Tema 1. Estadística Descriptiva (2ª parte)

Estadística

Ángel Serrano Sánchez de León

Índice

- Introducción
- Variables estadísticas
- Distribuciones de frecuencias
- Introducción a la representación gráfica de datos
- Medidas de tendencia central: media (aritmética, geométrica, armónica, cuadrática), mediana, moda, cuartiles, deciles, percentiles
- Medidas de dispersión: Recorrido (total/intercuartílico/semiintercuartílico), desviación media, desviación típica, varianza, coeficientes de variación
- Momentos. Medidas de asimetría: Sesgo, curtosis
- Variables estadísticas bidimensionales

2

Medidas de resumen o sumarización

- Dada una variable cuantitativa, las **medidas de resumen** permiten expresar toda la información característica de una muestra de datos con menos esfuerzo.
- Tipos:
  - **Medidas de tendencia central:** expresan el valor promedio o central de los datos.
  - **Medidas de dispersión:** indican cómo se reparten los datos respecto del promedio.
  - **Medidas de asimetría:** indican si los datos se reparten de manera simétrica respecto del promedio o están sesgado hacia uno de los dos lados.

Medidas de tendencia central

- También: medidas de centralización.
- Expresan el promedio o valor típico representativo de un conjunto de datos.
- Ejemplos:
  - Media aritmética.
  - Media geométrica.
  - Media armónica.
  - Media cuadrática.
  - Mediana.
  - Moda.
  - Cuartiles, deciles, percentiles.

Media aritmética

- Sea una muestra de tamaño *N* donde la variable *X* toma los valores x_1, x_2, …, x_N.
- Definición:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{\sum_{i=1}^{N} x_i}{N}$$

(se lee "*x* barra")

(se lee "sumatorio desde *i* igual a 1 hasta *N* de *x* sub *i* dividido todo ello por *N*")

Media aritmética

- Conocida la distribución de frecuencias absolutas n_i o relativas f_i se calcula como:

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i x_i}{N} = \sum_{i=1}^{k} \frac{n_i}{N} x_i = \sum_{i=1}^{k} f_i x_i \qquad (k \text{ valores diferentes de } x_i)$$

- Para datos agrupados en intervalos, cada uno con marca de clase c_i y frecuencia absoluta n_i, la media aritmética es aproximadamente:

$$\bar{x} = \frac{\sum_{i=1}^{k} c_i n_i}{N} \qquad (k \text{ marcas de clase } c_i)$$

Media aritmética

- Ejemplo: los pesos en kg (con precisión de 0.01 kg) de unas barras de acero fabricadas en una empresa vienen agrupados en la siguiente tabla:

Intervalos	a_i	a_i+1	c_i	n_i
[7.41, 8.10]	7.405	8.105	7.755	7
[8.11, 8.80]	8.105	8.805	8.455	9
[8.81, 9.50]	8.805	9.505	9.155	2
[9.51, 10,20]	9.505	10.205	9.855	2
[10.21, 10.90]	10.205	10.905	10.555	1

$$\bar{x} = \frac{7.755 \cdot 7 + 8.455 \cdot 9 + 9.155 \cdot 2 + 9.855 \cdot 2 + 10.555 \cdot 1}{21} = \frac{178.955}{21} = 8.52 \text{ kg}$$

Media aritmética: propiedades

- La suma de las desviaciones de un conjunto de datos respecto a su media es cero.

$$\sum_{i=1}^{N}(x_i-\bar{x})=\sum_{i=1}^{N}x_i-\sum_{i=1}^{N}\bar{x}=\sum_{i=1}^{N}x_i-\bar{x}\sum_{i=1}^{N}1=\sum_{i=1}^{N}x_i-N\bar{x}=0$$

- La suma de los cuadrados de las desviaciones de un conjunto de datos respecto a su media es mínima.

$$\sum_{i=1}^{N}(x_i-\bar{x})^2 \text{ es mínimo}$$

- La media aritmética no es **robusta**, porque es muy sensible a valores extremos o anómalos.
  - **Media aritmética recortada:** no se consideran en la media los valores situados en un determinado % superior e inferior de la muestra.

Media aritmética ponderada

- En una media aritmética normal todos los valores tienen la misma importancia.
- Si queremos dar más importancia a unos valores respecto de otros, se les asignan unos pesos w_i, tal que definimos la **media ponderada** como:

$$\bar{x}_W = \frac{\sum_{i=1}^{k} w_i x_i}{\sum_{i=1}^{k} w_i} \qquad (k \text{ valores diferentes de } x_i)$$

- Si los pesos están normalizados tal que su suma total es 1, entonces:

$$\bar{x}_W = \sum_{i=1}^{k} w_i x_i, \quad \sum_{i=1}^{k} w_i = 1$$

Media geométrica

- Para *N* datos x_i, la media geométrica es la raíz *N*-ésima del producto de los datos.

$$\bar{x}_G = \sqrt[N]{x_1 \cdots x_N} = \sqrt[N]{\prod_{i=1}^{N} x_i}$$

(Se lee raíz enésima del productorio desde *i* igual a 1 hasta *N* de equis sub *i*)

- Para datos agrupados con *k* valores diferentes de x_i :

$$\bar{x}_G = \sqrt[N]{x_1^{n_1} \cdots x_k^{n_k}} = \sqrt[N]{\prod_{i=1}^{k} x_i^{n_i}}$$

- Inconvenientes de la media geométrica:
  - Si algún x_i vale 0, la media geométrica también.
  - Si hay valores negativos, la media geométrica podría no estar definida.

Media armónica

- Para N datos x_i, la inversa de la media armónica es la media aritmética de las inversas de x_i. Es decir:

$$\frac{1}{\bar{x}_A} = \frac{\sum_{i=1}^{N} \frac{1}{x_i}}{N} \Rightarrow \bar{x}_A = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}}$$

- Para datos agrupados con k valores diferentes de x_i :

$$\frac{1}{\bar{x}_A} = \frac{\sum_{i=1}^{k} \frac{n_i}{x_i}}{N} \Rightarrow \bar{x}_A = \frac{N}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$$

- Poco afectada por valores grandes de x_i, pero mucho por los valores pequeños.
- Solo válida si todos los valores de x_i son $\neq 0$.

Media cuadrática

- Se define como la raíz cuadrada de la media aritmética de los cuadrados:

$$\bar{x}_Q = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N}}$$

- Para datos agrupados con k valores diferentes de x_i:

$$\bar{x}_Q = \sqrt{\frac{\sum_{i=1}^{k} n_i x_i^2}{N}}$$

- Se aplica con frecuencia en aplicaciones físicas.
- Poco afectada por valores pequeños de x_i.

Relación entre las medias

- En general:

$$\bar{x}_A \le \bar{x}_G \le \bar{x} \le \bar{x}_Q$$

La igualdad ocurre si y solo si todos los números x_1, ..., x_N son idénticos.

Mediana

- Para un conjunto de N datos, la **mediana** M_e es aquel valor tal que la mitad de los valores son menores que él y la otra mitad son mayores.
  - Divide la distribución de frecuencias o el histograma en dos partes iguales.
- Cálculo de la mediana para los valores $x_1 \ldots x_N$:
  - Se ordenan los valores de menor a mayor.

$$\hat{x}_1 \cdots \hat{x}_N, \quad \hat{x}_1 = \min_i x_i, \quad \hat{x}_N = \max_i x_i$$

  - Si N es impar, es el valor central.

$$M_e = \hat{x}_{(N+1)/2}$$

  - Si N es par, es la media aritmética de los dos valores centrales.

$$M_e = \frac{\hat{x}_{N/2} + \hat{x}_{1+N/2}}{2}$$

Mediana

- Sea el vector:

2	6	7	3	9	5	1	8	4

- Son $N = 9$ elementos (impar), luego la mediana es el elemento que, al ordenar el vector, queda en la posición $(9+1)/2 = 5$. En este caso es el valor 5.

- Sea el vector:

5	9	6	7	8	1	3	2	10	4

- Son $N = 10$ elementos (par), luego la mediana es la media aritmética de los elementos, al ordenar el vector, situados en las posiciones $10/2 = 5$ y $1+10/2 = 6$. En este caso: $(5 + 6)/2 = 5.5$.

Mediana para datos agrupados

- Para una variable discreta con N valores agrupados, de los que se calcula la distribución de frecuencias absolutas acumuladas.
  - Si $N/2$ no coincide con ninguna frecuencia absoluta acumulada N_i, la mediana sería el primer valor de x_i con frecuencia absoluta acumulada N_i mayor que $N/2$.

x_i	N_i
1	6
2	13
3	17
4	19
5	20

Estos datos agrupados equivalen a los datos desagrupados siguientes: 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 5

$N = 20$, $N/2 = 10$
Ningún N_i coincide con $N/2$
El primer N_i que supera a $N/2$ es $N_2 = 13$
Luego $M_e = x_2 = 2$

Mediana para datos agrupados

- Para una variable discreta con N valores agrupados, donde $N/2$ coincide con la frecuencia absoluta acumulada N_i de algún valor de x_i, entonces la mediana es la media aritmética entre x_i y x_{i+1}.

x_i	N_i
1	6
2	10
3	15
4	17
5	20

Estos datos agrupados equivalen a los datos desagrupados siguientes: 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 5, 5, 5

$N = 20$, $N/2 = 10$
La frecuencia acumulada N_2 coincide con $N/2$

Luego $M_e = (x_2 + x_3)/2 = (2 + 3)/2 = 2.5$

Mediana para datos agrupados

- Para una variable continua con N valores agrupados en intervalos, de los que se calcula la distribución de frecuencias absolutas acumuladas.

  - Si $N/2$ no coincide con ninguna frecuencia absoluta acumulada, y este valor se encuentra entre las frecuencias N_{i-1} y N_i, correspondientes a los intervalos (a_{i-1}, a_i) y (a_i, a_{i+1}), entonces la mediana se calcula por **interpolación** como:

$$M_e = a_i + \frac{N/2 - N_{i-1}}{n_i}(a_{i+1} - a_i)$$

  - Si $N/2$ coincide con la frecuencia absoluta acumulada N_i del intervalo (a_i, a_{i+1}), entonces la mediana es a_{i+1}, el extremo superior del intervalo.

Mediana para datos agrupados

- Ejemplo: los pesos en kg (con precisión de 0.01 kg) de unas barras de acero fabricadas en una empresa vienen agrupados en la siguiente tabla:

Intervalos	a_i	a_i+1	c_i	n_i	N_i
[7.41, 8.10]	7.405	8.105	7.755	7	7
[8.11, 8.80]	8.105	8.805	8.455	9	16
[8.81, 9.50]	8.805	9.505	9.155	2	18
[9.51, 10,20]	9.505	10.205	9.855	2	20
[10.21, 10.90]	10.205	10.905	10.555	1	21

N = 21
N/2 = 21/2 = 10.5 no coincide con ningún N_i

Este valor está entre N_1 = 7 y N_2 = 16, luego la mediana está en el intervalo [8.11, 8.80]

a_2 = 8.105, a_3 = 8.805

$$M_e = 8.105 + \frac{21/2-7}{9}(8.805-8.105) = 8.38\,\text{kg}$$

Mediana

- Ventajas:
  - Más robusta frente a valores anómalos que la media aritmética.
  - Utilizando un algoritmo inspirado en quicksort se consigue la misma complejidad computacional que en el cálculo de la media aritmética (lineal).
- Desventajas:
  - Precisamente un cambio real en los datos (que no sea anómalo) podría no verse reflejado en la mediana, pero sí en la media.

Moda

- La moda M_o es el valor con la frecuencia máxima (el más repetido).
- Válido tanto para variables categóricas y cuantitativas.
- Inconveniente:
  - Podría no ser única (distribución bimodal, trimodal, multimodal, etc.).
  - Podría no existir (no se repite ningún valor), especialmente para variables cuantitativas (salvo que las agrupemos).

Moda: datos discretos

- En el caso de una variable que toma valores discretos, la moda es simplemente el valor con la mayor frecuencia absoluta (la barra más alta del histograma), es decir, el valor más repetido.
- Es el caso de las variables categóricas o las cuantitativas discretas.
- Ejemplo: analizando el sexo (H, M) de los clientes de una tienda:

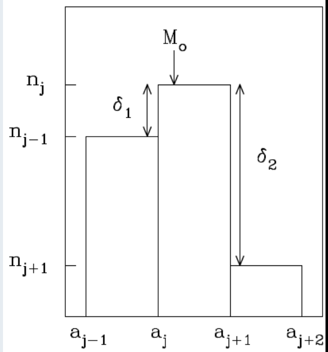
H	M	M	M	H	M	H	M	M	H

- La moda es M (mujer).

Moda: datos agrupados

- En el caso de una variable cuantitativa continua, debemos primero agrupar los valores en intervalos.
- El intervalo de fronteras (a_j, a_{j+1}) con la mayor frecuencia absoluta n_j se llama **intervalo modal**.
- La frecuencia de dicho intervalo supera al intervalo premodal en δ_1, y supera al intervalo postmodal en δ_2.
- La moda se calcula por **interpolación** como:

$$M_o = a_j + \frac{\delta_1}{\delta_1 + \delta_2}\left(a_{j+1} - a_j\right)$$



Moda: datos agrupados

- Ejemplo: los pesos en kg (con precisión de 0.01 kg) de unas barras de acero fabricadas en una empresa vienen agrupados en la siguiente tabla:

En amarillo marcado el intervalo modal

Intervalos	a_i	a_i+1	c_i	n_i
[7.41, 8.10]	7.405	8.105	7.755	7
[8.11, 8.80]	8.105	8.805	8.455	9
[8.81, 9.50]	8.805	9.505	9.155	2
[9.51, 10,20]	9.505	10.205	9.855	2
[10.21, 10.90]	10.205	10.905	10.555	1

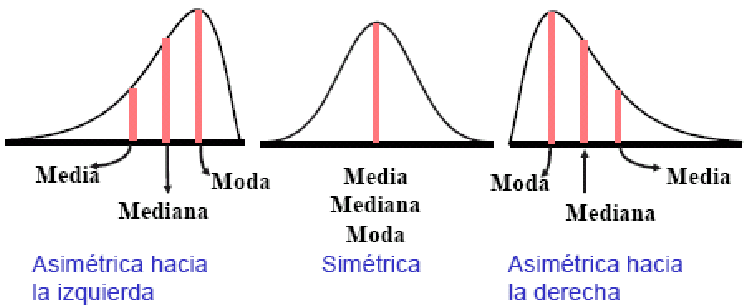
$\delta_1 = 9 - 7 = 2$
$\delta_2 = 9 - 2 = 7$
$a_2 = 8.105$
$a_3 = 8.805$
Luego la moda está en el intervalo [8.11, 8.80]

$$M_o = 8.105 + \frac{2_1}{2+7}(8.105 - 8.805) = 8.26\ \text{kg}$$

Relación entre media, mediana y moda

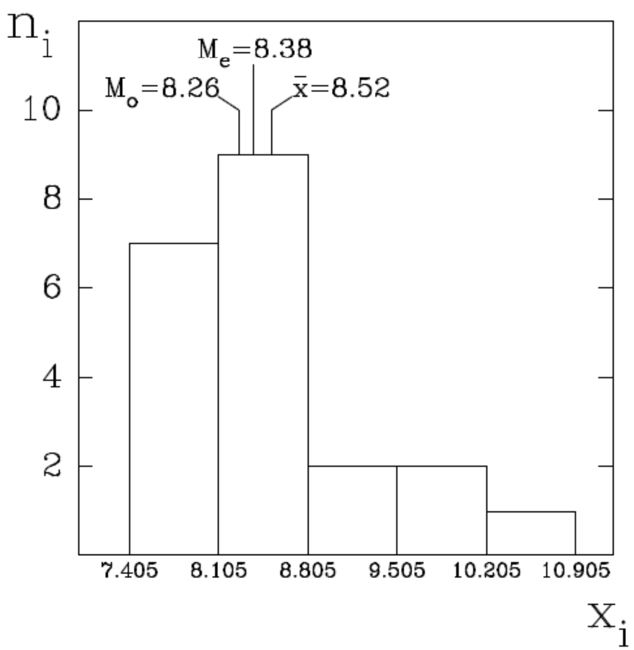
- Para curvas de frecuencias perfectamente simétricas, los tres valores coinciden.
- Para curvas de frecuencias unimodales poco asimétricas, se cumple la regla empírica:

$$\text{media} - \text{moda} = 3\,(\text{media} - \text{mediana})$$



Relación entre media, mediana y moda

- En el ejemplo de las barras de acero, hemos obtenido los siguientes resultados:



Como la media es mayor que la mediana y la moda, la distribución es asimétrica hacia la derecha.

Esto quiere decir que hay una tendencia a valores mayores que la media muy dispersos y a valores menores que la media muy concentrados.

Cuartiles, deciles, percentiles

- **Mediana:** valor que divide a la muestra en 2 partes iguales (50%-50%).
- **Cuartiles:** 3 valores que dividen a la muestra en 4 partes iguales.
  - Primer cuartil $Q_{1/4}$: 25%-75%.
  - Segundo cuartil $Q_{1/2}$: 50%-50% (=M_e).
  - Tercer cuartil $Q_{3/4}$: 75%-25%.
- **Deciles:** 9 valores que dividen a la muestra en 10 partes iguales: $D_1, \ldots, D_9$.
- **Percentiles:** 99 valores que dividen a la muestra en 100 partes iguales: $P_1, \ldots, P_{99}$.

Diagrama de caja y bigotes

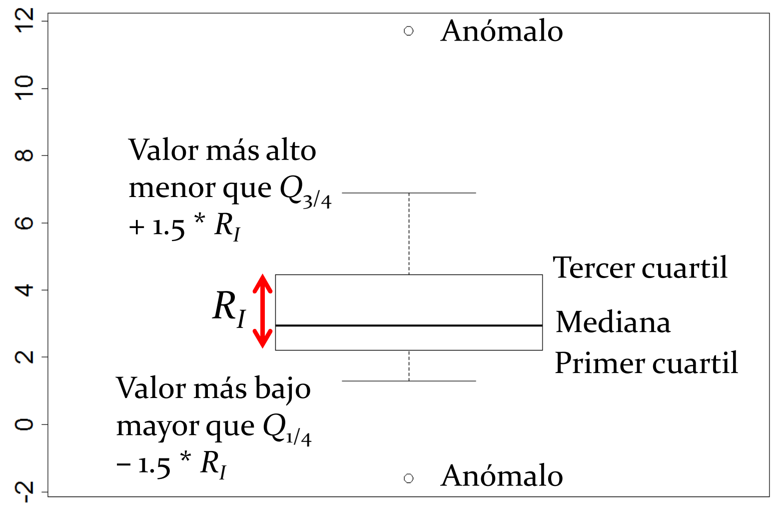
- Representación gráfica de la posición de los cuartiles y de la presencia de valores atípicos.
- Caja (rectángulo) formada por el primer cuartil $Q_{1/4}$ y por el tercer cuartil $Q_{3/4}$.
- La mediana se indica con una línea que divide la caja en dos.
- Normalmente los bigotes se extienden a ambos lados desde el valor mínimo hasta el máximo.
- Los valores situados más allá de la posición indicada por el cuartil correspondiente más 1.5 veces el recorrido intercuartílico (diferencia entre $Q_{3/4}$ y $Q_{1/4}$) son marcados como anómalos (circulitos).

Diagrama de caja y bigotes

6.9	2.9	3.2	1.7	2.9	6.8	1.3	4.4	2.3	2.1
5.6	3.1	3.0	11.7	2.3	4.5	1.3	3.7	2.6	-1.6

-1.6	1.3	1.3	1.7	2.1	2.3	2.3	2.6	2.9	2.9
3.0	3.1	3.2	3.7	4.4	4.5	5.6	6.8	6.9	11.7

$Q_{1/4} = 2.250$
$Q_{1/2} = M_e = 2.950$
$Q_{3/4} = 4.425$
$R_I = Q_{3/4} - Q_{1/4} = 2.175$
Límite teórico bigote inferior = -1.0125
Límite bigote inferior = 1.3
Límite teórico bigote superior = 7.6875
Límite bigote superior = 6.9



Medidas de dispersión

- Las medidas de centralidad reducen la muestra de datos a un solo valor.
- Las **medidas de dispersión** indican cómo se reparten o dispersan los valores de la muestra respecto de ese valor central.
- Ejemplos:
  - Recorrido (o rango)
  - Recorrido intercuartílico, semiintercuartílico
  - Desviación media absoluta
  - Desviación típica y Varianza

Recorrido (o rango)

- Para un conjunto de N datos, el **recorrido** es:

$$R = \text{máx}(x) - \text{mín}(x)$$

- Variante: para eliminar la influencia de valores extremos, se define el **recorrido intercuartílico** (rango del 50% central de los datos):

$$R_I = Q_{3/4} - Q_{1/4}$$

- Variante: **recorrido semiintercuartílico** (mitad del anterior):

$$R_{S\,I} = \frac{Q_{3/4} - Q_{1/4}}{2}$$

Desviación absoluta

- **Desviación absoluta respecto de la media aritmética:** media aritmética de las diferencias absolutas entre los valores de la variable y la media aritmética de la muestra.

$$D_{\bar{x}} = \frac{\sum_{i=1}^{N} |x_i - \bar{x}|}{N} \qquad D_{\bar{x}} = \frac{\sum_{i=1}^{k} |x_i - \bar{x}| n_i}{N}$$

- **Desviación absoluta respecto de la mediana:**

$$D_{M_e} = \underset{i}{\text{mediana}}\left(|x_i - M_e|\right)$$

32

Desviación típica y Varianza

- Las medidas de dispersión más importantes.
- **Varianza:** Media aritmética de las desviaciones cuadráticas respecto de la media aritmética.

$$s^2 = s_x^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N} \qquad s^2 = s_x^2 = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x})^2}{N}$$

- **Desviación típica:** Raíz cuadrada de la varianza.

$$s = s_x = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}} \qquad s = s_x = \sqrt{\frac{\sum_{i=1}^{k} n_i (x_i - \bar{x})^2}{N}}$$

Desviación típica y Varianza

- Si solo tenemos un dato (N=1), la definición anterior de desviación típica da 0, pero no tiene sentido el concepto de "dispersión de un solo dato".
- Como veremos en el tema de Inferencia Estadística, para estimar la desviación típica de una población a partir de la desviación típica de una muestra, es mejor corregir el denominador (**desviación típica y varianza "sin sesgo"**):

$$\hat{s}^2 = \hat{s}_x^2 = \frac{\sum_{i=1}^{N} (x_i - \overline{x})^2}{N-1} \qquad \hat{s} = \hat{s}_x = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \overline{x})^2}{N-1}}$$

Desviación típica y Varianza

- Relación entre ambas definiciones:

$$\hat{s}^2 = \frac{N}{N-1}s^2 \qquad \hat{s} = \sqrt{\frac{N}{N-1}}s$$

- Para valores altos de *N* (>30), ambas definiciones coinciden en la práctica.

Desviación típica y Varianza

- Se puede demostrar fácilmente (¡ejercicio!) que la varianza sesgada es igual a la media de los cuadrados de x menos el cuadrado de la media de los valores de x.

$$s^2 = \overline{x^2} - \bar{x}^2 = \frac{\sum_{i=1}^{N} x_i^2}{N} - \left(\frac{\sum_{i=1}^{N} x_i}{N} \right)^2$$

Desviación típica y Varianza

- Propiedades:
  - Ambas son siempre positivas o cero.
  - La desviación típica tiene las mismas unidades que los valores de x, la varianza tiene unidades al cuadrado.
  - Usando la desviación cuadrática respecto de la media aritmética como medida de dispersión, ambas dan la desviación cuadrática mínima.
  - Afectadas por valores extremos (mejor el recorrido intercuartílico).
  - Importancia por su conexión con la distribución normal (gaussiana).

Variables tipificadas

- Dada una variable x, que toma N valores, con media $\bar{x}$ y desviación típica s, podemos realizar un cambio de variables a **unidades estándar** (**variables tipificadas**):

$$z = \frac{x - \bar{x}}{s}$$

- Variable adimensional.
- De gran importancia para la distribución normal (gaussiana).

Relación entre medidas de dispersión

- Para datos poco sesgados (bastante simétricos), se cumplen estas reglas empíricas:
  - Relación entre la desviación media y la desviación típica:

$$D_{\bar{x}} = \frac{4}{5} s$$

  - Relación entre el recorrido semiintercuartílico y la desviación típica:

$$R_{SI} = \frac{2}{3} s$$

Coeficientes de variación

- La desviación típica es una medida de dispersión absoluta (tiene las mismas unidades que los datos originales).
- El **coeficiente de variación de Pearson** la expresa de manera relativa respecto del valor medio:

$$CV = \frac{s}{|\bar{x}|} \quad \text{(en tanto por uno)} \qquad CV = \frac{s}{|\bar{x}|} \cdot 100 \quad \text{(en \%)}$$

- Positivo o cero, igual que *s*.
- Independiente de las unidades (comparación de variables con unidades diferentes).
- Inconveniente: no válido si el promedio de *x* es próximo a cero.

Coeficientes de variación

- Definición alternativa basada en la desviación media respecto de la media:

$$CVM_{\bar{x}} = \frac{D_{\bar{x}}}{|\bar{x}|}$$

- Definición alternativa basada en la desviación media respecto de la mediana:

$$CVM_{M_e} = \frac{D_{M_e}}{|M_e|}$$

Medidas de asimetría

- Dan cuenta de cómo de simétricos son los datos respecto del valor central.
- Ejemplos:
  - Asimetría o sesgo.
  - Curtosis.

Momento

- Generalizando la definición de varianza, se denomina **momento de orden *r* respecto de la media**:

$$m_r = \frac{\sum_{i=1}^{N} \left(x_i - \bar{x}\right)^r}{N} \qquad m_r = \frac{\sum_{i=1}^{k} n_i \left(x_i - \bar{x}\right)^r}{N}$$

- Propiedades:

$$m_0 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^0}{N} = \frac{\sum_{i=1}^{N} 1}{N} = \frac{N}{N} = 1 \qquad m_1 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})}{N} = \frac{\sum_{i=1}^{N} x_i}{N} - \frac{\bar{x}\sum_{i=1}^{N} 1}{N} = \bar{x} - \bar{x} = 0$$

$$m_2 = \frac{\sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2}{N} = s^2 = \frac{N-1}{N}\hat{s}^2$$

43

Coeficientes de asimetría (o sesgo)

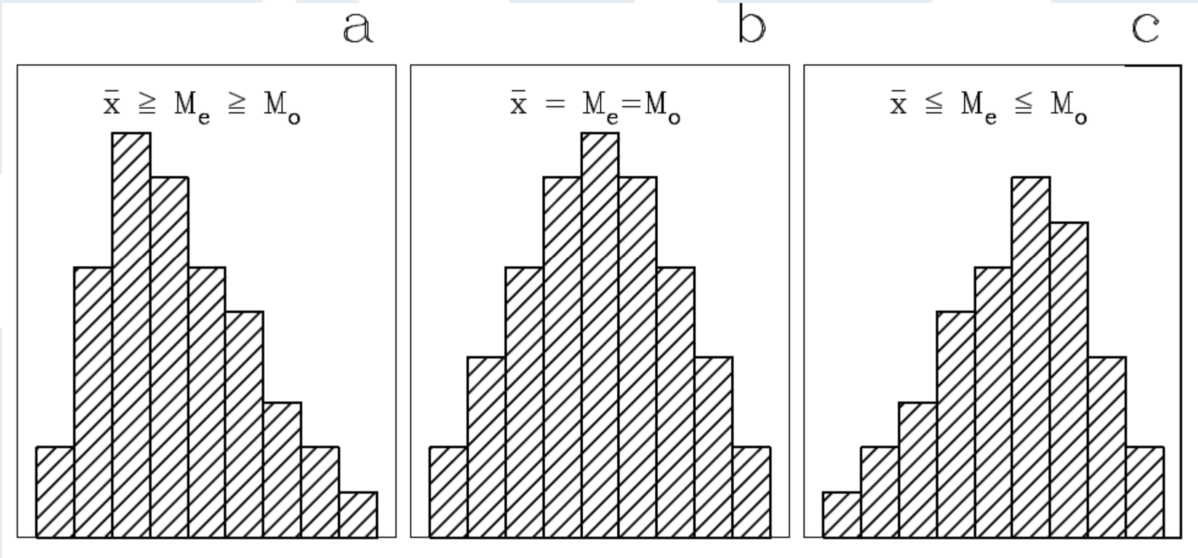
- **1er y 2º coeficientes de asimetría de Pearson** (adimensionales):

$$A_{P_1} = \frac{\bar{x} - M_o}{s} \qquad A_{P_2} = \frac{3(\bar{x} - M_e)}{s}$$

- **Coeficiente de asimetría de Fisher** (adimensional):

$$g_1 = \frac{m_3}{s^3}$$

Coeficientes de asimetría (o sesgo)



Asimetría hacia la derecha (o positiva), o cola derecha

$g_1 > 0$

Simétrica

$g_1 = 0$

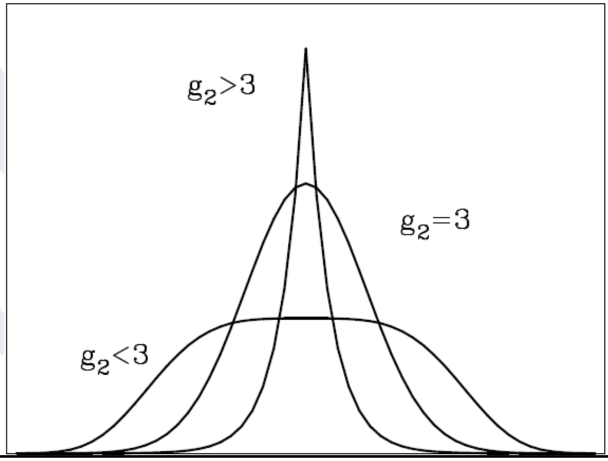
Asimetría hacia la izquierda (o negativa), o cola izquierda

$g_1 < 0$

Curtosis

- Compara la distribución de los valores respecto de la distribución normal (gaussiana). Tipos:
  - Leptocúrtica (picuda): $g_2 > 3$.
  - Mesocúrtica (como la distribución normal): $g_2 = 3$.
  - Platicúrtica (achatada): $g_2 < 3$.

$$g_2 = \frac{m_4}{s^4}$$



Exceso de curtosis = $g_2 - 3$

46