P[\,\{\omega\,|\,X(\omega)=3\}\,] \;=\; P[\{(\bullet,\bullet,\bullet)\}] = \frac{1}{8}.$$

Este ejemplo ilustra que π es una función del espacio de los valores de la variable X al conjunto de números reales en el intervalo [0,1]. Efectivamente, π asigna a cada valor de X una probabilidad entre 0 y 1. Gráficamente se puede representar el proceso completo (que consiste en definir el espacio muestral Ω, la variable aleatoria X y la función π) como sigue:



Considérese, a continuación, el caso del experimento aleatorio anterior, pero donde se cambia el supuesto de la imparcialidad de la moneda. En particular, se cambia el supuesto (a) por: Para cada lanzamiento, $P(\bullet) = .70$ y $P(\circ) = .30$. Si se mantiene el supuesto de independencia, entonces se puede calcular, por ejemplo, que

$$\begin{aligned} P[\{(\circ,\bullet,\bullet)\}] \;&=\; .30 \times .70 \times .70 \\ &=\; .147. \end{aligned}$$

Siguiendo el mismo razonamiento, las probabilidades $\pi(x_i)$ se modifican a:

$$\begin{aligned} \pi(0) \;&= P[\{(\circ,\circ,\circ)\}] \;=\; 0.027, \\ \pi(1) \;&= P[\{(\bullet,\circ,\circ),(\circ,\bullet,\circ),(\circ,\circ,\bullet)\}] \;=\; 0.189, \\ \pi(2) \;&= P[\{(\bullet,\bullet,\circ),(\bullet,\circ,\bullet),(\circ,\bullet,\bullet)\}] \;=\; 0.441, \\ \pi(3) \;&= P[\{(\bullet,\bullet,\bullet)\}] \;=\; 0.343. \end{aligned}$$

Finalmente, volvamos a considerar el mismo experimento aleatorio, suponiendo que la moneda no es imparcial y que los tres lanzamientos son independientes. Contrario al ejemplo anterior, *no* conocemos las probabilidades $P(\bullet)$ y $P(\circ)$. Procedemos escribiendo que $P(\bullet) = \theta$ y, por consiguiente, $P(\circ) = 1 - \theta$, donde θ representa el valor incógnito de la probabilidad de que salga águila en un lanzamiento de la moneda. Un incógnito de este tipo se llama un *parámetro*, es decir, una constante que representa algún aspecto númerico del modelo estadístico. Aunque desconocemos el valor exacto de θ, podemos derivar una expresión matemática para las probabilidades de cada evento. Por ejemplo, siguiendo la misma lógica del ejemplo anterior:

$$\begin{aligned} P[\{(\circ, \bullet, \bullet)\}] &= (1-\theta) \times \theta \times \theta \\ &= (1-\theta)\,\theta^2. \end{aligned}$$

Las probabilidades asociadas por la función de probabilidad π son las siguientes:

$$\begin{aligned} \pi(0) &= P[\{(\circ,\circ,\circ)\}] = (1-\theta)^3, \\ \pi(1) &= P[\{(\bullet,\circ,\circ),(\circ,\bullet,\circ),(\circ,\circ,\bullet)\}] = 3\theta(1-\theta)^2, \\ \pi(2) &= P[\{(\bullet,\bullet,\circ),(\bullet,\circ,\bullet),(\circ,\bullet,\bullet)\}] = 3\theta^2(1-\theta), \\ \pi(3) &= P[\{(\bullet,\bullet,\bullet)\}] = \theta^3. \end{aligned}$$

En el último ejemplo del bloque anterior se presentó un modelo estadístico incompleto, debido a que una característica no se ha especificado. Es muy común en la estadística inferencial especificar modelos estadísticos incompletos y utilizar datos muestrales para llegar a estimar los valores desconicidos de los parámetros.

El espacio muestral y el espacio de los valores de la variable X en los ejemplos anteriores eran finitos, y por lo tanto, discretos. Sin embargo, también es posible que una variable discreta tenga un número infinito de valores (véase la p. 294). El siguiente ejemplo introduce el caso de la función de probabilidad de una variable discreta con un espacio de valores infinito (numerable).

Ejemplo

Considérese el experimento aleatorio que consiste en lanzar una moneda hasta que salga águila y donde se consideran como resultados la secuencia de sol y águila que se produjo. Entonces, $\Omega = \{(\bullet), (\circ,\bullet), (\circ,\circ,\bullet), (\circ,\circ,\circ,\bullet), (\circ,\circ,\circ,\circ,\bullet), \ldots\}$. Defínase la variable aleatoria X que asocia cada resultado con el número de lanzamientos realizados.

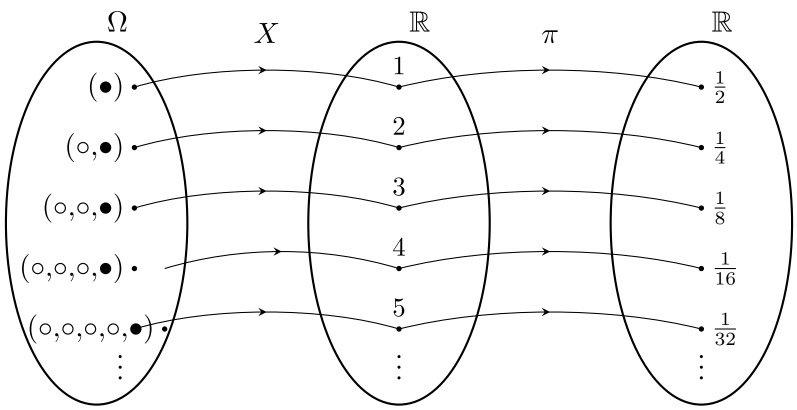
Si, como en el ejemplo anterior, se supone que la moneda es imparcial y los lanzamientos son independientes, se puede calcular la probabilidad de cada resultado y de los valores de la variable X. En particular, se encuentra:

$$\begin{aligned} \pi(1) &= P[(\bullet)] = \frac{1}{2}, \\ \pi(2) &= P[(\circ,\bullet)] = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}, \\ \pi(3) &= P[(\circ,\circ,\bullet)] = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}, \\ \pi(4) &= P[(\circ,\circ,\circ,\bullet)] = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}, \\ &\vdots \end{aligned}$$

O bien, para cualquier valor x en el espacio de los valores de X:

$$\pi(x) = \frac{1}{2^x}$$

Gráficamente, se representa el espacio muestral Ω, la variable aleatoria X y la función de probabilidad π de este ejemplo por:



10.1.2 Función de densidad para variables aleatorias continuas

10.1.3 Función de distribución

10.1.4 Cuantiles poblacionales

10.2 Características poblacionales de una variable aleatoria

10.2.1 Tendencia central

10.2.2 Variabilidad

10.3 Funciones de probabilidad y densidad para dos variables aleatorias

10.3.1 Función de probabilidad para variables aleatorias discretas

10.3.2 Función de densidad para variables aleatorias continuas

10.4 Características poblacionales de dos variables aleatorias

10.4.1 Independencia de variables aleatorias

10.4.2 Covariación lineal entre variables aleatorias

10.5 Algunas funciones de probabilidad y densidad importantes

10.5.1 La distribución de Bernoulli

10.5.2 La distribución binomial

10.5.3 La distribución de Poisson

10.5.4 La distribución normal

Problemas

1. Considérese el siguiente experimento aleatorio, que consiste en:
   - Extraer de forma aleatoria un individuo de la población de estudiantes de Medicina que se presentan al examen profesional para titularse;
   - Aplicar a este individuo un examen de 330 preguntas de opción multiple (con tres opciones de respuesta);
   - Registrar para cada pregunta si esta persona ha dado la respuesta correcta o no.

   Se definen variables aleatorias $X_1, X_2, \ldots, X_{330}$ que assumen valores de 0 ó 1; $X_j = 1$ significa que la persona extraída contesta correctamente el ítem j y $X_j = 0$ que no da la respuesta correcta en este ítem. Para el análisis de respuestas en este examen, un psicómetra considera los siguientes dos modelos:

   Modelo A:
   Cada una de las variables X_j tiene la siguiente distribución:

   $$X_j \;\sim\; \text{Bernoulli}(\theta),$$

   y las variables $X_1, X_2, \ldots, X_{330}$ son conjuntamente independientes.

   Modelo B:
   Cada una de las variables X_j tiene la siguiente distribución:

   $$X_j \;\sim\; \text{Bernoulli}(\theta_j),$$

   y la correlación entre cualquier par de variables X_j y $X_{j'}$ es igual a ρ.

   (a) Explica la teoría subyacente al Modelo A y el Modelo B y, especialmente, la diferencia entre ambos.

   (b) ¿Cuál de los dos modelos es el más parsimonio (es decir, tiene el menor número de parámetros)? ¿Cuántos parámetros desconocidos tiene cada modelo?

   (c) Si el Modelo A fuese una representación apropiada de la realidad, entonces ¿qué conclusión se podría sacar sobre lo apropiado del Modelo B?

   (d) Argumenta, a la luz del experimento aleatorio, por qué probablemente el Modelo B es el más adecuado de los dos modelos presentados.

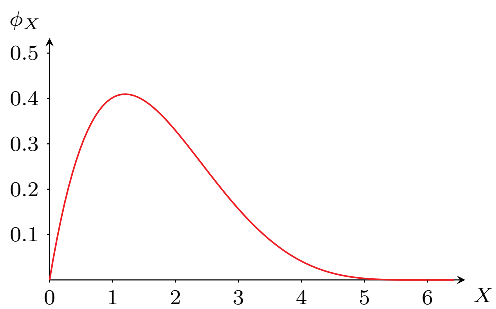
   (e) Si se define la variable Y

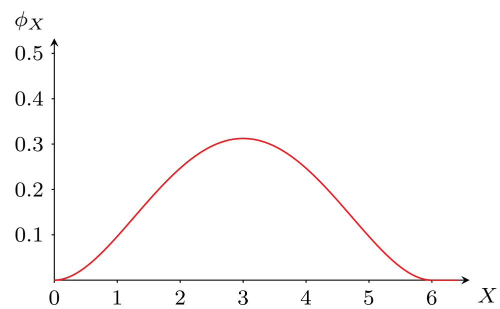
   $$Y \;=\; \sum_{j=1}^{330} X_j,$$

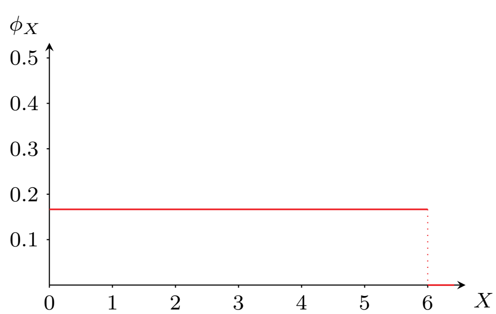
   entonces ¿cuál sería el valor esperado $\mathscr{E}(Y)$ bajo los supuestos de cada modelo?

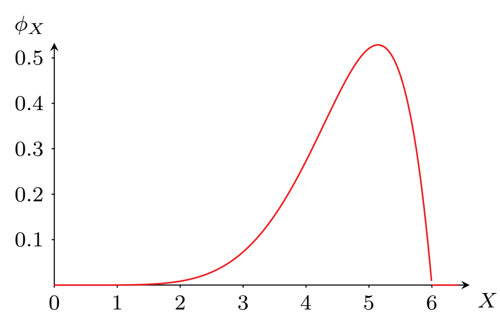
2. En un examen clínico objetivo estructurado (ECOE), los sustentantes pasan por una serie de *estaciones* y en cada estación tienen que realizar una tarea específica. Por fines organizativas es necesario restringir el tiempo total por estación, por ejemplo a seis minutos. Esto quiere decir que cada estudiante tendrá un máximo de seis minutos para completar la tarea que se pide en cada estación.

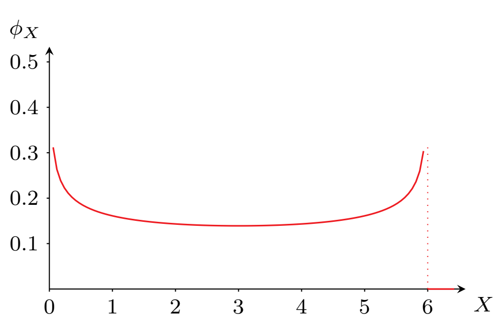
   El responsable del ECOE de la Facultad de Medicina de la UNAM se pregunta si el tiempo máximo de seis minutos por estación es suficiente y, para investigarlo, pide al evaluador de cada estación que anote para todos los estudiantes el tiempo en que terminan la tarea. En el análisis de datos se ajusta un modelo estadístico a los tiempos de cada estación. A través de los parámetros de este modelo, se pueden obtener diferentes distribuciones para la variable aleatoria X, "tiempo en minutos que los estudiantes requieren para completar la estación". Las figuras que se muestran a continuación presentan seis de estas distribuciones:

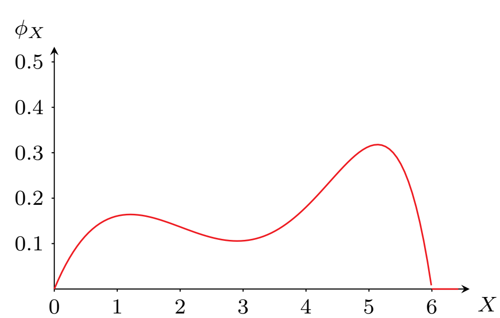












   ¿Qué significan estos modelos a la luz de la pregunta de investigación y cuál consideras como el modelo más adecuado?

Tema 11
Estimación de parámetros

Tema 12
Prueba de hipótesis