Unidad 4

Ejercicios y respuestas

I. Medidas de centralización (moda, media y mediana).

1.a) Calcular, para los siguientes ocho datos de una variable x, los estadísticos moda, media y mediana. (*No olviden, al calcular la mediana, primero ordenar los datos de menor a mayor*)

5 7 5 7 10 5 4 6

b) ¿Cuánto valdrán ahora los mismos estadísticos si, al conjunto de datos anteriores, le agregamos una **novena** observación que valga "8"?

5 7 5 7 10 5 4 6 **8**

2. Los empleados de una empresa tienen los siguientes sueldos en pesos:

800 1050 900 850 1200 1900 1000

a) ¿Cuál es el sueldo promedio (o media) de los empleados?
b) ¿Le parece que la media calculada es representativa de los sueldos que se pagan? ¿Por qué?
c) ¿Qué otro estadístico propondría para tener una mejor representación del sueldo central? Realice el cálculo con esse nuevo estimador propuesto.

3. En el ejercicio 2) anterior cambie el valor "1900" por "1300" y calcule nuevamente la media y la mediana. ¿Qué valores se obtienen ahora? ¿Qué observa al comprar la media con la mediana?

4. A cada uno de los cinco estudiantes de un curso se le midió su peso, obteniéndose los siguientes valores en kilogramos:

61 60 63 68 58

a) Calcular la media y la mediana.
b) Suponga que al cabo de un mes los cinco estudiantes pesan cada uno exactamente **dos** kilos menos, o sea, los nuevos valores ahora son:

59 58 61 66 56

Sin hacer cuentas: ¿Cuánto piensa que valdrán ahora la media y la mediana? Recién después de contestar la pregunta compruebe, realizando las cuentas correspondientes, cuanto dan, efectivamente, ambos estadísticos mencionados.

5. Se registraron las edades de una muestra de 18 estudiantes y se obtuvo:

17	17	18	20	18	20	20	17	18
19	19	21	18	21	19	18	18	19

a) Calcular la moda, la media y la mediana (¡no se olvide, antes de calcular la mediana, de ordenar los datos de menor a mayor!).

b) Realizar una tabla agrupando los datos y en la cual figuren las frecuencias absolutas y las acumuladas, para así volver a calcular la moda, la media y la mediana, **pero esta vez con la fórmula de datos agrupados**. Luego, verificar que realmente se obtuvieron los mismos resultados que antes.

6. Con el objetivo de tener una idea de cuánto le dedica la gente a ver televisión durante el fin de semana, se eligió al azar a 60 personas adultas de un pueblo llamado "Trulalá" en que viven un total de 350 personas, y se les preguntó cuánto tiempo dedicaron a ver televisión durante el fin de semana. Las respuestas obtenidas se sintetizan en la siguiente tabla:

Tiempo en hs (x)	Personas (f_i)
[0; 1)	8
[1; 2)	12
[2; 3)	18
[3; 4)	17
[4; 5)	5
Total:	**60**

a) ¿Cuál es la variable en estudio? ¿Es, continua o discreta?

b) ¿Cuál es la población en estudio?

c) ¿Qué tamaño tiene la muestra elegida?

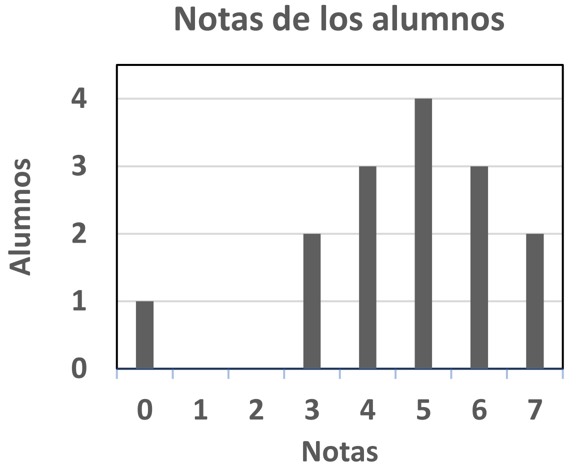
d) ¿Cuál es el tiempo promedio que las personas miraron televisión durante el fin de semana?

7. La siguiente tabla de frecuencias describe cuantos pesos por hora ganan los 60 trabajadores de una empresa, desde los obreros hasta el gerente:

x_i	50	100	150	450	700	4000	4500	5600
f_i	17	15	5	18	1	2	1	1

Calcular el promedio y la mediana. ¿Cuál de ambas es más representativa de lo que se gana en la empresa?

8. El siguiente gráfico de barras representa las notas obtenidas por N alumnos en un examen. Calcular cuantos alumnos son en total, la moda, la mediana, y la media.



9. Un pediatra obtuvo la siguiente tabla sobre los meses de edad de 50 niños de su consulta en el momento de andar por primera vez. Calcular la moda, la media y la mediana:

Meses	Niños
9	1
10	4
11	9
12	16
13	11
14	8
15	1

II Medidas de posición

10. Calcular el tercer cuartil de los siguientes datos:

14 9 4 4 9 7 8 8 14 9 7 11

11. A partir de una muestra de datos observados de una variable cuantitativa "X" se calcularon los siguientes estadísticos de posición:

P_{10}= 5 P_{30}=14 P_{70}= 33 $cuartil\ Q_1$ = 10 $quintil\ Q_3$ = 29 Me = 20

- **a)** ¿Es cierto que entre los valores 14 y 33 se encuentran el 40% de los datos de la muestra?
- **b)** ¿Cuánto vale Q_2?
- **c)** ¿Es verdad que el 40% de los datos de la muestra son mayores que 29?

d) Supongamos que alguien dice que calculó el P_{75} y obtuvo el valor 31 ¿Le parece posible? (Pista: observe el valor de P_{70} y compare…)

e) Supongamos que se supiera que la cantidad total de observaciones es de N=160. ¿Cuántos datos observados calcula que entonces son menores que el valor 10?

12. Supongamos que se tiene un conjunto de 120 observaciones de una variable numérica. ¿Cuántas observaciones serán menores que el **quintil** $\boldsymbol{Q_1}$? ¿Qué cantidad de observaciones quedarán entre el P_{10} y el P_{20}?

13. Se midió en mayo el peso a 40 alumnos de una escuela y a partir de los datos obtenidos, se calcularon los cuartiles y el desvío, obteniéndose

$$cuartil\ \boldsymbol{Q_2 = 50\ kg} \qquad \text{cuartil}\ \ \boldsymbol{Q_3 = 62\ kg} \qquad \boldsymbol{\sigma = 5\ kg}$$

a) ¿Cuántos alumnos pesan entre 50 y 62 kilogramos?

b) Supongamos que un alumno que pesaba 60 kg en mayo, al mes siguiente, junio, aumentó hasta llegar a los 64 kg, pero que el resto de los alumnos siguieron pesando exactamente lo mismo. Si entonces se recalcularan los valores de los cuartiles. ¿El nuevo valor para junio de Q_3 será, mayor, menor o igual al de mayo?

c) Supongamos, en cambio, que en junio todos los alumnos hubieran disminuido exactamente un kilo con respecto a mayo. ¿Cuánto valdrían entonces $\boldsymbol{Q_2}, \boldsymbol{Q_3}\ y\ \boldsymbol{\sigma}$?

III. Medidas de dispersión (Media, varianza, desvío estándar, intervalos, coeficiente de variación).

14. El gerente de personal entrevistó a cinco personas para su contratación; el tiempo en minutos que duró la entrevista de cada aspirante fue de:

11 8 10 7 11

a) Determinar la media de las entrevistas.

b) Calcular la varianza y el desvío estándar (desviación típica).

c) Supongamos que los datos originales no hubiesen sido los dados, si no que **a todos se les restaran 2 minutos**. O sea, imaginemos que los datos de la entrevista hubiesen sido:

9 6 8 5 9

¿Cuál imagina, **sin realizar ningún tipo de cálculo,** que van a ser los valores de la media y el desvío?

d) Realice, ahora sí, el cálculo de la media y el desvío para comprobar si su respuesta al punto c) fue correcta.

15. Se tienen los datos de la cantidad de pasajeros transportados en cada unidad de colectivo, tanto de la línea denominada “BONDI S.A.” como de la “CORRECAMINOS”. En ambas líneas las unidades transportan un promedio de 18 pasajeros. Sin embargo, al calcular los estadísticos para la variable, $X = Cantidad\ de\ pasajeros\ de\ la\ empresa\ BONDI$, se obtuvo un desvío estándar de 2 pasajeros mientras que, para la CORRECAMINOS, se obtuvo un desvío de 4 pasajeros.

Se sabe que una de las dos líneas recibe quejas de los pasajeros quienes afirman que, si bien muchas veces sus colectivos vienen bastante vacíos, en otras vienen tan repletos que no pueden sentarse. ¿Cuál de las dos líneas cree que es? ¿Por qué?

16. Se les tomó el mismo examen a dos comisiones de Matemática, A y B, de 50 alumnos cada una. Las notas medias de cada clase y sus respectivas desviaciones fueron:

$\bar{X}_A = 6$ alumnos $\sigma_A = 1$ alumnos

$\bar{X}_B = 6$ alumnos $\sigma_B = 3$ alumnos

a) ¿En qué comisión habrá más notas comprendidas entre 5 y 7?

b) En una de las comisiones se sabe que hay 15 desaprobados y 6 sobresalientes; en la otra hay 5 desaprobados y 1 sobresaliente (se desaprueba con menos de 4) ¿Cuál es la A y cuál la B?

c) Recordemos que en un intervalo del tipo $(\bar{x} - \sigma;\ \bar{x} + \sigma)$ suelen estar comprendidos el 68% de las observaciones. En tal caso ¿Cuántos alumnos estima que sacaron una nota de entre 5 y 7 puntos en la comisión A?

17. Se sabe que en una clase de matemáticas la calificación media de un examen ha sido de 5 y el desvío estándar de 1,8. En la misma clase, pero en otro examen, la calificación media también fue de 5, pero el desvío de 0,9. Si un alumno obtuvo 6,9 en el primer examen y 6,5 en el segundo ¿Qué nota le parece más meritoria, teniendo en cuenta los desvíos obtenidos?

18. Se calcularon la media y la desviación típica de las alturas (en cm) de alumnos pertenecientes a tres escuelas diferentes y se obtuvo:

Escuela	Promedio $\overline{x}$	Desvío σ
Alsina	157	12
Belgrano	160	5
Castelli	171	3

Calcule los intervalos $(\overline{x} - \sigma;\ \overline{x} + \sigma)$ de cada curso y a partir de ellos decida a cuál escuela es más probable que pertenezca un alumno que mida 167 cm.

19. En una fábrica se calcularon los sueldos en pesos de sus 300 operarios para el mes de agosto y se obtuvieron los siguientes datos:

$\overline{X} = 23.000$	$\sigma = 6900$	Cuartil Q_3= 26.000	P_{60} = 25.000

a) ¿Cuántos operarios ganan más de 25.000 pesos?
b) ¿Qué porcentaje de operarios gana entre 25.000 y 26.000 pesos?
c) Suponga al mes siguiente, septiembre, la empresa les paga a los operarios el mismo sueldo más un plus de $4.000. Dicho de otra manera, todos ganarán lo mismo que en agosto más otros $4.000 pesos. ¿Cuánto valdrán ahora los valores de los estadísticos?

Respuestas

1. a) Para calcular la moda nos fijamos cual es el valor más frecuente, o sea, el más repetido. **Mo= 5**

Para la media promediamos:

$$\overline{X} = \frac{5+7+5+7+10+5+4+6}{8} = \frac{49}{8} = \mathbf{6,125}$$

Para la mediana, primero ordenamos de menor a mayor los datos:

4 5 5 **5 6** 7 7 10

Como tenemos una cantidad par de observaciones (N= 8) debemos tomar por mediana el valor del promedio entre las observaciones quedan en la posición $\frac{N}{2}$ y $\frac{N}{2}$+1

Como $\frac{N}{2} = \frac{8}{2} = 4$; y $\frac{N}{2}$+1 = 4+1=5, promediamos entonces los números que quedaron en la 4° y 5° posición, que son el 5 y el 6.

$$\boldsymbol{Me = \frac{5+6}{2} = 5,5}$$

b) Mo= 5 $\boldsymbol{\bar{X} = \frac{57}{9} = 6,\widehat{3}}$

Para la **mediana**, primero ordenamos de menor a mayor los datos:

4 5 5 5 <u>**6**</u> 7 7 8 10

Como ahora tenemos $N =\ 9$ datos, tenemos entonces una cantidad impar de datos, debemos tomar por mediana el valor de la observación que queda en la posición que se obtiene al redondear hacia arriba $\frac{N}{2}$. Como $\frac{N}{2} = \frac{9}{2} = 4{,}5\ \Rightarrow$ redondeado hacia arriba da 5. Por lo tanto, tomamos como valor de la mediana el del dato que queda en la 5° posición: **Me= 6**

2.

- **a)** $\bar{X} = 1100$ pesos.
- **b)** No, pues solo dos personas ganan más de 1100, no es por lo tanto la media una buena representación de lo que están pagando. Este valor del promedio resultó poco representativo debido a que hay un empleado que gana una cifra muy por encima del resto, lo que ocasiona que el promedio resulte poco representativo.
- **c)** La mediana Me es más representativa: nos da un valor de 1000.

3. La media es ahora de 1014,29 pesos y la mediana se mantiene en 1000. Los dos estadísticos se parecen mucho porque ahora no hay valores "demasiado" extremos al haber reducido el valor de 1900.

4. **a) Media** $\boldsymbol{\bar{x} = 62}$ **Mediana Me=61**

b) Media $\boldsymbol{\bar{x} = 60}$ **Mediana Me=59**

5. Moda **Mo= 18** Media $\bar{X} \cong$ **18,72**

Para calcular la mediana primero ordenamos los datos de menor a mayor:

17 17 17 18 18 18 18 18 **18 19** 19 19 19 20 20 20 21 21

Como $\frac{N}{2} = \frac{18}{2} = 9$ entonces la mediana es el promedio de los datos que quedaron en el 9° y 10° lugar, es decir: Me= $\frac{18+19}{2} = \boldsymbol{18,5}$

Los datos agrupados quedarían así:

Edad	f	F
17	3	3
18	6	9
19	4	13
20	3	16
21	2	18
Total (N)	18	

Y, obviamente, utilizando la fórmula de datos agrupados, deberían obtenerse los mismos resultados de la moda, la media y la mediana que antes.

6.

a) La variable en estudio es "tiempo dedicado a ver la televisión durante el fin de semana". Estamos <u>midiendo</u> algo: el tiempo. La respuesta podría ser 3,45 horas, 4,56 horas, etc... Por lo tanto, la variable es cuantitativa continua.

b) La población está conformada por las personas que viven en el pueblo de Trulalá (o sea, por las 350 personas que viven en Trulalá).

c) El tamaño de la muestra es la cantidad de personas que forman mi muestra, o sea, N=60.

d) Para el cálculo, tenemos que utilizar las marcas de clase y las fórmulas de datos agrupados. Recordemos que la marca de clase (X_i) es el promedio de los valores extremos de cada intervalo:

Tiempo en hs (x)	Personas (f_i)	X_i	$f_i\ X_i$
[0; 1)	8	0,5	4
[1; 2)	12	1,5	18
[2; 3)	18	2,5	45
[3; 4)	17	3,5	59,5
[4; 5)	5	4,5	22,5
Total:	**60**		**149**

$$\bar{X} = \frac{149}{60} \cong \mathbf{2,48}$$

7. $\bar{X} = \mathbf{500} \quad \boldsymbol{Me} = \mathbf{100}$

Notemos que apenas 5 personas ganan más de 500 pesos, mientras que, por ejemplo, hay 32 personas (más de la mitad) que ganan 100 pesos o menos. Por lo tanto, decir que en esa empresa se gana unos 500 pesos, es una mala referencia. Es más representativo de lo que se gana término medio decir que se gana unos **100 pesos**.

8. $\boldsymbol{N} = \mathbf{15}$ $\boldsymbol{Mo} = \mathbf{5}$ $\boldsymbol{Me} = \mathbf{5}$ $\bar{\boldsymbol{x}} \cong$ **4,67**

9. Mo=12 Me=12 $\bar{X}$=12,2

Meses	**Niños**	**Fi**	**fi · xi**
9	1	1	9
10	4	5	40
11	9	14	99
12	16	30	192
13	11	41	143
14	8	49	112
15	1	50	15
		Total fi · xi :	610

10. Primero **<u>ordenamos de menor a mayor</u>**:

4 4 7 7 8 8 9 9 **9** **11** 14 14

El cuartil Q_3 debe dejar el 75% de los datos más grandes a la izquierda (y por lo tanto el 25%, una cuarta parte, a la derecha). Como son N= 12 observaciones, quiere decir que quiero que queden 9 datos a la izquierda, y 3 a la derecha. Si tomo el promedio de los datos cuyo valor es 9 y 11, eso da 10. Entonces este valor, **Q_3= 10,** cumple lo pedido.

11.

a) Como un 30% de los datos observados son menores que P_{30} = 14, y un 70% son menores que P_{70}= 33, entonces entre P_{30} y P_{70} hay un $70\% - 30\% = 40\%$ de datos.

b) Q_2 y la mediana son la misma cosa, así que $\boldsymbol{Q_2 = 20}$

c) Sí pues el $quintil\ Q_3$ es igual a 29, y por definición de quintil, el 60% de los datos son menores que su valor y el 40% serán mayores.

d) Imposible, porque al ser P_{70} = 33, quiere decir que el 70% de datos son menores que 33, así que no puede ser que haya un 75% de datos menores que 31, sería absurdo.

e) Al ser el cuartil Q_1= 10 quiere decir que el 25% de los datos son menores que 10, y como el 25% de N=160 es igual a 40, quiere decir que 40 observaciones de la muestra son menores que 10.

12. Como menores al quintil Q_1 son el 20% de las observaciones, entonces dado que N=120 quiere decir que menores a Q_1 son $\frac{20}{100} \cdot 120 = 24$ observaciones.

Entre P_{10} y P_{20} caen el 10% de las observaciones, o sea que habrá 12 observaciones entre ellos.

13.

a) **10**
b) **Mayor**
c) $\boldsymbol{Q_2 = 49\ kg, Q_3 = 61\ \mathrm{kg}} \quad \boldsymbol{\sigma = 5\ kg}\ (LO\ MISMO\ que\ en\ mayo)$

14.

a) $\bar{X} = \frac{47}{5} = 9{,}4$ errores

b) Varianza:

$$\boldsymbol{\sigma^2} = \frac{(11-9{,}4)^2 + (8-9{,}4)^2 + (10-9{,}4)^2 + (7-9{,}4)^2 + (11-9{,}4)^2}{5}$$

$$\boldsymbol{\sigma^2} = \frac{(1{,}6)^2 + (-1{,}4)^2 + (0{,}6)^2 + (-2{,}4)^2 + (1{,}6)^2}{5} =$$

$$\boldsymbol{\sigma^2} = \frac{13{,}2}{5} = \boldsymbol{2{,}64\ minutos^2}$$

Desvío: $\boldsymbol{\sigma} = \sqrt{\sigma^2} = \sqrt{2{,}64} = \mathbf{1{,}62\ minutos}$

c) La media es 2 unidades menor y el desvío debe dar **igual** que antes, o sea:

$$\bar{X} = \mathbf{7{,}4\ minutos} \qquad \boldsymbol{\sigma} = \mathbf{1{,}62\ minutos}$$

d) $\bar{x} = \frac{37}{5} = 7{,}4$

$$\boldsymbol{\sigma^2} = \frac{(9-7{,}4)^2+(6-7{,}4)^2+(8-7{,}4)^2+(5-7{,}4)^2+(9-9{,}4)^2}{5} = \mathbf{2{,}64}\ \boldsymbol{minutos^2}$$

$$\boldsymbol{\sigma} = \sqrt{\sigma^2} = \sqrt{2{,}64} = \mathbf{1{,}62\ minutos}$$

15. La CORRECAMINOS es la que recibe más quejas. El desvío dio más grande, eso indica que la cantidad de pasajeros es menos homogénea con respecto al promedio que transporta, o sea que puede haber unidades de colectivos de muy diferente cantidad de pasajeros. Mientras que en BONDI es raro que haya más de 18 pasajeros (y por lo tanto casi siempre irán todos sentados) en la CORRECAMINOS no es para nada raro que haya unidades con bastante más que 18 pasajeros y por lo tanto no puedan sentarse.

16.

a) En la primera porque el desvío es menor.

b) La comisión A es a la que pertenecen los 5 desaprobados y 1 sobresaliente (dado que el desvío es menor, quiere decir que es raro que alguien desapruebe, o que alguien saque nota alta. En cambio, en la B, hay casos muchos más variados, por eso encontraremos más gente con notas o muy altas, o muy bajas.

c) Para la comisión A es $(\overline{x}-\sigma;\ \overline{x}+\sigma) = (6-1;\ 6+1) = (\mathbf{5};\ \mathbf{7})$. Quiere decir que cerca de un 68% de los 50 alumnos sacaron una nota entre esos valores, o sea que aproximadamente habrá $0{,}68 \cdot 50 = \mathbf{34}$ alumnos con esas notas.

17. Calculemos primero los intervalos $(\overline{x}-\sigma;\ \overline{x}+\sigma)$ para cada examen:

Primer examen: $(5-1{,}8;\ 5+1{,}8) = (\mathbf{3{,}2};\ \mathbf{6{,}8})$

Segundo examen: $(5-0{,}9;\ 5+0{,}9) = (\mathbf{4{,}1};\ \mathbf{5{,}9})$

Observemos lo siguiente: el intervalo obtenido nos está diciendo que un 68% de los alumnos obtuvo una nota entre 3,2 y 6,8. Es decir que la nota de 6,9 es apenas un $6{,}9-6{,}8 = 0{,}1$ más alta que el 68% de la clase.

En cambio, en el segundo examen, si bien obtuvo una nota algo más baja (6,5) sin embargo la mayoría de los alumnos obtuvo una calificación de entre 4 y 6 puntos. Es decir que su nota fue $6{,}5-5{,}9 = 0{,}6$ punto más alta que la mayoría. Es decir, su nota de 6,5 ha sido muy alta, teniendo en cuenta como le fue al resto de la clase.

Por lo tanto, pese a que en el segundo examen obtuvo una calificación algo menor, al tener en cuenta como le fue al resto, se puede decir que fue más meritoria que la nota del primer examen.

18. Alsina $(\mathbf{145};\mathbf{169})$ **Belgrano** $(\mathbf{155};\mathbf{165})$ **Castelli** $(\mathbf{168};\mathbf{174})$

Como el 167 está dentro solo del intervalo de Alsina, y fuera de los restantes, quiere decir que lo más probable es que pertenezca a la escuela **Alsina,** ya que los intervalos nos dan una idea de entre que valores se encuentra cerca del el 70% de la población considerada.

19.

a) El 40% de $N = 300$, o sea, 120 operarios.

b) $75\% - 60\% = 15\%$

c) $\boldsymbol{\overline{x}} = \mathbf{27.000}$ $\boldsymbol{\sigma} = \mathbf{6.900}$ $\boldsymbol{Q_3} = \mathbf{30.000}$ $P_{60} = 29.000$

Explicación: Si a todos les aumentaron el sueldo en la misma cifra de 4000, entonces el promedio será ahora unos 4000 unidades más grande: $23.000 + 4.000 = 27.000$

El desvío mide cuan dispersas, cuan alejadas, están las observaciones del promedio. Pero como tanto el promedio como las observaciones se incrementaron en 4000, la distancia al promedio de cada observación al promedio sigue siendo la misma, por lo tanto, el desvío sigue siendo de 6.900.

Con el cuartil y el percentil es similar que con la media: si todos los datos aumentaron en 4000, entonces ahora, para que el 75% de los valores sean más chicos que el cuartil, debemos desplazarlo: $26.000 + 4000 = 30.000$, y similarmente $P_{60} = 29.000$.