Pensando antes de actuar: matemáticas para decidir.

 [1]

Beatriz Rumbos Pellicer
Instituto Tecnológico Autónomo de México

22 de junio de 2009

[1] *La Partie d'échecs* 1508, *Lucas van Leyden* (1494 – 1533)

Contenido

Prefacio

El estudio formal de la toma de decisiones es una parte fundamental de cualquier disciplina que involucre interacciones entre individuos. Nuestra vida cotidiana está llena de decisiones; algunas triviales, como la elección entre comer carne o pescado, y otras trascendentes, como la determinación de la inocencia o culpabilidad de un individuo en un juicio. Formalmente, una decisión consiste en la elección de una acción dentro de un conjunto de acciones posibles.

Este libro ha sido concebido como una introducción a los métodos cuantitativos que surgen en los procesos de decisión cotidianos. Para modelar estos procesos, es necesario cuantificar nuestras creencias, deseos y preferencias; asimismo, debe tomarse en cuenta que vivimos en un entorno en el cual reinan el riesgo y la incertidumbre. Como motivación para enriquecer nuestra cultura cuantitativa, se recomienda la lectura del magnífico libro de John Allen Paulos: *Innumeracy: Mathematical Illiteracy and its Consequences*[1].

El marco formal de la teoría de la decisión se origina con la probabilidad y la estadística matemática y, a partir del siglo XIX, ha sido explotada por los economistas en lo que se conoce como teoría de la utilidad. Durante el siglo XX, esta formalización se extendió a otras disciplinas y hoy en día, lo que se conoce como teoría de elección racional, es utilizada en campos tan diversos como la Biología, la Psicología y el Derecho.

Independientemente del tipo de decisión, asumimos que los individuos siguen un proceso de inferencia racional que relaciona premisas con conclu-

[1]Hill and Wang, New York, NY, 2001.

siones. El primer capítulo del libro proporciona los elementos fundamentales de la lógica deductiva. Ésta permite verificar la consistencia de los procesos de razonamiento y nos ofrece una estructura formal para la construcción de cualquier tipo de conocimiento.

A partir del segundo capítulo se consideran premisas inciertas, que son la base de la lógica inductiva. A fin de precisar el concepto de incertidumbre se proporcionan, en los siguientes capítulos, los principios de la teoría de la probabilidad. En el camino, se explica el concepto de valor esperado, el cómo los individuos evalúan prospectos inciertos, sus actitudes ante el riesgo y la teoría de juegos como un proceso de decisión colectiva. Finalmente, se proporcionan los elementos básicos de estadística descriptiva e inferencia: clasificación e interpretación de conjuntos de datos, así como pruebas de hipótesis y regresión lineal.

En el desarrollo de los temas descritos se hace hincapié en los métodos bayesianos de aprendizaje e inferencia. Personalmente, creo que la subjetividad implícita de estos métodos va en beneficio de la toma de decisiones bajo ambientes de incertidumbre -casi en todos los casos-. La inferencia bayesiana nos proporciona un marco formal para distinguir entre lo incierto y lo más incierto, entre la opinión de un experto y la de un charlatán, entre un rumor y una evidencia fundamentada. Su estructura nos obliga a examinar constantemente nuestras creencias y, de ser necesario, a modificarlas. Idealmente, ésta debe ser la base de toda adquisición del conocimiento. Aunque los seres humanos nos aferrarnos fácilmente a lo que consideramos *verdades* y el proceso de actualización bayesiana puede conflictuarnos, "es de sabios ser bayesianos".

¿Qué prerequisitos asume este texto? Conocimientos de álgebra de preparatoria: solución de ecuaciones lineales, manejo y simplificación de expresiones, representación geométrica de funciones lineales, interpretación de gráficas simples y lenguaje elemental de conjuntos. ¿A quienes está dirigido? A todos aquellos que tengan un interés por la teoría de elección racional y no requieran o deseen una gran profundidad matemática. Se incluyen ejemplos diversos y en particular un buen número de ellos relacionados con el Derecho. La razón es que ésta es una disciplina usualmente olvidada por los métodos cuantitativos y aquí se pretende compensar esta omisión.

A pesar de que la teoría de elección racional proporciona una estructura formal para describir el comportamiento, debe siempre tenerse en cuenta que se trata de modelos sumamente estilizados, los cuales simplemente nos ayudan a entender ciertos rasgos generales de dicho comportamiento. Los seres humanos somos extremadamente complejos y sería imposible modelar y predecir las acciones de cualquier individuo. No obstante, aún con esta limitación, la teoría de elección racional nos permite hacer predicciones y entender un gran número de fenómenos. Nunca seremos autómatas decidiendo de acuerdo a fórmulas preestablecidas; sin embargo, seremos más sabios si reflexionamos y pensamos antes de actuar. Espero que este libro sea una ayuda en este proceso.

Quiero agradecer a mis alumnos de la carrera de Derecho, que tomaron el curso de *Comprensión Matemática* en la primavera 2008, y al ministro José Ramón Cossío Díaz, pues gracias a ellos existe este texto. Asimismo, agradezco a mi colega y amigo Manuel Mendoza por su ayuda en la revisión del manuscrito, y a Carmen López y Leticia Barreiro por sus atinadas correcciones y sugerencias.

México, D.F.

Mayo 2009

Capítulo 1

Introducción a la lógica

1.1. Introducción

Como recordaremos, la palabra *lógica* proviene del griego *logos*, que significa palabra, idea o argumento. Como disciplina, sin entrar en detalle, podemos decir que la lógica estudia proposiciones, argumentos y sistemas de inferencia. Los primeros estudios serios de lógica que conocemos son los de Aristóteles (384-322 ac.). A partir de entonces, los filósofos griegos desarrollaron gran parte de lo que hoy conocemos como **lógica proposicional**. El siguiente avance significativo surgió más de mil años después, a mediados del siglo XIX. Fue entonces cuando los británicos, Augustus De Morgan (1806-1871) y George Boole (1815 - 1864), desarrollaron el lenguaje simbólico para la lógica proposicional. De esta forma, la manipulación de las proposiciones se transformó en algo más "algebraico" y la lógica se transformó en una disciplina matemática.

Una **proposición** es un enunciado que es verdadero o falso, sin ambigüedad alguna. Las proposiciones las denotaremos por letras como P, Q, R, seguidas del enunciado que puede estar descrito en lenguaje natural o en lenguaje matemático. Dos o más proposiciones pueden combinarse utilizando **conectivos lógicos**, éstos son palabras como *"y"*, *"o"*, o *"si...entonces"*. Asimismo, la palabra *"no"* sirve para negar una proposición. Los símbolos $\wedge$ y $\vee$ toman el lugar de los conectivos y y o, respectivamente. El símbolo

$\sim$ antes de una proposición denota su negación.

Así, dadas proposiciones P y Q, se tiene que $P \wedge Q$ es P y Q (o **conjunción** de P y Q), $P \vee Q$ es P o Q (o **disyunción** de P y Q) y $\sim P$ significa "no P" (**negación** de P). Es común que las proposiciones contengan palabras llamadas **cuantificadores**; éstos se pueden dividir en dos, los **cuantificadores existenciales** como: *alguno, existe, (para) al menos uno, etcétera.* y los **cuantificadores universales o categóricos** como, *(para) todo , ninguno, cada uno, algunos, etcétera.*

Ejemplos

Ej 1.1.1 Las siguientes son proposiciones:

- El perro es un animal doméstico. (Verdadera.)
- $5 \times 2 = 10$. (Verdadera.)
- $5 \times 2 = 12$. (Falsa)
- Todos los países son democráticos. (Falsa)

Ej 1.1.2 Las siguientes no son proposiciones:

- ¿Cómo te llamas?
- $\frac{5}{2}$.
- Los patos son animales más simpáticos que los conejos

Ej 1.1.3 Las siguientes son proposiciones compuestas

- Soy estudiante de licenciatura y tengo 23 años.
- Homero está enojado y confundido. (Notemos que es la forma abreviada de decir "Homero está enojado y Homero está confundido").
- Si $x + 3 = 5$, entonces $x = 2$.

Ej 1.1.4 Dada P : Hoy es viernes se tiene que su negación es $\sim P$: Hoy no es viernes.

Ej 1.1.5 Dada $P : x > 5$, su negación es $\sim P : x \leq 5$. (Esto último es equivalente a $x \not> 5$.)

Ej 1.1.6 Si P : Tenemos clase de economía y Q : Tenemos clase de matemáticas, entonces la conjunción y la disyunción están dadas por

$P \wedge Q$: Tenemos clase de economía y tenemos clase de matemáticas,

$P \vee Q$: Tenemos clase de economía o tenemos clase de matemáticas.

Ej 1.1.7 El conectivo *pero* es utilizado en lugar del conectivo *y* para dar un énfasis diferente a la proposición compuesta; sin embargo, en términos lógicos, ambos conectivos son equivalentes, es decir, tienen el mismo valor de verdad en cualquier situación. Así, P : Verónica quiere ir al cine pero Archie quiere ir al juego de fútbol, representa lo mismo que Q : Verónica quiere ir al cine y Archie quiere ir al juego de fútbol.

♠♠♠♠♠

1.2. Tablas de verdad y proposiciones equivalentes

Dadas las proposiciones P y Q, con sus respectivos valores de verdad, pueden determinarse los valores de verdad de las proposiciones compuestas que pueden obtenerse a partir de P y Q. Por ejemplo, si una proposición es verdadera (falsa), su negación es falsa (verdadera). En general, las **tablas de verdad** asignan valores de verdad a las proposiciones compuestas. A continuación presentamos las tablas de verdad para las proposiciones $\sim P, P \vee Q$ y $P \wedge Q$, en donde V y F denotan verdadero y falso, respectivamente.

P	$\sim P$
V	F
F	V

P	Q	$P \vee Q$	$P \wedge Q$
V	V	V	V
V	F	V	F
F	V	V	F
F	F	F	F

(I)

Dada una proposición compuesta, se dice que es una **tautología** si ésta es verdadera independientemente de los valores de verdad que tengan las proposiciones que la componen. Asimismo, se dice que es una **contradicción** si es falsa independientemente de los valores de verdad que tomen sus componentes.

Ejemplos

Ej 1.2.1 Dada la proposición falsa

$$P: \text{Todas las ranas son verdes}$$

su negación no es "Todas las ranas no son verdes" o "Ninguna rana es verde", ya que estas proposiciones también son falsas. Sabemos que la negación de una proposición falsa debe ser verdadera. La forma correcta de negar P es

$\sim P$: No todas las ranas son verdes (verdadera)

o bien

$\sim P$: Existe una rana que no es verde (verdadera)

o

$\sim P$: Al menos una rana no es verde (verdadera)

o

$\sim P$: Algunas ranas no son verdes (verdadera)

Asimismo, dada la proposición,

$$Q: \text{Algunas ranas son verdes (verdadera)},$$

tenemos que su negación es,

$$\sim Q: \text{Ninguna rana es verde (falsa)}$$

Ej 1.2.2 Sean P : Hamlet nació en Dinamarca y Q : La indecisión mató a Hamlet. Sabemos que ambas proposiciones son verdaderas; así, de acuerdo a las tablas de verdad I, se tiene que

$P \wedge Q$: Hamlet nació en Dinamarca y la indecisión mató a Hamlet (V),

$\sim Q$: La indecisión no mató a Hamlet (F),

$P \wedge \sim Q$: Hamlet nació en Dinamarca y la indecisión no mató a Hamlet (F)

♠♠♠♠♠

Dada cualquier proposición P, $P\vee \sim P$ es una tautología y $P\wedge \sim P$ es una contradicción. Esto es claro de acuerdo a la tabla de verdad,

P	$\sim P$	$P\vee \sim P$	$P\wedge \sim P$
V	F	V	F
F	V	V	F

Dadas las proposiciones P y Q, la tabla de verdad para $\sim P \vee Q$ se obtiene como,

P	Q	$\sim P$	$\sim P \vee Q$
V	V	F	V
V	F	F	F
F	V	V	V
F	F	V	V

Asimismo, la tabla de verdad para $(\sim P \vee Q) \wedge P$ es,

P	Q	$\sim P$	$\sim P \vee Q$	$(\sim P \vee Q) \wedge P$
V	V	F	V	V
V	F	F	F	F
F	V	V	V	F
F	F	V	V	F

Las operaciones $\vee$ y $\wedge$ son **conmutativas**, es decir, se satisfacen las siguientes equivalencias:

$$P \vee Q = Q \vee P,$$
$$P \wedge Q = Q \wedge P.$$

La igualdad entre dos proposiciones significa que éstas son **equivalentes.** Para probar la equivalencia entre proposiciones obtenemos sus tablas de verdad y vemos que los valores de verdad coinciden en todas las situaciones.

Por ejemplo, para probar $P \vee Q = Q \vee P$ se tiene

P	Q	$P \vee Q$	$Q \vee P$
V	V	**V**	**V**
V	F	**V**	**V**
F	V	**V**	**V**
F	F	**F**	**F**

A continuación se demuestran las llamadas **leyes de De Morgan**, que éstan dadas por las siguientes equivalencias:

$$\sim (P \vee Q) =\sim P \wedge \sim Q, \tag{1.1}$$
$$\sim (P \wedge Q) =\sim P \vee \sim Q.$$

En efecto, para la primera equivalencia de De Morgan tenemos,

P	Q	$\sim P$	$\sim Q$	$P \vee Q$	$\sim (P \vee Q)$	$\sim P \wedge \sim Q$
V	V	F	F	V	**F**	**F**
V	F	F	V	V	**F**	**F**
F	V	V	F	V	**F**	**F**
F	F	V	V	F	**V**	**V**

La otra equivalencia se deja como ejercicio al lector.

Podemos negar la proposición P : Se me antoja comer salmón o atún, utilizando la primera de las leyes de De Morgan como sigue:

$\sim P$: No se me antoja comer salmón ni atún.

Asimismo, la segunda ley de De Morgan se utiliza para negar la proposición Q : Es inteligente y está alerta, para obtener

$\sim Q$: No es inteligente o no está alerta.

Tres o más proposiciones pueden combinarse para obtener una tabla de verdad de la misma forma que antes. Por ejemplo, Dadas P, Q y R tenemos que la tabla de verdad para la proposición compuesta $P \wedge (Q \vee R)$ está dada como,

P	Q	R	$Q \vee R$	$P \wedge (Q \vee R)$
V	V	V	V	V
V	V	F	V	V
V	F	V	V	V
V	F	F	F	F
F	V	V	V	F
F	V	F	V	F
F	F	V	V	F
F	F	F	F	F

Notemos que se requiere de ocho renglones para incluir todas las posibles combinaciones de valores de verdad para las tres proposiciones P, Q y R.

1.3. Proposiciones condicionales

Dadas las proposiciones P y Q, la proposición compuesta obtenida al unirlas mediante el uso de *si P entonces Q*, se conoce como **proposición condicional**. Esta proposición se denota por $P \Rightarrow Q$ en donde P es el **antecedente** o **hipótesis** y Q el **consecuente** o **tesis**.

Ejemplos

Ej 1.3.1 "Si estudio mucho, entonces obtendré una buena calificación en el examen". Ésta es una proposición condicional con antecedente P : estudio mucho y consecuente Q : obtendré una buena calificación en el examen.

Ej 1.3.2 El conectivo *si...entonces* puede no ser explícito en una proposición condicional. Por ejemplo, "toda la gente trabajadora obtiene empleo" y "es fácil concentrarse cuando hay silencio" son en realidad proposiciones condicionales. En el primer caso la proposición puede reescribirse como: "si la gente es trabajadora, entonces obtiene un empleo" y en el segundo como, "si hay silencio, entonces es fácil concentrarse".

♠♠♠♠♠

La tabla de verdad para la proposición condicional puede construirse imaginando que alguien declara la proposición y queremos determinar si esta persona miente, en cuyo caso la condicional es falsa, o dice la verdad y la condicional es verdadera. Tomemos el caso de “Si estudio mucho, entonces obtendré una buena calificación en el examen”, dado en el ejemplo 1.3.1.

Si P es verdadera y estudié mucho y Q también es verdadera y obtuve una buena nota en el examen, entonces digo la verdad al declarar la condicional y $P \Rightarrow Q$ es verdadera. Si estudié mucho y no obtuve una buena calificación entonces mentí al declarar la condicional y $P \Rightarrow Q$ es falsa. Si no estudié mucho y aún así obtengo una buena calificación, no se me puede acusar de mentir pues, por ejemplo, es posible que por una afortunada coincidencia el examen consistiera únicamente de los temas que había estudiado. Finalmente, si no estudié mucho y no obtengo una buena calificación, tampoco se me puede acusar de mentir pues el antecedente, “estudié mucho” nunca se materializó. De esta forma, la tabla de verdad es la siguiente:

P	Q	$P \Rightarrow Q$
V	V	V
V	F	F
F	V	V
F	F	V

Ejemplo

Ej 1.3.3 Dado que la proposición $P \Rightarrow Q$ es verdadera cuando el antecedente P es falso, tenemos que las proposiciones

Si los conejos vuelan, entonces yo soy un humano,

Si los conejos vuelan, entonces yo soy un conejo,

son ambas verdaderas puesto que la proposición “Los conejos vuelan” es falsa.

Ej 1.3.4 Supongamos que siempre manejamos con extrema precaución. La condicional, "si manejamos con precaución (entonces) no tendremos accidentes", por desgracia es falsa ya que aunque sea verdadero que manejamos con precaución, siempre puede haber accidentes ocasionados por causas ajenas a nosotros, es decir, no tendremos accidentes es falsa y por lo tanto también lo es la condicional.

♠♠♠♠♠

Las proposiciones $P \Rightarrow Q$ y $\sim P \vee Q$ son equivalentes. En efecto, puesto que la tabla de verdad es la siguiente:

P	Q	$\sim P$	$P \Rightarrow Q$	$\sim P \vee Q$
V	V	F	**V**	**V**
V	F	F	**F**	**F**
F	V	V	**V**	**V**
F	F	V	**V**	**V**

Dada esta observación, podemos expresar

$$P \Rightarrow Q: \text{ Si no pido sopa entonces pido ensalada,}$$

como

$$\sim P \vee Q: \text{ Pido sopa o pido ensalada.}$$

De acuerdo a las leyes de De Morgan, la negación de esta última es,

$$P \wedge \sim Q: \text{No pido sopa y no pido ensalada (No pido sopa ni ensalada).}$$

La proposición "$x + 3 = 5 \Rightarrow x = 2$" puede expresarse como "$x + 3 \neq 5$ ó $x = 2$". Asimismo su negación queda dada por "$x + 3 = 5$ y $x \neq 2$".

Las proposiciones $\sim (P \Rightarrow Q)$ y $P \wedge \sim Q$ son equivalentes. En efecto, puesto que la tabla de verdad es,

P	Q	$\sim Q$	$P \Rightarrow Q$	$\sim (P \Rightarrow Q)$	$P \wedge \sim Q$
V	V	F	V	**F**	**F**
V	F	V	F	**V**	**V**
F	V	F	V	**F**	**F**
F	F	V	V	**F**	**F**

De esta forma, la negación de una proposición condicional es una conjunción.

La condicional $P \Rightarrow Q$: Si hay frío, entonces usaré el abrigo, puede negarse utilizando lo anterior como

$$\sim (P \Rightarrow Q) = P \wedge \sim Q : \text{Hay frio y no uso un abrigo.}$$

La condicional $P \Rightarrow Q$: Si madrugas mucho, (entonces) amanece más temprano, puede negarse como

$$\sim (P \Rightarrow Q) = P \wedge \sim Q : \text{madrugas mucho y no amanece más temprano.}$$

El lector identificará en ésta última al refrán "No por mucho madrugar amanece más temprano".

En ocasiones es semánticamente más claro expresar a $\sim P \vee Q$ como $Q \vee \sim P$ o bien a $P \wedge \sim Q$ como $\sim Q \wedge P$ (recordar que $\vee$ y $\wedge$ son conmutativos). Por ejemplo, puede decirse *no uso abrigo y hay frío*, en lugar de, *hay frío y no uso abrigo.*

En el ejemplo 1.2.1 se buscaba la negación de la proposición "Todas las ranas son verdes". Una forma de hacerlo es proceder transformando la proposición en la condicional $P \Rightarrow Q$: Si es rana, entonces es verde. La negación está dada como

$$\sim (P \Rightarrow Q) = P \wedge \sim Q : \text{ Es rana y no es verde.}$$

Esta última proposición equivale a las posibles negaciones enunciadas en el ejemplo 1.2.1.

De forma similar, la proposición "Ninguna rana es verde" equivale a la condicional $P \Rightarrow \sim Q$: Si es rana, entonces no es verde. La negación de ésta última es

$$\sim (P \Rightarrow \sim Q) = P \wedge Q : \text{Es rana y es verde,}$$

que es otra forma de expresar "Existe (al menos) una rana que es verde" o "Algunas ranas son verdes".

Dada la proposición condicional directa $P \Rightarrow Q$, la proposición obtenida al intercambiar el antecedente y el consecuente, $Q \Rightarrow P$, se llama proposición **recíproca**. Asimismo, a la proposición $\sim P \Rightarrow \sim Q$ se la denomina la **inversa** y a $\sim Q \Rightarrow \sim P$ la **contrapositiva**.

Ejemplos

Ej 1.3.5 Dada $P \Rightarrow Q$: Si vive en México, entonces es mexicano, se tiene que la recíproca es $Q \Rightarrow P$: Si es mexicano, entonces vive en México, la inversa es $\sim P \Rightarrow \sim Q$: Si no vive en México entonces no es mexicano, y la contrapositiva es $\sim Q \Rightarrow \sim P$: Si no es mexicano, entonces no vive en México.

Ej 1.3.6 Dada $P \Rightarrow Q$: Si cometió un delito, (entonces) tendrá que ser castigado, se tiene que la recíproca es $Q \Rightarrow P$: Si tiene que ser castigado, entonces cometió un delito, la inversa es $\sim P \Rightarrow \sim Q$: Si no cometió un delito, (entonces) no tendrá que ser castigado, y la contrapositiva es $\sim Q \Rightarrow \sim P$: Si no tiene que ser castigado, entonces no cometió un delito.

♠♠♠♠♠

Los valores de verdad para la recíproca, inversa y contrapositiva se obtienen a partir de la siguiente tabla de verdad:

P	Q	$\sim P$	$\sim Q$	$P \Rightarrow Q$	$Q \Rightarrow P$	$\sim P \Rightarrow \sim Q$	$\sim Q \Rightarrow \sim P$
V	V	F	F	**V**	*V*	*V*	**V**
V	F	F	V	**F**	*V*	*V*	**F**
F	V	V	F	**V**	*F*	*F*	**V**
F	F	V	V	**V**	*V*	*V*	**V**

Notamos que la condicional directa y su contrapositiva son equivalentes, así como también lo son la recíproca y la inversa.

Sean

P : El Congreso aprobó la Reforma Fiscal,

Q : El Gobierno Federal tiene mayores ingresos.

Supongamos que P es verdadera y Q es falsa. Entonces las proposiciones condicional directa, recíproca, inversa y contrapositiva y sus valores de verdad son,

$P \Rightarrow Q$: Si el Congreso aprobó la Reforma Fiscal,
entonces el Gobierno Federal tiene mayores ingresos (F),
$Q \Rightarrow P$: Si el Gobierno Federal tiene mayores ingresos,
entonces el Congreso aprobó la Reforma Fiscal (V),
$\sim P \Rightarrow \sim Q$: Si el Congreso no aprobó la Reforma Fiscal,
entonces el Gobierno Federal no tiene mayores ingresos (V),
$\sim Q \Rightarrow \sim P$: Si el Gobierno Federal no tiene mayores ingresos,
entonces el Congreso no aprobó la Reforma Fiscal (F).

1.3.1. Necesidad y suficiencia

La proposición condicional $P \Rightarrow Q$, "Si P, entonces Q", puede reinterpretarse como "P es **suficiente** para Q" o bien "Q es **necesaria** para P" (P no puede ocurrir sin Q). Los siguientes ejemplos ilustran esta situación.

Ejemplos

Ej 1.3.7 $P \Rightarrow Q$: Si tiras el agua, (entonces) mojarás el piso. De aquí que el antecedente "tirar el agua" sea una condición suficiente para mojar el piso, y el piso mojado es una condición necesaria para haber tirado el agua, ya que no puede tenerse un piso seco después de haber tirado el agua.

Ej 1.3.8 Dada $P \Rightarrow Q : x + 3 = 5 \Rightarrow x = 2$ tenemos que $x + 3 = 5$ es suficiente para tener $x = 2$, y $x = 2$ es necesaria para tener $x + 3 = 5$.

Ej 1.3.9 En el conocido refrán "Si madrugas, (entonces) Dios te ayuda", madrugar es condición suficiente para obtener la ayuda de Dios. Asimismo, el tener ayuda de Dios es una condición necesaria para haber madrugado ya que no puede carecerse de la ayuda divina después de haber madrugado.

Ej 1.3.10 La proposición "El Gobierno Federal tendrá mayores ingresos, sólo si se aprueba la Reforma Fiscal" hace énfasis en que la Reforma Fiscal es una condición necesaria para tener más ingresos. La proposición es equivalente a la condicional: "Si el Gobierno Federal tiene mayores ingresos, entonces se aprobó la Reforma Fiscal".

♠♠♠♠♠

La proposición compuesta $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$ se denota por $P \iff Q$ y se conoce como proposición **bicondicional**. Normalmente se lee como "P si y sólo si Q". Notemos que en este caso Q es condición necesaria y suficiente para P, y viceversa. La tabla de verdad asociada es

P	Q	$P \Rightarrow Q$	$Q \Rightarrow P$	$(P \Rightarrow Q) \wedge (Q \Rightarrow P)$
V	V	V	V	V
V	F	F	V	F
F	V	V	F	F
F	F	V	V	V

de manera que la bicondicional es verdadera siempre y cuando P y Q tengan el mismo valor de verdad. Cuando $P \iff Q$ es verdadera, es común decir que P y Q son equivalentes.

Ejemplos

Ej 1.3.11 "Si eres hijo de mis padres, entonces eres mi hermano" y "Si eres mi hermano, entonces eres hijo de mis padres" puede expresarse como "Eres hijo de mis padres si y sólo si eres mi hermano". Ser "mi hermano" es condición necesaria y suficiente para ser "hijo de mis padres".

Ej 1.3.12 La proposición $P \iff Q : x + 3 = 5 \iff x = 2$ con P y Q verdaderas, es verdadera.

Ej 1.3.13 La proposición "Todos los jueces son honestos si y sólo si todos los científicos son inteligentes", es verdadera ya que es de la forma $P \iff Q$ con P y Q falsas.

♠♠♠♠♠

1.4. Argumentos lógicos

Un **argumento** o **razonamiento lógico** se puede pensar como un arreglo de proposiciones

$$\begin{array}{c} P_1 \\ \vdots \\ P_k \\ \hline C \end{array}$$

en donde $P_1, ..., P_k$ se denominan **premisas** y C **conclusión**. Decimos que el argumento es válido si siempre que $P_1, ..., P_k$ son verdaderas, entonces C también lo es. Esto equivale a decir que el argumento es válido cuando $P_1 \wedge \cdots \wedge P_k \Rightarrow C$ es una tautología. Notemos que las proposiciones pueden ser verdaderas o falsas y los argumentos son válidos o inválidos.

Podemos ver que el argumento, conocido como *modus ponens*

$$\begin{array}{c} P \Rightarrow Q \\ P \\ \hline Q \end{array} \tag{MP}$$

es válido de dos formas. Primero, construyendo la tabla de verdad:

P	Q	$P \Rightarrow Q$
V	**V**	**V**
V	F	F
F	V	V
F	F	V

en la cual observamos que si $P \Rightarrow Q$ y P son verdaderas (1^{er} renglón), entonces Q también lo es. La segunda forma es verificando que $[(P \Rightarrow Q) \wedge P] \Rightarrow Q$ es una tautología como se muestra en la siguiente tabla de verdad:

P	Q	$P \Rightarrow Q$	$(P \Rightarrow Q) \wedge P$	$[(P \Rightarrow Q) \wedge P] \Rightarrow Q$
V	V	V	V	**V**
V	F	F	F	**V**
F	V	V	F	**V**
F	F	V	F	**V**

La equivalencia de $P \Rightarrow Q$ y su contrapositiva $\backsim Q \Rightarrow \backsim P$ sugiere que el argumento,

$$\begin{array}{c} P \Rightarrow Q \\ \sim Q \\ \hline \sim P \end{array} \tag{MT}$$

conocido como *modus tollens*, o argumento por contraposición, es válido. En efecto, la tabla de verdad

P	Q	$\sim Q$	$P \Rightarrow Q$	$\sim P$
V	V	F	V	F
V	F	V	F	F
F	V	F	V	V
F	F	**V**	**V**	**V**

muestra que si $P \Rightarrow Q$ y $\sim Q$ son verdaderos, entonces $\sim P$ también lo es. También podemos ver que $[(P \Rightarrow Q) \wedge \sim Q] \Rightarrow \sim P$ es una tautología completando la tabla de verdad como sigue:

P	Q	$\sim Q$	$P \Rightarrow Q$	$(P \Rightarrow Q) \wedge \sim Q$	$\sim P$	$[(P \Rightarrow Q) \wedge \sim Q] \Rightarrow \sim P$
V	V	F	V	F	F	**V**
V	F	V	F	F	F	**V**
F	V	F	V	F	V	**V**
F	F	V	V	V	V	**V**

En matemáticas es común realizar demostraciones *por contradicción o reducción al absurdo*, utilizando (MT). Una de las primeras demostraciones de este estilo se le atribuye a Euclides (alrededor del año 300aC) quién demostró que el conjunto de números primos es infinito[1]. Es decir,

$$\begin{array}{c} \text{Si } \Pi \text{ es el conjunto de números primos, entonces } \Pi \text{ es infinito,} \\ \Pi \text{ es el conjunto de números primos,} \\ \hline \Pi \text{ es infinito.} \end{array}$$

Evidentemente, demostrar directamente que un conjunto es infinito puede ser complicado por lo que se toma el argumento contrapositivo

$$\begin{array}{c} \text{Si } \Pi \text{ es el conjunto de números primos, entonces } \Pi \text{ es infinito,} \\ \Pi \text{ es finito,} \\ \hline \Pi \text{ no es el conjunto de números primos.} \end{array}$$

Euclides asumió entonces que Π era un conjunto finito. De esta forma podía enlistar sus elementos como

$$\Pi = \{p_1, p_2, \ldots, p_n\}.$$

El número

$$q = p_1 \times p_2 \times, \ldots \times p_n + 1$$

no es divisible entre ninguno de los primos de la lista, de manera que por definición q es primo pero $q \notin \Pi$. Así, ¡Π no es el conjunto de números primos! Se dice que se origina una contradicción al asumir "Π es un conjunto finito", con lo cual Π debe ser infinito.

Al argumento

$$\begin{array}{c} P \vee Q \\ \sim P \\ \hline Q \end{array} \tag{SD}$$

se lo conoce como *silogismo disyuntivo*. Éste es un argumento válido ya que la tabla de verdad

[1]Un número entero positivo diferente del 1 es *primo* si no es divisible mas que entre sí mismo y la unidad. Así, 2,3,5,7,11,13,... son primos.

P	Q	$P \vee Q$	$\sim P$	$(P \vee Q) \wedge \sim P$	$[(P \vee Q) \wedge \sim P] \Rightarrow Q$
V	V	V	F	F	**V**
V	F	V	F	F	**V**
F	**V**	**V**	**V**	V	**V**
F	F	F	V	F	**V**

muestra que Q es verdadero cuando $P \vee Q$ y $\sim P$ lo son, o bien, que $[(P \vee Q) \wedge \sim P] \Rightarrow Q$ es una tautología.

Dadas proposiciones P, Q y R, el argumento

$$\begin{array}{c} P \Rightarrow Q \\ \underline{Q \Rightarrow R} \\ P \Rightarrow R \end{array} \tag{T}$$

es válido. A este tipo de razonamiento se lo denomina razonamiento *transitivo*. La demostración de la validez requiere de 8 renglones en la tabla de verdad a manera de incluir todas las combinaciones de los valores de verdad de P, Q y R. La tabla de verdad es

P	Q	R	$P \Rightarrow Q$	$Q \Rightarrow R$	$P \Rightarrow R$
V	V	V	**V**	**V**	**V**
V	V	F	V	F	F
V	F	V	F	V	V
V	F	F	F	V	F
F	V	V	**V**	**V**	**V**
F	V	F	V	F	V
F	F	V	**V**	**V**	**V**
F	F	F	**V**	**V**	**V**

Notamos que cuando ambos, $P \Rightarrow Q$ y $Q \Rightarrow R$, son verdaderos, $P \Rightarrow R$ también lo es de manera que el argumento es válido.

1.4.1. Un matemático divertido: Lewis Carrol

Charles Lutwidge Dodgson (1832-1898) , mejor conocido como Lewis Carroll autor de *Alicia en el País de las Maravillas*, era matemático en la Universidad de Oxford y diácono de la iglesia anglicana. Carroll era un apasionado de la lógica simbólica y tenía un gran sentido del humor. Esta combinación de atributos lo llevó a elaborar una gran cantidad de acertijos lógicos, muchos de los cuales forman parte del folklore de la lógica. Veamos un ejemplo de argumento *carrollianos*. Tomemos las siguientes proposiciones:

Los bebés son ilógicos.
Nadie que pueda controlar a un cocodrilo es despreciado.
Las personas ilógicas son despreciadas.

¿Qué podemos concluir? Lo primero que hacemos es expresarlas mediante las siguientes proposiciones simples:

P : Es un bebé,
Q : Es lógico,
R : Puede controlar un cocodrilo,
S : Es despreciado.

Las tres proposiciones iniciales pueden reescribirse simbólicamente como,

$$\begin{gathered} P \Rightarrow \sim Q \\ R \Rightarrow \sim S \\ \sim Q \Rightarrow S \end{gathered}$$

Utilizando transitividad tenemos el argumento transitivo:

$$\begin{array}{c} P \Rightarrow \sim Q \\ \sim Q \Rightarrow S \\ \hline P \Rightarrow S \end{array}$$

La proposición $R \Rightarrow \sim S$ es equivalente a su contrapositiva dada por $S \Rightarrow \sim R$ y utilizando una vez más transitividad se tiene,

$$\begin{array}{c} P \Rightarrow S \\ S \Rightarrow \sim R \\ \hline P \Rightarrow \sim R \end{array}$$

Así, podemos concluir que "si se es bebé no se puede controlar un cocodrilo", o bien

Ningún bebé puede controlar un cocodrilo.

Esto último también puede expresarse como su contrapositiva $R \Rightarrow \sim P$, es decir,

Si alguien puede controlar un cocodrilo no es un bebé.

Ejemplos

Ej 1.4.1 El argumento

$$\frac{\begin{array}{c}\text{Todas las aves tienen plumas,}\\ \text{(el perico) es un ave,}\end{array}}{\text{(el perico) tiene plumas,}}$$

es válido ya que se aplica (MP). (Recordar que "todas las aves tienen plumas" se puede escribir en forma condicional como "si es ave, entonces tiene plumas".)

Ej 1.4.2 Podemos ver que el argumento

$$\frac{\begin{array}{c}P \Rightarrow Q\\ Q\end{array}}{P}$$

no es válido ya que el tercer renglón de la tabla de verdad

P	Q	$P \Rightarrow Q$
V	V	V
V	F	F
F	**V**	**V**
F	F	V

muestra que $P \Rightarrow Q$ y Q son verdaderos pero P no lo es.

Ej 1.4.3 El argumento

$$\frac{\begin{array}{c}\text{Todos los elefantes tienen cola,}\\ \text{(Lassie) tiene cola,}\end{array}}{\text{(Lassie) es un elefante,}}$$

no es válido pues es de la forma indicada por el ejemplo anterior.

Ej 1.4.4 El argumento

$$\frac{\begin{array}{c}\text{Todas las aves tienen plumas,}\\ \text{el cocodrilo no tiene plumas,}\end{array}}{\text{el cocodrilo no es un ave,}}$$

es válido aplicando (MT).

Ej 1.4.5 El argumento

$$\frac{\begin{array}{c}\text{Es águila o es sol,}\\ \text{no es águila,}\end{array}}{\text{(por lo tanto) es sol,}}$$

es válido aplicando (SD).

Ej 1.4.6 El argumento

$$\frac{\begin{array}{c}\text{Los cocodrilos son reptiles,}\\ \text{(la tortuga) no es un cocodrilo,}\end{array}}{\text{(la tortuga) no es reptil ,}}$$

no es válido ya que corresponde, en forma simbólica al argumento

$$\frac{\begin{array}{c}P \Rightarrow Q\\ \sim P\end{array}}{\sim Q,}$$

en donde P y Q se refieren, respectivamente, a ser cocodrilo y ser reptil. La tabla de verdad

P	Q	$P \Rightarrow Q$	$\sim P$	$\sim Q$
V	V	V	F	F
V	F	F	F	V
F	V	**V**	**V**	**F**
F	F	V	V	V

muestra, que $P \Rightarrow Q$ y $\sim P$ pueden ser verdaderos y, sin embargo, $\sim Q$ es falso.

Ej 1.4.7 El argumento

$$\frac{\text{Si es Navidad es diciembre,}\quad \text{si es diciembre es invierno,}}{\text{si es Navidad es invierno,}}$$

es válido puesto que es un razonamiento transitivo.

Ej 1.4.8 El argumento *genérico* para un fiscal es el siguiente argumento transitivo:

$$\frac{\text{Cierta acción viola la ley,}\quad \text{el acusado llevó a cabo dicha acción,}}{\text{el acusado violó la ley.}}$$

Ej 1.4.9 (Roe vs. Wade) El famoso caso Roe vs. Wade[2] llevó a la legalización del aborto en los Estados Unidos. La opinión del juez Blackmun puede traducirse como sigue:

"El derecho a la privacidad, fundamentado en el concepto de libertad personal y restricciones a la acción del Estado expuestos en la catorceava enmienda o bien, como determinó la Corte de Distrito, en la asignación de derechos a las personas otorgada por la novena enmienda, es suficientemente amplio para incluir la decisión de una mujer a poner o no fin a su embarazo."

Este es un argumento transitivo que puede expresarse como

$$\frac{\text{La catorceava o novena enmienda garantizan el derecho a la privacidad,}\quad \text{el derecho a la privacidad proteje la decisión de una mujer a poner fin a su embarazo,}}{\text{La catorceaba o novena enmienda protejen la decisión de una mujer a terminar su embarazo.}}$$

♠♠♠♠♠

Consideremos los siguientes argumentos:

$$\frac{\text{Si no es ave, (entonces) es reptil}\quad \text{el cocodrilo no es ave,}}{\text{el cocodrilo es reptil.}}$$

[2] *Roe vs. Wade* 410 U.S. 113, 153, 1973.

$$\frac{\text{Si no es ave, (entonces) es reptil,} \quad \text{el perro no es ave}}{\text{el perro es reptil}}$$

Ambos son válidos pues se trata de *modus ponens*. En el primero observamos que la primera premisa es falsa, sin embargo la conclusión es verdadera. En el segundo argumento la primera premisa sigue siendo falsa pero ahora la conclusión es falsa. Es común cometer el error de enunciar alguno de estos argumentos y decir que no es válido. Sin embargo, recordemos que lo que importa es la validez de un argumento, no el valor de verdad que tengan las proposiciones que lo componen.

También podemos tener el caso de un argumento inválido en el cual la conclusión sea verdadera como

$$\frac{\begin{array}{c}\text{Todos los elefantes tienen cola,}\\ \text{(Dumbo) tiene cola,}\end{array}}{\text{(Dumbo) es un elefante.}}$$

Dadas estas observaciones, podemos concluir que la validez no implica veracidad y la invalidez no implica falsedad.

En general cuando debatimos o criticamos un argumento puede ser por varias razones. La primera es que el argumento no sea válido como lo hemos hecho en varios ejemplos. En el caso de argumentos inválidos, el debate es sencillo. Sin embargo, si el argumento es válido la conclusión puede ser falsa, como vimos arriba, cuando alguna de las premisas es falsa. Por lo tanto, debatir la veracidad de las premisas es una forma de cuestionar la conclusión. Por ejemplo, consideremos los siguientes argumentos acerca del aborto:

$$\frac{\begin{array}{c}\text{La mujer tiene derecho a decidir sobre su cuerpo,}\\ \text{(el embrión) es parte del cuerpo de la mujer,}\end{array}}{\text{(la mujer) tiene derecho al aborto.}}$$

$$\frac{\begin{array}{c}\text{La constitución protege la vida humana,}\\ \text{(el embrión) tiene vida humana,}\end{array}}{\text{(la constitución) protege al embrión.}}$$

Ambos argumentos son válidos, no obstante, el debate se centra en la veracidad de las premisas. En el primer argumento, la veracidad de la primera premisa puede ser atacada si pensamos que la mujer no siempre tiene derecho a decidir sobre su cuerpo. Si así fuera, la prostitución sería legal y también podría traficar con sus órganos -vender legalmente un riñón, por ejemplo-. Para defender la veracidad de la premisa, podrían precisarse las instancias en las cuales puede decidirse sobre el cuerpo. En el segundo argumento, la segunda premisa es la que ocasiona controversia. ¿Qué es exáctamente poseer vida humana? Los más militantes alegan que se posee a partir de la concepción, otros dan como inicio de la vida el momento en el cual el embrión podría, potencialmente, tener vida independiente de la madre. Como se ve, el debatir la veracidad de argumentos válidos es una tarea no trivial.

Ejemplo

Ej 1.4.10 (Lemon vs. McDonnell Douglas Corp (1978)) [3] Una azafata demandó a un productor de aviones por lesiones sufridas cuando cayó por una escotilla de emergencia mal cerrada de un DC 10. El testigo experto testificó que la escotilla era "excesivamente peligrosa", sin sustanciar su opinion. Se tiene un simple argumento tipo *modus ponens* dado por,

$$\begin{array}{c}\text{Si la escotilla es excesivamente peligrosa el productor es responsable,}\\ \text{la escotilla es excesivamente peligrosa ,}\\ \hline \text{el productor es responsable.}\end{array}$$

La demanda fue rechazada, aún después de una apelación en la cual se enfatizó que la opinión del experto sembraba la duda acerca de si la escotilla estaba defectuosa. La razón fue que la proposición: "la escotilla es excesivamente peligrosa" no era verdadera simplemente porque un experto puso el calificativo a la escotilla (sin mayores detalles técnicos). El argumento es válido pero la conclusión no es verdadera pues una de las premisas no lo es.

[3] Tomado de "Logic for Lawyers: a Guide to Clear Legal Thinking", R.J. Aldisert, National Institute for Trial Advocacy, Notre Dame, IN, 1997.

♠♠♠♠♠

1.4.2. Argumentos con proposiciones categóricas

Considérese el argumento

$$\begin{array}{c} \text{Algunos duendes son magos,} \\ \text{los magos son entes imaginarios,} \\ \hline \text{algunos entes imaginarios son duendes,} \end{array} \tag{1.2}$$

que puede expresarse simbólicamente como

$$\begin{array}{c} \sim (P \Rightarrow \sim Q) \\ Q \Rightarrow R \\ \hline \sim (R \Rightarrow \sim P), \end{array}$$

en donde P, Q y R representan, respectivamente, el ser duende, mago o ente imaginario. Notemos que $P \Rightarrow \sim Q$ representa "Si se es duende no se es mago" que equivale a "Ningún duende es mago". La negación $\sim (P \Rightarrow \sim Q)$ puede expresarse como "Algunos duendes son magos". Un argumento semejante se aplica para obtener $\sim (R \Rightarrow \sim P)$: Algunos entes imaginarios son duendes.

La tabla de verdad

P	Q	R	$\sim P$	$\sim Q$	$P \Rightarrow \sim Q$	$R \Rightarrow \sim P$	$\sim (P \Rightarrow \sim Q)$	$Q \Rightarrow R$	$\sim (R \Rightarrow \sim P)$
V	V	V	F	F	F	F	**V**	**V**	**V**
V	V	F	F	F	F	V	V	F	F
V	F	V	F	V	V	F	F	V	V
V	F	F	F	V	V	V	F	V	F
F	V	V	V	F	V	V	F	V	F
F	V	F	V	F	V	V	F	F	F
F	F	V	V	V	V	V	F	V	F
F	F	F	V	V	V	V	F	V	F

implica la validez del argumento ya que si $\sim (P \Rightarrow \sim Q)$ y $Q \Rightarrow R$ son verdaderos, $\sim (R \Rightarrow \sim P)$ también lo es. Como podemos ver, puede ser bastante tedioso el realizar tablas de verdad cuando tratamos con argumentos del tipo "Algunos..." y más de dos proposiciones. A continuación describiremos un método gráfico más simple para analizar la validez de cierto tipo de argumentos.

Recordemos algo de la notación básica de conjuntos. Aquí los conjuntos se denotarán por $C_1, ...C_k$ y cada uno de ellos está formado por elementos con ciertas características, digamos los miembros de cierta clase. Si la clase representada por un conjunto, digamos C_1, está contenida en otro conjunto, digamos C_2, entonces escribimos $C_1 \subset C_2$ (**contención**). Se pueden formar nuevos conjuntos, a partir de éstos, mediante las siguientes operaciones: $C_1 \cap C_2$ (**interseccción**) que representa todos los elementos que están tanto en C_1 como en C_2 y $C_1 \cup C_2$ (**unión**) que representa los elementos que están en C_1 o en C_2. Asimismo, $\emptyset$ denota al **conjunto vacío** (sin elementos), y dado cualquier conjunto C su **complemento**, denotado por $(C)^c$, representa todos los elementos que no pertenecen a la clase representada por C.

Ejemplo

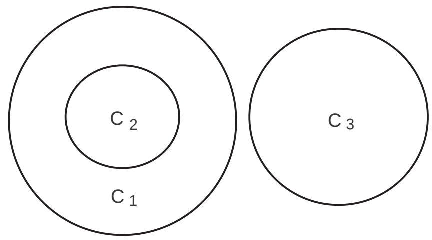
Ej 1.4.11 Supongamos que C_1, C_2 y C_3 denotan, respectivamente, los conjuntos de mamíferos, primates y reptiles. Entonces se cumple $C_2 \subset C_1$, ya que todo primate es mamífero, $C_1 \cup C_3$ representa a todos los animales que son mamíferos o reptiles y $C_1 \cap C_3 = \emptyset$ ya que no existe animal alguno que sea mamífero y reptil.

♠♠♠♠♠

Podemos ilustrar algunas relaciones básicas entre conjuntos mediante los llamados **diagramas de Venn**[4], en donde el conjunto de todos los elementos de una clase dada se representa por un círculo. Por ejemplo, la

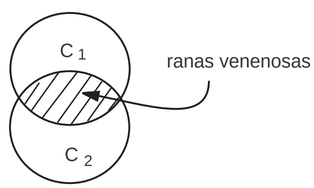
[4]Introducidos por el filósofo británico John Venn (1834-1923).

siguiente figura ilustra la situación del ejemplo anterior:



$C_2(primates) \subset C_1(mamíferos)$,
$C_3(reptiles)\ \cap C_1(mamíferos) = \emptyset$.

Si ahora C_1 representa al conjunto de ranas y C_2 al conjunto de animales venenosos, entonces $C_1 \cap C_2$ representa a las ranas venenosas. Esto se muestra en la siguiente figura:



$C_1 \cap C_2$ representa a la clase
de ranas venenosas.

Sean C_1 y C_2 dos clases de individuos (u objetos). Las proposiciones que relacionan a los individuos de una clase con los individuos de la otra se denominan **proposiciones categóricas**. Éstas utilizan palabras como "todo", "algún" o "ningún" y pueden ser de los siguientes tipos:

1. Algún miembro de C_1 está en C_2 ($C_1 \cap C_2 \neq \emptyset$).
2. Algún miembro de C_1 no está en C_2 ($C_1 \cap (C_2)^c \neq \emptyset$).
3. Ningún miembro de C_1 está en C_2 ($C_1 \cap C_2 = \emptyset$), equivalentemente, Todo C_1 no es C_2 ($C_1 \subset (C_2)^c$).

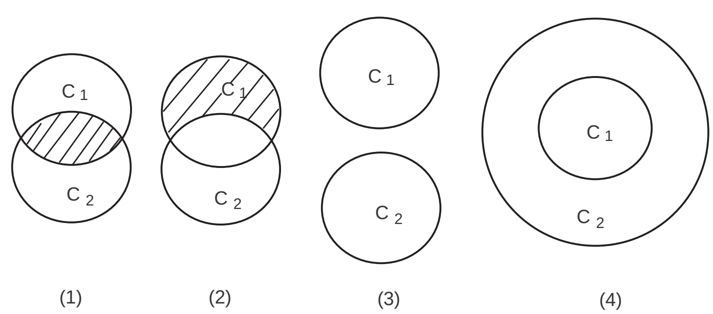
4. Todos los miembros de C_1 están en C_2 ($C_1 \subset C_2$).

Toda proposición categórica tiene dos atributos,

- Cantidad: ésta puede ser universal utilizando los cuantificadores *todos* o *ninguno*, o bien particular utilizando *alguno (s).*

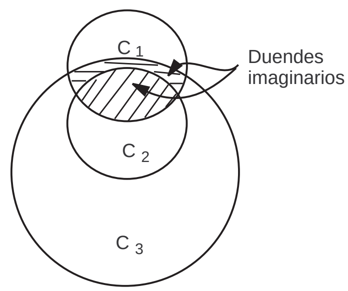
- Calidad: ésta se refiere a si la proporsición es *afirmativa* o *negativa*.

Podemos ilustrar las proposiciones categóricas por medio de diagramas de Venn. Los cuatro tipos de proposiciones categóricas pueden representarse como en la siguiente figura:



Los diagramas de Venn representan los cuatro tipos de proposiciones categóricas.

Regresemos al argumento dado por (1.2) en donde se tienen tres clases: $C_1 = duendes$, $C_2 = magos$ y $C_3 = entes\ imaginarios$. El diagrama de la siguiente figura ilustra esta situación, en donde la clase de duendes que son entes imaginarios al menos contiene a $C_1 \cap C_2$.



El análisis de la figura es:

- Algunos duendes son magos ($C_1 \cap C_2 \neq \emptyset$),
- los magos son entes imaginarios ($C_2 \subset C_3$),

Concluimos que las dos áreas sombreadas representan a los entes imaginarios que son duendes de manera que se cumple $C_1 \cap C_3 \neq \emptyset$. Así, puede concluirse la validez de,

"algunos entes imaginarios son duendes".

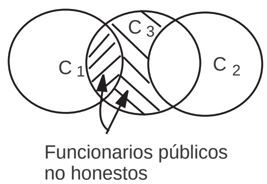
Ejemplos

Ej 1.4.12 La proposición categórica "todos los jueces son honestos" es universal y afirmativa. La proposición "algunos jueces no son honestos" es particular y negativa.

Ej 1.4.13 El argumento

$$\begin{array}{c} \text{Ningún narcotraficante es honesto,} \\ \text{algunos narcotraficantes son funcionarios públicos,} \\ \hline \text{algunos funcionarios públicos no son honestos,} \end{array}$$

es válido. Esto es fácil de ver mediante el diagrama de abajo, en donde C_1 representa a los narcotraficantes, C_2 a los individuos honestos y C_3 a los funcionarios públicos. La clase de funcionarios públicos deshonestos al menos contiene a $C_1 \cap C_3$.

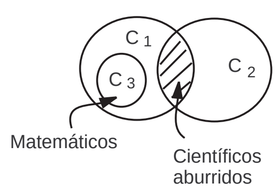


Área sombreada representa a funcionarios públicos deshonestos.

Ej 1.4.14 El argumento

$$\frac{\begin{array}{c}\text{Algunos científicos son aburridos,}\\ \text{Los matemáticos son científicos,}\end{array}}{\text{algunos matemáticos son aburridos,}}$$

no es válido. Esto se ilustra en el siguiente diagrama, en donde C_1, C_2 y C_3 representan, respectivamente, a los científicos, a los aburridos y a los matemáticos. Como vemos el conjunto de matemáticos puede no tener elementos comunes con los aburridos.



Como vemos, los matemáticos no son necesariamente aburridos.

♠♠♠♠♠

Bibliografía complementaria

1. Aldisert R.J., *Logic for Lawyers: a Guide to Clear Legal Thinking*, National Institute for Trial Advocacy, Notre Dame, IN, 1997.

2. Aldisert Ruggero, Clowney Stephen and Peterson Jeremy, "Logic for Law Students: How to Think Like a Lawyer", *University of Pittsburgh Law Review*, 2007. Disponible en forma electrónica en:

$$http : //ssrn.com/abstract = 966597.$$

3. Carroll Lewis, *Symbolic Logic and the Game of Logic*, Dover Publications, New York, NY, 1958.

4. Fiedler Herbert, "Derecho, Lógica y Matemática", *Biblioteca de Ética, Filosofía del Derecho y Política*, Distribuciones Fontamara, S.A., México, D.F. 1992.

5. Kneebone G.T., *Mathematical Logic and the Foundation of Mathematics: An Introductory Survey*, Dover Publications, New York, NY, 2001.

6. Schreiber Rupert, "Lógica del Derecho", *Biblioteca de Ética, Filosofía del Derecho y Política*, Distribuciones Fontamara, S.A., México, D.F. 1992.

7. Stolyar A.A, *Introduction to Elementary Mathematical Logic*, Dover Publications, New York, NY, 1984.

Ejercicios

▷ **1.1** Determinar si los siguientes enunciados son o no proposiciones. En caso afirmativo determinar sus valores de verdad.

1. $5 + 2(3) = 10$,
2. ¡No tires basura!
3. En México se hablan muchas lenguas indígenas.
4. El bosque es más bello que la playa.
5. Las próximas Olimpiadas serán en Beijing, China.

▷ **1.2** Dadas las proposiciones P y Q, encontrar $\sim P, \sim Q$, $P \vee \sim Q$ y $\sim P \wedge Q$.

1. P : Ninguna planta es carnívora, Q :La planta devoró a la hormiga.
2. P : Al menos un estudiante obtuvo un 10 en el examen, Q :El examen no era difícil.
3. P : Todos los modelos están disponibles en todas las tiendas, Q :Algunos zapatos tienen 50 % de descuento.
4. P : Algunas personas concentran toda la riqueza, Q :La riqueza hace la felicidad.
5. P : Todos los números racionales son reales, Q :Al menos un número real no es racional.

▷ **1.3** Elaborar las tablas de verdad para las siguientes proposiciones compuestas:

1. $\sim P \wedge Q$,
2. $\sim P \vee \sim Q$,
3. $P \vee (\sim Q \wedge P)$,
4. $\sim (\sim P \vee \sim Q)$,
5. $(\sim P \wedge \sim Q) \vee \sim Q$,
6. $(P \wedge R) \vee \sim Q$,
7. $(\sim P \wedge Q) \vee \sim R$.

▷ **1.4** Probar que las siguientes proposiciones son equivalentes:

1. $P \vee P, P \wedge P$ yP,
2. $\sim (\sim P)$ y P,
3. $P \vee (P \wedge Q)$ y $P \wedge (P \vee Q)$,
4. $P \vee (Q \wedge R)$ y $(P \vee Q) \wedge (P \vee R)$,
5. $P \wedge (Q \vee R)$ y $(P \wedge Q) \vee (P \wedge R)$.

▷ **1.5** Probar la segunda ley de De Morgan de las equivalencias 1.1: $\sim (P \wedge Q) = \sim P \vee \sim Q$.

▷ **1.6** Dadas las proposiciones P y Q, encontrar $\sim (P \vee Q)$.

1. P : yo pediré pasta, Q : yo pediré pizza.
2. P : estudiaré diariamente, Q : no tendré vida social.

▷ **1.7** Dadas las proposiciones P y Q, encontrar $\sim (P \wedge Q)$.

1. P : yo tomaré vino Q :yo tomaré cerveza.
2. P : Alfonso es deshonesto, Q : Alfonso no es buen actor.

▷ **1.8** Expresar las siguientes proposiciones en forma condicional: "Si P, entonces Q"

1. Las marchas callejeras implican descontento social.
2. Todos los duendes son pequeños.
3. Ningún duende es pequeño.
4. Ninguna rana no es verde.
5. Es necesario comprar baterías para utilizar la linterna.
6. Es suficiente tener un promedio de 9 para conservar la beca.
7. La constitución rige la vida de los mexicanos.
8. Te presto el balón si me dejas jugar.
9. El que es perico, dondequiera es verde.
10. Te arrepentirás, si no vas.
11. La ley Federal del trabajo protege a los trabajadores.
12. El país saldrá adelante cuando tenga instituciones sólidas.
13. Los cítricos contienen vitamina C.
14. Si no repruebo el examen me voy a festejar.
15. Si no te informas no opines.

▷ **1.9** Expresar las condicionales obtenidas en el ejercicio anterior como disyunciones.

▷ **1.10** Negar las condicionales del ejercicio 1.8.

▷ **1.11** Encontrar las proposiciones recíprocas, inversas y contrapositivas para las condicionales del ejercicio 1.8.

▷ **1.12** Determinar si las siguientes proposiciones bicondicionales son verdaderas o falsas:

1. Un polígono regular es un pentágono si y sólo si (el polígono) tiene cinco lados iguales.
2. $5 + 5 = 10 \Longleftrightarrow 1 = 0$.
3. $1 = 0 \Longleftrightarrow 5 = 2$.
4. $P \Longleftrightarrow Q$, en donde P : Es mi coche, Q : El coche es rojo.

▷ **1.13** Determinar, mediante una tabla de verdad, si los siguientes argumentos son o no válidos.

1. $$\begin{array}{c} P\wedge \sim Q \\ P \\ \hline \sim Q. \end{array}$$

2. $$\begin{array}{c} \sim P \Rightarrow \sim Q \\ Q \\ \hline P. \end{array}$$

3. $$\begin{array}{c} P\vee \sim Q \\ P \\ \hline \sim Q. \end{array}$$

▷ **1.14** Utilizando modus tollens, demostrar que el número de enteros pares es infinito.

▷ **1.15** Expresar en forma simbólica y determinar, utilizando argumentos ya conocidos o mediante una tabla de verdad, si los siguientes argumentos son o no válidos.

1. $$\begin{array}{c} \text{Es Hérdez,} \\ \hline \text{es bueno.} \end{array}$$

2. $$\begin{array}{c} \text{Si cumple con los requisitos, obtendrá el crédito.} \\ \text{Si obtiene el crédito comprará la casa} \\ \hline \text{Si cumple con los requisitos, comprará la casa.} \end{array}$$

3. $$\begin{array}{c} \text{Si cumple con los requisitos, obtendrá el crédito.} \\ \text{Cumple con los requisitos.} \\ \hline \text{Obtendrá el crédito.} \end{array}$$

4. $$\begin{array}{c} \text{Los leones son carnívoros.} \\ \text{El perro no es león.} \\ \hline \text{El perro no es carnívoro.} \end{array}$$

5. $$\begin{array}{c} \text{Las ballenas son mamíferos.} \\ \text{Las ballenas viven en el mar.} \\ \hline \text{Los mamíferos viven en el mar.} \end{array}$$

6. $$\begin{array}{c} \text{Ulises conseguirá trabajo o estudiará un posgrado.} \\ \text{Ulises no consiguió un trabajo.} \\ \hline \text{Ulises estudiará un posgrado.} \end{array}$$

7.
Si los cangrejos son escritores, entonces la Tierra no es un planeta.
Si la Tierra no es un planeta, entonces el sol es una estrella.
___
Si los cangrejos son escritores, entonces el sol es una estrella.

8.
El que es perico, dondequiera es verde
No soy verde.
___
No soy perico.

▷ **1.16** Determinar la validez de los siguientes argumentos y la veracidad de sus conclusiones. Justificar.

1.
Toda buena ley debe ser obedecida por los ciudadanos,
las leyes de Newton son buenas ,
___
las leyes de Newton deben ser obedecidas por los ciudadanos.

2.
Las marcas de aguja en el brazo son signo de consumo de drogas,
La detenida tiene marcas de aguja en el brazo,
___
La detenida es drogadicta.

3.
Todo ser vivo debe respetarse,
la propiedad privada no es un ser vivo,
___
la propiedad privada no tiene porque respetarse.

4.
Si el hombre estaba loco, su testamento es inválido,
el hombre no estaba loco,
___
su testamento es válido.

5. $$\frac{\begin{array}{c}\text{O es un patán o es un santo,}\\ \text{no es un patán,}\end{array}}{\text{es un santo.}}$$

▷ **1.17** Determinar, mediante diagramas de Venn, si los siguientes argumentos son o no válidos.

1. $$\frac{\begin{array}{c}\text{Todos los rockeros son músicos.}\\ \text{Algunos poetas son músicos.}\end{array}}{\text{Algunos rockeros son poetas.}}$$

2. $$\frac{\begin{array}{c}\text{Todos los rockeros son músicos.}\\ \text{Algunos poetas son rockeros.}\end{array}}{\text{Algunos poetas son músicos.}}$$

3. $$\frac{\begin{array}{c}\text{Ningún poeta es músico.}\\ \text{Todos los rockeros son músicos.}\end{array}}{\text{Ningún rockero es poeta.}}$$

4. $$\frac{\begin{array}{c}\text{Todas las ranas son pequeñas.}\\ \text{Algunas ranas son príncipes.}\end{array}}{\text{Todos los príncipes son pequeños.}}$$

5. $$\frac{\begin{array}{c}\text{Ningún fantasma es real.}\\ \text{Algunas sombras son fanatasmas.}\end{array}}{\text{Algunas sombras no son reales.}}$$

Capítulo 2

Elementos de Probabilidad

2.1. Introducción

Hasta el momento, los argumentos que hemos examinado no contienen incertidumbre y la conclusión se sigue o no de las premisas inequívocamente. La vida real, sin embargo, casi siempre es incierta y para modelarla utilizamos lo que se conoce como **lógica inductiva.** La diferencia radica en que ésta última trata con argumentos de validez dudosa. En un argumento inductivo, la veracidad de las premisas nos da cierto *grado de certeza* en la veracidad de la conclusión. Para poder cuantificar este grado de certeza se requiere del estudio de algunos conceptos de probabilidad.

En general, los argumentos inciertos no se estudian por el placer que esto conlleva, sino que lo hacemos para poder tomar alguna decisión adecuada. Usualmente, la toma de decisiones implica el análisis de lo que creemos que puede suceder y la elección óptima de acuerdo a lo que nos proporcione mayor beneficio o *utilidad.* En el capítulo 6 examinaremos estos conceptos con mayor detalle.

Un argumento incierto puede parecer válido y, sin embargo, aunque las premisas sean ciertas la conclusión puede no serlo. Supongamos que vamos un día al zoológico y observamos a un león comiendo un gran trozo de carne. Consideremos el siguiente argumento:

Este león come carne,

entonces los leones son carnívoros.

La premisa es verdadera, sin embargo, la conclusión podría no serlo. Vamos a otro zoológico y observamos que en la jaula de los leones, tres de ellos están devorando carne. Tenemos así,

Algunos leones comen carne,
entonces los leones son carnívoros.

La premisa es verdadera, la conclusión podría ser falsa, sin embargo, ahora tenemos más evidencia para sustentarla. Habría que observar más acerca de los hábitos alimenticios de los leones. Si en este proceso observamos que nuestra muestra extendida de felinos es, en efecto carnívora, entonces podemos ir removiendo la incertidumbre acerca de la validez del argumento y tener más confianza en la veracidad de la conclusión.

Notemos que el argumento de arriba toma una característica de una *muestra* (los leones que observamos) y la generaliza para toda la *población* de leones. Veamos otro ejemplo.Vamos al mercado un día por la tarde cuando los puestos están a punto de cerrar. Nos ofrecen 10 kg de mangos por 100 pesos. Los mangos se encuentran en un costal y dudamos de su calidad, ¿qué tal si están pasados o apachurrados? Expresamos nuestra preocupación al vendedor. Acto seguido, toma un mango de arriba del costal y nos lo da para que lo revisemos. El mango huele y se siente delicioso. Pensamos,

Este mango está bueno,
entonces, los mangos del costal están buenos.

La conclusión de este argumento es bastante incierta. El vendedor seguramente puso los mangos buenos hasta arriba para efecto visual. A manera de reducir la incertidumbre, metemos la mano en el costal y tomamos un mango de en medio (al azar). Lo examinamos y también se ve buenísimo. Ahora el argumento es,

Algunos mangos están buenos,
entonces, los mangos del costal están buenos.

La conclusión ya no es tan incierta y compramos el costal de mangos.

Al igual que en el argumento de los leones, tenemos una característica de una muestra y la generalizamos a toda la población. Podemos incrementar el tamaño de la muestra para fortalecer la validez del argumento. Una observación importante es que los elementos de la muestra deben tratar de ser elegidos al azar de entre la población. Evidentemente esto es sencillo en el caso de los mangos, aunque no tanto en el caso de los leones. Estos ejemplos son del tipo

Una característica se cumple para una muestra de la población,
entonces, la carácterística se cumple para toda la población.

Debido a la incertidumbre implícita, podemos redactar estos argumentos como,

Una característica se cumple para una muestra de la población,
entonces, la característica se cumple para casi toda la población.

o podríamos ir también de una muestra a otra,

Estos dos mangos están buenos,
entonces, ***probablemente*** *si tomo dos más, también estarán buenos.*

En el capítulo anterior la inferencia se realizaba de lo general a lo particular (*todos los leones son carnívoros, si es león come carne*), es decir, de la población a la muestra. Se trataba de argumentos deductivos válidos, en los cuales si las premisas son verdaderas la conclusión siempre es verdadera. Los ejemplos de arriba tratan con argumentos inductivos: las premisas pueden ser todas verdaderas y la conclusión es incierta. Es claro que sería deseable asignar un número específico a la validez de estos argumentos. Dado que las premisas son verdaderas, ¿qué tan probable es que la conclusión también lo sea? En otras palabras: ¿qué tan probable es que el argumento sea válido?

Observemos que un león dado será o no carnívoro, igualmente un mango estará bueno o no. La vacilación sobre la veracidad de la conclusión radica en que puede que existan algunos leones que no sean carnívoros o mangos apachurrados. ¿Cuántos? No sabemos con exactitud, ésta es la incertidumbre.

Existen otros argumentos inciertos que no tienen nada que ver con la probabilidad. Todas las disciplinas científicas ofrecen explicaciones acerca de fenómenos que observamos. Cuando se observa algún fenómeno nuevo que no puede explicarse con la teoría existente, ésta se revisa y se cambia o extiende para explicar el nuevo fenómeno. Todas estas teorías ofrecen **inferencia con la mejor explicación posible**, es decir, nunca se tiene la verdad absoluta en términos de ciencia, sólo explicamos y predecimos de la mejor forma posible y por lo tanto siempre hay incertidumbre en los argumentos.

Otro tipo de argumentos inciertos es el que se sustenta en **hechos testimoniales**. En este caso podemos evaluar su grado de incertidumbre, dependiendo de quién es el autor del testimonio: nuestros padres, expertos en el tema, amigos, etcétera.

2.2. Algo de historia

La incertidumbre siempre ha sido parte de la vida humana, su mejor representación es quizás la frase "el capricho de los dioses". El no poder predecir con exactitud los eventos futuros es parte de nuestra razón de ser. La mitología griega tenía a *Tykhe* -también conocida como *Eutykhia*- como diosa de la fortuna, la suerte, la providencia y el destino. *Tykhe* era usualmente representada junto con su compañera *Némesis*, quien era la que imponía límites a los favores otorgados por *Tykhe*.

Nuestra convivencia cotidiana con la incertidumbre genera cierta atracción natural por los juegos de azar. Los hombres primitivos lanzaban huesos u objetos semejantes, en lugar de dados, para decidir su suerte. El intento más antigüo que se conoce de asignar un patrón numérico a un juego de azar fue alrededor del año 960. En esa época, el obispo Wibold de Cambrai describió correctamente los resultados posibles que pueden obtenerse al lanzar tres dados simultaneamente. Sin embargo, el estudio formal de la teoría de la probabilidad es relativamente reciente y data de mediados del siglo XVII.

Su nacimiento oficial se originó con el análisis de un juego de azar en el año de 1654.

Antoine Gombauld Chevalier de Méré (1607 - 1684) era un inteligente y encantador noble francés de la época. Pasaba su tiempo entre su castillo en Poitou, la corte en Paris y las casas de apuestas. Existía en aquel entonces un juego que consistía en lo siguiente: la casa pagaba la apuesta del jugador si éste no obtenía ningún seis al lanzar un dado cuatro veces. A de Méré le intrigaba que con la misma lógica, la casa no diera la misma opción cuando no se obtenía ningún doble seis al lanzar dos dados veinticuatro veces. El sentido común indicaba que como hay 6 veces más resultados posibles, entonces simplemente, en lugar de lanzar un dado cuatro veces se debían lanzar dos dados $4 \times 6 = 24$ veces (en el siguiente capítulo estudiaremos este problema con detalle).

Chevalier de Méré estudió el asunto y llegó a la conclusión de que, en efecto, a pesar de lo que indicaba el sentido común no era redituable para la casa ofrecer el juego alternativo con dos dados. Para cerciorarse consultó con un amigo suyo, un joven matemático de nombre Blaise Pascal (1623- 1662). Pascal resolvió formalmente el problema probando lo siguiente: para el caso de dos dados lanzados veinticuatro veces los momios no favorecen a la casa; no obstante, los momios son ligeramente favorables a la casa si se consideran veinticinco lanzamientos.

Chevalier de Méré estaba interesado también en otro problema conocido como el *problema del juego inconcluso.* Éste consiste en lo siguiente: si después de varias rondas o manos un juego de azar -digamos dados o cartas- tiene que terminarse abruptamente antes de que haya un ganador, ¿cómo deben dividirse las apuestas? En el año de 1654 Pascal le comunicó por escrito éste y otros problemas probabilísticos a Pierre de Fermat (1601-1665)[1]. La correspondencia epistolar que existió entre ellos acerca de estos temas contiene los elementos de la teoría moderna de la probabilidad.

[1]Pierre de Fermat era abogado de profesión y era un simple aficionado a las matemáticas. Hoy en día es reconocido como una de las mentes matemáticas más brillantes, ¡sin embargo para él eran una simple afición!

Una aplicación inmediata de esta teoría fue para evaluar las primas de seguros de vida y de propiedad. En estos casos, es necesario saber la probabilidad de muerte de un individuo o bien la probabilidad de accidente de algún medio de transporte (el hundimiento de un barco, por ejemplo) o la destrucción de algún inmueble[2]. En la segunda mitad del siglo XIX, George Boole y John Venn establecieron la relación entre los conceptos de teoría de la probabilidad existentes y la lógica.

En 1925 John Maynard Keynes (1883-1946) impuso aún mayor rigor a esta relación en su libro *A Treatise on Probabilility,* en el cual el teorema de Bayes, que estudiaremos más adelante, tiene gran preponderancia en el desarrollo inductivo. Casi simultáneamente, un concepto alternativo de razonamiento inductivo fue desarrollado principalmente por Frank Ramsey (1903-1930), Bruno De Finetti (1906-1985) y Leonard Savage (1917-1971): la teoría Bayesiana de la decisión. Entre otras cosas, ésta formaliza el concepto de probabilidad subjetiva que forma una parte estructural de la teoría moderna de decisión.

2.3. Sesgo, independencia y justicia

Todos hemos jugado volados, dados o algo análogo. En ocasiones decimos que el juego no es justo cuando sospechamos que "nos están haciendo trampa". Por ejemplo, si en un juego que involucra lanzar un dado, nuestro oponente casi siempre obtiene el número deseado, no nos parece un resultado normal. Si al lanzar una moneda repetidamente notamos que una cara aparece con el doble de frecuencia que la otra, entonces, probablemente intuímos que la moneda está truqueada. Asimismo, si al tirar el dado obtenemos siempre la sucesión de números:

$$1, 2, 3, 4, 5, 6, 1, 2, 3, 4, 5, 6, 1, 2, 3...,$$

[2]Se recomienda el libro *Against the Gods: The Remarkable Story of Risk* de Peter Bernstein, como una introducción a los intentos para manejar y convivir con el riesgo.

o al lanzar la moneda obtenemos sol, águila, sol, águila, sol, águila,...., repetidamente, también nos parecería extraño[3]. Los dos primeros ejemplos describen un sesgo en los resultados y los dos últimos un patrón o dependencia entre ellos.

Decimos que un juego de azar, como los descritos arriba, es **justo** u **honesto** si sus resultados no presentan asimetría (aparecen con la misma frecuencia) y son **independientes** si no presentan patrón alguno. Tirar una moneda, un dado o girar una ruleta son juegos justos siempre y cuando no estén alterados de alguna manera.

Ejemplos

Ej 2.3.1 (Lanzamiento de oso) Lanzamos un oso de peluche al aire como el siguiente:

El oso gira varias veces y cae al suelo de cuatro posibles maneras, panza abajo, panza arriba, sentado o de cabeza. Al lanzarlo 100 veces se obtiene el número de veces que cae de cada forma como se observa en la siguiente tabla:

resultado	panza abajo	panza arriba	sentado	de cabeza
# de veces	54	40	5	1

Claramente se trata de resultados asimétricos ya que el oso cae panza abajo más de la mitad de las veces y sólo cae de cabeza uno de cada cien lanzamientos. Los resultados, sin embargo, son independientes pues si el oso cae en alguna posición, esto es irrelevante para el siguiente lanzamiento.

[3]En México águila y sol se refieren a las dos caras de una moneda. En otros países podrían ser: cara y cruz, cabezas y colas o algo semejante.

Ej 2.3.2 Consideremos una urna con 25 canicas blancas, 25 rojas, 25 amarillas y 25 azules. Sacamos una canica de la urna y observamos que es amarilla. Sin reemplazar la canica amarilla en la urna tomamos otra canica. Claramente la urna ya no es la misma pues ahora contiene 25 canicas blancas, rojas y azules y 24 canicas amarillas. Nuestras expectativas para el color de la segunda canica no son independientes del resultado de haber tomado una canica amarilla inicialmente. Si la segunda canica es, por ejemplo roja, y tampoco la reemplazamos, entonces la urna contiene 25 canicas blancas y azules y 24 rojas y amarillas. Las expectativas para el color de la tercera canica cambian. En este ejemplo, los resultados de sacar canicas de colores en sucesión y sin reemplazo no son independientes.

Ej 2.3.3 Consideremos la misma urna del ejemplo 2.3.2. Observemos que si cada vez que sacamos una canica y anotamos su color la reemplazamos nuevamente en la urna, entonces el sacar la segunda canica es independiente de lo que hayamos hecho antes. La razón es que tenemos, esencialmente, la misma urna inicial. Este modelo de la urna con canicas es utilizado con frecuencia ya que es sumamente práctico para ciertas abstracciones de la realidad.

Ej 2.3.4 En un casino observamos que los últimos cinco resultados de la ruleta han sido los siguientes: 10 negro, 17 negro, 4 negro, 15 negro y 22 negro. Al observar esto escuchamos el consejo de un experimentado apostador: "ponga todo su dinero en el rojo pues ya toca que salga rojo". Sabiamente no le hacemos caso. La razón es simple: la ruleta no es una urna sin reemplazo sino más bien es una urna con reemplazo. En cada giro, cada número tiene la misma probabilidad de aparecer y se trata de la misma ruleta. Cada giro es independiente de los demás por lo que no hay un patrón definido y los resultados previos no modifican la habilidad para predecir el resultado del siguiente giro.

♠♠♠♠♠

2.4. Concepto de probabilidad: ¿con qué frecuen-

cia?

Vamos ahora a formalizar un poco lo que hemos visto para introducir el concepto de probabilidad. Supongamos que una acción o *experimento* puede tener distintas consecuencias o resultados y sea $\Omega = \{r_1, r_2, ..., r_n\}$ el conjunto de resultados posibles. A este conjunto se le conoce como **espacio de resultados** o **espacio muestral**. Por ejemplo, si lanzamos una moneda al aire el espacio muestral es {águila, sol} y al tirar un dado de seis caras el espacio muestral es {1,2,3,4,5,6}. Es importante notar que, en cada caso, los resultados son **mutuamente excluyentes**, es decir, no pueden ocurrir simultáneamente. Asimismo, el espacio muestral comprende a *todos* los resultados posibles.

Supongamos que el experimento se repite un total de N veces y denotamos por $N(i)$ al número de veces que ocurre el resultado r_i, con $i = 1, ...n$, respectivamente. Entonces, la **probabilidad** (empírica u objetiva) de que ocurra el resultado o consecuencia r_i se define como

$$P(r_i) = \frac{N(i)}{N}. \tag{2.1}$$

Nótese que N puede expresarse como $N = \sum\limits_{i=1}^{n} N(i)$, donde $\sum\limits_{i=1}^{n} N(i)$ es la forma abreviada de denotar sumas, es decir,

$$\sum_{i=1}^{n} N(i) = N(1) + N(2) + \cdots + N(n).$$

Asimismo, obsérvese que

$$0 \leq P(r_i) \leq 1,$$

para toda $i = 1, ..., n$, y también que

$$\sum_{i=1}^{n} P(r_i) = \frac{\sum\limits_{i=1}^{n} N(i)}{\sum\limits_{i=1}^{n} N(i)} = \frac{N}{N} = 1.$$

De lo anterior, las probabilidades son números no negativos, menores que 1 y la suma de las probabilidades de todos los resultados posibles es la unidad.

¿Qué significa esta probabilidad? La definición nos da simplemente la proporción -del total de ocasiones que se realiza el experimento- de veces que aparece la $i-$ ésima consecuencia o resultado. Así, si este número es cercano a 0, entonces el resultado *casi nunca* aparece, o bien es *poco probable*; análogamente, si la probabilidad es cercana a 1, el resultado es *muy probable*. La probabilidad proporciona una medida cuantitativa de qué tan frecuentemente podemos esperar que ocurra una consecuencia o resultado.

En términos de proposiciones, podemos pensar que $P(r_i)$ es la probabilidad de que la proposición

Al realizar cierta acción o experimento se obtiene r_i como resultado,

sea verdadera.

Ejemplos

Ej 2.4.1 Durante 5 años el profesor X ha impartido el mismo curso. En total ha tenido 500 alumnos, de los cuales 80 se han dado de baja, 200 han reprobado y 220 han aprobado. Nos encontramos a uno de estos exalumnos. ¿Cuál es la probabilidad de que éste se haya dado de baja en el curso? En este caso el *experimento* consiste en la inscripción de un estudiante al curso y los resultados posibles son: baja, reprobar o aprobar. El espacio muestral es $\{baja, reprobar, aprobar\}$ y la definición dada por (2.1) nos dice que

$$\begin{aligned} P(baja) &= \frac{80}{500} = 0.16, \\ P(reprobar) &= \frac{200}{500} = 0.4 \text{ y} \\ P(aprobar) &= \frac{220}{500} = 0.44; \end{aligned}$$

en efecto, se verifica que $0{,}16+0{,}4+0{,}44 = 1$. La proposición: "Un estudiante inscrito en el curso del profesor X se habrá dado de baja" será verdadera un 16 % de las veces.

Ej 2.4.2 Una moneda se lanza al aire 300 veces y se obtienen 152 águilas y

148 soles. Entonces,

$$P(\acute{a}guila) = \frac{152}{300} = 0.50\bar{6},$$
$$P(sol) = \frac{148}{300} = 0.49\bar{3}.$$

Ej 2.4.3 El año pasado nacieron 1613 mujeres y 1531 hombres en la Delegación Coyoacán. Si se selecciona uno de estos bebes al azar, ¿cuál es la probabilidad de que sea niña? El experimento es "nacer" con dos posibles resultados: niño o niña. Tenemos así que,

$$P(ni\tilde{n}a) = \frac{1613}{1613 + 1531} = 0.513.$$

Podemos obtener también $P(ni\tilde{n}o)$ como

$$P(ni\tilde{n}o) = 1 - P(ni\tilde{n}a) = 1 - 0.513 = 0.487.$$

Así, la proposición: "un bebe nacido en la Delagación Coyoacán es niña" será verdadera con probabilidad 0.513.

♠♠♠♠♠

En ocasiones el experimento, por su naturaleza, es tal que las consecuencias aparecen con la misma frecuencia -son equiprobables-. Este sería el caso de lanzar una moneda al aire. Notemos que no es necesario realizar el experimento repetidamente para darnos cuenta de que, si la moneda está perfectamente balanceada, entonces se espera que la mitad de las veces aparezca *águila* y la otra mitad *sol*, es decir,

$$P(aguila) = \frac{1}{2} = P(sol).$$

En general, si el experimento tiene n consecuencias equiprobables, entonces esperaríamos que, $N(1) = N(2) = \cdots = N(n)$ y la probabilidad de cada una de ellas es

$$P(r_i) = \frac{N(i)}{nN(i)} = \frac{1}{n}.$$

2.5. Probabilidad subjetiva: ¿con qué certeza?

La definición de probabilidad dada por la ecuación 2.1 tiene algunos problemas. Es común que el *experimento* no pueda realizarse; quizás porque los eventos son poco frecuentes o se carece de datos confiables o simplemente es imposible[4]. En estos casos la definición de probabilidad en términos de frecuencias no tiene sentido.

Cuando alguien nos dice que la probabilidad de que los Cavaliers de Cleveland sean campeones de la NBA es del 60 %, esta proposición no está basada en resultados pasados ya que el equipo actual de los Cavaliers nunca a llegado a disputar el partido Final. Esta probabilidad está basada en la opinión *experta* de los analistas deportivos y tiene que ver con la cantidad de dinero que estarían dispuestos a apostar a favor de Cleveland.

Si escuchamos en las noticias que la probabilidad de que haya una recesión económica en China es del 15 %, este dato se basa en la opinión de algunos economistas. Similarmente, podríamos escuchar a alguien dando un estimado acerca de la probabilidad de que haya existido vida en Marte. Es una estimación personal, basada en observaciones y estudios académicos. Para complicar las cosas, la estimación varía entre los expertos y cada quien dará sus razones para defender su estimación.

Pensemos ahora en un paciente al cual le están practicando una cirujía. Surge una complicación inesperada y en cuestión de segundos el cirujano a cargo toma una decisión acerca del procedimiento médico a seguir. El médico seguramente considera que la probabilidad de que el paciente sobreviva es mayor al tomar esta decisión. Una vez más, se trata de una estimación personal. Lo que distingue al buen médico del mediocre, es que con mayor frecuencia tomará decisiones acertadas ante lo inesperado.

Al tipo de probabilidades descritas en estos ejemplos se les denomina probabilidades subjetivas (o personales). Éstas no se construyen mediante frecuencias de resultados sino en base al *grado de creencia* subjetivo acerca

[4]Por ejemplo, la probabilidad de que un meteorito impacte la ciudad de México, no podría definirse de esta manera.

de la veracidad de una proposición dada. Las teorías de Ramsey, de Finetti y Savage, mencionadas en la introducción, estudian este tipo de probabilidades subjetivas. Es posible definir el concepto general de probabilidad de esta forma como sigue:

Definición 2.5.1 *La probabilidad de que ocurra cualquier evento es el grado (personal) de creencia o certeza que se tiene de que el evento suceda.*

Si se trata de lanzar una moneda, seguramente no necesitamos lanzarla miles de veces para llegar a la conclusión de que la probabilidad de que caiga de una u otra cara es del 50 %. La física del problema y algunos lanzamientos bastarán para llegar a este grado de certeza. Claro está que ésta sería una conclusión subjetiva, basada en una opinión personal, ¿podemos confiar en ella?

En el caso de lanzar una moneda, seguramente estas probabilidades personales y subjetivas serán muy parecidas unas a otras. En general, si se tienen suficientes datos o si hay un consenso entre los diferentes grados de creencia personales, habrá poca incertidumbre en cuanto al valor que debe tomar esta probabilidad subjetiva. Sin embargo, cuando se trata de analizar situaciones para las cuales los datos son escasos, cuestionables o inexistentes, entonces estas probabilidades subjetivas difieren enormemente. Un analista deportivo puede pensar que los Cavaliers ganarán el campeonato con un 60 % de certeza, mientras que otro puede asegurar que los Lakers de Los Ángeles serán campeones con 95 % de certeza.

El tener todo un rango de probabilidades subjetivas para un mismo evento no debe incomodarnos. Esto simplemente refleja las diferencias existentes en la realidad con respecto al grado de creencia acerca de la ocurrencia de un evento. Por ejemplo, la probabilidad de éxito de una nueva técnica médica puede variar enormemente según el médico tratante. Cuando esta técnica se utiliza con mayor frecuencia y se acumula evidencia a su favor -o en su contra-, seguramente la comunidad médica comenzará a coincidir en el grado de creencia que se tiene en dicha técnica. La dispersión en las probabilidades iniciales es buena, ya que es un indicador acerca del desconocimiento o ignorancia que se tiene en la técnica.

Existen dos escuelas para el análisis estadístico de datos: la llamada **escuela frecuentista** que, como su nombre lo indica, se basa en la definición de probabilidad como la proporción de veces que ocurre un evento. Y, la **escuela Bayesiana,** que utilza la definición de probabilidad como grado de creencia o certeza en la ocurrencia de un evento. Cuando se trabaja con un gran número de datos, ambos análisis tienden a coincidir; pero, cuando los datos son escasos o prácticamente inexistentes, las interpretaciones difieren. En general, el análisis frecuentista tiende a ser más simple de realizar, en particular porque existe una gran cantidad de software comercial para llevarlo a cabo. Sin embargo, el análisis Bayesiano es más informativo y permite colocar a los resultados dentro del contexto de resultados previos. El buen estadístico debe poder realizar ambos análisis, según convenga.

2.5.1. Momios

Si visitamos alguna página de apuestas deportivas seguramente estarán llenas de *momios*. Aclaremos que no se trata de criaturas fantásticas ni terroríficas, sino simplemente de una forma alternativa de expresar la probabilidad de un resultado -como el resultado de una contienda deportiva-.

El origen de la palabra es de la región de Navarra, en España, en donde en el juego de Frontón se denomina *momio* a la ventaja de dinero que se obtiene de apostar al equipo con menos posibilidades de ganar. En inglés, los momios son *odds*, que puede traducirse como *posibilidades* u *oportunidades*[5]. Claramente, los momios deportivos provienen del grado de creencia de su autor y representan probabilidades subjetivas.

La definición formal es la siguiente:

Definición 2.5.2 *Supongamos que en una contienda hay dos competidores, A y B. Si los momios se colocan $x : y$ a favor del competidor A (léase "x a y"), esto significa que si A y B compitieran entre sí $x + y$ veces, se esperaría que A ganara x veces. Si $x = y$, se dice que los momios están parejos.*

[5] Ver "Posibilidades, oportunidades, momios: un comentario sobre la traducción del término odds", José A Tapia-Granados, Actualizaciones, vol 39 No 1, pp.69-71, 1997.

De esta definición es inmediato que las probabilidades de ganar la contienda para A y B son

$$P(A \text{ gana}) = \frac{x}{x+y},$$
$$P(B \text{ gana}) = \frac{y}{x+y}.$$

Así, $x > y$ equivale a tener una mayor probabilidad de ganar para A. Si los momios están parejos, cada competidor tiene la misma probabilidad de ganar. En general, la asignación de momios es un ejemplo de probabilidad subjetiva ya que usualmente está basada en el análisis de expertos y no en la frecuencia real de los resultados.

La definición de momios como una razón $x : y$ no es casual, pues éstos representan la razón de las probabilidades entre un evento, digamos, "A gana", y el evento complementario: "A pierde" (o equivalentemente, "B gana"). En efecto, se tiene:

$$\frac{P(A \text{ gana})}{P(A \text{ pierde})} = \frac{\frac{x}{x+y}}{\frac{y}{x+y}} = \frac{x}{y} = x : y.$$

Cuando hablamos acerca de los momios a favor de que una proposición P sea verdadera, simplemente nos referimos a expresar la probabilidad de que P sea verdadera como los momios entre la veracidad de P y $\sim P$.

Ejemplos

Ej 2.5.1 Observemos que los momios de $x : y$ son equivalentes a $nx : ny$. Así los momios de 2 a 3 son iguales que 4 a 6. Claramente es equivalente ganar 2 de cada 5 partidos que 4 de cada 10.

Ej 2.5.2 Los momios a favor de que el equipo brasileño gane el próximo Mundial de futbol son de 3 a 2. La probabilidad de que Brasil gane es de $\frac{3}{3+2} = \frac{3}{5} = 0.6$.

Ej 2.5.3 Los momios de contraer la influenza esta temporada de invierno son de $1 : 9$, es decir, la probabilidad de contraer esta enfermedad es de $\frac{1}{1+9} = \frac{1}{10} = 0.1$.

Ej 2.5.4 La Suprema Corte debe decidir si es legal gravar ciertas prestaciones con un impuesto especial. La probabilidad de que la decisión sea favorable es del 40 %. Esto es, la probabilidad de que la proposición P : "es legal gravar ciertas prestaciones con un impuesto especial" sea verdadera, es de 40 %. Tenemos así que $0{,}40 = \frac{40}{100} = \frac{2}{5} = \frac{x}{x+y}$, en donde $x : y$ son los momios a favor de que P sea verdadera. Entonces, resolviendo para x y y se tiene que $x = 2, y = 3$ y los momios a favor de tener una decisión favorable son de 2 a 3.

♠♠♠♠♠

2.6. Eventos

Consideremos el espacio muestral $\Omega = \{r_1, r_2, ..., r_n\}$. Un **evento** se define como un subconjunto de este espacio, de manera que todos los eventos posibles son:

$$
\begin{aligned}
&\emptyset,\\
&\{r_1\}, \{r_2\}, \ldots, \{r_n\},\\
&\{r_1, r_2\}, \{r_1, r_3\}, \ldots, \{r_2, r_3\}, \{r_2, r_4\}, \ldots, \{r_{n-1}, r_n\},\\
&\vdots\\
&\{r_1, r_2, ..., r_n\}.
\end{aligned}
$$

Lo usual es que los eventos se refieran a resultados con alguna característica de interés, por ejemplo, si lanzamos dos dados podrían interesarnos todas las parejas de números cuya suma sea mayor a cinco. Si se trata de una población de individuos, podríamos querer saber algo acerca de todos los que tienen cierto nivel de ingreso, o los que adquieren cierto nivel educativo o los que tuvieron sarampion de niños, etcétera.

Sea $E = \{r_{i_1}, ...r_{i_k}\}$ con $k \leq n$, un evento cualquiera[6]. Notemos que

[6]El conjunto de subíndices $\{i_1, \ldots, i_k\}$ es un subconjunto de k elementos del conjunto $\{1, \ldots, n\}$.

$P(\{r_{i_1},...r_{i_k}\})$ es la probabilidad de que la proposición compuesta

Se obtiene r_{i_1} o r_{i_2} o... o r_{i_k} como resultado,

sea verdadera. Una vez más, esta probabilidad se obtiene como la proporción de veces que esperamos que este evento ocurra, es decir,

$$\begin{aligned} P(E) = P(\{r_{i_1},...r_{i_k}\}) &= \frac{\sum_{j=1}^{k} N(i_j)}{N} \qquad (2.2)\\ &= \frac{N(i_1)}{N} + \cdots + \frac{N(i_k)}{N} = \sum_{j=1}^{k} P(r_{i_j}). \end{aligned}$$

En el caso de que todos los resultados sean equiprobables sabemos que,

$$P(r_{i_1}) = P(r_{i_2}) = \cdots = P(r_{i_k}) = \frac{1}{n}$$

y la fórmula 2.2 nos dice que.

$$P(E) = P(\{r_{i_1},...r_{i_k}\}) = \frac{k}{n}.$$

Sean E y F dos eventos y sea E^c el evento complementario de E, es decir $E^c = \Omega - E$. Las siguientes observaciones son fáciles de comprobar e intuitivamente claras.

$$\begin{aligned} P(\emptyset) &= 0,\\ P(\Omega) &= P(\{r_1, r_2, ..., r_n\}) = 1,\\ P(E) + P(E^c) &= 1,\\ P(E \cup F) &= P(E) + P(F) - P(E \cap F). \qquad (2.3) \end{aligned}$$

Notemos que en la fórmula para la unión de dos eventos, tenemos que restar la probabilidad de la intersección. La razón es simple: los resultados que pertenecen a la intersección de ambos se contabilizan dos veces, de manera que restamos el término $P(E \cap F)$ para evitar esta doble contabilidad.

Ejemplos

Ej 2.6.1 ¿Cuál es la probabilidad de que aparezca un 5 al lanzar un dado? Aquí, existen 6 consecuencias equiprobables de manera que $P(5) = \frac{1}{6}$.

Ej 2.6.2 ¿Cuál es la probabilidad de tener exactamente dos hijos varones en una familia de tres hijos? Se asume que los hombres y mujeres nacen con la misma probabilidad. El espacio muestral puede expresarse como

$$\{FFF, FFM, FMF, FMM, MFF, MFM, MMF, MMM\}.$$

Aquí, F o M representan sexo femenino o masculino y la “palabra” formada con estas letras representa, de izquierda a derecha, al primer hijo, segundo hijo y tercer hijo. Así, FMF representa una familia con el primogénito y el menor mujeres y un hombre en medio. Vemos que existen 8 consecuencias equiprobables y tres de ellas: FMM, MFM y MMF nos dan una familia con exactamente dos hijos varones. El evento de interés es $E = \{FMM, MFM, MMF\}$ y $P(E) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8} = 0.375$ que es la probabilidad de este tipo de familia.

Ej 2.6.3 Se lanza un dado de seis caras al aire. Sea E el evento, “el resultado es un número par” y F, “el resultado es un número primo”. Calcular la probabilidad de que la proposición, “el resultado es un número par o un número primo” sea verdadera. El espacio muestral y los eventos E y F son, respectivamente: $\Omega = \{1, 2, 3, 4, 5, 6\}$, $E = \{2, 4, 6\}$, $F = \{2, 3, 5\}$. Notemos que $E \cap F = \{2\}$. Entonces,

$$\begin{aligned} P(E \cup F) &= P(E) + P(F) - P(E \cap F) \\ &= \frac{3}{6} + \frac{3}{6} - \frac{1}{6} = \frac{5}{6} = 0.8\bar{3}. \end{aligned}$$

Concluimos que la probabilidad de obtener un par o un primo es de $0.8\bar{3}$ que es la probabilidad de que la proposición sea verdadera.

Ej 2.6.4 El evento complementario[7] del ejemplo 2.6.3 es $(E \cup F)^c = E^c \cap$

[7]Utilizando las leyes de de Morgan para conjuntos.

$F^c = \{1,3,5\} \cap \{1,4,6\} = \{1\}$. Corroboramos que

$$P((E \cup F)^c) = 1 - P(E \cup F) = 1 - \frac{5}{6} = \frac{1}{6} = P(\{1\}).$$

Ej 2.6.5 Un sondeo de opinión entre la población adulta acerca de si se desea una nueva línea del metro nos da los siguientes resultados:

	A favor	En contra	Total
Mujeres	210	50	260
Hombres	180	60	240
Total	390	110	500

Sea F el evento "el encuestado es mujer" y C el evento "está en contra de la nueva línea del Metro". Entonces,

$$\begin{aligned} P(F) &= \frac{260}{500} = \frac{13}{25}, \\ P(C) &= \frac{110}{500} = \frac{11}{50}, \\ P(F \cap C) &= \frac{50}{500} = \frac{1}{10}, \\ P(F \cup C) &= P(F) + P(C) - P(F \cap C) = \frac{13}{25} + \frac{11}{50} - \frac{1}{10} = \frac{16}{25} = 0.64. \end{aligned}$$

Ésta última es la probabilidad de que la proposición: "el encuestado es mujer o está en contra de la nueva línea del Metro" sea verdadera.

♠♠♠♠♠

Bibliografía Complementaria

1. Aguirre Víctor et al., "Fundamentos de Probabilidad y Estadística", 2^a edición. Jit Press, México, D.F. 2006.

2. Bernstein Peter, *Against the Gods: The Remarkable Story of Risk*, Wiley, 1998.

3. Devlin Keith, "The Unfinished Game", Basic Books, New York, NY, 2008.

4. Hacking Ian, "Probability and Inductive Logic", Cambridge University Press, New York, NY, 2001.

5. Keynes John M., "A Treatise on Probability", Rough Draft Printing, 2008.

6. Savage Leonard, "The Foundations of Statistics", Dover Publications, New York, NY, 1972.

7. Tapia-Granados José A., "Posibilidades, oportunidades, momios: un comentario sobre la traducción del término odds", *Actualizaciones*, vol 39 No 1, pp.69-71, 1997.

8. Warren Weaver, "Lady Luck: the Theory of Probability", Science Study Series, Dover Publications, New York, NY, 1982.

Ejercicios

▷ **2.1** Supongamos que Aerolineas Patito ha tenido 2 accidentes en la última semana. Un amigo te incita a elegir esta aerolínea para ir a Acapulco pues, según él, es muy poco probable que ocurra un accidente pues acaban de ocurrir dos de ellos recientemente. ¿Le debes hacer caso? Explica.

▷ **2.2** La ruleta en los casinos norteamericanos, aparte de los números del 1 al 36, contienen un cero y un doble cero, para los cuales la casa siempre gana. Los casinos europeos sólo tienen un cero. Calcular la probabilidad de que ocurra un cero o doble cero en EUA o un cero en europa.

▷ **2.3** Un dado está "cargado", de manera que los números pares son dos veces más probables de obtener que los nones.

1. Determinar $P(3)$.

2. Determinar $P(impar)$

▷ **2.4** La mesa de una ruleta sufre un pequeño desperfecto en una de sus patas de manera que queda ligeramente inclinada hacia una esquina. El juego no será justo esa noche, ¿porqué?

▷ **2.5** Considerar los siguientes resultados: "el *Arsenal* gana la copa de campeones" y "el *Barcelona* gana la copa de campeones". ¿Son independientes? Justificar.

▷ **2.6** Encontrar el espacio muestral de cada uno de los siguientes eventos:

1. Lanzar tres monedas al aire. Cada moneda está perfectamente balanceada tiene dos caras: águila y sol.

2. Se lanzan un dado de seis caras y una moneda.

3. Se lanzan dos dados de seis caras, obtener el espacio muestra que representa la suma de las caras.

4. Las posibles composiciones (en cuanto a género) del conjunto de niños en una familia con 4 hijos.

▷ **2.7** Obtener los espacios muestrales y la probabilidad empírica de cada elemento de las siguientes situaciones:

1. Después de años de escribir libros y notar los errores que comete un autor obtiene lo siguiente: en total ha escrito 22 000 cuartillas y se detectaron 1020 errores en los cuales invirtió dos letras (error disléxico), 835 errores en los cuales se omitió una letra, 620 errores en los cuales se omitió una palabra entera, 548 errores en los cuales se omitió una oración, 400 errores de concordancia de tiempo verbal y 350 errores de concordancia de género o número.

2. La recaudación de datos meteorológicos en los últimos 100 años para la zona metropolitana de la cd. de México, nos indica que para el día 24 de abril la temperatura máxima (en grados Celsius) ha sido como sigue: se han observado 45 veces temperaturas de entre 24 y 28 grados, 22 veces temperaturas mayores a los 28 grados, 26 veces temperaturas entre 17 y 24 grados y 5 veces temperaturas menores a 17 grados.

▷ **2.8** ¿Cuál es la probabilidad de obtener un número diferente de 5 al lanzar un dado?

▷ **2.9** Se lanza una moneda tres veces. ¿Cuál es la probabilidad de obtener al menos dos soles?

▷ **2.10** Durante los últimos 100 años se tienen datos para la temperatura de la ciudad de México, cada hora del día. En particular, para las 14 horas del día 21 de marzo se tiene la siguiente tabla:

grados °C	# de veces	grados °C	# de veces
39	1	25	10
36	2	24	8
33	4	21	7
32	8	20	4
31	11	19	2
29	13	18	2
28	11	15	3
27	13	8	1

Basándonos en esta muestra, ¿cuál es la probabilidad de que la temperatura el día 21 de marzo sea de 32 °C?

▷ **2.11** ¿Cuál es la probabilidad de no obtener ningún sol en los siguientes casos:

1. Se lanza una moneda.

2. Se lanzan dos monedas.
3. Se lanzan tres monedas.
4. Se lanzan cuatro monedas.
5. Se lanzan n monedas.

▷ **2.12** Encontrar el espacio muestral que se obtiene al lanzar dos dados

▷ **2.13** ¿Cuál es la probabilidad de que al lanzar dos dados los números sumen

1. 5?
2. 3?
3. 6?

▷ **2.14** ¿Cuál es la probabilidad de que una familia con 4 hijos tenga dos niños y dos niñas?

▷ **2.15** En el H. Congreso de la Unión se puso a votación una reforma que reduciría sustancialmente el gasto en comunicación social de los gobiernos estatales. Los resultados obtenidos fueron

	A favor	En contra	Total
PRI	25	97	122
PAN	120	3	123
PRD	5	110	115
Otros	10	0	10
Total	160	210	370

Sea E = "el congresista que votó pertenece al PRD", F = "el congresista votó en contra" . Encontrar

1. $P(E)$.
2. $P(F)$.
3. $P(E \cap F)$.
4. $P(E \cup F)$.

▷ **2.16** Dar una probabilidad para los siguientes eventos y una breve explicación de la estimación.

1. La probabilidad de aprobar este curso de matemáticas.
2. La probabilidad de pasar todas las materias este semestre.
3. La probabilidad de que no se vaya la luz en tú casa ningún día de este año.
4. La probabilidad de tener una enfermedad respiratoria este año.
5. La probabilidad de que caiga un meteorito de gran tamaño sobre la Tierra.

▷ **2.17** Un estudiante cree que la probabilidad de obtener más de nueve en un examen es de 0.20. ¿Cuales son los momios a favor de obtener más de un nueve en el examen?

▷ **2.18** Si los momios a favor de que pasen el curso de Teoría del derecho son 4:10, ¿cuál es la probabilidad de que pasen el curso?

▷ **2.19** Si hay un 75 % de probabilidad de que llueva mañana, ¿cuáles son los momios a favor de que llueva mañana?

Capítulo 3

Contando resultados

3.1. Los dados de Chevalier de Méré

Cuando pensamos en *contar* esencialmente estamos tratando de responder a la pregunta: ¿cuántos hay? Esto puede sonar sencillo, pero determinar el número de elementos de ciertos conjuntos puede no ser trivial, sobre todo si el conjunto es grande. Usualmente se tiene una descripción de los elementos a contar y a partir de ésta, debe obtenerse algún método para determinar su número. En esta sección se describen algunas formas de realizar este procedimiento de conteo.

Regresemos a los juegos de dados de la época de Chevalier de Méré. Un jugador apuesta 100 francos a que no obtendrá ningún seis al lanzar un dado cuatro veces seguidas. Si gana, la casa le regresa sus 100 francos y 100 adicionales; si pierde, la casa se queda con sus 100 francos. Para analizar el juego hay que conocer la probabilidad del evento: "que el jugador no obtenga un seis al lanzar cuatro veces el dado". Equivalentemente, podemos encontrar la probabilidad de que obtenga al menos un seis al lanzar el dado cuatro veces, ya que éste es el evento complementario.

El espacio muestral que se obtiene al lanzar un dado cuatro veces puede representarse como sigue[1]:

$$\Omega = \{abcd, \text{ tales que, } a, b, c, d \in \{1, 2, 3, 4, 5, 6\}\}.$$

Es decir, son todas las cuartetas de números en los que cada uno puede tomar cualquier valor entre el 1 y el 6. ¿Cuántos elementos tiene Ω? La respuesta es $6 \times 6 \times 6 \times 6 = 1296$. Esto ilustra un principio básico de conteo:

Si un procedimiento consiste de k pasos o etapas independientes y cada paso puede ocurrir de n_i formas distintas, entonces, el procedimiento puede llevarse a cabo de

$$n_1 \times n_2 \times \cdots \times n_k$$

maneras diferentes. En particular, si $n_i = n$ para toda i, entonces hay n^k formas distintas.

Ejemplos

Ej 3.1.1 Hay 5 caminos diferentes para ir del D.F. a Toluca, 3 caminos para ir de Toluca a Morelia y 4 para ir de Morelia a Guadalajara. ¿Cuántos posibles caminos hay para ir del D.F. a Guadalajara? La respuesta es

$$5 \times 3 \times 4 = 60.$$

Ej 3.1.2 ¿Cuántos números telefónicos posibles hay, si cada número tiene ocho cifras? Aquí se tienen ocho etapas y cada etapa consiste en asignar un número del conjunto $\{0, 1, ..., 9\}$. El total de números telefónicos posibles es,

$$\underbrace{10 \times 10 \times \cdots \times 10}_{\text{8 veces}} = 10^8.$$

Ej 3.1.3 ¿Cuántas contraseñas posibles de cuatro caracteres existen, si los primeros dos caracteres deben ser letras y los dos restantes números entre el

[1]Recordemos que la notación

$$a, b, c, d \in \{1, 2, 3, 4, 5, 6\},$$

simplemente quiere decir que a, b, c y d son elementos del conjunto (o pertenecen al conjunto): $\{1, 2, 3, 4, 5, 6\}$.

0 y el 9? La contraseña es sensible al uso de mayúsculas. Las primeras dos etapas pueden elegirse de 52 formas ya que hay 26 letras en el alfabeto y éstas pueden ser mayúsculas o minúsculas. Las siguientes dos etapas pueden elegirse de diez maneras que son las cifras del 0 al 9. Así, el número de contraseñas posibles es,

$$52 \times 52 \times 10 \times 10 = 27040000.$$

Ej 3.1.4 Supongamos que en el menú de una casa se ofrecen dos platillos: una entrada y un plato principal. Se quiere que la familia tenga una comida balanceada, de manera que si la entrada es rica en carbohidratos, entonces, el plato principal debe ser bajo en carbohidratos y rico en fibra y vitaminas y viceversa. Las opciones para cada plato son: rico en carbohidratos o rico en fibras y vitaminas. ¿De cuántas formas pueden escogerse los dos platos? En este caso $2 \times 2 = 4$ es incorrecto puesto que las etapas no son independientes. Si escogemos la entrada rica en carbohidratos, entonces el plato principal debe ser rico en fibra y vitaminas y si escogemos la entrada rica en fibra y vitaminas, entonces el plato principal debe ser rico en carbohidratos. Tenemos así, sólo 2 opciones y no 4.

♠♠♠♠♠

Podemos ahora calcular el número de elementos del evento E = "no obtener ningún seis al lanzar un dado cuatro veces". En este caso, cada etapa tiene 5 posibilidades ya que sólo admitimos los números del 1 al 5. Tenemos así que E tiene

$$5 \times 5 \times 5 \times 5 = 625$$

elementos.

Como vimos anteriormente, el espacio muestral tiene 6^3 =1296 elementos. Así, el evento "obtener al menos un seis al lanzar el dado cuatro veces" es el evento complementario E^c que tiene $1296 - 625 = 671$ elementos. De aquí se siguen fácilmente las siguientes probabilidades:

$$P(E^c) \ = \frac{671}{1296} = 0.51775,$$
$$P(E) = \frac{625}{1296} = 0.482\,25.$$

Concluimos que la casa tiene una pequeña ventaja y es por eso que está dispuesta a participar en la apuesta.

Examinemos el otro juego que preocupaba a de Méré: se lanzan dos dados y el evento que queremos analizar es, "obtener un doble seis en 24 lanzamientos de los dados". El espacio muestral es ahora,

$$\Omega = \{((r_i, s_j)_1, (r_i, s_j)_2, \ldots, (r_i, s_j)_{24}), \text{ tales que, } r_i, s_j \in \{1,2,3,4,5,6\}\}.$$

Aquí, las parejas $(r_i, s_j)_k$ representan los resultados de los dados en el k-ésimo lanazamiento; por ejemplo, $(r_i, s_j)_k = (2,5)_3$ significa que se obtuvo "dos-cinco" con los dados en el tercer lanzamiento. Notemos que cada entrada $(r_i, s_j)_k$ puede elegirse de $6 \times 6 = 36$ formas diferentes por lo que al lanzar los dados 24 veces tenemos

$$\underbrace{36 \times 36 \times \cdots \times 36}_{24 \text{ veces}} = 36^{24}$$

posibilidades. El evento, "no obtener ningún doble seis en 24 lanzamientos" tiene

$$\underbrace{35 \times 35 \times \cdots \times 35}_{24 \text{ veces}} = 35^{24}$$

elementos, pues en cada lanzamiento hay 35 parejas $(r_i, s_j)_k$ diferentes de la doble seis: $(6,6)_k$. Así, la probabilidad de no obtener ningún doble seis en 24 lanzamientos es,

$$\frac{35^{24}}{36^{24}} = 0.5086,$$

y por lo tanto, la probabilidad de obtener al menos un doble seis en el mismo número de lanzamientos es $1-0.5086$ =0.491 4. Es claro que aquí la casa tiene los momios en contra pues es ligeramente más probable que no salga un doble seis a que si salga.

3.2. Permutaciones y combinaciones

El ejemplo 3.1.4 sugiere que hay otra forma de contar cuando las etapas no son independientes. Si queremos formar números de tres cifras, sin repetir ningún número, entonces lo podemos realizar de

$$10 \times 9 \times 8 = 720$$

maneras distintas. La razón es que el primer número lo podemos elegir de 10 formas, el segundo únicamente de 9 pues ya no podemos utilizar el que hemos elegido antes y el tercero de 8, pues hay dos números que ya no son elegibles.

En general, si $n > r$ son enteros positivos, se definen las **permutaciones de n objetos tomadas en grupos de r** como,

$${}_nP_r = n(n-1)(n-2)\cdots(n-r+1).$$

Para el caso que acabamos de describir, $n = 10$ y $r = 3$. Existe otra forma de representar a ${}_nP_r$ utilizando la notación factorial.

Dado cualquier entero positivo n, se define n **factorial** y se denota por $n!$ como,

$$n! = n(n-1)(n-2)\cdots 1.$$

Entonces, es fácil ver que se cumple

$$\begin{aligned} {}_nP_r &= n(n-1)(n-2)\cdots(n-r+1) \\ &= \frac{n!}{(n-r)!}. \end{aligned}$$

Ejemplos

Ej 3.2.1 ¿Cuántas palabras de tres letras pueden formarse con las letras de la palabra *hojas* si no puede repetirse ninguna letra? La respuesta es

$$5 \times 4 \times 3 = {}_5P_3 = 60.$$

Ej 3.2.2 Cuatro personas ordenan diferentes bebidas en un bar. El cantinero sirve las bebidas pero por falta de atención olvida completamente quién ordenó qué. ¿Cuál es la probabilidad de que el cantinero coloque las bebidas correctamente? Las posibles permutaciones de las cuatro bebidas entre las cuatro personas son

$$_4P_4 = 4! = 24.$$

Sólo uno de estos resultados es el correcto, de manera que la probabilidad de colocarlas adecuadamente es de $\frac{1}{24} = 0.041\bar{6}$.

Ej 3.2.3 ¿Cuál es la probabilidad de que en un grupo de r $(2 \leq r \leq 365)$ personas todas tengan cumpleaños diferentes? Asumir que no hay años bisiestos. El espacio muestral consiste de todos los posibles arreglos de días del año en grupos de r:

$$\Omega = \{c_1 c_2 \cdots c_r \text{ tales que } c_i \in \{1, 2, \ldots, 365\}\}.$$

Aquí, el conjunto $\{1, 2, \ldots, 365\}$ representa al conjunto de fechas {enero 1, enero 2,...,diciembre 31} y c_i es el cumpleaños del i-ésimo individuo que puede elegirse de 365 formas diferentes. El número de elementos del espacio muestral es 365^r. El evento en el cual todos los cumpleaños son diferentes tiene

$$_{365}P_r = \frac{365!}{(365-r)!}$$

elementos, de manera que la probabilidad de que todos tengan cumpleaños diferentes es,

$$P = \frac{\frac{365!}{(365-r)!}}{365^r} = \frac{365!}{365^r(365-r)!}$$

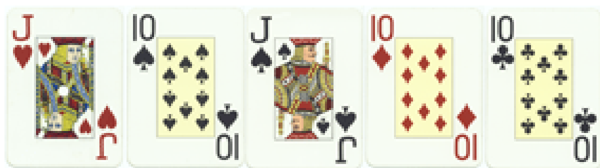
y la probabilidad de que al menos dos cumpleaños coincidan es $1 - P$. Mostramos una tabla de valores de P y $1 - P$ para algunos valores de r.

r	P	1-P
2	0.99726	0.0027397
10	0.88305	0.11695
20	0.58856	0.41144
23	0.4927	0.5073
30	0.29368	0.70632

De aquí, observamos que en un grupo de 10 personas hay aproximadamente el 11.7 % de probabilidad de que al menos dos de ellas tengan el mismo cumpleaños. Si el grupo es de 23 personas, la probabilidad es del 50.73 % y con 30 personas es del 70.63 %.

♠♠♠♠♠

Una mano de poker consta de 5 cartas que pueden elegirse de una baraja con 52 cartas[2]. En este caso no tiene importancia si nuestras cartas son



o bien

pues tenemos exáctamente el mismo *Full House*, aunque el orden de las cartas sea diferente. El valor de la mano de poker depende de las cartas que contenga y no del orden en que aparezcan. ¿Cómo podemos contar todas las manos posibles de poker que pueden repartirse con una baraja?

Si tomaramos las permutaciones de 52 cartas en grupos de 5, es decir ${}_{52}P_5$, las dos manos de arriba aparecerían como elementos diferentes y estaríamos contando cada mano varias veces. Concretamente, ${}_5P_5 = 5!$ veces, que son todas las posibles maneras de ordenar 5 cartas. Entonces, para

[2]Para aquellos que no están familiarizados con una baraja (inglesa), ésta tiene 13 números diferentes (del 1 al 13) y cada uno de ellos viene en 4 posibles "palos" o figuras (tréboles, corazones, diamantes y espadas) para un total de $13 \times 4 = 52$ cartas. Los números del 2 al 10, aparecen explícitamente en la baraja. El 11 corresponde al *Jack* (J) o *Jota*, el 12 a la *Reina* (Q) y el 13 al *Rey* (K). El número 1 es el *As* (A), que puede también tomar el lugar del número 14, después del *Rey*.

contar cada mano una sóla vez simplemente tomamos

$$\frac{{}_{52}P_5}{{}_5P_5}$$

y este número nos proporciona todas las posibles manos de poker. Notemos que éste puede expresarse como

$$\frac{52!}{(52-5)!5!} = 2598\,960.$$

Este número corresponde formalmente a todos los subconjuntos de 5 cartas que pueden obtenerse de un conjunto de 52.

El razonamiento anterior puede generalizarse y definimos así las **combinaciones de n objetos tomados en grupos de r**, denotadas por ${}_nC_r$, como

$${}_nC_r = \frac{n!}{(n-r)!r!}.$$

Ejemplos

Ej 3.2.4 Si se tienen 12 accionistas, ¿de cuántas formas puede elegirse entre ellos un comité de 3 miembros? Queremos todos los subconjuntos de tres accionistas elegidos de un total de doce. Esto es,

$${}_{12}C_3 = \frac{12!}{9!3!} = 220.$$

Ej 3.2.5 La lotería *Melate* consiste en elegir una combinación de seis números diferentes entre el 1 y el 56. ¿Cuál es la probabilidad de escoger la combinación ganadora? El número total de combinaciones es

$${}_{56}C_6 = \frac{56!}{50!6!} = 32468000,$$

como sólo hay una combinación ganadora, la probabilidad de tenerla es

$$\frac{1}{32468000} = 0.0000000308$$

o aproximadamente 1 en 32 millones.

Ej 3.2.6 En una mano de poker, ¿cuál es la probabilidad de obtener una corrida del mismo palo? Para cada palo, hay 10 formas de elegir una corrida pues, esencialmente, una vez determinada la carta inicial las demás están fijas. Las corridas pueden iniciar en cualquier número del 1 al 10 de manera que hay 10 de ellas por cada palo y por lo tanto hay 40 en total. Recordando que hay 2598960 manos posibles de poker, la probabilidad de obtener una corrida del mismo palo es,

$$\frac{40}{2598960} = 0.0000154.$$

Ej 3.2.7 En una mano de poker, ¿cuál es la probabilidad de obtener cuatro cartas iguales? Hay 13 cartas diferentes a elegir, una vez elegida la primera las tres restantes quedan determinadas. La quinta carta puede ser cualquiera de las 48 que quedan de manera que hay $13 \times 48 = 624$ manos con cuatro cartas iguales. La probabilidad de tener una de ellas es

$$\frac{624}{2598960} = 0.00024.$$

Ej 3.2.8 Ahora queremos saber, ¿cuántas manos de poker contienen (exactamente) una tercia? Notemos que la mano debe contener tres números diferentes: el de la tercia y los de las otras dos cartas. Con esto eliminamos la posibilidad de tener cuatro cartas iguales o bien un *Full*: tercia y par. El número de la tercia puede elegirse de 13 maneras, una vez elegido éste, escogemos 3 palos con ese número, de los cuatro posibles, o bien ${}_4C_3$. Así, los tres números iguales pueden elegirse de

$$13 \times_4 C_3$$

formas. De los doce números restantes elegimos dos de ${}_{12}C_2$ formas y después elegimos uno de los cuatro palos posibles para cada uno de ellos. Así, las posibles formas de elegir una tercia son,

$$\begin{aligned}(13 \times_4 C_3) \times ({}_{12}C_2 \times 4 \times 4) &= 13 \times \frac{4!}{1!(3!)} \times \frac{12!}{10!2!} \times 4 \times 4 \\ &= 54912.\end{aligned}$$

De aquí, la probabilidad de obtener una tercia es

$$\frac{54912}{2598960} = 0.21.$$

♠♠♠♠♠

Bibliografía Complementaria

1. Grimaldi Ralph P. *Matemática Discreta y Combinatoria*, 3^a edición, Pearson Prentice Hall, México, D.F., 2002.

2. Haeussler Ernest F. Jr., Paul Richard S., Wood Richard J. *Matemáticas para Administración y Economía*, Pearson, 12^a ed. México D.F., 2008.

Ejercicios

▷ **3.1** Se lanza una moneda siete veces.

1. ¿Cuántos elementos tiene el espacio muestra?
2. ¿Cuál es la probabilidad de obtener siete águilas consecutivas?

▷ **3.2** Verificar las siguientes identidades:

1. ${}_nP_n = n!$
2. ${}_nP_1 = n$.
3. ${}_nC_0 =_n C_n = 1$.
4. ${}_nC_1 = n$.
5. ${}_nC_r =_n C_{n-r}$.

▷ **3.3** Las placas de los autos contienen tres letras seguidas por tres números. Suponiendo que el alfabeto consiste de 26 letras.

1. ¿Cuántas placas posibles hay?
2. ¿Cuántas placas hay que contengan la palabra "oso"?
3. ¿Cuál es la probabilidad de que alguien tenga una placa que comienze con "oso"?

▷ **3.4** En la cafetería del ITAM se ofrecen dos opciones de primer plato, tres opciones de plato principal, dos opciones de verdura y dos opciones de postre. ¿Cuántos posibles menús diferentes hay?

▷ **3.5** Considérense las letras de la palabra "contiguas". Una **palabra** será cualquier reordenamiento de estas letras.

1. ¿Cuántas palabras pueden formarse con estas letras?
2. ¿Cuántas palabras pueden formarse que empiecen con t?
3. ¿Cuántas palabras pueden formarse que empiezen con g y terminen con s?

▷ **3.6** ¿De cuántas maneras pueden formarse 100 personas en una fila?

▷ **3.7** ¿De cuántas maneras pueden formarse 100 personas en una fila de 50 lugares?

▷ **3.8** ¿De cuántas maneras podemos sentar a siete personas en una mesa circular? (nótese que no se distingue entre un arreglo de personas y otro obtenido mediante un giro, sin cambiar el orden de las personas.)

▷ **3.9** En la liga nacional de beisbol los nueve jugadores batean. ¿De cuántas formas puede darse el orden al bat?

▷ **3.10** ¿Cuántas contraseñas de 5 caracteres pueden formarse si el primer caracter debe ser una letra mayúscula y los siguientes pueden ser cualquier letra o número? Se debe distinguir entre mayúsculas y minúsculas y se asume que el alfabeto consta de 26 letras.

▷ **3.11** Si alguien tiene un software que puede probar 10^6 de las contraseñas del ejercio anterior por minuto, ¿en cuánto tiempo podría adivinar una contraseña específica?

▷ **3.12** Consideremos el juego de dos dados de Chevalier de Méré con la siguiente variante: se lanzan dos dados 25 veces seguidas. ¿Es redituable para la casa apostar a que al menos saldrá un doble seis?

▷ **3.13** ¿De cuántas formas pueden elegirse comités de 3 miembros tomados de un grupo de 18 personas?

▷ **3.14** Volver a realizar el ejercicio anterior si cada comité consiste de un Presidente, un secretario y un tesorero.

▷ **3.15** La lotería 6/49 consiste en elegir combinaciones de seis números de los números del 1 al 49. ¿Cuál es la probabilidad de elegir la combinación ganadora?

▷ **3.16** Una mano de poker tiene cinco cartas. ¿De cuántas formas puede obtenerse una mano con

1. un par?
2. dos pares?
3. un *Full* (tercia y par)?

▷ **3.17** Considerar la siguiente situación: 4 personas se encuentran en una fiesta. Durante la plática descubren que dos de ellas, un hombre y una mujer, son Capricornio (el signo del Zodiaco). Se miran estupefactos y comentan que es una coicidencia asombrosa, seguramente están destinados el uno para el otro.

1. Realizar una tabla parecida a la del ejemplo 3.2.3 con la probabilidad de que en un grupo de n personas, al menos dos de ellas pertenezcan al mismo signo del Zodiaco. Utilizar $n = 2, 3, 4,$ y 5. Recordar que hay 12 signos del Zodiaco en un año.

2. En base a los resultados obtenidos, comentar acerca de esta asombrosa coincidencia.

Capítulo 4

Probabilidad condicional y el mundo *Bayesiano*

4.1. Introducción

Supongamos que nos hacen la siguiente pregunta: ¿cuál es la probabilidad de obtener un 6 si lanzamos un dado? La respuesta es simple, la probabilidad de obtener un 6 es

$$P(6) = \frac{1}{6}.$$

Pero si la pregunta es ¿cuál es la probabilidad de obtener un 6 dado que el número es par? La respuesta cambia pues ahora el espacio muestral se reduce a {2,4,6} y la probabilidad de tener un 6 dado que el número es par se denota por $P(6 \mid par)$ y es,

$$P(6 \mid par) = \frac{1}{3}.$$

Consideremos el espacio muestral de lanzar el dado: {1,2,...,6}. El evento "seis y par": $\{6\} \cap \{2,4,6\}$, ocurre con una frecuencia o probabilidad que puede pensarse como el producto

$$P(par)P(6 \mid par) = \frac{1}{2} \times \frac{1}{3},$$

donde

$$P(par) = \text{ porcentaje de veces que obtenemos un par,}$$
$$P(6 \mid par) = \text{porcentaje de veces que obtenemos 6 a partir de un par.}$$

Es decir,

$$P(6 \text{ y par}) = P(par)P(6 \mid par)$$

o bien,

$$P(6 \mid par) = \frac{P(6 \text{ y par})}{P(par)}$$

Obsérvese que, por simetría, también se tiene,

$$P(6 \text{ y par}) = P(6)P(par \mid 6) = \frac{1}{6} \times 1$$

o bien,

$$P(par \mid 6) = \frac{P(6 \text{ y par})}{P(6)}$$

Éstas observaciones, aparentemente sencillas dentro de los juegos de azar, como los dados o las cartas, dieron lugar a uno de los conceptos más elegantes en la teoría de la probabilidad mismo que estudiaremos en la siguiente sección.

4.2. Probabilidad condicional

El concepto de frecuencia de un resultado, dado que ha sucedido previamente otro resultado, puede precisarse y generalizarse de la siguiente forma: si E y F son eventos de un espacio muestral, entonces, la **probabilidad condicional** de E dado F, se denota por $P(E \mid F)$ y está dada por

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}. \tag{4.1}$$

Claramente, si cambiamos los papeles de E y F obtenemos,

$$P(F \mid E) = \frac{P(E \cap F)}{P(E)}. \tag{4.2}$$

En términos de proposiciones, $P(F \mid E)$ es la probabilidad de que la proposición F sea verdadera dado que E es verdadera.

Ejemplos

Ej 4.2.1 Se lanzan dos dados y nos dicen que la suma de los números que se obtuvieron es 7. ¿Cuál es la probabilidad de que los números obtenidos sean 5 y 2? Denotamos a esta probabilidad como $P(5 \text{ y } 2|\text{la suma es } 7)$. El espacio muestral completo tiene $6 \times 6 = 36$ elementos. Podemos, sin embargo, tomar el espacio muestral reducido de los números que suman 7 dado por,

$$\{(1,6),(2{,}5),(3,4),(4,3),(5,2),(6,1)\}.$$

Este espacio tiene 6 elementos y de éstos, los que nos interesan son dos: el (2,5) y el (5,2), de aquí,

$$P(5 \text{ y } 2 \mid \text{la suma es } 7) = \frac{2}{6} = \frac{1}{3}.$$

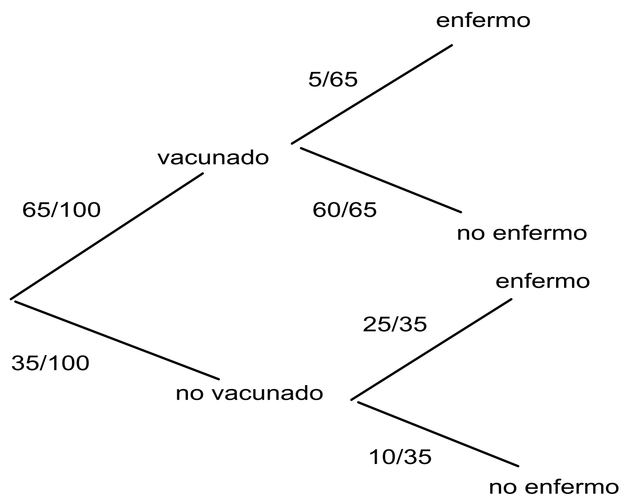
Alternativamente, de la fórmula 4.1 obtenemos,

$$P(5 \text{ y } 2 \mid \text{la suma es } 7) = \frac{P\left((5 \text{ y } 2) \text{ y } (\text{la suma es } 7)\right)}{P(\text{la suma es } 7)} = \frac{\frac{2}{36}}{\frac{6}{36}} = \frac{1}{3}.$$

Ej 4.2.2 Supongamos que en una muestra de 100 personas, 65 de ellas son vacunadas contra la influenza esta temporada de invierno. Cinco de las personas vacunadas contraen la enfermedad. De las 35 personas que no son vacunadas 25 caen enfermas. La siguiente tabla resume esta información:

	enfermo	no enfermo	Total
vacunados	5	60	65
no vacunados	25	10	35
Total	30	70	100

Esto también podríamos expresarlo con el diagrama de árbol presentado a continuación:



En los nodos están los posibles resultados y en las ramas las probabilidades de llegar a ellos. Notemos que la suma de las probabilidades de las ramas que salen de un mismo nodo es igual a uno. Las probabilidades de las ramas son las siguientes:

- P(vacunado)$= \frac{65}{100}$,
- P(no vacunado)$= \frac{35}{100}$,
- P(enfermo |vacunado) $= \frac{5}{65}$,
- P(no enfermo |vacunado) $= \frac{60}{65}$,
- P(enfermo |no vacunado) $= \frac{25}{35}$,
- P(no enfermo |no vacunado) $= \frac{10}{35}$.

Ahora bien, podemos inferir otras probabilidades como, ¿cuál es la probabilidad de que alguien haya estado vacunado y no haya contraido la enfermedad? Esto es,

$$P(\text{vacunado y no enfermo}).$$

- **Ej 4.2.3** Esta probabilidad es simplemente el producto de las probabilidades de las ramas que pasan por el nodo "vacunado" y terminan en en el nodo "no enfermo", es decir,

$$\begin{aligned} P(\text{vacunado y no enfermo}) &= P(\text{vacunado})P(\text{no enfermo} \mid \text{vacunado}) \\ &= \frac{65}{100}\frac{60}{65} = \frac{60}{100}, \end{aligned}$$

lo cual también puede inferirse de la tabla. De forma análoga, la probabilidad de que no haya sido vacunado y no haya caído enfermo es

$$\begin{aligned} P(\text{no vacunado y no enfermo}) &= P(\text{ no vacunado})P(\text{no enfermo} \mid \text{no vacunado}) \\ &= \frac{35}{100}\frac{10}{35} = \frac{10}{100} \end{aligned}$$

De estas dos últimas probabilidades podemos calcular la probabilidad de que un individuo elegido al azar de esta muestra no se enferme ya que, utilizando el resultado dado por 2.3 de la sección 2.6 , tenemos,

$$\begin{aligned} P(\text{no enfermo}) &= P(\text{vacunado y no enfermo}) + P(\text{no vacunado y no enfermo}) \\ &= \frac{60}{100} + \frac{10}{100} = \frac{70}{100}. \end{aligned}$$

Notemos que esta probabilidad también puede inferirse de la tabla.

Ej 4.2.4 Con la misma tabla del ejemplo anterior, ¿cuál es la probabilidad de que alguien que no contrajo la enfermedad haya estado vacunado? Aquí lo que se busca es

$$P(\text{vacunado} \mid \text{no enfermo}).$$

De la tabla vemos que 70 personas no presentaron la enfermedad y de éstos, 60 estaban vacunados por lo que

$$P(\text{vacunado} \mid \text{no enfermo}) = \frac{60}{70}.$$

O alternativamente, podemos utilizar la fórmula 4.1 para obtener

$$\begin{aligned} P(\text{vacunado} \mid \text{no enfermo}) &= \frac{P(\text{vacunado y no enfermo})}{P(\text{no enfermo})} \\ &= \frac{\frac{60}{100}}{\frac{70}{100}} = \frac{60}{70}. \end{aligned}$$

¡Claro que coincide con lo acabamos de obtener! Como observamos, podemos obtener gran cantidad de información a partir de la probabilidad condicional.

♠♠♠♠♠

4.3. Independencia

Recordemos que dos eventos son independientes si la ocurrencia de uno de ellos es irrelevante para el otro. La probabilidad condicional nos proporciona una manera simple de definir este concepto.

Definición 4.3.1 *Sean E y F eventos con probabilidades positivas. Entonces, los eventos son **independientes** si y sólo si se cumple alguna de las siguientes:*

$$P(E \mid F) = P(E),$$
$$P(F \mid E) = P(F).$$

Observemos que si E y F son independientes de acuerdo a la definición anterior, entonces, de la fórmula 4.1 obtenemos,

$$P(E \cap F) = P(E)P(F), \tag{4.3}$$

lo cual nos da una forma alternativa de verificar la independencia. Así, la probabilidad de obtener dos águilas seguidas al lanzar una moneda dos veces es $\frac{1}{2} \times \frac{1}{2}$, ya que si a_1 y a_2 son los eventos de obtener águila en el primer y segundo lanzamiento, respectivamente, éstos son independientes y

$$P(a_1 \cap a_2) = P(a_1)P(a_2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}.$$

Es importante notar que si los eventos son mutuamente excluyentes, es decir, no pueden ocurrir simultáneamente y tienen probabilidades positivas, entonces, no pueden ser independientes. En efecto, por un lado, si $P(E) \neq 0$

y $P(F) \neq 0$, entonces $P(E)P(F) \neq 0$. Por otro lado, si E y F no pueden ocurrir simultáneamente se tiene que $P(E \cap F) = 0$; entonces, $P(E \cap F) \neq P(E)P(F)$ y los eventos no pueden ser independientes. En otras palabras, cualesquiera eventos independientes con probabilidades positivas no pueden ser mutuamente excluyentes.

El concepto de independencia puede extenderse a independencia condicional; concretamente, tenemos la siguiente definición:

Definición 4.3.2 *Sean E, F y D eventos con probabilidades positivas. Entonces, los eventos E y F son* ***condicionalmente independientes*** *dado D si y sólo si,*

$$P(E \cap F \mid D) = P(E \mid D)P(F \mid D).$$

Esto equivale a decir que "E dado D" y "F dado D" son eventos independientes.

Ejemplos

Ej 4.3.1 Supongamos que se tiene una baraja de 52 cartas. La probabilidad de seleccionar un corazón es de $\frac{13}{52}$ y la probabilidad de seleccionar un as es de $\frac{4}{52}$. La selección de un as y la selección de un corazón son eventos independientes de manera que

$$\begin{aligned} P(\text{ as y corazón}) &= P(as)P(corazón) = P(\text{as de corazones}) \\ &= \frac{4}{52} \times \frac{13}{52} = \frac{1}{52}. \end{aligned}$$

Ej 4.3.2 En ocasiones no es obvio que dos eventos son independientes. La fórmula 4.3 nos proporciona una forma sencilla de verificar la independencia, como lo ilustra este ejemplo. Se lanzan dos dados, uno rojo y otro verde. Considérense los eventos,

$$\begin{aligned} E &= \{\text{los números suman 5}\}, \\ F &= \{\text{el número en el dado verde es impar}\}. \end{aligned}$$

¿Son independientes estos eventos? Vamos a verificar si satisfacen la fórmula 4.3. Supongamos que cada pareja (a,b) representa el número del dado verde

(a) y el del dado rojo (b). Los eventos anteriores están descritos por,

$$\begin{aligned} E &= \{(1,4),(2,3),(3,2),(4,1)\}, \\ F &= \{(1,1),(1,2),\cdots,(3,1),\cdots,(5,1),\cdots,(5,6)\}. \\ E\cap F &= \{(1,4),(3,2)\}. \end{aligned}$$

Notemos que F contiene 18 parejas (3×6). El espacio muestral de todas las parejas tiene 36 (6×6) elementos, se cumple entonces,

$$P(E\cap F)=\frac{2}{36}=\frac{4}{36}\frac{18}{36}=P(E)P(F),$$

de manera que los eventos son independientes.

Ej 4.3.3 Supongamos que cuando salimos a comer nuestra elección de postre es independiente del platillo principal que pidamos. La mitad de las veces que comemos en el *Bistró Mosaico* ordenamos salmón a la parrilla y la mitad de las veces pedimos tarta de piñón. ¿Cuál es la probabilidad de ordenar salmón a la parrilla y tarta de piñón dado que comemos en el *Bistró*? La probabilidad que queremos es,

$$P(\text{salmón y tarta} \mid \text{comer en el } \textit{Bistró}).$$

Los eventos de comer salmón y tarta son condicionalmente independientes dado que comemos en el Bistró, así, la definición 4.3.2 nos dice que,

$$\begin{gathered} P(\text{salmón y tarta} \mid \text{comer en el } \textit{Bistró}) = \\ P(\text{salmón} \mid \text{comer en el } \textit{Bistró})P(\text{tarta} \mid \text{comer en el } \textit{Bistró}) = \\ 0.5 \ \times 0.5 \ = 0.25. \end{gathered}$$

♠♠♠♠♠

4.4. Teorema de Bayes

Thomas Bayes (1702-1761) fue un ministro presbiteriano y matemático inglés. Hoy en día es reconocido por su influencia en la teoría de la probabilidad a pesar de que escribió sólo un par de trabajos al respecto. El más importante, *An Essay towards Solving a Problem in the Doctrine of Chances* (1763), fue editado y publicado por un amigo suyo después de su muerte. La relación entre este trabajo y el teorema o fórmula de Bayes que enunciaremos abajo es bastante tenue, aunque implícitamente el teorema puede derivarse de los resultados ahí descritos. Sin embargo, la mayor aportación de Bayes radica en la idea de *inferencia inversa*, que va de los efectos a las causas, en lugar de ir de las causas a los efectos.

Supongamos que E y E^c representan los eventos de tener o no una enfermedad viral y que existe una prueba sanguínea para detectar este virus. Denotamos por $+$ al evento de tener una prueba positiva y por $-$ al que sea negativa. Las probabilidades $P(E)$ y $P(E^c)$ pueden pensarse como las **probabilidades *a priori*** de tener o no la enfermedad; asimismo, $P(E \mid +)$, $P(E^c \mid +)$, $P(E \mid -)$ y $P(E^c \mid -)$ son las **probabilidades *a posteriori***, es decir, cuando tenemos información acerca de la prueba de sangre. Una forma de calcularlas, utilizando las fórmulas 4.1 y 4.2 es la siguiente:

$$\begin{aligned} P(E \mid +) &= \frac{P(E \cap +)}{P(+)} \\ &= \frac{P(+ \mid E)P(E)}{P(+)} \\ &= \frac{P(+ \mid E)P(E)}{P(+ \cap E) + P(+ \cap E^c)} \\ &= \frac{P(+ \mid E)P(E)}{P(+ \mid E)P(E) + P(+ \mid E^c)P(E^c)} \end{aligned} \tag{4.4}$$

y similarmente para los otros casos. Las probabilidades *a posteriori* son una revisión de las probabilidades originales, actualizadas con la nueva información del resultado de la prueba de sangre.

Notemos que $P(+ \mid E)$ es la probabilidad de que la prueba resulte positiva dado que tenemos la enfermedad. Este número nos proporciona una idea de qué tan confiable es la prueba de sangre. En contraste, $P(E \mid +)$ es la probabilidad de tener la enfermedad dado que la prueba de sangre resultó positiva para el virus, es decir, nos indica qué tan confiable es el resultado de la prueba. La fórmula 4.4 nos indica como calcular esta probabilidad *a posteriori*.

El resultado anterior es válido para el caso general de dos eventos E y F y es una versión simplificada del llamado teorema o fórmula de Bayes que ahora enunciamos.

Teorema 4.4.1 (Bayes) *Sean E, E^c eventos complementarios de un espacio muestral y sea F algún evento tal que $P(F) \neq 0$, entonces, la probabilidad a posteriori de E dado F está dada por,*

$$P(E \mid F) = \frac{P(F \mid E)P(E)}{P(F)},$$

o equivalentemente,

$$P(E \mid F) = \frac{P(F \mid E)P(E)}{P(F \mid E)P(E) + P(F \mid E^c)P(E^c)}.$$

En general, si el espacio muestral es la unión de los eventos $E_1, E_2, \ldots, E_n$ que son mutuamente excluyentes y F es algún evento con $P(F) \neq 0$, la fórmula anterior se generaliza como sigue[1]*:*

$$\begin{aligned} P(E_i \mid F) &= \frac{P(F \mid E_i)P(E_i)}{P(F)} \\ &= \frac{P(F \mid E_i)P(E_i)}{\sum_{j=1}^{n} P(F \mid E_j)P(E_j)}. \end{aligned}$$

Si E es la causa y F el efecto, $P(E \mid F)$ representa la probabilidad de la causa dado que se observó el efecto. El teorema de Bayes nos puede contestar preguntas como, dado que el pavimento está mojado, ¿cuál es la probabilidad de que hubiese llovido? Dado que un testigo es confiable, ¿qué tan confiable es su testimonio? Dado que aumentó el desempleo, ¿qué tan probable es que haya una recesión económica? A continuación vemos algunos ejemplos.

[1] Recordemos que

$$\sum_{j=1}^{n} P(F \mid E_j)P(E_j)$$

es simplemente una forma abreviada para denotar a la suma

$$P(F \mid E_1)P(E_1) + P(F \mid E_2)P(E_2) + \cdots + P(F \mid En)P(En).$$

Ejemplos

Ej 4.4.1 La prueba para detectar la presencia de ciertos esteroides en el cuerpo de los atletas tiene *falsos positivos*, si la prueba es positiva sin haber utilizado esteroides y *falsos negativos*, si la prueba es negativa a pesar de haber utilizado esteroides. Denotemos por E y E^c a los eventos de haber utilizado o no esteroides y por $+$ y $-$ al resultado de una prueba de sangre positiva o negativa, respectivamente. Supongamos que se estima que el 4 % de cierto grupo de atletas ha utilizado esteroides, de manera que

$$\begin{aligned} P(E) &= 0.04, \\ P(E^c) &= 0.96 \end{aligned}$$

y que los falsos positivos y negativos están dados por

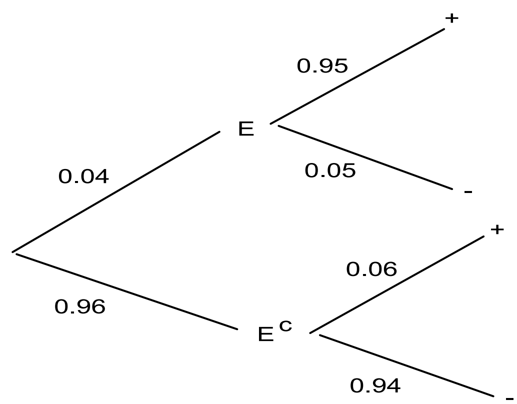
$$\begin{aligned} P(+ \mid E^c) &= 0.06, \\ P(- \mid E) &= 0.05. \end{aligned}$$

Esto nos dice que en un 6 % de los casos un atleta que no utiliza esteroides tiene una prueba positiva y, en un 5 % un atleta que si los ha utilizado tiene una prueba negativa.

Supongamos que un atleta es sometido a la prueba y ésta sale positiva, ¿cuál es la probabilidad de que haya utilizado esteroides? Aquí la probabilidad que quiere encontrarse es

$$P(E \mid +).$$

El siguiente diagrama de árbol ilustra esta situación:



Dado que $P(+ \mid E) = 1 - P(- \mid E) = 0.95$, el teorema de Bayes nos dice que $P(E \mid +)$ puede encontrarse como,

$$P(E \mid +) = \frac{P(+ \mid E)P(E)}{P(+)} \tag{4.5}$$

$$= \frac{P(+ \mid E)P(E)}{P(+ \mid E)P(E) + P(+ \mid E^c)P(E^c)} \tag{4.6}$$

$$= \frac{0.95 \times 0.04}{0.95 \times 0.04 + 0.06 \times 0.96} = \frac{0.038}{0.0956} = 0.39749.$$

La probabilidad del evento complementario: "no haber utilizado esteroides dado que la prueba resultó positiva", es de

$$P(E^c \mid +) = 1 - 0.39749 = 0.60251$$

Esta es una probabilidad demasiado alta y sería injusto penalizar al atleta en base a esta evidencia[2].

[2] La corredora de larga distancia, Mary Slaney, obtuvo un resultado positivo para la presencia de esteroides en los Juegos Olímpicos de 1996. Las autoridades deportivas de EUA, le prohibieron volver a competir. Slaney apeló la decisión y sus abogados utilizaron un proceso de inferencia bayesiana para demostrar que la prueba no tomó en cuenta la probabilidad a priori y asumía implícitamente que la atleta era culpable. Slaney ganó el caso.

Ej 4.4.2 (La falacia del fiscal) Por desgracia, muchos jueces confunden las probabilidades condicionales dadas por,

$$P(I \mid E) = \text{probabilidad de inocencia dada la evidencia}$$

y

$$P(E \mid I) = \text{probabilidad de la evidencia dada la inocencia.}$$

A esta confusión se le conoce comunmente como *la falacia del fiscal.* Para determinar la inocencia de un acusado dada la evidencia debe utilizarse $P(I \mid E)$. Sin embargo, es común que el fiscal proporcione la probabilidad de haber encontrado la evidencia si el acusado es inocente, es decir, $P(E \mid I)$.

En el ejemplo 4.4.1 vimos que La probabilidad de resultar positivo para esteroides (evidencia) dado que el atleta no los utilizó (es inocente) era

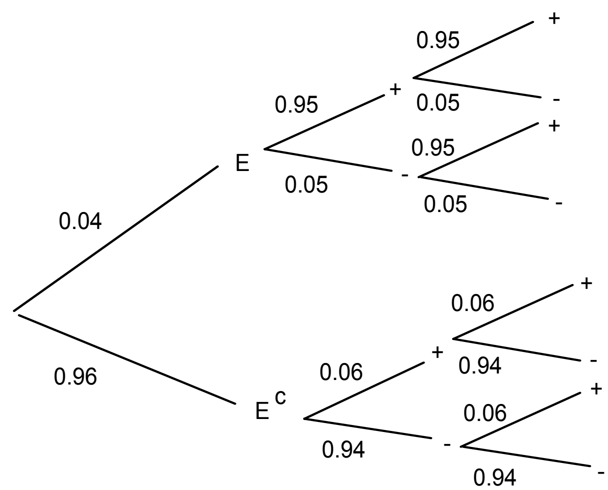
$$P(+ \mid E^c) = 0.06,$$

o bien la probabilidad de un falso positivo. Por otro lado, la probabilidad de que el atleta no hubiese utilizado esteroides (es inocente) dado que la prueba resultó positiva era

$$P(E^c \mid +) = 0.6.$$

La diferencia entre estas dos probabilidades es de un orden de magnitud y sería inaceptable que el atleta fuese declarado culpable por simple incompetencia de un juez.

Ej 4.4.3 Retomemos el escenario del ejercicio anterior. Para tener mayor certidumbre en el resultado de la prueba de sangre, se realiza una segunda prueba, (condicionalmente) independiente de la primera. Ahora queremos saber la probabilidad de haber utilizado esteroides dadas dos pruebas positivas, es decir, queremos encontrar $P(E \mid + \cap +)$. El diagrama árbol correspondiente es,



De acuerdo al teorema de Bayes y a la definición 4.3.2, $P(E \mid +\cap+)$ es

$$
\begin{aligned}
P(E \mid +\cap+) &= \frac{P(+\cap+ \mid E)P(E)}{P(+\cap+)} \\
&= \frac{P(+\cap+ \mid E)P(E)}{P(+\cap+ \mid E)P(E) + P(+\cap+ \mid E^c)P(E^c)} \\
&= \frac{P(+ \mid E)P(+ \mid E)P(E)}{P(+ \mid E)P(+ \mid E)P(E) + P(+ \mid E^c)P(+ \mid E^c)P(E^c)} \\
&= \frac{0.95 \times 0.95 \times 0.04}{0.95 \times 0.95 \times 0.04 + 0.06 \times 0.06 \times 0.96} \\
&= \frac{0.0361}{0.0361 + 0.003456} \\
&= \frac{0.0361}{0.039556} = 0.9126.
\end{aligned}
$$

Es decir, con un 91 % de certidumbre el atleta utilizó esteroides. Este resultado seguramente ameritaría una sanción.

Ej 4.4.4 Un albino es enjuiciado por haber cometido un crimen. El testigo principal asegura haber visto los hechos a través de una ventana de su casa e identifica a un individuo albino como el criminal. Realizando pruebas para la credibilidad del testigo, se determina que a la distancia y en las condiciones de luz desde las cuales presenció el crimen, identifica correctamente a una persona albina o no albina un 95 % de las veces. Aproximadamente una de cada 17 mil personas presenta la condición de albinismo. ¿Cuál es la probabilidad de que el albino haya cometido el crimen dado que el testigo lo identificó?

Sea T_a el evento “el testigo identifica a una persona como albina” y $\sim T_a$ “el testigo identifica a una persona como no albina”. Dado que el testigo identifica correctamente a una persona albina o no albina el 95 % de las veces, se tiene que

$$\begin{aligned} P(T_a \mid albino) &= 0.95, \\ P(\sim T_a \mid \text{ no albino}) &= 0.95. \end{aligned}$$

Asimismo, la probabilidad de albinismo es $P(albino) = \frac{1}{17000} = 0.000058824$. Sea $P(albino \mid T_a)$ la probabilidad de que la persona sea albina dado que el testigo la identifica como tal. De acuerdo al teorema de Bayes y dado que $P(\text{no albino}) = 1 - P(albino)$ y $P(T_a \mid \text{no albino}) = 1 - P(\sim T_a \mid \text{no albino})$ tenemos,

$$\begin{aligned} P(albino \mid T_a) &= \frac{P(T_a \mid albino)P(albino)}{P(T_a)} \\ &= \frac{P(T_a \mid albino)P(albino)}{P(T_a \mid albino)P(albino) + P(T_a \mid \text{no albino})P(\text{ no albino})} \\ &= \frac{0.95 \times 0.000058824}{0.95 \times 0.000058824 + 0.05 \times 0.999\,94} = 0.001165, \end{aligned}$$

esta es la probabilidad de que el testimonio del testigo sea creíble, ¡bajísima! Sin embargo, el testigo parecía sumamente confiable ya que identificaba correctamente a una persona albina o no albina en un 95 % de los casos. ¿Por qué sucede esto? La razón es que la probabilidad de que un individuo sea albino es sumamente baja. En el ejercicio 4.10 se ilustran otros casos para ver como influye esta probabilidad.

Ej 4.4.5 (La Catafixia) Este es un juego común de concurso: el animador del show -Chabelo, en el caso de La Catafixia en México y Monty Hall en la TV americana- nos presenta tres puertas cerradas. Detrás de una de ellas hay un coche último modelo y detrás de las otras dos nada. Si escogemos la puerta correcta el coche es nuestro. Numeremos a las puertas por 1, 2 y 3. El animador nos pide escoger una puerta sin abrirla, digamos que escogemos la puerta 3. El animador entonces escoge una de las otras dos puertas, digamos la puerta 2, la abre y muestra que no hay nada. Acto seguido nos dice: ¿quieres mantener tú decisión de abrir la puerta 3 o quieres *catafixearla* por la puerta 1? ¿Qué debe hacerse?

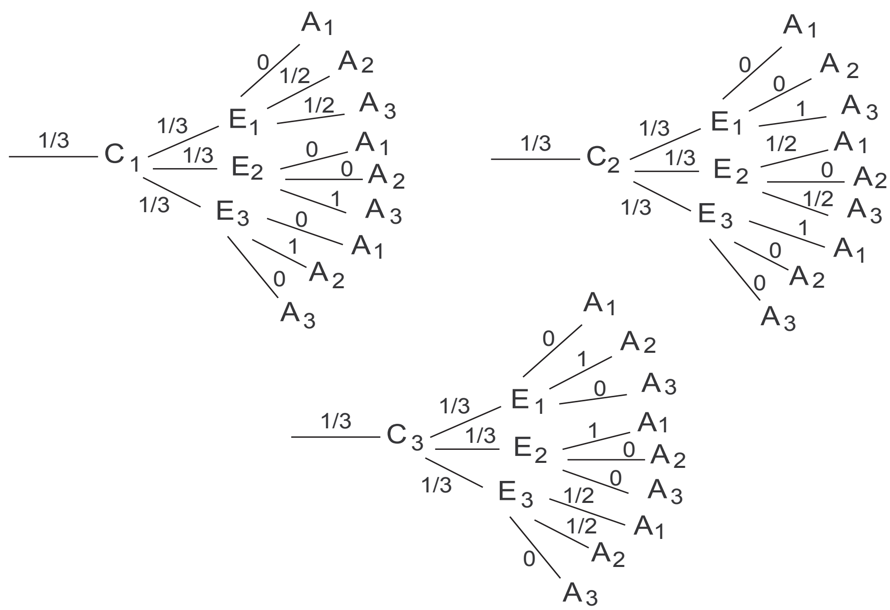
Asumimos que el animador sabe en donde se encuentra el coche y que el coche puede estar en cada puerta con probabilidad $\frac{1}{3}$. Adicionalmente, si elegimos la puerta correcta inicialmente, el animador elegirá con probabilidad 1/2 alguna de las otras dos puertas para mostrarnos que no hay nada. Sean C_1, C_2 y C_3 los eventos que representan que el coche esté en la puerta 1, 2 ó 3, respectivamente. Asimismo, E_1, E_2 y E_3 representan escoger las puertas 1,2 ó 3 por parte del jugador y A_1, A_2 y A_3 que el animador abra la puerta 1,2 ó 3 y muestre que no hay nada. Lo que nos interesa saber para tomar una decisión son las siguientes probabilidades:

$$
\begin{aligned}
&P(C_1 \mid E_3 \cap A_2),\\
&P(C_3 \mid E_3 \cap A_2) = 1 - P(C_1 \mid E_3 \cap A_2).
\end{aligned}
$$

Es decir, la probabilidad de que el coche esté en la puerta 1 si elegimos 3 y el animador abrió 2 y no había nada. Ésta la comparamos con el evento complementario que es, el que el coche esté en la puerta 3 si elegimos 3 y el animador abrió 2 y no había nada. Notemos que sabemos que

$$P(C_2 \mid E_3 \cap A_2) = 0.$$

Para mayor claridad, realizamos un diagrama de árbol y separamos las ramas correspondientes a cada puerta como sigue:



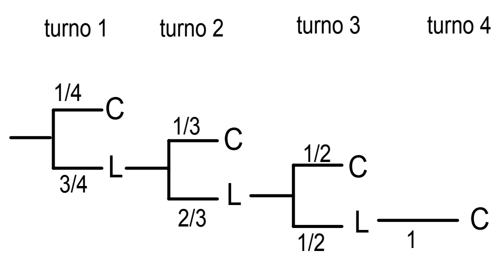
Notemos que para calcular $P(C_1 \mid E_3 \cap A_2)$ utilizamos el teorema de Bayes y el diagrama superior para encontrar las probabilidades, obteniendo:

$$
\begin{aligned}
&P(C_1 \mid E_3 \cap A_2) \\
&\quad = \frac{P(E_3 \cap A_2 \mid C_1)P(C_1)}{P(E_3 \cap A_2 \mid C_1)P(C_1) + P(E_3 \cap A_2 \mid C_2)P(C_2) + P(E_3 \cap A_2 \mid C_3)P(C_3)} \\
&\quad = \frac{\frac{1}{3} \times \frac{1}{3} \times 1}{\left(\frac{1}{3} \times \frac{1}{3} \times 1\right) + \left(\frac{1}{3} \times \frac{1}{3} \times 0\right) + \left(\frac{1}{3} \times \frac{1}{3} \times \frac{1}{2}\right)} = \frac{2}{3}.
\end{aligned}
$$

De manera que $P(C_3 \mid E_3 \cap A_2) = \frac{1}{3}$ y lo óptimo es *catafixear* y cambiar nuestra elección a la puerta 1.

La intuición es la siguiente: la probabilidad de elegir una puerta vacía inicialmente es de 2/3. Cuando esto sucede, el animador siempre nos mostrará la otra puerta vacía y por lo tanto la puerta restante será la que contiene al premio. Es decir, con 2/3 de probabilidad la puerta que no ha sido elegida ni por el concursante ni por el animador es la correcta. De esta forma, cambiar de puerta es la estrategia óptima a seguir.

Ej 4.4.6 (Sacando palillos) El siguiente es un mecanismo conocido para elegir a una persona de un grupo pequeño. Digamos que en un grupo de 4 personas quiere seleccionarse a una de ellas para que realice alguna actividad, digamos ir por cervezas a la hielera. Para este efecto, se toman 4 palillos y uno de ellos se recorta; así, se tienen tres palillos largos y uno corto. Las personas toman turnos tomando palillos, ignorando cuál es el corto, y el que saca el palillo corto es el elegido para ir por las cervezas. ¿Tiene alguna ventaja escoger primero, segundo, tercero o cuarto? La mayor parte de las personas dirán que quieren elegir al principio, después de todo, en el cuarto turno sólo queda un palillo y no hay posibilidad de elección. El siguiente diagrama de árbol ilustra la situación:



Si $P(C,i)$ representa la probabilidad de elegir el palillo corto en el $i-\acute{e}simo$ turno, el diagrama indica queLas probabilidades de elegir el palillo corto en los diferentes tunos son

$$
\begin{aligned}
&P(C,1) = \frac{1}{4},\\
&P(C,2) = P(C \mid L)P(L) = \frac{1}{3} \times \frac{3}{4} = \frac{1}{4},\\
&P(C,3) = P(C \mid L \cap L)P(L \cap L) = \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} = \frac{1}{4},\\
&P(C,4) = P(C \mid L \cap L \cap L)P(L \cap L \cap L) = 1 \times \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} = \frac{1}{4}.
\end{aligned}
$$

Es decir, ¡la probabilidad de obtener el palillo corto es $\frac{1}{4}$ en cualquier turno!

♠♠♠♠♠

4.5. Aprendizaje *Bayesiano*

La toma de decisiones casi siempre sucede dentro de un entorno de incertidumbre. Así, tenemos que tomar decisiones inciertas con información imperfecta. En general, tenemos varias hipótesis y comparamos la evidencia a favor o en contra de cada una de ellas para tomar la decisión más adecuada. Estas hipótesis no provienen de experimentos como lanzar monedas o de observaciones repetidas, sino de nuestra experiencia. En particular, en la jurisprudencia, la mayor parte de las hipótesis son de este estilo.

Cada una de estas hipótesis tiene alguna probabilidad subjetiva, asignada por nosotros, de ser verdadera. De acuerdo a la definición 2.5.1, podemos pensar en ésta como el grado de creencia o certidumbre que tenemos en la hipótesis. Esta es la probabilidad *a priori* y tiene que existir dado que el aprendizaje debe tener una base, no podemos empezar a aprender de la nada. El editor del trabajo de Bayes añadió un apéndice para tratar de explicar este procedimiento de aprendizaje con un ejemplo:

Imaginemos una persona que aparece repentinamente en este mundo, sin ningún conocimiento previo, y es abandonada en medio de la nada para observarlo. Probablemente el sol sería lo primero que llamaría su atención. Sorpresivamente, la primera noche el sol desaparece. ¿Volverá a aparecer? se pregunta. En ese momento este individuo se propone calcular la probabilidad de volver a ver el sol.

Esa primera noche piensa que la probabilidad de que el sol vuelva a salir no es muy buena, a esta primera probabilidad se le denomina la probabilidad *a priori*. Al día siguiente el sol aparece al amanecer y el solitario individuo actualiza esta probabilidad con la nueva evidencia, obteniendo una probabilidad *a posteriori*. Ésta es el nuevo punto de partida para la probabilidad de que el sol regrese.

Así, pasan los días y diariamente el sol aparece. Los momios en favor de que el sol vuelva a salir van incrementándose con cada amanecer. Quizás inicialmente se asignó una probabilidad del 50 % de que el sol volviese a salir y ésta aumentó al 65 % después del tercer día y al 75 % después del quinto día. Cada día, al experimentar el amanecer, se actualiza la probabilidad y con el transcurso del tiempo el individuo elimina la incertidumbre. ¿Cuándo debe dejar de actualizar sus probabilidades? No hay una regla fija para esto, deja de actualizar cuando cree que ya no es necesario hacerlo.

En el aprendizaje *Bayesiano*, cada vez que observamos evidencia nueva, actualizamos la probabilidad de que las hipótesis iniciales sean verdaderas. Si denotamos por H_i a la i-ésima hipótesis y por E a la nueva evidencia, entonces la actualización se realiza mediante el cálculo de la probabilidad *a posteriori*

$$P(H_i \mid E),$$

utilizando el teorema de Bayes. Este proceso se continúa hasta tener un grado de creencia adecuado en alguna de las hipótesis iniciales. Veamos como funciona este proceso con más detalle.

Un médico recibe a un paciente que dice tener apendicitis[3]. Sin saber aún nada, el médico asigna una probabilidad $P(H_1) = \frac{1}{1000}$ =0.001 de que tenga apendicitis. El paciente reporta haber tenido fiebre, dolor de cabeza y vómito, llamemos a este conjunto de síntomas la evidencia E_1. La probabilidad de tener estos síntomas dado que se tiene apendicitis es

$$P(E_1 \mid H_1) = 0.8.$$

El paciente también reporta haber comido un pastel con crema el día anterior, de manera que el médico considera también la posibilidad de que la crema haya estado descompuesta y haya contraído *E. Choli.* Llamemos a esta hipótesis H_2. Por simplicidad, asúmase que estas hipótesis son las únicas dos posibilidades de manera que

$$P(H_2) = 1 - P(H_1).$$

La probabilidad de presentar los síntomas descritos arriba, dado que se tiene *E. Choli* es de

$$P(E_1 \mid H_2) = 0.9.$$

[3]Adaptado de un ejemplo de Hacking Ian, *Probability and Inductive Logic,* Cambridge University Press, New York, NY, 2001.

Con esta información, el médico estima la probabilidad *a posteriori* de que el paciente tenga apendicitis dados los síntomas. Por el teorema de Bayes ésta es,

$$\begin{aligned} P(H_1 \mid E_1) &= \frac{P(E_1 \mid H_1)P(H_1)}{P(E_1 \mid H_1)P(H_1) + P(E_1 \mid H_2)P(H_2)} \\ &= \frac{0.8 \times 0.001}{0.8 \times 0.001 + 0.9 \times (1 - 0.001)} = 0.00089, \end{aligned}$$

demasiado baja para internar al paciente en el hospital.

Al día siguiente el paciente se queja de dolor en el costado derecho, llamamos a este nuevo síntoma E_2. El médico estima que las probabilidades de tener este dolor y los síntomas anteriores dado alguno de los diagnósticos posibles es,

$$\begin{aligned} P(E_1 \cap E_2 \mid H_1) &= 0.7, \\ P(E_1 \cap E_2 \mid H_2) &= 0.0002. \end{aligned}$$

La probabilidad *a posteriori* de que el paciente tenga apendicitis dada toda esta evidencia es de,

$$\begin{aligned} P(H_1 \mid E_1 \cap E_2) &= \frac{P(E_1 \cap E_2 \mid H_1)P(H_1)}{P(E_1 \cap E_2 \mid H_1)P(H_1) + P(E_1 \cap E_2 \mid H_2)P(H_2)} \\ &= \frac{0.7 \times 0.00\,1}{0.7 \times 0.001 + 0.002 \times (1 - 0.001)} = 0.78, \end{aligned}$$

con lo cual el médico interna al paciente en el hospital.

Evidentemente el ejemplo anterior está sumamente estilizado y simplificado, no obstante, nos da una idea de lo que es el proceso de aprendizaje *Bayesiano*. Dado un conjunto de hipótesis $\mathbf{H} = \{H_1, H_2, \ldots, H_n\}$, exhaustivas y mutuamente exclusivas, la cantidad $P(E \mid H_i)$ indica que tan probable es la evidencia observada (E) si la hipótesis H_i fuese cierta y se conoce como **verosimilitud** de H_i. La regla de correspondencia que asigna a cada hipótesis del conjunto $\mathbf{H}$ el valor $P(E \mid H_i)$, se denomina **función de verosimilitud**[4]. En la sección 10.8, retomaremos estas probabilidades y veremos su relación con las pruebas de hipótesis.

[4]Explícitamente, una vez obtenida la evidencia E, si F_E es la función de verosimilitud, ésta se calcula como $F_E(H_i) = P(E \mid H_i)$ para toda H_i en el conjunto $\mathbf{H}$.

Desgraciadamente, la mayor parte de la evidencia real también contiene incertidumbre; así, podría ser que el paciente descrito arriba sea hipocondriaco y su credibilidad, al describir los síntomas que le aquejan, no sea muy alta. Existe, sin embargo, una generalización para estos casos propuesta por Richard Jeffrey (1926-2002), en la cual se da una revisión de la fórmula de Bayes cuando la evidencia es incierta. Formalizar el proceso de aprendizaje y la toma de decisiones es una tarea sumamente compleja pero afortunadamente el enfoque *Bayesiano* nos ayuda a esclarecer un poco las cosas.

Ejemplo

Ej 4.5.1 [5] Supongamos que les digo que mi hermano tiene dos descendientes, sin especificar el género. Las posibles combinaciones para sus descendientes son:

$$\begin{aligned} A &: \text{dos hombres,} \\ B &: \text{dos mujeres,} \\ C &: \text{ un hombre y una mujer.} \end{aligned}$$

Si h y m denotan hombre y mujer respectivamente, el espacio muestral de las posibles combinaciones de los descendientes es

$$\{hh, mm, hm, mh\}$$

De aquí, las probabilidades *a priori* son

$$\begin{aligned} P(A) &= \frac{1}{4}, \\ P(B) &= \frac{1}{4}, \\ P(C) &= \frac{1}{2}. \end{aligned}$$

Si mi hermano les informa que al menos uno de sus descendientes es mujer, ¿cuál es la probabilidad *a posteriori* de que su otro descendiente sea hombre? Seguramente pensarán que es $\frac{1}{2}$, pero recordando el problema de "La Catafixia" deben pensarlo con más cuidado, veamos:

[5]Tomado de Devlin Keith, "The Unfinished Game", Basic Books, NY, 2008.

El espacio muestral reducido es ahora

$$\{mm, hm, mh\},$$

de manera que la probabilidad de que el otro descendiente sea hombre es $\frac{2}{3}$ y no $\frac{1}{2}$. Este ejemplo nos puede parecer trivial, sin embargo, ilustra que nuestra intuición puede ser errónea cuando actualizamos nuestras creencias con nueva información o evidencia.

♠♠♠♠♠

Bibliografía Complementaria

1. Finkelstein Michael O. and Fairley William B., "A Bayesian Approach to Identification Evidence", *Harvard Law Review,* Vol 83, No 3, Enero 1970, pp. 489-517

2. Devlin Keith, *The Unfinished Game*, Basic Books, New York, NY, 2008.

3. Gill Jeff, *Bayesian Methods: a Social and Behavioral Sciences Approach*, Chapman and Hall, Boca Raton, FL, 2002.

4. Hacking Ian, *Probability and Inductive Logic*, Cambridge University Press, New York, NY, 2001.

5. Hively Will, "The Mathematics of Making up your Mind (Bayesian Theorem)", *Discover*, Mayo 1996, publicado en línea en:

   http://discovermagazine.com/1996/may/themathematicsof760.

6. Jeffrey Richard, *Subjective Probability: the Real Thing*, Cambridge University Press, New York, NY, 2004.

Ejercicios

▷ **4.1** Se lanzan dos dados. ¿Cuál es la probabilidad de que la suma sea 4 si al menos uno de los números es 3?

▷ **4.2** Se realizó una encuesta entre 100 personas, hombres y mujeres, para determinar si fumaban o no. Los resultados se muestran en la siguiente tabla

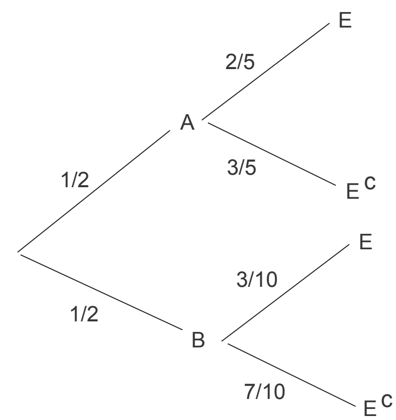
	fuma	no fuma	Total
hombre	15	25	40
mujer	10	50	60
Total	25	75	100

Encontrar,

1. $P(\text{mujer}|\text{no fuma})$,
2. $P(\text{fuma}|\text{hombre})$,
3. $P(\text{mujer y fuma})$,
4. $P(\text{fumar})$.

▷ **4.3** Realiza un diagrama de árbol para el problema 4.2.

▷ **4.4**



Utilizar el diagrama de arriba para responder las siguientes:

1. $P(A)$
2. $P(E \mid A)$
3. $P(E)$
4. $P(A \cap E)$
5. ¿Son A y E independientes?

▷ **4.5** Se lanzan dos dados, uno rojo y uno verde. Sea E el evento "el número en el dado verde no es ni uno ni dos" y F el evento "la suma de los números es igual a seis". Determinar si estos eventos son independientes.

▷ **4.6** Se lanzan dos dados. ¿Cuál es la probabilidad de que la suma de los números sea mayor a 9 si uno de los números es 6?

▷ **4.7** Sea E el evento de que el número 111111 sea ganador en el sorteo Melate y F el evento de que el número 555555 sea el ganador.

1. Determina si estos eventos son mutuamente excluyentes.
2. Determina si estos eventos son independientes.

▷ **4.8** [6]El 60 % de las papayas que se producen en México proviene de Veracruz y el 40 % proviene del estado de Guerrero. Desafortunadamente, se han encontrado tarántulas en el 2 % de los contenedores de papaya de Veracruz y en el 1 % de los contenedores provenientes de Guerrero. Suponga que una tarántula fue encontrada en un contenedor seleccionado al azar. ¿Cuál es la probabilidad de que este contenedor haya provenido de Veracruz?

▷ **4.9** Con los mismos supuestos del problema 4.4.3, calcula la probabilidad de que el atleta haya utilizado esteroides dado que se le hicieron cinco pruebas (condicionalmente) independientes obteniendo los resultados: $+$, $-$, $+$, $-$ y $+$.

▷ **4.10** Tomemos el mismo escenario y datos del ejemplo 4.4.4 pero en lugar de que el sospechoso sea un individuo albino ahora se tiene:

1. El sospechoso del crimen es un individuo pelirrojo y la probabilidad de ser pelirrojo es del 2 %. Encontrar $P(pelirrojo \mid T_p)$, en donde T_p representa: "el testigo identifica a una persona como pelirroja".
2. El sospechoso es un individuo de pelo rubio y la probabilidad de tener el pelo rubio es del 15 %. Encontrar $P(rubio \mid T_r)$, en donde T_r representa: "el testigo identifica a una persona como rubia".

▷ **4.11** En México se estima que aproximadamente el 1 % de la población sexualmente activa es portador del virus del SIDA. ¿Sería sensato proponer como política de salud pública que se le realizara

[6]Adaptado de un ejemplo de Hacking Ian, "Probability and Inductive Logic", Cambridge University Press, New York, NY, 2001.

la prueba del SIDA a todos los adultos sexualmente activos? Esta prueba tiene un 5 % de falsos positivos y negativos. Fundamentar la respuesta.

▷ **4.12** Los datos de la PGR indican que de las personas arrestadas por narcotráfico en 2007, el 90 % son hombres y el 10 % mujeres. Adicionalmente, el 10 % de los hombres y el 5 % de las mujeres eran menores de edad, respectivamente.

1. ¿Cuál es la probabilidad de que alguien arrestado por narcotráfico en 2007 haya sido menor de edad?
2. Si una persona arrestada por narcotráfico en 2007 era menor de edad, ¿cuál es la probabilidad de que haya sido mujer?

▷ **4.13** Supongamos que la probabilidad de una recesión económica es del 75 % y la de una expansión es del 25 %. Dada una recesión, un aumento en el desempleo suele observarse en el 90 % de los casos. Asimismo, dada una expansión económica el desempleo aumenta sólo con un 5 % de probabilidad. Dado que se observa un aumento en el desempleo, ¿cuál es la probabilidad de que la economía esté en recesión?

▷ **4.14 (Aprendizaje Bayesiano)** [7]Imagina que tus notas del curso de Comprensión Matemática se perdieron. Tienes dos hipótesis: H_1 : "se quedaron en la biblioteca" y H_2 : "se quedaron en el salón de clase".

1. Asignar probabilidades (sensatas) a priori para estas hipótesis.
2. Una amiga reporta haber visto unas notas en la biblioteca pero podrían ser de otra persona. Llamemos a esta información la evidencia E. Asignar valores (sensatos) para la función de verosimilitud: $P(E \mid H_1)$ y $P(E \mid H_2)$.
3. Calcular las probabilidades a posteriori $P(H_1 \mid E)$ y $P(H_2 \mid E)$.

[7]Adaptado de un ejercicio de "An Introduction to Probability and Inductive Logic", Ian Hacking, Cambridge University Press, 2001.

4. Otra amiga dice que vio algo, que parecía un cuaderno, con tu nombre en la biblioteca. Denotemos por F a esta evidencia. Asigna valores (sensatos) para las probabilidades $P(E \cap F \mid H_1)$ y $P(E \cap F \mid H_2)$.

5. Calcular las probabilidades a posteriori $P(H_1 \mid E \cap F)$ y $P(H_2 \mid E \cap F)$ y decide a donde vas a ir a buscar tus notas.

Capítulo 5

Cuantificando nuestras esperanzas

5.1. Introducción

¿Llevaré un paraguas?, ¿estudiaré para el examen?, ¿invertiré mi dinero en la bolsa?, ¿me casaré? Cada vez que nos hacemos una pregunta como las anteriores debemos tomar una decisión. Ésta puede ser tan simple como llevar o no un paraguas o bien tan compleja como decidir compartir nuestra vida con otra persona. Asumiendo que somos racionales, nuestra decisión la tomaremos en base a evaluar las consecuencias que se derivan de la misma.

Lo ideal sería asignar algún número al *valor* que las consecuencias tienen para nosotros. Por ejemplo, si llevo el paraguas y llueve, esto me proporciona una satisfacción de 5 puntos. Si no llevo el paraguas y llueve, entonces el mojarse me causa un disgusto de -6 puntos. Si llevo paraguas y no llueve, la inconveniencia me causa -1 punto y si no llevo paraguas y no llueve, entonces mi agrado vale 4 puntos. Si adicionalmente, el meteorólogo nos dice que la probabilidad de que llueva es del 45 %, nuestra decisión también tomaría esto en cuenta.

Pensemos ahora en una versión simplificada del sorteo Melate, en la cual sólo hay una combinación ganadora y un premio[1]. En el ejemplo 3.2.5 vimos que había 32468000 combinaciones posibles de números. La probabilidad de ganar la calculamos como $\frac{1}{32468000}$. El precio de cada combinación de seis números es de \$15. Si hay N millones de pesos en la bolsa Melate, entonces podemos ganar N millones con probabilidad $\frac{1}{32468000}$.

Comparemos el valor
$$N \times \frac{1}{32468000},$$
que representa el premio ponderado por la probabilidad de ganarlo o *premio esperado*, con \$15, que es el costo de adquirir un boleto. Observamos que para $N = \$487\,020\,000$ ambos valores coinciden. Para entender mejor qué significa esto, simplifiquemos el sorteo suponiendo que el premio no se va a compartir. Si tuviésemos \$487 020 000 pesos, podríamos comprar todas y cada una de las combinaciones de Melate, ya que

$$15 \times 32468000 = 487\,020\,000.$$

Si la bolsa fuese mayor a esta cantidad, entonces obtendríamos una ganancia segura al comprar todas las combinaciones. Claro está que en la práctica, más de una persona puede tener la combinación ganadora y la probabilidad de ganar todo el premio es aún menor, de manera que el valor esperado sería negativo. Adicionalmente, los pocos afortunados que cuentan con dicha cantidad de dinero, ¡seguramente tienen oportunidades menos riesgosas de inversión!

5.2. Variables aleatorias

Si una variable, digamos X, puede tomar valores numéricos, cada uno con cierta probabilidad, decimos que se trata de una **variable aleatoria**. El mejor ejemplo es pensar en X como los premios o pérdidas asociados a un evento aleatorio como podría ser el lanzar una moneda o girar una ruleta. Formalmente, una variable aleatoria es una regla que asigna un valor

[1]En realidad la bolsa acumulada se reparte entre el número de ganadores y varios premios menores, según el número de cifras acertadas entre dos y seis.

numérico a los elementos de un espacio muestral. Si X es una variable aleatoria y r es algún resultado de un espacio muestral, denotamos por $X(r)$ al valor numérico asociado a este resultado.

Supongamos que X es una variable aleatoria tal que dado un conjunto de resultados inciertos $\{r_1, ..., r_n\}$, ésta toma valores $X_1 = X(r_1), ..., X_n = X(r_n)$. Notemos que éstos no necesariamente son distintos. La probabilidad de que la variable aleatoria tome el valor X_i se denota por,

$$P(X = X_i).$$

A la regla de correspondencia que asigna a cada valor X_i su probabilidad $P(X = X_i)$, se le llama la **distribución de probabilidad** de la variable aleatoria. Si denotamos a esta regla por f tenemos que

$$f(X_i) = P(X = X_i).$$

Ejemplos

Ej 5.2.1 Vamos a utilizar la información del párrafo introductorio acerca de los valores obtenidos cuando tomamos la decisión de llevar o no paraguas. Pensemos en "llevar un paraguas" y "no llevar un paraguas" como dos variables aleatorias X y $\sim X$. El evento aleatorio es la lluvia. Podemos representar estas variables aleatorias como las siguientes funciones o correspondencias:

$$X(lluvia) = 5, \; X(\text{no lluvia}) = -1,$$
$$\sim X(lluvia) = -6, \;\; \sim X(\text{no lluvia}) = 4.$$

Aquí, las distribuciones de probabilidad, digamos f y g, son

$$f(5) = P(X = 5) = 0.45,$$
$$f(-1) = P(X = -1) = 0.55,$$
$$g(-6) = P(\sim X = -6) = 0.45,$$
$$g(4) = P(\sim X = 4) = 0.55.$$

Ej 5.2.2 Supongamos que un volado ofrece una ganancia de \$100 si sale águila y una pérdida de \$60 si se sale sol. La variable aleatoria X es la función dada por: $X(\acute{a}guila) = 100$, $X(sol) = -60$. La distribución de probabilidad f es,

$$\begin{aligned} f(100) &= P(X = 100) = \frac{1}{2}, \\ f(-60) &= P(X = -60) = \frac{1}{2}. \end{aligned}$$

Ej 5.2.3 Pensemos en una urna con 10 canicas rojas, 2 amarillas y 3 azules. Se nos ofrecen los siguientes premios: perder \$100 si sacamos una canica roja, ganar \$150 si sacamos un canica azul o amarilla. La variable aleatoria Y es $Y(roja) = -100$, $Y(azul) = Y(amarilla) = 150$. La distribución de probabilidad f es,

$$\begin{aligned} f(-100) &= P(Y = -100) = \frac{10}{15} = \frac{2}{3}, \\ f(150) &= P(Y = 150) = \frac{2}{15} + \frac{3}{15} = \frac{1}{3}. \end{aligned}$$

Ej 5.2.4 Se lanzan dos monedas. La variable aleatoria X es el número de águilas que aparecen. Si denotamos por a y s, respectivamente, a obtener águila o sol, el espacio muestral es,

$$\{(a, a), (a, s), (s, a), (s, s)\}.$$

de aquí,

$$\begin{aligned} X((a, a)) &= 2, \\ X((a, s)) &= 1, \\ X((s, a)) &= 1, \\ X((s, s)) &= 0. \end{aligned}$$

La distribución de probabilidad asociada f es,

$$\begin{aligned} f(0) &= P(X = 0) = \frac{1}{4}, \\ f(2) &= P(X = 2) = \frac{1}{4}, \\ f(1) &= P(X = 1) = \frac{1}{2}. \end{aligned}$$

Ej 5.2.5 La variable aleatoria asociada al sorteo Melate para una bolsa acumulada de \$N, es

$$X = \begin{cases} \$N, & \text{para la combinación ganadora,} \\ \$0, & \text{en cualquier otro caso.} \end{cases}$$

De acuerdo a lo visto en la introducción de esta sección, la distribución de probabilidad es

$$f(N) = \frac{1}{32468000},$$
$$f(0) = 1 - \frac{1}{32468000}.$$

de esta forma, el valor esperado del premio es lo que habíamos calculado como

$$E(X) = N\frac{1}{32468000}.$$

Vimos que éste es igual a \$15 si el premio es de

$$N = \$487\,020\,000.$$

Podemos interpretar ésto como que el precio *justo* de un boleto para participar en el sorteo es de \$15 si el premio es de \$487 020 000.

Ej 5.2.6 Se quiere hacer una rifa para recaudar fondos para alguna buena causa. Hay un premio único de \$10 000 y los boletos de la rifa cuestan \$100. Sea n el número de boletos que se venden, de manera que la probabilidad de ganar el premio es de $\frac{1}{n}$. De esta forma, el comprador de un boleto tiene como valor esperado para esta rifa,

$$\frac{1}{n} \times (10000 - 100) + \left(1 - \frac{1}{n}\right)(-100)$$

y el organizador de la misma tiene la ganancia segura de,

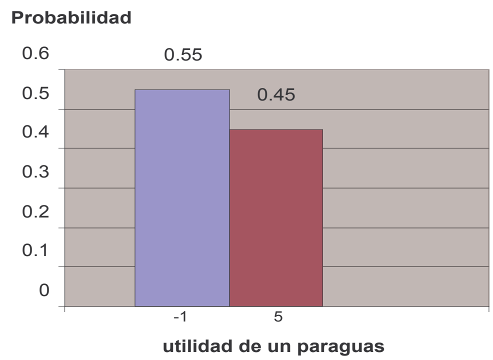
$$(100 \times n) - 10000.$$

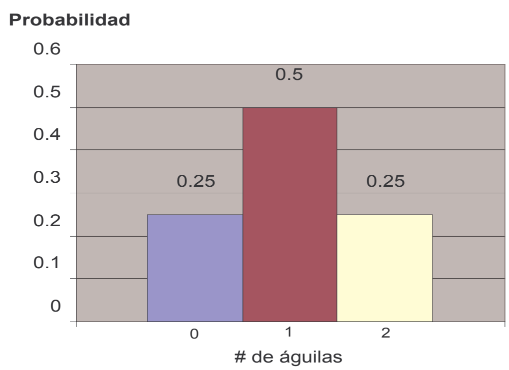
Evidentemente la rifa no debe realizarse si no se venden al menos 100 boletos, ¡ya que el organizador perdería dinero! La siguiente tabla muestra el valor esperado de la rifa para el comprador del boleto y la ganancia del organizador para algunos valores de n.

n	Valor Esperado (comprador)	Ganacia (organizador)
100	0	0
150	$-33.\bar{3}$	5000
200	-50	10000
500	-80	40000

♠♠♠♠♠

La distribución de probabilidad puede representarse visualmente en un diagrama al que llamamos **histograma de probabilidad.** Un histograma es simplemente una gráfica de barras. Para este caso, en el eje horizontal se representan los valores que puede tomar la variable aleatoria y en el eje vertical la probabilidad asociada a cada valor. Por ejemplo, para los casos descritos en los ejemplos 5.2.1 y 5.2.4 tenemos,





Vamos a asumir que la base de los rectángulos, por convención mide una unidad. De esta forma, observemos que la suma de las áreas de los rectángulos siempre es 1 pues ésta es simplemente la suma de las probabilidades de todos los valores posibles que toma la variable aleatoria, concretamente,

$$f(X_1) + \cdots + f(X_n) = \sum_{i=1}^{n} f(X_i) = 1. \tag{5.1}$$

5.3. Valor esperado

Supongamos que la variable aleatoria X toma valores $X_1, ..., X_n$ y sean $f(X_1), \ldots, f(X_n)$ los valores de la distribución de probabilidad asociada.

Definimos el **valor esperado o esperanza** de X como

$$E(X) = \sum_{i=1}^{n} X_i P(X = X_i) = \sum_{i=1}^{n} X_i f(X_i).$$

Éste nos da una idea del valor promedio de la variable aleatoria X.

Notemos que si todos los valores son equiprobables, de manera que $P(X = X_1) = \cdots = P(X = X_n) = \frac{1}{n}$, el valor esperado de X es

$$E(X) = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + \cdots X_n}{n},$$

es decir, lo que conocemos como el promedio de los valores $X_1, ..., X_n$. Así, el valor esperado no es más que un promedio ponderado por la probabilidad de obtener X_i como resultado.

Ejemplos

Ej 5.3.1 En el ejemplo 5.2.1 los valores esperados de $X =$"llevar paraguas" y de $\sim X =$"no llevar paraguas" son los siguientes:

$$E(X) = 0.45 \times 5 + 0.55 \times (-1) = 1.7,$$
$$E(\sim X) = 0.45 \times (-6) + 0.55 \times 4 = -0.5,$$

Interpretemos esta esperanza como el valor que nos proporcionan las acciones de llevar o no el paraguas. De acuerdo a los valores obtenidos, $E(X) > E(\sim X)$, equivalentemente, la acción de "llevar el paraguas" nos lleva a una consecuencia con mayor valor esperado. Si somos racionales, elegiremos llevar el paraguas. Más adelante estudiaremos este método de toma de decisiones con mayor detalle.

Ej 5.3.2 Para los ejemplos 5.2.2 y 5.2.3 tenemos que los valores esperados son,

$$E(X) = \frac{1}{2}(100 - 60) = \$20,$$
$$E(Y) = -100 \times \frac{2}{3} + 150 \times \frac{1}{3} = -\$16.667.$$

¿Si nos ofrecieran jugar uno de estos dos juegos cuál deberíamos elegir? El primero, ya que su valor esperado es mayor. Notemos que ninguno de los premios o pérdidas es igual al valor esperado, éste simplemente nos da una idea del promedio de los premios y pérdidas que resultarían si el juego se repitiese un gran número de veces.

Ej 5.3.3 Los juegos de azar en los casinos pueden representarse por medio de variables aleatorias. Consideremos una versión simplificada de una máquina de palanca: existen dos figuras, digamos un cuadrado y un triángulo, que pueden aparecer en tres columnas de una cinta circular. El jugador baja la palanca, las cintas giran y al detenerse cada columna muestra un cuadrado o un triángulo. Las 8 posibilidades equiprobables son:

$$\begin{array}{ccc}
\blacksquare & \blacksquare & \blacksquare \\
\blacksquare & \blacksquare & \blacktriangle \\
\blacksquare & \blacktriangle & \blacksquare \\
\blacksquare & \blacktriangle & \blacktriangle \\
\blacktriangle & \blacksquare & \blacksquare \\
\blacktriangle & \blacksquare & \blacktriangle \\
\blacktriangle & \blacktriangle & \blacksquare \\
\blacktriangle & \blacktriangle & \blacktriangle
\end{array}$$

Supongamos que para jugar se introduce una ficha de \$300 y los premios asociados son: \$500 si salen tres figuras iguales y \$0 de cualquier otra forma. Dos de las posibles consecuencias tienen tres figuras iguales de manera que la probabilidad de que salgan tres figuras iguales es $\frac{2}{8} = \frac{1}{4}$. La variable aleatoria asociada X tiene como premio $500 - 300 = 200$ con probabilidad $\frac{1}{4}$, y -300 con probabilidad $\frac{3}{4}$. El valor esperado es,

$$E(X) = (200)\frac{1}{4} - (300)\frac{3}{4} = -175.$$

Otra forma de calcularlo sería pensar que se ganan 500 con probabilidad $\frac{1}{4}$

y el costo seguro es de 300, entonces

$$E(X) = 500 \times \frac{1}{4} - 300 = -175.$$

Esto implica que un jugador puede esperar perder \$175 si juega un gran número de veces. Desde el punto de vista del casino, éste espera ganar \$175. Todos los juegos de los casinos tienen un valor esperado negativo para el jugador de manera que la casa, al *final del día*, siempre gana.

Ej 5.3.4 Recordemos que para el primer juego de dados de Chevalier de Méré se tiene

$$P(\text{obtener al menos un seis en cuatro lanzamientos}) = 0.51775,$$
$$P(\text{no obtener ningún seis en cuatro lanzamientos}) = 0.482\,25.$$

Si un jugador apuesta 100 francos, su **ganancia esperada** es el valor esperado de la variable aleatoria asociada, o bien,

$$0.48225 \times 100 + 0.51775 \times (-100) = -3.55$$

y análogamente, la ganancia esperada para la casa es,

$$0.48225 \times (-100) + 0.51775 \times (100) = 3.55.$$

Ej 5.3.5 Las variables económicas son, en su mayoría, variables aleatorias pues toman algún valor que depende de consecuencias inciertas. Por ejemplo, el crecimiento del PIB para el año entrante tomará valores dependientes de un gran número de eventos inciertos como el precio del petróleo, la tasa de interés en EUA, el índice de confianza del consumidor, la inflación, etcétera. Una forma sencilla de pensar al crecimiento del PIB como una variable aleatoria es reducir el conjunto de consecuencias a, digamos tres escenarios: malo, regular bueno, cada uno con cierta probabilidad de materializarse, digamos 0.75, 0.20 y 0.05, respectivamente. La variable aleatoria asociada al crecimiento del PIB podría ser algo así como $X(malo) = -1\,\%$, $X(regular) = 2\,\%$ y $X(bueno) = 7\,\%$. El valor esperado del crecimiento PIB sería,

$$-1 \times 0.75 + 2 \times 0.2 + 7 \times 0.05 = 0\,\%.$$

♠♠♠♠♠

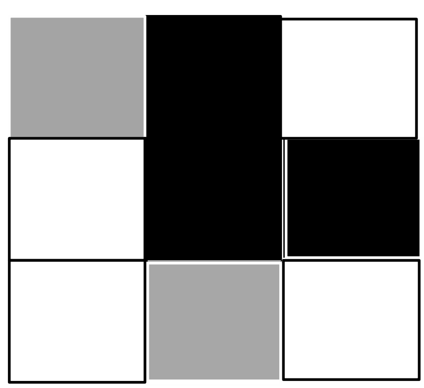
5.4. El caso continuo: midiendo resultados

Pensemos en el siguiente juego de azar: se toma el intervalo $[0,1]$ de la recta numérica, es decir, todos los números (reales) entre el 0 y el 1 y se elige aleatoriamente un número del intervalo. Podemos pensar a este juego como una *ruleta continua* y asociarle una variable aletoria X. La probabilidad de que X tome algún valor específico en el intervalo, digamos 0.322785038 no tiene mucho sentido, sin embargo, podemos pensar en la probabilidad de que X tome un valor entre 0.3 y 0.33 y denotamos a ésta por

$$P(0.3 < X < 0.33).$$

Las variables aleatorias que hemos estudiado pueden tomar solamente un número discreto de valores, es decir, representan datos que pueden contarse. En la vida real no todo es contable pues existen datos de naturaleza continua que más bien se miden. Por ejemplo, distancias, pesos, tiempos, etcétera. En lugar de utilizar técnicas de conteo como en la sección 3, ahora se utilizan técnicas de medida. Esto no es enteramente trivial de manera que aquí nos restringiremos a ver algunos casos sencillos que involucran mediciones geométricas.

Imaginemos un juego de dardos, en el cual el tirador es malísimo y los lanza sin ton ni son. El tablero es un cuadro formado por nueve cuadrados más pequeños, como se muestra a continuación:



La probabilidad de que el dardo caiga en algún color puede calcularse como la proporción del área que dicho color representa como parte del área total. Así,

$$P(blanco) = \frac{4}{9},$$
$$P(negro) = \frac{3}{9},$$
$$P(gris) = \frac{2}{9}.$$

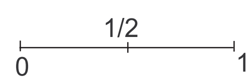
Notemos que la probabilidad de que el dardo caiga en un punto particular sería nula pues un punto carece de área. La solución sería pensar en una pequeña superficie alrededor del punto y calcular su área.

Ejemplos

Ej 5.4.1 Pensemos en la *ruleta continua* descrita al inicio de esta sección y sea X la variable aleatoria que representa al número obtenido. Todos los números son equiprobables de manera que la probabilidad de que X caiga en el intervalo $[0, \frac{1}{2}]$ es,

$$P(0 \leq X \leq \frac{1}{2}) = \frac{1}{2}.$$

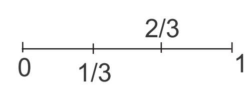
La razón es que $[0, \frac{1}{2}]$ mide la mitad del intervalo $[0, 1]$, en donde medida se refiere a la longitud (ver figura).



Asimismo,

$$P(\frac{1}{3} < X < 1) = \frac{2}{3},$$

ya que la longitud del intervalo $(\frac{1}{3}, 1)$ es $\frac{2}{3}$ partes la del intervalo $[0, 1]$ (ver figura).



Ej 5.4.2 Supongamos que ahora la ruleta continua consiste en elegir numeros en el intervalo $[0, 4]$ y X es la variable aleatoria que representa al número resultante. Entonces,

$$P(0 < X < 3) = \frac{3}{4}.$$

en donde podemos pensar a este número como

$$\frac{\textit{longitud} \text{ de } (0,3)}{\textit{longitud} \text{ de } [0,4]}.$$

♠♠♠♠♠♠

En los ejemplos 5.4.1 y 5.4.2, notemos que la medida o longitud del intervalo (a, b) es la misma que la del intervalo cerrado $[a, b]$. En general, la longitud de cualquiera de los intervalos[2] $(a, b), [a, b], (a, b]$ y $[a, b)$ es de $b-a$. Si pensamos en el intervalo original $[a, b]$ como nuestro espacio muestral y tomamos c y d tales que $a \leq c \leq d \leq b$, entonces, la probabilidad del evento "el número cae entre c y d" es,

$$P(c < X < d) = \frac{\textit{longitud} \text{ de } [c,d]}{\textit{longitud} \text{ de } [a,b]} = \frac{d-c}{b-a}.$$

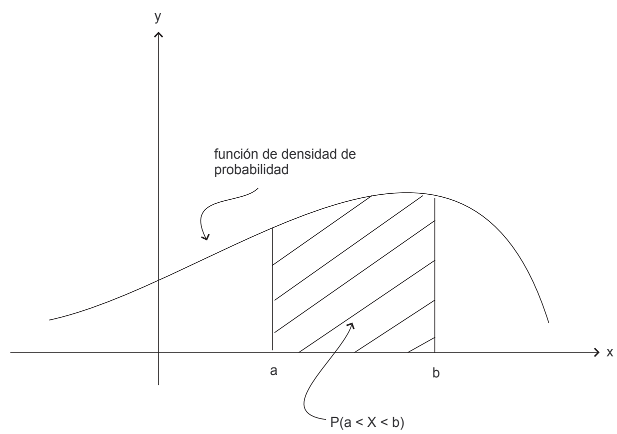
Al igual que en el tablero de colores para los dardos, aquí la probabilidad $P(X = X_i)$ también es nula pues un punto tampoco tiene longitud. En el caso continuo, la longitud o medida del intervalo toma el lugar del número de elementos en el caso discreto. Evidentemente, en el caso del tablero de dardos descrito arriba medimos áreas en lugar de medidas lineales de longitud.

Si X es una variable aleatoria continua, no podemos definir la distribución de probabilidad como en el caso discreto ya que no podemos hablar

[2]Formalmente,

$$\begin{aligned}
(a,b) &= \{x \in \mathbb{R} \text{ tales que } a < x < b\},\\
[a,b] &= \{x \in \mathbb{R} \text{ tales que } a \leq x \leq b\},\\
[a,b) &= \{x \in \mathbb{R} \text{ tales que } a \leq x < b\},\\
(a,b] &= \{x \in \mathbb{R} \text{ tales que } a < x \leq b\}.
\end{aligned}$$

de la probabilidad que X tome un valor específico, digamos X_i. Existe, sin embargo, una densidad de probabilidad, que es una curva tal que el área debajo de la curva, entre a y b, representa a la probabilidad de que X tome valores entre a y b, es decir, $P(a < X < b)$. Visualmente tenemos,

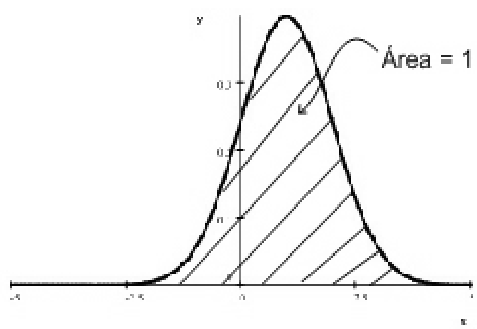


Como en el caso discreto, la curva es también el conjunto de puntos $(X, f(X))$ para la función de densidad f. La diferencia es que en el caso continuo, $f(a) \neq P(X = a)$.

En el caso discreto la suma de las probabilidades de todos los valores que toma la variable aleatoria es la unidad (ecuación 5.1). Aquí se tiene el resultado análogo: el área total bajo la función de densidad que representa la probabilidad de que la variable X tome cualquier valor (real), es la unidad, es decir,

$$P(-\infty < X < \infty) = 1. \tag{5.2}$$

Esto lo representamos en la siguiente figura:



El cálculo del valor esperado de una variable aleatoria continua no es enteramente trivial y es necesario utilizar el cálculo integral para evaluarlo. Sin embargo, la intuición sobre su significado es la misma y el valor esperado representa un promedio ponderado de los valores que toma la variable aleatoria. La notación que se utiliza es la misma que antes, es decir, si X es una variable aleatoria discreta o continua, su valor esperado se representa por $E(X)$.

Bibliografía Complementaria

1. Aguirre Víctor et al., *Fundamentos de Probabilidad y Estadística*, 2^a edición. Jit Press, México, D.F. 2006.

2. Haeussler Ernest F. Jr., Paul Richard S., Wood Richard J. *Matemáticas para Administración y Economía,* Pearson, 12^a ed. México D.F., 2008.

3. Warren Weaver, *Lady Luck: the Theory of Probability*, Science Study Series, Dover Publications, New York, NY, 1982.

Ejercicios

▷ **5.1** Determinar la variable aleatoria que resulta de los siguientes experimentos y obtener la distribución de probabilidad asociada:

1. El número de soles que ocurren al lanzar tres monedas al aire.
2. De un grupo de tres hombres y dos mujeres se elige un comité de dos personas al azar. La variable aleatoria es el número de hombres en el comité elegido.

▷ **5.2** Obtener el valor esperado para el jugador del juego de De Méré con 24 lanzamientos y con 25 lanzamientos. Suponer que la apuesta es de 100 francos.

▷ **5.3** Considérese el siguiente juego: se lanza un dado y se pagan \$100 si sale un 6, \$50 si sale un 5 y se pierden \$30 si sale cualquier otro número. Encontrar la variable aleatoria asociada a este juego, su distribución de probabilidad y calcular su valor esperado.

▷ **5.4** Un jardinero gana \$200 diarios si trabaja y pierde \$30 si no trabaja. La probabilidad de que tenga trabajo cualquier día dado es de $\frac{4}{7}$. Obtener el valor esperado de sus ganancias diarias.

▷ **5.5** Un paquete de cereal contiene un cupón que puede enviarse por correo registrado para participar en un sorteo. El premio es de una televisión de pantalla plana que tiene un valor de \$10000. Se espera que se reciban 40000 cupones y sólo uno será el ganador. El costo de enviar el cupón por correo registrado es de \$25. Comparar los valores esperados de los siguientes actos: “enviar el cupón” y “no enviar el cupón”.

▷ **5.6** Se escogen, al azar, dos números (distintos) del 1 al 5. Construir la variable aleatoria asociada a la suma de los dos números, encontrar su distribución de probabilidad y calcular su valor esperado.

▷ **5.7** Encontrar los histogramas de probabilidad para las distribuciones asociadas a las siguientes variables aleatorias:

1. El número de águilas que aparecen cuando lanzamos tres monedas al aire.

2. La suma de los números de las caras cuando se lanzan dos dados.

▷ **5.8** Se venden 8000 boletos para una rifa de \$5000.00 y cada boleto cuesta \$2.00.

1. Encontrar la ganancia esperada del comprador de un boleto.

2. ¿Cuál debería de ser el premio mínimo para que se pudiese garantizar "salir a mano" al comprar todos los boletos.

▷ **5.9** Determina el valor esperado de la siguiente variable aleatoria X: espacio muestra $= \{0, 1, 2, 3\}$ con resultados equiprobables y $X(0) = 0.1, X(1) = 0.4, X(2) = 0.2, X(3) = 0.3$.

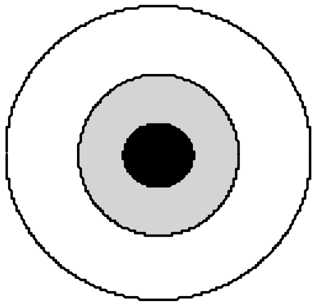
▷ **5.10** El precio del petróleo para el año entrante depende de toda una fauna de eventos aleatorios. Para reducir el problema se construyen 3 escenarios posibles: bueno, regular y malo y se estima que éstos serán los escenarios reales con probabilidades de 0.1, 0.3 y 0.6, respectivamente. El precio del petróleo en cada uno de estos escenarios se estima que será de \$25 dólares por barril en el escenario bueno, \$17 en el regular y \$10 en el malo. Encontrar el valor esperado del precio del petróleo para el año entrante.

▷ **5.11** La siguiente tabla obtenida por una compañía de renta de coches, nos da la probabilidad de que se renten n coches en un día

n	$prob(n)$
0	0.05
1	0.1
2	0.15
3	0.2
4	0.15
5	0.15
6	0.1
7	0.05
8	0.05

Determinar el valor esperado de la demanda diaria de carros.

▷ **5.12** Considera el siguiente "blanco", en donde los círculos tienen radio 1/2, 1 y 2, respectivamente.



Si se lanzan dardos de forma aleatoria, calcular la probabilidad de que el dardo caiga en la zona blanca, en la zona gris y en la zona negra.

▷ **5.13** Supongamos que X es la variable aleatoria que representa la elección de cualquier número en el intervalo $[0, 2]$ con igual probabilidad. Calcular,

1. $P(0 < X < 1)$,
2. $P(1 \leq X \leq \frac{4}{3})$,
3. $P(\frac{2}{3} < X \leq \frac{7}{6})$.

Capítulo 6

Valuando nuestras esperanzas

6.1. Introducción

La familia Bernoulli, de origen suizo, dio al mundo varios matemáticos y científicos destacados. Aquí vamos a concentrarnos únicamente en dos de ellos: Jacobo (1654-1705) y su sobrino Daniel (1700-1782). En la sección 9.3 hablaremos un poco del trabajo de Jacobo pero por el momento, nos concierne Daniel Bernoulli y algunos de sus estudios relacionados con el valor esperado.

Daniel realizó sus mejores aportaciones cuando era profesor de matemáticas en San Petersburgo y es, sin duda alguna, el padre de la teoría moderna de la decisión. En 1738 publicó un artículo seminal en relación a la medición del riesgo[1]. Este trabajo es, quizás, el más profundo que se ha escrito en relación al comportamiento humano ante la incertidumbre. En él, se toma como punto de partida el siguiente problema conocido como la **Paradoja de San Petersburgo.**

Paradoja de San Petersburgo

[1] Éste fue vuelto a publicar como: "Exposition of a New Theory on the Measurement of Risk", Econometrica vol 22 (1954), pp23-36.

> *Considérese el siguiente juego: se lanza una moneda repetidamente y en el momento que aparece la primer águila se termina el juego. Si ésta cae en el primer lanzamiento, ganamos 2 pesos, si cae en el segundo ganamos 4 pesos, si aparece hasta el tercero ganamos 8 pesos y así sucesivamente. En resumen, si la primer águila ocurre en el lanzamiento n, entonces ganamos 2^n pesos.*

Si nos presentan este juego, ¿cuánto estaríamos dispuestos a pagar por participar en él? De acuerdo a la sección 5.3, podríamos pensar que algo sensato sería su valor esperado. Pero antes de comprometernos, calculémoslo.

La probabilidad de que la primer águila aparezca en el $n-\textit{ésimo}$ lanzamiento es de

$$\underbrace{\frac{1}{2} \times \frac{1}{2} \times \ldots \times \frac{1}{2}}_{n-\text{ veces}},$$

o bien $\frac{1}{2^n}$. De aquí que el valor esperado de la variable aleatoria X asociada a este juego sea

$$\begin{aligned} E(X) &= 2\left(\frac{1}{2}\right) + 4\left(\frac{1}{4}\right) + 8\left(\frac{1}{8}\right) + \cdots + 2^n\left(\frac{1}{2^n}\right) + \cdots \\ &= 1 + 1 + 1 + \cdots , \end{aligned}$$

es decir, ¡una cantidad infinita! Sin embargo, nadie en su sano juicio estaría dispuesto a pagar más de una cantidad modesta de dinero para participar en este juego. Esta paradoja se resolverá más adelante, en la siguiente sección.

La solución que propuso Daniel Bernoulli involucra dos conceptos, que ahora son fundamentales en la teoría moderna de la decisión:

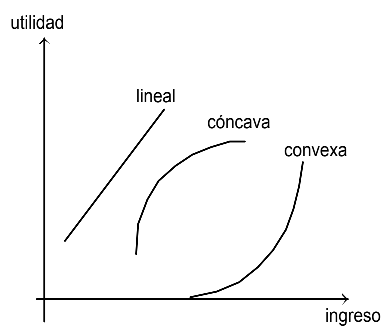
- Los individuos no valúan los premios de forma lineal y el valor del peso (unidad monetaria \$) adicional -o peso **marginal**- disminuye conforme nuestra riqueza va aumentando. Así, \$1 puede tener gran valor para un individuo en pobreza extrema pero muy poco para un millonario.

- Las personas no valúan una variable aleatoria de acuerdo a su valor esperado sino, de acuerdo al valor esperado subjetivo o utilidad esperada que perciben individualmente.

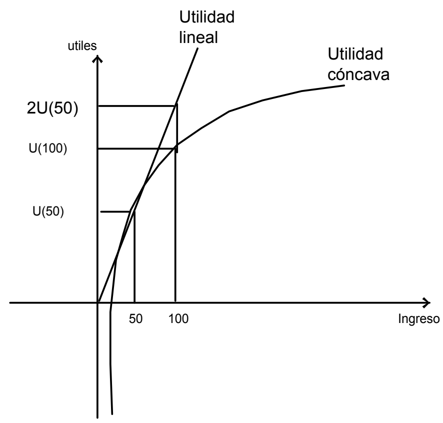
6.2. Utilidad

El ejemplo 5.2.1 de la sección anterior muestra que no sólo los premios monetarios son relevantes para la toma de decisiones. Asimismo, la paradoja de San Petersburgo y el análisis de Daniel Bernoulli sugieren que los individuos valúan los premios o pérdidas de forma individual. Daniel Bernoulli, tuvo la genialidad de medir y cuantificar algo que no es esencialmente cuantitativo: las preferencias individuales. Esto se hace mediante un concepto que llamamos **función de utilidad**.

Vamos a utilizar tres formas geométricas para las funciones de utilidad: lineal, cóncava y convexa. Cualitativamente, si tomamos la utilidad que nos proporciona un cierto nivel de ingreso, éstas son como se ilustra a continuación:



Daniel Bernoulli, esencialmente comparó la utilidad lineal con la cóncava. Si una cantidad de ingreso I nos proporciona *utilidad* $u(I)$, entonces, representamos gráficamente a estas funciones de utilidad como en la siguiente figura, en donde la utilidad o beneficio se mide en *utiles* o *unidades de utilidad*. La recta y la curva representan todos los puntos $(I, U(I))$, y pueden pensarse como gráficas de funciones de utilidad.



Si la utilidad fuese lineal, estaría representada por la línea recta, pero de acuerdo a Daniel Bernoulli, la representación debe ser como la curva cóncava que se muestra, de manera que $U(100) < 2U(50)$, es decir, el valor de \$100 es menor a dos veces el valor de \$50.

La formalización de esta función como una medida cuantitativa del valor subjetivo de cualquier bien económico fue postulada por el economista William S. Jevons (1835 -1882) en la segunda mitad del siglo XIX. En 1944, John Von Neumann (1903-1957) y Oskar Morgernstern (1902 -1976) retomaron el concepto de utilidad esperada de Daniel Bernoulli y la dotaron de una axiomatización en términos de las preferencias individuales. Esta generalización es una parte crucial del desarrollo de la teoría de juegos originada por ambos de la cual hablaremos en el capítulo 7.

¿Qué se quiere decir por *preferencias individuales*? Una forma simple de pensarlo es la siguiente:

Supongamos que un individuo quiere ordenar a los elementos de un conjunto $\mathcal{A}$ de acuerdo a como los valúa subjetivamente. Sea "$\prec$" una **relación de preferencia** de manera que si A y B son elementos de $\mathcal{A}$, entonces, $B \prec A$ significa "A es estrictamente preferido a B ". Por ejemplo, si vamos a un restaurante y nos fijamos en el conjunto de platillos principales, podemos ordenarlos de acuerdo a nuestro gusto particular en ese momento. Supongamos que hay cinco platillos y los ordenamos de la siguiente forma:

1. (F) Filete de atún en costra de amaranto con salsa de chipotle,
2. (S) Sábana de filete de ternera a la tampiqueña,
3. (E) Enchiladas de pato en mole de pepita,
4. (P) Pulpos en achiote con papas al romero,
5. (R) Rollos de pechuga rellenos de queso de cabra y pimiento,

es decir, tenemos $R \prec P \prec E \prec S \prec F$.

Si la relación de preferencia cumple ciertas condiciones, entonces podemos representarla por una función de utilidad. Concretamente, para los platillos mencionados, existirían números n_R, n_P, n_E, n_S, y n_F, tales que

$$n_R < n_P < n_E < n_S < n_F$$

y la utilidad estaría dada por la siguiente regla de correspondencia:

$$\begin{aligned} U(F) &= n_F, \\ U(S) &= n_S, \\ U(E) &= n_E, \\ U(P) &= n_P, \\ U(R) &= n_R. \end{aligned}$$

En otras palabras, la función de utilidad convierte a las preferencias subjetivas en cantidades numéricas. El número mayor representa nuestro platillo preferido y el menor el que menos nos gusta. El cuantificar las preferencias individuales permite justificar la elección de un objeto u acción sobre otra.

En el caso de la paradoja de San Petersburgo, la utilidad está definida sobre los valores de la variable aleatoria X, que son cantidades de dinero. Tenemos, entonces que la utilidad esperada es,

$$E(U(X)) = U(2)\left(\frac{1}{2}\right) + U(4)\left(\frac{1}{4}\right) + \cdots + U(2^n)\left(\frac{1}{2^n}\right) + \cdots$$

U puede escogerse de manera que esta suma infinita tenga un valor finito determinado y ya no se genere la paradoja.

Dependiendo de los supuestos, la utilidad puede ser simplemente **ordinal**, solamente ordenando numéricamente a los objetos o bien **cardinal**, es decir, indica que tanta más utilidad nos proporciona un objeto comparado con otro. La paradoja de San Petersburgo se presenta por un problema de cardinalidad: no es suficiente ordenar la riqueza de menor a mayor (asumimos implícitamente que cualquier individuo prefiere más riqueza que menos), sino se tiene que asumir también que el valor del peso ($) adicional va disminuyendo.

La función de utilidad de Von Neumann y Morgerstern, que es probablemente la más utilizada hoy en día en el estudio de decisiones bajo incertidumbre, define la utilidad sobre **loterías**. ¿Qué se quiere decir con ésto? Si $\{X_1, X_2, \dots, X_n\}$ es un conjunto de valores de una variable aleatoria, en dónde $P(X = X_i) = p_i$, entonces la lotería L es "obtener X_1 con probabilidad p_1, X_2 con probabilidad $p_2, \dots$ y X_n con probabilidad p_n"[2].

Por ejemplo, sea {\$0, \$1000, \$10 000} un conjunto de premios monetarios que ocurren con probabilidades $\frac{1}{2}, \frac{1}{4}$ y $\frac{1}{4}$, respectivamente. Una lotería L podría representarse por pares ordenados de números como:

$$L = \left((0, \frac{1}{2}), (1000, \frac{1}{4}), (10000, \frac{1}{4}) \right).$$

Esto significa: "obtener \$0 con probabiidad $\frac{1}{2}$, \$1000 con probabilidad $\frac{1}{4}$ y \$10 000 con probabilidad $\frac{1}{4}$". La utilidad esperada de la lotería es

$$EU(L) = \frac{1}{2}u(0) + \frac{1}{4}u(1000) + \frac{1}{4}u(10000),$$

en donde u es simplemente una función de utilidad sobre el conjunto de premios.

Las loterías de Von Neumann y Morgerstern asumen que las probabilidades asociadas son objetivas -o frecuentistas-, proporcionadas por una fuente externa al tomador de decisiones. Por esta razón, en ocasiones se les llama **loterías de ruleta**. La teoría análoga, con probabilidades subjetivas o personales fue desarrollada por Leonard Savage (1917 -1971) en su libro *Foundations of Statistics* (1954).

[2]Esencialmente, las loterías pueden pensarse como variables aleatorias.

Savage axiomatiza elegantemente, a partir de las preferencias del individuo, la obtención de probabilidades subjetivas y la función de utilidad. El resultado es una función de **utilidad esperada subjetiva**[3]. A las loterías que incorporan probabilidades subjetivas se les denomina **loterías de caballos**. Todo lo anterior puede extenderse al caso continuo, utilizando técnicas de integración para obtener los valores esperados.

Ejemplos

Ej 6.2.1 Para los platillos mencionados anteriormente, podemos dar la siguiente función de utilidad:

$$\begin{aligned} U(F) &= 10, \\ U(S) &= 8, \\ U(E) &= 5, \\ U(P) &= 3, \\ U(R) &= 1. \end{aligned}$$

En este ejemplo, si la utilidad fuese ordinal, podríamos haber utilizado los números del 1 al 5 para ordenar las preferencias. Suponiendo que es cardinal, tenemos que $U(F) = 2U(E)$ significa que el filete de atún nos proporciona el doble de utilidad que las enchiladas de pato.

Ej 6.2.2 Consideremos las siguientes dos loterías:

1. L_1 : Ganar \$100 con probabilidad $\frac{2}{3}$, ganar \$1 con probabilidad $\frac{1}{3}$.
2. L_2 : Ganar \$500 con probabilidad $\frac{1}{3}$, ganar \$1 con probabilidad $\frac{2}{3}$.

Suponiendo que la utilidad para estos valores de riqueza están dados por,

$$\begin{aligned} u(1) &= 0, \\ u(100) &= 4.6, \\ u(500) &= 6.2, \end{aligned}$$

[3]F. J. Anscombe y Robert Aumann dieron, en 1963, una axiomatización que incorpora ambos casos: probabilidades objetivas y subjetivas.

¿qué lotería proporciona la mayor utilidad esperada? Evaluando ambas, tenemos que

$$EU(L_1) = \frac{2}{3} \times 4.6 + \frac{1}{3} \times 0 = 3.07,$$
$$EU(L_2) = \frac{1}{3} \times 6.2 + \frac{2}{3} \times 0 = 2.07,$$

de manera que L_1 es la lotería cuya utilidad esperada es mayor.

Ej 6.2.3 Con las mismas loterías del ejemplo anterior, ahora la función de utilidad del individuo es lineal y cada peso le proporciona una unidad de utilidad. ¿Qué lotería le proporciona la mayor utilidad esperada?

$$EU(L_1) = \frac{2}{3} \times 100 + \frac{1}{3} \times 1 = 67,$$
$$EU(L_2) = \frac{1}{3} \times 500 + \frac{2}{3} \times 1 = 84,$$

de manera que ahora la segunda lotería es la que le proporciona mayor utilidad esperada.

♠♠♠♠♠

6.3. Tomando decisiones

Las teorías de utilidad descritas arriba presentan varios problemas cuando se realizan pruebas empíricas, sin embargo, los modelos alternativos no están exentos de dificultades y son mucho más complejos para operar. Hoy en día, las teorías de preferencias y utilidad que hemos expuesto son lo mejor que se tiene y, en general, representan correctamente la mayoría de las situaciones reales.

Estas teorías asignan probabilidades a lo que creemos que va a suceder y valores de utilidad a las consecuencias o resultados que experimentaremos. Las diferencias entre individuos pueden deberse a distintas estimaciones de probabilidad, si ésta es subjetiva, o bien a diferentes funciones de utilidad.

A lo largo de nuestra existencia, tenemos que elegir acciones o alternativas de acuerdo a las posibles consecuencias que se derivan de éstas. Las consecuencias pueden ser deterministas, o aleatorias. Si las probabilidades involucradas son objetivas, como sucedería al participar en cualquier juego de azar, decimos que tomamos una decisión bajo un entorno de riesgo. En el caso en que las probabilidades sean subjetivas, lo cual es casi siempre, se dice que la decisión se toma bajo un escenario de incertidumbre. Independientemente del tipo de probabilidad el proceso de decisión es básicamente el mismo, se pretende elegir la acción óptima de la siguiente forma:

- Se tiene algún conjunto de acciones posibles a elegir. El no hacer nada, o quedarse con el *status quo*, se considera una acción.
- Se evalúa la utilidad, en el caso determinista, o la utilidad esperada en el caso de consecuencias aleatorias, para las consecuencias que se derivan de estas acciones.
- Se elige la (s) acción (es) que resulta (n) en la mayor utilidad o utilidad esperada posible.

¿Cómo funciona este proceso en la práctica? Recordemos el ejemplo 5.3.1; ahí podemos pensar a las acciones de llevar o no paraguas como dos loterías. Los valores, aparentemente arbitrarios, que asignamos a $X(lluvia)$, $X(no\ lluvia)$, $\sim X(lluvia)$ y $\sim X(no\ lluvia)$ representan las utilidades de las consecuencias de los actos correspondientes. Elegimos el *llevar paraguas* pues su utilidad esperada es mayor.

6.3.1. Pascal y la existencia de Dios

En 1658 Pascal publicó una obra filosófica llamada *Pensées*. En ella utilizó el concepto de valor esperado para justificar el porqué se debe llevar una vida piadosa honrando a Dios. Veamos cual era su argumento.

Se tienen dos estados posibles del mundo:

$$s_1 : \text{ Dios existe}$$

y su negación,

$$s_2 : \text{ Dios no existe.}$$

Sea p la probabilidad de que s_1 sea verdadero, es decir, la probabilidad de que Dios exista y supongamos que $p \neq 0$. Entonces, $1-p$ es la probabilidad de que Dios no exista.

Tenemos dos opciones de vida o acciones posibles a elegir:

$$a_1: \text{ llevar una vida piadosa}$$

o bien

$$a_2: \text{ llevar una vida pecadora.}$$

Si Dios existe y vivimos una vida piadosa, seremos recompensados por la vida eterna en el cielo. La utilidad de esta recompensa es *infinita* y la denotamos por ∞. Si Dios existe y llevamos una vida pecadora, recibiremos una recompensa menor o seguramente un castigo (recompensa negativa) el cual nos proporciona una utilidad de U_1.

Si Dios no existe y llevamos una vida piadosa, estaremos sacrificando muchos placeres mundanos sin razón. La recompensa recibida también puede ser negativa y la utilidad derivada de la misma la denotamos por U_2. Finalmente, si Dios no existe y llevamos una vida pecadora, la pasaremos bien durante nuestra vida y tendremos una recompensa que nos proporciona una utilidad de U_3, en donde $U_3 > U_2$ (una suposición sensata dado que no hay porqué sacrificarse si no hay castigo).

La utilidad esperada de llevar una vida piadosa es[4],

$$EU(a_1) = p \times \infty + (1-p)U_2 = \infty.$$

Asimismo, la utilidad esperada de llevar una vida pecadora es,

$$EU(a_2) = p \times U_1 + (1-p)U_3 < \infty$$

[4]Utilizamos las siguientes propiedades del concepto de ∞ ("infinito"):

- $\infty > n$ para cualquier número n
- $m \times \infty = \infty$ para cualquier número positivo m, en particular, cualquier fracción de "infinito" sigue siendo "infinito".
- $\infty + k = \infty$ para cualquier número k.

de manera que

$$EU(a_2) < EU(a_1)$$

y la elección racional es elegir llevar una vida piadosa.

El llevar una vida pecadora sólo tendría sentido si la probabilidad de que Dios exista es nula ($p = 0$), puesto que $U_3 > U_2$. De acuerdo a Pascal, nadie puede tener la certeza absoluta de que Dios no exista, de manera que aunque la probabilidad de su existencia sea muy pequeña, ésta no es nula ($p \neq 0$). Por lo tanto, es óptimo llevar una vida piadosa honrando a Dios si nos cabe la menor duda de su existencia.

Ejemplos

Ej 6.3.1 Las loterías y sorteos están diseñados de tal manera que su valor esperado es negativo para el participante. Por lo tanto, los organizadores siempre ganan y estos mecanismos son populares para recaudar fondos. No obstante, millones de personas compran diariamente boletos para sorteos y loterías, acción que no parece óptima a todas luces. ¿Por qué alguien, que apenas gana el salario mínimo, gasta \$15 en un boleto para participar en Melate? Pensemos en las consecuencias de comprar dicho boleto:

1. Una oportunidad casi nula de ganar un gran premio.

2. Una esperanza o sueño, que puede durar algunos días, mientras se realiza el sorteo.

El premio esperado podrá ser negativo, pero, ¿cuál es la utilidad de un sueño? Por lo visto, mayor a \$15.

Ej 6.3.2 Este ejemplo es similar al ejercicio 5.5. En una caja de Corn Flakes hay un cupón para participar en el sorteo de un televisor LCD de 36 pulgadas, cuyo precio comercial se estima en \$14000. Se prevee que unas 50000 personas envíen cupones para participar, de manera que la probabilidad de ganar el televisor es de $\frac{1}{50000}$. El procedimiento parece simple, se llena el cupón con nuestros datos y posteriormente se envía por correo registrado o mensajería a la dirección indicada.

Queremos decidir entre tomarnos la molestia de recortar, llenar, poner en un sobre y pagar por el envío del cupón -aproximadamente \$20-, o simplemente no hacer nada y comer tranquilamente nuestros Corn Flakes. Denotemos al acto de enviar el cupón por A_1 y a no hacer nada por A_2. Supongamos que aquí la utilidad es lineal y cada peso proporciona una unidad de utilidad. La utilidad esperada de estos actos es,

$$\begin{aligned} EU(A_1) &= \left(1 - \frac{1}{50000}\right)(-20) + \frac{1}{50000} \times (14000 - 20) = -19.72, \\ EU(A_2) &= 0, \end{aligned}$$

¡y esto sin tomar en cuenta nuestro tiempo! Evidentemente lo óptimo será elegir A_2 y no hacer nada.

Ej 6.3.3 Supongamos que compramos un coche con valor de \$100 000. La probabilidad de que el coche sea robado y no recuperado es de $\frac{1}{100}$. La aseguradora **A** nos ofrece un seguro contra robo por tres años, con un costo total de \$2000 y un deducible de \$15 000. Es decir, si nos roban el coche, entonces recuperamos 100 000 - 15 000 = 85 000 pesos. La aseguradora **B** ofrece un seguro contra robo, también por tres años, con un costo de \$2500 y sin deducible, de manera que recuperaríamos el valor total del auto en caso de que éste sea robado. Alternativamente, podemos no asegurar el auto y encomendarlo a la buena providencia.

Sean A_1, A_2 y A_3, respectivamente, los actos que corresponden a comprar el seguro de la aseguradora, **A**, de la **B** o no comprar seguro alguno. Supongamos que cada peso nos da una unidad de utilidad. Las utilidades esperadas de estos actos son,

$$\begin{aligned} EU(A_1) &= \left(1 - \frac{1}{100}\right)(-2000) + \frac{1}{100}(85000 - 2000) = -1150, \\ EU(A_2) &= \left(1 - \frac{1}{100}\right)(-2500) + \frac{1}{100}(100000 - 2500) = -1500, \\ EU(A_3) &= \frac{-100000}{100} = -1000. \end{aligned}$$

De aquí que deba elegirse no comprar seguro.

♠♠♠♠♠

6.4. Atrevidos y cautos

El ecologista Steve Irwin (1962-2006), dirigía un zoológico en Australia y era mejor conocido por su personalidad televisiva del *cazador de cocodrilos.* Durante años protagonizó una serie documental en la cual luchaba contra enormes cocodrilos, gigantescas boas, cobras extremadamente venenosas, insectos mortales, tiburones y, en general, cualquier tipo de alimaña que pudiese ser peligrosa. Steve Irwin no parecia tenerle miedo a nada. En septiembre de 2006, mientras filmaba uno de los episodios de su programa bajo el agua del mar de Queensland, fue accidentalmente picado en el pecho por una mantarraya y murió casi instantáneamente.

Existen los personajes intrépidos como Steve Irwin y existen aquellos que viven con cautela extrema. Hay quienes pierden grandes cantidades de dinero en los casinos y hay quienes nunca invierten en algo más riesgoso que una cuenta de ahorros. Existen los que realizan salto de bungee y paracaidismo, otros simplemente juegan al golf. Hay quienes jamás comprarán un seguro de ningún tipo, hay quienes aseguran todo lo asegurable. Existen los que derivan placer al producir adrenalina y hay quienes se estresan con las emociones fuertes. Podemos concluir que los seres humanos somos sumamente diversos en nuestras actitudes ante el riesgo. ¿Habrá forma de reflejar esto en la función de utilidad individual?

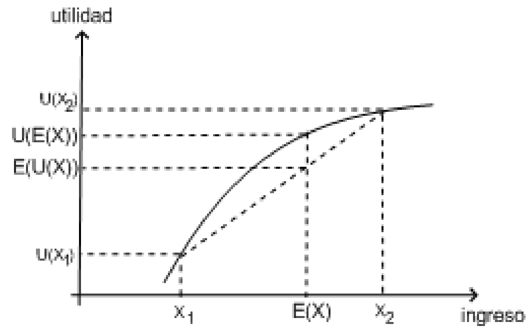
Supongamos que el ingreso x nos proporciona utilidad. Este ingreso es una variable aleatoria y toma valores x_1 con probabilidad p_1 y x_2 con probabilidad p_2. Su valor esperado es

$$E(x) = p_1 x_1 + p_2 x_2.$$

Si la función de utilidad de un individuo es cóncava, como en el caso propuesto por Daniel Bernoulli, entonces se tiene que

$$E(U(x)) \leq U(E(x)), \tag{6.1}$$

como se muestra en la figura 6.2 a continuación:



$$\textit{Aversión al riesgo}: E(U(x)) \leq U(E(x)) \tag{6.2}$$

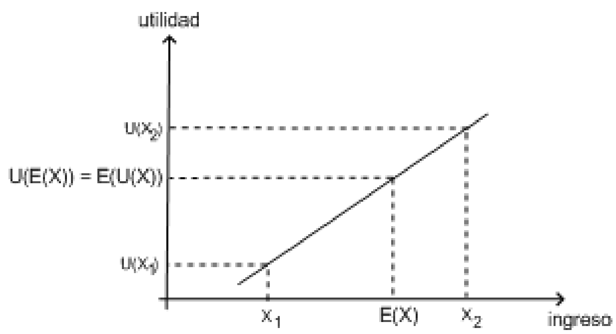
Esto significa que este individuo prefiere obtener la cantidad $E(x)$ con seguridad y tener utilidad $U(E(x))$, a participar en la lotería que le proporciona x_1 con probabilidad p_1 y x_2 con probabilidad p_2, en cuyo caso obtendría la utilidad esperada $E(U(x))$.

Definición 6.4.1 *Con la notación anterior, se dice de un individuo cuya función de utilidad satisface la desigualdad*[5] *dada por 6.1, que es* ***adverso al riesgo****.*

Si la función de utilidad fuese lineal, entonces tendríamos que se cumple la igualdad $E(U(x)) = U(E(x))$ y el individuo sería indiferente entre tener

[5] A esta desigualdad se le conoce como **desigualdad de Jensen**, por el matemático danés Johan Jensen (1859-1925), quien caracterizó a las funciones cóncavas y convexas mediante esta propiedad.

$E(x)$ con seguridad o participar en la lotería. Esto se ilustra en la figura 6.3



Neutralidad al riesgo: $E(U(x)) = U(E(x))$

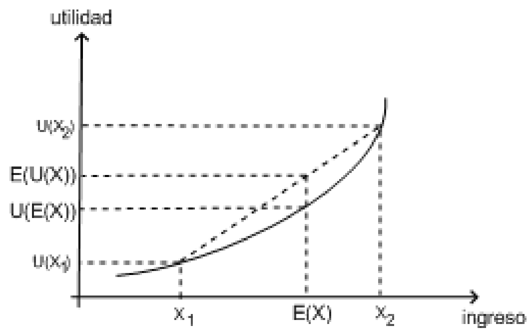
(6.3)

La definición formal de este tipo de perfil de riesgo es,

Definición 6.4.2 *Con la notación anterior, se dice de un individuo cuya función de utilidad satisface $E(U(x)) = U(E(x))$, que es* ***neutral al riesgo****.*

Finalmente, la función de utilidad podría ser convexa, como se ilustra en la figura 6.4. En este caso se tiene la siguiente definición:

Definición 6.4.3 *Con la notación anterior, se dice de un individuo cuya función de utilidad satisface $E(U(x)) \geq U(E(x))$, que es* ***amante del riesgo****.*



Amor al riesgo: $E(U(x)) \geq U(E(x))$

(6.4)

Notemos que un individuo amante del riesgo prefiere tomar la lotería que obtener el valor esperado con seguridad.

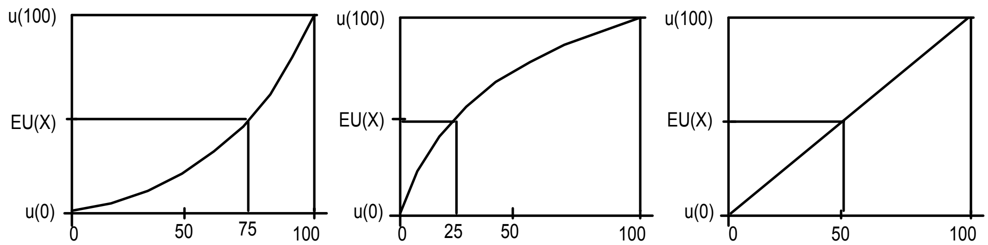
Ejemplos

Ej 6.4.1 El individuo del ejemplo 6.2.2 es adverso al riesgo y el del ejemplo 6.2.3 es neutral al riesgo. Para estos valores de los premios, los individuos tienden a comportarse con neutralidad al riesgo -o poca aversión al riesgo, es decir una utilidad casi lineal- y en general escogen la lotería L_2. Si las loterías fuesen las siguientes:

1. L_1 : Ganar \$1 000 000 con probabilidad $\frac{2}{3}$, ganar \$1 con probabilidad $\frac{1}{3}$.
2. L_2 : Ganar \$5 000 000 con probabilidad $\frac{1}{3}$, ganar \$1 con probabilidad $\frac{2}{3}$.

Los individuos se comportarían con mayor aversión al riesgo y en general escogerían L_1. Esto implica que los individuos tienden a ser más cautos cuando se trata de premios (o pérdidas) significativas.

Ej 6.4.2 Supongamos que alguien nos ofrece la siguiente lotería: ganar \$100 con probabilidad $\frac{1}{2}$ o ganar \$0 con probabilidad $\frac{1}{2}$. ¿Cuánto estaríamos dispuestos a pagar por participar en este juego? Como veremos a continuación depende de nuestra actitud ante el riesgo. Tomemos tres casos como se muestra en la siguiente figura:



Las tres gráficas ilustran a un individuo amante del riesgo, otro adverso al riesgo y el último neutral al riesgo, respectivamente. La utilidad esperada de la ganancia X es

$$EU(X) = \frac{1}{2}u(100) + \frac{1}{2}u(0),$$

que, como se muestra, se encuentra en el punto medio entre $u(0)$ y $u(100)$. Esta utilidad esperada es la misma para los tres individuos. Notemos, sin embargo, que cada uno valúa esta lotería de forma diferente.

Si V es la cantidad de dinero tal que $U(V) = \frac{U(100)+U(0)}{2}$, entonces, V es el valor asociado a la lotería ya que el individuo obtiene la misma utilidad entre la lotería y la cantidad V. Según la figura, el amante del riesgo considera que $V = \$75$, el adverso al riesgo toma $V = \$25$ y el neutral al riesgo $V = \$50$, que es exactamente el valor esperado de la lotería. Esta sería la cantidad que cada uno estaría dispuesto a pagar por participar.

Ej 6.4.3 En el ejemplo anterior, nótese que podemos interpretar la cantidad V en la que cada individuo valúa la lotería, es decir, el ingreso tal que $U(V) = EU(X)$, como la cantidad de dinero que acepta para ser indiferente entre dicha cantidad y la lotería. A estas cantidades se les llama **equivalentes ciertos**. Al amante del riesgo le es indiferente tener \$75 pesos o una lotería que le da \$100 o \$0 con probabilidad 1/2. El individuo neutral al riesgo le da lo mismo tener \$50 o la misma lotería. La persona adversa al riesgo, intercambiaría la lotería por \$25.

En el lenguaje de seguros, la diferencia entre el valor esperado y el equivalente cierto se denomina la **prima de riesgo**. Ésta representa la cantidad que el individuo está dispuesto a pagar por evitar participar en la lotería. En el ejemplo anterior, el individuo adverso al riesgo estaría dispuesto a pagar \$50 - \$25 = \$25 con tal de evitar la lotería, el individuo amante al riesgo pagaría \$75 - \$50 = \$25 para participar en la lotería y el individuo neutral al riesgo es totalmente indiferente y no paga nada.

La lotería podría representar el perder un activo por medio de un evento catastrófico con cierta probabilidad o no perderlo con la probabilidad complementaria. El individuo adverso al riesgo sería la persona que estaría dispuesta a pagar por asegurar el activo.

♠♠♠♠♠

6.5. Nuevas tendencias: *paternalismo libertario*

El proceso de toma de decisiones descrito en los últimos dos capítulos no garantiza que los individuos puedan elegir siempre en forma adecuada. En ocasiones la información es insuficiente o bien se carece de experiencia ante una situación nueva y se cometen errores inevitablemente. Adicionalmente, los procesos cerebrales naturales pueden ocasionar la toma de decisiones subóptimas, por ejemplo, la propensión a comportamientos adictivos y la dificultad para cuantificar la incertidumbre en forma apropiada.

El proceso de decisión de un individuo es más complejo que un simple análisis racional de *costo – beneficio*. Existen procesos fisiológicos en el cerebro humano que son determinantes para la percepción del beneficio o utilidad producto de una elección dada. Las decisiones involucran una parte racional, lógica y objetiva, procesada en una zona del cerebro, y otra emocional e intuitiva, procesada en otra región distinta. Así, si crecimos utilizando el sistema métrico y alguién nos dice que una persona mide 5 pies y 7 pulgadas, tenemos que reflexionar utilizando la parte racional del cerebro, para imaginar que tan alta es esta persona. Sin embargo, si nos dicen que mide 1.70 metros, automáticamente visualizamos su altura utilizando la región cerebral correspondiente a la intuición.

Así como existen ilusiones ópticas, el cerebro juega con nuestras percepciones y en ocasiones favorecemos el uso de la parte incorrecta del mismo para tomar alguna decisión. Cuando esto sucede, nos basamos fundamentalmente en la intuición cuando deberíamos también reflexionar y racionalizar, o lo opuesto: reflexionamos en exceso cuando simplemente deberíamos dejarnos llevar por nuestro mejor instinto. Como consecuencia, realizamos gastos innecesarios, ponemos nuestra salud en peligro y pensamos demasiado antes de lanzar un tiro penal o decidir si una persona nos atrae. Veamos un ejemplo específico:

Cuando una persona realiza un pago en efectivo, se genera actividad cerebral asociada con sentimientos negativos. Lo curioso es que si el mismo pago se efectúa con tarjeta de crédito, entonces se atenúa esta actividad y, literalmente, nos duele menos el desembolso[6]. Esta sensación sesga nuestro análisis racional acerca del precio del producto en cuestión y, en consecuencia, estamos dispuestos a pagar precios más altos si utilizamos nuestra tarjeta de crédito. Para corregir esto, debemos hacer conciencia del sesgo, no dejarnos llevar por la intuición y reflexionar antes de utilizar nuestra tarjeta.

En los últimos años el estudio del comportamiento humano y la toma de decisiones se ha convertido en una ciencia interdisciplinario. En ella se mezcla la teoría económica con la psicología la neurobiología y la modelación matemática. Recientemente ha surgido una corriente nueva de política y derecho públicos denominada *paternalismo libertario*[7], que intenta echar mano de esta teoría integral del comportamiento en beneficio de los ciudadanos. Veamos brevemente que es lo que se pretende.

La teoría económica tradicional nos dice que si tenemos acceso a un mayor número de bienes y servicios, entonces nuestro bienestar se incrementa.

[6]Drazen Prelec, Duncan Simester, “Always Leave Home Without It: A Further Investigation of the Credit-Card Effect on Willingness to Pay”, *Marketing Letters,* Kluwer Academic Publishers, 12:1, pp. 5-12, 2001.

[7]Del Inglés: *libertarian paternalism.* Se recomienda ver el libro de Thaler y Cass proporcionado en la bibliografía del capítulo. Es interesante notar que la administración del presidente Barack Obama está llena de profesionales estudiosos de la economía del comportamiento.

Así, las políticas públicas deben concentrarse en garantizar las mejores condiciones para la producción de estos bienes y servicios y procurar el que sean accesibles a la mayor cantidad de personas posible. Los individuos, si poseen los medios, realizarán la mejor elección posible en cuanto a lo que deben producir y consumir y lo óptimo es que el Estado no intervenga en este proceso.

Los proponentes del *paternalismo libertario*, la mayoría pertenecientes al área de economía denominada *economía del comportamiento*[8], sugieren que como somos propensos a errar en virtud de nuestra neurobiología, el Estado debe intervenir ocasionalmente para incentivarnos a tomar las decisiones correctas. El término de *libertario* implica que no hay coerción alguna sobre el individuo; simplemente, el "menú" de opciones disponibles se manipula para incrementar la probabilidad de una elección óptima. Evidentemente, la configuración de dicho menú de opciones no es trivial y requiere de expertos en la teoría del comportamiento.

Es claro que esta manipulación por medio "expertos" potencialmente podría ser peligrosa. ¿Qué tal si estos *arquitectos de la decisión* carecen de valores éticos y no desean el bienestar de la sociedad? Ciertamente es una preocupación válida para aquellos que desconfían de cualquier intervención gubernamental. Sin embargo, independientemente de su aplicación al ámbito de la política pública, la nueva teoría del comportamiento es una disciplina científica seria y fascinante que nos ayuda a entender la complejidad detrás de la toma de decisiones.

Bibliografía Complementaria

1. Bernoulli Daniel, "Exposition of a New Theory on the Measurement of Risk", *Econometrica* vol 22, pp23-36, 1954.

2. Bernstein Peter, *Against the Gods: The Remarkable Story of Risk*, Wiley, 1998.

[8]Entre ellos están Dan Ariely de MIT, George Akerlof de Berkeley, Richard Thaler de la Universidad de Chicago y Robert Shiller de Yale, entre otros.

3. Devlin Keith, *The Unfinished Game*, Basic Books, New York, NY, 2008.

4. "Utility and Probability", *The New Palgrave Collection*, editado por Eatwell John, Milgate Murray and Newman Peter, Macmillan Press, New York, NY, 1990.

5. Hirshleifer Jack, Riley John G., *The Analytics of Uncertainty and Information*, Cambridge University Press, Cambridge, UK, 1997.

6. Jeffrey Richard, *The Logic of Decision*, University of Chicago Press, Chicago, Ill, 1983.

7. Jonah Lehrer, *How We Decide,* Houghton Mifflin Co, New York, 2009.

8. Drazen Prelec, Duncan Simester, "Always Leave Home Without It: A Further Investigation of the Credit-Card Effect on Willingness to Pay", *Marketing Letters,* Kluwer Academic Publishers, 12:1, pp. 5-12, 2001.

9. Savage Leonard, *The Foundations of Statistics*, Dover Publications, New York, NY, 1972.

10. Thaler H. Richard, Sunstein R. Cass, *Nudge: Improving Decisions about Health, Wealth and Happiness,* penguin Books, New York, NY, 2009.

11. Varian Hal R., *Microeconomía Intermedia*, 7^a edición, Antoni Bosch Editores, Barcelona, España, 2007.

Ejercicios

▷ **6.1** Supongamos que cada peso equivale a una unidad de utilidad. Un individuo tiene que estacionarse unos minutos para recoger una pizza. En la calle está prohibido estacionarse y se deberá pagar una multa de \$200 si llega un agente de tránsito al lugar. La probabilidad de que esto suceda es de $\frac{1}{50}$. Junto a la Pizzería hay un estacionamiento público que cobra \$10, por la primera media hora. Dada esta información, ¿dónde se estacionará este individuo?

▷ **6.2** Considerar las siguientes dos loterías: L_1 : ganar \$500 con probabilidad 1, L_2 : ganar 2500 con probabilidad 0.1, 500 con probabilidad 0.89 y nada con probabilidad 0.01.

1. Sin evaluar las ganacias esperadas, escoger una de ellas (la respuesta puede varíar según el individuo).
2. Asumiendo que \$1 equivale a un util, evaluar la utilidad esperada de ambas loterías. ¿Es consistente el resultado con el inciso anterior?

▷ **6.3** Considerar las siguientes dos loterías. L_1 : ganar \$500 con probabilidad 0.11 y nada con probabilidad 0.89, L_2 : ganar 2500 con probabilidad 0.1 y nada con probabilidad 0.9.

1. Sin evaluar las ganacias esperadas, escoger una de ellas (la respuesta puede varíar según el individuo).
2. Asumiendo que \$1 equivale a un util, evaluar la utilidad esperada de ambas loterías. ¿Es consistente el resultado con el inciso anterior?[9]

▷ **6.4** Un criminal en la ciudad de México, apodado “el Pecas”, tiene un flujo de ingreso tal que le proporciona una utilidad de 500 utiles mientras no lo atrapen. La probabilidad de ser atrapado y condenado a una pena mayor es de $\frac{1}{100}$. En

[9] A los ejercicios 6.3 y 6.2 se les conoce como “Paradoja de Allais”. El economista francés Maurice Allais (1911-) propuso este ejemplo para probar que la teoría de la utilidad esperada era inservible. Según la teoría, en ambos ejemplos debe elegirse la primera lotería.

caso de que esto sucediera, su utilidad sería de -1000 utiles. Si "el Pecas" se rehabilitara y se convirtiera en una persona honesta, podría conseguir un trabajo en un taller mecánico que le pagaría poco, pero le permitiría dormir tranquilo. La utilidad que esta vida honesta le proporcionaría es de 400 utiles. Sea A_1 el acto de elegir la vida criminal y A_2 el de elegir una vida honesta.

1. ¿Qué acto le proporciona la mayor utilidad esperada al Pecas?
2. Proponer alguna solución, en términos de políticas públicas.

▷ **6.5** Proporcionar una explicación al comportamiento de Hamlet: no tomar decisiones o bien, decidir quedarse con el status quo. Explicar en términos de su función de utilidad o de su asignación de probabilidades subjetivas o ambas.

▷ **6.6** Asumir que en el ejemplo 6.3.3, el auto vale \$100 000 y la probabilidad de robo sigue siendo $\frac{1}{100}$. Sin embargo, compramos el seguro de la compañía **A** pues en caso de robo, el quedarnos sin nuestro medio transporte y sin dinero para reponerlo sería una complicación. ¿Al menos cuánto estamos dispuestos a pagar para eliminar esta inconveniencia?

▷ **6.7** Asumir que en el ejemplo 6.3.3 la probabilidad de que el coche sea robado aumenta a $\frac{1}{50}$, evaluar las utilidades esperadas y determinar la elección óptima.

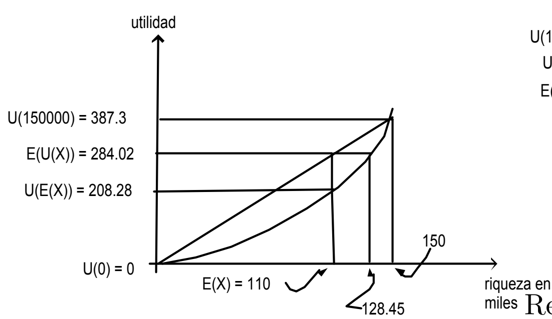
▷ **6.8** Asumir que en el ejemplo 6.3.3, la probabilidad de robo sigue siendo $\frac{1}{50}$, pero la aseguradora **B** baja el costo del seguro por tres años a \$2 250 . Determinar la elección óptima.

▷ **6.9** Pensemos en dos tipos de individuos: los verdes y los cafés. Los verdes consideran que es urgente que las medidas que restringen las emisiones industriales se implementen cuanto antes. Las consecuencias de no hacerlo resultarían desastrosas para la humanidad por el calentamiento global asociado. Los cafés alegan que es muy costoso adoptar medidas extremas para reducir las emisiones. Éstas se irán adoptando naturalmente durante los próximos años. A pesar de que hay que estar concientes del problema no hay razón para alarmarse en el corto plazo. La diferencia entre los verdes y los cafés es su estimación de las probabilidades de las consecuencias, su función de utilidad o ambas? Explicar.

▷ **6.10** Un individuo posee un automóvil valuado en \$150 000. La probabilidad de robo total del auto es de $\frac{4}{15}$ de manera que si X es la variable aleatoria que representa el valor del vehículo, se tiene que

$$X = \begin{cases} 0 & \text{con probabilidad } \frac{4}{15}, \\ 150000 & \text{con probabilidad } \frac{11}{15}. \end{cases}$$

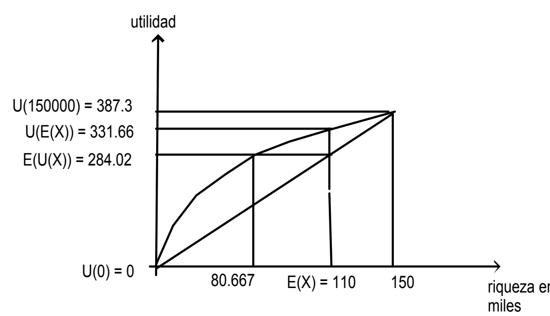
Considerar la siguiente figura que representa la función de utilidad de este individuo y nos proporciona todos los valores relevantes:



Responder a las siguientes preguntas:

1. Verificar que $E(X) = 110000$. ¿Qué representa este valor para el individuo?
2. ¿Cómo se calculó el valor de 284.02 unidades de utilidad?
3. ¿Comprará seguro este individuo? Si la respuesta es afirmativa, ¿cuánto estaría dispuesto a pagar por él? Justificar.
4. Determinar el perfil de riesgo de este individuo (adverso, neutral o amante del riesgo). Justificar.

▷ **6.11** Supongamos que todos los datos son idénticos al problema anterior, pero la función de utilidad del individuo está descrita por la siguiente figura:



Responder a las siguientes preguntas:

1. ¿Comprará seguro este individuo? Si la respuesta es afirmativa, ¿cuánto estaría dispuesto a pagar por él? Justificar.

2. Determinar el perfil de riesgo de este individuo (adverso, neutral o amante del riesgo). Justificar.

▷ **6.12** Sea X la variable aleatoria que representa la riqueza de un individuo. Considerar la siguiente lotería : perder \$1000 con probabilidad $\frac{1}{3}$ o perder \$0 con probabilidad $\frac{2}{3}$. Si el individuo tiene una cantidad inicial de riqueza de \$1500, encontrar $E(X), U(E(X)), E(U(X))$, el equivalente cierto y la prima de riesgo para los siguientes casos:

1. La función de utilidad es $U(X) = \sqrt{X}$,

2. la función de utilidad es de $U(X) = X^2$.

Capítulo 7

Introducción a los juegos

7.1. Introducción

En la sección 6.3 se describió brevemente el proceso de decisión de un individuo. Con frecuencia, las decisiones no son tomadas de forma aislada y dos o más individuos interactúan entre sí en este proceso. El estudio de la toma de decisiones de un conjunto de individuos es, esencialmente, lo que se conoce como **teoría de juegos**.

La mención más antigua de una estrategia de teoría de juegos data del Talmud Babilónico (0 - 500 DC), en donde se describe la división de una herencia; sin embargo, el nacimiento oficial de la teoría ocurrió hasta el siglo XX. En 1921, el matemático francés, Émile Borel (1871 - 1958), publicó varios artículos en los cuales sugería el uso de un sistema estructurado para algunos juegos de salón, pero fue el matemático húngaro John Von Neumann, mencionado en la sección 6.2, el primero en formalizar estas ideas en su artículo "Theory of Parlour Games" (1928).

La inspiración de Von Neumann fue el juego de poker: se dio cuenta que éste no era el típico juego de azar, guiado por la teoría de la probabilidad. En el poker, no necesariamente gana el jugador que tiene la mejor mano[1],

[1] Para aquellos que nunca hayan jugado cartas, la palabra "mano" denomina al conjunto de cartas que posee un jugador y con las cuales llevará a cabo el juego.

pues el ganador puede ser aquel que ha convencido a los demás de que posee la mejor mano. Para representar su mano, los jugadores se valen del recurso de apostar diferentes cantidades de dinero. Por ejemplo, apostarían una cantidad baja si quieren representar una mano mediocre o bien una cantidad alta si quieren representar una buena mano. El juego sería trivial de no ser porque puede utilizarse el engaño o *bluff*, representando una mano distinta a la que se posee.

Von Neumann emigró a los Estados Unidos para incorporarse a la Universidad de Princeton. Allí conoció al economista Oskar Morgerstern, con quien publicó el libro que revolucionó a la teoría económica y convirtió a la teoría de juegos en una disciplina académica: *Theory of Games and Economic Behavior* (1944). Es en este mismo libro en donde exponen la teoría de utilidad esperada sobre loterías mencionada en la sección 6.2.

Von Neumann aplicó la nueva teoría para estrategias militares y para modelar el conflicto entre Estados Unidos y la Unión Soviética durante la Guerra Fría. Adicionalemente, revolucionó a la computación con el desarrollo de la computadora digital que realizaba miles de operaciones por segundo. Ésta fue una pieza instrumental para el desarrollo de la bomba atómica en el proyecto Manhattan, del cual formó parte Von Neumann.

El matemático fue criticado por varios de sus colegas académicos, ya que apoyó la utilización de la bomba atómica y ayudó a seleccionar las ciudades japonesas que serían blancos de la misma. Como muchos científicos que habían realizado investigación con materiales radioactivos, Von Neumann murió de cancer en 1957, después de haber sido una de las mentes más brillantes y más controvertidas del siglo XX.

En 1948, otro hombre genial entró a Princeton para doctorarse en matemáticas: John Nash. El joven Nash publicó entre 1950 y 1953, varios trabajos que aportaron las ideas fundamentales acerca del concepto de solución en la teoría de juegos. A partir de las aportaciones de Nash, la teoría de juegos se convirtió en una disciplina independiente cuyas aplicaciones van desde la mesa de poker hasta la biología evolutiva.

7.2. ¿Qué es un juego?

La definición general de **juego** podría ser: *cualquier situación en la cual existe una interacción entre dos o más individuos.* Nos podrían venir a la mente los siguientes ejemplos: "gato", "piedra papel y tijera", futbol, damas, ajedrez, poker, etcétera. Sin embargo, de acuerdo con la definición dada, cualquier tipo de negociación, una batalla militar y la repartición de un pastel también son juegos. La **solución** de un juego es la descripción formal de los resultados posibles del juego.

Los participantes de un juego se denominan, ¿qué más? ¡jugadores! Cada jugador tiene una serie de **opciones o estrategias** que puede llevar a cabo en cada etapa del juego y éstas se ejercitan de acuerdo a **reglas** preestablecidas. Las diferentes combinaciones de estrategias de los jugadores dan lugar a los resultados posibles del juego. Las preferencias de cada jugador están representadas por una función de utilidad -o utilidad esperada- y los jugadores actúan racionalmente de manera que siempre intentarán maximizar su utilidad.

Las estrategias pueden ser **puras**, cuando la única opción es llevarlas cabo con probabilidad uno o cero, o bien **mixtas** cuando se mezclan en diferentes proporciones. Así, si un pitcher siempre lanza bolas rápidas contra cierto bateador, se trata de una estrategia pura. Si, en cambio, una tercera parte de sus lanzamientos son curvas y el resto son bolas rápidas, entonces su estrategia contra el bateador es mixta.

Los juegos son **cooperativos**, cuando los jugadores pueden negociar acerca de las estrategias a seguir o bien, **no cooperativos,** si los jugadores siguen su estrategia, independientemente del otro jugador. Si los participantes del juego tienen toda la información relativa a las jugadas que realizan los demás y sus preferencias, entonces, decimos que se trata de un juego de **información perfecta**. Si esto no sucede, entonces el juego es de **información imperfecta**.

Veamos algunos ejemplos para clarificar las clasificaciones de los diferentes tipos de juegos expuestas en el párrafo anterior.

- El juego de gato es un juego no cooperativo -no hay negociación previa entre los jugadores- y de información perfecta -ambos jugadores conocen las opciones que tiene el otro-. El ajedrez puede clasificarse de la misma forma, aunque las reglas son más complejas y, por supuesto, el número de estrategias posibles es enorme.

- El poker es un juego no cooperativo de información imperfecta. En el ajedrez, ambos jugadores ven el mismo tablero y pueden, en principio, calcular todas las posibilidades que se derivan de una jugada específica -aunque evidentemente esto no es trivial-. En el poker, no conocemos las cartas de los demás jugadores ni las cartas que faltan por aparecer, de manera que el problema no es meramente computacional.

- El juego de "piedra papel y tijera" es no cooperativo y de información perfecta. Una decisión acerca de cómo librar una batalla militar podría clasificarse de la misma forma.

- La negociación para la compra de una casa, la repartición de una herencia o el cabildeo parlamentario son juegos cooperativos, de información imperfecta.

Si los jugadores escogen sus estrategias de una buena vez -no hay marcha atrás- y las decisiones de todos los participantes se realizan simultáneamente, entonces el juego está en **forma estratégica o normal.** En contraste, si los jugadores toman acciones en varias etapas y no sólo al principio del juego, se denomina un juego en **forma extensiva o secuencial** y se especifica el orden de los eventos. Éstas son dos formas de presentar la estructura de un juego.

Se recomienda al lector el sitio *www.gametheory.net* que contiene gran cantidad de información útil acerca de diversos aspectos de la teoría de juegos.

7.3. Juegos suma cero

La mayor parte de los deportes y los juegos de salón imitan a los conflictos militares. La forma más simple de describirlos es que cuentan con dos jugadores -individuales o equipos- cuyos intereses son totalmente opuestos: la ganancia de un jugador es la pérdida del otro ya que compiten por el mismo premio. A este tipo de situación de conflicto se le denomina **juego suma cero de dos jugadores**. Comenzaremos por este tipo de juego pues es el más simple de describir, al menos en su forma estratégica que es la que mostraremos aquí.

Cuando tomamos decisiones que involucran a otras personas, el problema no es simplemente el que cada quien maximize su utilidad (esperada). La razón es que las decisiones de los demás participantes afectan, a su vez, nuestra decisión óptima. Así, tomamos una decisión previendo que nuestro oponente tomará otra, que a su vez tomará en cuenta la nuestra y así sucesivamente.

Von Neumann demostró en 1928, que para los juegos suma cero entre dos jugadores, existe al menos una pareja de estrategias que da solución al juego. Estas estrategias cumplen con el **principio minimax**: *cada jugador escoge la estrategia que minimiza la pérdida máxima.* En otras palabras, queremos perder lo menos posible. Podríamos haber también expresado este principio como su equivalente: el **maximín**, en donde se persigue *maximizar la ganancia mínima.*

Podemos enunciar el principio minimax (maximin) como sigue:

Teorema 7.3.1 (Minimax) *Considerar un juego suma cero entre dos jugadores. Supongamos que las ganancias y pérdidas de los jugadores*[2] *están dadas por las parejas* $(N, -N)$. *Entonces, existirá al menos un número* $N^* \geq 0$, *tal que* $-N^*$ *minimiza la pérdida máxima de un jugador y* N^* *maximiza la ganancia mínima del otro. A la pareja* $(N^*, -N^*)$ *se le conoce como solución minimax.*

Este tipo de estrategia es la que aplica a múltiples juegos, desde el gato hasta el ajedrez. Así mismo, se utiliza en estrategias militares como la que describimos a continuación.

[2]Si un jugador gana N, el otro pierde la misma cantidad, es decir, gana $-N$.

Durante la segunda guerra mundial, la mitad norte de la isla de Nueva Guinea estaba controlada por el ejercito japonés, mientras que los aliados controlaban el sur. Los aliados tuvieron noticias de que los japoneses enviarían un convoy de ayuda desde la costa Este de la isla vecina de Nueva Bretaña hasta la costa Oeste de Nueva Guinea. Las rutas a elegir eran, la ruta corta por el norte de Nueva Bretaña, o bien, la ruta larga hacia el sur. La ruta norte tomaría 2 días y la sur 3 días.

El objetivo de las fuerzas aliadas era infringir el mayor daño posible al convoy japonés mediante un ataque aéreo, ¡pero primero había que encontrarlo! Tenían que decidir si comenzar la búsqueda hacía el norte o hacía el sur de Nueva Bretaña. Si erraban la dirección, les llevaría casi todo un día cambiar de ruta y encontrar al convoy. Analicemos ahora esta situación desde el punto de vista de la teoría de juegos.

- Se tienen dos jugadores: los aliados y los japoneses.
- Cada día de bombardeo es una ganacia de 1 para los aliados y una pérdida de -1 para los japoneses, es decir un juego suma cero.
- Las estrategias que pueden llevarse a cabo por parte de los aliados son: buscar al norte o buscar al sur.
- Las estrategias que pueden llevarse a cabo por parte del convoy japonés son: zarpar al norte o zarpar al sur.

El juego se presenta en forma estratégica mediante lo que se conoce como una **tabla o matriz de pagos**. Los renglones representan las estrategias para los aliados -jugador renglón- y las columnas para los japoneses -jugador columna-. En cada celda de la tabla hay una pareja de números: el primero representa el **pago** o utilidad para los aliados resultante de la combinación de estrategias correspondiente a la celda. El segundo número, representa la utilidad que derivan los japoneses de la misma pareja de estrategias. En este caso la utilidad se mide en días de bombardeo a favor o en contra.

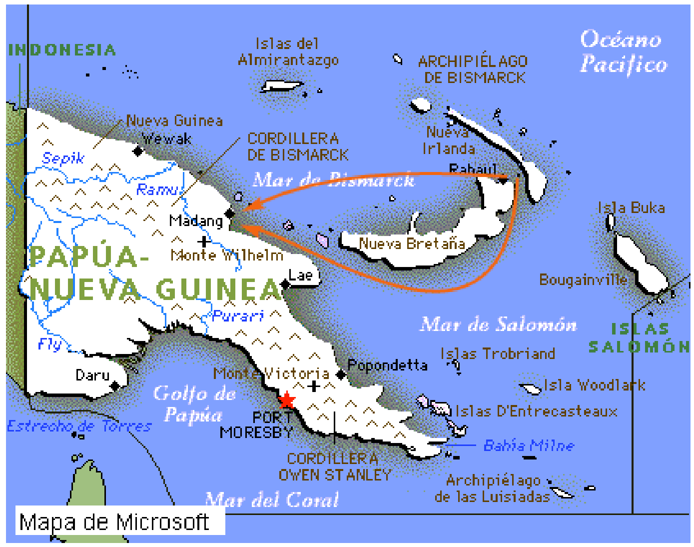
La tabla de pagos asociada que representa los días de bombardeo es la siguiente:

Japón / Aliados	Zarpar al norte	Zarpar al sur
Buscar al norte	(2,-2)	(2,-2)
Buscar al sur	(1,-1)	(3,-3)

Así, si los japoneses zarpan al norte y los aliados buscan al norte, el bombardeo dura 2 días. Si los japoneses zarpan al sur pero los aliados buscan al norte, entonces les llevará un día a los aliados cambiar de ruta y tendrán también 2 días de bombardeo. Si los japoneses zarpan al norte y los aliados buscan al sur, también se pierde un día de bombardeo en el cambio de ruta y sólo pueden bombardear durante un día. Finalmente, si el convoy japonés va al sur al igual que los aliados, el bombardeo dura 3 días.

Japón minimiza sus daños zarpando al norte, evitando así la posibilidad de 3 días de bombardeo en el sur. Los aliados, sabiendo esto, buscan al norte pues así maximizan su ganancia mínima. El equilibrio minimax es buscar y zarpar al norte, resultando en 2 días de bombardeo. Aquí, $N^* = 2$ en la solución minimax.

La situación que acabamos de describir fue lo que tuvo lugar en febrero de 1943 y se conoce en la historia militar como la *batalla del Mar de Bismark.*



La batalla del Mar del Bismark

7.3.0.1. Equilibrio de Nash

Notemos que la solución encontrada para la batalla del mar de Bismark satisface el siguiente concepto de equilibrio: ninguno de los jugadores tiene incentivo de cambiar unilateralmente su estrategia. Ilustremos esto sobre la matriz de pagos como sigue,

Japón / Aliados	Zarpar Norte	Zarpar Sur
Buscar Norte	(2,-2)	(2,-2)
Buscar Sur	(1,-1)	(3,-3)

en donde las flechas indican la dirección hacia la cual hay un incentivo para moverse.

El jugador renglón (aliados) busca cambiar hacia la estrategia que le proporcione mayor utilidad (flechas verticales) y lo mismo pasa con el jugador columna (japón), para el cual las flechas horizontales indican la estrategia preferida -la flecha $\longleftrightarrow$ denota indiferencia entre dos estrategias-. Observamos que en la pareja (2,-2) correspondiente a las estrategias (buscar al norte, zarpar al norte), ninguno de los participantes tiene incentivo para cambiar de estrategia. En este caso se obtiene un equilibrio con estrategias puras, es decir, se busca y se zarpa al norte con probabilidad igual a 1.

La noción de equilibrio que satisface la solución minimax fue precisada por Nash y es, probablemente, el concepto más importante en la teoría de juegos. Su definición intuitiva es la siguiente:

Definición 7.3.2 (Equilibrio de Nash) *Considérese un juego de dos personas y denotemos por (a,b) a todas las parejas de estrategias posibles. Aquí, "a" corresponde a la estrategia del primer jugador y "b" a la del segundo. Una pareja de estrategias (a^*,b^*) forma un* ***equilibrio de Nash*** *del juego, si cada una de ellas es la mejor respuesta para la otra.*

Esta definición puede extenderse para cualquier número n de jugadores. Un equilibrio de Nash, no necesariamente es la pareja de estrategias que proporciona la máxima utilidad a cada jugador. Sin embargo, tiene la característica de que los jugadores no tienen incentivo alguno para desviarse unilateralmente de este equilibrio. En otras palabras, ningún jugador se arrepiente de seguir la estrategia de equilibrio de Nash.

Nash extendió el resultado de Von Neumann para cualquier juego entre dos personas, es decir, demostró el siguiente teorema:

Teorema 7.3.3 *Considérese un juego de dos personas y denotemos por (a,b) a todas las parejas de estrategias posibles. Entonces, existe al menos una pareja de estrategias (a^*,b^*) que forma un equilibrio de Nash del juego.*

7.4. El dilema del prisionero

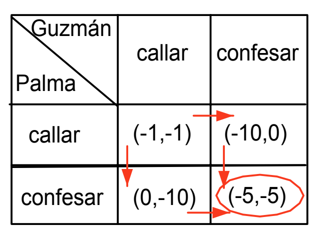
Quizás uno de los juegos más famosos es el llamado *Dilema del Prisionero.* Éste fue desarrollado por M. Flood y M. Dresher de la Rand Corporation en 1950. Se trata de un juego no cooperativo que no es suma cero, cuya descripción se plantea a continuación.

Dos individuos, llamados Palma -alias “El Güero”- y Guzmán -alias “El Chapo”-, cometen un crimen y son aprehendidos y puestos en confinamiento, incomunicados entre sí. El procurador les ofrece, a cada uno por separado, el siguiente trato:

Pueden o no confesar al crimen; si uno de ellos confiesa y el otro no, entonces el que confesó sale libre y su compañero tiene una sentencia de 10 años de prisión. Si ambos confiesan, entonces se les da una sentencia reducida de 5 años a cada uno. Finalmente, si ninguno confiesa, sólo se les puede acusar de un crimen menor y cada uno recibe un año de cárcel. Analicemos la matriz de pagos correspondiente en donde los pagos corresponden a los años de carcel.

Palma \ **Guzmán**	Callar	Confesar
Callar	(-1,-1)	(-10,0)
Confesar	(0,-10)	(-5,-5)

El siguiente diagrama nos indica las estrategias que buscan los jugadores para perseguir el mejor resultado posible:



Independientemente de lo que haga Guzmán, lo mejor que puede hacer Palma es confesar y se tiene la situación análoga para Guzmán. Así, su estrategia óptima es también confesar. Aquí confesar es una **estrategia dominante** para ambos ya que es mejor que callar en cualquier caso. De esta forma, (confesar, confesar) representa las estrategias de equilibrio con pagos de (-5,-5), es decir, los dos detenidos van 5 años a la cárcel.

Observando la tabla, es claro que ambos individuos estarían mejor con las estrategias (callar, callar) con resultado de (-1,-1), puesto que sólo irían un año a prisión. Sin embargo, en la ausencia de cooperación o cualquier esquema de coordinación entre Palma y Guzmán -ninguno confía en que el otro calle-, el resultado es que ambos confiesan y ninguno se arriesga a que el otro lo delate.

El mismo dilema que tienen los *prisioneros*, surge en una gran cantidad de situaciones de la vida real. En este tipo de juego, la falta de cooperación o la desconfianza hacia los demás nos conduce a un equilibrio que no es el deseable. Esto podría solucionarse si hubiese comunicación y alguna forma legal para garantizar que los jugadores cumplieran sus promesas. Es curioso que en ocasiones, el concepto de que ayudando a los demás nos ayudamos a nosotros mismos, no es algo natural. Nuestro espíritu de competencia nos hace pensar que todos los juegos son suma cero de manera que la ganancia del otro será inevitablemente nuestra pérdida. El dilema del prisionero nos demuestra que esto no siempre es así.

Un ejemplo clásico es el que involucra *bienes o lugares públicos*. Para ilustrar, supongamos que en una calle de la ciudad de México hay varios puestos ambulantes de comida en la banqueta. Al final del día, se acumulan las bolsas de basura, mismas que deberían desecharse en algún basurero. De no ser así, la banqueta se convertirá en un lugar sucio al cual nadie querrá ir a comer y todos los ambulantes pierden. Si todos los vendedores ambulantes recogen un poco de basura, el esfuerzo es menor y gozan de un espacio limpio al día siguiente. Si sólo unos cuantos hacen el trabajo, tendrán que realizar un esfuerzo muy grande para limpiar el lugar. Aquí, la banqueta es el bien público, es decir, pertenece a la sociedad en general y no a un individuo.

Para simplificar examinemos el caso de dos vendedores llamados v1 y v2, en el cual la matriz de pagos está dada por,

v2 \ **v1**	limpiar	no limpiar
limpiar	(5, 5)	(-1,6)
no limpiar	(6,-1)	(0, 0)

(7.1)

Notemos que cuando un vendedor limpia y el otro no, el que limpia obtiene un pago de -1, debido al esfuerzo que realiza, mientras que el otro obtiene un pago de 6, pues disfruta de los beneficios sin esfuerzo alguno. A semejanza del dilema del prisionero, la estrategia de "no limpiar" es dominante para ambos jugadores y el equilibrio de Nash es el (0,0).

En el ejemplo anterior, el equilibrio obtenido es racional desde el punto de vista individual, sin embargo, evidentemente ambos jugadores (la sociedad) estarían mejor si limpiaran cada quién su parte. Aunque este resultado sería racional desde el punto de vista colectivo, no puede llegarse a él a menos que exista un mecanismo institucional que los obligue a limpiar, por ejemplo, imponiendo multas.

Una forma general de expresar el juego del dilema del prisionero es pensar en las estrategias como, cooperar y no cooperar, de manera que el equilibrio de Nash es aquel en el cual ninguno de los jugadores coopera. Ciertamente no es muy alentador el tener este equilibrio, cuando sabemos que la cooperación mutua nos podría llevar a un mejor resultado. Sin embargo, no todo está perdido cuando el juego se repite más de una vez. A este juego se le conoce como el **dilema del prisionero iterado (DPI)** y consiste en que los jugadores tienen memoria de las estrategias que se jugaron en las versiones pasadas del juego.

El politólogo norteamericano, Robert Axelrod (1943 -) interesado en los incentivos a la cooperación, organizó torneos entre la comunidad académica para proponer estrategias para competir en juegos del DPI. La estrategia ganadora fue diseñada por el matemático ruso Anatol Rapoport (1911 - 2007) y es conocida como **Tit for Tat** o **Uno por Otro** Ésta consiste en que el jugador coopera la primera vez que se realiza el juego y posteriormente elige la misma acción que su oponente tomó en la versión anterior del juego. Esta estrategia es lo que se conoce como una estrategia **amable** pues el jugador que la sigue nunca será el primero en no cooperar, y **clemente** pues los jugadores están dispuestos a perdonar.

7.5. Coordinación y gallinas

Podemos estudiar variantes del dilema del prisionero al considerar la siguiente matriz de pagos para los jugadores C (columna) y R (renglón) que tienen estrategias puras a elegir α y β:

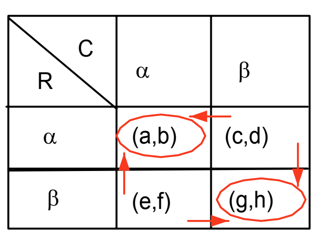
$R \backslash C$	α	β
α	(a,b)	(c,d)
β	(e,f)	(g,h)

Veamos algunos casos comunes en los cuales se obtienen equilibrios de estrategias puras[3].

- Si $e > a, g > c, d > b$ y $h > f$, la estrategia β es dominante para ambos jugadores y el equilibrio es (β, β), obteniendo pagos (g, h). Este es simplemente el caso del dilema del prisionero visto anteriormente.

- Si $a > e, c > g, b > d$ y $f > h$, la estrategia α es dominante para ambos jugadores y el equilibrio es (α, α), obteniendo pagos (a, b). En realidad, por simetría, este es el mismo caso del dilema el prisionero, con ambos jugadores compartiendo la misma estrategia dominante.

- Si $a > e, g > c$, $b > d$ y $h > f$, existen dos equilibrios: (α, α) con pagos (a, b) y (β, β) con pagos (g, h). A este tipo de juego se le conoce como juego de coordinación pues los equilibrios corresponden a jugar estrategias iguales. La siguiente figura ilustra la situación descrita:



Coordinación

[3] Los equilibrios de estrategias mixtas se tratarán en la siguiente sección.

- Si $c > g, e > a,\ f > h$ y $d > b$, existen dos equilibrios: $(\ \beta, \alpha)$ con pagos (e, f) y (α, β) con pagos (c, d). A este tipo de juego se le conoce como juego de no coordinación pues los equilibrios corresponden a jugar estrategias diferentes. Esto se ilustra en la siguiente figura:

R \ C	α	β
α	(a,b)	(c,d)
β	(e,f)	(g,h)

No coordinación

Como ejemplo de juego de coordinación está el siguiente que se conoce como *batalla de los sexos*:

Carlos y Rosita están tratando de decidir si van a un partido de futbol o a la ópera. La matriz de pagos está dada por,

Rosita \ Carlos	Ópera	Futbol
Ópera	$\boxed{\mathbf{(2,1)}}$	(0,0)
Futbol	(0,0)	$\boxed{\mathbf{(1,2)}}$

De aquí, Carlos prefiere que Rosita lo acompañe al futbol y Rosita prefiere que Carlos la acompañe a la ópera, pero a ninguno de ellos le gusta ir solo a cualquiera de los eventos.

Notemos que si Rosita va a la ópera, entonces la mejor respuesta para Carlos es acompañarla. Lo mismo pasa si Carlos va al futbol. De aquí que este juego tenga dos equilibrios de Nash dados por dos estrategias: que ambos vayan a la ópera o bien que ambos vayan al futbol. Se trata de un juego de coordinación. El problema de tener más de un equilibrio es que no sabemos como escoger entre ellos. Adicionalmente, no hay comunicación entre Carlos y Rosita antes de realizar su elección. El tener más de un equilibrio puede pensarse como el tener más de una solución en algún sistema de ecuaciones representando un problema real. Algunas soluciones tendrán sentido y otras no, dependiendo del problema en cuestión.

Si el equilibrio es ir siempre a la ópera, entonces Rosita obtiene mayor beneficio que Carlos, asimismo, si siempre van al futbol, Carlos es el que obtiene un pago mayor. Es claro que ésta es una deficiencia de ambos equilibrios pues no son *justos* para uno de los jugadores.¿Qué será lo mejor para Carlos y Rosita?

Cualquier promesa que ambos realicen antes de tomar su decisión, tendrá que resultar en uno de los dos equilibrios. De otra forma, decimos que se trata de una **promesa vacía** pues no existe incentivo alguno para cumplirla. Adicionalmene, si Rosita ya compró boletos para la ópera y Carlos amenaza con ir al futbol, ésta será una **amenaza increíble**, ya que evidentemente prefiere ir a la ópera con Rosita. Ambos pueden anunciar a qué evento piensan ir antes de tomar una decisión. A este tipo de comunicación se le conoce como **plática barata o parloteo**: no tiene ningún costo ni compromete a los jugadores a seguir una estrategia definida, sin embargo, puede ayudar a coordinar de manera que ambos se beneficien.

Supongamos que Carlos y Rosita deciden dejar la elección al azar, en este caso pueden determinar lanzar una moneda para elegir si ambos van a ir a la ópera o al futbol. Por ejemplo, *águila* podría determinar ir a la ópera y *sol* al futbol. Así, la utilidad esperada sería,

$$\frac{1}{2}(1) + \frac{1}{2}(2) = \frac{3}{2},$$

para cada uno, lo cual parece ser una solución justa para ambos. Éste resulta ser una extensión al equilibrio de Nash y se denomina **equilibrio correlacionado**. Una vez que Carlos y Rosita observan la cara que resulta de lanzar la moneda, ninguno de los dos tiene incentivo de cambiar su estrategia propuesta puesto que esto resultaría en la falta de coordinación con el otro (jugador) y en un pago menor.

El ejemplo análogo al de la batalla de los sexos para juegos de no coordinación es el juego conocido como *gallina.* Este juego se inmortalizó con James Dean en la película clásica: "Rebelde sin Causa", en la cual dos jóvenes -Jim y Buzz- lanzan sus automóviles, a toda velocidad, uno contra el otro. El que se desvía antes del impacto inminente pierde alguna apuesta y, ¡por supuesto su dignidad ya que será llamado gallina!

Veamos la matriz de pago para Jim y Buzz:

Jim \ **Buzz**	desviarse	no desviarse
desviarse	(0,0)	$\boxed{\textbf{(-1,2)}}$
no desviarse	$\boxed{\textbf{(2,-1)}}$	(-100,-100)

Éste corresponde a un juego de no coordinación con dos equilibrios en los cuales uno de los jóvenes se desvía y el otro no.

Al igual que en la *batalla de los sexos,* estos equilibrios no parecen ser justos para el jugador que queda como *gallina.* Sin embargo, en este juego también existe un equilibrio correlacionado que corresponde a que un agente externo eliga aleatoriamente entre las estrategias:

(desviarse, desviarse),
(desviarse, no desviarse),
(no desviarse, desviarse),

cada una con probabilidad de $\frac{1}{3}$. El agente externo le comunica a cada jugador, por separado, la acción que debe tomar, es decir los jugadores no saben lo que hará el contrincante. Notemos que si el agente externo le dice a Jim que no se desvíe, éste no tendrá incentivo para desviarse pues sabe que Buzz lo hará. Asimismo, si le pide que se desvíe, entonces Buzz puede desviarse o no con probabilidad $\frac{1}{2}$ y la utilidad esperada de Jim al desviarse es

$$EU(desviarse) = \frac{1}{2} \times 0 + \frac{1}{2}(-1) = -\frac{1}{2}.$$

Si cambia de estrategia y no se desvía, entonces su utilidad esperada es

$$EU(\text{no desviarse}) = \frac{1}{2} \times 2 + \frac{1}{2}(-100) = -49$$

que es mucho menor que la anterior y por lo tanto no tiene incentivo para cambiar.

Dado que el juego es simétrico, lo que acabamos de describir también aplica para Buzz y esta selección aleatoria de estrategias por un agente externo es un equilibrio. Cada uno de los participantes obtiene una utilidad esperada de

$$\frac{1}{3} \times 0 + \frac{1}{3} \times (-1) + \frac{1}{3} \times 2 = \frac{1}{3}.$$

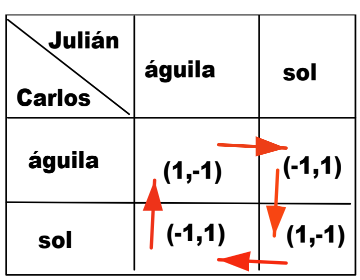
7.6. Estrategias mixtas

Consideremos el siguiente juego suma cero: Carlos y Julián tienen cada uno monedas de un peso, mismas que deberán colocar sobre la mesa. Si las caras coinciden, Carlos se queda con ambas monedas, si las caras difieren, entonces Julián se queda con las dos monedas. Este juego se conoce como el juego de *hacer coincidir las monedas*. Las estrategias son elegir alguna de las caras de la moneda. Supongamos que la utilidad es simplemente la ganancia o pérdida monetaria asociada.

La forma estratégica de este juego por medio de su matriz de pagos es la siguiente:

Carlos \ **Julián**	águila	sol
águila	(1,-1)	(-1,1)
sol	(-1,1)	(1,-1)

en donde el primer número de cada pareja representa el pago de Carlos y el segundo el de Julián. Si analizamos los incentivos de los jugadores para buscar su mejor opción, observamos que no hay una estrategia de equilibrio. Veamos el diagrama:



Carlos sigue la dirección que indican las flechas verticales y Julián sigue las horizontales. No existe una pareja de estrategias para la cual ambos carezcan de incentivo para cambiar de estrategia. Sin embargo, según Von Neumann debe existir una solución que cumpla con el principio minimax y de acuerdo a Nash, ésta es un equilibrio de Nash. ¿Cuál es dicha solución? En ocasiones como ésta, no hay parejas de estrategias puras que nos proporcionen una solución de equilibrio del juego. Para encontrar la(s) solución(es) es necesario buscar en el conjunto de estrategias mixtas.

¿Qué se quiere decir con una estrategia mixta? Simplemente que, de forma aleatoria, utilizaremos el conjunto de estrategias disponibles, cada una con alguna probabilidad. Como se ilustró en la sección 7.2, el mezclar distintos tipos de lanzamiento es una estrategia mixta para un pitcher. Una observación importante es que, a diferencia de los equilibrios correlacionados de la sección 7.5, la adopción de una estrategia mixta no requiere de agentes externos para su elección.

En el juego de las monedas descrito arriba, se intuye que la mejor estrategia es utilizar águilas y soles aleatoriamente, con probabilidad de $\frac{1}{2}$. Para obtener formalmente el resultado se procede como sigue:

- Se denotan las proporciones o probabilidades con las cuales cada jugador utilizará sus estrategias. En este caso, las probabilidades o proporciones de águilas elegidas por Carlos y Julián se denotan por p y q, respectivamente. En consecuencia, $1-p$, $1-q$ denotan las proporciones de soles que cada uno elige.

- Se calcula la utilidad esperada de los jugadores en términos de las probabilidades anteriores. Esta utilidad es simplemente, la suma sobre todos los resultados posibles de la probabilidad de que ocurra el resultado por el pago al jugador.

- Un equilibrio será una pareja (p^*, q^*), para la cual p^* es la mejor opción dado q^* y q^* es la mejor opción dado p^*.

- La interpretación de (p^*, q^*) es que Carlos utilizará águilas en proporción p^* y Julián en proporción q^*, en forma aleatoria.

En este caso, la utilidad esperada para Carlos $EUC(p)$ está dada por,

$$\begin{aligned} EUC(p) &= pq(1) + p(1-q)(-1) + (1-p)q(-1) + (1-p)(1-q)(1) \\ &= 2(2q-1)p + (1-2q) \end{aligned}$$

y la utilidad esperada para Julián es,

$$\begin{aligned} EUJ(q) &= pq(-1) + p(1-q)(1) + (1-p)q(1) + (1-p)(1-q)(-1) \\ &= 2(1-2p)q + (2p-1). \end{aligned}$$

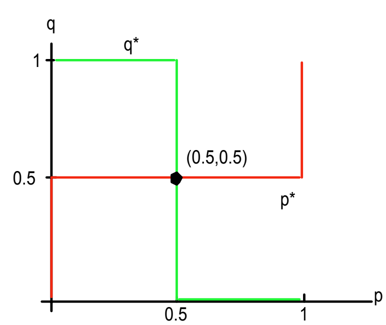
Carlos elige p^* tomando a q como un valor externo dado. La elección óptima es la siguiente

$$p^* = \begin{cases} 1, \text{ si } q \text{ es mayor que } \frac{1}{2}, \\ \text{cualquier valor entre 0 y 1, si } q = \frac{1}{2}, \\ 0, \text{ si } q \text{ es menor que } \frac{1}{2}. \end{cases}$$

La razón es que esta elección ocasiona que el término $2(2q-1)p$, -el único término que contiene a p-, sea lo más grande posible. De la misma forma, las elección óptima de q^* para Julián es,

$$q^* = \begin{cases} 0, \text{ si } p \text{ es mayor que } \frac{1}{2}, \\ \text{cualquier valor entre 0 y 1, si } p = \frac{1}{2}, \\ 1, \text{ si } p \text{ es menor que } \frac{1}{2}. \end{cases}$$

Para que ambos realicen simultáneamente una elección óptima graficamos en el plano pq las estrategias p^* y q^*, obteniendo:



De la figura anterior se observa que cuando $p^* = q^* = 0.5$, cada estrategia es la mejor respuesta para la otra, de manera que ésta es la solución de equilibrio.

Jugar águila y sol aleatoriamente con probabilidad 0.5, no quiere decir que si el juego se repite diez veces, cada jugador elija 5 águilas y 5 soles. Más bien, cada jugador debe tener un mecanismo que elija aleatoriamente águilas y soles con probabilidad 0.5. En este caso, los jugadores podrían simplemente lanzar su moneda al aire y jugar con la cara que resulte.

Dado que se trata de un juego suma cero, deberíamos poder expresar el resultado en términos de una solución minimax. En efecto, dadas $p^* = q^* = 0.5$, las utilidades esperadas de Carlos y Julián son,

$$EUC(0.5) = EUJ(0.5) = 0,$$

es decir $N^* = 0 = -N^*$.

Los juegos de la batalla de los sexos y la gallina poseen equilibrios de estrategias mixtas, adicionalmente a los equilibrios descritos de estrategias puras. Por ejemplo, en el caso de Carlos y Rosita si p y q denotan, respectivamente, la proporción de veces que Rosita y Carlos eligen ir a la ópera, entonces las utilidades esperadas de Rosita y Carlos están dadas por,

$$\begin{aligned} EUR(p) &= pq2 + p(1-q)0 + (1-p)q0 + (1-p)(1-q)1 \\ &= (3q-1)p + (1-q), \\ EUC(q) &= pq1 + p(1-q)0 + (1-p)q0 + (1-p)(1-q)2 \\ &= (3p-2)q + 2(1-p). \end{aligned}$$

Rosita elige p^* de manera que su utilidad esperada sea la mayor posible para q dado, así se tiene[4],

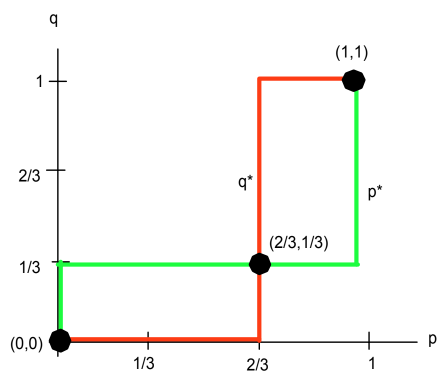
$$p^* = \begin{cases} 1, \text{ si } q \text{ es mayor que } \frac{1}{3}, \\ \text{cualquier valor entre 0 y 1, si } q = \frac{1}{3}, \\ 0, \text{ si } q \text{ es menor que } \frac{1}{3}. \end{cases}$$

[4]El único término sensible a la elección de p es $(3q-1)p$.

Análogamente, Carlos elige q^* :

$$q^* : \begin{cases} 1, \text{ si } p \text{ es mayor que } \frac{2}{3}, \\ \text{cualquier valor entre 0 y 1, si } p = \frac{2}{3}, \\ 0, \text{ si } p \text{ es menor que } \frac{2}{3}. \end{cases}$$

Al igual que en el caso anterior podemos visualizar ambas estrategias obteniendo,



De la figura tenemos tres equilibrios: los dos anteriores de estrategias puras en los que $p^* = q^* = 0$ y ambos van al futbol o bien $p^* = q^* = 1$ y ambos van a la ópera. El nuevo equilibrio de estrategias mixtas está dado cuando $p^* = \frac{2}{3}$ y $q^* = \frac{1}{3}$ o bien $1 - q^* = \frac{2}{3}$, es decir, ambos eligen su evento preferido con el doble de frecuencia. En la ausencia de agentes externos, estos son los tres equilibrios posibles.

¿Cómo se compara la estrategia mixta que acabamos de encontrar con el equilibrio correlacionado de la sección 7.5? Veamos, para las estrategias mixtas la utilidad esperada de Carlos y Rosita es:

$$\frac{2}{3} \times \frac{1}{3} \times 2 + \frac{2}{3} \times \frac{1}{3} \times 1 = \frac{2}{3}$$

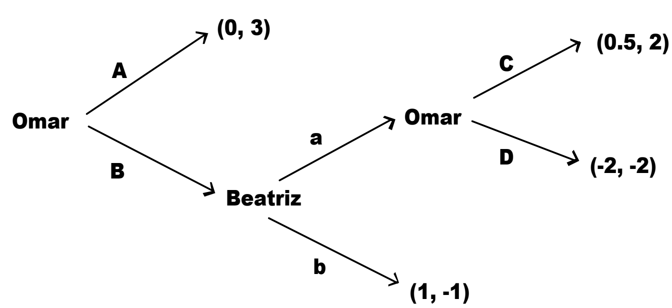
que es menor que la utilidad esperada del equilibrio correlacionado de $\frac{3}{2}$. La conclusión es que los participantes de este juego se benefician del mecanismo aleatorio externo que propicia el equilibrio correlacionado.

7.7. Juegos en forma extensiva

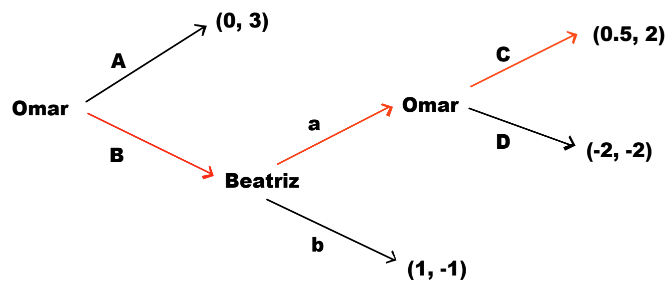
Hasta ahora todos nuestros ejemplos han sido de juegos en forma normal o estratégica. Para representar un juego en forma extensiva se utiliza un **árbol de juego**, en lugar de matrices de pago. Las ramas representan las acciones posibles, los nodos a cada jugador y los nodos terminales los pagos de los jugadores. En un juego de información perfecta, los jugadores saben en que parte del árbol se encuentran en todo momento. Una estrategia será el conjunto de todas las acciones que tome un jugador. El ejemplo que se describe a continuación ilustra el procedimiento.

Omar y Beatriz realizan el siguiente juego: primero Omar puede elegir entre las estrategias **A** o **B**. Si elige **A**, el juego se termina y los pagos son 0 para Omar y 3 para Beatriz. Si elige **B**, entonces el turno es de Beatriz quien puede elegir **a** o **b**. Si elige **b**, el juego termina y los pagos son 1 para Omar y -1 para Beatriz, si elige **a**, entonces el turno siguiente es de Omar. Sus elecciones posibles son **C** o **D**. Si elige **C**, el pago final para Omar es de 0.5 y para Beatriz de 2. Si elige **D**, los pagos finales son de -2 para ambos.

Las consideraciones anteriores se ilustran en el árbol de juego a continuación, en donde el primer miembro de cada pareja de pagos corresponde a la ganancia de Omar y el segundo a la de Beatriz.



¿Cómo se resuelve este juego? Las acciones iniciales de los jugadores dependen de lo que hacen los demás posteriormente en el juego. Así, para encontrar estrategias de equilibrio se utiliza el procedimiento de **inducción inversa.** Esto significa que comenzamos al final del juego y vamos encontrando estrategias óptimas hasta llegar al comienzo del árbol. En este caso, en los últimos nodos correspondientes a Omar, se tiene que 0.5 > -2, por lo que escogería **C** sobre **D**. En el paso anterior, Beatriz, sabiendo que Omar jugará **C**, escoge **a** puesto que 2 > -1. Finalmente, en la primera etapa Omar, sabiendo todo lo anterior, escoge **B** puesto que 0.5 > 0. Ilustramos con líneas rojas las acciones de cada jugador en la siguiente figura:



De lo anterior, la estrategia seguida por Omar es **B** seguida de **C** y Beatriz sigue **a**. Puede demostrarse que las estrategias obtenidas por inducción inversa constituyen un equilibrio de Nash para el juego[5]. Observemos que este método funciona puesto que ambos jugadores saben exáctamente en donde se encuentran en el árbol y tienen la misma información. De no ser así, tendríamos un juego de información imperfecta. Para este tipo de juegos existe un refinamiento de la inducción inversa que es la **perfección de subjuego**, sin embargo, aquí no nos ocuparemos de dicho caso.

7.8. Juegos biológicos

En 1972 John Maynard Smith (1920 - 2004) introdujo el concepto de juegos evolutivos, aplicando la teoría de juegos a la biología. Maynard Smith definió lo que se llama una estrategia *evolutivamente estable*, como aquella que garantiza la supervivencia de una población en el siguiente sentido: si la mayoría de los individuos adoptan la estrategia, entonces ningún individuo con una estrategia nueva -llamado *mutante*- será exitoso. Uno de los ejemplos más conocidos es el del *halcón y la paloma* que va como sigue.

[5]Sin embargo, no todo equilibrio de Nash puede obtenerse por inducción inversa.

Se tiene una población con dos tipos de individuos que pueden comportarse como *halcones* o *palomas*. Estos compiten por un recurso limitado que tiene un valor fijo en términos de bienestar, digamos V. Los halcones siempre eligen mostrar agresividad y posteriormente no paran hasta que son lastimados o el oponente se retira. Las palomas, en cambio, se retiran inmediatamente si el oponente muestra agresividad. Podemos imaginar que se trata de una población en donde los individuos pueden adoptar dos tipos de comportamiento: agresivo (halcón) o pasivo (paloma). La especie biológica puede ser cualquiera, bacterias, aves, humanos o extraterrestres.

Adicionalmente, se tienen los siguientes supuestos:

1. Si dos individuos muestran agresividad, siempre terminan peleando y ambos se lastiman.

2. El costo del conflicto se cuantifica por una pérdida de bienestar fija, digamos C.

3. Cuando un halcón se encuentra con una paloma, la paloma inmediatamente se retira y el halcón obtiene todo el recurso.

4. Cuando dos palomas se encuentran, comparten el recurso.

Si los individuos son R (renglón) y C (columna), la tabla de pagos queda como,

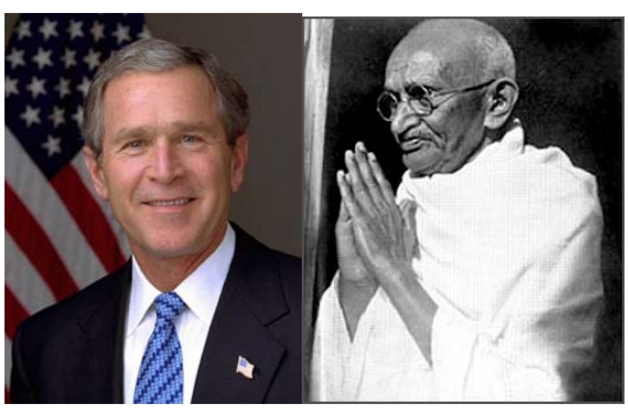
R \ C	Halcón	Paloma
Halcón	$\left(\frac{1}{2}(V-C), \frac{1}{2}(V-C)\right)$	$(V, 0)$
Paloma	$(0, V)$	$\left(\frac{V}{2}, \frac{V}{2}\right)$

Así, si dos halcones se encuentran, acabarán peleandose y cada uno termina con una cantidad $\frac{1}{2}(V-C)$ de bienestar. Si dos palomas se encuentran, comparten el recurso y terminan cada una con $\frac{V}{2}$ unidades de bienestar y finalmente, si se encuentran un halcón y una paloma, el halcón se queda con todo el bienestar V y la paloma con 0.

La estrategia de comportarse como paloma resulta no ser evolutivamente estable, ¿porqué? Si en una población de palomas surge un individuo mutante que adopta la estrategia del halcón, entonces se quedará con todo el recurso y las palomas acabarán desapareciendo. Si se cumple que $V > C$ -el valor del recurso es mayor al costo de pelear-, entonces la estrategia de comportarse como halcón es evolutivamente estable.

Alternativamente, si $V < C$, entonces se requiere de una estrategia mixta para ser evolutivamente estable. En este caso, la estrategia consiste en comportarse como halcón con probabilidad $p = \frac{V}{C}$ y como paloma con probabilidad $1 - p$. Una forma de interpretar esto es que un porcentaje p de la población se comporte como halcón y el resto como palomas. Por ejemplo, si el recurso tuviese un valor de 5 y el costo fuese 10, entonces $p = \frac{5}{10} = \frac{1}{2} = 0.5$. En este caso la mitad de la población se comportaría como halcón y la otra mitad como paloma.

En términos de la teoría de la evolución, este tipo de análisis explica la existencia de varios tipos de comportamiento en una población. El ejemplo que acabamos de analizar nos muestra que para que esta población hipotética sea exitosa en términos de supervivencia, debe existir algún porcentaje de individuos agresivos. Sin embargo también puede utilizarse un análisis semejante para explicar la existencia de la variedad genética dentro de una misma especie, el porqué de la desaparición de algunas características o especies y la dominancia de otras. De esta forma la teoría de juegos se aplica a la biología de la evolución, al comportamiento animal y a la genética de poblaciones.



Halcones y palomas en la especie humana.

7.9. Las dificultades de cooperar

Todos los juegos que se han discutido hasta el momento han sido no cooperativos. Aunque hubiese comunicación previa al juego entre los participantes, éstos no pueden obligarse a cumplir una promesa, es decir, se trata de simple parloteo (*cheap talk*). En el caso de los juegos cooperativos, no sólo se permite la comunicación, sino que existen mecanismos externos para obligar a los jugadores a cumplir sus promesas y acuerdos. Cuando dos o más jugadores cooperan entre sí, decimos que forman una **coalición**.

Por ejemplo, en el caso del dilema del prisionero, los jugadores pueden obligarse a cooperar mediante un agente externo que les cobre multas o imponga algún otro tipo de castigo si no lo hacen; el resultado será que ambos jugadores se benefician de este arreglo. Los seres humanos han buscado este tipo de agente externo mediante la imposición de convenciones, reglas, leyes y principios religiosos, la convivencia no sería posible de otra forma.

El análisis de los juegos cooperativos es más complejo debido a la interacción entre los jugadores. Aquí nos restringimos a analizar algunos casos sencillos. Ilustremos con un ejemplo.

Una propuesta de ley se pondrá a votación en el Congreso de la Unión. Supongamos que el partido A representa al gobierno federal y el partido B representa a todos los partidos de oposición. La ley puede ser rechazada o aceptada por cada uno de estos partidos y la única forma de que la iniciativa prospere es que ambos la acepten. La matriz de pagos está dada como sigue:

B \ **A**	rechazar	aceptar
rechazar	$\boxed{\mathbf{(5,1)}}$	(0,-1)
aceptar	(0,0)	$\boxed{\mathbf{(3,\ 6)}}$

Este es un juego de coordinación en donde los dos equilibrios están dados cuando ambos partidos rechazan o bien, cuando los dos aceptan la iniciativa.

Observemos que cuando ambos aceptan la iniciativa, el total de los pagos o pago agregado, es de $3 + 6 = 9$. Sin embargo, esto no sería relevante para una negociación si los pagos finales no pueden reasignarse entre los participantes. Decimos que un juego cooperativo acepta **transferencia de**

utilidad (TU) si es posible repartir el total de los pagos finales de cualquier forma entre los jugadores. Se asume que existe un mecanismo externo que hace posible la repartición y obliga a los participantes a cumplir sus promesas.

Suponiendo que existe transferencia de utilidad, evidentemente es mejor repartir 9 unidades que 5 + 1 = 6 que son las que se obtienen si ambos rechazan la iniciativa. En la negociación, los partidos realizan propuestas y contrapropuestas amenazando con cambiar su estrategia si no se llega a un acuerdo satisfactorio. Asumiendo que ambos son racionales, se debe llegar a un arreglo pues éste les proporciona mayor utilidad.

En este caso, los partidos se dan cuenta que pueden repartirse 9 unidades de utilidad aceptando la iniciativa. El partido **A** puede ofrecer la transferencia de 1.5 unidades a manera de que ambos tengan 4.5 unidades de utilidad al final: 3 + 1.5 para **B** y 6 - 1.5 para **A**. Sin embargo, el partido **B** podría rechazar esta oferta, amenazando con votar en contra de la iniciativa a menos que se le transfieran 2 unidades para tener un total de 5. Esto significaría que el partido **A** tendría un pago final de 6 - 2 = 4.

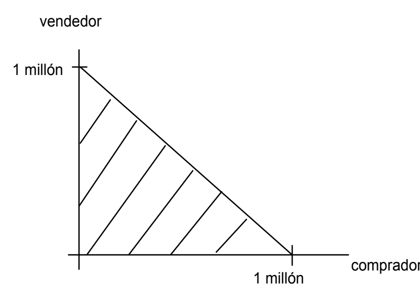
Para desgracia del partido **A**, esta amenaza es creible pues si el partido **B** rechaza la iniciativa el partido A no puede amenzar con votar a favor puesto que obtendría un pago de -1 en lugar de 1 -el pago para A cuando ambos rechazan-. Como el partido **B** obtiene un pago de 5 cuando ambos rechazan, no aceptará nada menor que esa cantidad a cambio de su voto a favor. En resumen, al final de la negociación, ambos partidos aceptarían la iniciativa y el partido **A** transferiría al **B** dos unidades de su pago final.

7.9.1. Repartiendo un pastel

Supongamos que un vendedor y un comprador están negociando la venta de una casa. Asumamos que el vendedor valora su casa en 2 millones de pesos y el comprador la valora en 3 millones. La valoración del comprador debe ser mayor o igual a la del vendedor para que una transacción sea posible. ¿En qué consiste la negociación?

La diferencia entre las valoraciones del vendedor y del comprador es, en este caso, 3 - 2 = 1 millón; éste es el *excedente* potencial que puede repartirse entre los jugadores si ambos acuerdan negociar una transacción. Este excedente no es dinero real -en este caso- sino su equivalente en unidades de utilidad. Si no hay transacción el excedente desaparece. Esencialmente, esta negociación es equivalente a la de repartir un pastel así que imaginemos que dicho excedente es un gran pastel.

En este juego, el pastel puede repartirse de cualquier forma entre los jugadores, cuando esto sucede decimos que el juego cooperativo acepta transferencia de utilidad. Podemos visualizar todos los posibles pagos si ponemos en el eje horizontal la ganancia para el comprador y en el vertical la ganacia del vendedor, como el área triangular sombreada que se muestra a continuación,



Si p y q son, respectivamente, las proporciones del excedente que reciben el comprador y el vendedor, éstas deben cumplir con la restricción

$$p + q \leq 1.$$

El área mostrada es precisamente el conjunto de parejas de pagos (p, q). Notemos que puede desperdiciarse parte o todo el pastel ya que si la transacción no se realiza, $p = q = 0$ y si se realiza *ineficientemente*, $p + q < 1$. El precio de venta determina la repartición del excedente y una vez acordado, los jugadores están obligados, mediante algún contrato, a cumplir su parte del pacto. Tenemos así, que si la casa se vende en 2.8 millones:

- El comprador obtiene un excedente equivalente a la utilidad que le proporcionan 200 mil pesos (se ahorró esta cantidad pues hubiera estado dispuesto a pagar hasta 3 millones).

- El vendedor obtiene un excedente que equivale a la utilidad de 800 mil pesos pues estaba dispuesto a vender por 2 millones.

John Nash resolvió este tipo de problemas de negociación. La solución depende del poder de negociación que posea cada jugador, mismo que puede cuantificarse y de la aversión al riesgo de los jugadores. Supongamos que α y β denotan el poder de negociación del comprador y vendedor, respectivamente. Éstos son números entre 0 y 1, tales que $\alpha + \beta = 1$. Lo más sencillo es suponer que $\alpha = \beta = \frac{1}{2}$, es decir, ningún jugador tiene una ventaja en la negociación. De acuerdo a la sección 6.4, la aversión al riesgo de cada jugador es una medida de que tan cóncava es su función de utilidad. El resultado cualitativo es que

1. Si un jugador tiene mayor poder de negociación, le corresponde una proporción del excedente mayor.

2. Si un jugador es más adverso al riesgo, le corresponde una proporción menor del excedente.

En particular, si ambos tienen el mismo poder de negociación y tienen la misma aversión al riesgo (por ejemplo ambos son neutrales al riesgo), entonces el excedente se reparte en partes iguales y a cada quien le tocan 500 mil pesos. En la vida real, los negociadores tienen incentivo para ocultar su aversión al riesgo (su urgencia por realizar una transacción exitosa) por esta razón.

7.10. Algunos ejemplos

Ejemplos

Ej 7.10.1 Carmen acusa a Rosita de fraude. Carmen puede o no levantar una demanda formal y Rosita puede declararse culpable o inocente. La matriz de pagos es la siguiente, en donde cada número representa millones de pesos:

Rosita \ Carmen	demandar	no demandar
culpable	(-3,3)	(-1,1)
inocente	(1,-1)	$\boxed{\textbf{(0,0)}}$

Notamos que declararse inocente es una estrategia dominante para Rosita. Carmen, conciente de esto, elige no demandar y ambas obtienen un pago de 0. Este es un juego suma cero y el equilibrio cumple con el principio minimax.

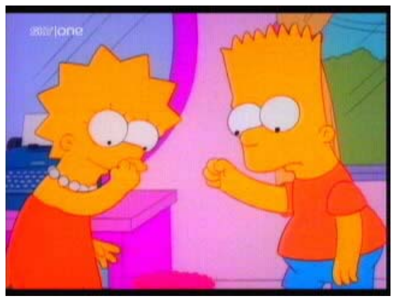
Ej 7.10.2 En la historia *El Problema Final* de Sir Arthur Conan Doyle el maligno profesor Moriarty persigue a Sherlock Holmes con la intención de matarlo. Holmes va en un tren y puede descender en Londres o en Paris mientras que Moriarty debe tomar la decisión de esperarlo en alguna de estas dos estaciones. Si ambos coinciden, Moriarty mata a Holmes, de otra froma Holmes logra escapar. Esta situación puede describirse con un juego como el descrito en la sección 7.6 representado por la siguiente matriz de pagos:

Holmes \ Moriarty	Londres	Paris
Londres	(-1,1)	(1,-1)
Paris	(1,-1)	(-1,1)

Aquí si ambos coinciden en la misma estación, Holmes recibe un pago de -1 (muere) y Moriarty de 1 (por el placer de matarlo). Si cada quien elige una estación diferente, entonces los pagos se invierten. Al igual que en la sección 7.6, no existe un equilibrio de estrategias puras. El equilibrio de estrategias mixtas es aquel en el cual cada jugador lanza una moneda para elegir la estación, es decir, eligen cada una con probabilidad 0.5. Evidentemenete pueden ponerse pagos distintos y así obtener estrategias mixtas diferentes a la descrita.

Ej 7.10.3 El conocido juego de *piedra, papel y tijera*, fue inventado en Japón en donde se le conoce como: "Jan-Ken-Po". La primera referencia que se tiene data de 200 años A.C. Más recientemente, en uno de los episodios de "Los

Simpsons", Bart y Lisa juegan piedra, papel y tijera. Bart, siempre predecible, cree que la "piedra es su amiga y siempre lo hará ganar". Lisa, siempre analítica, razona: "Bart siempre escoge piedra así que escogeré papel". ¡Por supuesto que el resultado es que Lisa gana!

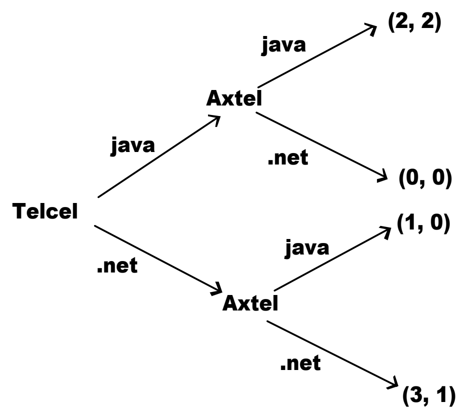


Aquí la solución es una estrategia mixta. Ésta consiste en jugar piedra o papel o tijera con la misma frecuencia y de forma aleatoria. Es decir, jugar cada una con probabilidad 1/3. La razón es que si jugamos con algún patrón reconocible -como lo hacía Bart-, entonces el jugador contrario puede contrarestrar el patrón. La mejor estrategia es que nuestro contrincante no tenga ni la más remota idea de qué vamos a jugar.

Este juego es más famoso de lo que nos imaginamos. Existe una asociación mundial de Piedra, Papel y Tijera y se realizan concursos internacionales periódicamente. Hace un par de años, Takashi Hashiyama, presidente de una compañía de electrónica en Japón, decidió vender su colección de arte valuada en unos 20 millones de dólares. Dos casas subastadoras de gran fama se ofrecieron a vender la colección: Christies y Sotheby's. El señor Hashiyama convocó a los representantes de ambas casas y les dijo que contrataría a aquella que ganara un juego de piedra papel y tijera. La selección se realizó de esa manera. ¡Christie's ganó eligiendo tijeras mientras que Sotheby's perdió eligiendo papel!

Ej 7.10.4 Telcel tomará la decisión de adoptar una tecnología de browser para sus celulares. Posteriormente, Axtel debe decidir si adoptar o no la misma tecnología. Las plataformas posibles para el browser son java y .net.

Supongamos que el juego en forma extensiva está dado por el siguiente árbol:



en donde los pagos de los nodos terminales corresponden, el primer número de la pareja a Telcel y el segundo a Axtel.

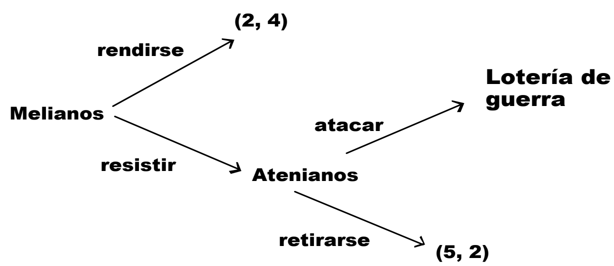
Comenzando por los nodos finales, se tiene que si Axtel adopta java, sus pagos finales son de 2 ó 0, dependiendo de si Telcel adoptó java o .net. Alternativamente, si Axtel adopta .net, entonces sus pagos finales son 0 (si Telcel adopta java) y 1 (si Telcel adopta .net). De esta forma, Axtel obtiene el máximo beneficio adoptando la misma plataforma que Telcel. Sabiendo que los nodos finales serán el (2,2) o el (3,1), Telcel elige .net pues es la tecnología que le reditúa mayor beneficio ($3 > 2$). Así, el equilibrio por inducción inversa es que ambos adopten .net.

Ej 7.10.5 (Diálogo Meliano) El diálogo meliano[6] se refiere a un pasaje de *La historia de las guerras del Peloponeso* del historiador griego de la antigüedad, Thucydides (460 - 395 BC). Durante las guerras del Peloponeso entre Atenas y Esparta, los atenienses invadieron la isla de Melos (416 BC). En el diálogo se describe la oferta que Atenas ofrece a Melos: pagar tributo y sobrevivir o bien enfrentar a Atenas en una guerra y quedar destruidos. Los melianos apelarón al derecho de ser neutrales y asegurando que tanto

[6] Basado en un ejercicio del curso "Theories of International Relations" de Curtis S. Signorino de la Universidad de Rochester.

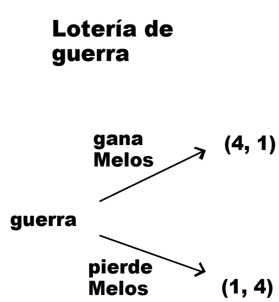
Esparta como los dioses se apiadarían de ellos, ayudándolos a ganar la guerra contra Atenas. El desenlace fue que Melos opuso resistencia, originando un ataque de Atenas que destruyó totalmente a Melos.

La situación descrita puede representarse como un juego en forma extensiva en el cual, si Atenas ataca, el resultado para ambos es una lotería, que es la guerra. El árbol de juego es el siguiente, en donde los pagos representan alguna función de utilidad ordinal:



Supongamos que los melianos creen que ganarán la guerra con probabilidad p y la perderán con probabilidad $1 - p$. Asimismo, los atenienses consideran que sus probabilidades de victoria y derrota son de q y $1 - q$, respectivamente. Ambas probabilidades son totalmente subjetivas y seguramente influenciadas por el análsis militar, algún oráculo, fé en los dioses y confianza en la ayuda de los espartanos.

Supongamos que la lotería que representa ir a la guerra es la siguiente,



Las utilidades esperadas de la guerra para melianos (EUM) y atenienses

(EUA) quedan dadas, respectivamente, por,

$$EUM(guerra) = p4 + (1-p)1,$$
$$EUA(guerra) = q4 + (1-q)1.$$

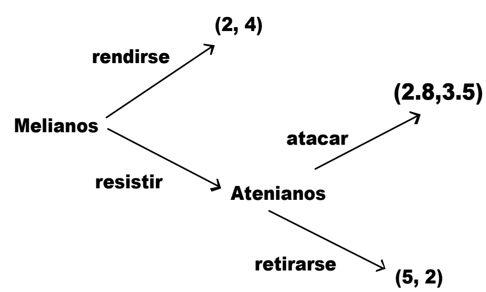
Veamos el resultado para las siguientes probabilidades:

$$p = \frac{3}{5},$$
$$q = \frac{5}{6}.$$

En este caso tenemos que,

$$EUM = \frac{14}{5} = 2.8,$$
$$EUA = \frac{21}{6} = 3.5.$$

De esta forma, el árbol puede completarse como,



En la última etapa $3.5 > 2$ por lo que Atenas ataca. Como $2.8 > 2$, entonces Melos resiste en la primera etapa. En el diálogo, Atenas intenta convencer a Melos que Esparta y los dioses no los ayudarán de manera que si ellos atacan el desenlace sería una victoria de Atenas. En otras palabras, los atenienses tratarón de que Melos redujera su probabilidad de victoria p a menos de $\frac{1}{3}$. En dicho caso, Melos se hubiese rendido inicialmente.

♠♠♠♠♠

Bibliografía Complementaria

1. Douglas G. Baird, Robert H. Gertner and Randal C. Picker, *Game Theory and the Law*, Harvard University Press, Cambridge, MA, 1998.

2. Binmore, Ken, *Fun and Games: A Text on Game Theory*, D.C. Heath and Co, Lexington, MA, 1992.

3. Conan Doyle Arthur, *El Problema Final* puede descargarse en

   www.librosparadescargar.com/ebook−1/1119865782−*el*−*problema*−*final*

4. Dixit Avinash and Skeath Susan, *Games of Strategy*, Norton, 1999.

5. Signorino Curtis S. Material en línea para el curso *Theories of International Relations*:

   http : *//www.rochester.edu/College/PSC/signorino/courses*/272/*default.html*

6. Ramesh C. Sachdeva, Daniel D. Blinka, "Improving the Odds of Success: Quantitative Methodology in Law Practice", *Wisconsin Lawyer*, Vol. 78, No. 12, December 2005. Puede obtenerse electrónicamente en: http://www.wisbar.org

7. Stahl Saul, *A Gentle Introduction to Game Theory*, American Mathematical Society, Providence, RI, 1999.

8. von Neumann John, Morgersetrn Oskar, *Theory of Games and Economic Behavior* (commemorative edition) Princeton University Press, Princeton, NJ, 2007.

9. www.gametheory.net

Ejercicios

▷ **7.1** Encontrar los equilibrios de estrategias puras (si existen) de los siguientes juegos, en donde R y C son los jugadores (renglón y columna, y a,b,α y β estrategias.

1.

R \ C	a	b
α	(1,-1)	(-2,2)
β	(2,-2)	(-1,1)

2.

R \ C	a	b
α	(-1,1)	(-3,3)
β	(-2,2)	(-1,1)

3.

R \ C	a	b
α	(10,10)	(15,5)
β	(5,15)	(12,12)

4.

R \ C	A	B
α	(0,0)	(0,2)
β	(3,0)	(0,0)

5.

R \ C	a	b
α	(30,30)	(60,0)
β	(0,60)	(35,35)

▷ **7.2** Supongamos que dos países, digamos México y Brasil, tienen la opción de abrirse mutuamente al intercambio comercial (apertura) o bien proteger sus economías cerrando sus fronteras (protección). La tabla que describe sus ganancias (pérdidas), en términos de unidades de bienestar para sus habitantes, es la siguiente.

México \ Brasil	Apertura	Protección
Apertura	(10,10)	(-10,20)
Protección	(20,-10)	(-5,-5)

1. Encontrar el equilibrio de Nash del juego e interpretar.
2. ¿Existe una estrategia dominante?
3. Discutir la necesidad de firmar tratados comerciales.
4. ¿Es el comercio un juego suma cero?

▷ **7.3** Consideremos la siguiente situación. Dos familias vecinas (los Rumbos y los Campos) pueden elegir entre respetar o no ciertas normas de convivencia básica. La tabla que describe los pagos de las familias, en unidades de bienestar está dada por,

R \ C	Respetar	No respetar
Respetar	(30,30)	(-10,40)
No respetar	(40,-10)	(-5,-5)

1. Encontrar las estrategias y pagos de equilibrio.

2. ¿Existe una estrategia dominante?

3. Comentar el resultado anterior y proponer alguna alternativa.

▷ **7.4** Tomemos el ejemplo dado por la matriz de pagos 7.1. Describir el equilibrio si el gobierno municipal impone una multa de 2 unidades a los vendedores que no limpien.

▷ **7.5** Carlos y Ramón viven en el mismo condominio y el arreglo del jardín común está a cargo de los dos. Los fines de semana, cada uno de ellos puede decidir entre ayudar a arreglar el jardín o no hacer nada. Supongamos que V es el valor de la ayuda y C es el costo (esfuerzo) de dicha ayuda. La matriz de pagos está dada por

R \ C	ayudar	no ayudar
ayudar	(V-C,V-C)	(V-2C,V)
no ayudar	(V,V-2C)	(0,0)

1. Calcula equilibrio(s) de estrategias puras si $V < 2C$.

2. Calcula equilibrio(s) de estrategias puras si $V > 2C$.

3. Si $V > 2C$, suponer que al vecino que no ayude se le impone una multa de S, que se refleja en un aumento del pago de mantenimiento del condominio. Encontrar la nueva matriz de pagos y equilibrio(s) de estrategias para los casos $S < C$ y $S > C$.

4. Si en lugar de multar al que no coopera se otorga un premio de S al que ayuda, encontrar la nueva matriz de pagos y equilibrio(s) de estrategias para los casos $S < C$ y $S > C$.

▷ **7.6** Sony y Samsung pueden adaptar dos tecnologías diferentes de alta definición, digamos A y B. La matriz de pagos es la siguiente:

Sony \ Samsung	A	B
A	(3,3)	(1,2)
B	(2,1)	(5,5)

1. Encontrar los equilibrios de estrategias puras.
2. Si el estatus quo es que los dos adoptan A, ¿qué sucede si una compañía cambia unilateralmente a B?
3. Suponer que no puede transferirse la utilidad entre ambas compañías. ¿Podrían beneficiarse al cooperar?

▷ **7.7** El juego de la *gallina* posee un equilibrio de Nash de estrategias mixtas.

1. Encontrar este equilibrio.
2. Comparar la utilidad esperada de este equilibrio con la del equilibrio correlacionado obtenido en la sección7.5.

▷ **7.8** Un oso y un alce se detectan a distancia y tienen las opciones y pagos dados por la siguiente matriz:

Alce \ Oso	correr	esconderse
correr	(60,20)	(0,0)
esconderse	(0,0)	(20,60)

Encontrar todos los equilibrios de Nash de este juego (estrategias puras y mixtas) si p y q son, respectivamente, los porcentajes (probabilidades) de veces que el alce y el oso corren.

▷ **7.9** Camila y Rodulfo viven juntos y acaban de comprar un sistema para climatizar su hogar. En general hay una batalla entre ellos para decidir la témperatura de su casa. Supongamos que las estrategias y pagos están dados como en la siguiente matriz:

Rodulfo \ Camila	cálido	frío
cálido	(20,0)	(0,10)
frío	(0,90)	(20,0)

1. Verificar que no existe un equilibrio de estrategias puras.
2. Verificar que existe un equilibrio de estrategias mixtas y encontrarlo.

3. Encontrar la utilidad esperada para Camila y Rodulfo para el equilibrio encontrado.

▷ **7.10** Veamos como la estructura de pagos es relevante en términos de políticas públicas. Los jugadores son el hampa (H) y la sociedad/sistema judicial (S). La matriz de estrategias y pagos podría estár dada por,

S \ H	honestidad	cimen
tolerar	(3,1)	(-3,2)
castigar	(-1,-1)	(-4,-3)

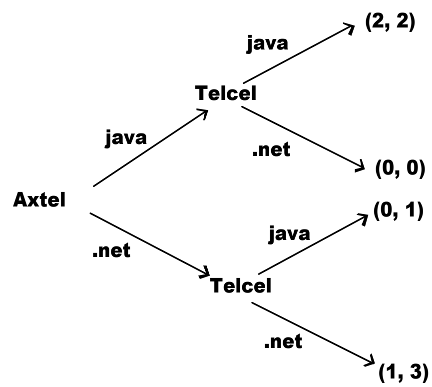
1. Encontrar equilibrio(s) de estrategias puras.

2. Ahora supongamos que los pagos cambian de la siguiente forma,

S \ H	honest.	crimen
tolerar	(3,3)	(-3,2)
castigar	(-1,-1)	(-4,-3)

Obtener equilibrio(s) de estrategias puras. ¿Cómo podría cambiarse el pago para lograr este nuevo equilibrio?

▷ **7.11** Tomar el ejemplo 7.10.4 y suponer que Axtel actúa primero y Telcel segundo de manera que el árbol de juego es el siguiente:



Encontrar el nuevo equilibrio por inducción inversa y comparar el resultado con el del ejemplo 7.10.4.

▷ **7.12** Encontrar una probabilidad q para los atenienses, de manera que el resultado del juego del ejemplo 7.10.5 sea la retirada de Atenas.

Capítulo 8

Elementos de estadística descriptiva

8.1. Introducción

El análisis matemático de datos poblacionales y económicos se formalizó con el surgimiento de la teoría de la probabilidad en el siglo XVII. Originalmente a este análisis se le conocía como *Aritmética Política* y en el siglo XIX se comenzó a utilizar el término: *estadística*. Éste se deriva del latín *statisticus* que quiere decir *del Estado*, ya que la recopilación de este tipo de información numérica fue iniciada por los gobernantes de la antigüedad.

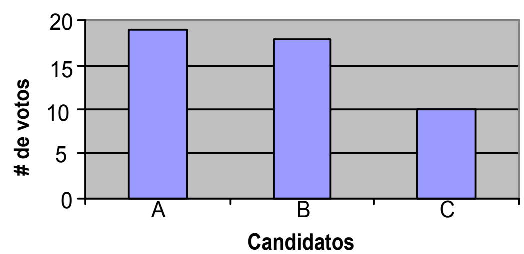
La tarea de la estadística es proporcionar las herramientas para el análisis de datos cuantitativos. Al conjunto de todos los datos que se desean estudiar se le denomina **población,** asimismo, a cualquier subconjunto del total de datos se le conoce como **muestra**. Por ejemplo, las encuestas pretenden encontrar alguna característica general de una población -digamos votantes- por medio del análisis de una muestra.

La **estadística descriptiva** obtiene, organiza y presenta la información. La **inferencia estadística** prueba la validez -o la invalidez- de hipótesis que se tienen acerca de una población, con base a información de una o varias muestras. En este texto nos ocuparemos simplemente de algunas técnicas sencillas de descripción e inferencia.

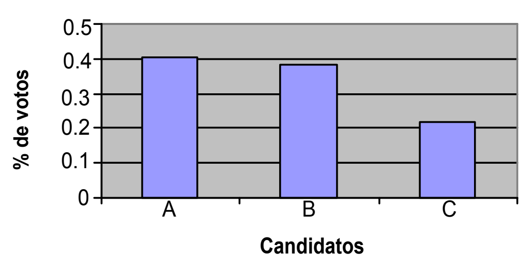
8.2. Histogramas y pays

Dado un conjunto de datos numéricos cuyos elementos se repiten, éstos pueden organizarse para mostrar el número de veces -o frecuencia- con la cual aparece cada elemento. La forma más popular de representar esto es mediante un **histograma de frecuencias** que consiste en una gráfica de barras donde se presentan los datos en orden ascendente -si se trata de datos que pueden ordenarse- en el eje horizontal y la frecuencias con la que aparecen en el eje vertical. Puede realizarse también un **histograma de frecuencias relativas**, poniendo en el eje vertical el porcentaje de veces que aparece cada dato.

Las siguientes figuras ilustran histogramas de frecuencia y frecuencia relativa para la votación obtenida por tres candidatos presidenciales: A, B y C.

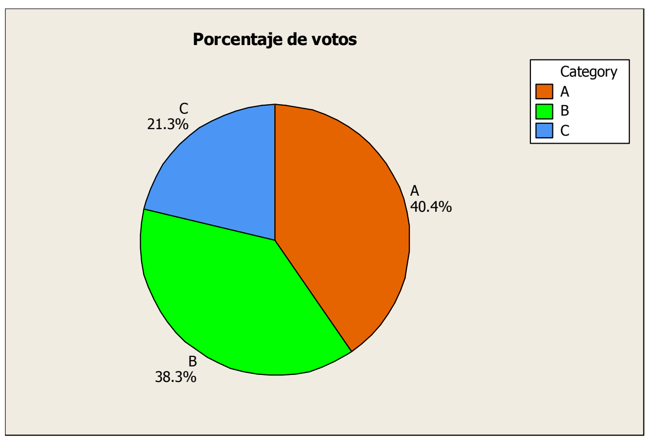


Votación en millones de votos.



Porcentaje de votos.

Cuando se trata de observaciones no numéricas las frecuencias o frecuencias relativas suelen también representarse por medio de un **diagrama de pay**. Este diagrama representa las frecuencias relativas de cada categoría: cada rebanada del pay muestra el porcentaje del total que representa una categoría. El ejemplo de los candidatos presidenciales del párrafo anterior podría ilustrarse, alternativamente, con el siguiente diagrama:

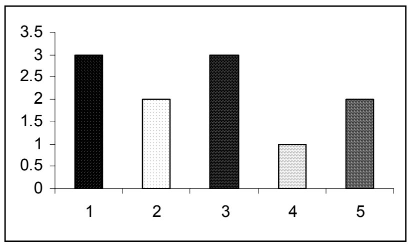


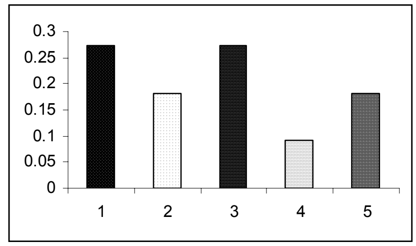
El diagrama de pay y el histograma contienen esencialmente la misma información, sin embargo, el histograma también nos da una idea visual de la distribución de los datos.

Existen otras formas de presentar los datos: gráficas de barras horizontales, que son como histogramas con las categorías en el eje vertical y los valores en el eje horizontal, *diagramas de caja* que muestran la mediana, el rango y puntos extremos, diagramas de dispersión que muestran la relación entre dos variables, etcétera. Aquí nos concentraremos en los histogramas pues son los que mejor representan el concepto de distribución de los datos.

Ejemplos

Ej 8.2.1 Realizar un histograma de frecuencias y de frecuencias relativas con el conjunto de datos $\{1, 3, 3, 5, 2, 1, 5, 4, 2, 1, 3\}$:





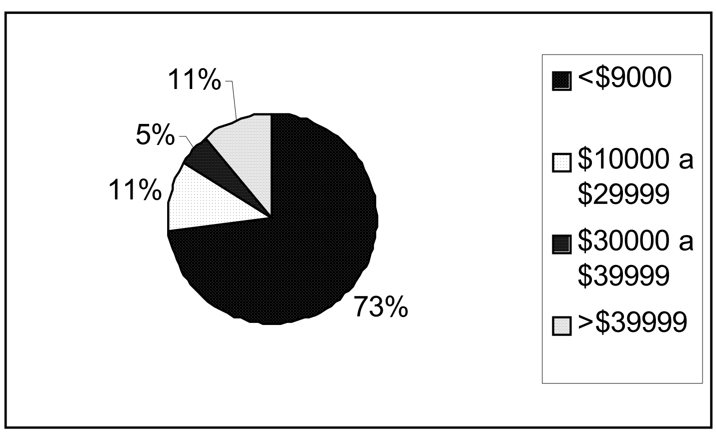
Ej 8.2.2 Consideremos los siguientes datos de una compañía

Salario mensual	*Número de empleados*
menos de \$9999	10
\$10,000 a \$29999	70
\$30,000 a \$39999	10
mayor a \$39999	5

Notemos que el número total de empleados es de

$$10 + 70 + 10 + 5 = 95.$$

Así, el diagrama de pay para las frecuencias relativas es el siguiente:



♠♠♠♠♠

8.3. Diagramas de tallos y hojas

Una alternativa al histograma de frecuencias es el llamado **diagrama de tallos y hojas**. Este diagrama es útil cuando los datos son de órdenes de magnitud semejantes. Por ejemplo, el diagrama de tallos y hojas para las calificaciones obtenidas por treinta y tres estudiantes en un examen podría ser el siguiente, en donde las decenas de la primera columna forman el tallo y las unidades, en las columnas subsecuentes, las hojas. Notemos que este diagrama nos da una buena idea de la distribución de frecuencias de las calificaciones de los 33 estudiantes.

10	0					
9	1	2	2			
8	4	5	8	8	9	
7	5	7	7	9	9	9
6	0	3	4	6	6	
5	2	5	6	6	8	
4	1	4	5	6		
3	3	6				
2	1					
1	8					

8.4. Media, mediana y moda

Dado un conjunto de datos, las llamadas medidas de tendencia central son números alrededor de los cuales se concentran los datos. La **media** o promedio es quizás la medida de tendencia central más conocida. Dado un conjunto de datos $\{x_1, ..., x_n\}$ que representa una muestra de alguna población, la media del conjunto se define como

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

es decir, el promedio de los datos.

Ejemplos

Ej 8.4.1 Un estudiante cursó seis materias en el semestre obteniendo calificaciones de 8,7,10,8,9 y 7. Su calificación media o promedio semestral es de

$$\frac{8+7+10+8+9+7}{6} = 8.17$$

Ej 8.4.2 Una pequeña compañía consultora tiene una secretaria un empleado de limpieza, un mensajero y un economista. Sus salarios mensuales respectivos son de \$5000, \$4000, \$3500 y \$50000. El salario promedio de la compañía es de

$$\frac{5000+4000+3500+50000}{4} = 15625.$$

El lector podría pensar que este salario no es una medida *representativa* de los salarios del personal. ¿Qué sucede? La razón es que existe un dato, el salario del economista, que está totalmente fuera del rango de los demás salarios. Así, al realizar el promedio, este dato *jala* a los demás.

♠♠♠♠♠

Para evitar problemas con datos alejados de los demás, como en el ejemplo 8.4.2, se utiliza otra medida de tendencia central llamada mediana. La **mediana** es el valor que parte al conjunto de datos ordenados en dos. Para encontrar la mediana, los datos se ordenan de menor a mayor y,

- si el conjunto de datos es impar, la mediana es el valor que se encuentra a la mitad del conjunto,
- si el conjunto de datos es par, la mediana es el promedio de los dos datos intermedios.

Ejemplos

Ej 8.4.3 Encontrar la mediana del conjunto de datos $\{1,3,8,15,5,9,2,1,4\}$. Se ordena el conjunto de menor a mayor obteniendo,

$$\{1, 1, 2, 3, 4, 5, 8, 9, 15\}.$$

Dado que es un conjunto impar de datos, la mediana es el dato intermedio, es decir, 4.

Ej 8.4.4 Encontrar la mediana del conjunto de salarios {\$5000, \$4000, \$3500 y \$50000}. Los ordenamos de menor a mayor obteniendo,

$$\{3500, 4000, 5000, 50000\}.$$

Este es un conjunto par de datos de manera que la mediana es el promedio de los dos datos intermedios, es decir la mediana es

$$\frac{4000 + 5000}{2} = 4500.$$

En ocasiones, este dato puede ser una mejor representación de los salarios de la empresa que el promedio obtenido anteriormente en el ejemplo 8.4.2.

♠♠♠♠♠

Notemos que en el ejemplo 8.4.3, el número 1 aparece con mayor frecuencia que los demás (se repite dos veces). La media y la mediana no toman en cuenta la repetición de los datos, se define para este efecto, la **moda.** Ésta es simplemente el valor o categoría que ocurre con mayor frecuencia en un conjunto de datos. Es claro que puede haber más de una moda ya que puede haber más de un dato que se repita con la misma frecuencia. Para el caso de dos modas, decimos que la distribución de los datos es **bimodal**. Cuando existen más de dos modas, la distribución se denomina **multimodal**.

Ejemplos

Ej 8.4.5 Para los datos del ejemplo 8.4.3 la moda es igual a 1.

Ej 8.4.6 La moda del conjunto de datos {100,55,60,100,70,100,55,60} es 100 ya que se repite el mayor número de veces: 3.

Ej 8.4.7 El conjunto {100,55,60,100,70,55} es bimodal con modas de 55 y 100.

Ej 8.4.8 El conjunto de juguetes {coche, canicas, trompo, pelota, coche, pelota, yoyo, pelota}, tiene como moda a "pelota".

♠♠♠♠♠

Para grandes poblaciones de datos numéricos, es común utilizar a la media para *representar* a la población. Si las muestras presentan datos extremos alejados de los demás, se eliminan estos datos de manera que la media sea un valor representativo. Un ejemplo de esto son las medias de las evaluaciones de los jueces en eventos deportivos: se elimina la calificación más alta y la más baja y se toma el promedio de las restantes como representativo.

En general, la media poblacional es desconocida y tenemos que inferirla mediante las medias de una o varias muestras de la población. A la media de la población total o **media real** la denotamos por μ. Si tenemos un gran número de muestras, el valor esperado de sus medias es una *buena* aproximación de la media de la población. En este caso decimos que $\bar{x}$ (media muestral) es un **estimador sin sesgo** de μ (media poblacional).

8.5. Rango y desviación estándar

Consideremos a los conjuntos $\{5,6,7,8,9\}$ y $\{1,2,7,12,13\}$. Ambos tienen media y mediana iguales a 7 y, sin embargo, nuestra intuición nos dice que los datos del segundo conjunto están más dispersos. ¿Cómo formalizar este concepto de dispersión? Una forma de hacerlo es considerando el **rango** o extensión de los datos que se define como la diferencia entre el dato más grande y el más pequeño.

Ejemplo

Ej 8.5.1 El rango del conjunto de datos $\{5,6,7,8,9\}$ es de $9-5=4$ y el rango del conjunto $\{1,2,7,12,13\}$ es de $13-1=12$. Observemos que el conjunto con el rango más grande es más disperso.

♠♠♠♠♠

Una forma muy utilizada para medir la dispersión de un conjunto de datos es la llamada **desviación estándar**. Ésta mide qué tanto los datos se desvían de la media y se denota comúnmente por s_n. Esta desviación se construye a partir de su cuadrado conocido como **varianza**, como sigue: dado un conjunto de datos $\{x_1, ..., x_n\}$ con media $\bar{x}$, la varianza se define como

$$s_n^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}. \tag{8.1}$$

Intuitivamente, $x_i - \bar{x}$ es la distancia a la media del dato x_i, ésta se eleva al cuadrado para tener siempre un valor positivo y se divide entre n para obtener el promedio de estas desviaciones.

Por definición, la varianza da un valor numérico para el promedio de los cuadrados de las distancias. Para que el número conserve las unidades originales de la variable, se toma la raíz cuadrada y se tiene así la desviación estándar

$$s_n = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}.$$

Esta desviación nos da una idea de que tan alejados están los datos de la media.

Es claro que si el conjunto de datos no es pequeño, tanto la varianza como la desviación estándar son tediosas de calcular. Afortunadamente, casi cualquier calculadora, toda hoja de cálculo y cualquier software estadístico, tienen ya funciones preprogramadas para el cálculo de estas medidas.

A la varianza de una población se le denota por σ^2 y a su desviación estándar por σ. Al igual que en el caso de la media, la varianza de una población puede inferirse a partir de las varianzas de las muestras. Desgraciadamente, en este caso se tiene una complicación: la varianza de la población no se aproxima bien por el valor esperado de las varianzas de las muestras y tiende a subestimarse. Decimos que s_n^2 es un **estimador con sesgo** de σ^2.

El problema de estimación de la varianza poblacional puede solucionarse utilizando, en lugar de las varianzas muestrales s_n^2, la siguiente expresión[1], en donde se divide entre $n-1$, en lugar de n :

$$s_{n-1}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}. \tag{8.2}$$

En este caso, el valor esperado de los números s_{n-1}^2 para las diferentes muestras si es un buen estimador -sin sesgo- para la varianza poblacional.

La observación anterior causa algo de confusión en algunos textos. En ocasiones se define la varianza de una muestra directamente como el estimador dado por la expresión 8.2. No obstante, cuando nos piden evaluar la varianza de un conjunto de datos y no estamos realizando ningún tipo de inferencia acerca de la población debe utilizarse la definición 8.1 para el cálculo de la varianza de un conjunto de datos.

Ejemplo

Ej 8.5.2 Calcular la varianza y la desviación estándar de los conjuntos de datos $\{5,6,7,8,9\}$ y $\{1,2,7,12,13\}$. Como se vio anteriormente, ambos conjuntos tienen media igual a 7. La varianza para el primer conjunto es

$$s_5^2 = \frac{(5-7)^2 + (6-7)^2 + (7-7)^2 + (8-7)^2 + (9-7)^2}{5} = 2,$$

y la desviación estándar

$$s_5 = \sqrt{2} = 1.4142.$$

para el segundo conjunto tenemos,

$$s_5^2 = \frac{(1-7)^2 + (2-7)^2 + (7-7)^2 + (12-7)^2 + (13-7)^2}{5} = 24.4$$

[1]Es fácil verificar que

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0,$$

de manera que uno de los términos $(x_i - \bar{x})$ puede siempre obtenerse a partir de los demás. Esto significa que sólo $n-1$ diferencias $(x_i - \bar{x})$ son independientes.

y

$$s_5 = \sqrt{24.4} = 4.94.$$

Estos resultados nos dicen que el segundo conjunto es aproximadamente cuatro veces más disperso que el primero.

Ej 8.5.3 Los rendimientos bursátiles se representan por un rendimiento esperado y su varianza. Las carteras se construyen combinando diversos instrumentos financieros, de manera que el rendimiento esperado y la desviación estándar de la combinación satisfaga el perfil de riesgo del cliente. Los rendimientos altos van acompañados de desviaciones altas y los rendimientos bajos sufren de menor variabilidad.

En un extremo tenemos los instrumentos libres de riesgo, digamos bonos gubernamentales como CETES con un rendimiento esperado de, digamos el 7 % y desviación estándar de 0. En otro extremo tendríamos un activo con un rendimiento esperado del 50 % pero una desviación estándar del 120 %.

♠♠♠♠♠

Bibliografía Complementaria

1. Agresti Alan, Finlay Barbara, *Statistical Methods for the Social Sciences*, 4^a ed, Pearson, Allyn & Bacon, New York, NY, 2008.

2. Aguirre Víctor et al., *Fundamentos de Probabilidad y Estadística*, 2^a edición. Jit Press, México, D.F. 2006.

3. Dekking F. Michel et al., *A Modern Introduction to Probability and Statistics : understanding why and how*, Springer-Verlang, , London, UK 2005.

4. Kaye, David H., Freedman, David A., "A Reference Guide on Statistics", *Reference Manual on Scientific Evidence*, 2nd edition, Federal Justice Center. Versión electrónica en,

   $http://www.fjc.gov/public/pdf.nsf/lookup/sciman01.pdf/\$file/sciman01.pdf$

Ejercicios

▷ **8.1** Un grupo de 20 estudiantes presentó un examen de matemáticas con 5 preguntas. Los siguientes datos muestran el número de preguntas correctas que contestó cada estudiante: {1,5,5,0,2,1,3,4,0,1,2,1,0,3,2,0,2,1,3,2}. Realizar un histograma de frecuencias y otro de frecuencias relativas.

▷ **8.2** La bolsa de trabajo del ITAM determinó que el año pasado se colocaron 220 estudiantes en profesiones repartidas en las siguientes categorías como se muestra en la tabla:

Categoría	# estudiantes
Economía	75
Administración	53
Ingenierías	15
Matemáticas	35
Otros	42

Representar los datos con un histograma de frecuencias y un diagrama de pay.

▷ **8.3** Los gastos de la familia Gómez se dividen en las siguientes categorías: alimentación (30 %), renta (25 %), gasolina y mantenimiento del auto (10 %), diversión (15 %) y pago de colegiaturas (20 %). Representar la aportación al gasto de cada una de las categorías mediante una gráfica de barras y un diagrama de pay.

▷ **8.4** Supongamos que queremos describir el color típico de ojos de los habitantes de la ciudad de México. La mejor descripción estaría dada por, ¿la media, la mediana o la moda? Explicar.

▷ **8.5** Alguien reporta un conjunto de datos con una media de 1500 y una desviación estándar de -80. ¿Cómo sabemos que cometió un error?

▷ **8.6** Para cada una de las siguientes listas de datos encontrar la media, la mediana, la (s) moda (s) (si existen), el rango y la desviación estándar.

1. {3,7,12,16,23}
2. {3,7,12,16,100}
3. {1,1,5,5,5,7,7,8,9,9,9}
4. {26,31,46,31,26,29,31}
5. {128,131,136,125,132,128,125,127}

▷ **8.7** Proporcionar una muestra de datos con:

1. Una media de 61.
2. Una mediana de 15.
3. Una media de de 10 y mediana de 4.
4. Una media de 10, mediana de 4 y moda de 2.
5. Una moda de 7.
6. Una moda de azul.
7. Un rango de 23.
8. Una media de 20 y desviación estándar de 0.
9. Una media de 20 y desviación estándar de 2.
10. Una media de 20 y desviación estándar de 8.

▷ **8.8** En una empresa, se tomó una muestra de 15 trabajadores y se obtuvo el siguiente conjunto de datos:

$\{5, 7, 0, 3, 15, 6, 5, 9, 4, 7, 11, 5, 2, 0, 11\}$.

Éste representa el número de días que faltaron a causa de enfermedad. Obtener la media, la mediana y la(s) moda(s) del conjunto.

▷ **8.9** Un grupo de 15 estudiantes mostró la cantidad de dinero que tenía cada uno en su cartera. Se obtuvieron los siguientes datos: {\$120,\$70,\$20,\$50,,\$0,\$50,\$200,\$0,\$20, \$500,\$100,\$150,\$0,\$50,\$80}.

1. Obtener la media, la mediana y la(s) moda(s). ¿Qué medida(s) crees qué representa(n) mejor estos datos? Explicar.
2. Obtener el rango y la desviación estándar. Interpretar.

▷ **8.10** Los siguientes datos representan el número de créditos que llevan 30 estudiantes de 5^0 semestre de la carrera de derecho en el ITAM:

$\{30, 12, 18, 12, 24, 30, 36, 18, 36, 42, 30, 24, 30, 18, 42, 6, 24, 30, 48, 30, 24, 18, 36, 24, 30, 30, 24, 36, 30, 24\}$.

1. Realizar un histograma de frecuencias y otro de frecuencias relativas.
2. Realizar un diagrama de tallo y hojas.
3. Encontrar la media, la mediana y la(s) moda(s).
4. Encontrar el rango y la desviación estándar.

Capítulo 9

Distribuciones comunes

9.1. Introducción

En la sección 5.2 del capítulo 2 vimos como asociarle una distribución de probabilidad a los valores de una variable aleatoria. En el análisis estadístico de datos de poblaciones aparecen ciertas distribuciones con mucha frecuencia. En esta sección vamos a analizar tres de las más comunes: la distribución uniforme (discreta y continua), que representa datos equiprobables, la distribución binomial y la distribución normal, que están íntimamente relacionadas y podemos pensar a la segunda como la versión continua de la primera. Ésta última es la reina de las distribuciones tanto por su importancia práctica como teórica.

Si X es una variable aleatoria con función de distribución -o densidad en el caso continuo- f, entonces decimos que los datos representados por los valores que toma X tienen una distribución dada por f.

En el caso discreto, en el cual X toma valores $X_1, X_2, \ldots, X_n$, éstos representan datos con media

$$\mu = E(X) = \sum_{i=1}^{n} X_i f(X_i),$$

en donde, recordemos que $f(X_i) = P(X = X_i)$ es la probabilidad de que X

tome el valor X_i. Asimismo, la varianza está dada por

$$\sigma^2 = E\left(\sum_{i=1}^{n}(X_i - E(X))^2\right) = E\left(\sum_{i=1}^{n}(X_i - \mu)^2\right)$$
$$= \sum_{i=1}^{n} f(X_i)(X_i - E(X))^2,$$

que es el valor esperado de las desviaciones -al cuadrado- alrededor de la media. Decimos que la distribución f tiene media μ y varianza σ^2. Claramente la desviación estándar es simplemente $\sigma = \sqrt{\sigma^2}$.

Las definiciones de media y varianza pueden extenderse al caso de distribuciones continuas, sin embargo, se requiere el concepto de integral para realizar la *suma de un continuo de valores* por lo que queda fuera del alcanze de este texto. No obstante, la intuición detrás de los conceptos es la misma y, abusando de esto, hablaremos de medias y varianzas -o desviaciones estándar- para algunos casos continuos.

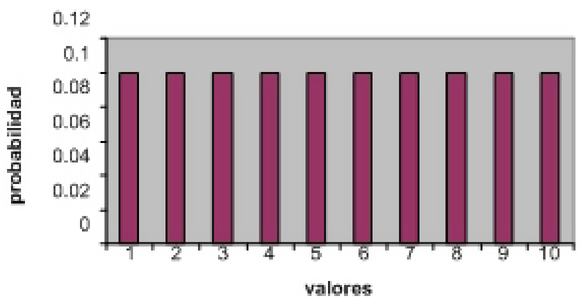
9.2. Uniforme

Cuando los datos de un conjunto ocurren, cada uno con la misma frecuencia, decimos que están uniformemente distribuidos. Éste es el caso de la distribución de probabilidad determinada por una variable aleatoria sobre un espacio muestra de datos equiprobables. Por ejemplo, los que resultan de lanzar una moneda, lanzar un dado, girar una ruleta o bien, elegir un número al azar en el intervalo $[0, 1]$. La distribución que resulta, ya sea en el caso discreto o continuo, se conoce como **distribución uniforme**.

En el caso discreto, dada una variable aleatoria X que toma valores $\{1, 2, \ldots, n\}$, cada uno con probabilidad $\frac{1}{n}$, la distribución de probabilidad queda dada por

$$f(i) = \frac{1}{n}, \text{ para toda } i = 1, 2, \ldots, n.$$

En el siguiente histograma se ilustra el caso para $n = 10$:



La media de esta distribución es

$$\mu = \frac{1}{n}\left(\sum_{i=1}^{n} i\right) = \frac{1}{n}\frac{n(n+1)}{2}$$
$$= \frac{n+1}{2}, \tag{9.1}$$

en donde se utiliza la conocida igualdad para la suma de los primeros n enteros:

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

Asimismo, la desviación estándar es

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(i-\mu)^2}$$
$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(i - \frac{n+1}{2}\right)^2}$$
$$= \sqrt{\frac{n^2-1}{12}}, \tag{9.2}$$

en donde se utiliza la igualdad para la suma de los cuadrados de los primeros n enteros dada por,

$$\left(\sum_{i=1}^{n} i^2\right) = \frac{n(n+1)(2n+1)}{6}.$$

Pensemos ahora en el caso continuo de elegir un número al azar en el intervalo $[a, b]$ de la recta numérica. Sea X la variable aleatoria asociada que toma cada uno de los valores del intervalo. La densidad de probabilidad correspondiente debe ser tal que el área bajo la gráfica sea la unidad y dados números c y d en el intervalo $[a, b]$, con $c < d$, la probabilidad

$$P(c < X < d) = \frac{d-c}{b-a} \tag{9.3}$$

es el área por debajo de la gráfica entre c y d. Podemos definir esta función de densidad como

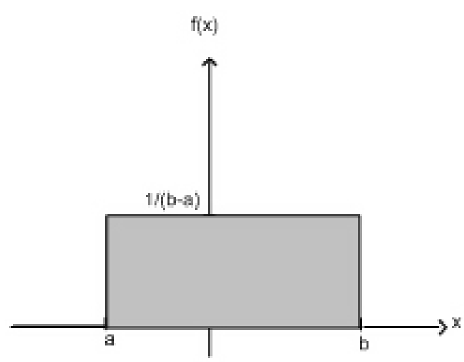
$$f(x) = \begin{cases} \frac{1}{b-a} \text{ si } a \leq x \leq b, \\ 0 \text{ de otra forma.} \end{cases} \tag{9.4}$$

La media y varianza están dadas por,

$$\mu = \frac{a+b}{2}, \tag{9.5}$$

$$\sigma^2 = \frac{(b-a)^2}{12}. \tag{9.6}$$

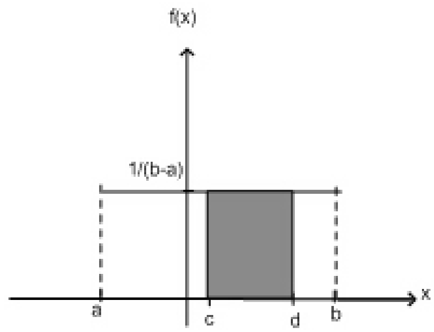
La gráfica de esta regla de correspondencia o función está dada por:



El área sombreada es

$$(b-a)\frac{1}{b-a} = 1$$

Asimismo, la probabilidad de que X tome valores entre c y d es el área sombreada de la siguiente figura:



Ejemplos

Ej 9.2.1 Sea X la variable aleatoria uniforme que representa el resultado observado al lanzar un dado. La probabilidad de obtener al menos un 3 es de

$$
\begin{aligned}
P(X \geq 3) &= P(X=3)+P(X=4)+P(X=5)+P(X=6) \\
&= \frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{4}{6}=\frac{2}{3}.
\end{aligned}
$$

El lanzamiento medio, utilizando 9.1 es de,

$$\frac{6+1}{2}=\frac{7}{2}=3.5,$$

y la desviación estándar según 9.2 de,

$$
\begin{aligned}
\sigma &= \sqrt{\sigma^2} \\
&= \sqrt{\frac{36-1}{12}} \\
&\simeq 1.1.
\end{aligned}
$$

Ej 9.2.2 Sea X el tiempo de espera (en minutos) en un consultorio dental y supongamos éste sigue una distribución uniforme descrita por la función de densidad,

$$f(x)=\begin{cases} \frac{1}{90} \text{ si } 0 \leq x \leq 90, \\ 0 \text{ de otra forma.} \end{cases}$$

La probabilidad de que un paciente espere entre 15 y 45 minutos está dada, de acuerdo a 9.3 por,

$$P(15 < X < 45) = \frac{45-15}{90} = \frac{30}{90} = \frac{1}{3}.$$

El tiempo medio de espera, utilizando 9.5 es de

$$\mu = \frac{90-0}{2} = 45 \text{ minutos}$$

con desviación estándar, de acuerdo a 9.6 de

$$\begin{aligned}\sigma &= \sqrt{\frac{(90-0)^2}{12}} \\ &= \frac{90}{\sqrt{12}} \simeq 26 \text{ minutos}\end{aligned}$$

Ej 9.2.3 El tiempo (en minutos) que una persona espera el Metrobús los días de semana sigue una distribución uniforme dada por,

$$f(x) = \begin{cases} \frac{1}{12} \text{ si } 0 \leq x \leq 12, \\ 0 \text{ de otra forma.} \end{cases}$$

La probabilidad de que una persona espere menos de seis minutos es,

$$P(0 < X < 6) = \frac{6}{12} = \frac{1}{2}.$$

El tiempo medio de espera es de

$$\mu = \frac{12}{2} = 6 \text{ minutos},$$

con desviación estándar de

$$\sigma = \sqrt{\frac{12^2}{12}} \simeq 3.5 \text{ minutos}.$$

♠♠♠♠♠

9.3. Binomial

En 1713, ocho años después de su muerte, fue publicado el libro *Ars Conjectandi* (el arte de la conjetura) de Jacobo Bernoulli. Este trabajo incluye varias aplicaciones de la teoría de la probabilidad a los juegos de azar. Entre otras cosas, se define lo que hoy se entiende por un ensayo Bernoulli que es el elemento clave para llevar a cabo un experimento binomial. Para entender de que se trata consideremos las siguientes dos situaciones:

- Se lanza un dado 10 veces, ¿cuál es la probabilidad de obtener 4 seises?
- Un nuevo medicamento para la diabetes mantiene el nivel de glucosa en la sangre del paciente, muy cercano a los niveles normales. Desgraciadamente, el 10 % de los pacientes que reciben este medicamento desarrolla hipertensión arterial. Un médico tiene 5 pacientes diabéticos -sin relación alguna entre ellos- y desea saber la probabilidad de que todos ellos desarrollen hipertensión arterial si son tratados con el medicamento.

Estas dos situaciones describen ejemplos de lo que se conoce como un experimento binomial cuya definición precisa es la siguiente:

Definición 9.3.1 *Un experimento binomial consiste de n etapas o ensayos idénticos e independientes, llamados* ***ensayos Bernoullli****, tales que,*

1. *En cada ensayo o etapa hay dos posibles resultados: éxito y fracaso.*[1]
2. *La probabilidad de éxito en cada ensayo es de p y la de fracaso de $1-p$.*

La variable aleatoria discreta X que asociamos a este experimento es el número de éxitos en los n ensayos. En las situaciones descritas tenemos que,

- $X = \#$ de seises en 10 lanzamientos de un dado.

[1] Podríamos llamarlos positivo y negativo, o 1 y 0, o cualquier otra cosa. La denominación de éxito y fracaso no tiene nada que ver con el significado literal de estos términos.

- $X = \#$ de pacientes, del grupo de 5 pacientes, que desarrollan hipertensión arterial.

Pensemos ahora en lanzar un dado n veces. Queremos calcular la probabilidad de que salgan -exactamente- k cuatros en estos n lanzamientos. El espacio muestral tiene 6^n elementos. Denotemos por $\{l_1, l_2, \ldots, l_n\}$ al conjunto de lanzamientos, es decir, l_i es el lanzamiento i. Si queremos que salgan k cuatros, éstos pueden aparecer distribuidos de diferentes formas en los n lanzamientos. ¿De cuántas maneras? Simplemente elegimos, del conjunto de n lanzamientos, los k lanzamientos en los cuales aparece un cuatro, éstos son

$$_nC_k.$$

Por ejemplo, si $k = 2$ y $n = 4$ los dos cuatros pueden aparecer en los siguientes subconjuntos de lanzamientos

$$\{l_1, l_2\}, \{l_1, l_3\}, \{l_1, l_4\}, \{l_2, l_3\}, \{l_2, l_4\} \text{ y } \{l_3, l_4\},$$

es decir, puede haber cuatros en los lanzamientos 1 y 2, en el 1 y 3, etcétera. Claramente se trata de $_4C_2$ posibilidades.

En general, una vez elegidos los k cuatros, de $_nC_k$ formas, tenemos 5^{n-k} maneras de elegir los $n-k$ números restantes, ya que puede aparecer cualquier otro número diferente del cuatro. En resumen, hay

$$_nC_k 5^{n-k}$$

maneras de obtener k cuatros en n lanzamientos. La probabilidad de este evento es,

$$\begin{aligned} P(k \text{ cuatros en } n \text{ lanzamientos}) = \\ \frac{_nC_k 5^{n-k}}{6^n} &=_n C_k \frac{1}{6^k} \frac{5^{n-k}}{6^{n-k}} \\ &=_n C_k \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k} \\ &=_n C_k P(\mathit{cuatro})^k P(\mathit{no} \text{ cuatro})^{n-k}. \end{aligned}$$

En el ejemplo anterior, interpretemos el evento: "obtener un cuatro", como el éxito y a "no obtener un cuatro" como el fracaso. Los lanzamientos son los ensayos del experimento binomial. La siguiente proposición es simplemente una generalización de lo que acabamos de obtener:

Proposición 9.3.2 *Sea X la variable aleatoria asociada a un experimento binomial y π la probabilidad de éxito, entonces,*

$$\begin{aligned} P(X=k) &=_n C_k P(\acute{e}xito)^k P(fracaso)^{n-k} \\ &=_n C_k p^k (1-p)^{n-k}. \end{aligned}$$

Ésta representa la probabilidad de obtener k éxitos en n etapas o ensayos. La distribución de probabilidad asociada es simplemente,

$$f(k) = P(X=k) =_n C_k p^k (1-p)^{n-k}.$$

La distribución de probabilidad que acabamos de describir se conoce como **distribución binomial**. Si representamos a los éxitos y fracasos, respectivamente por los números 1 y 0; entonces, la variable binomial consiste en contar el número de éxitos en los n ensayos. La media y la desviación estándar de la distribición binomial con parámetros n y p están dadas por,

$$\mu = np, \tag{9.7}$$

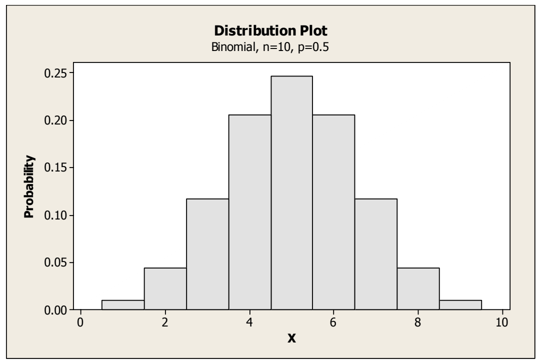
$$\sigma = \sqrt{\sigma^2} = \sqrt{np(1-p)}. \tag{9.8}$$

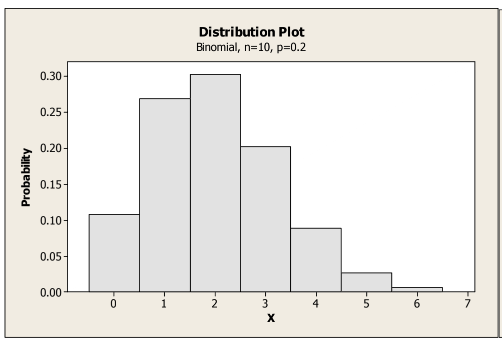
Notemos que dado un número n de ensayos, la varianza tomará su valor máximo cuando el producto $p(1-p)$ sea lo más grande posible[2] y esto sucede cuando $p = \frac{1}{2}$. De esta forma, para un número dado de ensayos, la mayor dispersión en una distribución binomial ocurre cuando $p = \frac{1}{2}$.

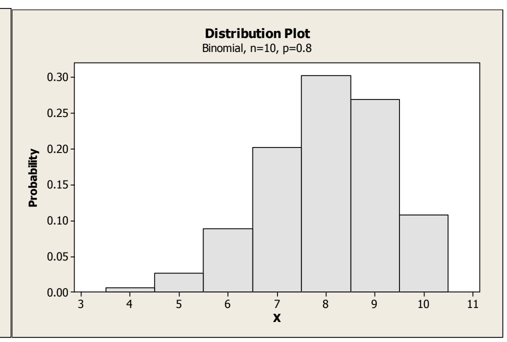
Podemos realizar los histogramas de probabilidad para la distribución binomial. A continuación mostramos los casos para el número de éxitos de un total de $n = 10$ ensayos. La probabilidad de éxito es de 0.5, de 0.2 y de

[2]Esto es un ejercicio sencillo de cálculo diferencial, verificando que el máximo de la curva $y = x(1-x)$ ocurre para $x = \frac{1}{2}$. Otra forma sencilla de comprobarlo es observando la curva $y = x(1-x)$ es una parábola con vértice en $x = \frac{1}{2}, y = \frac{1}{4}$.

0.8, respectivamente. Notemos que el histograma es simétrico para $p = 0.5$, presenta un sesgo hacia la izquierda para $p = 0.2$ y hacia la derecha para $p = 0.8$.







Ejemplos

Ej 9.3.1 La probabilidad de obtener 4 seises en 10 lanzamientos de un dado se calcula poniendo: $n = 10, k = 4$ y $p = \frac{1}{6}$ para obtener,

$$\begin{aligned} P(X = 4) &=_{10} C_4 \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^{10-4} \\ &= \frac{10!}{6!4!} \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^6 = 0.0543. \end{aligned}$$

El número medio de seises que se obtiene en 10 lanzamientos, de acuerdo a 9.7 es de

$$\mu = \frac{10}{6} = 1.67$$

y según 9.8 la desviación estándar es de

$$\sigma = \sqrt{\frac{10}{6} \times \frac{5}{6}} = 1.18.$$

Ej 9.3.2 La probabilidad de obtener al menos 3 águilas en 10 lanzamientos de una moneda honesta se representa por $P(X \geq 3)$, en donde X es una variable aleatoria binomial con $n = 10$ y $p = \frac{1}{2}$. En este caso notemos que el evento complementario es el de obtener menos de 3 águilas en 10 lanzamientos, por lo tanto,

$$P(X \geq 3) = 1 - P(X < 3).$$

Se tiene que

$$\begin{aligned}
&P(X < 3) = \\
&P(X = 0) + P(X = 1) + P(X = 2) \\
&\quad = {}_{10}C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{10-0} + {}_{10}C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{10-1} + {}_{10}C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{10-2} \\
&\quad = \left(1 \times 1 \times \frac{1}{1024}\right) + \left(10 \times \frac{1}{2} \times \frac{1}{512}\right) + \left(45 \times \frac{1}{4} \times \frac{1}{256}\right) = 0.0547.
\end{aligned}$$

De aquí obtenemos

$$P(X \geq 3) = 1 - 0.0547 = 0.9453,$$

es decir, con un 94.53 % de probabilidad aparecen al menos 3 águilas en 10 lanzamientos de una moneda.

El número medio de águilas que esperamos obtener en 10 lanzamientos es de

$$\mu = \frac{10}{2} = 5$$

con desviación estándar de

$$\sigma = \sqrt{\frac{10}{2} \times \frac{1}{2}} = 1.58.$$

Ej 9.3.3 La probabilidad de que los 5 pacientes tratados con el medicamento desarrollen hipertensión arterial se calcula poniendo: $n = 5, k = 5$ y $p = 0.1$ para obtener,

$$\begin{aligned} P(X = 5) &= {}_5 C_5 \, (0.1)^5 \, (0.9)^{5-5} \\ &= \frac{5!}{5!0!} \, (0.1)^5 = 0.00001. \end{aligned}$$

El número medio de pacientes que desarrollan hipertensión arterial es de

$$\mu = \frac{5}{10} = 0.5$$

con desviación estándar de

$$\sigma = \sqrt{\frac{5}{10} \times \frac{9}{10}} = 0.67.$$

Ej 9.3.4 (Teorema de Bernoulli) Un experimento binomial puede representarse por la siguiente urna con canicas: en la urna colocamos r canicas rojas y a azules[3]. El experimento consiste en extraer repetidamente canicas de la urna, con remplazo, de manera que la urna siempre contiene las mismas canicas. Si al extraer una canica al azar, ésta es roja, entonces se contabiliza como un éxito. La probabilidad de éxito está dada por la proporción de canicas rojas, es decir,

$$p = \frac{r}{r + a}.$$

Supongamos que no sabemos de antemano la proporción de canicas de cada color, por lo tanto desconocemos p. Para tratar de averiguar esta proporción empezamos a extraer canicas con remplazo. Sea n el número de veces que repetimos este proceso de extracción y k el número de canicas rojas o éxitos que obtenemos. Jacobo Bernoulli se dio cuenta de que si n era suficientemente grande, entonces la proporción $\frac{k}{n}$ se parecía cada vez más a la probabilidad p, es decir, a la proporción real de canicas. La prueba formal de este hecho, que nos puede parecer trivial, es uno de los grandes teoremas de la estadística por su importancia para la inferencia y se conoce como el **Teorema de Bernoulli** o la **Ley de los Grandes Números**.

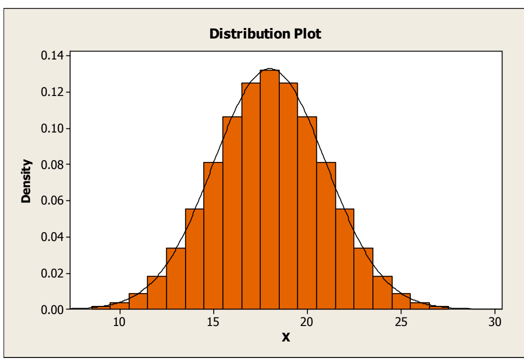
[3]La elección de los colores específicos es irrelevante.

♠♠♠♠♠

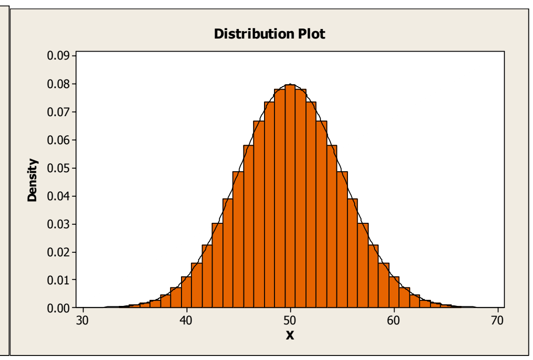
9.4. Normal

En el siglo XVIII, el matemático y estadístico francés Abraham de Moivre (1667 - 1754) fungía como consultor de los grandes apostadores europeos. Como consecuencia, tenía que realizar cálculos del tipo: "la probabilidad de obtener al menos 60 caras iguales al lanzar una moneda 100 veces", involucrando a la distribución binomial. Este cálculo, que ahora nos puede parecer trivial, era sumamente tedioso sin la ayuda de la tecnología actual.

De Moivre notó que cuando la distribución binomial involucraba una gran número de ensayos, su forma podía reproducirse con una curva suave en forma de campana, como se muestra en la figura siguiente:



Binomial con $n = 36, p = 0{,}5$.



Binomial con $n = 100, p = 0{,}5$.

Al observar esto, De Moivre razonó que si pudiese obtener una fórmula para describir esta curva, entonces podría aproximar las probabilidades binomiales cuando se tenía un gran número de ensayos.

Fue así como De Moivre descubrió una expresión para la curva que hoy se conoce como la **curva normal**. La primera mención de ésta, aparece en la edición de 1756 de su libro *The Doctrine of Chance*, publicado en Londres[4].

[4] De Moivre realizó practicamente todo su trabajo científico en Inglaterra y entabló gran amistad con Isaac Newton.

Esta curva fue descubierta en forma independiente por su compatriota Pierre S. Laplace (1749 - 1827), en 1778 y por el alemán Carl F. Gauss (1777 - 1855) en 1809. Hoy en día también es conocida como la distribución Gaussiana o campana de Gauss.

Al medir cualquier fenómeno siempre hay algún error en la medición. En general, los errores pueden ser tanto positivos como negativos, es decir, podemos sobreestimar o subestimar la cantidad real. Asimismo, los errores pequeños ocurren con mayor frecuencia que los grandes. La distribución de estos errores suele ser normal y esto fue observado y estudiado por Gauss en el siglo XIX.

Adicionalmente a la distribución de errores, una gran cantidad de datos reales suelen estar distribuidos en forma normal: las calificaciones de un grupo de alumnos, el rendimiento de una vaca lechera, la altura de los individuos de una población, el nivel de consumo de un producto, el tiempo dedicado para tomar ciertas decisiones y una infinidad de ejemplos más.

La curva que descubrió De Moivre correponde a la función de densidad de lo que llamamos la **distribución normal**. La ecuación que describe esta curva está dada por una expresión bastante compleja, en términos de la media μ y la desviación estándar σ, que es,

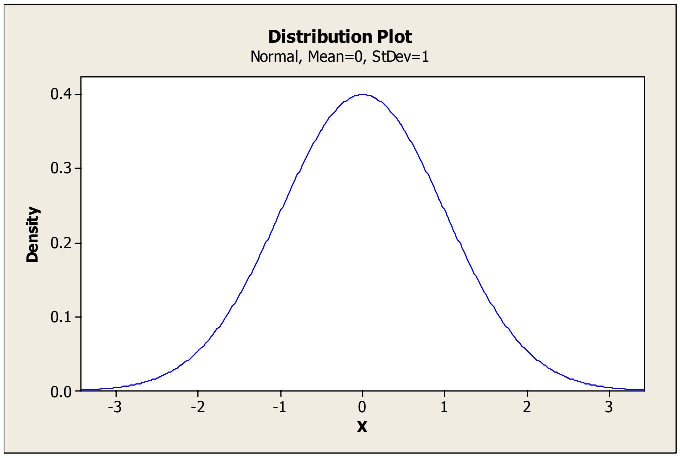
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{\frac{-[(x-\mu)/\sigma]^2}{2}} . \tag{9.9}$$

Aquí π y exp son los números irracionales 3.1416... y 2.7182..., respectivamente. La media μ nos proporciona geométricamente el centro de la curva. La desviación estándar σ, determina el ancho de la misma.

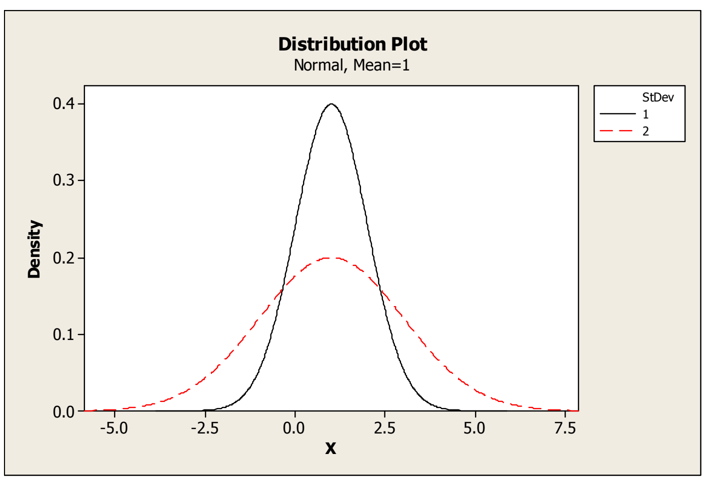
La **distribución normal estándar** es aquella en la cual la función de densidad tiene $\mu = 0$ y $\sigma = 1$, de manera que la expresión 9.9 se simplifica a

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp^{\frac{-x^2}{2}},$$

cuya gráfica es,

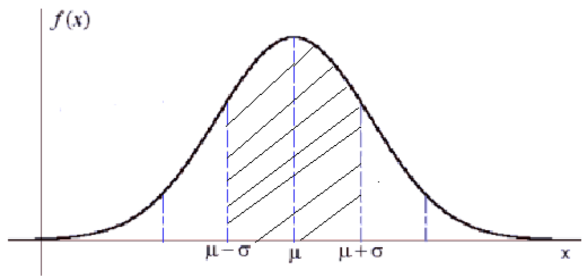


Si cambiamos a $\mu = 1$, se tiene que la curva se centra en el 1 y si ponemos $\mu = 1$ y $\sigma = 2$, la curva se vuelve más ancha. Estas consideraciones se muestran a continuación:



La simetría de la curva normal tiene como consecuencia que su media, su mediana y su moda sean todas iguales.

Dada una distribución normal, el 68 % de los datos se encuentran a una distancia de σ de la media, el 95 % a una distancia de 1.96σ y el 99 % a una distancia[5] de 2.58σ. Es decir, el área bajo la curva entre $\mu-\sigma$ y $\mu+\sigma$ es de 0.68, el área entre $\mu-1.96\sigma$ y $\mu+1.96\sigma$ de 0.95 y el área entre $\mu-2.58\sigma$ y $\mu+2.58\sigma$ de 0.99. En la siguiente figura ilustramos el área entre $\mu-\sigma$ y $\mu+\sigma$:



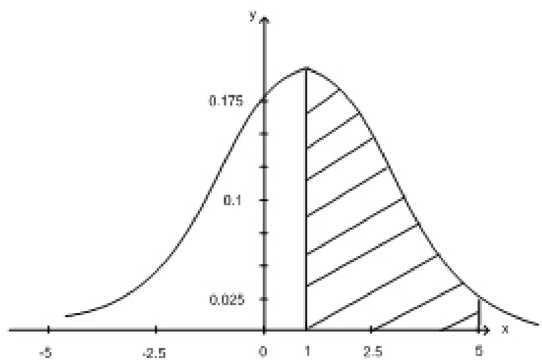
Cuando una variable aleatoria X está distribuida normalmente con media μ y desviación estándar σ, denotamos esto por

$$X \sim N(\mu, \sigma).$$

La probabilidad de que X tome valores entre dos números dados es el área bajo la curva entre dichos números. Por ejemplo, si $X \sim N(1,2)$, entonces la probabilidad de que X tome valores entre 1 y 5, dada por $P(1 < X < 5)$,

[5]Con frecuencia se aproxima y se dice que el 95 % de los datos está a una distancia de 2σ y el 99 % de los datos está a una distancia de 3σ.

es el área sombreada de la siguiente figura:



Hay que notar que mediante un cambio de variable adecuado, concretamente,

$$Z = \frac{X - \mu}{\sigma},$$

si $X \sim N(\mu, \sigma)$, entonces $Z \sim N(0, 1)$. Intiutivamente, tomamos a la distribución original, con media μ y desviación estándar σ, primero la trasladamos μ unidades para centrarla en el cero $(X - \mu)$ y posteriormente la adelgazamos o ensanchamos para que su desviación estándar sea la unidad $(\frac{X-\mu}{\sigma})$. De esta forma, cualquier distribución normal X puede transformarse en una normal estándar Z. Es usual denotar por Z a la variable aleatoria que sigue una distribución normal estándar.

El cálculo de las áreas bajo la curva no es trivial, de manera que existen tablas para encontrarlas. Al final del capítulo se incluye una de estas tablas con los valores de $P(Z < a)$ para la distribución normal estándar. Alternativamente, los cálculos pueden efectuarse con *Excel, Minitab* o con cualquier software matemático o estadístico.

Ejemplos

Ej 9.4.1 Utilizando la tabla al final del capítulo vemos que si Z sigue una distribución normal estándar, entonces

$$
\begin{aligned}
P(0 < Z < 2.5) &= 0.4938, \\
P(Z > 0) &= 0.5, \\
P(0 < Z < 3.15) &= 0.4992. \\
P(0.5 < Z < 2) &= P(0 < Z < 2) - P(0 < Z < 0.5) \\
&= 0.4772 - 0.1915 = 0.285\,7 \\
P(-1 < Z < 0) &= P(0 < Z < 1) = 0.3413.
\end{aligned}
$$

Notemos que, aunque la tabla no contiene los valores para números negativos, la simetría de la gráfica nos permite utilizar el área correspondiente al valor positivo, es decir si $a > 0$,

$$
\begin{aligned}
P(-a < Z < 0) &= P(0 < Z < a), \\
P(Z < -a) &= P(Z > a)\ .
\end{aligned}
$$

Ej 9.4.2 Supongamos que el ingreso anual de un imigrante mexicano en los EUA se distribuye de forma normal con una media de \$30 000 y una desviación estándar de \$10000 dólares. ¿Cuál es la probabilidad de que un inmigrante mexicano gane menos de \$20 000 dólares anuales? ¿Cuál es la probabilidad de que su ingreso sea mayor a los \$50 000 dólares anuales?

La variable aleatoria X tiene que transformarse en una normal estándar. Esto se hace considerando $Z = \frac{X-30000}{10000}$, así, despejando X,

$$X = 5000Z + 30000$$

y por lo tanto

$$\begin{aligned} P(X < 20000) &= \\ P\left(Z < \frac{20000 - 30000}{10000}\right) &= \\ P(Z < -1) &= \\ P(Z > 1) &= \\ P(Z > 0) - P(0 < Z < 1) &= \\ 0.5 - 0.3413 = 0.1587. \end{aligned}$$

Entonces, se espera que el 15.87 % de los imigrantes mexicanos ganen menos de \$20 000 dólares al año. Similarmente,

$$\begin{aligned} P(X > 50000) &= \\ P\left(Z > \frac{50000 - 30000}{10000}\right) &= \\ P(Z > 2) &= \\ P(Z > 0) - P(0 < Z < 2) &= 0.5 - 0.4772 \\ &= 0.0228. \end{aligned}$$

En este caso, el porcentaje de los imigrantes que se espera ganen más de \$50000 es del 2.28 %.

♠♠♠♠♠

9.5. En el límite, todo es normal.

En la sección anterior mencionamos que De Moivre utilizaba la distribución normal para simplificar los cálculos de la distribución binomial. El hecho de que para un gran número de ensayos una distribución binomial se parezca, cada vez más a una normal, es un resultado parecido al Teorema de Bernoulli visto en el ejercicio 9.3.4. Esencialmente es también una *ley de los grandes números*. La versión más general de este tipo de resultados es el llamado **Teorema del Límite Central.**

Este teorema se le atribuye al matemático ruso Aleksandr Lyapunov (1857 - 1918). En términos simples, el teorema nos dice que si tenemos cierto número de variables aleatorias independientes con la misma distribución, al aumentar su número, la distribución de la suma de sus resultados se aproxima a una distribución normal. ¡Lo increíble es que esto sucede independientemente de cual sea la distribución original de las variables!

En particular, la media es simplemente una suma, de manera que el resultado del teorema aplica si cambiamos la palabra suma por media. Podemos, por ejemplo, iniciar con una distribución uniforme, tomar muestras aleatorias de tamaño dos y calcular sus medias. Después procedemos con muestras de tamaño tres, cuatro y así sucesivamente aumentamos el tamaño de la muestra. La distribución de estas medias se aproxima a una normal con la misma media de la distribución original. La varianza de esta distribución normal disminuye conforme aumentamos el tamaño de la muestra.

Ilustremos lo que sucede con la distribución binomial. Ésta, por construcción, es una suma de variables aleatorias o ensayos independientes, de manera que aumentando el número de ensayos incrementamos el tamaño de la muestra. Supongamos que comenzamos con una distribución binomial con parámetros n y p, es decir, nuestra variable aleatoria es el número de éxitos en n ensayos cuando la probabilidad de éxito para cada ensayo es de p. De acuerdo a las expresiones 9.7 y 9.8, la distribución normal que aproxima a esta binomial tendrá los mismos parámetros, es decir,

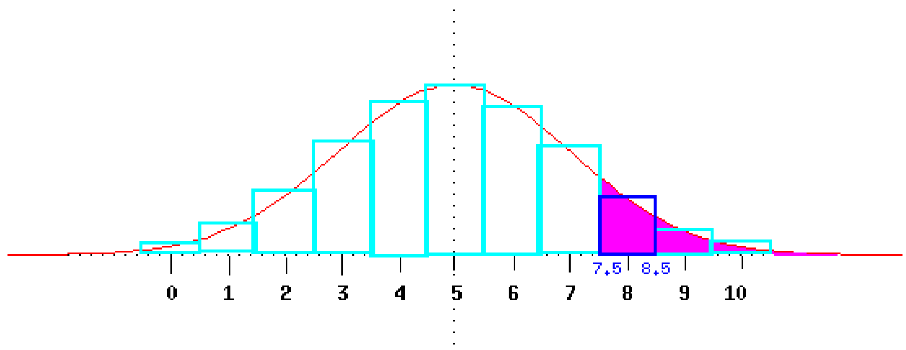
$$\begin{aligned} &\mu = np, \\ &\sigma = \sqrt{\sigma^2} = \sqrt{np(1-p)}. \end{aligned}$$

En general la aproximación normal es considerada *buena* si $np \geq 10$ y $np(1-p) \geq 10$.

Sea B una variable con distribución binomial y X su aproximación normal. ¿Cómo se calcula la probabilidad de k éxitos $P(B = k)$ con la aproximación normal? Sabemos que en el modelo normal, $P(X = k)$ no tiene sentido pues se trata de una distribución continua. Para calcular una aproximación razonable se aplica la **corrección por continuidad** y se piensa al entero k como representado por el intervalo $(k - 0.5, k + 0.5)$, de manera que

$$P(B = k) = P(k - 0.5 < X < k + 0.5).$$

En la siguiente figura se representa la aproximación $P(B > 7) = P(X > 7.5)$. Los rectángulos representan a la variable binomial B y la curva continua es la aproximación normal X.



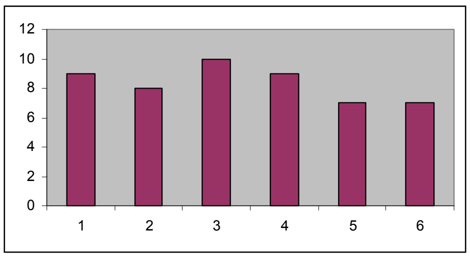
Notemos que se tendrían también las siguientes probabilidades:

$$P(B = 7) = P(6.5 < X < 7.5),$$
$$P(B \geq 7) = P(X \geq 6.5).$$

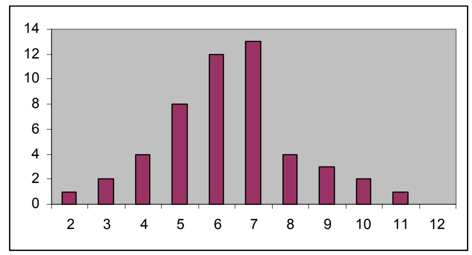
Es de resaltar que $P(B > 7) = P(X > 7.5)$ y $P(B \geq 7) = P(X \geq 6.5)$, ya que en éste último tiene que incluir al área del rectángulo correspondiente a $P(B = 7)$.

Ejemplos

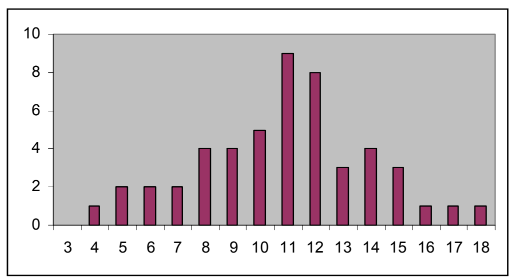
Ej 9.5.1 Este ejemplo ilustra el teorema del límite central cuando las distribuciones iniciales son uniformes. Lanzamos un dado 50 veces y hacemos un histograma con los resultados. Después tomamos dos dados y los lanzamos 50 veces. Sumamos los números de las dos caras que aparecen en cada lanzamiento y realizamos un histograma con ellos. Tomamos ahora tres dados y lanzamos los tres juntos 50 ocasiones. Sumamos los números de las tres caras que aparecen en cada lanzamiento y realizamos un histograma con los resultados. Las siguientes figuras muestran estos histogramas. Visualmente observamos como la distribución de la suma de las caras tiende a ser normal.



Un dado

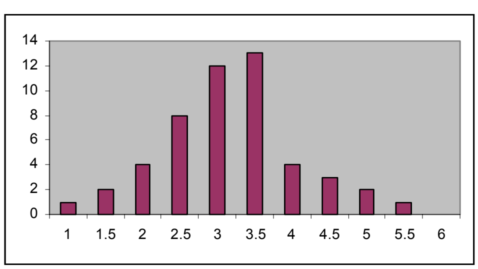


Dos dados

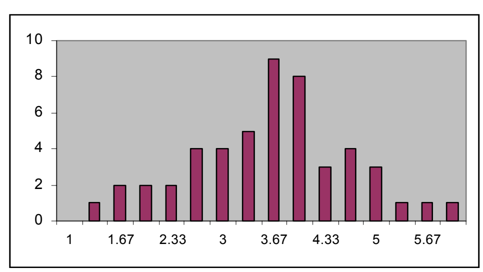


Tres dados

Las histogramas de las medias para dos y tres dados se ven igual, pero en el eje horizontal se encuentran los promedios y no la suma de los números de las caras. Vemos que las medias se concentran entre 3 y 4.



Medias de dos dados



Medias de tres dados

Estos histogramas convergen a los de una distribución normal con media de 3.5, igual a la de la distribución uniforme original.

Ej 9.5.2 Consideremos una población de votantes en la ciudad de México. Se reporta que la proporción de votantes que favorecen al candidato del partido Naranja es igual a 0.40. Dada una muestra aleatoria de 200 votantes, ¿cuál es la probabilidad de que más de la mitad de ellos tengan intención de voto por el candidato naranja? En este caso la variable binomial B tiene parámetros $n = 200$ y $p = 0.4$. Utilizando 9.7 y 9.8 obtenemos,

$$\begin{aligned} \mu &= 80, \\ \sigma &= \sqrt{48} = 6.93 \end{aligned}$$

Sea $X \sim N(80, 6{,}93)$ la aproximación normal de B. Por lo tanto, la normal estándar asociada es,

$$Z = \frac{X - 80}{\sqrt{48}}.$$

La mitad de los votantes son 100 y podemos calcular, aplicando la corrección por continuidad, la probabilidad correspondiente como

$$\begin{aligned} P(B > 100) &= 1 - P(B \leq 100) \\ &= 1 - P(X \leq 100.5) \\ &= 1 - P\left(Z \leq \frac{100.5 - 80}{\sqrt{48}}\right) \\ &= 1 - P(Z \leq 2.96) \\ &= 1 - 0.9985 = 0.0015. \end{aligned}$$

De aquí que la probabilidad de que en la muestra de votantes, más de la mitad expresen preferencia por el candidato naranja es bajísima: el 0.15 %.

Ej 9.5.3 Si queremos calcular la probabilidad de obtener 60 águilas en cien lanzamientos de una moneda, entonces $n = 100$ y $p = 0.5$. La distribución normal X que aproxima esta binomial -digamos B- tiene como parámetros[6],

$$\mu = 50,$$
$$\sigma = \sqrt{25} = 5.$$

La normal estándar asociada es

$$Z = \frac{X - 50}{5},$$

por lo tanto,

$$\begin{aligned} P(B = 60) &= P(59.5 < X < 60.5) \\ &= P(\frac{59.5 - 50}{5} < Z < \frac{60.5 - 50}{5}) \\ &= P(1.9 < Z < 3.1) \\ &= P(Z < 3.1) - P(Z < 1.9) \\ &= 0.999 - 0.9713 = 0.0277. \end{aligned}$$

Esto nos dice que la probabilidad de obtener 60 águilas en 100 lanzamientos es de aproximadamente 2.77 %.

♠♠♠♠♠

9.6. Distribución muestral de la media

El teorema del límite central resalta la conveniencia de elegir muestras aleatorias de una población, para inferir alguna característica de la población a partir de las mismas. Así, los elementos que forman parte de la muestra tienen la misma distribución de probabilidad y son independientes entre sí. Cuando la muestra se compone de, supongamos algún artefacto producido masivamente, es sencillo tomar muestras aleatorias y someterlas a pruebas de control de calidad para inferir, digamos una calidad mínima en la población total.

[6] Una vez más utilizando 9.7 y 9.8.

Cuando se trata de poblaciones humanas y se eligen muestras para realizar encuestas acerca de alguna opinión personal, no es fácil que la muestra sea verdaderamente aleatoria. Por ejemplo, si elegimos números de teléfono aleatoriamente, la muestra sólo incluirá a los habitantes con teléfono. De éstos, los que aceptan ser encuestados seguramente quieren expresar su opinión sobre el tema en cuestión. Adicionalmente, el encuestado puede mentir en su respuesta. El buen encuestador debe saber elegir las muestras y compensar por los posibles errores de muestreo.

Supongamos que tenemos una población y consideremos todas las posibles muestras de tamaño n de la misma. Siempre pensaremos que la población es suficientemente grande para que la muestra nunca exceda el 5 % de la población[7]. Para cada una de estas muestras puede calcularse la media y obtenerse una **distribución muestral de medias**. Por el teorema del límite central, sabemos que si las muestras son suficientemente grandes, esta distribución se aproximará por una distribición normal. El siguiente resultado nos proporciona la media y la desviación estándar para esta distribución muestral.

Proposición 9.6.1 *Comencemos con una población cuya media es μ y desviación estándar σ. Las medias $\bar{X}$ de las muestras aleatorias de tamaño n siguen una distribución con media μ y desviación estándar $\frac{\sigma}{\sqrt{n}}$. Si n es suficientemente grande (usualmente $n \geq 30$), entonces, aproximadamente se tiene que,*

$$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}}),$$

es decir, las medias están distribuidas normalmente y la desviación estándar de las medias de la muestra: $\frac{\sigma}{\sqrt{n}}$, se hace muy pequeña cuando n es grande.

[7]Cuando la población es relativamente pequeña y el tamaño de la muestra excede el 5 % del tamaño de la población, tienen que aplicarse algunos factores de corrección. En este texto, todas las poblaciones son sufucientemente grandes (infinitas), de manera que esto no sucede.

Ejemplo

Ej 9.6.1 Una muestra aleatoria de 36 individuos se toma de una población cuyo ingreso mensual promedio es de \$25000 con una desviación estándar de \$6000. ¿Cual es la probabilidad de que la media de la muestra esté entre 23000 y 27000? La media $\bar{X}$ de las muestras de tamaño $n = 36$ sigue aproximadamente una distribución normal $N(25000, \frac{6000}{\sqrt{36}}) = N(25000, 1000)$. La probabilidad buscada es,

$$P(23000 < \bar{X} < 27000).$$

Sin necesidad de buscar en la tabla sabemos que ésta es aproximadamente 0.95, pues para cualquier distribución normal el 95 % del área se encuentra a 1.96 desviaciones estándar de la media que aquí aproximamos a 2 desviaciones estándar o bien $2 \times 1000 = 2000$.

♠♠♠♠♠

Bibliografía Complementaria

1. Agresti Alan, Finlay Barbara, *Statistical Methods for the Social Sciences*,4^a ed, Pearson, Allyn & Bacon, New York, NY, 2008.

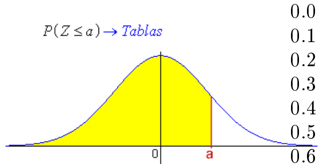
2. Aguirre Víctor et al., *Fundamentos de Probabilidad y Estadística*, 2^a edición. Jit Press, México, D.F. 2006.

3. Dekking F. Michel et al., *A Modern Introduction to Probability and Statistics : understanding why and how*, Springer-Verlang, , London, UK 2005.

4. Kaye, David H., Freedman, David A., "A Reference Guide on Statistics", *Reference Manual on Scientific Evidence*, 2nd edition, Federal Justice Center. Versión electrónica en,

   http://www.fjc.gov/public/pdf.nsf/lookup/sciman01.pdf/\$file/sciman01.pdf

NORMAL CUMULATIVE DISTRIBUTION FUNCTION



x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Ejercicios

▷ **9.1** ¿Cuál es la probabilidad de que en una lotería en la cual se eligen aleatoriamente los números del 1 al 100, el número ganador sea mayor que 65?

▷ **9.2** Sea X una variable aleatoria que se distribuye de manera uniforme en el intervalo $[0, 3]$. Encontrar:

1. $P(X < 2)$,
2. $P(X \geq 1)$,
3. $P(1.5 < X \leq 3)$.

▷ **9.3** Dibujar la función de densidad para la variable aleatoria del ejercicio anterior.

▷ **9.4** El número de mililitros de café en un vaso tamaño "alto" en *Starbucks®*, se distribuye uniformemente entre 235 y 245 ml.

1. ¿Cuál es la probabilidad de comprar un café alto que contenga menos de 240 ml?
2. ¿Cuántos mililitros contiene el vaso medio de café?
3. ¿Cuál es la desviación estándar?

▷ **9.5** Sea X una variable aleatoria binomial, n el número de ensayos y π la probabilidad de éxito de cada ensayo. Encontrar:

1. $P(X = 3)$ si $n = 8$ y $\pi = 0.2$,
2. $P(X = 5)$ si $n = 10$ y $\pi = 0.3$,
3. $P(X > 4)$ si $n = 12$ y $\pi = \frac{2}{3}$,
4. $P(X \leq 5)$ si $n = 7$ y $\pi = \frac{1}{2}$.

▷ **9.6** La probabilidad de que el jugador de basketball Shaquille O'Neal enceste un tiro libre es del 60 %. Utilizar la aproximación normal para calcular la probabilidad de que en un juego, en el cual va a línea 25 veces, enceste al menos en 15 ocasiones.

▷ **9.7** Supongamos que el examen final de Economía es de opción múltiple. Éste consiste de 30 preguntas, cada una con 4 opciones de las cuales sólo una es la correcta.

1. ¿Cuál es la probabilidad de que un estudiante obtenga al menos 18 preguntas correctas en el examen si responde todas de forma aleatoria?
2. Calcular la media y la desviación estándar del número de respuestas correctas.

3. Estimar la misma probabilidad del inciso 1 utilizando la aproximación normal.

▷ **9.8** Realizar un histograma de probabilidad para las siguientes variables binomiales:

1. X es el número de soles que se obtiene al lanzar una moneda 6 veces.
2. X es el número de unos que se obtiene al lanzar un dado 5 veces.

▷ **9.9** Sea X una variable aleatoria que sigue una distribución normal estándar, encontrar:

1. $P(0 < X < 3.1)$,
2. $P(X < -1)$,
3. $P(X > 2)$,
4. $P(-1.1 < X < 2.2)$,
5. $P(X \leq 1.35)$.

▷ **9.10** Sea X una variable aleatoria distribuida normalmente con media μ y desviación estándar σ. Encontrar las siguientes probabilidades:

1. $P(110 < X < 180)$ si $\mu = 100$ y $\sigma = 20$,
2. $P(50 < X < 100)$ si $\mu = 75$ y $\sigma = 10$,
3. $P(X > 3)$ si $\mu = 0$ y $\sigma = 1$,
4. $P(X < 25)$ si $\mu = 15$ y $\sigma = 4$,
5. $P(X > -1)$ si $\mu = -2$ y $\sigma = 1$.

▷ **9.11** ¿Porqué las tablas para la distribución normal estándar sólo llegan hasta $Z = 4$? Justifica.

▷ **9.12** En un curso de Teoría del Derecho, las calificaciones finales quedaron distribuidas normalmente con media de 75 y desviación estándar de 12.

1. Aproximadamente, ¿qué porcentaje de los alumnos obtuvo más de 60? (Es decir, pasaron el curso).
2. Aproximadamente, ¿qué porcentaje del grupo obtuvo más de 90?

▷ **9.13** El peso promedio al nacer de un bebé de 40 semanas de gestación está distribuido normalmente con media de 2.8 kg y desviación estándar de 0.5 kg. ¿Cuál es la probabilidad de que un bebé pese más de 4 kg?

▷ **9.14** Tomar los resultados del ejemplo 9.5.2. Supongamos que el equipo de campaña del candidato naranja cita una encuesta con una muestra de 200 personas. “Nuestro candidato es preferido por el 51 % de los votantes”, dicen a la prensa.

1. Trabajamos para el candidato del partido opositor y queremos demostrar que mienten. Proporcionar argumentos válidos.

2. Trabajamos para el candidato naranja y queremos responder a las críticas del equipo del candidato B. Proporcionar argumentos válidos.

▷ **9.15** Tomar tres papeles, numerarlos con los números 1,2 y 3 y doblarlos de manera que no pueda reconocerse uno de otro. Poner los tres papeles en algún recipiente (urna) y repetir el siguiente experimento (ensayo) diez veces: se extrae un papel al azar, se anota su número, se regresa a la urna y se extrae otro papel al azar, anotando su número. Calcular el promedio de estos números.

1. Llenar la siguiente tabla con los datos obtenidos:

# ensayo	# 1	# 2	promedio
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

2. ¿Cuáles son los valores posibles para el promedio de los números?

3. Realiza una tabla de frecuencias y de frecuencias relativas con los datos obtenidos para los promedios.

4. Con los datos de la tabla anterior, realiza un histograma de frecuencias relativas para los promedios. Este es el histograma que representa a la distribución muestral de la media.

▷ **9.16** Los miembros del H. Congreso de la Unión tienen un promedio de ausentismo del 22 días con una desviación estándar de 5 días. Si se toma una muestra aleatoria de 25 congresistas (asumir que los días de ausentismo siguen una distribución normal) ¿Cuál es la probabilidad de que su promedio de ausentismo sea,

1. menos del 19 días?
2. más del 24 días?

▷ **9.17** El promedio académico de un estudiante del ITAM es de 7.5 con una desviación estándar de 2. Se toma una muestra aleatoria de 49 estudiantes, ¿aproximadamente, cuál es la probabilidad de que su promedio

1. sea menor que 6.6?
2. mayor que 8.1?

Capítulo 10

Elementos de inferencia

10.1. Introducción

El objetivo principal de la inferencia estadística es el describir alguna característica de una población, en base a la información que se tiene de una o varias muestras de la misma. De esta forma, podemos inferir que el salario promedio mensual de un abogado en México es de $9500, sin necesidad de considerar a toda la población de abogados en el país, ya que es suficiente el estudio de una muestra aleatoria de éstos. Para llevar a cabo una encuesta de opinión política, no es necesario consultar a toda la población de votantes para concluir que cierto porcentaje de ésta favorece a un candidato. Un estadístico destaca sobre los demás por sus aportaciones al marco formal para realizar este tipo de inferencias: Ronald Aylmer Fisher (1890 - 1962), matemático inglés, cuyo interés por la agricultura y la genética lo llevó a dirigir el centro de investigación agrícola más importante de su época en el Reino Unido[1].

Fisher contribuyó gran parte de las ideas fundamentales que utiliza la estadística moderna y originó lo que se conoce como *diseño de experimentos*. La aleatoriedad del muestreo que hoy nos puede parecer obvio fue introduci-

[1] El *Rothamsted Agricultural Experiment Station* establecido en 1837 para estudiar los efectos de fertilizantes, nutrientes y variedades genéticas en la productividad agrícola.

da por Fisher. Asimismo, formalizó lo que hoy conocemos como pruebas de hipótesis y acuñó el término de **hipótesis nula**, sólo por mencionar algunas de sus aportaciones.

Como individuo Fisher fue controversial. Entre otras cosas, proponía que la ayuda gubernamental fuese sesgada hacia quienes eran genéticamente más fuertes para optimizar su eficacia. Nunca le perdonó a Karl Pearson (1857 - 1936), otro estadístico famoso de la época, el haber publicado una crítica a uno de sus trabajos y se sentía incomprendido pues los biólogos no sabían matemáticas y los matemáticos no sabían biología. Dejando la controversia aparte, es incuestionable que Fisher era un individuo fuera de serie que revolucionó la ciencia estadística.

Usualmente las hipótesis surgen para tratar de explicar algún fenómeno. En la sección 4.5 vimos como actualizar las probabilidades *a priori* de un conjunto de hipótesis, conforme se obtiene evidencia nueva. Para este propósito se requerían las probabilidades condicionales de (observar) la evidencia dada la hipótesis, es decir:

$$P(E \mid H).$$

Cuando la hipótesis está enunciada en términos de algún parámetro poblacional, entonces su *validez* puede ser puesta a prueba con la evidencia obtenida. Esta evidencia consiste en datos obtenidos a partir de una o de varias muestras.

El caso clásico o frecuentista, parte de una hipótesis inicial llamada **hipótesis nula**, usualmente denotada por H_0. Ésta representa el *status quo* y usualmente no es rechazada, en favor de la **hipótesis alterna o rival** (denotada por H_1 o H_a), a menos que la evidencia sugiera *fuertemente* que es falsa. La hipótesis nula se refiere a alguna característica de una población, como la media ("la estatura media de una mujer mexicana es de..."), la desviación estándar o la proporción de individuos perteneciente a cierta categoría ("el 47 % de los votantes favorecen al candidato del partido..."). Veamos como se procede.

10.2. Significancia

Imaginemos que un estudiante realiza un examen de matemáticas de opción múltiple. El examen consta de 10 preguntas, cada una con 5 opciones de respuesta, de las cuales sólo una es la correcta. Este estudiante obtiene 5 preguntas correctas. Notemos que si las respuestas se eligen en forma aleatoria, la probabilidad de éxito es de $\frac{1}{5}$ para cada pregunta. Podemos pensar al proceso de contestar el examen aleatoriamente como un experimento binomial con 10 ensayos y probabilidad de éxito de $\frac{1}{5}$ en cada ensayo. De acuerdo a la fórmula 9.7, la media de esta distribución es de

$$\mu = 10 \times \frac{1}{5} = 2,$$

de manera que éste es el número de aciertos esperados al contestar de forma aleatoria. ¿Qué tan plausible es que el estudiante haya obtenido al menos 5 aciertos? ¿Será meramente producto del azar o quizás posee algún conocimiento de matemáticas?

Para contestar estas preguntas planteamos las siguientes hipótesis complementarias:

$$H_0 : \text{el estudiante contestó al azar,}$$
$$H_1 : \text{el estudiante no contestó al azar.}$$

La probabilidad de tener al menos 5 aciertos puede calcularse como,

$$P(X \geq 5) = 1 - P(X < 5).$$

En donde

$$P(X < 5) = \sum_{i=0}^{4} P(X = i)$$

De acuerdo a la proposición 9.3.2 tenemos que si H_0 es cierta y $p = \frac{1}{5}$,

$$
\begin{aligned}
P(X=4) &= {}_{10}C_4\,(0.2)^4\,(0.8)^{10-4} = \frac{10!}{6!4!}\,(0.2)^4\,(0.8)^6 = 0.088\,08 \\
P(X=3) &= \frac{10!}{7!3!}\,(0.2)^3\,(0.8)^7 = 0.201\,33, \\
P(X=2) &= \frac{10!}{8!2!}\,(0.2)^2\,(0.8)^8 = 0.301\,99, \\
P(X=1) &= \frac{10!}{9!1!}\,(0.2)^1\,(0.8)^9 = 0.268\,44, \\
P(X=0) &= \frac{10!}{10!0!}\,(0.2)^0\,(0.8)^{10} = 0.107\,37.
\end{aligned}
$$

Por lo tanto, sumando éstas probabilidades se tiene,

$$P(X < 5) = 0.967\,21$$

y de aquí,

$$P(X \geq 5) = 1 - 0.967\,21 = 0.033.$$

En conclusión, la probabilidad de que el estudiante haya obtenido al menos 5 aciertos contestando de manera aletoria es de 3.3 %. Existen dos posibilidades:

- H_0 es verdadera y estamos ante un acontecimiento inusual que sólo ocurre el 3.3 % del tiempo o bien,
- H_0 es falsa.

En el segundo caso, dado que las hipótesis son complementarias, tenemos que H_0 se rechazaría en favor de H_1. Si consideramos que los 5 aciertos no fueron obra de la casualidad, entonces, ¿podemos concluir que el estudiante sabe algo de matemáticas? Desafortunadamente no, ya que no puede descartarse que haya copiado algunas de las respuestas, ¡pero al menos podríamos darle el beneficio de la duda! Intuitivamente, los 5 aciertos obtenidos podrían ser suficientemente *significativos* para descartar que son producto del azar. ¿Qué tan significativos? A continuación trataremos de precisar este concepto.

Definición 10.2.1 *La probabilidad de obtener los datos o evidencia, a partir de una muestra, dado que la hipótesis nula es verdadera, se conoce como* ***valor* p**.

El valor p para la hipótesis H_0 referente al estudiante que contesta el examen de matemáticas es de $p = 0.033$. Este valor también puede interpretarse como la probabilidad de rechazar erróneamente la hipótesis nula, es decir, cuando ésta es verdadera pero es rechazada por considerar a la evidencia demasiado improbable. ¿Con qué valor p podemos rechazar la hipótesis nula? En general, es estándar tomar $p < 0.05$ o bien $p < 0.01$; sin embargo, depende de cada experimento y de la trascendencia del rechazo. ¿Se trata de una situación de vida o muerte o de algo con consecuencias menos drásticas? Cada caso será distinto.

Un valor p *pequeño*, no favorece a la hipótesis nula, un valor p *grande* indica que los datos obtenidos son compatibles con la hipótesis nula. Supongamos que una trabajadora demanda a su empleador por discriminación salarial en contra de las mujeres. El abogado que toma su caso realiza un estudio estadístico comparando la media salarial entre hombres y mujeres que realizan el mismo tipo de trabajo y tienen educación y experiencia similares. Se pone a prueba la hipótesis nula,

$$H_0: \text{ no existe diferencia entre las medias salariales de hombres y mujeres.}$$

en contra de la hipótesis alterna,

$$H_1: \text{existe diferencia entre las medias salariales de hombres y mujeres.}$$

Entre más pequeño sea el valor p que resulte de la prueba, mayor oportunidad tendrá la demandante de ganar el caso. De forma análoga, los valores p grandes favorecen a la defensa.

El siguiente concepto de **significancia** proporciona una forma de precisar que se quiere decir por *pequeño* y *grande*:

Definición 10.2.2 *La probabilidad máxima con la cual podemos rechazar la hipótesis nula se denota por α y se denomina el* ***nivel de significancia*** *de la prueba. Es decir, la hipótesis nula se rechaza si $p < \alpha$.*

Los valores de $\alpha = 0.05$ y $\alpha = 0.01$ son los más comunes y no son casuales. Recordemos que para una distribución normal, el 95 % de los datos se encuentran a una distancia de 1.96σ de la media y el 99 % a una distancia de 2.58σ y estos valores suelen aproximarse por 2σ y 3σ. Dado que en la mayoría de los casos la distribución de los datos es aproximada por una normal, éstas probabilidades son particularmente fáciles de calcular sin ninguna ayuda computacional.

En las ciencias sociales es común tomar el valor $\alpha = 0.05$ como nivel de significancia estándar; así, cuando se habla de una prueba *significativa* se tiene que $p < 0.05$. Analogamente, cuando hay referencia a una prueba *altamente significativa*, usualmente se trata de $p < 0.01$. La significancia estadística se determina comparando el valor p obtenido con un valor preestablecido α, que es el nivel de significancia. Si $p \geq \alpha$ entonces se dice que la evidencia no es significativa y la hipótesis nula no puede rechazarse. De forma análoga, si $p < \alpha$ entonces la evidencia es significativa y la hipótesis nula se rechaza.

10.3. Cómo desconfiar de la hipótesis nula

La noción de rechazar la hipótesis nula -en favor de la hipótesis alterna- como forma de proceder, sigue el principio de falsificación introducido por el filósofo austriaco Karl Popper (1902 - 1994). Este principio estipula que podemos negar pero no afirmar algo en forma concluyente. El conocimiento se construye, entonces, a partir de la imposibilidad de la negación de las hipótesis que se introducen para explicar la realidad. Una hipótesis no puede desecharse, a menos que haya evidencia suficiente para ponerla en duda.

Originalmente, Fisher construyó las pruebas de hipótesis para aplicaciones sumamente prácticas, concretamente: experimentos agrícolas. Para aumentar la productividad en el campo se utilizan fertilizantes y pesticidas y cuando se introduce un nuevo producto es deseable determinar su eficacia. Supongamos que quiere determinarse si la adopción del uso de cierto fertilizante para el maiz es deseable. Los pasos a seguir son los siguientes:

- Diseñar un experimento con una muestra de parcelas en las que se aplica el fertilizante.

- Elegir una cantidad que represente a los datos que nos interesa medir, usualmente se trata de una media, de una desviación estándar o de una proporción. En este caso queremos representar el rendimiento o productividad de las parcelas. La cantidad a medir podría ser la cantidad promedio de maíz producida, digamos que $\mu =$ rendimiento promedio. Sea $\mu = R_0$ el valor actual (*status quo*) del rendimiento promedio del maíz.

- Determinar la hipótesis nula y la alterna, que en este caso son,

$$H_0 : \text{El fertilizante no sirve,}$$
$$H_1 : \text{El fertilizante sirve.}$$

  En términos del rendimiento promedio éstas se reescriben como,

$$H_0 : \mu \leq R_0,$$
$$H_1 : \mu > R_0.$$

- Establecer un nivel de significancia, por ejemplo $\alpha = 0.05$.

- Obtener los datos acerca de los rendimientos de cada parcela de la muestra y calcular el rendimiento promedio $\bar{R}$. Digamos que éste es de $\bar{R} > R_0$. Evidentemente, si sucediera que $\bar{R} \leq R_0$, ¡entonces ya no tenemos nada más que hacer pues H_0 no será rechazada!

- Calcular la probabilidad p de obtener un rendimiento de $\bar{R}$ o mayor, dada la hipótesis nula ($\mu \leq R_0$), éste es el valor p.

Si $p < 0.05$, entonces concluimos que estamos ante algo muy inusual, que sólo ocurre con probabilidad menor al 5 % o bien, la hipótesis nula es falsa. Se dice que los resultados son significativos a un nivel de 0.05. Dado que 0.05 es el nivel de significancia preestablecido, se procede a rechazar la hipótesis nula -en favor de la alterna-, es decir, el fertilizante sirve.

Este mismo procedimiento puede aplicarse a una gran variedad de situaciones: pruebas de control de calidad, nuevos medicamentos, efectos de cambios en la alimentación, efectos de una nueva legislación, efectos de políticas económicas y casi cualquier cosa que podamos imaginar. Desgraciadamente, existe gran descuido y abuso de este tipo de pruebas.

10.4. Mal uso y abuso

Un problema grave es la interpretación incorrecta del resultado de una prueba de hipótesis. Es común observar casos legales en los cuales un juez interpreta el nivel de significancia, digamos $\alpha = 0.05$, como la probabilidad de que la hipótesis nula sea válida. Otro problema común es que la hipótesis alterna no esté especificada correctamente. En el caso de los granjeros no hay duda de cuál es ésta; sin embargo, no siempre es tan simple.

Lo anterior es importante pues, en general, desea rechazarse la hipótesis nula en favor de la alterna. Por ejemplo, si se especifica la hipótesis nula: "el 90 % de los crímenes cometidos en la ciudad de México quedan impunes", la hipótesis alterna es que el porcentaje que queda impune es distinto del 90 % y el rechazo de la hipótesis nula en favor de la alterna no es muy útil. La opción es especificar la hipótesis nula de alguna otra forma de manera que su rechazo nos proporcione información. En este ejemplo podría ser: "Al menos un 90 % de los crímenes cometidos en la ciudad de México quedan impunes".

Otro problema que surge, por desgracia con gran frecuencia, es el llevar a cabo experimentos en los cuales el resultado es un valor p, digamos $p < 0.01$ y ahí concluye la investigación. Para ilustrar, supongamos que realizamos un estudio en el cual se hace un cuestionario a una muestra aleatoria de adultos. Se especifica la hipótesis nula como:

$$H_0 : \text{comer brócoli no afecta la incidencia de cáncer de colon},$$

estableciendo un nivel de significancia de 0.01. Se obtiene un valor $p < 0.01$ y por lo tanto la hipotesis es rechazada. A los pocos días una revista de gran circulación nos incita a comer grandes cantidades de brócoli para prevenir el cancer de colon.

¿Será el brócoli preventivo del cancer de colón? En realidad no sabemos. Es muy probable que los pacientes que reportan brócoli en su dieta cotidiana, sean personas que cuidan su alimentación y, adicionalmente, llevan una vida generalmente sana. El comer brócoli podría ser trascendente para prevenir el cancer de colón o podría no serlo. El punto es que habría que investigar más a fondo. Las pruebas de significancia, como su nombre lo indica, deben ser una señal para detectar que vale la pena estudiar algo, es decir, deben ser el inicio de un proyecto de investigación y no el fin.

Con frecuencia, el experimento para la prueba de una hipótesis se repite un gran número de veces. Por ejemplo, supongamos que se quiere determinar si una moneda es honesta. Sea π la proporción de águilas que se obtienen al lanzar una moneda. La hipótesis nula es,

$$H_0 : \pi = 0.5.$$

Lanzamos la moneda 10 veces y obtenemos 10 águilas. La probabilidad de obtener este resultado dado H_0 es de $p = \frac{1}{2^{10}} = \frac{1}{1024} = 0.000976$. Este valor nos proporciona una evidencia sólida a favor de que la moneda no sea honesta. Ahora supongamos que este experimento de 10 lanzamientos fue uno más de, digamos otros 1000 experimentos idénticos. Adicionalmente, únicamente hubo resultados significativos -por ejemplo $p < 0.01$- en unos cuantos de ellos. En este caso, no habría evidencia para desconfiar de la moneda ¿Porqué? La obtención de la misma cara de la moneda en forma consecutiva sería un producto natural del azar y no consecuencia de una moneda cargada. Esto no sería un problema, siempre y cuando se reportaran todos los resultados: los significativos y los no significativos. Desgraciadamente ésto no siempre sucede y en ocasiones sólo son reportados los resultados significativos.

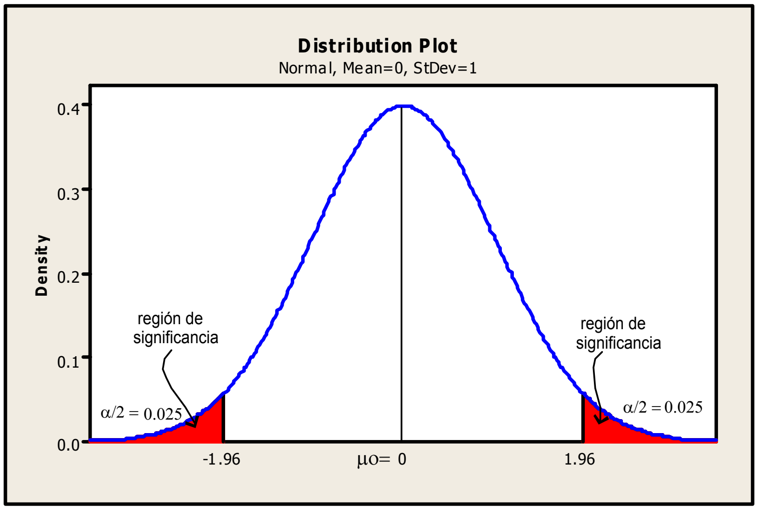
10.5. Una o dos colas

El nivel de significancia α determina una región, llamada **región de significancia** o **región de rechazo**, en la distribución de probabilidad de la estadística de interés especificada por la hipótesis nula. Si los datos nos proporcionan un valor para la estadística dentro de dicha región, entonces podemos rechazar la hipótesis nula con el nivel de significancia α. Intuitivamente, el estar dentro de la región significa que estamos suficientemente alejados del valor que especifica la hipótesis nula y, por lo tanto, podemos rechazarla. En el caso continuo, el área de dicha región es la probabilidad de estar *suficientemente* alejados de este valor y es precisamente el valor α.

Por ejemplo, si la población está distribuída normalmente -o la muestra es de más de 30 individuos- y las hipótesis nula y alterna acerca de la media poblacional son,

$$H_0 : \mu = \mu_0,$$
$$H_1 : \mu \neq \mu_0.$$

Entonces, dado un nivel de significancia de α, cualquier muestra que proporcione una media $\bar{X}$ dentro de la región determinada por α será significativa y H_0 sería rechazada pues $\bar{X}$ estaría suficientemente alejada de μ_0. De forma análoga, si $\bar{X}$ queda fuera de la región, entonces el resultado de la muestra no es significativo y no hay elementos para rechazar H_0. Como la región de significancia se encuentra de ambos lados de la distribución, decimos que se trata de una **prueba de dos colas**: $\bar{X}$ es significativo, si está suficientemente alejado de un lado o de otro de μ_0. Esto se ilustra en la siguiente figura para $\alpha = 0.05$ que representa el área total de la región de significancia, que en este caso es la suma de dos regiones:



Los valores de ± 1.96 se conocen como **valores críticos z** y se denotan por

$$\pm z_{0.025} = \pm 1.96$$

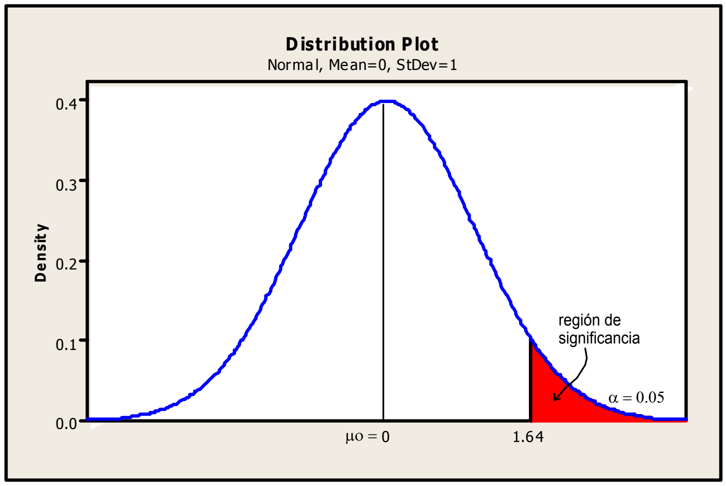
En general, para cualquier área A de una cola de la distribución, z_A denota el valor crítico z de manera que para la distribución normal estándar se cumple:

$$P(Z \geq z_A) = A = P(Z \leq -z_A).$$

En ocasiones, consideramos que el valor obtenido a partir de la evidencia es significativo sólo si está de un lado de la media especificada por la hipótesis nula. Este era el caso del ejemplo del fertilizante en la sección 10.3, en donde expresamos a las hipótesis nula y alterna como

$$H_0 : \mu \leq \mu_0,$$
$$H_1 : \mu > \mu_0.$$

La región de significancia correspondiente se encuentra a la derecha de μ_0, pues un valor será significativo sólo si es suficientemente grande. Decimos entonces, que se trata de una **prueba de una cola**. Esto se muestra en la siguiente figura para $\alpha = 0.05$ en donde la región de significancia está de un sólo lado de la distribución y el valor crítico z es de $z_{0.05} = 1.64$:



Para ilustrar lo anterior pensemos en las siguientes dos situaciones:

- En la primera, queremos determinar si una moneda es justa (honesta) y la lanzamos 10 veces, obteniendo 8 águilas y 2 soles.

- En la segunda, queremos determinar si una gitana tiene poderes de predicción. Para este propósito lanzamos una moneda justa 10 veces y le preguntamos acerca del resultado. Se observa que la gitana reporta el resultado correcto en ocho ocasiones.

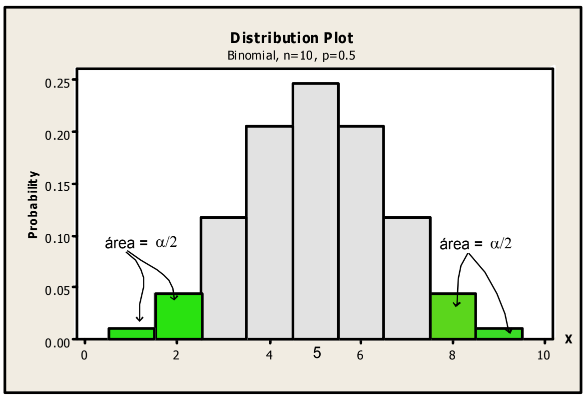
Ambas situaciones pueden modelarse mediante una distribución binomial con $n = 10$, $k = 8$ y $p = 0.5$. En el primer caso podemos tomar cada águila como un éxito. Si μ es el número promedio de éxitos, la hipótesis nula sería,

$$H_0 : \ \mu = 5$$

y la alterna

$$H_1 : \ \mu \neq 5.$$

Aquí, sospechamos de la moneda si obtenemos más éxitos (águilas) o fracasos (soles) que el esperado de 5. Si el nivel de significancia es de α, la hipótesis nula puede rechazarse si estamos de un lado o del otro de la distribución para los valores de las medias muestrales; es decir, en ambas colas de la distribución. El área de cada cola es igual a $\frac{\alpha}{2}$ -asumiendo una distribución simétrica como la normal o binomial si $p = 0.5$-. En este caso tenemos una prueba de dos colas. La región (consistente en la suma de dos regiones) de significancia con área α se ilustra en la siguiente figura:



El valor p es

$$
\begin{aligned}
p &= P(\mu \geq 8 \mid H_0) = \\
&= \frac{10!}{2!8!}(0.5)^{10} + \frac{10!}{1!9!}(0.5)^{10} + \frac{10!}{0!10!}(0.5)^{10} \\
&= 0.055.
\end{aligned}
$$

De esta forma, la hipótesis nula sólo puede rechazarse con un nivel de significancia de $2 \times 0.055 = 0.110$ o mayor.

La situación de la gitana es diferente. Sabemos que la moneda es honesta pero queremos saber si la gitana tiene algún poder de predicción. Designemos por un éxito cada vez que la gitana acierta el resultado. Una vez más sea μ el número promedio de aciertos. Las hipótesis nula y alterna son,

H_0 : la gitana no tiene poderes de predicción.

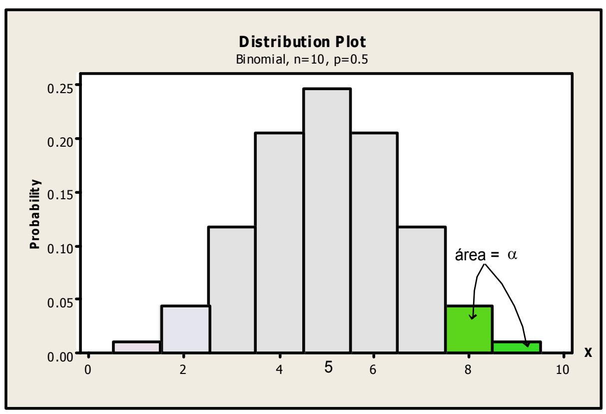
H_1 : la gitana si tiene algún poder de predicción.

o en términos estadísticos,

$$H_0: \ \mu \leq 5,$$
$$H_1: \ \mu > 5.$$

Aquí, si obtenemos más éxitos que los esperados podríamos rechazar la hipótesis nula; sin embargo, no queremos rechazarla si obtenemos 5 éxitos o menos. Por lo tanto, la prueba no es simétrica[2].

La hipótesis nula es rechazada únicamente si estamos de un lado de la media de 5 (a la derecha pues la hipótesis alterna es $H_1 : \mu > 5$). Tenemos así, una prueba de una cola y el área de dicha cola es igual al nivel de significancia α. Esto se muestra en la siguiente figura:



El valor p es idéntico al que acabamos de calcular y es

$$P(\mu \geq 8 \mid H_0) = 0.055.$$

Así, la hipótesis nula se rechaza con un nivel de significancia de 0.055 o mayor.

Un error común es el preferir de antemano una prueba de una o dos colas pues, según lo visto arriba, es la naturaleza del problema lo que debe

[2]No consideramos el hecho de que al fallar más de la mitad de la veces consistentemente, la gitana también "tendría poderes de predicción". ¡La razón es que estaría prediciendo la cara que no sale más de la mitad de las veces!

determinar esto. En el primer caso nos interesan los datos observados que quedan de ambos lados del valor dado por la hipótesis nula; en el segundo, nos interesan sólo los que están de un lado.

10.6. Potencia

El método de Fisher para la prueba de hipótesis tiene como base el nivel de significancia o la probabilidad de rechazar la hipótesis nula cuando ésta es verdadera. El experimentador puede elegir una región de significancia a manera de forzar que el valor p sea tan pequeño como se desee. Así, si la hipótesis nula fuese

$$H_0: \text{ una dieta con frambuesas mejora la memoria a corto plazo,}$$

seguramente querríamos minimizar su rechazo si fuese verdadera. En otras palabras, buscaríamos un nivel de significancia muy pequeño. Por otro lado, si aceptaramos H_0, no pasaría gran cosa, aún si fuese falsa, pues a nadie le haría daño comer muchas frambuesas.

Supongamos, sin embargo, que ahora estamos probando un medicamento nuevo para la diabetes. Se sospecha que éste podría ocasionar trombosis cerebral. La hipótesis nula es,

$$H_0: \text{ el medicamento no causa trombosis cerebral.}$$

En este caso, sería deseable el rechazar esta hipótesis si es que es falsa. Idealmente, nos gustaría maximizar la probabilidad de rechazarla -equivalentemente minimizar la probabilidad de aceptarla- cuando es falsa.

Jerzy Neyman (1894 - 1981) y Egon Pearson (1895 - 1980, hijo de Karl Pearson) idearon una variante al método de Fisher identificando los dos tipos posibles de errores en las pruebas de hipótesis, concretamente,

- **Error tipo I** cuando H_0 se rechaza a pesar de ser cierta.
- **Error tipo II**, cuando H_0 se acepta a pesar de ser falsa.

Estos son semejantes a los falsos negativos y falsos positivos de la prueba para detectar esteroides del ejemplo 4.4.1. Usualmente se denotan por α y β, respectivamente, a las probabilidades de cometer un error tipo I o tipo II. Observemos que la probabilidad de cometer un error tipo I es simplemente el nivel de significancia α de la definición 10.2.2.

El método de Neyman y Pearson procede como sigue: se fija la probabilidad α de error tipo I en algún valor convenientemente pequeño para el experimentador, por ejemplo 0.05 ó 0.01. Una vez fija esta probabilidad, se minimiza la probabilidad de error tipo II dada por,

$$\begin{aligned}\beta &= P(\text{aceptar } H_0 \mid H_0 \text{ falsa}) \\ &= 1 - P(\text{rechazar } H_0 \mid H_0 \text{ falsa}).\end{aligned}$$

Ésto equivale a maximizar,

$$P(\text{rechazar } H_0 \mid H_0 \text{ falsa}).$$

A esta última probabilidad se le conoce como el **potencia** de la prueba y está dada por $1-\beta$. Maximizar la potencia es simplemente maximizar la probabilidad de rechazo de una hipótesis falsa. Como vimos en el ejemplo del medicamento de la diabetes, esto sería lo deseable en dicho caso.

Evidentemente nos gustaría que la probabilidad de ambos tipos de error fuese lo más baja posible, sin embargo, para una muestra dada existe un trueque inevitable entre las dos probabilidades α y β. Si una disminuye la otra aumenta y viceversa. Una vez más, depende del experimentador el tomar la decisión adecuada. Si dado un tamaño de muestra y un valor de significancia α, puede elegirse entre pruebas con distinta potencia, entonces lo óptimo es escoger la que tiene la potencia máxima.

El siguiente párrafo, tomado del artículo original de Neyman y Pearson[3] expresa las consideraciones anteriores:

[3]Neyman, Jerzy y Egon S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypothesis", *Philosophical Transactions of the Royal Society* A, pp. 289-337, 1933.

¿Qué es más serio, condenar a un inocente o exonerar a un culpable? Esto dependerá de las consecuencias del error: ¿se castigará con la muerte o simplemente con una multa? ¿Cuál es el peligro a la comunidad de tener un criminal suelto? Desde el punto de vista de la teoría matemática, lo único que podemos lograr es demostrar como el riesgo de errores puede ser controlado y minimizado.

Notemos que la potencia de la prueba puede expresarse como

$$
\begin{aligned}
&P(\text{rechazar } H_0 \mid H_0 \text{ falsa}) = \\
&P(\text{rechazar } H_0 \mid H_1 \text{ verdadera}).
\end{aligned}
$$

Siempre y cuando las hipótesis estén bien especíificadas y una sea la negación de la otra ($H_1 =\sim H_0$). Es conveniente observar que esta probabilidad, al igual que el valor p, puede calcularse si H_0 y H_1 determinan por completo la distribución de probabilidades de la estadística de prueba. Por ejemplo, si las medias muestrales cumplen $\bar{X} \sim N(\mu, 1)$ y a partir de un valor observado de $\bar{X}$ se desea comparar la hipótesis,

$$H_0 : \mu = 5$$

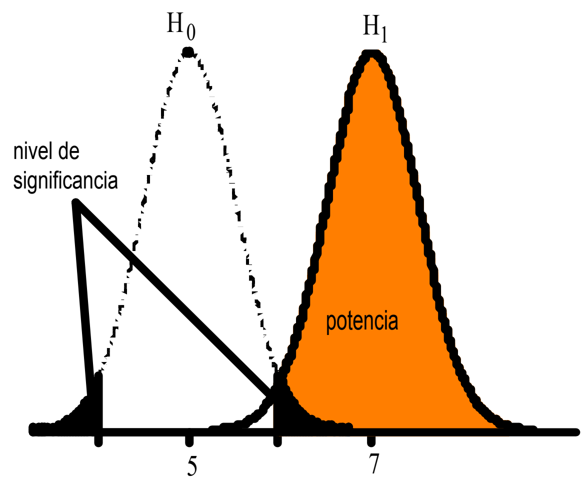
contra la hipótesis alterna,

$$H_1 : \mu \neq 5,$$

entonces,

$$
\begin{aligned}
\beta &= P(\text{aceptar } H_0 \mid H_0 \text{ falsa}) \\
&= P(\text{aceptar } H_0 \mid H_1 \text{ cierta}) \\
&= P(\text{aceptar } H_0 \mid \bar{X} \sim N(\mu, 1), \text{ con } \mu \neq 5).
\end{aligned}
$$

En este caso, β depende de todos los valores de μ que son diferentes de 5 y es imposible calcular β para todos ellos. En la práctica, se eligen algunos valores de μ compatibles con la hipótesis alterna y se calcula la potencia para cada uno de ellos[4]. Por ejemplo, para $H_1 : \mu = 7$, visualmente se tendría,



10.7. Intervalos de confianza

La significancia del resultado de una prueba de hipótesis depende de las probabilidades α y β. De esta forma, pueden tenerse efectos significativos que son de una magnitud muy pequeña o bien efectos no significativos de gran magnitud. Esta situación puede ser confusa, sobre todo cuando los resultados son utilizados en disciplinas poco cuantitativas.

Para evitar esta confusión con el significado real de los valores p, o la potencia de una prueba, es común utilizar intervalos de confianza para expresar los resultados. Esencialmente los intervalos de confianza nos proporcionan la misma información que una prueba de hipótesis; sin embargo, la información se expresa de forma visual lo cual puede ser útil en ocasiones.

Supongamos que μ es el parámetro poblacional en cuyo valor estamos

[4]En realidad, β es una función de μ y ésta se evalúa para algunos valores de μ.

interesados con las siguientes hipótesis nula y alterna:

$$H_0 : \mu = \mu_0,$$
$$H_1 : \mu \neq \mu_0.$$

Al tomar una muestra se obtiene el valor $\bar{X}$ para la media de la muestra. Si H_0 se rechaza con un nivel de significancia α, entonces el valor μ_0 no está en el **intervalo con nivel de confianza** igual a $1-\alpha$. Este intervalo está dado por

$$(\bar{X} - sz_{\frac{\alpha}{2}}, \bar{X} + sz_{\frac{\alpha}{2}}),$$

en donde s es la desviación estándar de las medias muestrales y $z_{\frac{\alpha}{2}}$ es el valor crítico z que determina un área de $\frac{\alpha}{2}$ en la cola de la distribución[5].

Equivalentemente, si un valor μ está en el intervalo con nivel de confianza $1 - \alpha$, entonces este valor no puede rechazarse con un nivel de significancia α. El intervalo de confianza para μ es un rango de valores dentro del cual se encuentra el verdadero valor del parámetro con un nivel de confianza $1 - \alpha$; es decir, la probabilidad de que el verdadero valor de μ esté en el intervalo es de $1 - \alpha$. En un escenario frecuentista, si tomamos muchas muestras de tamaño n, medimos su media $\bar{X}$ y construímos un intervalo de 95 % de confianza alrededor de cada media, entonces se espera que la media real esté en el 95 % de estos intervalos.

Veamos como se construyen estos intervalos. Supongamos que nos informan que el tiempo promedio de respuesta de Locatel (en minutos) está distribuido normalmente con media de 25 minutos y desviación estándar de 18 minutos. Se toma una muestra aleatoria de 36 llamadas a Locatel, la media de esta muestra es de $\bar{X} = 30$ minutos y la desviación estándar de las medias muestrales es de $s = \frac{18}{\sqrt{36}} = 3$ minutos. El intervalo

$$(30 - (1.96 \times 3), 30 + (1.96 \times 3)) \simeq (24.12, 35.88)$$

es un intervalo de 95 % de confianza para el tiempo promedio de respuesta.

[5]El valor crítico z se introdujo en la sección 10.5.

Cualquier valor del tiempo promedio entre 24.12 y 35.88 minutos no puede rechazarse con un nivel de significancia del 0.05. Así, este intervalo de confianza ilustra que podemos rechazar la hipótesis nula,

$$H_0 : \mu = \mu_0$$

con un nivel de significancia de 0.05, si μ_0 está fuera de este intervalo. El valor informado de $\mu_0 = 25$ minutos no puede rechazarse con un 0.05 de significancia puesto que está dentro del intervalo dado.

Los extremos del intervalo dependen de la media de la muestra, de la desviación estándar de la muestra y del grado de confianza. Sabemos que el 95 % de los datos se encuentran a una distancia de 1.96 veces la desviación estándar. Concretamente, los extremos de un intervalo de 95 % de confianza son:

$$(\bar{X} - 1.96s, \bar{X} + 1.96s).$$

De acuerdo a la proposición 9.6.1, la desviación estándar s disminuye al aumentar el tamaño de la muestra. Entonces, aumentando el número de elementos de la muestra podemos, con el mismo grado de confianza, disminuir la longitud del intervalo. Alternativamente, para el mismo intervalo, puede aumentarse el grado de confianza cambiando el valor crítico z: $z_{\frac{\alpha}{2}}$. Observemos que como el valor $\bar{X}$ es aleatorio -depende de la muestra-, los extremos del intervalo también lo serán ya que varían con cada muestra.

Ejemplos

Ej 10.7.1 Un casino reclama a su aseguradora varios millones de pesos por desfalco de sus empleados. La prueba presentada es el rechazo de la hipótesis nula: "no hay desfalco", con un nivel de significancia del 1 %. La hipótesis se presenta como,

$$H_0 : \ \mu \geq 20\,\%,$$

en donde μ es la ganancia esperada de las mesas de Black Jack. La aseguradora se niega a pagar aduciendo que la prueba no es concluyente.

En efecto, la hipótesis alterna es

$$H_1 : \mu < 20\,\%,$$

pero ésta no es equivalente a decir que si hubo desfalco por parte de los empleados, pues el resultado podría deberse a otras causas. La prueba determina que algo está pasando en las mesas de Black Jack, pero no nos dice la causa. De esta forma, la aseguradora tiene razón en no pagar.

Ej 10.7.2 (Una cola) El gobierno de la ciudad de México estima que el tiempo de recorrido del metrobús, a lo largo de cierta ruta, en días hábiles y entre 8 am y 8 pm, es aleatorio con un promedio de 60 minutos y una desviación estándar de 20 minutos. Un ciudadano incrédulo realiza 30 viajes en dicho medio de transporte y determina que el tiempo promedio de recorrido es de 70 minutos. Con esta evidencia quiere poner a prueba la hipótesis nula,

$$H_0 : \mu \leq 60.$$

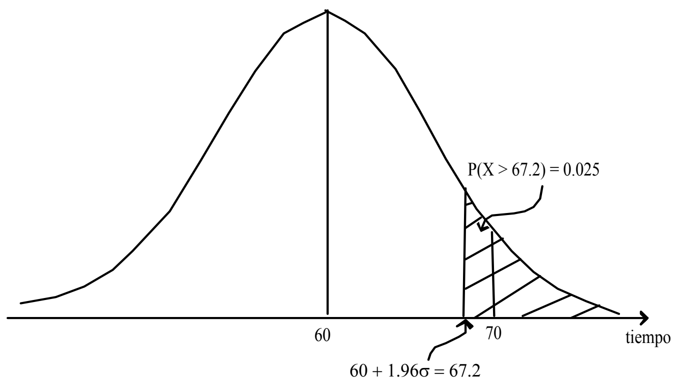
La hipótesis alterna en este caso es,

$$H_1 : \mu > 60,$$

en donde μ es el tiempo promedio de recorrido. Digamos que se especifica un nivel de significancia de $\alpha = 0.025$.

Denotemos por $\bar{X}$ a la variable aleatoria que representa los tiempos promedio de recorrido de las muestras de 30 viajes. Sabemos, de la sección 9.6.1, que $\bar{X}$ está aproximadamente distribuida normalmente con media μ y desviación estándar de$\frac{20}{\sqrt{30}} = 3.65$. La estadística que representa nuestra evidencia es, en este caso, $\bar{X} = 70$.

Como $\alpha = 0.025$, sabemos que ésta es la probabilidad de obtener valores mayores que $60 + 1.96 \times 3.65 = 67.\,2$ si H_0 es cierta. El valor p es la probabilidad $P(\ \bar{X} \geq 70)$ y como $70 > 67.3$, claramente se cumple $p < 0.025$. No es necesario calcular explícitamente el valor p pues simplemente queremos saber si es menor al nivel de significancia. Visualmente se tiene:



Si el ciudadano hubiese obtenido cualquier tiempo promedio de recorrido menor a 67.2, entonces la hipótesis nula no podría ser rechazada con el nivel de significancia especificado, es decir, no puede rechazarse la información que proporciona el gobierno acerca del tiempo promedio de recorrido del Metrobús.

Ej 10.7.3 (Dos colas) Una máquina empaca bolsas de almendras, cuyo contenido es aleatorio y distribuido normalmente con media de 250 grs. y desviación estándar de 4 grs. El departamento de control de calidad ha establecido un procedimiento para determinar si la máquina funciona correctamente. Éste consiste en tomar una muestra aleatoria de 16 bolsas y detener el proceso para ajustar la máquina si en la muestra el contenido promedio es inferior a 248 grs. o mayor a 252 grs. de almendras. ¿Cuál es la probabilidad de determinar erróneamente que la máquina necesita ajustarse?

Notemos que el procedimiento del departamento de control de calidad equivale a rechazar la hipótesis $H_0 : \mu = 250$ si $\bar{X} < 248$ o bien $\bar{X} > 252$, en donde $\bar{X}$ es el contenido promedio en una muestra de 16 bolsas. En consecuencia, lo que se solicita es el nivel de significancia de la prueba.

Las hipótesis nula y alterna son

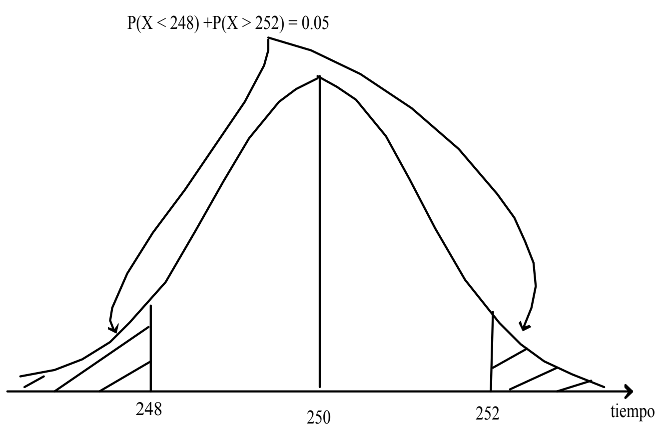
$$H_0 : \mu = 250 \text{ (La máquina no necesita ajuste)},$$
$$H_1 : \ \mu \neq 250 \text{ (La máquina necesita ajuste)}.$$

Las medias muestrales para muestras de tamaño $n = 16$ se aproximan por una distribución normal con media de 250 y desviación estándar de $\frac{4}{\sqrt{16}} = 1$.

La probabilidad de obtener una media de menos de 248 grs. o más de 252 grs. es de

$$\begin{aligned} & P(\bar{X} < 248) + P(\bar{X} > 252) \\ & \quad = P(\bar{X} < 250 - 2) + P(\bar{X} > 250 + 2) \\ & \quad \simeq 0.05. \end{aligned}$$

Aquí aproximamos el 95 % del área de una distribución normal como el área entre $\mu + 2\sigma$ y $\mu - 2\sigma$. Esto nos dice que el nivel de significancia es de aproximadamente 0.05 y ésta es la probabilidad de rechazar la hipótesis nula erróneamente. Visualmente tenemos,



Si los datos proporcionaran una cantidad media fuera del rango 248 a 252 grs., entonces el valor p será menor a 0.025 y la hipótesis nula se rechazaría. En dicho caso se determinaría que la máquina debe ajustarse.

Ej 10.7.4 Para el ejemplo 10.7.3, podemos calcular la potencia de la prueba para un valor específico de la hipótesis alterna. Digamos que

$$H_1: \ \mu = 247,$$

entonces X tiene ahora media de 247 y desviación estándar de 1. Tomando $Z = \frac{\bar{X}-247}{1}$, la distribución normal estándar asociada, se tiene que,

$$\begin{aligned}\beta &= P(248 < \bar{X} < 252)\\ &= P\left(\frac{248-247}{1} < Z < \frac{252-247}{1}\right)\\ &= P(1 < Z < 5) \simeq 0.1587.\end{aligned}$$

La potencia de la prueba es de $1-\beta = 0.841\,3$ para el valor $\mu = 247$ de la hipótesis alterna. Si se hubiese supuesto que

$$H_1: \ \mu = 248,$$

entonces de forma similar se obtiene que

$$\begin{aligned}\beta &= P(248 < \bar{X} < 252)\\ &= P(0 < Z < 4) \simeq 0.5.\end{aligned}$$

La potencia de la prueba es ahora $1-\beta = 0.5$ cuando $\mu = 248$, de manera que la prueba es más potente en el primer caso.

Ej 10.7.5 Para el ejemplo 10.7.3, supongamos que la muestra de 16 bolsas de almendras proporciona una cantidad media de $\bar{X} = 247$ grs. De acuerdo a los resultados obtenidos, la hipótesis nula es rechazada con un nivel de significancia de $\alpha = 0.05$. Con los resultados obtenidos de la muestra, el intervalo de 95 % de confianza asociado para la cantidad media de almendras μ es de aproximadamente,

$$(247-2s, 247+2s),$$

o bien,

$$(245, 249);$$

es decir, especificar cualquier valor $\mu = \mu_0$ en este intervalo es *sensato* pues la hipótesis

$$H_0 : \mu = \mu_0$$

no puede rechazarse con un nivel de significancia de $\alpha = 0.05$.

Si la muestra hubiese sido de tamaño 64 con la misma media $\bar{X} = 247$ grs., entonces, la desviación estándar de las medias muestrales sería $s = \frac{4}{\sqrt{64}} = 0.5$. En este caso, el intervalo de 95 % de confianza estaría dado aproximadamente por

$$(246, 248)$$

y el intervalo anterior sería de más del 99.9 % de confianza pues estaríamos a una distancia de cuatro desviaciones estándar de la media muestral de 247.

♠♠♠♠♠

Todos los ejemplos anteriores se refieren a pruebas de hipótesis acerca de la media en base a los resultados de una sóla muestra. Adicionalmente, siempre se conocía la desviación estándar de la población. Es posible extender estos resultados a otro tipo de pruebas de hipótesis, referentes a desviaciones estándar o proporciones o bien cuando no conocemos la desviación estándar de la población (muy usual), cuando tenemos muestras pequeñas ($n < 30$) o cuando tenemos más de una muestra. Referimos al lector a los libros de Agresti y Finlay, Aguirre et al. o Dekking et al para estos y otros casos de pruebas de hipótesis.

10.8. El regreso de Bayes

En la interpretación frecuentista de probabilidad, los valores p o la potencia de una prueba, no nos dicen que tan probable es que se cumplan las hipótesis. Ni siquiera tiene sentido el hacernos la pregunta ya que las hipótesis son ciertas o falsas, no hay medias tintas.

El análisis Bayesiano parte de una probabilidad *a priori* para cada hipótesis. La prueba de hipótesis evalúa la probabilidad de cada una de ellas, actualizada con la evidencia obtenida. El propósito es elegir la hipótesis más apropiada, tomando en cuenta sus probabilidades y las consecuencias de las distintas elecciones posibles. Como parte de este mecanismo, se sigue el proceso de aprendizaje Bayesiano expuesto en la sección 4.5.

Para ilustrar, tomemos los siguientes dos ejemplos:

- Un conocido catador de vinos es presentado con diez muestras de vino tinto y se le pide clasificarlas como Merlot o Cabernet Sauvignon. Las muestras fueron elegidas aleatoriamente entre el mismo número de botellas de cada cepa. Al finalizar, clasifica correctamente 8 de las muestras.

- Tomamos el mismo experimento de la gitana de la sección 10.5, en el cual ésta predice correctamente el resultado de 8 de diez lanzamientos de una moneda.

Claramente ambos casos pueden modelarse de igual forma con ensayos Bernoulli. Si p es la probabilidad de éxito del catador y de la gitana, las hipótesis nula y alterna pueden plantearse como,[6]

$$\begin{aligned} H_0 &: \ p \leq 0.5, \\ H_1 &: \ p > 0.5. \end{aligned}$$

Ahora bien, la probabilidad *a priori* de estas hipótesis no tiene por que ser la misma. El análisis frecuentista no puede tomar esto en cuenta; sin embargo, el análisis Bayesiano permite partir de distintas probabilidades para las hipótesis. Por ejemplo, para el experto catador podríamos pensar que

$$\begin{aligned} P(H_0) &= 0.3, \\ P(H_1) &= 0.7, \end{aligned}$$

en cambio para la gitana,

$$\begin{aligned} P(H_0) &= 0.9, \\ P(H_1) &= 0.1. \end{aligned}$$

Se procedería entonces al análisis y éste sería diferente en ambos casos, ¡como debiera ser!

[6] En la sección 10.5, las hipótesis para el problema de la gitana se plantearon en términos del número promedio de aciertos. Aquí lo hacemos en términos de la probabilidad de éxito.

Sea E la evidencia obtenida, los momios *a priori* se definen como,

$$\frac{P(H_1)}{P(H_0)},$$

y los momios *a posteriori* como,

$$\frac{P(H_1 \mid E)}{P(H_0 \mid E)}.$$

Para el catador tenemos,

$$\frac{P(H_1)}{P(H_0)} = \frac{0.7}{0.3} = 2.\bar{3},$$

de manera que H_1 es un poco más de dos veces más probable que H_0. Para la gitana se tiene

$$\frac{P(H_1)}{P(H_0)} = \frac{0.1}{0.9} = 0.11,$$

así que ahora H_0 es casi diez veces más probable que H_1.

Por otra parte, la evidencia es que ambos obtuvieron 8 éxitos en 10 ensayos. Para simplificar, supongamos que las dos hipotesis -nula y alterna- son,

$$\begin{aligned} H_0 &: \; p = 0.5, \\ H_1 &: \; p = 0.8. \end{aligned}$$

Calculamos las probabilidades de la evidencia, como las probabilidades de obtener 8 éxitos en diez intentos para las probabilidades de éxito específicadas por las hipótesis. Estas quedan dadas por,

$$\begin{aligned} P(E \mid H_0) &=_{10} C_8 \, (0.5)^8 \, (0.5)^{10-8} = 0.044, \\ P(E \mid H_1) &=_{10} C_8 \, (0.8)^8 \, (0.2)^{10-8} = 0.3. \end{aligned}$$

Después utilizamos el teorema de Bayes para obtener las probabilidades *a posteriori* de las hipótesis. Para el catador de vinos tenemos,

$$\begin{aligned} P(H_0 \mid E) &= \frac{P(E \mid H_0)P(H_0)}{P(E \mid H_0)P(H_0) + P(E \mid H_1)P(H_1)} \\ &= \frac{0.044 \times 0.3}{0.044 \times 0.3 + 0.3 \times 0.7} = 0.06, \\ P(H_1 \mid E) &= 1 - P(H_0 \mid E) = 0.94. \end{aligned}$$

Así, los momios de las hipótesis *a posteriori* son,

$$\frac{P(H_1 \mid E)}{P(H_0 \mid E)} = \frac{0.94}{0.06} = 15.67.$$

Ahora los momios en favor de H_1 son de aproximadamente 16 a 1 con lo cual, si las consecuencias de cometer errores tipo I o tipo II son igualmente costosas, podríamos descartar H_0 en favor de H_1.

Las probabilidades *a posteriori* para la gitana se evalúan de forma similar obteniéndose,

$$\begin{aligned} P(H_0 \mid E) &= \frac{P(E \mid H_0)P(H_0)}{P(E \mid H_0)P(H_0) + P(E \mid H_1)P(H_1)} \\ &= \frac{0.044 \times 0.9}{0.044 \times 0.9 + 0.3 \times 0.1} = 0.57, \\ P(H_1 \mid E) &= 1 - P(H_0 \mid E) = 0.43. \end{aligned}$$

Los momios de las hipótesis *a posteriori* son,

$$\frac{P(H_1 \mid E)}{P(H_0 \mid E)} = \frac{0.43}{0.57} = 0.75$$

y H_0 sigue siendo ligeramente más probable que H_1. Los momios no son como los del catador, no obstante le damos algo de crédito a la gitana actualizando las probabilidades *a priori*. Considerando estas probabilidades actualizadas como las nuevas *a priori*, podríamos realizar otro experimento. Suponga que ahora se observa que la gitana adivina correctamente sólo 4 de los diez lanzamientos. Realizando el mismo proceso y calculando los momios de las probabilidades *a posteriori* de H_1 vs. H_0, puede continuarse así para eventualmente arribar a que los momios fuesen muy favorables a alguna de las hipótesis.

En general, el análisis Bayesiano es mucho más completo; sin embargo, su grado de complejidad es mayor pues no puede automatizarse mediante el uso de algún paquete de software estadístico. En virtud de esta limitación práctica, no hay una receta para decidir cuando es conveniente realizarlo, pero claramente, si se posee la capacidad, el tiempo, las herramientas y la importancia del problema lo amerita, vale la pena el esfuerzo. Afortunadamente, en áreas en donde las pruebas de hipótesis son de gran trascendencia, como en la medicina, los métodos Bayesianos han comenzado a tomar suma importancia.

En general, los costos asociados a los errores tipo I y tipo II no tienen por que ser iguales. Cuando esto sucede, debe ponderarse el costo del error por la probabilidad del mismo para obtener el costo esperado. Los momios deben también ponderarse adecuadamente para reflejar los diferentes costos. Por ejemplo, en la sección 10.6 se presentó la siguiente hipótesis:

$$H_0: \text{ el medicamento no causa trombosis cerebral.}$$

Si ésta es aceptada cuando es falsa -error tipo II- el costo puede ser muy alto ya que se pone en peligro la vida de los pacientes. La hipótesis alterna es

$$H_1: \text{el medicamento causa trombosis cerebral.}$$

Si ésta es aceptada cuando es falsa -error tipo I-, el costo sería menor ya que los pacientes serían tratados con un medicamento diferente. Como vemos, estos costos miden la gravedad de las consecuencias que se derivan de aceptar una hipótesis falsa.

Ejemplos

Ej 10.8.1 El Departamento de Economía de una prestigiosa Universidad contrata a un doctor en economía llamado Pancho Keynes. Es difícil saber *a priori* si su desempeño será bueno o no. Por un lado, se sabe por experiencia en otros casos, que si Pancho es un buen economista, entonces con probabilidad de 0.8 tendrá un buen desempeño y con probabilidad 0.2 será un desastre. Por otro lado, si Pancho es un mal economista, se desempeñará bien con probabilidad 0.4 y mal con probabilidad 0.6.

La Universidad quiere cerciorarse de que contrató a un buen académico y ponen a prueba a Pancho durante un año antes de ofrecerle un contrato definitivo. El Departamento de Economía está convencido que el encontrar buenos economistas es un *volado* es decir, aproximadamente sólo la mitad son buenos.

Durante este año Pancho publica varios artículos interesantes y las evaluaciones de sus cursos son excelentes. Como el Departamento es sumamente serio realiza un análisis cuidadoso para poner a prueba la hipótesis nula:

$$H_0: \text{ Pancho es un buen economista,}$$

versus la alterna

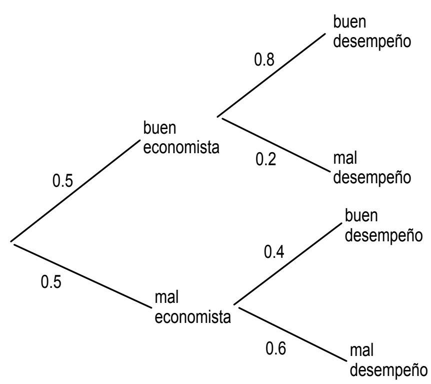
$$H_1: \text{ Pancho es un mal economista.}$$

Las probabilidades *a priori* del Departamento de Economía son $P(H_0) = 0.5 = P(H_1)$, de manera que

$$\frac{P(H_0)}{P(H_1)} = 1$$

y ambas hipótesis son equiprobables.

Veamos el árbol que representa la situación de Pancho



Al final del reporte del primer año, el comité evaluador procede igual que en la sección 4.5 para obtener,

$$P(\text{buen economista} \mid \text{buen desempeño}) =$$
$$\frac{P(\text{buen desempeño} \mid \text{buen economista})P(\text{buen economista})}{P(\text{buen desempeño})} =$$
$$\frac{0.8 \times 0.5}{(0.8 \times 0.5) + (0.4 \times 0.5)} = 0.\bar{6}$$

y

$$P(\text{mal economista} \mid \text{buen desempeño}) =$$
$$1 - P(\text{buen economista} \mid \text{buen desempeño}) = 0.\bar{3}.$$

Los momios *a posteriori* para las hipótesis son

$$\frac{P(H_0 \mid \text{buen desempeño})}{P(H_1 \mid \text{buen desempeño})} = \frac{0.\bar{6}}{0.\bar{3}} = 2.$$

Así, después del primer año, el Departamento de Economía considera que los momios son 2 a 1 a favor de que Pancho sea un buen economista y, si las consecuencias de cualquier decisión equivocada son igualmente graves, deciden extenderle el contrato.

Ej 10.8.2 Un famoso deportista es acusado de matar a su esposa. Dado que es un individuo carismático y su conducta, antes de este desagradable incidente, ha sido ejemplar, el jurado tiene las siguientes probabilidades *a priori* para su culpabilidad (C) e inocencia (I):

$$P(C) = 0.1,$$
$$P(I) = 0.9.$$

De aquí, la razón entre ellas es de,

$$\frac{P(C)}{P(I)} = \frac{0.1}{0.9} = \frac{1}{9}$$

y los momios son de 1 a 9 a favor de C, o bien de 9 a 1 a favor de su inocencia.

Los investigadores forenses encuentran, en la escena del crimen, un guante ensangrentado y realizan una prueba de sangre para determinar si pertenece al deportista. El fenotipo sanguíneo detectado se encuentra en el 1 % de la población y coincide con el del atleta. Llamémosle E a esta evidencia.

La probabilidad de tener la evidencia si el deportista es culpable es de 1 pues la sangre en el guante sería, sin duda, suya. La probabilidad de la evidencia si el individuo es inocente es de 0.01. Esto es porque sería simplemente producto de la casualidad el haber encontrado dicho fenotipo en la sangre del guante. Se tiene entonces,

$$P(E \mid C) = 1,$$
$$P(E \mid I) = 0.01.$$

Utilizando el teorema de Bayes, calculamos la razón de las probabilidades actualizadas de culpabilidad e inocencia como,

$$\begin{aligned}\frac{P(C \mid E)}{P(I \mid E)} &= \frac{\frac{P(E|C)P(C)}{P(E|C)P(C)+P(E|I)P(I)}}{\frac{P(E|I)P(I)}{P(E|C)P(C)+P(E|I)P(I)}} \\ &= \frac{P(E \mid C)P(C)}{P(E \mid I)P(I)} = \frac{1}{0.01}\frac{1}{9} = \frac{100}{9}.\end{aligned}$$

Es decir, la probabilidad de culpabilidad se multiplica por 100 y ahora los momios a favor de la culpabilidad son de 100 a 9 o aproximadamente de 11 a 1.

Ej 10.8.3 Ciertamente, el atleta del ejemplo anterior está en serios problemas. Sus probabilidades actualizadas de culpabilidad e inocencia son,

$$\begin{aligned}P(C \mid E) &= \frac{P(E \mid C)P(C)}{P(E \mid C)P(C) + P(E \mid I)P(I)} \\ &= \frac{0.1}{0.1 + 0{,}01 \times 0.9} = 0.917\,43, \\ P(I \mid E) &= 1 - P(C \mid E) = 0.082\,57.\end{aligned}$$

La hipótesis: "el atleta es inocente", está en peligro de rechazarse. Sin embargo, la defensa sorprende al fiscal con la siguiente evidencia que llamamos E^* : ¡el guante es demasiado pequeño y en realidad pertence al verdadero criminal! Esta nueva evidencia es (condicionalmente) independiente de la anterior ya que se refiere al tamaño del guante y no a la mancha de sangre.

La probabilidad de esta evidencia, si el atleta es inocente, es 1, y la probabilidad de la nueva evidencia si es culpable es de 0.01, pues el guante sería de otra persona inocente que por casualidad tiene el mismo fenotipo sanguineo. Tenemos así,

$$\begin{aligned}P(E^* \mid C) &= 0.01, \\ P(E^* \mid I) &= 1.\end{aligned}$$

Procediendo igual que antes llegamos a que las probabilidades actualizadas de culpabilidad e inocencia son:

$$
\begin{aligned}
P(C \mid E^*) &= \frac{0.01 \times 0.91743}{0.082\,57 + 0.01 \times 0.91743} = 0.1, \\
P(I \mid E^*) &= 0.9,
\end{aligned}
$$

¡por lo que los momios ahora son una vez más los originales de 9 a 1 en favor de la inocencia del atleta!

♠♠♠♠♠

10.9. Bayes va a la corte

Es común el hacer una analogía entre las pruebas de hipótesis estadísticas y los juicios criminales como sigue:

En el lenguaje frecuentista, lo usual, y lo que se considera políticamente correcto, es partir de la hipótesis nula de que el individuo no es culpable. Si la evidencia es significativa, entonces se rechaza esta hipótesis a favor de la culpabilidad: *el individuo es culpable si hay evidencia más allá de la duda razonable.* Un error tipo I se comete cuando se condena a una persona inocente, un error tipo II surge si se exonera a un culpable. El término *más allá de la duda razonable* representa "para un determinado nivel de significancia". Si la probabilidad de condenar a un culpable es alta, entonces la potencia de la prueba es alta. El veredicto de *no culpable* se obtiene cuando la evidencia no es suficientemente significativa para rechazar la hipótesis nula. Notemos que esto no implica necesariamente que el individuo sea inocente a menos que la probabilidad de cometer un error tipo II fuese cero.

Es interesante notar, que dentro de este contexto, las reformas recientes (marzo 2008) al sistema judicial mexicano cambian la hipótesis nula de: "el individuo es culpable", a favor de: "el individuo es no culpable". El trabajo de probar la significancia de la evidencia pasa de la defensa a la fiscalia. Anteriormente se era culpable a menos que la defensa recaudara evidencia significativa de lo contrario. Ahora, el status quo es de no culpabilidad y es trabajo de la fiscalia el probar que la evidencia es significativa para rechazar esta hipótesis.

La analogía entre los juicios criminales y las pruebas de hipótesis parece muy esclarecedora, ¡pero incorrecta dentro de la interpretación frecuentista de probabilidad! Asimismo, los errores tipo I y II también se refieren a probabilidades numéricas que serían imposibles de especificar en una corte pues ni siquiera tenemos la distribución muestral de probabilidad. ¡Qué cómodo sería poder asignar *a priori* que la probabilidad de condenar a un inocente fuese menor al 1 %!

No todo está perdido pues, afortunadamente para los Bayesianos las probabilidades subjetivas no son un problema. Se tiene así, que puede seguirse un análisis Bayesiano de prueba de hipótesis dentro de las cortes. Los ejemplos 10.8.2 y 10.8.3 muestran una situación, que aunque estilizada, ilustra el tipo de estrategia que podría seguirse por la fiscalia y la defensa. Asimismo, refleja lo complejo que puede ser el proceso de decisión de un jurado o de un juez, pues en algún momento habrán de tomar una decisión y sentirse satisfechos con ésta.

Bibliografía Complementaria

1. Agresti Alan, Finlay Barbara, *Statistical Methods for the Social Sciences*,4^a ed, Pearson, Allyn & Bacon, New York, NY, 2008.

2. Aguirre Víctor et al., *Fundamentos de Probabilidad y Estadística*, 2^a edición. Jit Press, México, D.F. 2006.

3. Dekking F. Michel et al., *A Modern Introduction to Probability and Statistics : Understanding Why and How*, Springer-Verlang, , London, UK 2005.

4. Finkelstein Michael O, Levin Bruce, *Statistics for Lawyers*, Springer Verlag, New York, NY, 2007.

5. Gill Jeff, *Bayesian Methods: a Social and Behavioral Sciences Approach*, Chapman and Hall, Boca Raton, FL, 2002.

6. Hacking Ian, *Probability and Inductive Logic*, Cambridge University Press, New York, NY, 2001.

7. Kadane J.B. and Schum D. A., *A Probabilistic Analysis Of The Sacco and Vanzetti Evidence*, Wiley-Interscience, 1996.

8. Kaye, David H., Freedman, David A., "A Reference Guide on Statistics", *Reference Manual on Scientific Evidence*, 2nd edition, Federal Justice Center. Versión electrónica en,

   http://www.fjc.gov/public/pdf.nsf/lookup/sciman01.pdf/$file/sciman01.pdf

9. Lee Peter M, *Bayesian Statistics: An Introduction*, Hodder Arnold, 3^a ed, London, UK, 2004.

10. Moreno Elías, Girón F. Javier, "On the frequentist and Bayesian approaches to hypothesis testing", *SORT* 30 (1) January-June 2006, pp. 3-28, versión electrónica en,

    http://dmle.cindoc.csic.es/pdf/SORT_2006_30_01_01.pdf

11. Neyman, Jerzy, Egon S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypothesis", *Philosophical Transactions of the Royal Society* A, pp. 289-337, 1933.

12. Zeisel Hans, Kaye David, *Prove it with Figures: Empirical Methods in Law and Litigation*, Springer Verlag, New York, NY, 1997.

Ejercicios

▷ **10.1** Describir un experimento para las siguientes situaciones. Adicionalmente, determinar las hipótesis nula y alterna más sensatas. Justificar las respuestas.

1. Se quiere determinar si una ruleta está truqueada.
2. Se quiere determinar si hay una diferencia entre los consumos promedios de tequila de hombres y mujeres.
3. Quiere determinarse si cierta hormona afecta la producción de leche en las vacas.
4. Quiere determinarse si un nuevo programa educativo afecta el rendimiento escolar de los estudiantes de primaria.
5. Se quiere determinar si existe una diferencia en la temperatura ambiente promedio entre zonas rurales y urbanas adjacentes.
6. Quiere determinarse si el jugo de arándano reduce la incidencia de infecciones de vías urinarias.
7. Se quiere determinar si la población tiene una opinión favorable acerca de cierta política.
8. Quiere determinarse si el 96 % de los focos ahorradores de luz de cierta marca, duran más de 10000 horas, como lo anuncia el productor.
9. Quiere determinarse si una astróloga puede determinar el signo del zodiaco de una persona, simplemente con verla a los ojos.
10. Se quiere determinar si existe discriminación en contra de la contratación de trabajadores mayores de 40 años en cierta dependencia del sector público.

▷ **10.2** El ingreso mensual de los profesionistas con 5 años de experiencia en 2002, tenía un promedio de \$ 14200 con una desviación estándar de \$2600. En 2007, se obtuvo que el salario mensual promedio de una muestra aleatoria de 75 profesionistas fue de \$15300 (ajustado a la inflación).

1. Determinar, con un nivel de significancia de 0.05, si el ingreso promedio de un profesionista ha cambiado entre 2002 y 2007.

2. Con los mismos supuestos determinar, con un nivel de significancia de 0.025 , si el ingreso promedio de un profesionista aumentó entre 2002 y 2007.

3. Con los mismos supuestos determinar, con un nivel de significancia de 0.025 , si el ingreso promedio de un profesionista disminuyó entre 2002 y 2007.

▷ **10.3** El tiempo promedio de procesamiento para abrir una cuenta de cheques en las sucursales de cierto banco es aleatorio con una media de 12.3 minutos y una desviación estándar de 3.5 minutos. Se pone a prueba un nuevo procedimiento con 100 clientes obteniéndose un tiempo promedio de 10.9 minutos. Determinar si el nuevo procedimiento reduce significativamente el tiempo de procesamiento, con un nivel de significancia de 0.05.

▷ **10.4** El alcoholímetro está calibrado de tal forma que de 1000 personas que poseen más del límite legal de alcohol en su organismo, 950 resultan positivos y 50 negativos. Asimismo, de 1000 personas que no han ingerido alcohol, 995 resultan negativas y 5 positivas. La calibración puede modificarse para variar estos niveles. Interpretar la calibración del alcoholímetro como errores tipo I y II.

▷ **10.5** Se tienen dos urnas idénticas en apariencia, excepto que la primera contiene 60 canicas blancas y 40 negras y la segunda 40 blancas y 60 negras. Están etiquetadas como $\pi = 0.6$ y $\pi = 0.4$ en referencia a la proporción de canicas blancas. El problema es que la persona que colocó las etiquetas no prestó atención y no se sabe si están colocadas correctamente. Supongamos que nos presentan la urna que dice $\pi = 0.4$ y se nos permite tomar una muestra de cinco canicas con reemplazo para ayudarnos a decidir si la etiqueta es correcta. Sea X el número de canicas blancas en esta muestra de cinco. Las hipótesis nula y alterna son

$$H_0 : \pi = 0.4,$$
$$H_1 : \pi = 0.6.$$

(Sugerencia: utilizar una distribución binomial para calcular todas las probabilidades solicitadas)

1. Supongamos que decidimos que la hipótesis nula será rechazada si obtenemos 5 canicas blancas en nuestra muestra de cinco. Calcular la probabilidad de error tipo I (α), la probabilidad de error tipo II (β) y la potencia de la prueba para esta situación.

2. Supongamos que decidimos que la hipótesis nula será rechazada si obtenemos 4 canicas blancas en nuestra muestra de cinco. Calcular la probabilidad de error tipo I, la probabilidad de error tipo II y la potencia de la prueba para esta situación.

3. ¿Qué pasa con α y β en las dos situaciones planteadas?

4. ¿Qué se tendría que hacer para que $\alpha = 0$? ¿Y para tener $\beta = 0$?

▷ **10.6** Un experimento de psicología se repite 200 veces. El investigador obtiene un nivel de significancia de 0.01 en dos ocasiones y niveles mucho mayores en las 198 veces restantes. Se envía para su publicación un artículo académico que reporta únicamente las dos pruebas significantes. ¿Se debe publicar este artículo o no? Justificar cuidadosamente.

▷ **10.7** Una encuesta de una muestra aleatoria de 1000 habitantes de la ciudad de México revela que el 22 % de ellos cree que ha tenido apariciones de la Virgen de Guadalupe. Estos resultados aproximan el porcentaje correspondiente de toda la población de la ciudad con una precisión del 3 % .

1. Proporcionar un intervalo de 95 % de confianza para el porcentaje de la población de la ciudad de México que dice haber visto a la virgen.

2. Una encuesta con las mismas características y precisión se aplica a los habitantes de la ciudad de Monterrey, en donde el 18 % de los habitantes

reporta haber visto a la virgen. Proporcionar el intervalo de 95 % de confianza para este caso.

3. Una encuesta con las mismas características y precisión se aplica a los habitantes de la ciudad de Puebla, en donde el 29 % de los habitantes reporta haber visto a la virgen. Proporcionar el intervalo de 95 % de confianza para este caso.

▷ **10.8** Considerar las siguientes hipótesis relativas a la exploración en aguas profundas del Golfo de México para incrementar las reservas probables de petróleo en México:

H_0 : No es necesario explorar en aguas profundas para encontrar petróleo,

H_1 : Es necesario explorar en aguas profundas para encontrar petróleo.

Proporcionar probabilidades a priori sensatas para la hipótesis nula (y consecuentemente para la alterna), suponiendo que la hipótesis nula es una aseveración que proviene de:

1. Un taxista.
2. Los miembros de la Comisión de energía del H. Congreso de la Unión.
3. Un político de oposición iluminado conocido como el "Rayo de Esperanza"
4. Un grupo de ingenieros especialistas en mantos petrolíferos del Golfo de México.
5. Un grupo de académicos de diversas disciplinas, principalmente ciencias sociales.
6. El director de Petróleos Mexicanos.
7. La Secretaría de Energía.
8. El presidente de la República.
9. Un grupo de compañías privadas dedicadas a la exploración en aguas profundas.
10. Un grupo de doctores en matemáticas.

▷ **10.9** Una corte está decidiendo un caso de paternidad atípico: 2 hombres se disputan la paternidad de una niña de 10 meses. La madre de la niña falleció recientemente en circunstancias trágicas y dejó una herencia de varios millones de dólares a su única hija. Los 2 hombres disputan ahora la paternidad con el fin de tener la custodia de la menor y poder administrar su fortuna.

Llamemos a los dos caballeros A y B. En el acta de nacimiento de la niña, el caballero A está registrado como el padre; el caballero B admite el haber tenido con ella una simple "aventura de una noche". Para complicar las cosas, la fallecida había tenido encuentros con un buen número de caballeros diferentes al A y B. Cualquiera de ellos podría también ser el padre de la bebé, a este posible caballero desconocido le llamamos X. Si X es el padre, la custodia sería otorgada al caballero A. Con esta información, las probabilidades a priori de ser el verdadero padre se asignan como

$$P(A \text{ es el padre}) = 0.6$$
$$P(B \text{ es el padre}) = 0.1,$$
$$P(X \text{ es el padre}) = 0.3.$$

1. Calcular los momios a priori para cada pareja de las hipótesis siguientes:

$$H_0 : A \text{ es el padre},$$
$$H_1 : B \text{ es el padre},$$
$$H_2 : X \text{ es el padre}$$

2. Una prueba de ADN seguramente aclararía la situación, sin embargo, la madre dejó especificado en el testamento que mientras su hija fuese menor de edad no se le podría practicar dicha prueba. La corte ordena que se realize la prueba alternativa ABO de sangre para determinar la posible paternidad. Esta prueba no tiene la precisión de la prueba de ADN, pero es la segunda mejor opción.

   La prueba determina que el tipo de sangre de la bebé es compatible con la del caballero B y no con la del A, llamémosle a ésta la evidencia E. Se estima que 1 de cada 10 individuos puede presentar la misma compatibilidad sanguínea. Las probabilidades de esta evidencia dadas las hipótesis son:

$$P(E \mid A \text{ es el padre}) = 0,$$
$$P(E \mid B \text{ es el padre}) = 1,$$
$$P(E \mid X \text{ es el padre}) = 0.1.$$

Actualizar las probabilidades a priori con esta evidencia, es decir, calcular

$$P(A \text{ es el padre} \mid E),$$
$$P(B \text{ es el padre} \mid E),$$
$$P(X \text{ es el padre} \mid E).$$

3. Obtener nuevamente los momios para las hipótesis con las probabilidades actualizadas. ¿Qué se concluye?

4. Suponer que la custodia se le otorgó al caballero B, el abogado del caballero A apela la decisión con el siguiente argumento: el caballero X representa a un conjunto más numeroso de individuos que el que inicialmente se suponía. Así, las probabilidades a priori deben rectificarse. Después de deliberar, el juez asigna las siguientes a priori:

$$P(A \text{ es el padre}) = 0.2$$
$$P(B \text{ es el padre}) = 0.05,$$
$$P(X \text{ es el padre}) = 0.75.$$

Recalcular los momios para las hipótesis con estas probabilidades y con las probabilidades actualizadas con la evidencia E ¿Qué se concluye?

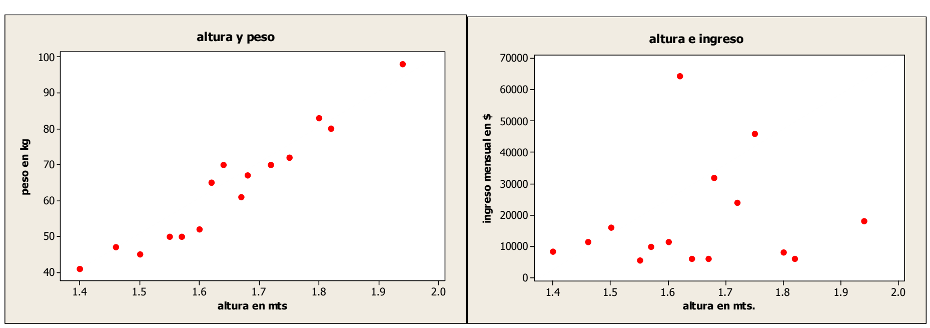
Capítulo 11

Regresión

11.1. Introducción: correlación

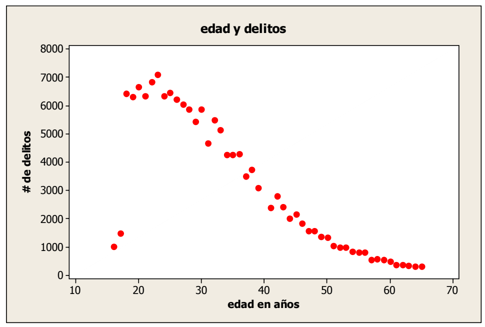
Con frecuencia es necesario determinar si dos variables (aleatorias) están *relacionadas* de alguna manera. Por ejemplo, ¿tendrán los años de educación efecto sobre el salario que percibe un individuo? La relación entre dos variables cuantitativas puede visualizarse en un **diagrama de dispersión** en el plano, representando los valores de las variables en los ejes horizontal y vertical.

La siguiente figura muestra diagramas de dispersión para datos de altura y peso y también para la altura y el ingreso mensual de una muestra de individuos adultos.



Visualmente observamos que la altura y el peso tienden a aumentar o disminuir juntos. Cuando dos variables se relacionan de esta forma decimos que están **positivamente correlacionadas**. La altura y el ingreso no parecen tener relación alguna entre ellas.

El siguiente diagrama de dispersión ilustra la relación entre la edad (16 a 65 años) y el número de delitos cometidos por dicho grupo de edad para hombres en México.

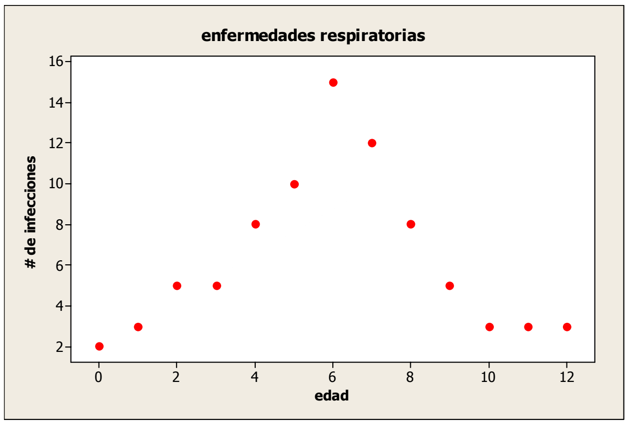


Los delitos parecen aumentar con la edad hasta aproximadamente los 25 años, para posteriormente disminuir conforme la edad aumenta. Si consideramos únicamente a la población mayor de 25 años, entonces decimos que el grupo de edad y el número de delitos están **negativamente correlacionados**.

La correlación puede darse entre variables sin ninguna implicación de causalidad entre ellas, por ejemplo: si tomamos una muestra de individuos y medimos los diámetros del antebrazo y del muslo, seguramente encontraremos que hay una correlación positiva alta. Evidentemente no hay ninguna relación de causalidad entre estas variables y más bien ambas dependen del peso y la altura del individuo. A este tipo de correlación entre variables se le conoce como **correlación espuria**.

El **análisis de regresión** es la herramienta estadística que estudia las relaciones entre variables. En general nos interesa, no sólo la relación entre las variables, sino algún tipo de causalidad entre ellas. En los ejemplos anteriores querríamos determinar si la altura de un individuo afecta su peso o si el grupo de edad al que pertenece lo hace más propenso a cometer un crimen. En estas dos situaciones es claro el sentido de la causalidad, sin embargo esto no siempre es evidente.

La asociación más simple entre variables es cuando éstas se relacionan en forma lineal, sin embargo, no siempre es posible establecer este tipo de relaciónl entre ellas. Por ejemplo, si tomamos el número promedio de infecciones respiratorias al año que sufre un niño de los 0 a los 12 años, la relación podría ser como en la siguiente figura:

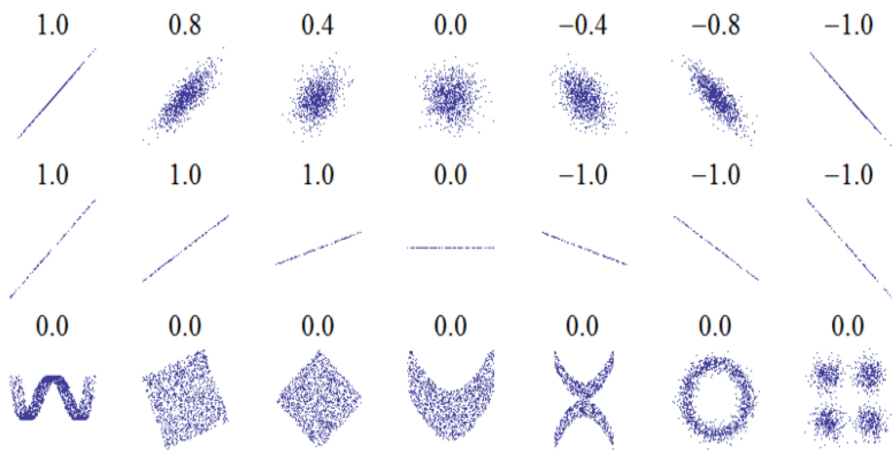


Es claro que una curva con un valor máximo alrededor de los 6 años sería la relación apropiada en este caso y no una recta. En general, el diagrama de dispersión nos dará una idea visual del tipo de relación. Aquí nos ocuparemos únicamente del caso lineal.

Para medir la magnitud de la asociación lineal entre dos variables, se utiliza comúnmente el **coeficiente de correlación** introducido por Karl Pearson. Éste es un número entre el -1 y el 1 denotado por la letra R. Si $R = -1$, se tiene una relación negativa perfecta y los puntos en el diagrama de dispersión se encuentran sobre una recta con pendiente negativa.

Si $R = 1$, la relación lineal es también perfecta pero positiva: los puntos en el diagrama de dispersión están sobre una recta con pendiente positiva. Si $R = 0$, entonces no hay relación lineal alguna y los puntos forman más bien una nube difusa o algún otro patrón evidentemente no lineal. Lo usual es tener casos intermedios, en donde existe algún grado moderado de corrrelación lineal entre las variables. En general, en las ciencias sociales es raro tener coeficientes de correlación mayores que 0.7 (o menores que -0.7).

La siguiente figura[1] ilustra las consideraciones anteriores.



Intuitivamente, el coeficiente de correlación nos da una idea de que *tan ruidosa* es la relación lineal entre las variables. Observemos que en el caso de la recta horizontal se tiene que $R = 0$ puesto que la variable en el eje vertical toma un valor constante, independiente de la otra variable.

Tomemos el ejemplo visto anteriormente que describe la asociación entre la altura y el peso de una muestra de individuos. A continuación se presentan los datos utilizados y se agrega un dato al final que corresponde a un individuo atípico ya que mide 1.70 mts. y pesa 200 kgs.

Altura	**Peso**
1.94	98
1.82	80
1.75	72
1.80	83
1.62	65
1.64	70
1.68	67

Altura	**Peso**
1.46	47
1.50	45
1.55	50
1.72	70
1.67	61
1.57	50
1.60	52
1.70	200

[1]This image has been (or is hereby) released into the public domain by its author, Imagecreator at the wikipedia project. This applies worldwide.

El resultado del coeficiente de correlación, sin el dato final, obtenido con *Minitab* ® es,

$$\text{Pearson correlation of } \textit{altura} \text{ and } \textit{peso} = 0.966$$
$$\text{P-Value} = 0.000$$

Esto significa que $R = 0.966$, que es un valor altísimo. El valor p se refiere a la pareja de hipótesis

$$H_0 : R = 0,$$
$$H_1 : R \neq 0.$$

Aquí p es tan pequeño que aparece como 0.000 ya que Minitab no muestra valores menores a 0.001. Así, $p < 0.001$ y hay fuerte evidencia en favor de que exista un alto grado de correlación lineal entre las variables.

Si agregamos el dato de la persona obesa se obtiene,

$$\text{Pearson correlation of } \textit{altura} \text{ and } \textit{peso} = 0.483$$
$$\text{P-Value} = 0.058$$

Como se observa, la correlación es mucho menor y no podemos rechazar H_0 con un nivel de significancia ni siquiera de 0.05, ya que $p = 0.058$. La conclusión es que el resultado es sensible a los puntos extremos. Estos puntos rompen con la asociación entre las variables y debe intentarse explicarlos. Por ejemplo, ¿es un error de medición?, ¿el individuo es atípicamente obeso? Para evitar este tipo de sesgos, idealmente deben eliminarse los puntos extremos y simplemente explicar el por qué ocurrieron.

11.2. Regresión lineal simple

Las relaciones lineales entre variables son sumamente populares por su simplicidad. Con frecuencia, aunque la relación entre las variables no sea inicialmente lineal, puede volverse lineal mediante alguna transformación. A continuación se describen los elementos básicos de la búsqueda y el análisis de las relaciones lineales.

Tomemos los siguientes datos que proporcionan los años de escolaridad, o instrucción formal y el ingreso mensual promedio, en pesos, para una muestra de adultos de la ciudad de México. Al observar los datos, se intuye que los años de instrucción tienen algún efecto en el ingreso.

Años de instrucción	**Ingreso**
0	1632
2	2124
6	2937
9	3120
12	5908
13	6881
16	14310
18 o más	14500

En este ejemplo, los años de instrucción o equivalentemente el nivel de educación (causa), representan la variable **independiente o predictor** y el ingreso (efecto) la varibale **dependiente o respuesta**. Llamémosle a estas variables E e I, respectivamente. El análisis de **regresión (lineal) simple,** no sólo se ocupa de la correlación entre las variables, sino determina la línea recta que *mejor aproxima* a los puntos que tenemos. Aquí, la correlación entre ambas variables está dada por,

$$\text{Pearson correlation of E and I} = 0.894$$
$$\text{P-Value} = 0.003$$

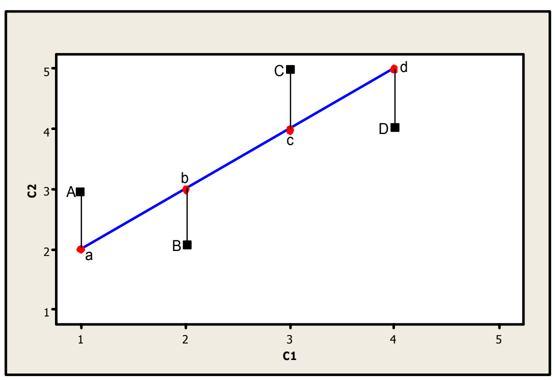
Dado que queremos encontrar una asociación lineal entre estas variables, ésta quedaría expresada como,

$$I = a + bE + \varepsilon, \tag{11.1}$$

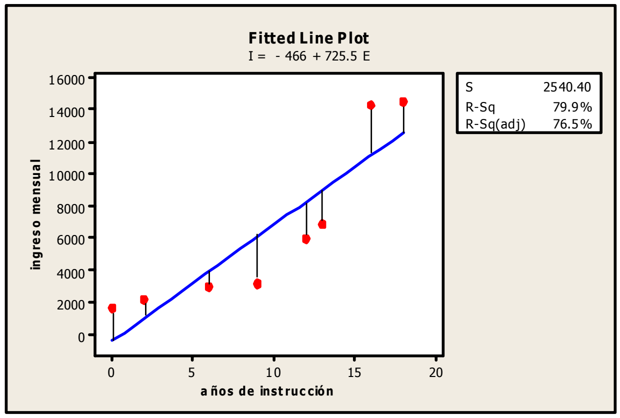
en donde el **término constante** a representa el ingreso de un individuo sin instrucción alguna o la intersección vertical de la recta. El **coeficiente** b representa el aumento en el ingreso de un año adicional de instrucción o bien, la pendiente (inclinación) de la recta. El término de *ruido* ε refleja todos los otros factores que pueden tener influencia en el ingreso y que no tomamos en cuenta, ya sea porque no son observables, o no son reportados.

La tarea de la regresión simple es la estimación de a y b, asumiendo ciertas características para el término de ruido ε. En otras palabras, proporciona la recta que da la mejor estimación de los datos. El ruido es el conjunto de **errores** o **residuos** que son las distancias de los datos reales a la recta estimada. Idealmente, la distribución de los errores es normal con media de cero, cada error es independiente de los demás y todos poseen la misma desviación estándar. En la práctica a veces no se satisfacen exactamente todas estas condiciones.

En realidad, tomamos los cuadrados de las distancias para evitar que una distancia positiva y una negativa se cancelen. Esto se ilustra en la siguiente figura: la suma de las distancias de los datos {a,b,c,d} a la recta es de cero pues la recta pasa sobre todos ellos. La suma de las distancias de los datos {A,B,C,D} a la recta sería también cero, a pesar de que la recta no pasa por ninguno de ellos. Por esta razón, se toman los cuadrados de las distancias de los datos a la recta para representar al error.



El método más común de regresión elige la recta que minimiza las sumas de los cuadrados de estas distancias o errores y se conoce como **método de mínimos cuadrados**. Éste fue introducido por el matemático francés Adrien Marie Legendre (1752 - 1833) en 1805 e independientemente por Gauss en 1809. La siguiente figura muestra el resultado de la regresión simple por mínimos cuadrados proporcionado por Minitab para los datos de escolaridad e ingreso.



Esta recta es la que minimiza la suma de los cuadrados de las distancias verticales mostradas y su ecuación es,

$$I = -466 + 725.5E.$$

La regresión también arroja la siguiente tabla:

Predictor	**Coef**	**SE Coef**	**T**	**P**
Constant	-466	1674	-0.28	0.790
E	725.5	148.7	4.88	0.003

$$S = 2540.40 \; R - Sq = 79.9\,\% \quad R - Sq(adj) = 76.5\,\%.$$

Supongamos que quiere determinarse que tan significativos son los coeficientes a y b, es decir, si podemos rechazar la hipótesis nula de que son iguales a cero. En este tipo de aproximación lineal, los errores y los coeficientes son variables aleatorias que siguen una distribución normal. La primera columna de números (Coef) se refiere a los valores estimados para a y b. La segunda columna (SE Coef) es la desviación estándar de los valores de la primera columna. La tercera columna requiere de explicación adicional.

Sea $coef$ el valor de alguno de los coeficientes de la primera columna. La variable estándar asociada para la prueba de hipótesis es

$$\frac{coef - 0}{\sigma}$$

y ésta parecería ser una normal estándar, ¿o no? La respuesta es negativa y la razón es que hasta ahora, en principio conocíamos (o al menos pretendíamos conocer) tanto la media como la desviación estándar de la población de donde provenían los datos. Aquí la situación es distinta, veamos:

En el análisis de regresión no conocemos la desviación estándar σ y ésta se infiere a partir de los datos (segunda columna de la tabla). La variable estandarizada ahora sigue una distribución parecida a la normal -salvo por tener *colas más gordas*-, llamada **distribución t** o **Student**. Esta distribución fue descubierta por William Gosset (1876 - 1937) un empleado de la cervecería irlandesa *Guinness.* Gosset requería una distribución para tamaños de muestra pequeños. La cervecería no permitía la publicación de resultados de investigación de manera que Gosset utilizó el seudónimo de *Student* para publicar su descubrimiento.

Regresemos a la tercera columna de la tabla. Ésta proporciona el **valor t** o **estadística t** que es el cociente del coeficiente estimado (columna 1) entre su desviación estándar (columna 2). Éste sigue una distribución t como se explicó arriba. La última columna es el valor p o bien la probabilidad de obtener el valor t (evidencia) dado que el coeficiente es cero (hipótesis nula).

Veamos que nos dice la tabla acerca de los coeficientes:

- El coeficiente constante a se estima como -466 con una desviación estándar de 1674. Esto correponde al valor t de -0.28 que ocurre con un 79 % de probabilidad con la hipótesis $a = 0$ (valor p). Evidentemente no podemos rechazar la hipótesis nula de $a = 0$ (como hemos mencionado, usualmente se requiere de valores p menores a 0.05 o 0.01 para un resultado significativo).

- El coeficiente b se estima como 725.5 con una desviación estándar de 148.7. De aquí el valor t es 4.88 y la probabilidad de que éste ocurra dado la hipótesis nula $b = 0$ es de 0.003. En este caso el coeficiente si es significativamente diferente del cero y puede rechazarse la hipótesis nula.

Debe tenerse cuidado con los valores p, ya que éstos también dependen del tamaño de la muestra: a mayor tamaño de muestra, menor el valor p.

El renglón debajo de la tabla se refiere a que tan bueno es el modelo lineal propuesto o bien la **bondad del ajuste**. Tenemos:

- $S = 2540.40$ es la desviación estándar de los datos dada la relación lineal estimada entre las variables. Un valor de S pequeño indica una relación lineal robusta. Aquí, dado que los ingresos van de 1632 a 14500, S es relativamente alta.

- $R - Sq = 79.9\,\%$ (también denotada por R^2) es el cuadrado del coeficiente de correlación R o la proporción de la variación en el ingreso -o variable dependiente- que puede explicarse por los años de instrucción -o variable independiente- y toma valores entre cero y uno. Evidentemente, es importante tener un valor alto para $R - sq$ si el modelo quiere utilizarse con propósitos predictivos. Sin embargo, esto no es tan importante cuando simplemente nos interesa determinar si algún coeficiente es significativamente diferente del cero. Un valor bajo de $R - sq$ puede reflejar una cantidad sustancial de ruido en los datos, pero también puede indicar que factores importantes se han omitido en el modelo. Aquí, el $79.9\,\%$ de la variación en el ingreso es explicada por los años de instrucción.

- $R - Sq(adj) = 76.5\,\%$ (también denotada por $R - sq(adj)$) tiene el mismo significado que $R - sq$, pero ajusta el valor al número de predictores (variables independientes) del modelo. Usualmente $R - sq$ sobreestima la fortaleza de la asociación lineal cuando hay más de un predictor, por esta razón $R - sq(adj)$ es un valor menor. El valor de $R - sq(adj)$ se utiliza cuando se comparan modelos con diferentes números de predictores.

El análisis realizado nos dice que, en efecto, los años de instrucción formal o educación son una determinante importante del ingreso. Sin embargo, probablemente no utilizaríamos el modelo como un predictor del ingreso pues la desviación estándar es alta. En las ciencias sociales no es usual utilizar el modelo lineal como un predictor ya que usualmente hay una gran cantidad de ruido en los datos, sin embargo, el análisis es útil para estudiar las interacciones entre variables.

11.3. Regresión lineal múltiple

La regresión lineal simple tiene una sola variable independiente o predictor. Sin embargo, en la mayor parte de las situaciones el efecto a describir tendrá más de una causa. En el ejemplo anterior, seguramente existen más determinantes del ingreso aparte de la educación. La **regresión lineal múltiple** nos permite incluir a más de un predictor en el análsis, de manera que pueda estimarse por separado el efecto de cada uno. Este análisis es sumamente útil para cuantificar el impacto de varios factores sobre una única variable dependiente.

El resultado de la regresión múltiple tiene la misma interpretación que la regresión simple, ilustramos con un ejemplo: sea y el número de puntos promedio por partido que anota un jugador de basketball. Las variables x_1 y x_2 son, respectivamente, el número de minutos adicionales al día que entrena -después de su entrenamiento normal- y el número de minutos diarios que dedica a su arreglo personal. Para una serie de datos de x_1 y x_2 *Minitab* nos proporciona la ecuación,

$$y = 9.31 + 0.787x_1 - 0.044x_2,$$

con el siguiente análisis:

Predictor	**Coef**	**SE Coef**	**T**	**P**
Constant	9.306	8.666	1.07	0.301
x_1	0.78726	0.05655	13.92	0.000
x_2	-0.0441	0.1049	-0.42	0.681

$$S = 13.52 \; R - Sq = 96.8\,\% \quad R - Sq(adj) = 96.4\,\%.$$

El valor p para x_1 aparece como 0, aunque en realidad es menor a 0.001 (recordar que *Minitab* no registra valores menores). De aquí que el coeficiente de x_1 es significativamente diferente de cero. Este coeficiente nos indica que cada minuto adicional de entrenamiento aumenta su promedio de anotaciónes en 0.78726. El coeficiente de x_2 es negativo por lo que el tiempo que dedica a su arreglo personal va en detrimiento de sus encestes. Sin embargo, este efecto es muy pequeño pues su promedio bajaría en apenas 0.0441 anotaciones por cada minuto de arreglo personal. Notemos que el valor p para el coeficiente de x_2 es sumamente alto, de manera que probablemente x_2 no es un buen predictor para y. El valor de $R - sq$ es alto lo cual indica que el modelo predice gran parte de la variación en y, concretamente, el 96.8 %.

11.3.1. Un ejemplo de discriminación

Para ilustrar el proceso de regresión múltiple tomemos el siguiente ejemplo estilizado, basado también en determinantes del ingreso[2].

Supongamos que se demanda a una compañía por discriminación salarial en contra de las mujeres. El abogado demandante presenta como prueba un análisis de regresión múltiple. El modelo estadístico se obtiene tomando los siguientes datos para una muestra de empleados de la compañía:

- salario anual (S),
- escolaridad (E),
- antigüedad en la empresa (A),
- índices de productividad individual (I)
- y lo que llamamos una **variable dummy** que representa el género (G) y toma valores $G = 1$ para hombres y $G = 0$ para mujeres.

Las variables dummy toman valores numéricos arbitrarios para identificar atributos no númericos de la población y son sumamente utilizadas en las ciencias sociales en donde abundan categorías no cuantitativas.

[2]Basado en un ejemplo de "An Introduction to Regression Analysis", Alan O. Sykes, Chicago Working Papers in Law and Economics, Oct, 1993.

La generalización de la ecuación 11.1 para el caso que nos concierne con tres predictores es,

$$S = a + bE + cA + dI + eG + \varepsilon, \tag{11.2}$$

en donde los coeficientes y el error tienen la misma interpretación que antes. Notemos que el propósito del abogado demandante es demostrar que el coeficiente correspondiente al género, e, es positivo y significativamente diferente de cero.

La variable de género toma únicamente valores de 0 y 1, ocasionando que la regresión proporcione, en realidad, dos ecuaciones lineales: una para hombres cuando $G = 1$, dada por,

$$S_H = (a + e) + bE + cA + dI + \varepsilon$$

y otra para mujeres cuando $G = 0$, que es,

$$S_M = a + bE + cA + dI + \varepsilon.$$

Aquí $S_{H,}$ y S_M denotan el salario para el caso de hombres y mujeres, respectivamente.

Observemos que la diferencia entre $S_{H,}$ y S_M está dada por el salario de entrada de un individuo sin educación y sin antigüedad en la empresa. Éste es de a en el caso de las mujeres y de $a + e$ para los hombres. En otras palabras, el término constante de la regresión depende del valor que tome la variable dummy, es decir, ésta interactúa con el término constante.

La regresión estima la siguiente ecuación:

$$S = 4784 + 1146.2E + 285.4A + 39.1I + 1867.6G, \tag{11.3}$$

de manera que los salarios por género (poniendo $G = 1$ o $G = 0$) son,

$$S_H = 6651.6 + 1146.2E + 285.4A + 39.1I,$$
$$S_M = 4784 + 1146.2E + 285.4A + 39.1I.$$

El análisis de los coeficientes y valor $R - sq$ quedan dados por,

Predictor	**Coef**	**SE Coef**	**T**	**P**
Constant	4784.2	945.4	5.060	0.000
E	1146.2	72.0	15.913	0.000
A	285.4	20.2	14.131	0.000
I	39.1	6.8	5.741	0.000
G	1867.6	350.5	5.328	0.000

$R - Sq = 96.4\,\%$.

Incluimos el valor para $R - sq$, simplemente para mostrar que las variables explican un alto porcentaje de la variación en el ingreso.

Todos los valores p en la tabla aparecen como cero, de manera que son menores que 0.001. Así, con un nivel de significancia $\alpha = 0.001$ podemos rechazar las hipótesis nulas de que los coeficientes sean cero. El coeficiente de 1867.6 para la variable dummy de género nos indica que se espera que los hombres ($G = 1$) reciban \$ 1867 más que las mujeres ($G = 0$), manteniendo los demás factores constantes. Esto indicaría, salvo nueva evidencia, que podemos rechazar el que no exista discriminación.

11.3.2. Multicolinearidad

Prosigamos con el ejemplo de discriminación de la sección anterior. Supongamos que en el análisis inicial se piensa que la discriminación también sucede mediante el valor que se da a la escolaridad. Es decir, cada año de escolaridad tendrá mayor impacto en el ingreso de los hombres que en el de las mujeres. Para modelar esto, se propone la estimación del siguiente modelo lineal:

$$S = a + bE + cA + dI + eG + f(E \times G) + \varepsilon,$$

en donde agregamos una nueva variable que es el producto $E \times G$.

Así, cuando $G = 1$, el salario es,

$$S_H = (a + e) + (b + f)E + cA + dI + \varepsilon$$

y cuando $G = 0$,

$$S_M = a + bE + cA + dI + \varepsilon.$$

Vemos que ahora la diferencia salarial entre hombres y mujeres, no sólo está en el término independiente sino en el coeficiente de E. La variable dummy ahora interactúa sobre dos coeficientes.

Supongamos que se obtiene la siguiente estimación:

$$S = 3500.2 + 962.3E + 288.3A + 61.2I + 4243.7G - 14.8(E \times G),$$

de manera que, poniendo $G = 1$ y $G = 0$ obtenemos,

$$S_H = 7743.9 + 947.5E + 288.3A + 61.2I,$$
$$S_M = 3500.2 + 962.3E + 288.3A + 61.2I.$$

La tabla de la regresión es,

Predictor	**Coef**	**SE Coef**	**T**	**P**
Constant	3500.2	2130.6	1.643	0.108
E	962.3	144	6.679	0.000
A	288.3	36.7	7.861	0.000
I	61.2	12.3	4.966	0.000
G	4243.7	2916.5	1.455	0.153
E×G	-14.8	198.2	-0.075	0.941

$R - Sq = 90.9\,\%$.

Los primeros cuatro coeficientes no parecen tener problema, sin embargo, los coeficientes de G y $E \times G$ nos deben llamar la atención. Primero, el coeficiente para $E \times G$ tiene un signo negativo lo cual indica que hay discriminación negativa con respecto a la escolaridad: cada año de escolaridad proporciona un mayor incremento en el ingreso de las mujeres que en el de los hombres. Segundo, el nivel de significancia es muy bajo, en especial para el término $E \times G$, ya que los valores p son demasido grandes. Tercero, el valor del coeficiente para G es muy alto, comparado con el de la regresión anterior, pero con una desviación estándar también grande. Finalmente, $R - sq = 90.9$, que a pesar de ser alto es menor al valor obtenido en la regresión anterior, es decir, el primer modelo proporciona un mejor ajuste lineal.

Los problemas descritos pueden derivarse del fenómeno llamado **multicolinearidad.** Éste sucede cuando dos o más de los predictores tienen una alta correlación entre ellos. En este caso, G y $E \times G$ tienen un coeficiente de correlación de 0.96, lo cual complica el análisis. Cuando esto sucede, la regresión no puede separar los efectos de las variables que están correlacionadas. Idealmente, todos los predictores deben ser independientes entre sí. Aquí podría cometerse el error de no rechazar la ausencia de discriminación, en base a esta última regresión.

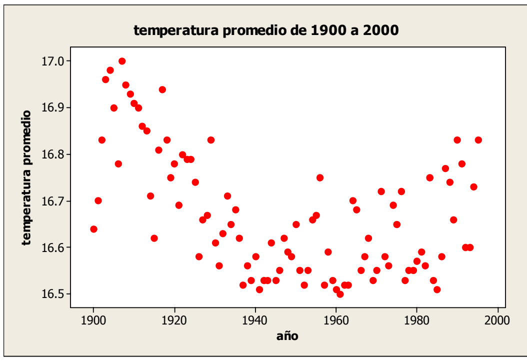
En el ejercicio 11.7 se ilustra cómo la omisión de variables independientes puede dañar el análisis de regresión. La solución parecería ser la introducción del mayor número de predictores posible, sin embargo, esto conlleva el peligro de multicolinearidad, pues los predictores podrían no ser independientes entre sí. ¿Hay alguna solución sencilla para este problema? La respuesta, desafortunadamente es negativa y hay que analizar cada caso para evaluar la mejor solución según convenga. La moraleja es que se debe ser sumamente cauto con los resultados de los análisis de regresión.

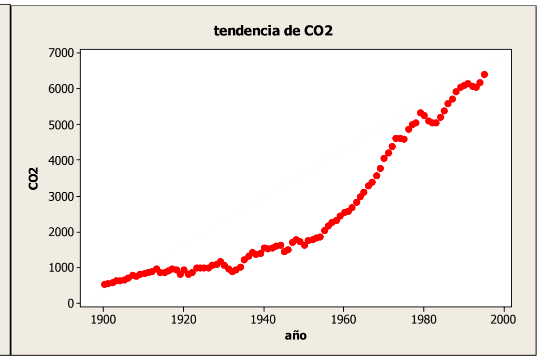
Ejemplos

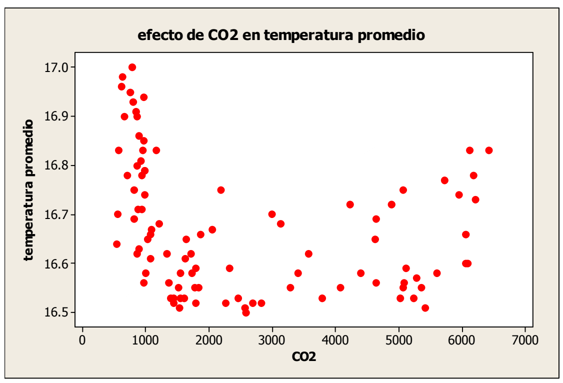
Ej 11.3.1 Los siguientes son coeficientes de correlación proporcionados por Minitab para datos acerca de la temperatura mundial promedio y la cantidad de CO_2 en la atmósfera entre los años de 1900 y 2000:

Pearson correlation of “año” and “temperatura promedio” $= -0.496$
P-Value $= 0.000$
Pearson correlation of “año” and $CO_2 = 0.937$
P-Value $= 0.000$
Pearson correlation of “temperatura promedio” and $CO_2 = -0.275$
P-Value $= 0.007$

Se concluye que en estos 100 años, la temperatura promedio ha bajado, el CO_2 ha aumentado y no hay evidencia de que el aumento en CO_2 incremente la temperatura, ¡sino más bien lo contrario! Los siguientes diagramas de dispersión ilustran estas correlaciones







Entonces, ¿el calentamiento global es un mito? En este caso, 100 años no son suficientes para llegar a una conclusión. Los cambios climáticos necesitan miles (hasta millones) de años para manifestar alguna tendencia, de manera que necesitaríamos datos de mucho más de 100 años para poder concluir algo.

Ej 11.3.2 Se realiza un análsis de regresión sobre una muestra de obreros de una empresa automotriz para determinar el impacto de los años de experiencia sobre su ingreso (incluidas prestaciones). Los años de experiencia se denotan por X y el ingreso trimestral (en miles de pesos) por I. La regresión en Minitab nos proporciona los siguientes resultados:

$$I = 11.0 + 2.66X$$

Predictor	**Coef**	**SE Coef**	**T**	**P**
Constant	11.031	2.050	5.38	0.000
X	2.6563	0.5870	4.52	0.000

$$S = 3.321 \; R - Sq = 71.9\,\% \quad R - Sq(adj) = 68.4\,\%.$$

El valor de $R - Sq = 71.9\,\%$ indica que éste es el porcentaje de variación en el ingreso explicado por los años de experiencia o bien que la correlación entre los datos de experiencia e ingreso es de $R = \sqrt{0.719} = 0.848$. Dado que el valor p para el coeficiente de X es menor que 0.001, se rechaza la hipótesis nula de que la experiencia no es determinante del ingreso.

Utilizemos la ecuación lineal estimada para hacer algunas predicciones: un trabajador sin experiencia se espera que perciba trimestralmente

$$I = 11.0 + 2.66 \times 0 = 11 \text{ mil pesos}$$

y un trabajador con 6 años de experiencia se espera que tenga un ingreso trimestral de

$$I = 11.0 + 2.66(6) = 29.96 \text{ miles de pesos.}$$

Ej 11.3.3 (Discriminación) Una empresa es demandada por discriminación salarial hacia las mujeres. Concretamente, se alega que dado el mismo nivel de experiencia, existe un sesgo salarial que favorece a los hombres. Para demostrar la presencia de discriminación, se toma una muestra de 100 empleados con datos acerca de los años de trabajo en la empresa (experiencia, denotada por X), su ingreso trimestral (denotado por I) y una variable dummy de género (denotada por G) que es cero para las mujeres y uno para los hombres.

Minitab estima la ecuación,

$$I = 17161 + 5244.4X + 6867G$$

con el análisis de coeficientes dado por la tabla,

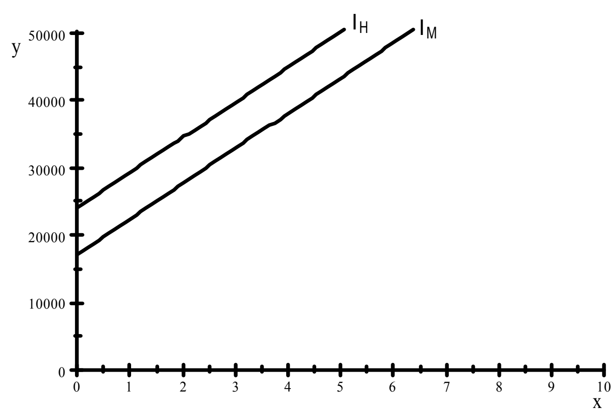
Predictor	**Coef**	**SE Coef**	**T**	**P**
Constant	17161	2770	6.19	0.000
X	5244.4	405.2	12.94	0.000
G	6867	2328	2.95	0.004

$S = 11638.0 \quad R - Sq = 64.5\,\% \quad R - Sq(adj) = 63.8\,\%$

Notamos que tomando $G = 1$ y $G = 0$ en la ecuación estimada, se tienen los ingresos por género (hombres y mujeres, respectivamente) dados por,

$$I_H = 24028 + 5244.4X,$$
$$I_M = 17161 + 5244.4X.$$

Graficando estas dos rectas se obtiene,



El abogado demandante alega que el coeficiente de género es significativamente diferente del cero, con un nivel de significancia alto ($p = 0.004$). Los hombres parecen percibir \$ 6867 más cada trimestre, lo cual se muestra en la gráfica. El modelo explica un 64.5 % de la variación en el ingreso dado por el valor de $R - sq$.

El abogado de la empresa se defiende con el siguiente argumento: por política de la empresa, las mujeres son favorecidas salarialmente al ser contratadas. Adicionalmente, la empresa otorga licencias de maternidad con sueldo parcial que pueden durar hasta un año. Dado que un gran número de mujeres hacen uso de esta prestación, los rendimientos a la experiencia son menores en las mujeres. Es decir, un año de experiencia tiene mayor impacto en el salario de un hombre que en el de una mujer.

Se utilizan los mismos datos de la muestra de 100 empleados pero ahora se introduce la variable $G \times X$. La ecuación estimada es ahora,

$$I = 23047 + 4174.3X - 4904G + 2140.2(G \times X)$$

y la tabla con el análisis de la regresión es,

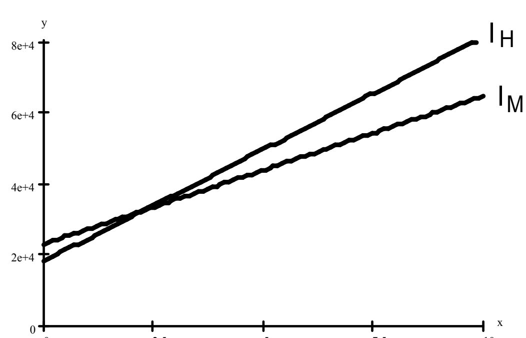
Predictor	**Coef**	**SE Coef**	**T**	**P**
Constant	23047	3443	6.69	0.000
X	4174.3	554.9	7.52	0.000
G	-4904	4869	-1.01	0.316
$G \times X$	2140.2	784.7	2.73	0.008

$$S = 11270.0 \quad R - Sq = 67.1\,\% \quad R - Sq(adj) = 66.0\,\%$$

Una vez más, pueden obtenerse las ecuaciones correspondientes al ingreso de hombres y mujeres poniendo $G = 1$ y $G = 0$. Éstas son,

$$I_H = 18143 + 6314.5X,$$
$$I_M = 23047 + 4174.3X.$$

Que se representan geométricamente como,



El abogado defensor alega que el segundo modelo describe mejor la situación. El coeficiente correspondiente al género es negativo, mostrando que el salario inicial de las mujeres es mayor. Este coeficiente no es significativo por el alto valor p ($p = 0.316$), pero el coeficiente de $G \times X$ si es significativo y positivo. El ajuste del segundo modelo es mejor pues presenta un valor menor de S y valores mayores de $R - sq$ y $R - sq(adj)$, de manera que no hay que preocuparse por la multicolinearidad. El caso se resuelve a favor de la empresa.

Ej 11.3.4 Supongamos que se quiere un modelo para predecir el precio de venta de un casa (P) en cierta zona de la ciudad de México. Las variables a considerar son las siguientes: metros cuadrados de terreno (T), metros cuadrados de construcción (C), número de habitaciones (H), tipo de construcción (L) y distancia al primer crucero importante (D). Se tiene la matriz de correlación entre todas estas variables, en donde cada entrada representa el coeficiente de correlación entre la variable columna y la variable renglón[3]:

$$\left(\begin{array}{ccccccc} & P & T & C & H & L & D \\ P & 1 & & & & & \\ T & 0.82 & 1 & & & & \\ C & 0.90 & 0.95 & 1 & & & \\ H & 0.88 & 0.90 & 0.92 & 1 & & \\ L & 0.85 & 0.60 & 0.62 & 0.35 & 1 & \\ D & 0.70 & 0.30 & 0.28 & 0.25 & 0.15 & 1 \end{array}\right)$$

Todos los predictores propuestos parecen estar altamente correlacionados con el precio de venta (primera columna). Sin embargo, sería un error utilizarlos todos puesto que se presentaría el problema de multicolinearidad. Los metros cuadrados de terreno (T), de construcción (C) y el número de habitaciones (H), también están altamente correlacionados entre sí, como es de esperarse. Una buena opción podría ser eliminar T y H y quedarnos

[3]Las entradas de la diagonal son unos puesto que representan la correlación de cada variable consigo misma.

con C, L y D. Los paquetes estadísticos realizan varias combinaciones de predictores y eligen la mejor regresión posible para casos como este.

♠♠♠♠♠

Bibliografía Complementaria

1. Agresti Alan, Finlay Barbara, *Statistical Methods for the Social Sciences*,4^a ed, Pearson, Allyn & Bacon, New York, NY, 2008.

2. Burstein, James A., and Jeri A. Lindahl, "Defending an Age Discrimination Case Involving Statistical Evidence" *Employee Relations Law Journal* , vol. 18, number 2, pp. 325-336, (Autumn 1992) .

3. Campbell Thomas J., "Regression Analysis in Title VII Cases: Minimum Standards, Comparable Worth, and Other Issues Where Law and Statistics Meet", *Stanford Law Review* vol 36, pp.1299-1324*,* July 1984.

4. Finkelstein Michael O., "The Judicial Reception of Multiple Regression Studies in Race and Sex Discrmination Cases", *Columbia Law Review* vol 80, No 12, pp. 737-754 (1980).

5. Finkelstein Michael O., Bruce Levin, *Statistics for Lawyers*, Springer Verlag, New York, NY, 2007.

6. Fisher Franklin M., "Multiple Regression in Legal Proceedings", *Columbia Law Review* vol 80, No 4, pp. 737-754, 1980.

7. Hazelwood School District v. United States, 433 U.S. 299, 1977. Puede encontrarse en,

   $http://caselaw.lp.findlaw.com/scripts/getcase.pl?court = US\&vol = 433\&invol = 299$

8. Jackson H., Kaplow L., Shavell S., Viscusi W. K. and Cope D., *Analytical Methods for Lawyers (Statistics)*, Foundation Press 2003.

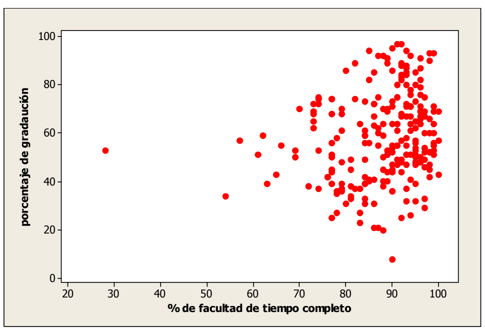
9. Rubinfeld Daniel L., "A Reference Guide on Multiple Regression", *Reference Manual on Scientific Evidence*, 2nd edition, Federal Justice Center. Versión electrónica en,

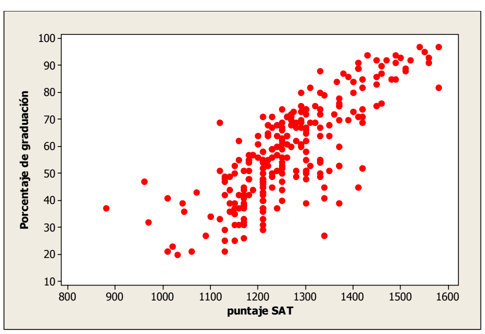
   $http://www.fjc.gov/public/pdf.nsf/lookup/sciman01.pdf/\$file/sciman01.pdf$

10. Zeisel Hans, Kaye David, *Prove it with Figures: Empirical Methods in Law and Litigation*, Springer Verlag, New York, NY, 1997.

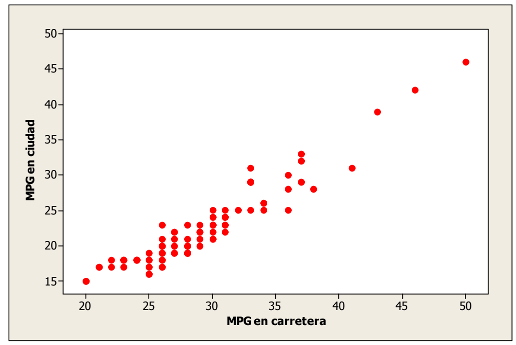
Ejercicios

▷ **11.1** Los siguientes diagramas de dispersión muestran el porcentaje de graduación vs el porcentaje de profesores de tiempo completo y el porcentaje de graduación vs el puntaje en el examen de admisión a la universidad (SAT). Los datos corresponden a 249 instituciones de educación superior en Estados Unidos. ¿Qué puede concluirse?





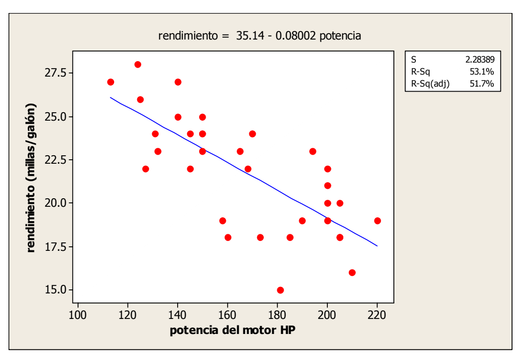
▷ **11.2** El siguiente es un diagrama de dispersión y el análisis de correlación para los datos de rendimiento en ciudad y rendimiento en carretera (en millas por galón, MPG), para una muestra de 93 automóviles



Pearson correlation of MPG ciudad and MPG carretera = 0.944, P-Value = 0.000. ¿Habrá alguna relación de causalidad entre estas variables? Justificar.

▷ **11.3** Se muestra la regresión para ver el efecto de la potencia del motor (en HP o caballos de fuerza) en el rendimiento de gasolina (en millas por galón) de 34 modelos de vehículos automotores.

rendimiento=35.1-0.08potencia



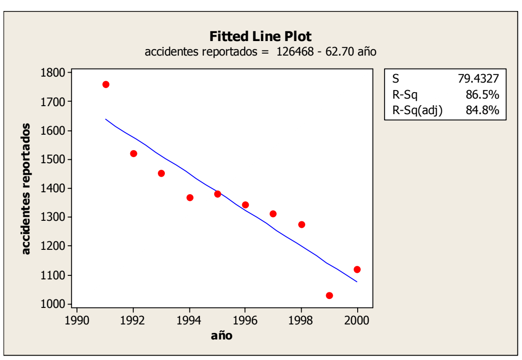
Predictor	Coef	SE Coef	T	P
Constant	35.136	2.254	15.59	0.000
Potencia	−0.08002	0.01329	−6.02	0.000

S=2.28389 R-Sq=53.1 % R-Sq(adj)=51.7 %

1. ¿Qué rendimiento se esperaría tener de un coche con una potencia de 240 HP?

2. Decribir e interpretar los resultados de la tabla de la regresión.

▷ **11.4** A partir de 1990 se comenzó a promover el uso del casco para los ciclistas en EUA. La siguiente figura de Minitab muestra el número de accidentes ciclistas por año entre 1990 y 2000. Se muestra también el ajuste de la regresión. ¿Qué puede concluirse? Justificar.

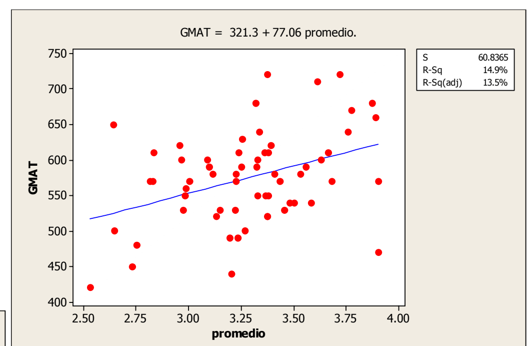


▷ **11.5** Quiere determinarse si el promedio de un estudiante en la licenciatura tiene impacto en la calificación que obtiene en el examen GMAT utilizado para los posgrados en ciencias de la administración y negocios. Se toman datos de 61 estudiantes y Minitab proporciona los siguientes resultados.

GMAT=321.3+77.06promedio

Predictor	Coef	SE Coef	T	P
Constant	321.31	79.11	4.06	0.000
Promedio	77.06	23.94	3.22	0.002

S=60.8365 R-Sq=14.1 % R-Sq(adj)=13.5 %



1. ¿Cuál es el coeficiente de correlación entre el promedio y el resultado del GMAT?

2. ¿Qué calificación se esperaría que obtuviera un estudiante con cero de promedio?

3. ¿Qué tan buen predictor es el promedio para el resultado del GMAT? Justificar

▷ **11.6** Con los mismos datos del ejercicio anterior, se añade una variable dummy G para denotar el género: $G = 0$ para mujeres y $G = 1$ para hombres. El resultado de la nueva regresión es,

GMAT=297+83.4promedio+14.3género,

Predictor	Coef	SE Coef	T	P
Constant	297.41	86.44	3.44	0.001
Promedio	83.40	25.68	3.25	0.002
Género	14.33	20.41	0.70	0485

S=61.0997 R-Sq=15.7 % R-Sq(adj)=12.7 %

¿Qué tan significativo es el género en el resultado del GMAT? Justificar.

▷ **11.7** En el análisis de la regresión dada por la ecuación 11.3 se omite la variable E, correpondiente a los años de escolaridad. Al realizar la regresión con los predictores restantes se obtiene el siguiente resultado:

I=9806.5+256.9A+107.5I+2445.5G

Predictor	Coef	SE Coef	T	P
Constant	9806.5	4653.8	2.107	0.041
A	256.9	103.3	2.487	0.017
I	107.5	25.6	4.173	0.000
G	2445.5	1779	1.375	0.176

R-Sq=40.8 %.

1. ¿Qué pasa con los coeficientes de las variables A, I y G en comparación con la regresión original? Sugerir una explicación para esta diferencia.

2. ¿Qué pasa con el valor de $R - sq$ en comparación con la regresión original? Sugerir una explicación para esta diferencia.

3. ¿Qué puede concluirse acerca de la elección de los predictores?

▷ **11.8** Se realizó un análisis, a partir de una muestra de 311 películas, para estudiar los determinantes de sus ganancias de taquilla (G), en millones de dólares.

1. Como primer paso se incluye como predictor el costo (C) de la película obteniendo,

   G =7.24+1.02C

Predictor	Coef	SE Coef	T	P
Constant	7.243	4.897	1.48	0.142
C	1.0191	0.1106	9.22	0.000

S=37.4953 R-Sq=39.9 % R-Sq(adj)=39.4 %

Describir y analizar el resultado.

2. En segunda instancia se incluye también como predictor la variable dummy E que vale 1 si la película contiene actores célebres y 0 en el caso contrario. Se obtiene,

G=6.76+0.859C+20.7E

Predictor	Coef	SE Coef	T	P
Constant	6.759	4.732	1.43	0.156
C	0.8591	0.1180	7.28	0.000
E	20.704	6.48	3.19	0.002

S=36.2185 R-Sq=44.4 % R-Sq(adj)=43.5 %

Describir y analizar el resultado y comparar con el inciso anterior.

3. Como tercer paso se incluye un nuevo predictor, que es el núemro de pantallas (N) en el cual se estrenó la película. Se obtiene ahora,

G=-6.89+0.375C+17E+0.0195N

Predictor	Coef	SE Coef	T	P
Constant	-6.89	5.128	-1.34	0.181
C	0.3747	0.1453	2.58	0.011
E	17.044	5.995	2.84	0.005
N	0.0195	0.0039	5	0.000

S=33.2184 R-Sq=53.6 % R-Sq(adj)=52.5 %

Describir y analizar el resultado y comparar con los incisos anteriores.

4. Finalmente se agregan las ganancias en el primer fin de semana de estreno (I) como predictor. El resultado de la regresión es,

G=1.91-0.089C+13.1E-0.00191P+4.13I

Predictor	Coef	SE Coef	T	P
Constant	1.911	3.420	0.56	0.577
C	-0.0886	0.1014	-0.87	0.384
E	13.084	3.931	3.33	0.001
N	-0.001906	0.003034	-0.63	0.531
I	4.1335	0.3172	13.03	0.000

S=21.7154 R-Sq=80.3 % R-Sq(adj)=79.7 %

Describir y analizar el resultado y comparar con los incisos anteriores. ¿Qué predictores tendrán una alta correlación entre sí? ¿Son lógicos los signos para C y P?

5. Después de realizar todas estas regresiones alguien sugiere simplemente tomar el ingreso en el primer fin de semana y la presencia de estrellas como predictores. Se ob-

tiene:

G=-0.392+3.80I+11.1E

Predictor	Coef	SE Coef	T	P
Constant	-0.3921	0.9478	-0.41	0.678
I	3.7978	0.1115	34.07	0.000
E	11.055	2.381	4.64	0.000

S=14.3176 R-Sq=84.9 % R-Sq(adj)=84.8 %

Describir y analizar el resultado y comparar con los incisos anteriores. ¿Porqué se obtiene un mejor ajuste con menos predictores?

Apéndice A

Respuestas a ejercicios

Capítulo1

1.1

1. Si, falso.
2. No.
3. Si, verdadero.
4. No (valor de verdad subjetivo).
5. Si, verdadero (o falso, después de agosto 2008).

1.2

1. P :Ninguna planta es carnívora, Q :La planta devoró a la hormiga. Res: Algunas plantas son carnívoras, La planta no devoró a la hormiga, Ninguna planta es carnívora o la planta no devoró a la hormiga, Algunas plantas son carnívoras y la planta devoró a la hormiga.
2. P :Al menos un estudiante obtuvo un 10 en el examen, Q :El examen no era difícil. Res: Ningún estudiante obtuvo 10 en el examen, El examen era difícil, Al menos un estudiante obtuvo un 10 en el examen

o el examen no era difícil, Ningún estudiante obtuvo 10 en el examen y el examen no era difícil.

3. P :Todos los modelos están disponibles en todas las tiendas, Q :Algunos zapatos tienen 50 % de descuento. Res: Algunos modelos no están disponibles en algunas las tiendas, Ningún zapato tiene 50 % de descuento, Todos los modelos están disponibles en todas las tiendas o Ningún zapato tiene 50 % de descuento, Algunos modelos no están disponibles en algunas las tiendas y Algunos zapatos tienen 50 % de descuento.

4. P :Algunas personas concentran toda la riqueza, Q :La riqueza hace la felicidad. Res: Todas las personas tienen algo de riqueza, La riqueza no hace la felicidad, Algunas personas concentran toda la riqueza o la riqueza hace la felicidad, Todas las personas tienen algo de riqueza y la riqueza hace la felicidad.

5. P :Todos los números racionales son reales, Q :Al menos un número real no es racional. Res: Algunos números racionales no son reales, Todos los números reales son racionales, Todos los números racionales son reales o todos los números reales son racionales, Algunos números racionales no son reales y al menos un número real no es racional.

1.3

1. $\sim P \wedge Q$,

P	Q	$\sim P$	$\sim P \wedge Q$
V	V	F	F
V	F	F	F
F	V	V	V
F	F	V	F

2. $\sim P \vee \sim Q$,

P	Q	$\sim P$	$\sim Q$	$\sim P \vee \sim Q$
V	V	F	F	F
V	F	F	V	V
F	V	V	F	V
F	F	V	V	V

3. $P \vee (\sim Q \wedge P)$,

P	Q	$\sim Q$	$\sim Q \wedge P$	$P \vee (\sim Q \wedge P)$
V	V	F	F	V
V	F	V	V	V
F	V	F	F	F
F	F	V	F	F

4. $\sim (\sim P \vee \sim Q)$,

P	Q	$\sim P$	$\sim Q$	$\sim P \vee \sim Q$	$\sim (\sim P \vee \sim Q)$
V	V	F	F	F	V
V	F	F	V	V	F
F	V	V	F	V	F
F	F	V	V	V	F

5. $(\sim P \wedge \sim Q) \vee \sim Q$,

P	Q	$\sim P$	$\sim Q$	$\sim P \wedge \sim Q$	$(\sim P \wedge \sim Q) \vee \sim Q$
V	V	F	F	F	F
V	F	F	V	F	V
F	V	V	F	F	F
F	F	V	V	V	V

6. $(P \wedge R) \vee \sim Q$,

P	Q	R	$P \wedge R$	$\sim Q$	$(P \wedge R) \vee \sim Q$
V	V	V	V	F	V
V	V	F	F	F	F
V	F	V	V	V	V
V	F	F	F	V	V
F	V	V	F	F	F
F	V	F	F	F	F
F	F	V	F	V	V
F	F	F	F	V	V

7. $(\sim P \wedge Q) \vee \sim R$.

P	Q	R	$\sim P$	$\sim R$	$\sim P \wedge Q$	$(\sim P \wedge Q) \vee \sim R$
V	V	V	F	F	F	F
V	V	F	F	V	F	V
V	F	V	F	F	F	F
V	F	F	F	V	F	V
F	V	V	V	F	V	V
F	V	F	V	V	V	V
F	F	V	V	F	F	F
F	F	F	V	V	F	V

1.4

1. $P \vee P, P \wedge P$ yP,

P	$P \vee P$	$P \wedge P$
V	**V**	**V**
F	**F**	**F**

2. $\sim (\sim P)$ y P,

P	$\sim P$	$\sim (\sim P)$
V	F	**V**
F	V	**F**

3. $P \vee (P \wedge Q)$ y $P \wedge (P \vee Q)$,

P	Q	$P \wedge Q$	$P \vee Q$	$P \vee (P \wedge Q)$	$P \wedge (P \vee Q)$
V	V	V	V	**V**	**V**
V	F	F	V	**V**	**V**
F	V	F	V	**F**	**F**
F	F	F	F	**F**	**F**

4. $P \vee (Q \wedge R)$ y $(P \vee Q) \wedge (P \vee R)$,

P	Q	R	$Q \wedge R$	$P \vee Q$	$P \vee R$	$P \vee (Q \wedge R)$	$(P \vee Q) \wedge (P \vee R)$
V	V	V	V	V	V	**V**	**V**
V	V	F	F	V	V	**V**	**V**
V	F	V	F	V	V	**V**	**V**
V	F	F	F	V	V	**V**	**V**
F	V	V	V	V	V	**V**	**V**
F	V	F	F	V	F	**F**	**F**
F	F	V	F	F	V	**F**	**F**
F	F	F	F	F	F	**F**	**F**

5. $P \wedge (Q \vee R)$ y $(P \wedge Q) \vee (P \wedge R)$.

P	Q	R	$Q \vee R$	$P \wedge Q$	$P \wedge R$	$P \wedge (Q \vee R)$	$(P \wedge Q) \vee (P \wedge R)$
V	V	V	V	V	V	**V**	**V**
V	V	F	V	V	F	**V**	**V**
V	F	V	V	F	V	**V**	**V**
V	F	F	F	F	F	**F**	**F**
F	V	V	V	F	F	**F**	**F**
F	V	F	V	F	F	**F**	**F**
F	F	V	V	F	F	**F**	**F**
F	F	F	F	F	F	**F**	**F**

1.5 $\sim (P \wedge Q)$ y $\sim P \vee \sim Q$.

P	Q	$\sim P$	$\sim Q$	$P \wedge Q$	$\sim (P \wedge Q)$	$\sim P \vee \sim Q$
V	V	F	F	V	**F**	**F**
V	F	F	V	F	**V**	**V**
F	V	V	F	F	**V**	**V**
F	F	V	V	F	**V**	**V**

1.6 Utilizamos $\sim (P \vee Q) = \sim P \wedge \sim Q$

1. No pediré pasta y no pediré pizza.
2. No estudiaré diariamente y tendré vida social

1.7 Utilizamos $\sim (P \wedge Q) = \sim P \vee \sim Q$

1. No tomaré vino o no tomaré cerveza.
2. Alfonso es honesto o es buen actor.

1.8

1. Si hay marchas callejeras entonces hay descontento social.
2. Si es duende, entonces es pequeño.
3. Si es duende, entonces no es pequeño.
4. Si es rana, entonces es verde.
5. Si se utiliza la linterna, entonces se compran baterías.
6. Si se tiene un promedio de 9, entonces se conserva la beca.
7. Si es mexicano, entonces la constitución rige su vida.
8. Si me dejas jugar, entonces te presto el balón.
9. Si es perico, entonces dondequiera es verde.

10. Si no vas, entonces te arrepentirás.

11. Si es trabajador, entonces lo proteje la ley Federal del Trabajo.

12. Si tiene instituciones sólidas, entonces el país saldrá adelante

13. Si es cítrico, entonces contiene vitamina C.

14. Si no repruebo el examen, entonces me voy a festejar.

15. Si no te informas, entonces no opines.

1.9 Observación: se utiliza la equivalencia $\sim P \vee Q = Q \vee \sim P$ en algunos casos.

1. No hay marchas callejeras o hay descontento social.

2. Es pequeño o no es duende.

3. No es duende o no es pequeño.

4. Es verde o no es rana.

5. Se compran baterías o no se utiliza la linterna.

6. Se conserva la beca o no se tiene promedio de 9.

7. La constitución rige su vida o no es mexicano.

8. Te presto el balón o no me dejas jugar.

9. Dondequiera es verde o no es perico.

10. Vas o te arrepentirás.

11. Lo proteje la ley Federal del trabajo o no es trabajador.

12. El país saldrá adelante o no se tienen instituciones sólidas.

13. No es cítrico o contiene vitamina C.

14. Repruebo el examen o me voy a festejar.

15. Te informas o no opines.

1.10 Observación: se utiliza la equivalencia $P \wedge \sim Q = \sim Q \wedge P$ en algunos casos.

1. Hay marchas callejeras y no hay descontento social.
2. Es duende y no es pequeño.
3. Es duende y es pequño.
4. Es rana y no es verde.
5. Se utiliza la linterna y no se compran baterías.
6. Se tiene promedio de 9 y no se conserva la beca.
7. Es mexicano y su vida no la rige la constitución.
8. No te presto el balón y me dejas jugar.
9. Es perico y no es verde dondequiera.
10. No vas y no te arrepentirás.
11. Es trabajador y no lo proteje la ley Federal del Trabajo.
12. Se tienen instituciones sólidas y el país no saldrá adelante.
13. Es cítrico y no contiene vitamina C.
14. No repruebo el examen y no me voy a festejar.
15. No te informas y opinas.

1.11

1. Si hay descontento social, entonces hay marchas callejeras; si no hay marchas callejeras, entonces no hay descontento social; si no hay descontento social, entonces no hay marchas callejeras.

2. Si es pequeño, entonces es duende; si no es duende, entonces no es pequeño, si no es pequeño, entonces no es duende.

3. Si no es pequeño, entonces es duende; si no es duende, entonces es pequeño; si es pequeño, entonces no es duende.

4. Si es verde, entonces es rana; si no es rana, entonces no es verde; si no es verde, entonces no es rana.

5. Si se compran baterías, entonces se utiliza la linterna; si no se utiliza la linterna, entonces no se compran baterías; si no se compran baterías, entonces no se utiliza la linterna.

6. Si se conserva la beca, entonces se tiene promedio de 9; si no se tiene promedio de 9, entonces no se conserva la beca, si no se conserva la beca, entonces no se tiene promedio de 9.

7. Si la constitución rige su vida, entonces es mexicano; si no es mexicano, entonces la constitución no rige su vida; si la constitución no rige su vida, entonces no es mexicano.

8. Si te presto el balón, entonces me dejas jugar; si no me dejas jugar, entonces no te presto el balón; si no te presto el balón, entonces no me dejas jugar.

9. Si dondequiera es verde, entonces es perico; si no es perico, entonces puede no ser verde; si a veces no es verde, entonces no es perico.

10. Si te arrepientes, entonces no fuiste; si vas, entonces no te arrepentirás; Si no te arrepientes, entonces fuiste.

11. Si lo proteje la Ley Federal del Trabajo, entonces es trabajador; si no es trabajador, entonces no lo proteje la Ley Federal del Trabajo; si no lo proteje la ley Federal del Trabajo, entonces no es trabajador.

12. Si el país sale adelante, entonces tiene instituciones sólidas; si no tiene instituciones sólidas, entonces el país no saldrá adelante; si el país no sale adelante, entonces no tiene instituciones sólidas.

13. Si contiene vitamina C, entonces es cítrico; si no es cítrico, entonces no contiene vitamina C; si no contiene vitamina C, entonces no es cítrico.

14. Si me voy a festejar, entonces no reprobé el examen; si repruebo, entonces no me voy a festejar; si no voy a festejar, entonces reprobé el examen.

15. Si no opinas, entonces no te informas; si te informas, entonces opinas; si opinas, entonces, te informas.

1.12

1. V
2. F
3. V
4. F

1.13

1. válido.
2. válido.
3. inválido.

1.14 Sea P el conjunto de enteros pares. Considérese el argumento:

$$\begin{array}{c} \text{Si } P \text{ es el conjunto de números pares, entonces } P \text{ es infinito,} \\ P \text{ es finito,} \\ \hline P \text{ no es el conjunto de números pares.} \end{array}$$

Supongamos que el conjunto de enteros pares es finito, digamos que tiene n elementos. Sea $P = \{p_1, p_2, \ldots, p_n\}$ el conjunto de TODOS los pares, ordenados de menor a mayor, es decir, $p_1 < p_2 < \ldots < p_n$. Entonces, el número $p_n + 2$ es un número par pues es divisible entre 2 y no está en este conjunto. De aquí que P no es el conjunto de números pares. La contradicción se origina al asumir que P es finito por lo que concluímos que P es infinito.

1.15

1. P : es Hérdez, Q : es bueno, el argumento es MP y es válido.

$$\begin{array}{c} P \Rightarrow Q \\ P, \\ \hline Q. \end{array}$$

   Se sobreentiende la condicional $P \Rightarrow Q$.

2. Válido, argumento transitivo.

$$\begin{array}{c} P \Rightarrow Q \\ Q \Rightarrow R, \\ \hline P \Rightarrow R. \end{array}$$

3. Válido, modus ponens.

4. No es válido, pues una condicional y su inversa no son equivalentes.

$$\begin{array}{c} P \Rightarrow Q \\ \sim P, \\ \hline \sim Q. \end{array}$$

5. No es válido, ambas premisas son verdaderas pero la conclusión es falsa.

$$\begin{array}{c} P \Rightarrow Q \\ P \Rightarrow R, \\ \hline Q \Rightarrow R. \end{array}$$

6. Válido, es silogismo disyuntivo.

$$\begin{array}{c} P \vee Q \\ \sim P, \\ \hline Q. \end{array}$$

7. Válido, es un argumento transitivo.

$$\begin{array}{c} P \Rightarrow Q \\ Q \Rightarrow R, \\ \hline P \Rightarrow R. \end{array}$$

8. Válido pues es del tipo modus tollens.

$$\begin{array}{c} P \Rightarrow Q \\ \sim Q, \\ \hline \sim P. \end{array}$$

1.16

1. Parecería modus ponens, pero la segunda premisa no es el antecedente de la primera premisa puesto que la palabra "ley" tiene otro significado. En realidad este argumento es

$$\begin{array}{c} P \Rightarrow Q \\ R, \\ \hline Q. \end{array}$$

   que no es válido.

2. El argumento es válido (MP) pero su conclusión es falsa puesto que la primera premisa es falsa: las marcas de aguja en el brazo son signo de consumo de drogas o de otro tipo de condición. Por ejemplo, la detenida podía haber estado sujeta a un tratamiento de quimioterapia.

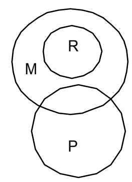
3. No es válido pues una condicional y su inversa no son equivalentes.

4. No es válido pues una condicional y su inversa no son equivalentes.

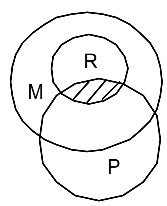
5. Es válido, silogismo disyuntivo.

1.17

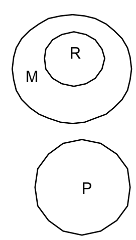
1. No es válido.



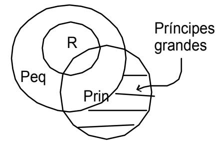
2. Válido.



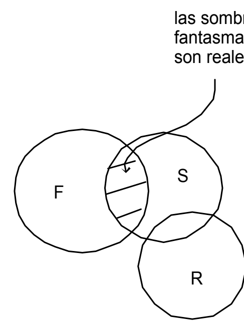
3. Válido.



4. Inválido.



5. Válido.



Capítulo 2

2.1 ¡No! Cada vuelo de aerolíneas Patito es independiente de los vuelos anteriores. Si la aerolínea ha sufrido accidentes últimamente es posible que esto refleje falta de mantenimiento en los aviones u algún otro problema. Más bien habría que investigar eso.

2.2 En EUA $P(0 \text{ ó } 00) = \frac{2}{38} = \frac{1}{19}$. En Europa $P(0) = \frac{1}{37}$.

2.3

1. $P(3) = \frac{1}{9}$.
2. $P(impar) = \frac{3}{9} = \frac{1}{3}$.

2.4 Los resultados estarán sesgados hacia los números de la esquina defectuosa.

2.5 No es posible asegurar que sean independientes ya que podría suceder que un equipo gane sólo si adquiere jugadores del otro durante la temporada.

2.6

1. {aaa,aas,asa,saa,ssa,sas,ass,sss}.
2. {1a,1s,2a,2s,3a,3s,4a,4s,5a,5s,6a,6s}.
3. $\{x + y \mid x, y \in \{1, 2, \ldots, 6\}\}$.
4. {ffff, fffm, ffmf, fmff, mfff, ffmm, fmfm, mfmf, mmff, fmmf, mffm, fmmm, mfmm, mmfm, mmmf, mmmm}.

2.7

1. $1020 + 835 + 620 + 548 + 400 + 350 = 3773$. así, $P(dislexia) = \frac{1020}{3773}, P(omisión1) = \frac{835}{3773}, P(omisión2) = \frac{620}{3773}, P(omisión3) = \frac{548}{3773}, P(concordancia1) = \frac{400}{3773}, P(concordancia2) = \frac{350}{3773}$.

2. $45 + 22 + 26 + 5 = 98$, así, $P(24 \leq T \leq 28) = \frac{45}{98}, P(28 \leq T) = \frac{22}{98}, P(17 \leq T \leq 24) = \frac{26}{98}, P(T \leq 17) = \frac{5}{98}$.

2.8 $\frac{5}{6}$.

2.9 $\frac{4}{8} = \frac{1}{2}$.

2.10 Total de observaciones: 100, $P(32) = \frac{8}{100} = \frac{2}{25}$.

2.11 Respuestas

1. $\frac{1}{2}$.
2. $\frac{1}{4}$.
3. $\frac{1}{8}$.
4. $\frac{1}{16}$.
5. $\frac{1}{2^n}$.

2.12 $\{(1,1),(1,2),...(1,6),(2,1),...,(2,6),...,(6,1),...,(6,6)\}$ con 36 elementos.

2.13 Respuestas

1. $\frac{4}{36} = \frac{1}{9}$.
2. $\frac{2}{36} = \frac{1}{18}$.
3. $\frac{5}{36}$.

2.14 $\frac{6}{16} = \frac{3}{8}$.

2.15 Respuestas

1. $P(E) = \frac{115}{370}$.
2. $P(F) = \frac{210}{370}$.
3. $P(E \cap F) = \frac{110}{370}$.
4. $P(E \cup F) = \frac{115}{370} + \frac{210}{370} - \frac{110}{370} = \frac{215}{370}$.

2.16 Todas son probabilidades subjetivas, la respuesta se basa en la experiencia y conocimiento personal.

2.17 $0.20 = \frac{2}{10} = \frac{1}{5}$, los momios son de 1 a 4.

2.18 $\text{P(pasar)} = \frac{4}{14} = \frac{2}{7}$.

2.19 $0.75 = \frac{75}{100} = \frac{3}{4}$, los momios son de 3 a 1.

Capítulo 3

3.1 Respuestas

1. 2^7.

2. P(7 águilas) $= \frac{1}{2^7}$.

3.2 Respuestas

1. ${}_nP_n = \frac{n!}{(n-n)!} = \frac{n!}{0!} = \frac{n!}{1} = n!$

2. ${}_nP_1 = \frac{n!}{(n-1)!} = n$.

3. ${}_nC_0 = \frac{n!}{(n-0)!0!} = \frac{n!}{n!1} = 1$.

4. ${}_nC_1 = \frac{n!}{(n-1)!1!} = \frac{n!}{(n-1)!1} = n$.

5. ${}_nC_{n-r} = \frac{n!}{(n-r)!(n-(n-r))!} = \frac{n!}{(n-r)!r!} =_n C_r$.

3.3

1. $26^3 \times 10^3$ placas posibles.

2. 10^3 con "oso".

3. P(oso en placa) $= \frac{10^3}{26^3 \times 10^3}$.

3.4 $2 \times 3 \times 2 \times 2 = 24$.

3.5

1. 9!

2. 8!

3. 7!

3.6 100!

3.7 ${}_{100}P_{50} = \frac{100!}{50!}$

3.8 7!

3.9 9!

3.10 $26 \times 62 \times 62 \times 62 \times 62 = 384\,184\,736 \simeq 3{,}842 \times 10^8$

3.11 $\frac{384\,184\,736}{10^6} = 384.18$ minutos.

3.12 Si, la probabilidad de no obtener ningún doble seis en 25 lanzamientos es $\frac{35^{25}}{36^{25}} = 0{,}494\,47$, por lo tanto la probabilidad de que al menos salga un doble seis es de $1 - 0{,}494\,47 = 0{,}505\,53$, ligeramente favorable a la casa.

3.13 ${}_{18}C_3 = \frac{18!}{15!3!}$.

3.14 ${}_{18}P_3 = \frac{18!}{15!}$.

3.15 ${}_{49}C_6 = \frac{49!}{43!6!}$.

3.16

1. Par: $13 \times_4 C_2 \times_{12} C_3 \times 4^3$.
2. Dos pares: ${}_{13}C_2 \times_4 C_2 \times_4 C_2 \times 11 \times 4$.
3. Full: $13 \times_4 C_3 \times 12 \times_4 C_2$.

3.17

1. El espacio muestral tiene 12^n elementos. Para un grupo de n personas, el evento en el cual todos ellos pertenecen a un distinto signo del Zodiaco tiene
$$ {}_{12}P_n = \frac{12!}{(12-n)!} $$
elementos. Así, la probabilidad P de que todos tengan signos diferentes es
$$ P = \frac{12!}{12^n(12-n)!} $$

El valor $1-P$ es la probabilidad de que al menos dos personas pertenezcan al mismo signo del Zodiaco. La tabla ilustra estas probabilidades para $n = 2, 3, 4$ y 5:

n	P	1-P
2	0.91667	0.08333
3	0.76389	0.23611
4	0.57292	0.42708
5	0.38194	0.61806

2. De acuerdo a la tabla, en un grupo de 4 personas, aproximadamente con un probabilidad del 43 %, al menos dos de ellas pertenecen al mismo signo del Zodiaco. ¡No se trata de una asombrosa coincidencia!

Capítulo 4

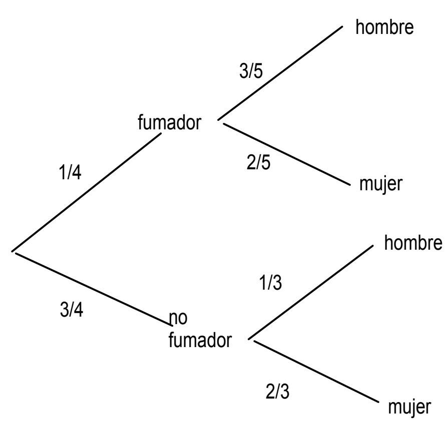
4.1 El espacio muestral reducido es {(3,1),(3,2),...,(3,6),(1,3),(2,3),...,(6,3)} con 11 elementos. El (3,1) y ((1,3) son los que suman 4 por lo que,

$$P(suma = 4 \mid \text{un número} = 3) = \frac{2}{11}.$$

4.2

1. $\frac{50}{75} = \frac{2}{3}, \frac{15}{40} = \frac{3}{8}$.
2. $\frac{10}{100} = \frac{1}{10}$.
3. $\frac{25}{100} = \frac{1}{4}$.

4.3



4.4

1. $P(A) = \frac{1}{2}$
2. $P(E \mid A) = \frac{2}{5}$
3. $P(E) = \left(\frac{2}{5} \times \frac{1}{2}\right) + \left(\frac{3}{10} \times \frac{1}{2}\right) = \frac{7}{20}$.
4. $P(A \cap E) = \frac{2}{5} \times \frac{1}{2} = \frac{1}{5}$.

5. No pues $P(E) \neq P(E \mid A)$.

4.5 $P(E) = \frac{24}{36} = \frac{2}{3}, P(F) = \frac{5}{36}, P(E \cap F) = \frac{3}{36} = \frac{1}{12} \neq P(E)P(F)$. por lo tanto no son independientes.

4.6 11 elementos del espacio muestral contienen al menos un 6, de éstos, la suma es mayor o igual a nueve para {(3,6),(6,3),(4,6),(6,4),(5,6),(6,5),(6,6)} que son 7 elementos, por lo tanto,

$$P(suma > 9 \mid \text{un número} = 6) = \frac{7}{11}.$$

4.7 Respuestas

1. Son mutuamente excluyentes.

2. Son independientes.

4.8 $P(Veracruz \mid tarántula) =$

$$\frac{P(tarántula \mid Veracruz)P(Veracruz)}{P(tarántula \mid Veracruz)P(Veracruz) + P(tarántula \mid Guerrero)P(Guerrero)}$$
$$= \frac{0.02(0.6)}{0.02(0.6) + 0.01(0.4)} = 0.75$$

La probabilidad es del 75 %.

4.9 $P(E \mid + \cap - \cap + \cap - \cap +) =$

$$\frac{P(+ \cap - \cap + \cap - \cap + \mid E)P(E)}{P(+ \cap - \cap + \cap - \cap + \mid E)P(E) + P(+ \cap - \cap + \cap - \cap + \mid E^c)P(E^c)}.$$

Sustituimos

$$\begin{aligned} P(+ \cap - \cap + \cap - \cap + \mid E) &= P(+ \mid E)P(- \mid E)P(+ \mid E)P(- \mid E)P(+ \mid E) \\ &= 0.002, \\ P(+ \cap - \cap + \cap - \cap + \mid E^c) &= P(+ \mid E^c)P(- \mid E^c)P(+ \mid E^c)P(- \mid E^c)P(+ \mid E^c) \\ &= 0.00019, \end{aligned}$$

en donde $P(+ \mid E) = 0{,}95, P(- \mid E) = 0{,}05, P(+ \mid E^c) = 0{,}06, P(- \mid E^c) = 0{,}94$ y $P(E) = 0{,}04, P(E^c) = 0{,}96$. De aquí que $P(E \mid + \cap - \cap + \cap - \cap +) =$

$$\frac{0.002 \times 0.04}{(0.002 \times 0{,}04) + (0.00019 \times 0.96)} = 0.305$$

o 30.5 %.

4.10

1.

$$\begin{aligned} P(pelirrojo \mid T_p) &= \frac{P(T_p \mid pelirrojo)P(pelirrojo)}{P(T_p \mid pelirrojo)P(pelirrojo) + P(T_p \mid \text{no pelirrojo})P(\text{ no pelirrojo})} \\ &= \frac{0.95 \times 0.02}{0.95 \times 0.02 + 0.05 \times 0.98} = 0.2794, \end{aligned}$$

o bien 27.94 %.

2.

$$\begin{aligned} P(rubio \mid T_r) &= \frac{P(T_r \mid rubio)P(rubio)}{P(T_r \mid rubio)P(rubio) + P(T_r \mid \text{no rubio})P(\text{ no rubio})} \\ &= \frac{0.95 \times 0.15}{0.95 \times 0.15 + 0.05 \times 0.85} = 0.7703, \end{aligned}$$

o bien 77.03 %.

4.11 No sería sensato pues la probabilidad de no tener la enfermedad dado que la prueba resulta positiva sería muy alta, concretamente:

$$\begin{aligned} P(sida \mid +) &= \frac{P(+ \mid sida)P(sida)}{P(+ \mid sida)P(sida) + P(+ \mid \text{ no } sida)P(\text{ no } sida)} \\ &= \frac{0.95 \times 0.01}{(0.95 \times 0.01) + (0.05 \times 0.99)} = 0.16, \end{aligned}$$

$$P(\text{no sida} \mid +) = 1 - 0.16 = 0.84,$$

es decir, con 84 % de probabilidad el individuo no tiene SIDA a pesar de haber tenido un resultado positivo en la prueba.

4.12

1. $P(menor) =$

$$\begin{aligned} &P(menor \mid hombre)p(hombre) + P(menor \mid mujer)p(mujer) \\ &= 0.1 \times 0.9 + 0.05 \times 0.1 = 0.095 \end{aligned}$$

   o el 9.5 %.

2. $P(mujer \mid menor) =$

$$\begin{aligned} &\frac{P(menor \mid mujer)P(mujer)}{P(menor)} \\ &= \frac{0.05 \times 0.1}{0.095} = 0.0526 \end{aligned}$$

   o el 5.26 %.

4.13 $P(recesión \mid desempleo) =$

$$\begin{aligned} &\frac{P(desempleo \mid recesión)P(recesión)}{P(desempleo \mid recesión)P(recesión) + P(desempleo \mid \text{expansión})P(\text{expansión})} \\ &= \frac{0.9 \times 0.75}{(0.9 \times 0.75) + (0.05 \times 0.25)} = 0.98, \end{aligned}$$

es decir el 98 % de probabilidad.

4.14 Proceso de aprendizaje Bayesiano La respuesta varía según las probabilidades subjetivas asignadas. sígase el procedimiento de la sección de aprendizaje Bayesiano.

Capítulo 5

5.1

1. Si a representa águila y s sol, el espacio muestral es

$$\{aaa, aas, asa, saa, ssa, sas, ass, sss\}.$$

La variable aleatoria X que representa el # de soles es

$$X(aaa) = 0, X(aas) = 1, X(asa) = 1, X(saa) = 1,$$
$$X(ssa) = 2, X(sas) = 2, X(ass) = 2, X(sss) = 3.$$

La distribución de probabilidad es

$$f(0) = P(X = 0) = \frac{1}{8},$$
$$f(1) = P(X = 1) = \frac{3}{8},$$
$$f(2) = P(X = 2) = \frac{3}{8},$$
$$f(3) = P(X = 3) = \frac{1}{8}.$$

2. El espacio muestral son los comités de dos personas de un grupo de cinco que pueden ser,

$${}_5C_2 = \frac{5!}{3!2!} = 10.$$

De estos diez, ${}_3C_2 = 3$ tienen dos hombres, ${}_2C_2 = 1$ tienen dos mujeres y los 6 restantes tienen un hombre y una mujer. Si el espacio muestral de comités es

$$\{c_1, c_2, \ldots, c_{10}\}$$

y de aquí los 3 primeros tienen dos hombres, c_4 tiene dos mujeres y los últimos seis tienen un hombre y una mujer, entonces la variable aleatoria X que representa el número de hombres en un comité es,

$$X(c_1) = X(c_2) = X(c_3) = 2,$$
$$X(c_4) = 0,$$
$$X(c_5) = X(c_6) = \cdots = X(c_{10}) = 1.$$

La función de distribución es,

$$\begin{aligned} f(0) &= P(X=0) = \frac{1}{10}, \\ f(1) &= P(X=1) = \frac{6}{10} = \frac{3}{5}, \\ f(2) &= P(X=2) = \frac{3}{10}. \end{aligned}$$

5.2 Con 24 lanzamientos el valor esperado del juego es

$$0.5086 \times 100 + 0.4914 \times (-100) = 1.72.$$

Con 25 lanzamientos es de,

$$0.49447 \times 100 + 0.50553 \times (-100) = -1.106.$$

5.3 La variable aleatoria X es

$$\begin{aligned} &X(1) = X(2) = X(3)) = X(4) = -30, \\ &X(5) = 50, \\ &X(6) = 100. \end{aligned}$$

La distribución de probabilidad es

$$\begin{aligned} f(-30) &= P(X=30) = \frac{4}{6} = \frac{2}{3}, \\ f(50) &= P(X=50) = \frac{1}{6}, \\ f(100) &= P(X=100) = \frac{1}{6}. \end{aligned}$$

El valor esperado es

$$E(X) = \frac{2}{3} \times (-30) + \frac{1}{6} \times 50 + \frac{1}{6} \times 100 = 5.$$

5.4 $200 \times \frac{4}{7} + (-30) \times \frac{3}{7} = \$101.\,43$ es su ganancia esperada diaria.

5.5 $E(enviar) = (10000 - 25) \times \frac{1}{40000} + (-25)\left(1 - \frac{1}{40000})\right) = -24.\,75$, $E($no enviar$) = 0$.

5.6 El espacio muestral tiene ${}_5C_2 = 10$ elementos y es:

$$\{\{1,2\},\{1,3\},\{1,4\},\{1,5\},\{2,3\},\{2,4\},\{2,5\},\{3,4\},\{3,5\},\{4,5\}\}.$$

La variable aleatoria X asociada a la suma de los números es,

$$\begin{aligned}
X(\{1,2\}) &= 3,\\
X(\{1,3\}) &= 4,\\
X(\{1,4\}) &= X(\{2,3\}) = 5,\\
X(\{1,5\}) &= X(\{2,4\}) = 6,\\
X(\{2,5\}) &= X(\{3,4\}) = 7,\\
X(\{3,5\}) &= 8,\\
X(\{4,5\}) &= 9.
\end{aligned}$$

La distribución de probabilidad,

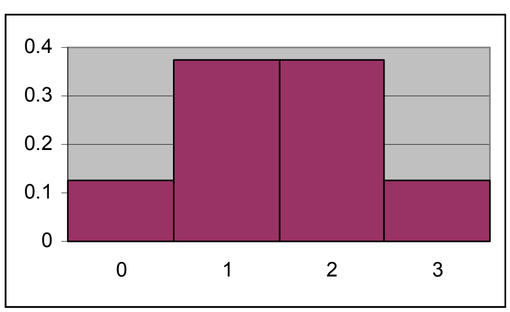
$$\begin{aligned}
f(3) = f(4) = f(8) = f(9) &= \frac{1}{10},\\
f(5) = f(6) = f(7) &= \frac{2}{10} = \frac{1}{5}.
\end{aligned}$$

El valor esperado,

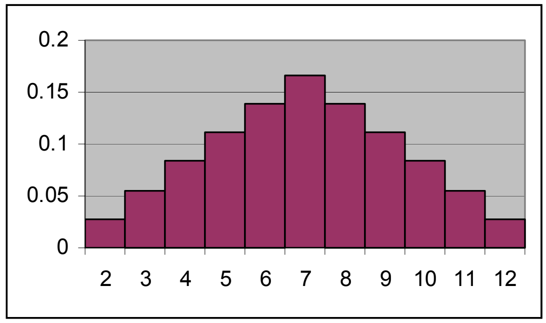
$$E(X) = (3+4+8+9)\frac{1}{10} + (5+6+7)\frac{2}{10} = 6.$$

5.7

1.



2.



5.8

1. Ganacia esperada $= \frac{1}{8000} \times (5000 - 2) + \left(1 - \frac{1}{8000}\right) \times (-2) = -1.375.$

2. El premio mínimo para salir "a mano" debe ser tal que la ganancia esperada es nula, es decir, si éste premio se denota por m, se tiene que

$$\frac{1}{8000} \times (m - 2) + \left(1 - \frac{1}{8000}\right) \times (-2) = 0,$$

de manera que $m = \$16000$.

5.9 $E(X) = \left(0.1 \times \frac{1}{4}\right) + \left(0.4 \times \frac{1}{4}\right) + \left(0.2 \times \frac{1}{4}\right) + \left(0.3 \times \frac{1}{4}\right) = 0.25.$

5.10 Precio esperado $= (25 \times 0.1) + (17 \times 0.3) + (10 \times 0.6) = 13.6.$

5.11 Demanda esperada diaria =

$$[0.05 \times (0 + 7 + 8)] + [0.1 \times (1 + 6)] + [0.15 \times (2 + 4 + 5)] + (0.2 \times 3)$$
$$= 3.7 \text{ autos.}$$

5.12 Recordando que el área del círculo es πr^2 con $r = radio$, se tiene,

$$P(verde) = \frac{\pi(0.5)^2}{\pi 2^2} = 0.0625,$$
$$P(gris) = \frac{\pi(1^2) - \pi(0.5)^2}{\pi 2^2} = 0.1875,$$
$$P(blanco) = \frac{\pi(2^2) - \pi(1)^2}{\pi 2^2} = 0.75.$$

5.13 Respuestas

1. $P(0 < X < 1) = \frac{1-0}{2-0} = \frac{1}{2}$,

2. $P(1 \leq X \leq \frac{4}{3}) = \frac{\frac{4}{3}-1}{2-0} = \frac{1}{6}$,

3. $P(\frac{2}{3} < X \leq \frac{7}{6}) = \frac{\frac{7}{6}-\frac{2}{3}}{2-0} = \frac{1}{4}$.

Capítulo 6

6.1

$$EU(calle) = -200 \times \left(\frac{1}{50}\right) + 0 \times \left(1 - \frac{1}{50}\right) = -4,$$
$$EU(estacionamiento) = -10.$$

Se estacionará en la calle pues -10 < -4.

6.2

1. La respuesta puede variar.
2. $EU(L_1) = 500, EU(L_2) = 2500 \times 0.1 + 500 \times 0.89 + 0 \times 0.01 = 695.$

6.3

1. La respuesta puede variar.
2. $$EU(L_1) = 500 \times 0.11 + 0 \times 0.89 = 55,$$
   $$EU(L_2) = 2500 \times 0.1 + 0 \times 0.9 = 250.$$

6.4

1. $$EU(A_1) = 500 \times \left(1 - \frac{1}{100}\right) + (-1000) \times \frac{1}{100} = 485,$$
   $$EU(A_2) = 400.$$

   La vida criminal le proporciona mayor utilidad.

2. Para desincentivar al Pecas podría incrementarse, por ejemplo, la probabilidad de ser atrapado, teniendo una policía y un sistema judicial más eficientes. Otra opción sería que la vida honesta le proporcionara mayor utilidad haciéndola más atractiva ya sea en términos salariales, de prestaciones sociales o algo similar.

6.5 El status quo puede darle siempre más utilidad que cualquier otra decisión. Si teme a lo desconocido, podría asignarle probabilidades muy bajas a cualquier consecuencia positiva que se deriva de alguna acción.

6.6 1150 - 1000 = \$150.

6.7

$$
\begin{aligned}
EU(A_1) &= \left(1 - \frac{1}{50}\right)(-2000) + \frac{1}{50} \times (85000 - 2000) = -300,\\
EU(A_2) &= \left(1 - \frac{1}{50}\right)(-2500) + \frac{1}{50}(100000 - 2500) = -500,\\
EU(A_3) &= \frac{-100000}{50} = -2000.
\end{aligned}
$$

La elección óptima es A_1 o comprar seguro de la aseguradora **A**.

6.8

$$
\begin{aligned}
EU(A_1) &= \left(1 - \frac{1}{50}\right)(-2000) + \frac{1}{50} \times (85000 - 2000) = -300,\\
EU(A_2) &= \left(1 - \frac{1}{50}\right)(-2250) + \frac{1}{50}(100000 - 2250) = -250,\\
EU(A_3) &= \frac{-100000}{50} = -2000.
\end{aligned}
$$

La elección óptima es A_2 o comprar el seguro de la compañía **B**.

6.9 Hay varias respuestas correctas. Por ejemplo, a los cafés no les importa el futuro de la humanidad tanto como a los verdes. O quizás los cafés valúan mucho más el presente que el futuro, o bien los cafés son igualmente concientes pero las probabilidades que asignan a la eminencia del calentamiento global no son tan altas (se asesoran de científicos distintos), o bien estiman consecuencias menos graves, etcétera.

6.10

1. E(X) = 110000 es el valor esperado del auto, tomando en cuenta que puede ser robado.

2. El valor 284.02 es la utilidad esperada o valor esperado de la utilidad:
$$U(150000) \times \frac{11}{15} + U(0) \times \frac{4}{15} = 284.02.$$

3. Su equivalente cierto es de 128.45 que es mayor a 110 por lo que este individuo prefiere tener la incertidumbre de perder su auto a comprar seguro.

4. No compra seguro y es amante del riesgo.

6.11

1. Su equivalente cierto es de 80.667 que es menor que 110. Este individuo prefiere la certidumbre de no tener que precuparse por perder su auto. Estaría dispuesto a pagar 110 - 80.667 = 29.333 o bien hasta \$29 333 de seguro.

2. Es averso al riesgo.

6.12

1. Si $U(X) = \sqrt{X}$:
$$\begin{aligned} E(X) &= 500 \times \frac{1}{3} + 1500 \times \frac{2}{3} = 1166.7, \\ U(E(X)) &= \sqrt{1166.7} = 34.157, \\ E(U(X)) &= \sqrt{500} \times \frac{1}{3} + \sqrt{1500} \times \frac{2}{3} = 33.273, \\ U(V) &= \sqrt{V} = 33.273 \Rightarrow V = 33.273^2 = 1107.1, \\ E(X) - V &= 59.6 \end{aligned}$$

2. Si $U(X) = X^2$:
$$\begin{aligned} E(X) &= 500 \times \frac{1}{3} + 1500 \times \frac{2}{3} = 1166.7, \\ U(E(X)) &= (1166.7)^2 = 1361200, \\ E(U(X)) &= 500^2 \times \frac{1}{3} + 1500^2 \times \frac{2}{3} = 1583300, \\ U(V) &= V^2 = 1583300 \Rightarrow V = 1258.3, \\ E(X) - V &= -91.6 \end{aligned}$$

Capítulo 7

7.1

1.

C / R	a	b
α	(1,-1)	(-2,2)
β	(2,-2)	$\boxed{\textbf{(-1,1)}}$

2.

C / R	a	b
α	(-1,1)	(-3,3)
β	(-2,2)	(-1,1)

No hay equilibrio de estrategias puras.

3.

C / R	a	b
α	$\boxed{\textbf{(10,10)}}$	(15,5)
β	(5,15)	(12,12)

4.

C / R	A	B
α	(0,0)	$\boxed{\textbf{(0,2)}}$
β	$\boxed{\textbf{(3,0)}}$	$\boxed{\textbf{(0,0)}}$

5.

C / R	a	b
α	$\boxed{\textbf{(30,30)}}$	(60,0)
β	(0,60)	(35,35)

7.2

Brasil / *México*	Apertura	Protección
Apertura	(10,10)	(-10,20)
Protección	(20,-10)	$\boxed{\textbf{(-5,-5)}}$

1. El equilibrio es cuando ambos protegen, se trata de un dilema del prisionero en el cual la apertura no es estable pues si el otro país cambia de estrategia y protege, se pierde.
2. Si, la protección.
3. Se necesita un mecanismo que obligue a los países a cumplir con la apertura, de esta forma ambos se benefician.
4. No, ambos países pueden beneficiarse simultáneamente.

7.3

Campos / *Rumbos*	Respetar	No respetar
Respetar	(30,30)	(-10,40)
No respetar	(40,-10)	$\boxed{\textbf{(-5,-5)}}$

1. Ninguna de las dos familias respeta y c/u obtiene -5 de pago.
2. Si, no respetar.
3. Es también un dilema del prisionero. Se requiere de un mecanismo que obligue a las familias a respetar para que ambas se beneficien.

7.4 La multa de 2 unidades por no limpiar cambia la matriz como sigue

v1 / **v2**	limpiar	no limpiar
limpiar	$\boxed{\textbf{(5, 5)}}$	(-1,4)
no limpiar	(4,-1)	(-2, -2)

Ahora el equilibrio es que ambos limpien.

7.5 Respuestas

1. Ninguno ayuda, es un dilema del prisionero.

2. Un vecino ayuda y el otro no, es un juego de gallina.

3. Si S < C uno ayuda y el otro no (gallina) pero si S > C, entonces el equilibrio es que ambos ayudan. La matriz es

Carlos \ *Ramón*	ayudar	no ayudar
ayudar	(V-C,V-C)	(V-2C,V-S)
no ayudar	(V-S,V-2C)	(0,0)

4. Se obtiene el mismo resultado que en el inciso anterior. La matriz es,

Carlos \ *Ramón*	ayudar	no ayudar
ayudar	(V-C+S,V-C+S)	(V-2C+S,V)
no ayudar	(V,V-2C+S)	(0,0)

7.6

Sony \ *Samsung*	A	B
A	$\boxed{\textbf{(3,3)}}$	(1,2)
B	(2,1)	$\boxed{\textbf{(5,5)}}$

1. Es un juego de coordinación con equilibrios en donde ambos adoptan A o B.

2. Pierde una unidad de utilidad.

3. Si, pues podrían elegir ambas la tecnología B, en lugar de la A.

7.7

1. Ambos jugadores se desvían con probabilidad $\frac{99}{101} = 0.98$ y no se desvían con probabilidad 0.02.

2. Aquí la utilidad esperada es -0.02, menor que en el equilibrio correlacionado que es de $\frac{1}{3}$.

7.8

Alce \ *Oso*	correr	esconderse
correr	$\boxed{\textbf{(60,20)}}$	(0,0)
esconderse	(0,0)	$\boxed{\textbf{(20,60)}}$

Es un juego de coordinación. Los equilibrios de estrategias puras son cuando ambos corren o ambos se esconden. Hay un equilibrio de estrategias mixtas en el cual el alce corre con probabilidad 0.75 y se esconde con probabilidad 0.25. El oso corre con probabilidad 0.25 y se esconde con probabilidad 0.75.

7.9

Rodulfo \ *Camila*	cálido	frío
cálido	(20,0)	(0,10)
frío	(0,90)	(20,0)

1. En cualquier combinación de estrategias puras, alguno de los jugadores tiene incentivo para cambiar de estrategia.

2. Rodulfo pone la temperatura cálida el 90 % del tiempo y fría el 10 %. Camila usa el termostato en cálido el 50 % del tiempo y lo mismo para frío.

3. Rodulfo tiene una utilidad esperada de 10 y Camila de 9.

7.10

Sociedad \ *Hampa*	obedecer ley	ser criminal
tolerar	(3,1)	$\boxed{\textbf{(-3,2)}}$
castigar	(-1,-1)	(-4,-3)

1. La sociedad tolera y el hampa tiene comportamiento criminal.

2. Ahora supongamos que los pagos cambian de la siguiente forma,

Sociedad \ *Hampa*	obedecer ley	ser criminal
tolerar	$\boxed{\textbf{(3,3)}}$	(-3,2)
castigar	(-1,-1)	(-4,-3)

La sociedad es tolerante y el hampa obedece la ley. Este equilibrio podría obtenerse si, en una sociedad tolerante, el hampa obtiene mayor beneficio siendo honesta que criminal.

7.11 Al igual que en el ejemplo 7.10.4 , en los nodos finales, Telcel obtiene mayor utilidad si adopta la misma plataforma que Axtel. Así, los nodos finales serán (2,2) o (1,3). Sabiendo esto, Axtel adopta java para obtener 2 unidades de utilidad. En el ejemplo 7.10.4 el resultado era adoptar .net. Se observa que la compañía que actúa primero tiene ventaja para imponer una de las tecnologías y a la otra compañía no le queda más que seguirla.

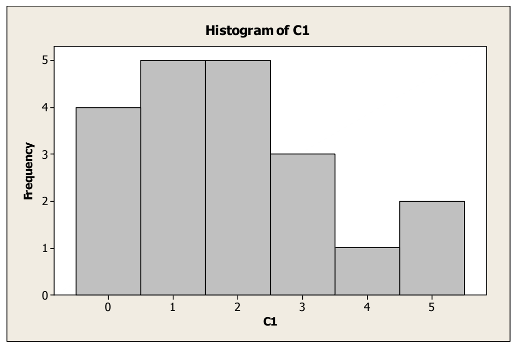
7.12 Se quiere que la utilidad de la lotería de guerra para Atenas sea menor que la utilidad que obtiene al retirarse. Es decir,

$$\begin{aligned} EUA(guerra) = q4 + (1-q)1 &\leq 2, \\ 3q + 1 &\leq 2, \\ q &\leq \frac{1}{3}. \end{aligned}$$

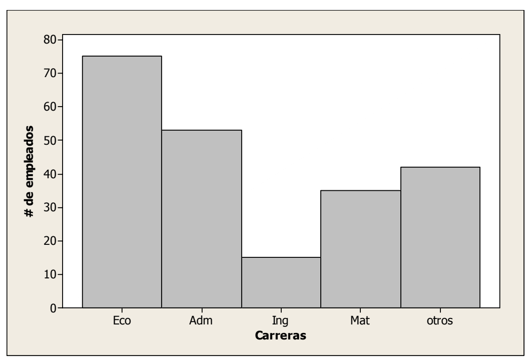
Así, cualquier probabilidad q menor a $\frac{1}{3}$, ocasionaría la retirada de Atenas.

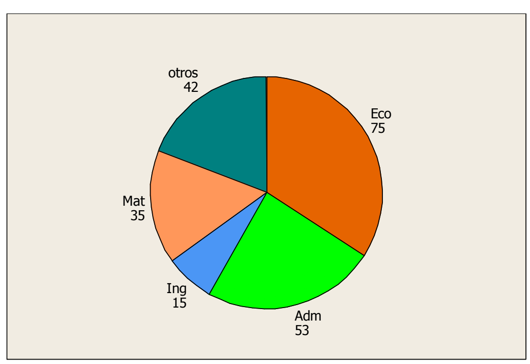
Capítulo 8

8.1

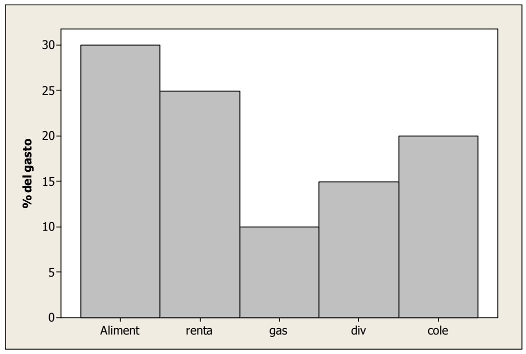


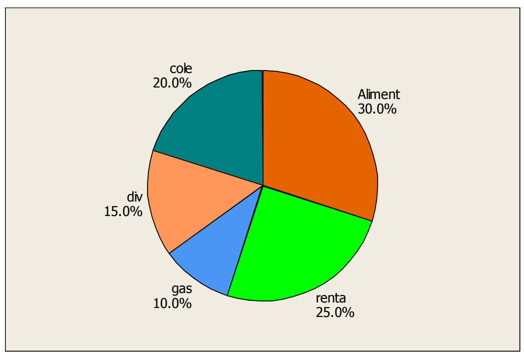
8.2





8.3





8.4 La moda, pues se trata de una característica no numérica.

8.5 La desviación estándar no puede ser negativa.

8.6

1. Media de 12.2, mediana de 12, no hay moda, rango de 20, desv. est. de 6.97.

2. Media de 27.6, mediana de 12, no hay moda, rango de 97, desv. est. de 36.47.

3. Media de 6, mediana de 7, moda de 5 y 9, rango de 8, desv est de 2.8.

4. Media 31.43, mediana 31, moda 31, rango 20, desv. est 6.3.

5. Media 129, mediana 128 modas 125, 128, rango 11, desv est 3.54.

8.7 Respuestas posibles (no hay respuesta única):

1. {60,61,62}.
2. {1,2,15,20,21}.
3. {1,4,25}.
4. {2,2,4,20,22}.
5. {1,2,7,7}.
6. {rojo,amarillo,azul,negro,azul}.
7. {1,3,24}.
8. {20,20}.
9. {18.59,21.41}.
10. {28,12}.

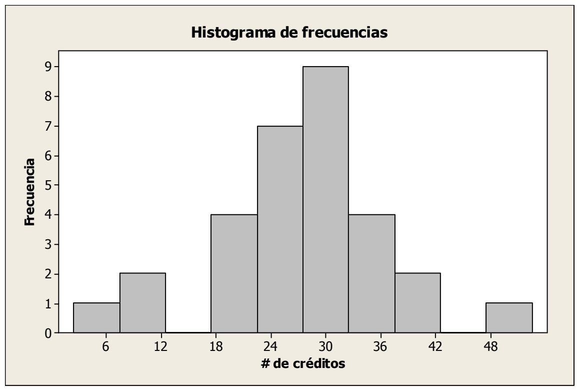
8.8 Media de 6, mediana de 5, moda de 5, desv est de 4.19.

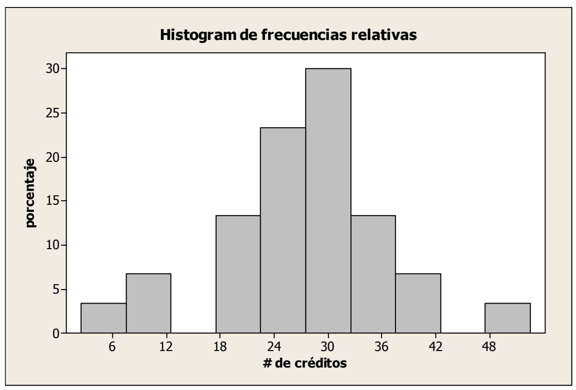
8.9

1. Media de 94, mediana de 50, modas de 0 y 50, 50 es el número que mejor representa a los datos.
2. Rango de 500 y desv est de 126.5. Hay gran dispersión.

8.10

1.





2. Realizar un diagrama de tallo y hojas.

0	6												
1	2	2	8	8	8	8							
2	4	4	4	4	4	4	4						
3	0	0	0	0	0	0	0	0	0	6	6	6	6
4	2	2	8										

3. Media 27.20, mediana 30, moda 30.

4. Rango 42, desv est 9.42.

Capítulo 9

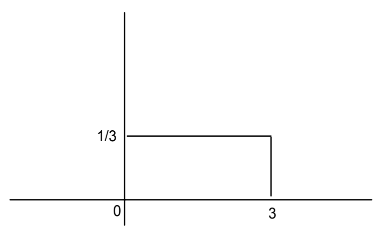
9.1 Se trata de una distribución uniforme discreta

$$\begin{aligned} P(i > 65) &= \sum_{i=66}^{100} P(i) \\ &= (100 - 66)\frac{1}{100} = \frac{34}{100} = 0.34 \end{aligned}$$

9.2

1. $P(X < 2) = \frac{2}{3}$,
2. $P(X \geq 1) = \frac{2}{3}$,
3. $P(1.5 < X \leq 3) = \frac{1.5}{3} = \frac{1}{2}$.

9.3 Dibujar la función de densidad para la variable aleatoria del ejercicio anterior.



9.4

1. $P(X < 240) = \frac{5}{10} = \frac{1}{2}$.
2. $\mu = \frac{235+245}{2} = 240$ ml.
3. $\sigma = \sqrt{\frac{(245-235)^2}{12}} = 2.887$ ml.

9.5 Dados n y π, utilizar la fórmula

$$P(X = k) =_n C_k \pi^k (1 - \pi)^{n-k}.$$

1. $P(X = 3) = 0.146\,8$.

2. $P(X = 5) = 0.102\,92$.

3. $P(X > 4) = 1 - P(X \leq 4) = 0.995$.

4. $P(X \leq 5) = 1 - P(X > 5) = 0.937\,5$.

9.6 La distribución binomial asociada tiene $n = 25, \pi = 0.6$. Por lo tanto $\mu = 25 \times 0.6 = 15$, $\sigma = \sqrt{15 \times 0{,}4} = 2.45$. La aproximación normal es $X \sim N(15, 2.45)$. Podríamos aproximar la probabilidad buscada por $P(X \geq 15) \simeq 0.5$, Sin embargo lo correcto es utilizar la corrección por continuidad. Así, la probabilidad de que enceste al menos en 15 ocasiones debe ser

$$P(X \geq 14.5) = P(Z \geq \frac{14.5 - 15}{2.45}) = P(Z \geq -0.204)$$
$$= 0.58$$

En donde Z es la normal estándar asociada.

9.7

1. Se modela con una distribución binomial con $n = 30, \pi = 0.25$. $P(X \geq 18) = 0.00005$.

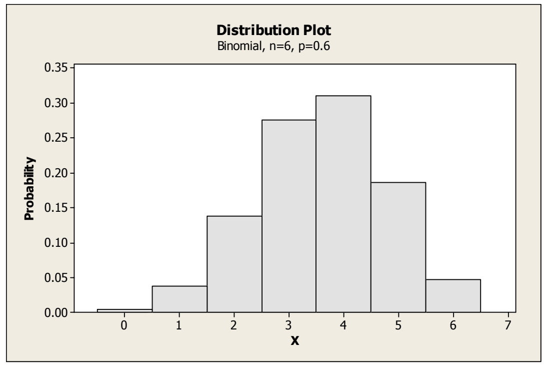
2. $\mu = 30 \times 0.25 = 7.5$, $\sigma = \sqrt{7{,}5(0{,}75)} = 2.37$.

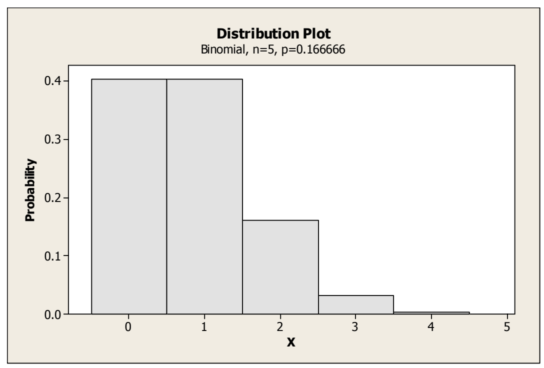
3. La aprox. normal es $X \sim N(7.5, 2.37)$, utilizando la corrección por continuidad la probabilidad buscada es

$$P(X \geq 17.5) = P(X \geq \frac{17.5 - 7.5}{2.37}) = P(Z \geq 4.22)$$
$$\simeq 0.$$

En donde Z es la distribución estándar asociada.

9.8

1. 

2. 

9.9

1. $P(0 < X < 3.1) = 0.499$.
2. $P(X < -1) = 0.158\,66$.
3. $P(X > 2) = 0.022\,75$.
4. $P(-1.1 < X < 2.2) = 0.986\,10 - (1 - 0.864\,33) = 0.850\,43$
5. $P(X \leq 1.35) = 0.871\,81$.

9.10 Z es la normal estándar asociada.

1. $P(110 < X < 180) = P(0.5 < Z < 4) = 0.308\,54$.
2. $P(50 < X < 100) = P(-2.5 < Z < 2.5) = 0.987\,6$.

3. $P(X > 3) = P(Z > 3) = 0.001\,35.$

4. $P(X < 25) = P(Z < 2.5) = 0.993\,79.$

5. $P(X > -1) = P(Z > 1) = 0.158\,66.$

9.11 La probabilidad $P(Z > 4) \simeq 0$ es muy pequeña.

9.12 $X \sim N(75, 12)$ la normal estándar asociada es Z.

1. $P(X > 60) = P(Z > -1.25) = 0.894\,35.$

2. $P(X > 90) = P(Z > 1.25) = 0.105\,65.$

9.13 $X \sim N(2.8, 0{,}5)$ la normal estándar asociada es Z.

$$P(X > 4) = P(Z > 2.4) = 0.008\,2.$$

9.14

1. La probabilidad de que más de la mitad de los votantes favorezcan al candidato naranja es bajísima. Los opositores al candidato naranja dudarían de la muestra tomada para la encuesta.

2. El equipo del candidato naranja defendería la encuesta y podría poner en duda el la media de 40 % de los votantes que le favorecen. Quizás la media real que favorece al candidato naranja no es del 40 % sino mayor.

9.15 Las respuestas dependen de los datos obtenidos.

9.16 La media $\bar{X}$ de la muestra de 25, está aproximadamente distribuida normalmente, $\bar{X} \sim N(22, \frac{5}{\sqrt{25}}) = N(22, 1)$. Entonces, si Z es la normal estándar asociada,

1. $P(X < 19) = P(Z < -3) = 0.00135.$

2. $P(X > 24) = P(Z > 2) = 0.022\,75.$

9.17 La media $\bar{X}$ de la muestra de 49, está aproximadamente distribuida normalmente, $\bar{X} \sim N(7.5, \frac{2}{\sqrt{49}}) = N(7.5, 0.2857)$, Entonces, si Z es la normal estándar asociada,

1. $P(X < 6.6) = P(Z < -3.15) = 0.00082$.

2. $P(X > 8.1) = P(Z > 2.1) = 0.01786$.

Capítulo 10

10.1

1. Realizar un experimento en el cual se gira la ruleta un buen número de veces, digamos 400-500. Hay 38 posibles resultados en una ruleta americana de manera que los números deben aparecer aproximadamente de forma uniforme con probabilidad 1/38, cada uno. Anotamos el número o números que aparecen con la frecuencia máxima. Si esta frecuencia máxima es F, las hipótesis nula y alterna son

$$H_0 : F \leq \frac{1}{38}\text{(la ruleta no está truqueada)},$$
$$H_1 : F > \frac{1}{38}\text{(la ruleta está truqueada)}.$$

2. Escoger aleatoriamente muestras de hombres y mujeres bebedores de tequila. Determinar la media mensual de mililitros de tequila que bebe cada género. Si $\bar{M}$ corresponde a la media de las mujeres y $\bar{H}$ a la de hombres, las hipóteses son:

$$H_0 : \bar{M} = \bar{H} \text{ (no existe diferencia en el consumo)},$$
$$H_1 : \bar{M} \neq \bar{H} \text{ (si existe diferencia en el consumo)}$$

3. Realizar un experimento en el cual se escoge una muestra aleatoria de vacas a las cuales se les administra la hormona. Se mide el rendimiento lechero de la muestra. Supongamos que el rendimiento medio de la población es de μ_0 y que el de la muestra es de $\bar{R}$. Las hipótesis son:

$$H_0 : \bar{R} \leq \mu_0\text{(la hormona no aumenta el rendimiento)},$$
$$H_1 : \bar{R} > \mu_0\text{(la hormona si aumenta el rendimiento)}.$$

4. Se elige una muestra aleatoria de estudiantes a los cuales se les imparte el nuevo programa educativo. Después de completar el programa, se comparan los resultados escolares (algún examen de diagnóstico, del cual conocemos la distribución de resultados) de los estudiantes sujetos

al programa con los de la población general de estudiantes. Digamos que se trata de la media obtenida en un examen. Si μ_0 es la media de la población y $\bar{X}$ la de la muestra, las hipótesis son,

$$H_0 : \bar{X} \leq \mu_0 (\text{el programa no aumenta el rendimiento}),$$
$$H_1 : \bar{X} > \mu_0 (\text{el programa si aumenta el rendimiento}).$$

5. Se recolectan promedios diarios de temperatura para ambas zonas. Si la media para la zona urbana es T_u y para la zona rural T_r, las hipótesis son

$$H_0 : T_u = T_r (\text{no existe diferencia en la temperatur}a),$$
$$H_1 : T_u \neq T_r (\text{si existe diferenciaen la temperatura}).$$

6. Se toma una muestra de personas que padecen de infecciones recurrentes en vías urinarias. A una submuestra aleatoria de éstas, se les administra jugo de arándano todos los días. Después de algún periodo de tiempo se comparan el número promedio de infecciones que presentaron los que tomaron jugo y los que no tomaron. Digamos que éstas medias son, N_a, y N. Las hipótesis son,

$$H_0 : N_a \geq N \text{ (el jugo no disminuye el \# de infecciones)},$$
$$H_1 : N_a < N \text{ (el jugo si disminuye el \# de infecciones)}.$$

7. Se toma una muestra aleatoria de personas y se les pregunta si favorecen o no la política. Si $\bar{P}$ es la proporción que muestra una opinión favorable. Digamos que la política se considera popular si es favorecida por más de la mitad de la población. Las hipótesis son,

$$H_0 : \bar{P} \leq 0.5 \text{ (la política no es popular)},$$
$$H_1 : \bar{P} > 0.5 \text{ (la política si es popular)}.$$

8. Se toma una muestra de focos del productor en cuestion. Se prueban y se determina la proporción de focos que duran menos de 10000 horas,

digamos que ésta es $1-p$. La hipótesis son,

$$H_0 : p \geq 96\,\% \text{ (al menos el } 96\,\% \text{ de los focos duran 10000 horas),}$$
$$H_1 : p < 96\,\% \text{ (menos del } 96\,\% \text{ de los focos duran 10000 horas).}$$

9. Se toma una muestra de personas, digamos 10 mismas que se le presentan a la astróloga. Se anotan las predicciones de la astróloga, digamos que acertó en k ocasiones (éxitos). Se trata de un experimento binomial con $n = 10$ y probabilidad de éxito de $\pi = \frac{1}{12}$, pues hay 12 signos del zodiaco. Las hipótesis, en términos de π son,

$$H_0 : \pi \leq \frac{1}{12} \text{ (la astróloga no tiene poderes),}$$
$$H_1 : \pi > \frac{1}{12} \text{ (la astróloga si tiene poderes).}$$

10. Se toma una muestra de solicitudes de trabajo para la dependencia y se separan en dos grupos, mayores de 40 años y menores de 40 años. Se estiman los porcentajes de solicitudes exitosas (las que resultaron en contratación) para cada grupo. Digamos que éstos son P_{40} y P. Las hipótesis son,

$$H_0 : P_{40} = P \text{ (no existe discriminación),}$$
$$H_1 : P_{40} \neq P \text{ (si existe discriminación).}$$

10.2

1. La hipótesis nula es $I = 14200$, en donde I es el ingreso promedio mensual, la hipótesis alterna es $I \neq 14200$. Digamos que X es la variable aleatoria que corresponde a los datos de la muestra con desviación estándar de $s = \frac{2600}{\sqrt{75}} = 300.\,22$ y media 14200. Sea Z la normal estándar asociada. El valor p es,

$$P(X > 15300) =$$
$$P\left(Z > \frac{15300 - 14200}{300.22}\right) = P(Z > 3.66) \simeq 0.0002$$

que evidentemente es menor que $\frac{0{,}05}{2} = 0.025$, por lo que se determina que el ingreso ha cambiado.

2. La hipótesis nula es ahora $I \leq 14200$ y la alterna $I > 14200$. El valor p es el mismo que acabamos de obtener, por lo que podemos concluir que el ingreso ha aumentado.

3. La hipótesis nula es ahora $I \geq 14200$ y la alterna $I < 14200$. El valor p es

$$P(X < 15300) = 1 - 0.0002 = 0.9998,$$

con lo cual evidentemente no podemos concluir que el salario disminuyó.

10.3 Sea T el tiempo promedio de procesamiento. La hipótesis nula es $T \geq 12.3$ y la alterna $T < 12.3$. Los tiempos de la muestra siguen una distribución X con media 12.3 y desviación estándar de $\frac{3{,}5}{\sqrt{100}} = 0.35$. Si Z es la normal estándar asociada, el valor p es

$$\begin{aligned} P(X < 10.9) &= P\left(Z < \frac{10.9 - 12.3}{0.35}\right) \\ &= P(Z < -4) \simeq 0, \end{aligned}$$

de manera que podemos rechazar la hipótesis nula y concluir que el nuevo procedimiento reduce el tiempo de procesamiento.

10.4 Supongamos que la hipótesis nula es,

$$H_0 : \text{“el sujeto no ingirió más del límite legal de alcohol”},$$

la probabilidad de error tipo I es aquella de rechazar H_0 siendo que es verdadera, es decir, cuando el sujeto no sobrepasa el límite legal de alcohol, pero, sin embargo el alcoholímetro proporciona un falso positivo. Según los datos ésta es $\alpha = \frac{5}{1000} = 0.005$. La probabilidad de error tipo II es la aquella de aceptar H_0, siendo que es falsa, es decir, el sujeto ingirió más del límite legal, pero el alcoholímetro no lo detecta, o bien un falso negativo. En este caso éste es de $\beta = \frac{50}{1000} = 0.05$.

10.5

1. La distribución binomial tiene probabilidad de éxito de 0.4, $n = 5 = k$. De aquí la probabilidad de error tipo I es

$$\alpha = P(k = 5 \mid \pi = 0.4) =_5 C_5 (0.4)^5 (0.6)^0 = 0.01024.$$

Para la probabilidad de error tipo II, H_0 es falsa de manera que la probabilidad de éxito es 0.6 y para aceptar H_0 tendríamos que obtener 4 canicas o menos, entonces,

$$\begin{aligned}\beta &= P(k \leq 4 \mid \pi = 0.6) = 1 - P(k = 5 \mid \pi = 0.6)\\ &= 1 -_5 C_5 (0.6)^5 (0.4)^0 = 0.9222.\end{aligned}$$

El poder de la prueba es de,

$$1 - \beta = 0.0778.$$

2. Ahora

$$\begin{aligned}\alpha &= P(k \geq 4 \mid \pi = 0.4)\\ &= P(k = 4 \mid \pi = 0.4) + P(k = 5 \mid \pi = 0.4)\\ &=_5 C_4 (0.4)^4 (0.6)^1 + 0.01024\\ &= 0.0768 + 0.01024 = 0.08704,\\ \beta &= P(k < 4 \mid \pi = 0.6) = 1 - P(k \geq 4 \mid \pi = 0.6)\\ &= 1 - (P(k = 4 \mid \pi = 0.6) + P(k = 5 \mid \pi = 0.6))\\ &= 1 - (0.2592 + 0.0778) = 0.663\\ \textit{poder} &= 1 - \beta = 0.337.\end{aligned}$$

3. α aumenta a costa de β.

4. Si nunca rechazamos H_0, entonces $\alpha = 0$, si siempre rechazamos H_0, entonces $\beta = 0$.

10.6 ¡Por supuesto que no debe publicarse! Al repetir 200 veces el experimento, eventos de probabilidad 0.01 seguramente aparecerán y en realidad no serían significativos. Deberían reportarse también los resultados no significativos, si es que se quiere aportar algo que ayude a la ciencia.

10.7 El intervalo aproximado es de dos desviaciones estándar a la derecha y a la izquierda de la media:

1. Entre 16 % y 28 %.
2. Entre 12 % y 24 %.
3. Entre 23 % y 35 %.

10.8 Son probabilidades subjetivas de manera que la respuesta varía. Podría sugerirse en confiar más en la aseveración de un ingeniero experto en la materia que en la de un político, pero esto sigue siendo subjetivo.

10.9

1. $\frac{P(H_0)}{P(H_1)} = \frac{6}{1}, \frac{P(H_0)}{P(H_2)} = \frac{6}{3} = \frac{2}{1}, \frac{P(H_2)}{P(H_1)} = \frac{3}{1}$.
2. $$\begin{aligned} P(A \text{ es el padre} \mid E) &= 0, \\ P(B \text{ es el padre} \mid E) &= \frac{1 \times 0.1}{0 + 1 \times 0.1 + 0.1 \times 0.3} \simeq 0.77, \\ P(X \text{ es el padre} \mid E) &= 1 - 0.77 = 0.23. \end{aligned}$$
3. $P(H_0) = 0, \frac{P(H_2)}{P(H_1)} = \frac{0.23}{0.77} = 0.298\,7$ o bien $\frac{P(H_1)}{P(H_2)} \simeq 3.35$.
4. $$\begin{aligned} P(A \text{ es el padre} \mid E) &= 0, \\ P(B \text{ es el padre} \mid E) &= \frac{1 \times 0.05}{0 + 1 \times 0.05 + 0.1 \times 0.75} \simeq 0.4, \\ P(X \text{ es el padre} \mid E) &= 1 - 0.4 = 0.6. \end{aligned}$$

   de aquí que se sigue teniendo $P(H_0) = 0$, pero $\frac{P(H_2)}{P(H_1)} = \frac{0.6}{0.4} = \frac{3}{2}$ de manera que ahora los momios son 3 a 2 a favor de que el padre sea el caballero X y la custodia pordría otorgársele al caballero A.

Capítulo 11

11.1 No parece existir correlación alguna entre el % de facultad de tiempo completo y el porcentaje de graduación. Se observa una correlación positiva entre el puntaje del exámen de admisión y el % de graduación, sugiriendo que si entran buenos alumnos, es más probable que éstos se gradúen.

11.2 A pesar que hay una correlación positiva entre ellas, se trata de una correlación espuria. Ambas variables dependen de la eficiencia y potencia del motor del auto.

11.3

1. Rendimiento $= 35.1 - 0.08 \times 240 = 15.9$ millas por galón.
2. Ambos coeficientes son significativamente distintos de cero, el signo negativo para el coeficiente de la potencia indica que a mayor potencia, menor rendimiento. El valor de $R-sq$ indica que el 53.1 de la variación en el rendimiento puede explicarse por la potencia del motor.

11.4 El uso del casco ayudó a disminuir los accidentes ciclistas. El alto valor de $R - sq$ (86.5) indica que este porcentaje de la disminución en los accidentes puede ser atribuido a la utilización generalizada del casco.

11.5

1. $R = \sqrt{R - sq} = \sqrt{0.141} = 0.3755 = 37.55\,\% =$ correlación.
2. GMAT = 321.3.
3. No es tan buen predictor pues sólo explica el 14.1 % de la variación en el resultado del GMAT.

11.6 Parecería que los hombres (G = 1) obtienen en promedio 14.33 puntos más, sin embargo, la desviación estándar es altísima (20.41) y el coeficiente no podemos rechazar la hipótesis de que el efecto del género es nulo pues el valor p es muy alto (0.485). No es significativo.

11.7

1. El coeficiente de A no cambia mucho pero los de I y G aumentan sustancialmente, así como el término constante. La razón es que al omitir la educación, se magnifica el efecto de la contribución de las otras variables. La antigüedad no cambia mucho pues seguramente no tiene gran correlación con la educación.
2. Es mucho menor pues se omitió un predictor importante.
3. Pueden ser muy pocos y no explicar suficientemente la variación o demasiados y ocasionar multicolinearidad.

11.8 ($ está en millones de dólares)

1. El coeficiente es significativo, aunque pequeño, $1 de costo incrementaría aproximadamente en $1 las ganancias. El costo explica el 39.9 % de la variación en las ganancias. El término constante no es significativamente diferente de cero ($p = 0.142$).
2. El coeficiente de E es significativamente diferente de cero ($p = 0.002$), se reduce el efecto del costo ($0.8591 por $1 de costo), aunque sigue siendo significativo. Se explica el 44.4 % de la variación en las ganancias con estos dos predictores. La constante sigue sin ser significativamente diferente de cero.
3. Al incluir el número de pantallas se reduce a $0.3747 la contribución marginal del costo y su nivel de significancia es mucho menor ($p = 0.011$). También se disminuye el efecto de las estrellas y disminuye ligeramente su nivel de significancia. El número de pantallas contribuye poco ($0.0195 por pantalla) pero es significativo ($p < 0.001$). Estos tres predictores predicen el 53.6 % de la variación en las ganancias. La constante continúa siendo no significativamente diferente de cero.
4. El costo ahora parece tener una contribución negativa (−$0.0886) pequeña, pero no significativa pues $p = 0.384$. Lo mismo sucede con el número de pantallas. La contribución de la presencia de estrellas es

menor pero sigue siendo significativa. Lo sobresaliente es que $R - sq$ aumenta significativamente de manera que ahora los predictores explican el 80.3 % de la variación.

5. Excepto por el término constante, los coeficientes del ingreso en el primer fin de semana y la presencia de estrellas son ambos positivos y significativos. Estas dos variables explican el 84.9 % de la variación en las ganancias, de manera que el ajuste es mejor que antes. Seguramente hay una correlación alta entre I, N y C de manera que omitiendo N y C logramos un mejor ajuste.