Facultad de Ingeniería
OBERA

ESTADÍSTICA APLICADA A LA INVESTIGACIÓN CIENTÍFICA

Unidad 2:
Estadística descriptiva aplicada

EL PENSAMIENTO ESTADÍSTICO

Puede entenderse como la capacidad para **observar de manera integral el problema**, para tomar en cuenta factores relevantes como el contexto, la fuente de los datos, el método de muestreo, y además, para interpretar resultados e identificar implicaciones prácticas.

El pensamiento estadístico **NO** es simplemente tener la habilidad de ejecutar cálculos complejos.

Propósito de la estadística

La estadística tiene muchos usos, pero quizás el más importante es que ayuda a **tomar decisiones** en situaciones que incluyen ***incertidumbre***.

ESTADÍSTICA

Es la ciencia que se encarga de planear estudios y/o experimentos, obtener, organizar, y resumir los datos, para luego generar, analizar e interpretar la información, y finalmente extraer conclusiones.

Población: es el conjunto completo de todos los elementos puntuaciones, personas, mediciones, etc.) que son sometidos a estudio.

Muestra: es un subconjunto de elementos seleccionados de una población.

Un ***parámetro*** es una medida numérica que describe una característica de una población.

Un ***estadístico*** es una medida numérica que describe una característica de una muestra.

Razonar

Para cada una de las situaciones siguientes, describa la población que se estudiará e identifique algunos de los parámetros de la población que serían de interés.

a. Usted trabaja para Seguros del Agricultor y ha investigado para determinar el monto promedio pagado a víctimas de accidente en automóviles sin bolsas de aire para impactos laterales.

b. Ha sido contratado por McDonald's para determinar los pesos de las papas enviadas cada semana para producir las papas fritas.

c. Usted es un reportero de negocios que está cubriendo los Laboratorios Genentech, e investiga si su nuevo tratamiento es efectivo contra la leucemia infantil.

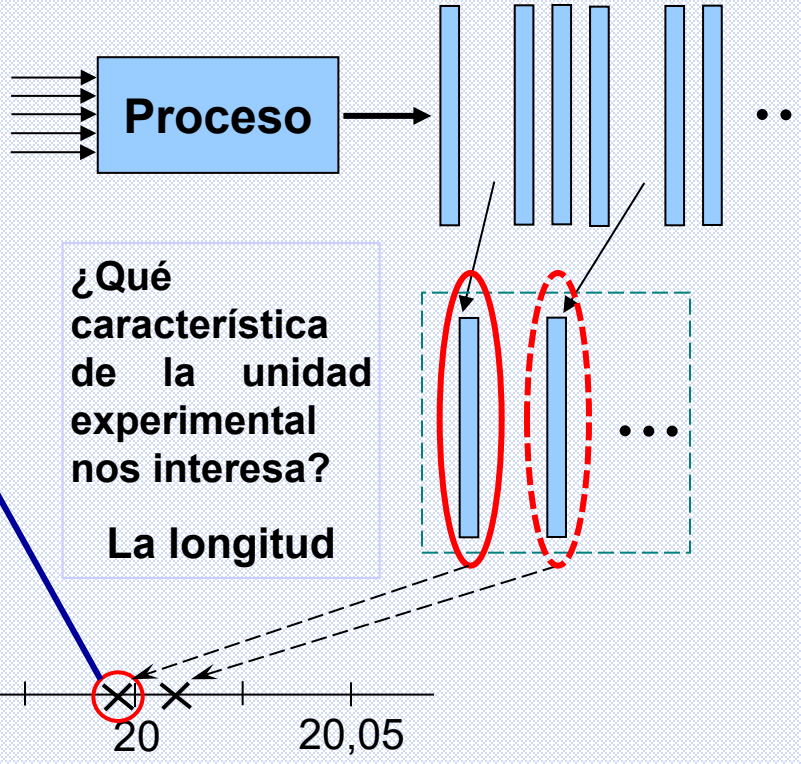
RECOLECCIÓN DE DATOS

Estudio observacional: vemos y medimos características específicas, pero no intentamos modificar a los sujetos que estamos estudiando

Experimento: aplicamos algunos tratamientos y luego procedemos a observar sus efectos sobre los sujetos. En los experimentos, los sujetos se denominan ***unidades experimentales***

Una ***unidad experimental*** es aquella sobre la cual se efectúan mediciones o se intenta clasificar en categorías

Un ***dato*** es toda característica de una unidad experimental que *se registra* en una *escala* durante un proceso de medición u observación.



MUESTREO (1)

Una ***muestra representativa*** es aquella en la cual las características relevantes de sus elementos, generalmente son muy similares a las características bajo estudio de la población.

No se debe aceptar a ciegas los cálculos matemáticos; es necesario tomar en cuenta los siguientes factores:

- Contexto de los datos
- Fuente de los datos
- Método de muestreo
- Conclusiones
- Implicaciones prácticas

Para aprender a ***pensar en términos estadísticos***, suelen ser más importantes el sentido común y las consideraciones prácticas, que la **aplicación irreflexiva** de expresiones matemáticas y sus cálculos derivados.

MÉTODOS DE MUESTREO (1)

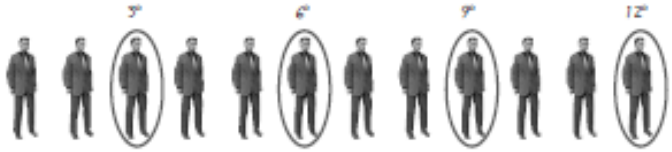
Métodos comunes de muestreo

- **Muestreo aleatorio simple:** elegimos una muestra de elementos de tal manera que cualquier muestra del mismo tamaño tenga igual oportunidad de ser seleccionada.
- **Muestreo sistemático:** utilizamos un sistema sencillo para seleccionar la muestra, tal como elegir a cada 10 o cada 50 elementos de la población.
- **Muestreo por conveniencia:** utilizamos una muestra que pueda ser conveniente de seleccionar.
- **Muestreo por conglomerados:** primero dividimos la población en grupos, o conglomerados, y seleccionamos al azar algunos de estos conglomerados. Luego obtenemos la muestra mediante la selección de *todos* los miembros en cada uno de los conglomerados seleccionados.
- **Muestreo estratificado:** utilizamos este método cuando nos interesan las diferencias entre subgrupos, o *estratos*, dentro de una población. Primero identificamos los estratos y luego sacamos una muestra aleatoria en cada uno de los estratos. La muestra total consiste en todas las muestras de cada uno de los estratos individuales.

MÉTODOS DE MUESTREO (2)

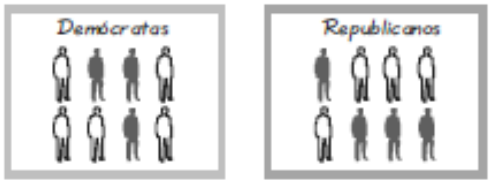
Muestreo aleatoria:
Cada miembro de la población tiene la misma probabilidad de resultar seleccionado. A menudo se usan computadoras para generar números telefónicos aleatorios.

Muestreo aleatorio simple:
Se selecciona una muestra de n sujetos, de manera que cada posible muestra del mismo tamaño n tenga la misma posibilidad de ser elegida.

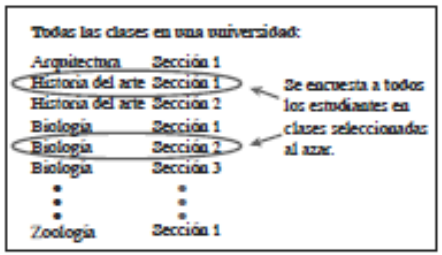


Muestreo sistemática:
Se selecciona un punto de partida, después se elige cada k-ésimo (por ejemplo, cada quincuagésimo) elemento de la población.

Muestreo de conveniencia:
Se utilizan resultados que son fáciles de obtener.



Muestreo estratificada:
Se subdivide a la población en al menos dos subgrupos (o estratos) diferentes, de manera que los sujetos del mismo subgrupo compartan las mismas características (como el género o la categoría de edad), y después se obtiene una muestra de cada subgrupo.



Muestreo por conglomerados:
Se divide a la población en secciones (o conglomerados), luego se eligen al azar algunos de estos conglomerados, y después se elige a todos los miembros de los conglomerados seleccionados.

MÉTODOS DE MUESTREO (3)

RAZONAR **Métodos de muestreo**

Identifique el tipo de muestreo utilizado en cada uno de los casos siguientes:

a. La cosecha de manzanas en un huerto se recolecta en 1 200 canastos. Un inspector de agricultura selecciona 25 canastos al azar y luego revisa si cada manzana en estos canastos tiene gusanos.

b. Una investigadora educativa quiere conocer si, en una escuela en particular, los hombres y las mujeres tienden a realizar más preguntas en clase. De los 10 000 estudiantes en el colegio, ella entrevista a 50 hombres y a 50 mujeres elegidas al azar.

c. Al tratar de aprender si los sistemas planetarios son comunes, los astrónomos realizan un sondeo buscando planetas entre 100 de las estrellas más cercanas.

d. Para determinar quién ganará balones de futbol autografiados, un programa de cómputo selecciona aleatoriamente los números de boleto de 11 personas en un estadio lleno con gente.

TIPOS DE DATOS (1)

Cualitativos (o categóricos o de atributo)

Nominales: sus valores se consignan mediante etiquetas o nombres que no se pueden ordenar (profesión; residencia; nacionalidad; etc.)

Ordinales: sus valores se consignan mediante etiquetas o nombres que se ordenan (clasificación de estaturas: baja, media y alta; clasificación de exámenes: malo, regular, y bueno; etc.)

Cuantitativos (o numéricos)

Discretos: corresponden a valores separados entre sí, por alguna cantidad, y surgen por conteo. Pueden comprender un número finito ó infinito de valores enteros (Número de defectos en un alambre de cobre, etc.)

Continuos: surgen de mediciones efectuadas sobre cada unidad experimental. Pueden tomar infinitos valores, que corresponden a una escala continua que cubre un rango de valores sin interrupciones o saltos (Longitud de varillas de acero, etc.)

TIPOS DE DATOS (2)

ANALIZAR Tipos de datos

Clasifique cada uno de los conjuntos de datos siguientes como cualitativos o cuantitativos.

a. Nombres de marcas de zapatos en una encuesta a consumidores.

b. Calificaciones en un examen de opción múltiple.

c. Calificaciones con letras en una tarea de ensayo.

d. Números en los uniformes que identifican a los jugadores en un equipo de básquetbol.

ESCALAS DE MEDICIÓN (o niveles de medida) (1)

Una ***medición*** es la asignación de un número ó código al registro obtenido a partir de la valoración de una característica de la unidad experimental.

Una ***escala*** es la herramienta sobre la que se referencian los registros obtenidos en una observación o medición. Un registro sobre una escala implica un ***dato***.

1 - PARA DATOS CUALITATIVOS o CATEGÓRICOS o DE ATRIBUTO

Escala nominal: los datos se clasifican en categorías (Ejemplos: tipo de actividad laboral, tipo de religión, etc.)

Escala ordinal: los datos muestran las propiedades de los datos nominales, pero además se establece un orden o jerarquía de esos datos (Ejemplos: nivel de instrucción: muy bueno, bueno, regular, malo; clase social: alta, media, baja; etc.)

ESCALAS DE MEDICIÓN (o niveles de medida) (2)

2 - PARA DATOS CUANTITATIVOS o NUMÉRICOS

Escala de intervalo: los datos tienen las propiedades de los datos ordinales, pero además se asignan distancias entre ellos. Establece intervalos iguales y números consecutivos de registro. Este tipo de datos siempre es numérico, y el valor cero no indica la ausencia de la característica evaluada (Ejemplos: lapsos de tiempo de producción, temperatura en ºC de una ciudad, etc.)

Escala de razón: los datos tienen todas las propiedades de los datos de intervalo, y la proporción entre ellos tiene sentido. Se requiere que el valor cero de la escala indique la ausencia de la propiedad a medir (Ejemplos: peso de una persona, estatura de una persona, tiempo utilizado para una tarea, temperatura en ºK de una ciudad, etc.)

ESCALAS DE MEDICIÓN (o niveles de medida) (3)

3 – ESCALA DE LIKERT

En teoría pura de medición, la escala Likert es **ordinal** (los números representan un orden, pero no garantizan distancias iguales).
Sin embargo, **en investigación cuantitativa aplicada** se asume que, con suficientes puntos (7 o más), las distancias entre categorías son aproximadamente iguales, lo que permite **tratarlas como datos de intervalo**.

En el enfoque de investigación cuantitativa, una encuesta que utilice una escala de Likert de 0 a 10 se considera una herramienta válida, ya que produce datos numéricos que pueden analizarse estadísticamente. Aunque la escala Likert es originalmente ordinal, al contar con once puntos de respuesta se aproxima a una escala de intervalo, lo que permite calcular medidas como la media, la mediana, la desviación estándar y aplicar pruebas inferenciales. Este tipo de instrumento transforma percepciones o actitudes subjetivas en valores cuantificables, lo que favorece la objetividad, la comparabilidad entre individuos y la replicabilidad del estudio.

En resumen:
El único cuidado es ser consciente del origen y la naturaleza de la escala, por lo que se aconseja justificar su uso como intervalo en el marco metodológico.

ESCALAS DE MEDICIÓN (o niveles de medida) (3)

RAZONAR **Niveles de medida**

Identifique el nivel de medida (nominal, ordinal, intervalo, razón) para cada uno de los conjuntos de datos siguientes.

a. Números en uniformes que identifican a los jugadores en un equipo de básquetbol.

b. Clasificación que los estudiantes dan a los alimentos de la cafetería como excelente, bueno, regular o malo.

c. Años del calendario para acontecimientos históricos, tal como 1776, 1945 o 2001.

d. Temperatura en la escala Celsius (o centígrada).

e. Tiempos de los corredores en la maratón de Boston.

DATOS / INFORMACIÓN

Datos = Hechos, realidades

Información = Respuestas a preguntas

La "Información incluye "datos"

Los "datos" no necesariamente incluyen "información"

La clave no está en como obtener los datos, sino en, para qué se los obtiene (que información se necesita)

El proceso de generar información:

1. Formular claramente la pregunta que se desea contestar.
2. Recopilar datos relativos a la cuestión planteada.
3. Analizar los datos para determinar la respuesta a la pregunta formulada.
4. Presentar los datos de forma adecuada para responder claramente cuál es la respuesta a la pregunta (**implica obtener información**).

MEDIDAS DESCRIPTIVAS NUMÉRICAS

1) Medidas de tendencia central

2) Medidas de dispersión

3) Medidas de orden

1) MEDIDAS DE TENDENCIA CENTRAL

Son promedios de un conjunto de datos. Los valores obtenidos tienden, en general, a situarse en el centro del conjunto ordenado. Se destacan: la media aritmética, la mediana y el modo.

A) MEDIA ARITMÉTICA (Para datos cuantitativos)

a) *Para datos sin agrupar*

Se la define como el cociente entre la suma de todos los valores ($x_1 + x_2 + ...x_n$) y el número total de ellos (n).

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

b) *Para datos agrupados en una tabla simple de frecuencias*

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

Donde

x_i: valores que asume la variable

f_i: frecuencias absolutas de los valores

n: núm. total de datos

Resistencia (ohms)	Tabulación	f	F
3,35	|	1	1
3,36	| |	3	4
3,37	| | | |	5	9
3,38	| | |	4	13
3,39	| |	2	15

Ejemplo:

Los datos obtenidos representan las medidas de la resistencia eléctrica (ohms) de 15 bobinas.

c) *Para datos agrupados en intervalos de clases*

$$\overline{X} = \frac{\sum_{i=1}^{K} Mc \,.\, f_i}{n}$$

Donde

Mc : marca de clase

f_i : frecuencia del intervalo

n : núm. total de datos

Ejemplo:

Los datos representan la resistencia a la tensión, en libras por pulgada cuadrada (psi), de 80 muestras de una nueva aleación de aluminio y litio, que está siendo evaluada como posible material para la fabricación de elementos estructurales de aeronaves.

Intervalo de clase (Psi)	Mc	f	F	fr	Fr
$70 \leq x < 90$	80	2	2	0,025	0,025
$90 \leq x < 110$	100	3	5	0,038	0,062
$110 \leq x < 130$	120	6	11	0,075	0,138
$130 \leq x < 150$	140	14	25	0,175	0,312
$150 \leq x < 170$	160	22	47	0,275	0,587
$170 \leq x < 190$	180	17	64	0,212	0,801
$190 \leq x < 210$	200	10	74	0,125	0,925
$210 \leq x < 230$	220	4	78	0,050	0,975
$230 \leq x < 250$	240	2	80	0,025	1,000
Total		80		1,000	

B) MEDIANA

Es la medida que ocupa la posición central de la distribución, en la cual los datos deben estar organizados en ***orden de magnitud*** (creciente o decreciente). Es el valor que divide a los datos en dos partes iguales.

Posición **Valor**

I. Datos *cualitativos – Ordinales* (Me)

Posición Me
- Datos impares: Posición = $\frac{n+1}{2}$
- Datos pares
  - Posición 1 = $\frac{n}{2}$
  - Posición 2 = $\frac{n}{2}+1$

a) *Para datos sin agrupar*

Ejemplo: Respuesta a encuesta sobre libro:
M, M, R, R, B, B, B, B, MB, MB, E .-

Posición 6 => **Me** = B

b) *Para datos agrupados*

Procedimiento:

1- Detectar la posición donde se encuentra la Me.

2- Identificar la categoría en la que se encuentra el valor.

Ejemplo: Encuesta a usuarios del servicio de transporte urbano publico de pasajeros.

Respuestas	Usuarios (f)	F
Muy malo	7	7
Malo	25	32
Regular	131	163
Bueno	327	490
Muy bueno	35	**525**
TOTAL	**525**	

1- Posición (*n* es impar): $\frac{n+1}{2}=\frac{525+1}{2}=263$

2- Categoría donde está la **Posición Me**

Me = Bueno

II. Datos cuantitativos (Me)

$$\text{Me} = \begin{cases} x_{(n+1)/2}, & \text{impar} \\ \dfrac{x_{(n/2)} + x_{(n/2)+1}}{2}, & \text{par} \end{cases}$$

$$\text{Posición} = \frac{n+1}{2}$$

a) *Para datos sin agrupar*

Ejemplo a: Determinar la mediana de los valores siguientes: 3, 1, 5, 17, 8, 12, 15

Ejemplo b: Determinar la mediana de los valores siguientes: 3, 1, 5, 17, 8, 12, 15, 20.

b) *Para datos agrupados en intervalos de clases*

Ejemplo:

a. $Posición = \dfrac{x_{(n/2)} + x_{(n/2)+1}}{2} = \dfrac{x_{(20)} + x_{(21)}}{2} = x_{(20,5)}$

$$Posición = \frac{40+1}{2} = 20{,}5$$

Interv. de clase	f	F
299 ≤ x < 685	11	11
20,5 (Posición Me) → 685 ≤ x < 1071	13	24
1071 ≤ x < 1457	6	30
1457 ≤ x < 1843	5	35
1843 ≤ x < 2229	3	38
2229 ≤ x < 2615	1	39
2615 ≤ x < 3001	1	**40**
	40	

b. La mediana está dentro de la clase 685 ≤ x < 1071,

c. Se calcula como: $\mathbf{Me} = \boldsymbol{c} + \dfrac{(Posición - F_a) \cdot h}{f_c}$

Donde:

c: extremo inferior de la clase a la cual pertenece la Me

F_a: frecuencia acumulada de la clase anterior a la que pertenece la Me

h: amplitud del intervalo de clase

f_c: frecuencia absoluta de la clase a la que pertenece la Me

Me = 967,076

C) MODO

Es aquel valor de la variable que presenta la mayor frecuencia en una distribución.

I. *Datos cualitativos – Nominales* (Mo)

a) *Para datos agrupados*

Ejemplo: En un análisis de profesiones se tiene:

Mo →

Profesión	Cantidad
Abogados	25
Ingenieros	45
Contadores	31
Total	101

II. *Datos cualitativos – Ordinales* (Mo)

a) *Para datos sin agrupar*

Ejemplo: Respuesta a encuesta sobre libro: M, M, R, R, B, B, B, B, MB, MB, E .-

Dato más frecuente

Mo = B

b) *Para datos agrupados*

Ejemplo: Encuesta a usuarios del servicio de transporte urbano publico de pasajeros.

Mo = Bueno

Mo →

Respuestas	Usuarios
Muy malo	7
Malo	25
Regular	131
Bueno	327
Muy bueno	35
TOTAL	**525**

III. *Datos cuantitativos* (Mo)

a) Para datos sin agrupar

Ejemplo:

Considérense los siguientes datos: 2, 3, 4, 4, 5, 6, 6, 6, 6, 6, 8, 8, 9, 9, 9 (Dato más frecuente)

Mo = 6

b) Para datos agrupados en intervalos de clase

El método más usado es el que se calcula mediante la fórmula siguiente:

$$Mo = c + \frac{f_2}{f_1 + f_2} \cdot h$$

Donde:

c : extremo inferior de la clase modal.

f_1 : frecuencia absoluta de la clase anterior a la modal

f_2 : frecuencia absoluta de la clase posterior a la modal

h : amplitud de la clase modal

$$Mo = 685 + \frac{6}{11+6} \cdot 386$$

Mo = 821,23

Ejemplo:

Intervalos de clase	f	F
299 ≤ x < 685	11	11
685 ≤ x < 1071	13	24
1071 ≤ x < 1457	6	30
1457 ≤ x < 1843	5	35
1843 ≤ x < 2229	3	38
2229 ≤ x < 2615	1	39
2615 ≤ x < 3001	1	40
	40	

Clase modal → 685 ≤ x < 1071

ACTIVIDAD AULICA 6

Consideraciones generales

1. La $\overline{X}$ utiliza más información que la Me, por cuanto al calcular la $\overline{X}$ nos servimos de la totalidad de los datos, en tanto que la Me sólo de los datos centrales.

2. La $\overline{X}$ resulta afectada por el cambio de los valores extremos, en tanto que la Me permanece inalterable, a menos que cambien los valores de los datos centrales.

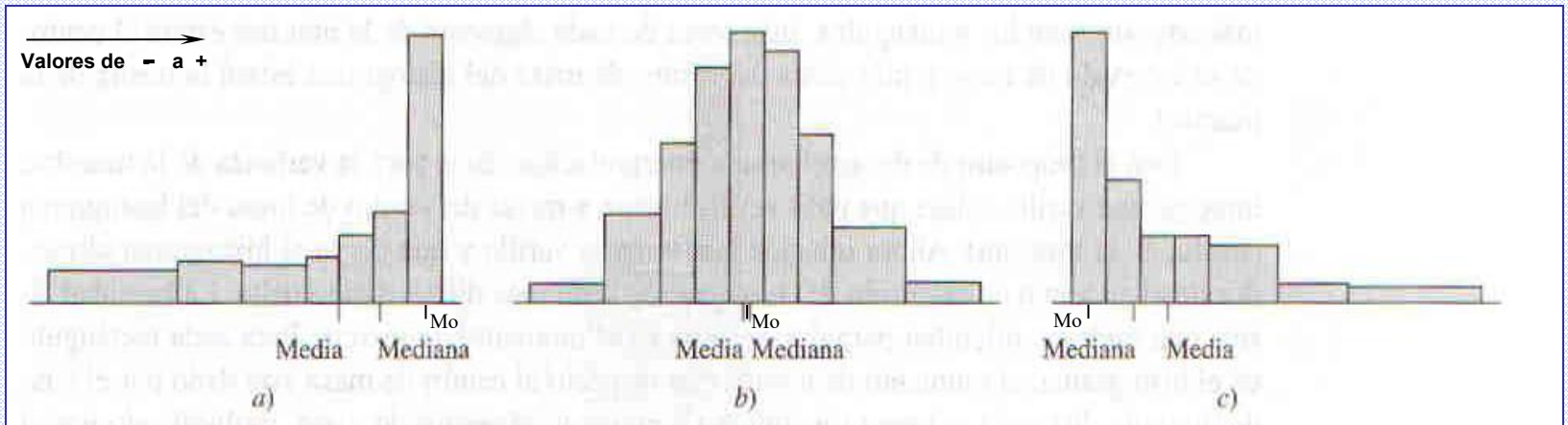
3. Siempre que una distribución sea fuertemente asimétrica, la Me será por lo general más apropiada que la $\overline{X}$.

4. Si los datos son simétricos, entonces la media y la mediana coinciden o están próximos. Si, además, los datos tienen sólo un modo (esto es, son u*nimodales*), entonces la media, la mediana y el modo coinciden o están próximos.

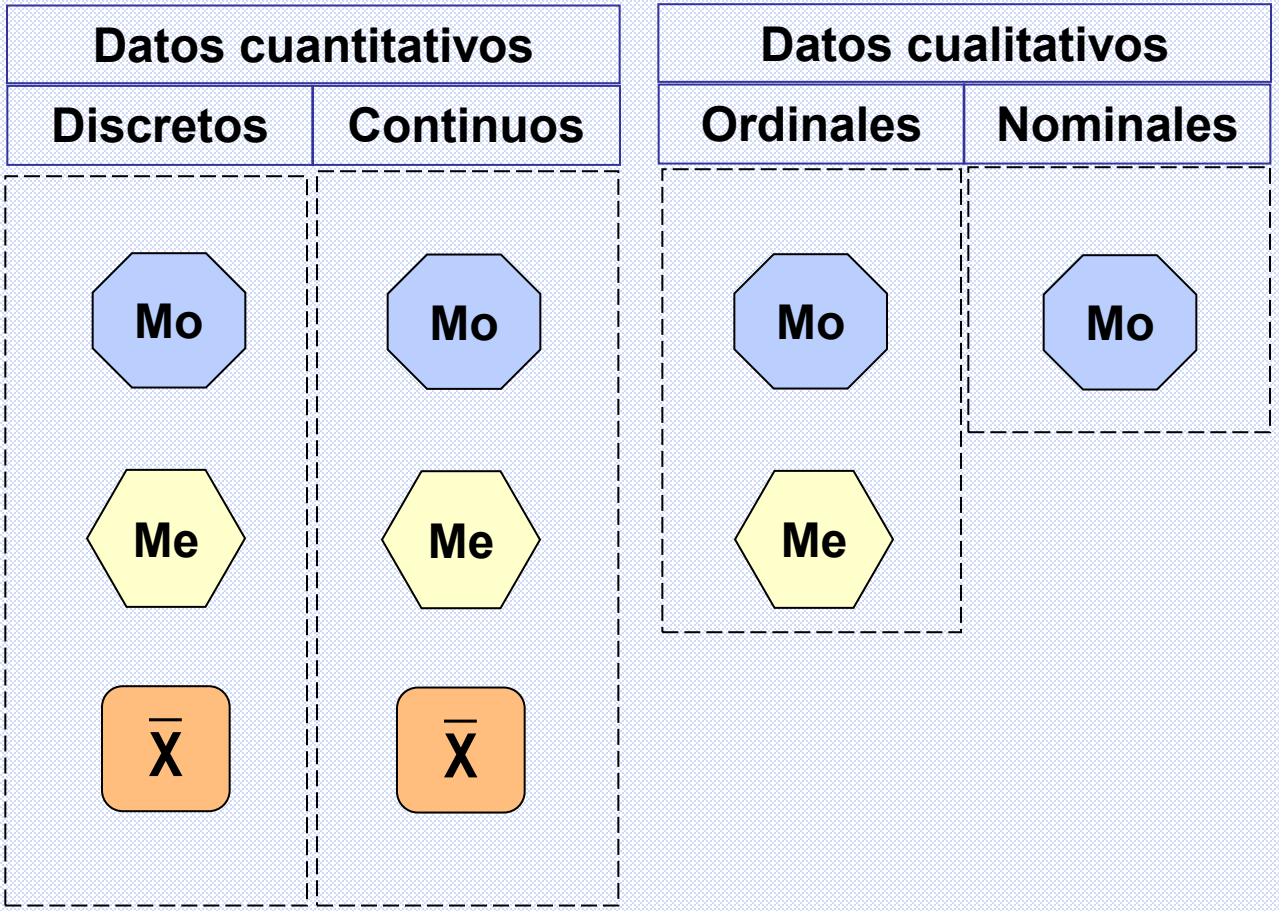
5. Si los datos son sesgados (esto es, son *asimétricos*, con una larga cola en uno de los extremos), entonces la media, la mediana y el modo no coinciden. Generalmente se encuentra que:

a) "modo $<$ mediana $<$ media" si la distribución está *sesgada hacia la derecha*;

b) "media < mediana < modo" si la distribución está *sesgada hacia la izquierda*.



Medidas de tendencia central que caracterizan a un conjunto de datos



2) MEDIDAS DE DISPERSIÓN – DATOS CUANTITATIVOS

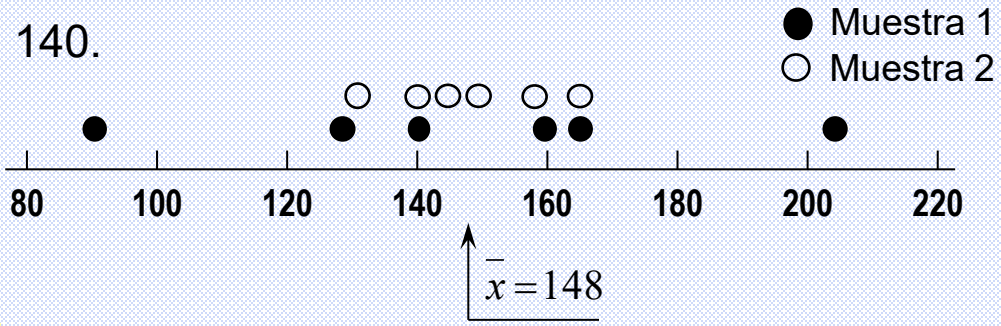
Dan idea de la separación o concentración de los valores de una variable alrededor de su media aritmética u otro promedio.

Ejemplo:

Considérese los datos de resistencia a la tensión (en psi) de dos muestras de aleación de aluminio-litio:

Muestra 1: 90, 128, 205, 140, 165, 160.

Muestra 2: 130, 150, 145, 158, 165, 140.



A) RANGO

Se define como la diferencia entre los valores extremos x_n y x_1 del conjunto de datos

$$R = x_n - x_1$$

Donde:

x_n: valor superior extremo

x_1 :valor inferior extremo

B) VARIANZA

Mide el grado de dispersión de los valores de la variable con respecto a su media.

Varianza poblacional se define como:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$$

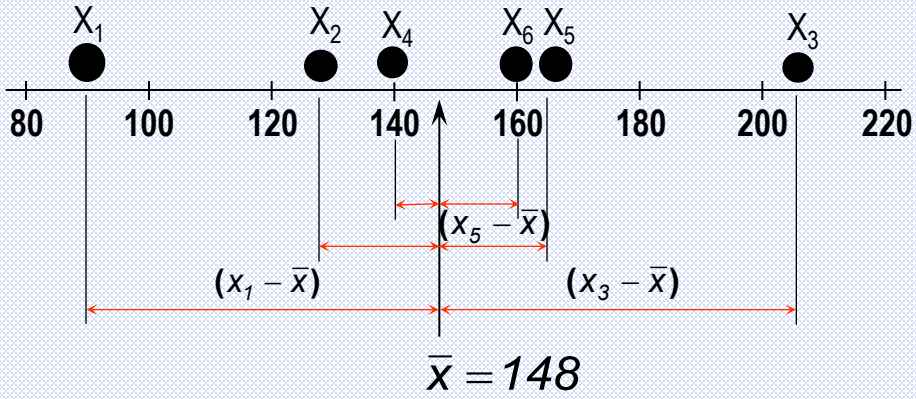
Varianza muestral

Si el punto de referencia es igual a la media aritmética muestral, y consideramos la distancia de las *n* observaciones respecto del valor de referencia, podríamos expresar dicha magnitud como la diferencia $x_i - \overline{X}$; que nos indicaría el grado de dispersión de los datos respecto de la media

a) *Para datos sin agrupar*

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1} = \frac{\sum_{i=1}^{n}x_i^2 - \frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}{n-1}$$

Ejemplo: Para la muestra **1**: 90, 128, 205, 140, 165, 160



Calcular la varianza muestral:

$$\sum_{i=1}^{6}(x_i - \bar{x})^2 = 7.510 \Rightarrow S^2 = \frac{7.510}{6-1} = \frac{7.510}{5} = 1.502 \text{ psi}^2$$

RAZONAR

Rango engañoso

Considere los dos conjuntos siguientes de calificaciones de exámenes rápidos para nueve estudiantes. ¿Cuál conjunto tiene el mayor rango? ¿También diría que este conjunto tiene la mayor variación?

Examen 1:	1	10	10	10	10	10	10	10	10
Examen 2:	2	3	4	5	6	7	8	9	10

ACTIVIDAD AULICA 7

3) MEDIDAS DE ORDEN – DATOS CUANTITATIVOS (1)

Cuantiles: Son valores que dividen al conjunto de datos en partes iguales.

Cuartiles: cuando se divide el conjunto ordenado de datos en cuatro partes iguales.

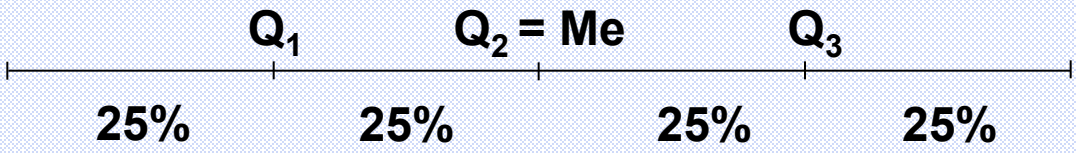
Octiles: cuando se divide el conjunto ordenado de datos en ocho partes iguales.

Deciles: cuando se divide el conjunto ordenado de datos en diez partes iguales.

Cuartiles

$$Q_1 = X_{(q_1)}$$

$$Q_3 = X_{(q_3)}$$



Donde:

$$q_1 = \frac{(n+1)}{4} \quad \text{y} \quad q_3 = \frac{3(n+1)}{4}$$

Ejemplo:

Hallar el primer y tercer cuartil para una muestra, con los siguientes valores de la resistencia al cortante de soldaduras de acero inoxidable (libras/cordón): 2.200, 2.285, 2.360, 2.385, 2.400, 2.410, 2.500, 2.550, 2.750, y 2.765.

En nuestro ejemplo: n = 10 ; q_1 = 11/4 = 2,75 y q_3 = 33/4 = 8,25

$$Q_1 = X_{(2,75)} = X_{(2)} + 0{,}75 \times (X_{(3)} - X_{(2)}) = 2.341$$

$$Q_3 = X_{(8,25)} = X_{(8)} + 0{,}25 \times (X_{(9)} - X_{(8)}) = 2.600$$

3) MEDIDAS DE ORDEN – DATOS CUANTITATIVOS (2)

Percentiles

Son medidas de ubicación, que se denotan con P_1, P_2, …, P_{99}, las cuales dividen un conjunto de datos en 100 grupos con aproximadamente 1% de los valores totales por cada grupo.

$$\text{Percentil del valor } X = \frac{\text{número de valores menores que } X}{\text{número total de valores}} \text{ x } 100$$

Ejemplo **Presupuestos de películas *ordenados* (en millones de dólares)**

4.5	5	6.5	7	20	20	29	30	35	40
40	41	50	52	60	65	68	68	70	70
70	72	74	75	80	100	113	116	120	125
132	150	160	200	225					

a) Calcular el percentil para el valor de u$s 29 millones de dólares
En la lista ordenada de presupuestos de la tabla anterior, observamos que hay 6 cantidades menores que 29, por lo tanto:

$$\text{Percentil de } 29 = \frac{6}{35} \text{ x } 100 = 17$$ (Redondeado al entero más cercano)

El presupuesto de u$s 29 millones corresponde al percentil 17; con lo cual el presupuesto de u$s 29 millones separa al 17% inferior de los presupuestos, del 83% superior de los presupuestos.

TÉCNICAS DE EXPLORACIÓN DE DATOS

DIAGRAMA DE CAJA Y BIGOTES

Constituye una gráfica que incluye la mediana, el primer y el tercer cuartil, y cualquier dato atípico que se presente en un lote de datos

Construcción del diagrama

1) De los datos se calculan los valores de la mediana (Me), los cuartiles $(Q_1 \text{ y } Q_3)$ y los extremos $x_{mín}$ y $x_{máx}$

2) Se calcula la distancia intercuartil $d_c = Q_3 - Q_1$ que constituye la "caja"

3) Se definen los puntos de corte para los valores distantes (extensión de "bigotes") como: $Q_3 + 1{,}5\, d_c$ (1) y $Q_1 - 1{,}5\, d_c$ (2) (*)

4) Encontrar el valor más grande de la muestra que esté en el rango de $Q_3 + 1{,}5\, d_c$ y el valor más pequeño de la muestra que esté en el rango de $Q_1 - 1{,}5\, d_c$. Extender líneas verticales ("bigotes") desde la "caja" hasta dichos puntos.

(*) Son datos ***atípicos*** los valores que son mayores que (1) ó menores que (2) .

(**) Son datos ***atípicos extremos*** a los valores que son mayores que $Q_3 + 3\, d_c$ (3) ó menores que $Q_1 - 3\, d_c$ (4)

Ejemplo: resistencia a la tensión (psi), de 80 muestras

7	6
8	7
9	7
10	51
11	580
12	103
13	413535
14	29583169
15	471340886808
16	3073050879
17	8544162106
18	0361410
19	960934
20	7108
21	8
22	189
23	7
24	5

$x_{mín} = 76$

$x_{máx} = 245$

$Me = 161{,}5$

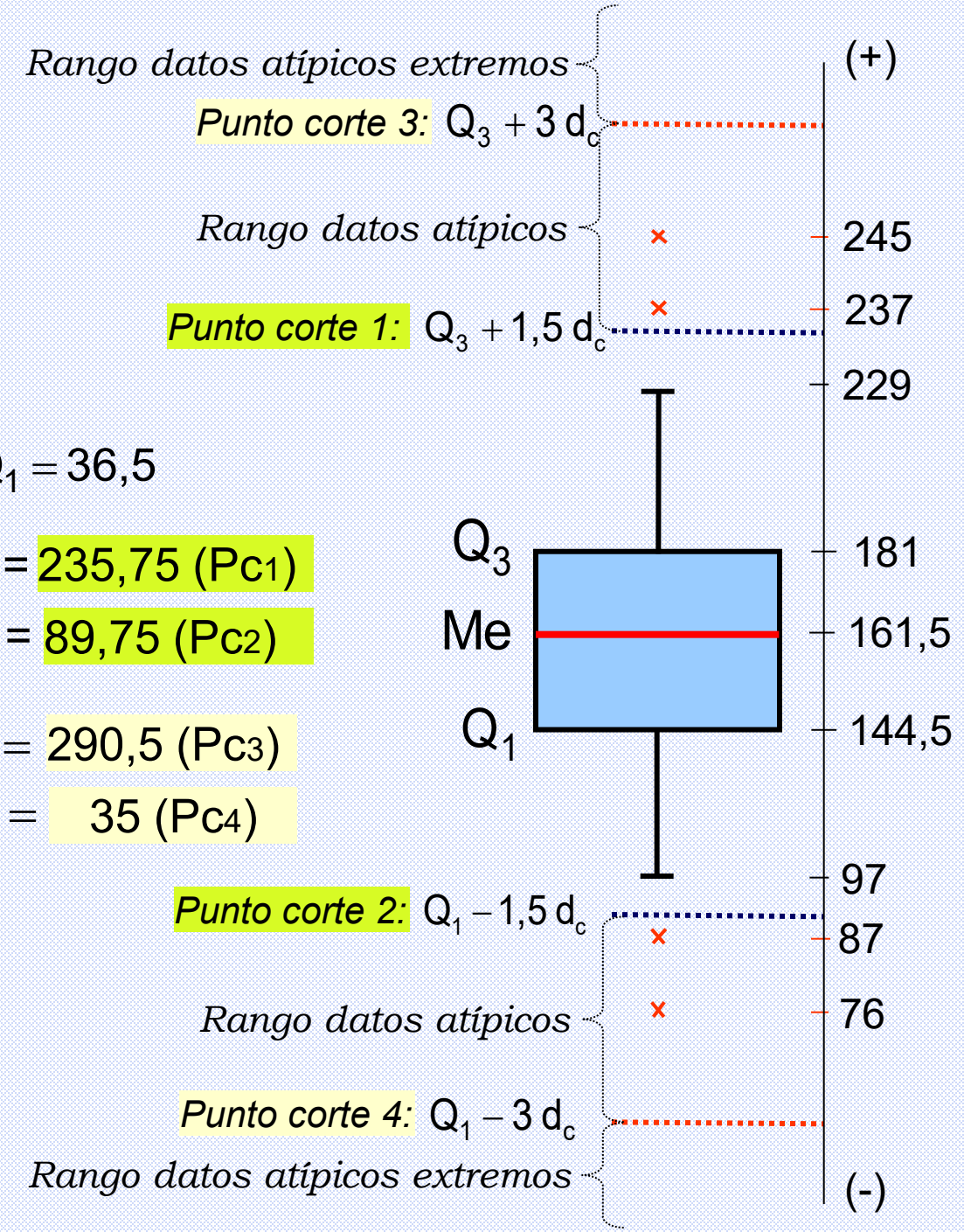
$Q_1 = 144{,}5$

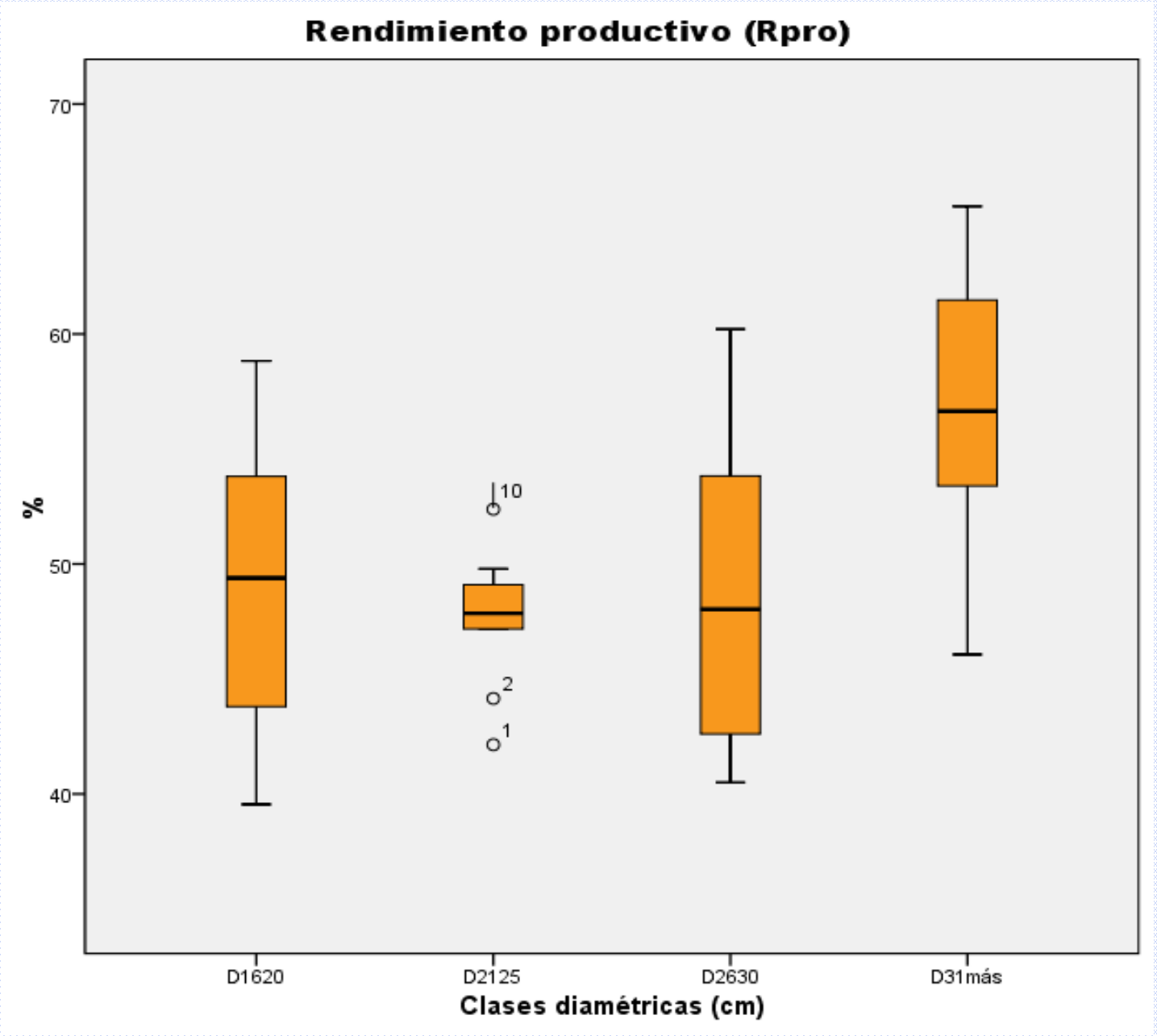
$Q_3 = 181$

$d_c = Q_3 - Q_1 = 36{,}5$

$$\begin{cases} Q_3 + 1{,}5\,d_c = 235{,}75\ (Pc_1) \\ Q_1 - 1{,}5\,d_c = 89{,}75\ (Pc_2) \end{cases}$$

$$\begin{cases} Q_3 + 3\,d_c = 290{,}5\ (Pc_3) \\ Q_1 - 3\,d_c = 35\ (Pc_4) \end{cases}$$

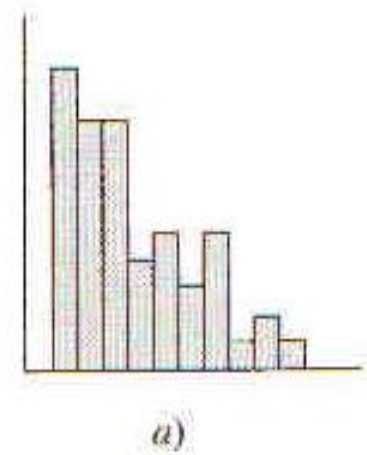




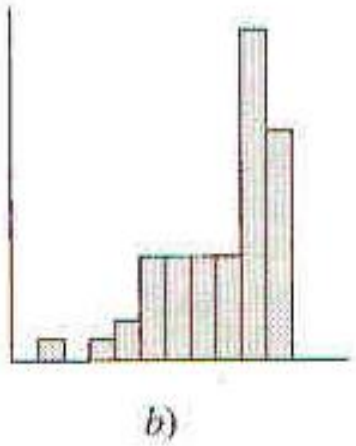
Fuente: Proy. de investigación - Determinación por métodos estadísticos del rendimiento en el proceso industrial primario de madera en un aserradero PyMe.

RAZONAR:

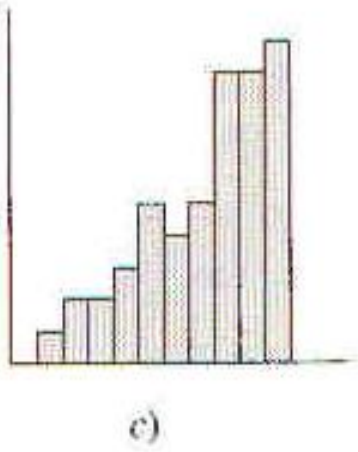
Vincular cada histograma con el diagrama de caja que representa al mismo conjunto de datos



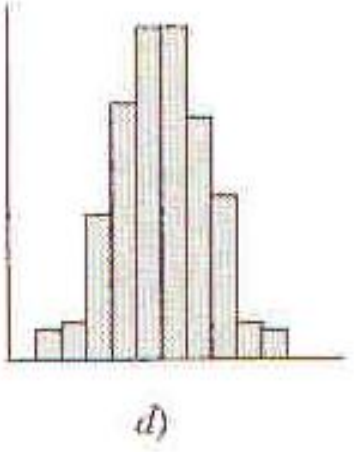
a)



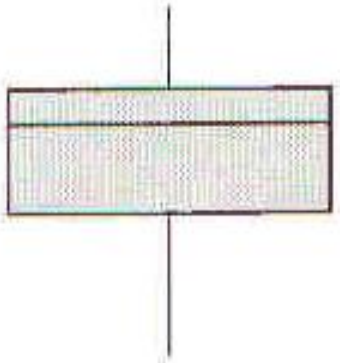
b)



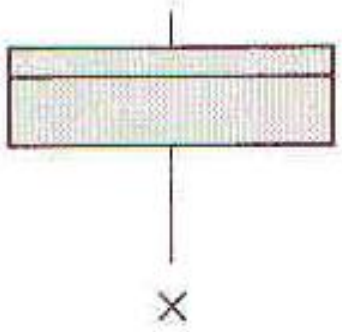
c)



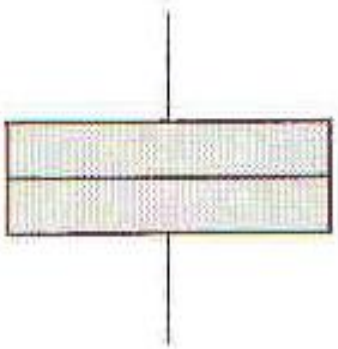
d)



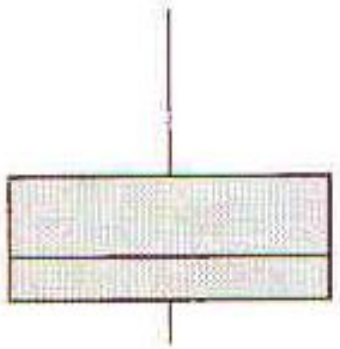
(1)



(2)



(3)



(4)