Introducción a la Estadística descriptiva

Toponautas

ÍNDICE GENERAL

CAPÍTULO 1

INTRODUCCIÓN A LA ESTADÍSTICA

1.1. ¿Qué es la estadística?

La estadística es la ciencia que se encarga de recolectar, describir e interpretar datos. Esta es una definición sencilla, pero bastante precisa. Ahora,

¿por qué nos interesa estudiar estadística?

En el mundo ocurren muchos fenómenos que no podemos predecir con exactitud. Por ejemplo, no podemos decir el día exacto en que habrá un huracán que pase por Yucatán. Sin embargo, que no podamos hacer predicciones exactas no significa que no podamos hacer buenas estimaciones. Aunque no sepamos la fecha exacta del próximo huracán en Yucatán, sí podemos inferir un rango de tiempo en el que es más probable que ocurra. Por ejemplo, podemos decir que existe una alta probabilidad de que los huracanes ocurran entre agosto y octubre.

¿Con qué tanta certeza sabemos esto?

Bueno, no nos adelantemos: eso lo mediremos más adelante con la probabilidad. Por ahora, demos un paso atrás y preguntémonos lo siguiente:

¿cómo sabemos que entre agosto y octubre hay más probabilidad de que ocurra un huracán?
¿Por qué no entre enero y marzo?
¿Estamos seguros de que en esos otros meses no puede ocurrir ningún huracán en Yucatán?

Observación 1

Estas estimaciones no se hacen al azar. Se realizan a partir de la observación de datos históricos como fechas de huracanes, temporadas, frecuencia, intensidad, etcétera.

El primer paso es entonces **recolectar datos**. Una vez que se tiene esta información, el siguiente paso es **organizarla y describirla**, es decir, eliminar datos irrelevantes, resumir la información y presentarla de forma clara, por ejemplo mediante tablas o gráficas.

Con los datos ya bien descritos, ahora sí podemos analizarlos, buscar patrones y, con ayuda de la probabilidad, extraer conclusiones sobre el fenómeno estudiado.

Todo este proceso **recolectar datos, describirlos e interpretarlos** es justamente lo que llamamos estadística. Gracias a la estadística podemos tener cierta certeza sobre en qué meses del año es más probable

que ocurran huracanes y en cuáles no.

La estadística se utiliza también para estudiar fenómenos en economía, sociología, biología, física y muchas otras áreas. Por eso decimos que la estadística es una herramienta universal.

1.2. Etapas de la estadística

Muchas personas suelen confundir que existen dos tipos de estadística con la idea de que se trata de una clasificación rígida o incluso que son dos trabajos completamente separados, donde distintas personas se encargan de cada uno. En realidad, a la estadística no la dividimos como tal en tipos, sino en **etapas**: la etapa descriptiva y la etapa inferencial.

1.2.1. Estadística descriptiva

Cuando estudiamos un fenómeno mediante la estadística, primero recolectamos datos, los organizamos, los limpiamos y los presentamos de forma clara. A esta primera etapa se le conoce como **estadística descriptiva**.

Observación 2

En la estadística descriptiva no sacamos conclusiones ni hacemos predicciones; simplemente describimos lo que dicen los datos. Tablas, gráficas, promedios, medianas y medidas de dispersión pertenecen a esta parte.

1.2.2. Estadística inferencial

Una vez que los datos están bien descritos —porque alguien hizo el trabajo pesado— entramos a la segunda etapa: la **estadística inferencial**.

Aquí es donde, basándonos en los datos y en sus descripciones, inferimos conclusiones sobre el fenómeno que estamos estudiando.

Ejemplo 1

Supongamos que tenemos una lobina negra y queremos estimar su edad. Evidentemente, no tenemos un acta de nacimiento de cada pez de este tipo, pero aun así podemos usar estadística para inferirla. Primero, recolectamos información relevante sobre lobinas negras: edad observada, longitud del pez, peso, entre otras variables. Luego revisamos qué información es útil y cuál no, organizamos los datos y los presentamos de forma clara.
Por ejemplo, podríamos obtener una tabla que relacione la edad del pez con su longitud:

- 1 año $\rightarrow$ 18 cm
- 2 años $\rightarrow$ 26 cm
- 3 años $\rightarrow$ 34 cm
- 4 años $\rightarrow$ 41 cm

Todo este proceso corresponde únicamente a estadística descriptiva.

Observación 3

Al analizar estos datos, podemos observar patrones, por ejemplo: a mayor longitud del pez, mayor es su edad. Con base en esta relación, podemos inferir que si una lobina negra mide, por ejemplo, 30 cm, entonces su edad probable es de alrededor de 2 a 3 años. Esta es la parte inferencial del proceso.

Entonces, en la estadística descriptiva surgen preguntas como: *¿qué información es relevante?, ¿cómo conviene resumir los datos?, ¿qué medidas son adecuadas?*

En la estadística inferencial surgen preguntas más delicadas: *¿qué tan confiables son nuestras conclusiones?, ¿cómo medimos la incertidumbre?, ¿con qué nivel de certeza podemos afirmar algo?*

Observación 4

La estadística descriptiva y la estadística inferencial no son mundos separados, sino partes complementarias de un mismo proceso.

1.3. Población y muestra

Cuando estudiamos un fenómeno usando estadística, surge una pregunta fundamental: **¿a quién o a qué le vamos a hacer el estudio?**

1.3.1. Población

Cuando somos capaces de responder con claridad **¿a quién o a qué estamos estudiando?**, entonces podemos hablar de **población**.

Ejemplo 2

Supongamos que queremos estudiar las enfermedades más recurrentes entre los enfermeros de México. Aquí es importante delimitar no solo al sujeto del estudio —los enfermeros—, sino también su contexto —México—.

Definición 1 Población

Es el conjunto o colección de individuos, objetos o eventos cuyas características o propiedades serán analizadas en un estudio.

1.3.2. Muestra

Ahora supongamos que ya hemos delimitado nuestra población: los enfermeros de México. La siguiente pregunta natural es: **¿vamos a obtener información de absolutamente todos los enfermeros de México?**

La respuesta es no. Hacerlo sería extremadamente costoso, tardado y poco realista. Además, la población puede cambiar con el tiempo. Por eso, en la práctica, se estudia solo una parte representativa de ella.

A este subconjunto de la población se le llama **muestra**.

Definición 2 Muestra

Es un grupo más pequeño de individuos seleccionado con cuidado, de tal forma que represente adecuadamente a la población.

Ejemplo 3

En lugar de estudiar a todos los enfermeros de México, podríamos estudiar a unos 4,000 enfermeros bien seleccionados.

Observación 5 Cuidado

Pero... *¿Qué significa "bien seleccionados"?.*

En la practica, no cualquier muestra sirve o es representativa de la población. Por ejemplo, si solo incluimos enfermeros de Saltillo, esta muestra no sería representativa de todo México. Recordemos que México es un país sumamente diverso, tanto en cultura como en clima, demografía, y otras características. Así que, una opción más fiable de muestra para nuestro ejemplo podría ser seleccionar 4,000 enfermeros distribuidos entre distintos estados o regiones de México.

CAPÍTULO 2

VARIABLES, VALORES Y DATOS

Ejemplo 4

Supongamos que se nos solicita realizar un estudio sobre los estudiantes de la Universidad Nacional Autónoma de México (UNAM). Por el momento no interesa el objetivo específico del estudio; lo que sí se nos pide es registrar determinadas características de los alumnos, tales como su edad, estatura, color de cabello o si disfrutan estudiar en la UNAM.

Observa que todas estas son características se refieren a aspectos observables de cada estudiante. A cada una de estas características se le denomina como **variable**. Así, la edad constituye una variable, al igual que la estatura, el color de cabello o el hecho de que un alumno manifieste agrado o desagrado por estudiar en la institución.

En los siguientes apartados se describirán los distintos tipos de variables que existen en estadística. Por ahora basta dar una definición de lo que es una variable.

Definición 3 Variable

Una variable es una característica que podemos observar o medir en cada sujeto de una muestra.

Notemos, además, que no todos los estudiantes poseen la misma edad, ni la misma estatura, ni el mismo color de cabello. Asimismo, no todos manifiestan la misma opinión respecto a estudiar en la UNAM. Esto implica que, para cada individuo de la muestra, cada variable puede adoptar distintos resultados.

Ejemplo 5

Un alumno podría medir 1,70 metros, mientras que otro puede medir 2 metros.

Entonces, a cada resultado asociado a un individuo se le denomina **valor de la variable**. El conjunto formado por todos los valores obtenidos de los distintos sujetos de estudio recibe el nombre de **datos**.

Ejemplo 6

Consideremos la variable "edad". Si se consultan tres alumnos y se obtienen los valores 18, 20 y 19 años, cada uno corresponde a un estudiante particular y, en conjunto, constituyen parte de los datos de la investigación.

Ejemplo 7

Si ahora se atiende a otra variable, por ejemplo "color de cabello", podrían registrarse valores como "café", "rubio" y "negro". Nuevamente, cada valor corresponde a un individuo y todos en conjunto conforman datos.

Entonces:

- **Variable:** Es una característica que deseamos medir u observar en los individuos.
- **Valor:** Es un resultado concreto que adopta la variable en cada individuo.
- **Datos:** Es el conjunto de todos los valores registrados.

2.1. Tipos de variables

Cada una de estas categorías admite subclasificaciones, las cuales se presentan a continuación.

2.1.1. Variables cualitativas

Aunque es posible clasificar las variables de múltiples maneras, la distinción inicial más utilizada en estadística divide a las variables en dos grandes grupos: **cualitativas** y **cuantitativas**.

Definición 4 Variables cualitativas

Las variables **cualitativas** son aquellas que describen cualidades o categorías.

En otras palabras, nos indican "qué es" una característica, y no "cuánto es".

Ejemplo 8

El género, la nacionalidad, el color de cabello o el hecho de consumir determinado producto.

Las variables cualitativas se subdividen en dos tipos:

- **Ordinales:** Cuando sus categorías poseen un **orden inherente**. Por ejemplo, en una escala de satisfacción del tipo *malo – regular – bueno*, es evidente que existe una gradación entre las categorías.
- **Nominales:** Cuando sus categorías no admiten un orden natural. Por ejemplo, en la variable "color de cabello" no tendría sentido establecer una jerarquía entre los valores "negro", "rubio" o "castaño".

Ejemplo 9

- **Ordinales:** En una escala de satisfacción del tipo *malo – regular – bueno*, es evidente que existe una gradación entre las categorías.
- **Nominales:** En la variable "color de cabello" no tendría sentido establecer una jerarquía entre los valores "negro", "rubio" o "castaño".

2.1.2. Variables cuantitativas

Por el contrario, las variables **cuantitativas** se expresan numéricamente y están relacionadas con cantidades medibles.

Definición 5 Variable cuantitativa

Una variable cuantitativa es una característica medible que se expresa numéricamente, permitiendo realizar operaciones matemáticas como sumar o promediar

Las variables cuantitativas pueden ser:

- **Discretas:** Cuando sus valores son *contables*, es decir, cuando puede enumerarse cada posible resultado. Esta situación es análoga a subir una escalera: se pasa de un valor a otro sin puntos intermedios.
- **Continuas:** Una variable continua puede tomar *infinitos valores dentro de un intervalo*, es decir, entre dos valores cualesquiera siempre es posible hallar un valor intermedio.

Ejemplo 10

- **Discretas:** Número de hijos, edad, número de empleados.
- **Continuas:** Estatura, peso, temperatura.

CAPÍTULO 3

DATOS AGRUPADOS

Cuando el tamaño de un conjunto de datos es muy grande, o bien cuando presenta valores muy dispersos, puede resultar poco práctico analizarlo considerando cada observación individual. En estos casos es conveniente cambiar el formato de presentación, agrupando los datos en **intervalos** o **clases**. De esta forma podemos facilitarnos la visualización del comportamiento general de la variable y nos será más fácil identificar patrones como concentraciones, asimetrías o valores atípicos.

3.1. ¿Cómo agrupar datos?

Para realizar este proceso debemos considerar que, en la mayoría de los casos, se agrupan **datos cuantitativos**. Supongamos que hemos obtenido los siguientes valores de la variable `edad` en una encuesta aplicada a un grupo de personas:

`32, 27, 41, 23, 38, 29, 45, 21, 36, 44, 25, 33, 40, 28, 42`

El primer paso consiste en **delimitar los extremos** del conjunto, es decir:

$$X_{\text{mín}} = \text{valor mínimo}, \qquad X_{\text{máx}} = \text{valor máximo}.$$

En este caso obtenemos:

$$X_{\text{mín}} = 21, \qquad X_{\text{máx}} = 45$$

3.1.1. Rango

Con ello podemos calcular el **rango** R mediante:

$$R = X_{\text{máx}} - X_{\text{mín}} = 45 - 21 = 24$$

3.1.2. Número de intervalos o clases

El siguiente paso es determinar cuántos intervalos o clases tendrá nuestra agrupación. Este número se denota por k y se conoce como **número de clases**. Una forma habitual de estimarlo es utilizando la raíz cuadrada del tamaño de la muestra:

$$k = \sqrt{n},$$

donde n es el número total de datos.

En nuestro ejemplo:

$$k = \sqrt{15} \approx 3{,}87 \approx 4 \quad \text{(redondeando)}.$$

Observación 6

La elección del número de clases k no es única y puede variar según el criterio del investigador o la naturaleza de los datos. Además de la regla de la raíz cuadrada, existen otras reglas:

- **Regla de Sturges:** $k = 1 + 3{,}322 \log_{10}(n)$, recomendada para distribuciones aproximadamente normales.
- **Regla de Rice:** $k = 2n^{1/3}$, que proporciona un número intermedio entre otras reglas.
- **Regla de Freedman-Diaconis:** $k = \frac{R}{2 \cdot \text{IQR} \cdot n^{-1/3}}$, donde R es el rango e IQR es el rango intercuartílico; bastante útil cuando hay valores atípicos.

En la práctica, es común probar con diferentes valores de k y elegir aquel que produzca una distribución más clara y significativa, recordando que muy pocas clases pueden ocultar patrones importantes, mientras que demasiadas pueden generar una representación fragmentada.

3.1.3. Amplitud de los intervalos

Conociendo k, podemos calcular la **amplitud** A de cada intervalo (el tamaño de cada clase) mediante la fórmula:

$$A = \frac{X_{\text{máx}} - X_{\text{mín}}}{k}$$

Sustituyendo valores:

$$A = \frac{45 - 21}{4} = \frac{24}{4} = 6$$

Esto significa que cada intervalo cubrirá un rango de 6 unidades.

3.1.4. Armando los intervalos

Para construir las clases:

- El primer intervalo inicia en $X_{\text{mín}}$.
- Su extremo derecho se obtiene sumando la amplitud.

Así, el primer intervalo es:

$$[21, 21 + 6) = [21, 27)$$

El siguiente intervalo comienza donde terminó el anterior:

$$[27, 27 + 6) = [27, 33)$$

Continuando este procedimiento obtenemos los intervalos:

$$[21, 27)$$
$$[27, 33)$$
$$[33, 39)$$
$$[39, 45]$$

Observación 7

Es habitual que el último intervalo **incluya** su extremo derecho para asegurar que todos los datos queden clasificados.

3.1.5. Frecuencia de clase

Una vez definidos los intervalos, el siguiente paso consiste en **contar cuántos datos caen en cada clase**. Por ejemplo, en el primer intervalo $[21, 27)$ se encuentran los valores:

$$21,\ 23,\ 25,$$

correspondientes a 3 datos.

Realizando el mismo proceso para todos los intervalos obtenemos:

$$[21, 27) \to 3 \text{ datos}$$
$$[27, 33) \to 4 \text{ datos}$$
$$[33, 39) \to 3 \text{ datos}$$
$$[39, 45] \to 5 \text{ datos}$$

Cada uno de estos números se conoce como **frecuencia de clase** y se denota por f_i, con $i = 1, \ldots, k$. Es decir:

$$f_1 = 3, \quad f_2 = 4, \quad f_3 = 3, \quad f_4 = 5.$$

Observación 8

Para comprobar que la agrupación se realizó correctamente, se puede verificar que:

$$\sum_{i=1}^{k} f_i = n$$

Si esta igualdad se cumple, entonces todos los datos han sido clasificados correctamente en sus respectivas clases.

3.1.6. Marcas de clase

Una vez que hayamos agrupado nuestros datos, podríamos olvidar los valores originales (los datos sin agrupar). Esto tiene una clara ventaja visual: podemos trabajar con intervalos en lugar de con decenas o cientos de observaciones. Sin embargo, también representa un riesgo conceptual. Surge la pregunta:

¿Qué representa cada intervalo?

Al olvidar cada dato individual, estamos perdiendo parte de la esencia de nuestra información. Entonces, para que cada intervalo tenga un significado, debemos asignarle un "vocero" que lo represente: a esta representación se le llama **marca de clase**.

Definición 6 Marca de clase

La **marca de clase** de un intervalo es un valor representativo que refleja el centro de los datos contenidos en ese intervalo.

Para calcular la marca de clase c_i de un intervalo i, usamos la fórmula:

$$c_i = \frac{\text{Límite inferior}_i + \text{Límite superior}_i}{2}$$

donde:

- **Límite inferior:** extremo izquierdo del intervalo.
- **Límite superior:** extremo derecho del intervalo.
- i: índice del intervalo, con $i = 1, \ldots, k$ y k el número total de intervalos.

Ejemplo 11

Usando los intervalos de nuestro ejemplo anterior:

- Para el primer intervalo $[21, 27)$:

$$c_1 = \frac{21 + 27}{2} = \frac{48}{2} = 24$$

- Para el segundo intervalo $[27, 33)$:

$$c_2 = \frac{27 + 33}{2} = \frac{60}{2} = 30$$

Estos números, 24 y 30, son las marcas de clase de los intervalos y representan el centro de los datos agrupados en cada uno.

Observación 9

Más adelante veremos que la marca de clase se utiliza para calcular estimadores como la **media de los datos agrupados**.

CAPÍTULO 4

FRECUENCIAS

El término **frecuencia** hace referencia al número de repeticiones por unidad de tiempo de cualquier proceso periódico. En estadística, cuando obtenemos una serie de observaciones de una variable, habrá ocasiones en que un determinado valor aparezca repetidas veces. A estos los llamaremos **datos repetidos**. Una de las primeras herramientas con las que trabajaremos será medir la frecuencia con la que aparecen dichos datos repetidos. En este caso, la frecuencia es un número entero que nos indica cuántas veces se repite un valor de este tipo, dentro del conjunto de datos.

Observación 10

Para efectos prácticos, utilizaremos el siguiente conjunto de datos para ejemplificar todas las secciones de este capítulo:

$$7,\ 12,\ 5,\ 7,\ 18,\ 12,\ 9,\ 23,\ 5,\ 14,\ 12,\ 7,\ 20,\ 9,\ 5,\ 1$$

4.1. Frecuencia absoluta

La frecuencia absoluta de una observación en un conjunto de datos es exactamente el número de veces que se repite dicha observación.

Definición 7 Frecuencia absoluta

Dado un conjunto de observaciones $x_1, x_2, \ldots, x_n$ donde algún x_i se repite $k \in \mathbb{N}$ veces, la frecuencia absoluta de la observación x_i es $f_i = k$.

Ejemplo 12

Dado el conjunto del capítulo, las frecuencias absolutas son:

Dato	**Frecuencia** f_i
1	1
5	3
7	3
9	2
12	3
14	1
18	1
20	1
23	1

Observación 11

En el ámbito de las frecuencias es sumamente útil anotar todo en una tabla como lo hicimos en el ejemplo. Posteriormente definiremos lo que se conoce como **tabla de frecuencias**.

4.2. Frecuencia acumulada

Para definir la frecuencia acumulada, primero ordenamos nuestros datos y las frecuencias absolutas obtenidas:

Dato	**Frecuencia** f_i
1	1
5	3
7	3
9	2
12	3
14	1
18	1
20	1
23	1

La frecuencia acumulada de cada observación se obtiene sumando iterativamente las frecuencias absolutas hasta esa observación.

Observación 12

Cuando hablamos de acumulación, es muy natural pensar en sumas.

Definición 8 Frecuencia acumulada

Dado un conjunto de observaciones $x_1, x_2, \ldots, x_n$ con frecuencia f_i, la frecuencia acumulada F_i de dicha observación está dada por

$$\begin{aligned} F_1 &= f_1 \\ F_2 &= f_2 + F_1 \\ &\vdots \\ F_n &= f_n + F_{n-1}. \end{aligned}$$

Observación 13

Una forma de eliminar el proceso iterativo es notar que

$$\begin{aligned} F_1 &= f_1 \\ F_2 &= f_2 + F_1 = f_2 + f_1 \\ F_3 &= f_3 + F_2 = f_3 + f_2 + f_1 \\ &\vdots \\ F_n &= \sum_{i=1}^{n} f_i. \end{aligned}$$

Dicho de otra forma, la k-ésima frecuencia acumulada F_k la obtenemos

$$F_m = \sum_{i=1}^{m} f_i,$$

donde $m \geq i$.

Ejemplo 13

Agregando la columna de frecuencia acumulada al ejemplo anterior:

Dato	**Frecuencia** f_i	**Frecuencia acumulada** F_i
1	1	1
5	3	4
7	3	7
9	2	9
12	3	12
14	1	13
18	1	14
20	1	15
23	1	16

La columna de frecuencia acumulada nos indica cómo se suman las ocurrencias de cada dato a medida que avanzamos por el conjunto de datos.

4.3. Frecuencia relativa

La frecuencia relativa indica qué proporción del total de observaciones representa cada dato.

Definición 9 Frecuencia relativa

Dado un conjunto de observaciones $x_1, x_2, \ldots, x_n$ con frecuencia f_i, la frecuencia relativa de dicha observación está dada por

$$f_{r_i} = \frac{f_i}{n}.$$

Ejemplo 14 **A**

gregando la columna de frecuencia relativa al ejemplo anterior:

Dato	f_i	F_i	**Frecuencia relativa** f_{r_i}
1	1	1	0.0625
5	3	4	0.1875
7	3	7	0.1875
9	2	9	0.125
12	3	12	0.1875
14	1	13	0.0625
18	1	14	0.0625
20	1	15	0.0625
23	1	16	0.0625

La frecuencia relativa nos permite ver la proporción de cada dato respecto al total de observaciones.

4.4. Frecuencia porcentual

La frecuencia porcentual expresa la frecuencia relativa en porcentaje.

Definición 10 **Frecuencia porcentual**

Dado un conjunto de observaciones $x_1, x_2, \ldots, x_n$ con frecuencia relativa f_{r_i}, la frecuencia porcentual está dada por

$$f_{\%_i} = f_{r_i} \cdot 100\,\%.$$

Ejemplo 15 **A**

gregando la columna de frecuencia porcentual al ejemplo anterior:

Dato	f_i	F_i	f_{r_i}	$f_{\%_i}$
1	1	1	0.0625	6.25 %
5	3	4	0.1875	18.75 %
7	3	7	0.1875	18.75 %
9	2	9	0.125	12.5 %
12	3	12	0.1875	18.75 %
14	1	13	0.0625	6.25 %
18	1	14	0.0625	6.25 %
20	1	15	0.0625	6.25 %
23	1	16	0.0625	6.25 %

La frecuencia porcentual permite interpretar la proporción de cada dato de manera más intuitiva, como un porcentaje del total.

4.5. Tabla de frecuencias

La tabla que nos quedó en el último ejemplo es una **tabla de frecuencias**. Esta resume toda la información anterior: datos, frecuencia absoluta, frecuencia acumulada, frecuencia relativa y frecuencia porcentual. Permite visualizar de manera la distribución de un conjunto de datos y es la base para muchos análisis estadísticos posteriores.

4.5.1. Tabla de frecuencias para datos agrupados

El proceso es análogo, pero ahora cada observación se sustituye por la **marca de clase** de cada intervalo o clase. Se agrupan los datos en rangos y se calculan las frecuencias absolutas, acumuladas, relativas y porcentuales para cada clase. Esto facilita el análisis cuando el conjunto de datos es muy grande o continuo.

CAPÍTULO 5

REPRESENTACIONES GRÁFICAS EN ESTADÍSTICA DESCRIPTIVA

En estadística descriptiva no basta con calcular promedios, medianas o desviaciones. Los datos **necesitan verse**. Una buena representación gráfica permite detectar patrones, comparar grupos, identificar valores atípicos y comprender la estructura general de la información de forma inmediata.

Dos conjuntos de datos pueden tener los mismos valores numéricos que los resumen (media, mediana, etc), pero también tener comportamientos muy distintos al representarlos gráficamente.

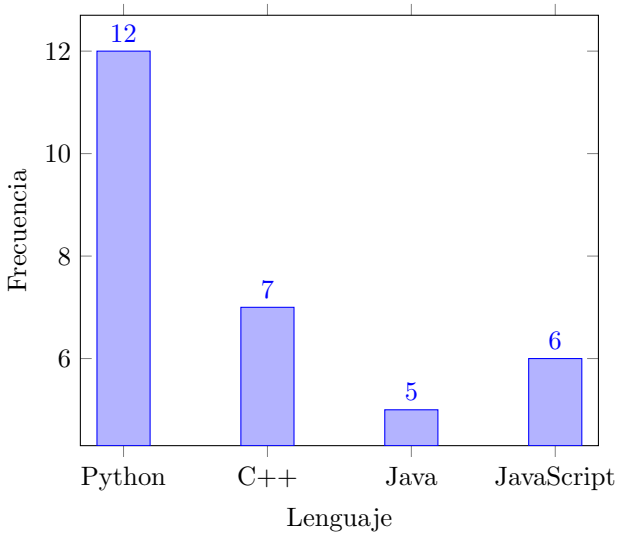
En este capítulo se presentan las principales representaciones gráficas utilizadas en estadística descriptiva, indicando **para qué tipo de datos son más adecuadas** y mostrando ejemplos ilustrativos.

5.1. Gráfica de barras

La gráfica de barras se utiliza principalmente para representar **datos cualitativos** o **datos cuantitativos discretos**. Cada barra representa una categoría y su altura está dada por la frecuencia absoluta o relativa.

Ejemplo 16 Gráfica de barras

Supongamos que se encuestó a un grupo de estudiantes sobre su lenguaje de programación favorito, obteniéndose los siguientes datos:



Observación 14

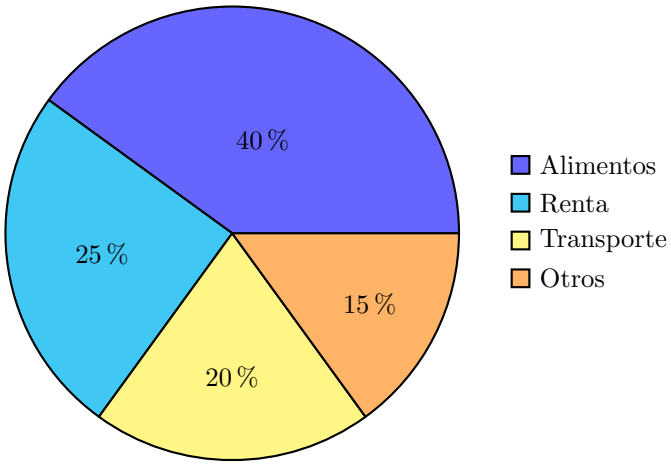
Las gráficas de barras no deben utilizarse para datos continuos, ya que no representan intervalos sino categorías bien definidas.

5.2. Gráfica circular (o de pastel)

La gráfica circular se emplea para mostrar la **proporción** que representa cada categoría respecto al total. Es especialmente útil cuando se desea enfatizar porcentajes.

Ejemplo 17 Gráfica circular

Consideremos la distribución porcentual del gasto mensual de una persona:



Observación 15

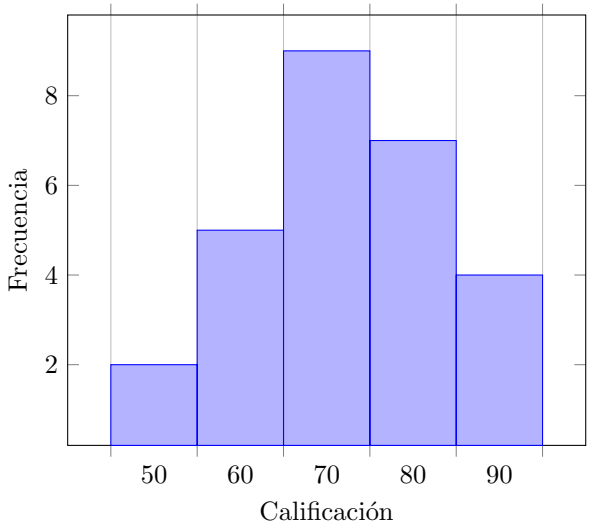
No es recomendable usar gráficas circulares cuando hay muchas categorías, pues la comparación visual se vuelve difícil.

5.3. Histograma

El histograma es una de las gráficas más importantes para datos **cuantitativos continuos**. Representa la distribución de los datos mediante intervalos (clases).

Ejemplo 18 Histograma

Supongamos que se registraron las calificaciones de un examen (en una escala de 0 a 100) para un grupo de estudiantes.



Observación 16

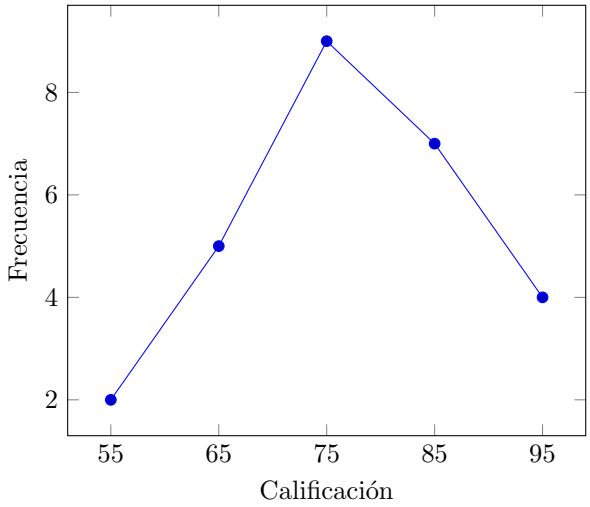
A diferencia de la gráfica de barras, en el histograma las barras **deben estar pegadas**, ya que representan intervalos contiguos.

5.4. Polígono de frecuencias

El polígono de frecuencias es una representación similar al histograma, pero conecta los puntos medios de cada intervalo mediante segmentos de recta. Se usa para analizar la forma de la distribución.

Ejemplo 19 Polígono de frecuencias

Usando los mismos intervalos del histograma anterior:



Observación 17

El polígono de frecuencias es útil para comparar varias distribuciones en una misma gráfica.

5.5. Diagrama de caja y bigotes

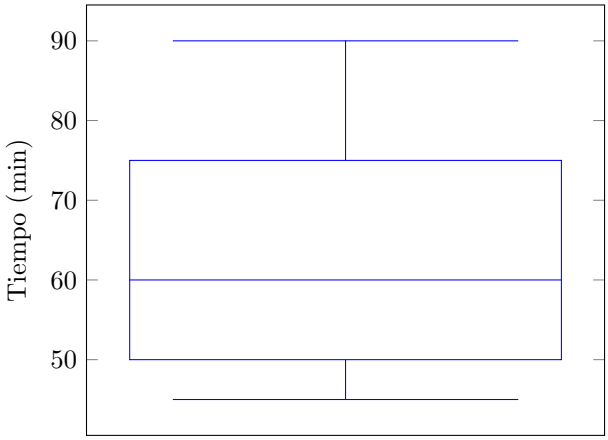
El diagrama de caja (boxplot) resume gráficamente un conjunto de datos cuantitativos mediante los cuartiles, la mediana y los valores extremos. Es especialmente útil para detectar valores atípicos.

Ejemplo 20 Diagrama de caja

Consideremos los siguientes tiempos (en minutos) que tardaron varias personas en resolver un examen. A partir de los datos recolectados, se obtuvieron los siguientes valores estadísticos:

- Tiempo mínimo observado: 45 minutos.
- Primer cuartil (Q_1): 50 minutos.
- Mediana (Q_2): 60 minutos.
- Tercer cuartil (Q_3): 75 minutos.
- Tiempo máximo observado: 90 minutos.

Con esta información se construye el siguiente diagrama de caja, el cual resume gráficamente la distribución de los tiempos de resolución del examen.



En el diagrama se observa que el 50 % central de los datos se encuentra entre 50 y 75 minutos, mientras que la mitad de las personas resolvió el examen en 60 minutos o menos. Los valores extremos indican que el tiempo mínimo fue de 45 minutos y el máximo de 90 minutos.

Observación 18

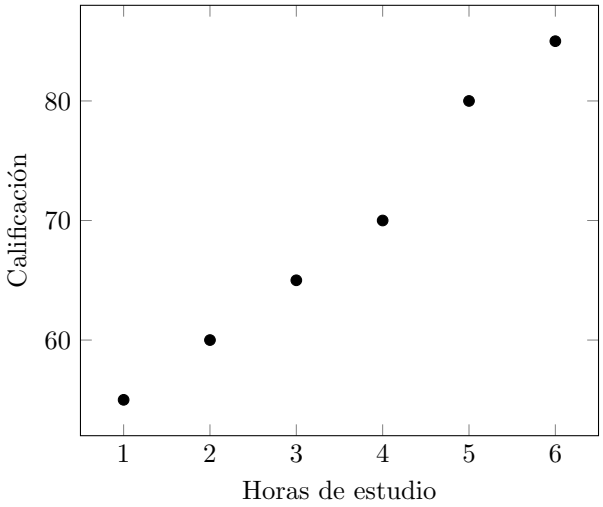
Los diagramas de caja permiten comparar rápidamente la dispersión y posición central entre varios conjuntos de datos.

5.6. Diagrama de dispersión

El diagrama de dispersión se utiliza para analizar la relación entre **dos variables cuantitativas**. Cada punto representa una observación.

Ejemplo 21 Diagrama de dispersión

Supongamos que se mide el número de horas de estudio y la calificación obtenida.



Observación 19

Un patrón creciente o decreciente en la nube de puntos puede indicar la existencia de correlación entre las variables.

CAPÍTULO 6

MEDIDAS DE TENDENCIA CENTRAL

Una **medida de tendencia central** (MTC) es un valor único que describe el comportamiento global de un conjunto de datos, representando su punto central o típico. Es decir, indica al rededor de que valor tienden a concentrarse las observaciones obtenidas.

Ejemplo 22

Supongamos que queremos entender cómo se desempeñan los estudiantes de la Licenciatura en Matemáticas de la Facultad de Ciencias de la UNAM durante su transición a la vida universitaria. Para ello, seleccionamos una muestra de quince alumnos que ya habían superado su primer semestre y les preguntamos por su promedio general en dicho período. Los datos recabados fueron los siguientes:

7.2, 8.5, 7.8, 6.9, 8.1, 7.5, 9.0, 7.4, 8.2, 7.0, 8.4, 7.6, 8.8, 7.1, 7.9

A primera vista, se trata de una sucesión numérica sin un orden aparente. Sin embargo, pero si observamos con atención, notamos que ningún promedio es inferior a 6.9 ni superior a 9.0, y la mayoría oscila entre 7.0 y 8.5. Si tuviéramos que señalar, a ojo de buen cubero, un valor representativo de este grupo, probablemente diríamos que "rondan alrededor del 8".

Si sumamos los quince valores obtenemos un total de 118.4. Al dividir esta suma entre el número de observaciones (15), encontramos que el promedio aritmético es 7.89. Este número, la media, actúa como un punto de equilibrio. Es un valor alrededor del cual tienden a concentrarse nuestros datos y que, en cierto modo, sintetiza la experiencia académica promedio de ese primer semestre, ayudándonos a afirmar que la mayoría tuvo un desempeño cercano al 8, con cierta variabilidad.

Este de hecho es un ejemplo de medida de tendencia central conocida como **media**.

Ejemplo 23

Supongamos ahora que queremos entender la situación económica de un pequeño pueblo a través de los ingresos mensuales de diez familias. Los datos, expresados en pesos, son:

16,500; 17,200; 15,000; 18,000; 16,000; 14,800; 17,000; 2,200,000; 15,500; 16,300

De nuevo, examinemos la lista. Nueve familias reportan ingresos que fluctúan entre 14,800 y 18,000 pesos, una variación esperada. Pero la décima familia reporta un ingreso de 2.2 millones de pesos, un valor que se aleja dramáticamente del resto.
Si aplicáramos ciegamente la fórmula de la media, como lo hicimos en el ejemplo anterior, sumaríamos todos los valores, lo cual da como resultado 2,367,300, que al dividirlos entre el número de observaciones, el resultado sería 236,730 pesos.

Este número ¿Refleja adecuadamente la situación economíca de cada habitante? Claramente no. Decir que "el ingreso familiar promedio es de casi 237 mil pesos" sería engañoso, pues nueve de cada diez familias ganan menos de una décima parte de esa cantidad. La media, aquí, ha sido secuestrada por un solo valor extremo.

Ahora, ordenemos los datos de menor a mayor.

14,800; 15,000; 15,500; 16,000; 16,300; 16,500; 17,000; 17,200; 18,000; 2,200,000

Y tratemos de buscar un número que se encuentre justo en medio de todos estos datos. Con diez observaciones (un número par), tomamos los dos valores del medio (16,300 y 16,500), calculamos su promedio y obtenemos 16,400 pesos. Esta es es otra medida de tendencia central conocida como mediana. Observe su virtud: es completamente inmune al valor atípico de los 2.2 millones. La mediana nos dice que la familia típica, la que se encuentra justo en el medio de la distribución, gana 16,400 pesos. Este valor está mucho menos alejado de la mayoría de los valores de la lista. El único del que está alejado es de nuestro caso atípico. Podremos notar entonces que esta medida ofrece una imagen mucho más realista y equitativa de la tendencia central cuando existen asimetrías o valores atípicos.

Ejemplo 24

Supongamos que el dueño de una zapatería, debe realizar un pedido importante de un determinado tipo de zapato. Para minimizar pérdidas, necesita saber qué tallas son las más demandadas. Afortunadamente el lleva un registro histórico de las ventas de este tipo de calzado y toma las últimas 20 ventas, obteniendo así las siguientes tallas vendidas:

24, 23, 25, 24, 26, 24, 25, 27, 24, 23, 25, 24, 26, 24, 28, 24, 25, 23, 24, 25

Calcular la media (24.4) o la mediana (24) de estas tallas podría darle una idea general, pero no responde la pregunta importante: ¿Cuál se vende más? Entonces, se le ocurre ponerse a contar cada caso y descubre que la talla 24 aparece ocho veces, es decir, aparece con más frecuencia que cualquier otra. Este valor, el que se repite con mayor frecuencia, es la moda.
La moda es como una medida de popularidad. Es la elección de la mayoría dentro del conjunto de datos, y resulta particularmente útil para datos cualitativos (como la marca de celular más preferida) o para identificar el pico de una distribución, señalando la categoría o valor "más típico.[en] términos de recurrencia.

Como hemos visto en estos ejemplos, la media, la mediana y la moda representan tres formas distintas —pero igualmente válidas— de responder a la misma pregunta : ¿hacia dónde se concentran nuestros datos?

Sin embargo, conocer sus nombres y propósitos es solo el primer paso. Lo que sigue es sumergirnos en el cómo.

6.1. Media aritmética

La primera medida de tendencia central que estudiaremos es la media, o siendo más específicos, la media aritmética.

Definición 11 Media

Sean las n observaciones $x_1, x_2, ..., x_n$ de un experimento. La **media aritmética** de dichas operaciones está dada por

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{x_1 + x_2 + ... + x_n}{n}$$

Observación 20

Es natural y correcto decir que si no tenemos datos ($n = 0$):

¡La media no está definida!.

Ejemplo 25

Un profesor quiere estudiar el rendimiento de un salón en el semestre concluido. Para ello toma todas las calificaciones de sus 10 alumnos:

8, 9, 9.1, 6, 6.4, 7, 7, 7.1, 10, 8.2

Para obtener la media de las calificaciones hace lo siguiente:

$$\bar{x} = \frac{8 + 9 + 9{,}1 + 6 + 6{,}4 + 7 + 7 + 7{,}1 + 10 + 8{,}2}{10} \approx 7{,}78$$

Esto quiere decir que la mayoría de calificaciones tienden a concentrarse alrededor del 7.68. Decidió quedarse con esta medida, pues la mayoría de observaciones no eran muy distantes de $\bar{x}$.

A pesar de su poder y sencillez, la media tiene una debilidad: es extremadamente sensible a valores atípicos. Esto significa que si en nuestro conjunto aparece un solo valor (o muy pocos en comparación al total) drásticamente distante del resto, el promedio aritmético se desplazará hacia él, perdiendo así su capacidad de representar fielmente al grueso de los datos.

Ejemplo 26

Juan realizó una encuesta a sus 6 compañeros de trabajo para determinar cuanto dinero en pesos mexicanos estaban dispuestos a pagar por los sobres de café para la oficina y obtuvo los resultados:

1000, 20, 35, 20, 40, 34

Como cada uno dio cantidades distintas, intentó calcular su promedio

$$\bar{x} = \frac{1000 + 20 + 35 + 20 + 40 + 34}{6} = 191{,}5$$

Sin embargo, noto que la mayoría de sus compañeros no estaban dispuestos a pagar más de 40 pesos, por lo tanto decidió no usar al promedio como medida que determine la cantidad de café a comprar, Juan es inteligente.

Por otro lado, la sensibilidad de la media a valores atípicos puede ser relativa. Si nuestro conjunto de datos tiene una variabilidad natural pequeña, o si los valores extremos no están tan alejados del resto, el promedio puede seguir siendo una medida informativa.

Ejemplo 27

El mismo profesor quiere estudiar el rendimiento de otro salón en el semestre concluido. Para ello toma todas las calificaciones de sus 10 alumnos:

$$0,\ 9,\ 3.1,\ 2,\ 2.6,\ 1,\ 1,\ 2.2,\ 10+,\ 3$$

Para obtener la media de las calificaciones hace lo siguiente:

$$\bar{x} = \frac{0+9+3{,}1+2+2{,}6+1+1+2{,}2+10+3}{10} \approx 3{,}39$$

Esto quiere decir que la mayoría de calificaciones tienden a concentrarse alrededor del 3.39. Decidió quedarse con esta medida, pues aunque habían 2 valores bastante atípicos, esto no causó un movimiento brusco del promedio general.

Pregunta de control 12:

En ambos ejemplos del profesor, la media obtenida fue una medida que representa el comportamiento de nuestro conjunto de datos. ¿Qué tipo de medida fue: un estadístico o un parámetro?

hint Toma en cuenta que la intención del profesor era estudiar solamente a cada salón por individual. Para obtener una respuesta, define entonces la población y la muestra en cada caso.

6.1.1. Media para datos agrupados

En los ejemplos anteriores definimos la **media aritmética** para datos no agrupados. Para datos agrupados, la idea es la misma: queremos un valor representativo del conjunto. Lo que cambia en este caso no es la definición, sino los elementos que utilizamos para calcularla.

Recordemos que, al agrupar datos, perdemos la información individual y sólo conservamos las clases y sus frecuencias. Es por esta razón que se introdujo el concepto de **marca de clase** c_i, que actúa como un representante numérico del intervalo i.

La media para datos agrupados se calcula entonces usando las marcas de clase en lugar de los datos originales. Más aún, como no todos los intervalos contienen la misma cantidad de datos, cada marca de clase debe ponderarse por la **frecuencia de clase** f_i.

Definición 13 Media para datos agrupados

Si una variable cuantitativa se ha agrupado en k clases con marcas de clase c_i y frecuencias de clase f_i, entonces la media se calcula como:

$$\bar{x} = \frac{\sum_{i=1}^{k} c_i f_i}{n}$$

donde $n = \sum_{i=1}^{k} f_i$ es el número total de observaciones.

Ejemplo 28

Se obtuvieron las edades de 100 estudiantes de ingeniería. Debido a la gran cantidad de datos, se optó por agruparlos de la siguiente manera:

$$[18, 25) \to 55 \text{ datos}$$
$$[25, 32) \to 25 \text{ datos}$$
$$[32, 39) \to 15 \text{ datos}$$
$$[39, 46] \to 5 \text{ datos}$$

Calculamos las marcas de clase:

$$c_1 = \frac{18 + 25}{2} = 21{,}5$$
$$c_2 = \frac{25 + 32}{2} = 28{,}5$$
$$c_3 = \frac{32 + 39}{2} = 35{,}5$$
$$c_4 = \frac{39 + 46}{2} = 42{,}5$$

Usando la fórmula para la media agrupada:

$$\bar{x} = \frac{c_1 f_1 + c_2 f_2 + c_3 f_3 + c_4 f_4}{n}$$

Sustituimos los valores:

$$\bar{x} = \frac{21{,}5(55) + 28{,}5(25) + 35{,}5(15) + 42{,}5(5)}{100}$$

Calculamos el numerador:

$$= \frac{1182{,}5 + 712{,}5 + 532{,}5 + 212{,}5}{100}$$

$$= \frac{2640}{100} = 26{,}4$$

Por lo tanto, la edad media estimada de los estudiantes es:

$$\boxed{\bar{x} = 26{,}4}$$

Observación 21

Si intentáramos calcular la media agrupada ignorando las frecuencias, obtendríamos resultados inconsistentes. Las frecuencias ponderan adecuadamente la contribución de cada intervalo.

6.2. Mediana

La **mediana** es una medida de tendencia central (MTC) que representa el valor que ocupa la posición central en un conjunto de datos ordenados. Es decir, al ordenar los datos de menor a mayor, la mediana divide al conjunto en dos partes: la misma cantidad de valores mayores que menores que ella.

6.2.1. Cálculo de la mediana

Supongamos que tenemos un conjunto de datos

$$x_1, x_2, \ldots, x_n$$

y que los hemos ordenado de menor a mayor (si no lo están, debemos ordenarlos primero):

$$x_1 \leq x_2 \leq \cdots \leq x_n.$$

La mediana es un valor x_i tal que existen k valores menores y k valores mayores que él. Matemáticamente, si n es impar:

$$x_1 \leq x_2 \leq \cdots \leq x_{i-1} \leq x_i \leq x_{i+1} \leq \cdots \leq x_n$$

con

$$k = i - 1 = n - i.$$

Si n es impar, el valor central x_i pertenece directamente al conjunto y es la mediana. En cambio, si n es par, no hay un valor único en el centro, sino dos valores que ocupan las posiciones centrales, y la mediana se calcula como el promedio de ambos.

6.2.2. Mediana para datos impares

Ejemplo 29

Se tienen las siguientes observaciones:

$$1, 5, 4, 5, 0, 1, 2, 4, 0.$$

Ordenamos los datos de menor a mayor:

$$0, 0, 1, 1, 2, 4, 4, 5, 5.$$

Podemos observar que hay 4 datos menores que 2 $(0, 0, 1, 1)$ y 4 datos mayores que 2 $(4, 4, 5, 5)$. Por lo tanto, la mediana es

$$M_e = 2.$$

6.2.3. Mediana para datos pares

Ejemplo 30

Se tienen las siguientes observaciones:

$$1, 5, 4, 5, 0, 1, 2, 4, 0, 6.$$

Ordenamos los datos de menor a mayor:

$$0, 0, 1, 1, 2, 4, 4, 5, 5, 6.$$

Como el número de datos es par, los valores centrales son 2 y 4. La mediana se calcula como el promedio de ambos:

$$M_e = \frac{2+4}{2} = 3,$$

que representa justamente el punto medio entre los dos valores centrales.

6.2.4. La mediana y los valores atípicos

La mediana, a diferencia de la media aritmética, es menos sensible a los **valores atípicos** en un conjunto de datos. Esto la convierte en una medida de tendencia central más robusta cuando existen observaciones extremas.

Ejemplo 31

Se tienen las siguientes observaciones:

$$1000, 1, 5, 4, 5, 0, 1, 2, 4, 0, 6.$$

Ordenamos los datos de menor a mayor:

$$0, 0, 1, 1, 2, 4, 4, 5, 5, 6, 1000.$$

Como el número de datos es impar, la mediana es el valor central de la lista, que es

$$M_e = 4,$$

ya que hay 5 valores menores y 5 valores mayores que él.
Si calculamos la media aritmética, obtenemos:

$$\bar{x} = \frac{0+0+1+1+2+4+4+5+5+6+1000}{11} \approx 93{,}45.$$

Esto muestra que la media se desplazó drásticamente debido al valor atípico 1000, mientras que la mediana refleja mucho mejor la concentración central de los datos. Por ello, la mediana es una herramienta más resistente frente a valores extremos.

6.3. Moda

La **moda** es una medida de tendencia central (MTC) muy útil, pues a diferencia de la media o la mediana, *tiene sentido para variables cualitativas nominales*. Dicho esto, podemos definir la moda como el o los valores que se repiten más veces en un conjunto de datos.

Definición 14 Moda

Se define la moda M_o de una muestra como aquel valor de la variable que tiene la **frecuencia máxima**.

Ejemplo 32 Moda en variables cualitativas

Se hizo una encuesta a 10 alumnos sobre su color de cabello y se obtuvieron las observaciones:

Negro, Castaño, Rojo, Morado, Negro, Castaño, Castaño, Rubio, Rojo, Castaño

Contando la frecuencia de cada valor:

- Castaño $\rightarrow 4$
- Negro $\rightarrow 2$
- Rojo $\rightarrow 2$
- Rubio $\rightarrow 1$
- Morado $\rightarrow 1$

Por lo tanto, el valor que más se repite es Castaño:

$$M_o = \text{“Castaño”}.$$

Si ningún dato se repite, la moda no existe.

Ejemplo 33 Moda inexistente

Se tienen las edades de 5 alumnos:

$$12, 13, 15, 10, 11$$

Cada valor es único, por lo que en este caso:

$$M_o \text{ no existe.}$$

Si hay dos o más valores con la misma frecuencia máxima, todos ellos se consideran moda.

Ejemplo 34 Moda múltiple

Se tienen los nombres de 7 alumnos:

Juan, Jimena, Mauricio, Juan, Pedro, Pedro, Roberto

Aquí Juan y Pedro tienen la misma frecuencia máxima (2) y ningún otro valor se repite más:

$$M_o = \text{“Juan”, “Pedro”}.$$

6.4. Medidas de posición relativa: Cuantiles

Las medidas como **cuartiles, deciles y percentiles** son una generalización de la mediana. Mientras que la mediana divide un conjunto de datos ordenados en dos partes iguales, los cuantiles permiten dividir

los datos en **más partes**, brindando información más detallada sobre la distribución y concentración de los datos.

6.4.1. Cuartiles

Definición 15 Cuartiles

Los **cuartiles** son valores que dividen un conjunto de datos ordenados en cuatro partes iguales. Cada cuartil indica la posición de un porcentaje determinado de los datos:

- Q_1 (primer cuartil) $\rightarrow$ 25 % de los datos son menores o iguales a este valor.
- Q_2 (segundo cuartil) $\rightarrow$ coincide con la mediana (50 % de los datos son menores o iguales).
- Q_3 (tercer cuartil) $\rightarrow$ 75 % de los datos son menores o iguales a este valor.

Ejemplo 35 Cuartiles

Se tiene el conjunto de datos:

$$3, 5, 7, 8, 12, 13, 14, 18, 21, 23$$

Ordenados de menor a mayor (ya están ordenados).

- Q_1: mediana de la primera mitad $(3, 5, 7, 8, 12) \Rightarrow Q_1 = 7$
- Q_2: mediana general $(3, 5, 7, 8, 12, 13, 14, 18, 21, 23) \Rightarrow Q_2 = 12$
- Q_3: mediana de la segunda mitad $(13, 14, 18, 21, 23) \Rightarrow Q_3 = 18$

6.4.2. Deciles

Definición 16 Deciles

Los **deciles** dividen un conjunto de datos ordenados en 10 partes iguales. Cada decil D_i indica que aproximadamente $i \cdot 10$ % de los datos son menores o iguales a ese valor.

- $D_1 \rightarrow 10$ %
- $D_2 \rightarrow 20$ %
- $\vdots$
- $D_9 \rightarrow 90$ %

Ejemplo 36 Deciles

Usando el mismo conjunto de datos anterior:

$$3, 5, 7, 8, 12, 13, 14, 18, 21, 23$$

- $D_1 \approx 3{,}9$ (10 % de los datos $\leq D_1$)
- $\vdots$
- $D_5 = Q_2 = 12$ (50 % de los datos $\leq D_5$)
- $\vdots$
- $D_9 \approx 22{,}1$ (90 % de los datos $\leq D_9$)

6.4.3. Percentiles

Definición 17 Percentiles

Los **percentiles** dividen un conjunto de datos ordenados en 100 partes iguales. Cada percentil P_i indica que aproximadamente i % de los datos son menores o iguales a ese valor. - P_{25} coincide con Q_1, P_{50} con Q_2, P_{75} con Q_3.

Ejemplo 37 Percentiles

Para el conjunto de datos:

$$3, 5, 7, 8, 12, 13, 14, 18, 21, 23$$

- $P_{10} \approx 4$
- $P_{25} = Q_1 = 7$
- $P_{50} = Q_2 = 12$
- $P_{75} = Q_3 = 18$
- $P_{90} \approx 22$

CAPÍTULO 7

MEDIDAS DE DISPERSIÓN

En el capítulo anterior definimos las **medidas de tendencia central**. Estas medidas permiten reducir toda la información de un conjunto de datos a un solo valor representativo.

Sin embargo, aunque en algunos casos estos valores son suficientes para resumir nuestros datos, en muchas situaciones más realistas los datos pueden estar tan **dispersos** que un solo número no logra capturar adecuadamente la variabilidad existente.

Por ello, en este capítulo nos enfocaremos en definir y analizar aquellos valores que permiten medir qué tan dispersos están nuestros datos. Estas medidas son fundamentales para entender la **variabilidad**, la **homogeneidad** o **heterogeneidad** de un conjunto de observaciones.

7.1. Rango

Una primera forma de medir la dispersión de los datos es mediante el **rango** o **recorrido**. El rango es básicamente una medida de separación entre los valores extremos de un conjunto de datos.

De manera intuitiva, podríamos definir el rango como la diferencia entre el valor más grande observado $x_{\max}$ y el valor más pequeño $x_{\min}$:

$$R = x_{\max} - x_{\min}$$

Sin embargo, esta medida puede resultar **sesgada**, ya que los valores extremadamente altos o bajos pueden dar una impresión exagerada de la dispersión real de los datos.

Ejemplo 38

Consideremos el conjunto de datos:

$$1,\ 4,\ 1,\ 2,\ 5,\ 10,\ 1000$$

Entonces la diferencia entre los extremos está dada por:

$$R = x_{\max} - x_{\min} = 1000 - 1 = 999$$

Del ejemplo anterior podría parecer que los datos están muy dispersos. Sin embargo, si omitimos el valor extremo 1000, la nueva diferencia entre extremos sería apenas 9 unidades. Esto muestra que, aunque el rango es intuitivo, puede ser **engañoso** si hay valores atípicos.

Para superar este problema, podemos utilizar otra medida de dispersión más robusta: el **rango intercuartílico**.

7.1.1. Rango intercuartílico

Definición 18 Rango intercuartílico

Dado un conjunto de datos, definimos el **rango intercuartílico** R_I como la diferencia entre el tercer cuartil Q_3 y el primer cuartil Q_1:

$$R_I = Q_3 - Q_1$$

Ejemplo 39

Consideremos el conjunto de datos:

$$999,\ 23,\ 1200,\ 998,\ 950,\ 1000,\ 1099,\ 991,\ 992$$

Calculando los cuartiles obtenemos:

$$Q_1 = 970{,}5, \quad Q_3 = 1049{,}5$$

Entonces el rango intercuartílico es:

$$R_I = Q_3 - Q_1 = 1049{,}5 - 970{,}5 = 79$$

Observación 22

El valor de 79 refleja mucho mejor la dispersión de los datos, ya que los valores extremos (como 23) no distorsionan significativamente la medida.

7.2. Desviación media

Otra medida de dispersión es la **desviación media**. Esta mide qué tan alejados están, en promedio, todos los datos de su **media aritmética**.

Definición 19 Desviación media

Dado un conjunto de n datos con media $\bar{x}$, k distintos valores y frecuencia absoluta f_i para $i = 1, \ldots, k$, la desviación media $D_{\bar{x}}$ se define como:

$$D_{\bar{x}} = \frac{\sum_{i=1}^{k} |x_i - \bar{x}| f_i}{n}$$

Ejemplo 40

Sea el conjunto de datos:

$$0,\ 10,\ 11,\ 4,\ 5,\ 10,\ 100$$

La media es:

$$\bar{x} = \frac{0+10+11+4+5+10+100}{7} = 20$$

Entonces su desviación media es:

$$D_{\bar{x}} = \frac{|0-20|\cdot 1 + |10-20|\cdot 2 + |11-20|\cdot 1 + |4-20|\cdot 1 + |5-20|\cdot 1 + |100-20|\cdot 1}{7}$$
$$\approx 22{,}86$$

Observación 23

Es intuitivo notar que si no hay datos repetidos, la fórmula se simplifica a:

$$D_{\bar{x}} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

Observación 24

La desviación media puede verse afectada por valores atípicos (como el 100 en este conjunto). Por ejemplo, aunque el valor más pequeño es 0, su distancia con respecto a la media (20) es menor que la desviación promedio. Esto sucede porque la media fue sesgada por el valor extremo desde el principio.

7.3. Desviación respecto a la mediana

Cuando, en lugar de usar la media como medida de tendencia central, se usa la **mediana**, obtenemos la llamada **desviación respecto a la mediana**.

Definición 20 Desviación respecto a la mediana

Dado un conjunto de n datos con mediana M_e, k distintos valores y frecuencia absoluta f_i para $i = 1, \ldots, k$, la desviación respecto a la mediana D_{M_e} se define como:

$$D_{M_e} = \frac{\sum_{i=1}^{k} |x_i - M_e| f_i}{n}$$

Ejemplo 41

Tomando el mismo conjunto de datos y ordenándolo:

$$0,\ 4,\ 5,\ 10,\ 10,\ 11,\ 100$$

La mediana es $M_e = 10$. Entonces su desviación respecto a la mediana es:

$$D_{M_e} = \frac{|0-10|\cdot 1 + |10-10|\cdot 2 + |11-10|\cdot 1 + |4-10|\cdot 1 + |5-10|\cdot 1 + |100-10|\cdot 1}{7}$$
$$= 16$$

Observación 25

Aunque la dispersión sigue siendo relativamente alta, se reduce con respecto a la calculada usando la media, mostrando que la mediana es menos sensible a valores atípicos.

7.4. Varianza

Cuando se utiliza a la media como medida de tendencia central, la medida de dispersión más empleada es la **varianza**. Conceptualmente, es muy similar a la desviación media; sin embargo, en lugar de utilizar valores absolutos para evitar la cancelación de términos, eleva al cuadrado las diferencias respecto a la media.

Este uso del cuadrado no es casual: además de penalizar más fuertemente las desviaciones grandes, tiene ventajas analíticas importantes. Por ejemplo, en optimización y *machine learning*, las funciones cuadráticas (como la varianza) son suaves y derivables en todo su dominio, a diferencia del valor absoluto, que no es derivable en el punto 0.

Definición 21 Varianza

Dado un conjunto de n datos con media $\bar{x}$, k valores distintos $x_1, \ldots, x_k$ y frecuencias absolutas f_i para $i = 1, \ldots, k$, la **varianza muestral**, denotada por s^2, se define como

$$s^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 f_i}{n-1}.$$

Observación 26

El divisor $n-1$ aparece porque esta expresión corresponde a la varianza muestral; su uso corrige el sesgo que se produciría al estimar la dispersión poblacional a partir de una muestra (corrección de Bessel).

Ejemplo 42

Consideremos el conjunto de datos

$$\{2, 4, 4, 6\}.$$

Aquí $n = 4$ y la media es

$$\bar{x} = \frac{2+4+4+6}{4} = 4.$$

Entonces,

$$s^2 = \frac{(2-4)^2 + (4-4)^2 + (4-4)^2 + (6-4)^2}{3} = \frac{4+0+0+4}{3} = \frac{8}{3}.$$

Una fórmula alternativa, que suele implicar un menor costo computacional, puede deducirse de la siguiente manera:

$$\begin{aligned} s^2 &= \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 f_i}{n-1} \\ &= \frac{\sum_{i=1}^{k}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) f_i}{n-1} \\ &= \frac{\sum_{i=1}^{k} f_i x_i^2 - 2\bar{x}\sum_{i=1}^{k} f_i x_i + \bar{x}^2 \sum_{i=1}^{k} f_i}{n-1}. \end{aligned}$$

Dado que $\sum_{i=1}^{k} f_i = n$, se obtiene

$$s^2 = \frac{\sum_{i=1}^{k} f_i x_i^2 - 2\bar{x}\sum_{i=1}^{k} f_i x_i + n\bar{x}^2}{n-1}.$$

7.5. Desviación estándar

Un inconveniente de la varianza es que, al elevar las diferencias al cuadrado, la medida resultante queda expresada en unidades al cuadrado, lo que puede dificultar su interpretación directa. Para solucionar esto se introduce la **desviación estándar**, que no es más que la raíz cuadrada de la varianza y, por tanto, se expresa en las mismas unidades que los datos originales.

Definición 22 Desviación estándar

Dado un conjunto de n datos con media $\bar{x}$, k valores distintos y frecuencias absolutas f_i para $i = 1, \ldots, k$, la **desviación estándar muestral**, denotada por s, se define como

$$s = \sqrt{\frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 f_i}{n-1}}.$$

Ejemplo 43

Para el conjunto de datos del ejemplo anterior, cuya varianza es $s^2 = \frac{8}{3}$, la desviación estándar es

$$s = \sqrt{\tfrac{8}{3}} \approx 1{,}63.$$

Esto indica que, de manera típica, los datos se desvían alrededor de 1,63 unidades respecto a la media.

7.6. Coeficientes de variación

Uno de los principales inconvenientes de las medidas de dispersión estudiadas hasta ahora es que sus resultados se expresan en las mismas unidades que los datos (o incluso en unidades al cuadrado, como en el caso de la varianza). Esto dificulta la comparación entre conjuntos de datos medidos en escalas distintas o con magnitudes muy diferentes.

Observación 27

Esto significa que, mediante los coeficientes de variación, es posible comparar el grado de dispersión entre distintos conjuntos de datos, incluso cuando están medidos en unidades diferentes.

Por ejemplo, podemos determinar si existe mayor o menor variabilidad en un conjunto de mediciones de estaturas (expresadas en metros) que en un conjunto de mediciones de edades (expresadas en años). Esto no sería posible utilizando directamente la varianza o la desviación estándar, ya que estaríamos comparando dispersión medida en metros con dispersión medida en años, lo cual carece de sentido.

Para subsanar este problema se introduce la noción de **coeficientes de variación**, los cuales permiten expresar la dispersión de manera relativa, generalmente en términos porcentuales. De esta forma, la variabilidad de los datos puede interpretarse de manera más clara y compararse entre distintos conjuntos.

7.6.1. Coeficiente de Pearson

Uno de los coeficientes de variación más conocidos es el **coeficiente de variación de Pearson**.

Definición 23 Coeficiente de Pearson

Sea $x_1, x_2, \ldots, x_n$ un conjunto de observaciones, $\bar{x}$ la media de dicho conjunto y s la desviación estándar muestral. El **coeficiente de variación de Pearson**, denotado por CV, se define como

$$CV = \frac{s}{|\bar{x}|}.$$

El valor absoluto en el denominador se utiliza para evitar ambigüedades cuando la media es negativa. Este coeficiente es adimensional y suele expresarse como porcentaje.

Ejemplo 44

Usando los datos del ejemplo 7.4, se tiene que

$$\bar{x} = 4 \quad \text{y} \quad s \approx 1{,}63.$$

Por lo tanto, el coeficiente de variación de Pearson es

$$CV = \frac{1{,}63}{4} \approx 0{,}41.$$

Interpretando este resultado como un porcentaje, podemos afirmar que la dispersión de los datos respecto a su media es aproximadamente del 41 % del valor de la media. Esto indica una variabilidad moderada en relación con el tamaño promedio de los datos.

7.6.2. Coeficiente de desviación media

Otro ejemplo de coeficiente de variación es el **coeficiente de desviación media**, que es muy parecido al coeficiente de Pearson, pero en lugar de usar la desviación estándar, utiliza la **desviación media** $D_{\bar{x}}$ del conjunto de datos.

Definición 24 Coeficiente de desviación media

Sea $x_1, x_2, \ldots, x_n$ un conjunto de observaciones, $\bar{x}$ su media y $D_{\bar{x}}$ la desviación media. El **coeficiente de desviación media**, denotado por CVM, se define como

$$CVM = \frac{D_{\bar{x}}}{|\bar{x}|}.$$

Este coeficiente también es adimensional y se puede expresar como porcentaje, lo que permite comparar la dispersión relativa de distintos conjuntos de datos de manera análoga al coeficiente de Pearson.

Ejemplo 45

La desviación media es

$$D_{\bar{x}} = \frac{|2-4| + |4-4| \cdot 2 + |6-4|}{4} = 1$$

Entonces

$$CVM = \frac{1}{1{,}63} = 0{,}61$$

es decir, los datos tenían una variación del %61 con respecto a la media.

CAPÍTULO 8

MEDIDAS DE ASIMETRÍA Y CURTOSIS

La estadística descriptiva no se limita al estudio de las medidas de tendencia central o de dispersión de un conjunto de datos. Existen otros atributos relevantes que permiten describir la *forma* de los datos. En particular, podemos preguntarnos: *¿qué tan equilibrados están los datos alrededor de su centro (la media)?*. Esta cuestión se aborda mediante las **medidas de asimetría**.

Por otro lado, la pregunta *¿qué tan concentrados están los datos alrededor de la media?* se responde a través de otra medida conocida como **curtosis**. En este capítulo nos enfocaremos primero en el estudio de la asimetría.

8.1. Asimetría

La **simetría**, o su contraparte, la **asimetría**, describe cómo se distribuyen los valores de un conjunto de datos con respecto a la media. De manera intuitiva, un conjunto es simétrico cuando valores de la variable equidistantes, a uno y otro lado, del valor central (media) tienen la misma frecuencia.

Es importante notar que la simetría es una condición estricta: un conjunto de datos es simétrico o no lo es. En cambio, la asimetría admite distintos grados, los cuales pueden ser cuantificados mediante medidas específicas.

Definición 25 Simetría

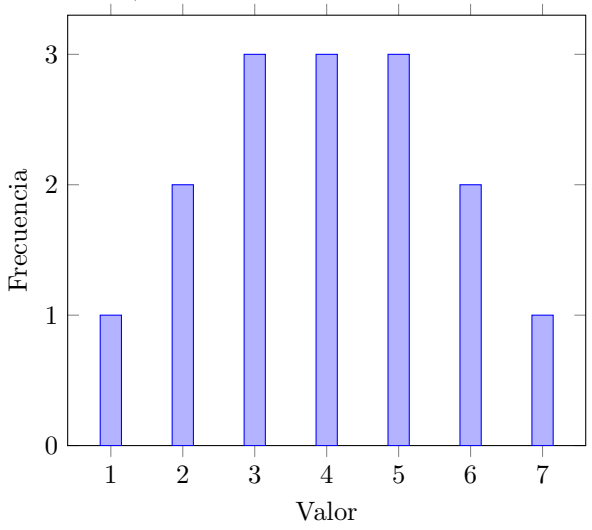
Decimos que un conjunto de datos $x_1, x_2, \ldots, x_n$ con media $\bar{x}$ es **simétrico** si para cada distancia $d > 0$, el número de observaciones situadas a una distancia d por debajo de la media es igual al número de observaciones situadas a la misma distancia d por encima de la media.

Ejemplo 46

Considérese el conjunto de datos

$$1,\ 2,\ 2,\ 3,\ 3,\ 3,\ 4,\ 4,\ 4,\ 5,\ 5,\ 5,\ 6,\ 6,\ 7$$

La media es $\bar{x} = 4$.
Los valores se distribuyen de manera equilibrada alrededor de la media: a cada valor menor que 4 le corresponde un valor mayor que 4 a la misma distancia y con la misma frecuencia (por ejemplo, 3 y 5, 2 y 6, 1 y 7). Por lo tanto, el conjunto es **simétrico**.



Cuando esta compensación de distancias no ocurre, el conjunto presenta **asimetría**.

Definición 26 Asimetría

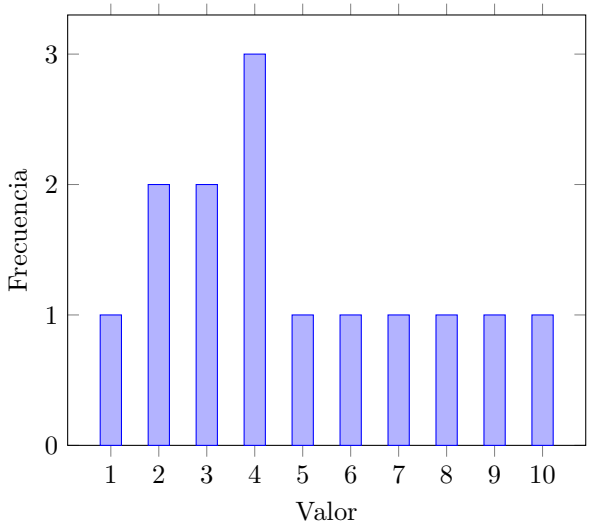
Decimos que un conjunto de datos es **asimétrico** si existen distancias a la media que no están compensadas por observaciones situadas al otro lado con la misma frecuencia.

Ejemplo 47

Considérese el conjunto

$$1,\ 2,\ 2,\ 3,\ 3,\ 4,\ 4,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10$$

La media es $\bar{x} \approx 4{,}86$.
Aunque existen valores a ambos lados de la media, las observaciones grandes a la derecha (como $8, 9, 10$) no están compensadas por valores igualmente alejados a la izquierda. Esto genera un desequilibrio claro, por lo que el conjunto es **asimétrico**.



8.1.1. Conjuntos sesgados

Cuando un conjunto de datos es asimétrico, es posible identificar la **dirección del desequilibrio**. Esta depende de hacia qué lado de la media se concentran las observaciones más alejadas.

Definición 27 Sesgo a la derecha

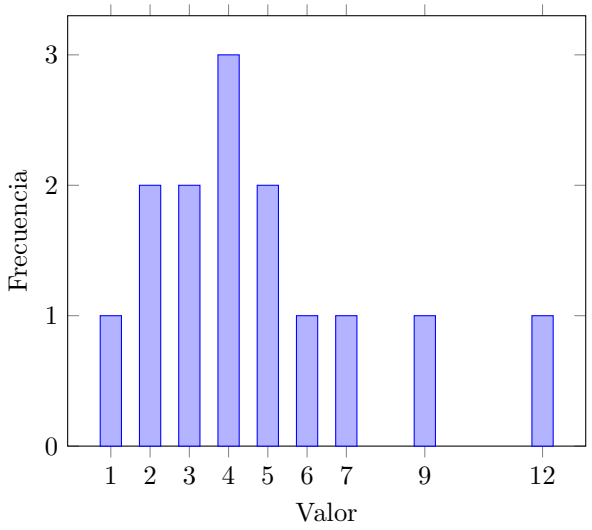
Decimos que un conjunto de datos presenta **sesgo a la derecha** si las observaciones situadas por encima de la media alcanzan distancias mayores que las observaciones situadas por debajo, sin que estas últimas logren compensarlas.

Ejemplo 48

Considérese el conjunto

$$1,\ 2,\ 2,\ 3,\ 3,\ 4,\ 4,\ 4,\ 5,\ 5,\ 6,\ 7,\ 9,\ 12$$

La media es $\bar{x} \approx 4{,}64$.
Las observaciones grandes $7, 9$ y 12 generan distancias importantes a la derecha de la media que no son compensadas por valores igualmente alejados a la izquierda. Por ello, el conjunto presenta un marcado **sesgo a la derecha**.



Definición 28 Sesgo a la izquierda

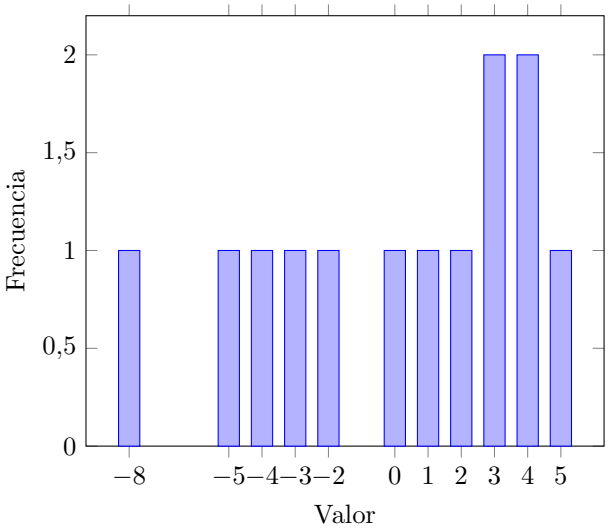
Decimos que un conjunto de datos presenta **sesgo a la izquierda** si las observaciones situadas por debajo de la media alcanzan distancias mayores que las observaciones situadas por encima, sin que estas últimas logren compensarlas.

Ejemplo 49

Considérese el conjunto

$$-8,\ -5,\ -4,\ -3,\ -2,\ 0,\ 1,\ 2,\ 3,\ 3,\ 4,\ 4,\ 5$$

La media es $\bar{x} \approx 0{,}08$.
Los valores negativos grandes $(-8, -5, -4)$ se encuentran mucho más alejados de la media que cualquier valor positivo, generando un desequilibrio hacia la izquierda. El conjunto presenta **sesgo a la izquierda**.



8.1.2. Coeficientes de asimetría

Las representaciones gráficas permiten identificar visualmente la asimetría, pero para cuantificarla se introducen coeficientes numéricos que resumen el desequilibrio de las distancias respecto a la media.

Coeficiente de asimetría de Fisher

Definición 29 Coeficiente de Fisher

Para un conjunto de k datos no repetidos, con media $\bar{x}$ y desviación estándar s se define como

$$g_1 = \frac{1}{s^3}\left[\frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^3 f_i\right],$$

donde

$$s^3 = \left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^3.$$

Ejemplo 50 Coeficiente de Fisher

Considérese el conjunto de datos

$$2,\ 3,\ 3,\ 4,\ 4,\ 5,\ 5,\ 6,\ 6$$

El tamaño de la muestra es $n = 9$. La media aritmética es

$$\bar{x} = \frac{2+3+3+4+4+5+5+6+6}{9} = \frac{38}{9} \approx 4{,}22.$$

Calculamos las desviaciones respecto a la media y sus potencias:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$
2	$-2{,}22$	4,93	$-10{,}95$
3	$-1{,}22$	1,49	$-1{,}82$
3	$-1{,}22$	1,49	$-1{,}82$
4	$-0{,}22$	0,05	$-0{,}01$
4	$-0{,}22$	0,05	$-0{,}01$
5	0,78	0,61	0,48
5	0,78	0,61	0,48
6	1,78	3,17	5,64
6	1,78	3,17	5,64

Sumando los cuadrados y los cubos:

$$\sum (x_i - \bar{x})^2 \approx 15{,}57, \qquad \sum (x_i - \bar{x})^3 \approx -2{,}37.$$

La desviación estándar es

$$s = \sqrt{\frac{1}{9}\sum (x_i - \bar{x})^2} = \sqrt{\frac{15{,}57}{9}} \approx 1{,}31, \qquad s^3 \approx (1{,}31)^3 \approx 2{,}25.$$

Finalmente, el coeficiente de asimetría de Fisher es

$$sk(x) = \frac{1}{s^3}\left[\frac{1}{9}\sum (x_i - \bar{x})^3\right] = \frac{-2{,}37/9}{2{,}25} \approx -0{,}12.$$

El valor es cercano a cero, lo que indica que el conjunto es **aproximadamente simétrico**.

Coeficiente de asimetría de Pearson

Definición 30 Coeficiente de Pearson

Se define como

$$A_p = \frac{\bar{x} - M_o}{s}.$$

Ejemplo 51 Coeficiente de Pearson

Considérese el conjunto de datos

$$1,\ 2,\ 2,\ 3,\ 3,\ 3,\ 4,\ 5$$

Se tiene $n = 8$. La media aritmética es

$$\bar{x} = \frac{1+2+2+3+3+3+4+5}{8} = \frac{23}{8} = 2{,}875.$$

La moda del conjunto es

$$M_o = 3,$$

ya que es el valor con mayor frecuencia.
Calculamos la desviación estándar:

$$s = \sqrt{\frac{1}{8}\sum (x_i - \bar{x})^2} \approx 1{,}17.$$

El coeficiente de asimetría de Pearson es entonces

$$A_p = \frac{\bar{x} - M_o}{s} = \frac{2{,}875 - 3}{1{,}17} \approx -0{,}11.$$

El valor negativo indica una **ligera asimetría a la izquierda**, aunque de magnitud pequeña.

Observación 28

Cuando tenemos un conjunto de datos con una sola moda es posible usar los siguientes criterios para determinar **asimetrías**:

1. **Simétrico:** Si $\bar{x} = M_e = M_o$.
2. **Asimetría positiva:** Si $\bar{x} \geq M_e \geq M_o$
3. **Asimetría negativa:** Si $\bar{x} \leq M_e \leq M_o$

8.2. Curtosis

Además de la simetría, otra característica importante de la forma en que se distribuyen los datos es **qué tan concentrados están alrededor del valor central** y **qué tan frecuentes son los valores extremos**.

Intuitivamente, dos conjuntos de datos pueden tener la misma media y la misma desviación típica, pero diferir notablemente en su forma: en algunos casos la mayoría de los datos se agrupa muy cerca del centro, formando un **pico pronunciado** en el histograma; en otros, los datos se dispersan de manera más uniforme, dando lugar a un histograma **más plano**.

Cuando los datos presentan un fuerte agrupamiento alrededor del valor central y colas relativamente largas, se dice que la distribución es **leptocúrtica**. En el extremo opuesto, cuando el histograma es bajo y aplanado, con menor concentración central, se habla de una distribución **platicúrtica**. Entre ambos casos se encuentra la distribución **mesocúrtica**, cuyo grado de agrupamiento coincide con el de la distribución normal o campana de Gauss.

Esta característica de la distribución, que describe el **grado de concentración de los datos y la presencia de valores extremos**, se denomina **curtosis**. Para cuantificarla, se introduce el coeficiente de curtosis, definido como el cociente entre el momento de cuarto orden respecto a la media y la cuarta potencia de la desviación típica.

8.2.1. Coeficiente de curtosis

Definición 31 Coeficiente de curtosis

Para un conjunto de k datos no repetidos, con media $\bar{x}$ y desviación estándar s, se define el coeficiente de curtosis como

$$g_2 = \frac{1}{s^4}\left[\frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^4 f_i\right],$$

donde f_i representa la frecuencia del dato x_i.

Ejemplo 52 Distribución leptocúrtica

Considérese el conjunto de datos

$$3,\ 4,\ 4,\ 4,\ 5,\ 4,\ 4,\ 4,\ 5$$

Se tiene $n = 9$, $\bar{x} = 4{,}11$ y $s = 0{,}57$. Entonces

$$g_2 = \frac{1}{(0{,}57)^4}\left[\frac{1}{9}\left((3-4{,}11)^4 + 6(4-4{,}11)^4 + 2(5-4{,}11)^4\right)\right] \approx 4{,}52$$

Como $g_2 > 3$, la distribución es **leptocúrtica**.

Ejemplo 53 Distribución mesocúrtica

Considérese el conjunto de datos

$$2,\ 3,\ 3,\ 4,\ 4,\ 4,\ 5,\ 5,\ 6$$

Aquí $n = 9$, $\bar{x} = 4$ y $s = 1{,}15$. Por tanto,

$$g_2 = \frac{1}{(1{,}15)^4}\left[\frac{1}{9}\left((2-4)^4 + 2(3-4)^4 + 3(4-4)^4 + 2(5-4)^4 + (6-4)^4\right)\right] \approx 3{,}05$$

El valor obtenido es cercano a 3, por lo que la distribución es **mesocúrtica**.

Ejemplo 54 Distribución platicúrtica

Considérese el conjunto de datos

$$1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8$$

Se tiene $n = 8$, $\bar{x} = 4{,}5$ y $s = 2{,}29$. Luego,

$$g_2 = \frac{1}{(2{,}29)^4}\left[\frac{1}{8}\sum_{i=1}^{8}(x_i - 4{,}5)^4\right] \approx 1{,}78$$

Dado que $g_2 < 3$, la distribución es **platicúrtica**.

Observación 29 Criterios de interpretación de la curtosis

Sea g_2 el coeficiente de curtosis de un conjunto de datos:

- Si $g_2 > 3$, la distribución es **leptocúrtica**.
- Si $g_2 = 3$, la distribución es **mesocúrtica**.
- Si $g_2 < 3$, la distribución es **platicúrtica**.

CAPÍTULO 9

MOMENTOS

Hasta ahora hemos definido varias cantidades importantes en estadística descriptiva: la media aritmética, la varianza, la asimetría y la curtosis. Aunque se han presentado como conceptos distintos, en realidad todas estas medidas son **casos particulares de una idea más general**: la de *momento*.

La noción de momento permite unificar muchas de las definiciones vistas anteriormente y proporciona un lenguaje común para describir la forma de una distribución.

Definición 32 Momento

Sea x una variable estadística que toma k valores distintos $x_1, \ldots, x_k$, con frecuencias $f_1, \ldots, f_k$ y tamaño total $n = \sum_{i=1}^{k} f_i$. Se define el **momento de orden r respecto al parámetro c** como

$$M_r(c) = \frac{1}{n} \sum_{i=1}^{k} (x_i - c)^r \, f_i.$$

Dependiendo de la elección del parámetro c, se obtienen distintos tipos de momentos con interpretaciones estadísticas concretas.

9.1. Momentos respecto al origen

Un caso particular de interés se obtiene al tomar $c = 0$. En este caso, el **momento de orden r respecto al origen** se define como

$$a_r = \frac{1}{n} \sum_{i=1}^{k} x_i^r \, f_i.$$

Los primeros momentos respecto al origen son especialmente ilustrativos:

$$a_0 = \frac{1}{n} \sum_{i=1}^{k} f_i = 1, \qquad a_1 = \frac{1}{n} \sum_{i=1}^{k} x_i f_i = \bar{x}, \qquad a_2 = \frac{1}{n} \sum_{i=1}^{k} x_i^2 f_i.$$

En particular, se observa que **la media aritmética es el momento de primer orden respecto al origen**. Esto muestra que incluso una medida tan básica como la media puede interpretarse dentro del marco general de los momentos.

9.2. Momentos respecto a la media

Si en lugar del origen se toma como referencia la media aritmética, es decir, $c = \bar{x}$, se obtienen los llamados **momentos respecto a la media**:

$$m_r = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^r \, f_i.$$

Los primeros momentos respecto a la media son

$$m_0 = 1, \qquad m_1 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x}) f_i = 0, \qquad m_2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 f_i.$$

El hecho de que $m_1 = 0$ es consecuencia directa de una propiedad fundamental de la media aritmética: las desviaciones respecto a la media se compensan. Además, el momento de orden 2 respecto a la media coincide, salvo un factor de corrección, con la varianza muestral.

> **Observación 30**
>
> Los momentos de orden superior respecto a la media permiten redefinir conceptos como la **asimetría** y la **curtosis** de forma más general. Desde este punto de vista, estas medidas no son objetos aislados, sino parte de una misma familia de descriptores basados en momentos.