

Unidad nº4: MEDIDAS de TENDENCIA CENTRAL y de DISPERSIÓN**Ejercicios resueltos**

Las siguientes observaciones, presentadas en una tabla de frecuencias básica, corresponden al tipo de enfermedad de pacientes que concurrieron a un hospital público la última semana del mes pasado.

Tipo de enfermedad	nº de pacientes
Cardiovasculares	12
Gastrointestinales	38
Vías respiratorias	24
Otras afecciones	22
Total	96

Establezca las medidas de tendencia central que considere oportunas y calcule la razón entre pacientes con enfermedades *cardiovasculares* y enfermedades en *vías respiratorias*.

Solución:

Como estamos en presencia de una variable cualitativa politómica (tipo de enfermedad), la única medida de tendencia central que se puede calcular, es la moda.

$$\hat{X} = \text{Gastrointestinales}$$

En este caso, la razón entre las enfermedades mencionadas es: Razón = $\frac{12}{24} = \frac{1}{2} = 0,5$. Es decir que por cada paciente con enfermedad *cardiovascular* concurrieron, al hospital público, dos pacientes con enfermedad en *vías respiratorias*.

A 54 madres de familia se les preguntó respecto al número de días, por semana, que incluyen carne vacuna en su menú del día. Las respuestas fueron las siguientes:

2 2 1 1 3 4 6 7 0 0 0 0 1 1 2 2 1 2
 5 1 2 2 1 1 1 2 1 3 4 0 4 1 2 1 2 5
 5 0 0 0 2 5 4 0 2 0 1 4 1 2 2 1 1 7

- Determine las medidas de tendencia central y dispersión que considere oportunas e interprete los resultados obtenidos en relación a la variable medida.
- Calcule P_{25} y P_{75} e interprete los resultados obtenidos.
- Calcule los coeficientes de asimetría y curtosis e interprete, de acuerdo a la variable, los resultados obtenidos.

Solución:

- Como estamos en presencia de una variable cuantitativa discreta (días por semana), podemos calcular todas las medidas de tendencia central y de dispersión.

Medidas de Tendencia Central

- Media aritmética (\bar{X}):

$$\bar{X} = \frac{2+2+1+1+3+4+6+7+0+0+0+0+1+1+2+2+1+2+5+1+2+2+1+1+2+2+1+1+7}{54} = 2,04$$

Cada familia, en promedio, consume carne vacuna 2,04 días por semana. También lo podemos expresar cómo que, cada familia consume, en promedio, 2 o 3 días por semana carne vacuna.

- *Mediana* (\tilde{X}): en primer lugar, tenemos que ordenar los datos:

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, ..., 3, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 6, 7, 7

La mediana se encuentra en la posición $(\tilde{X})_0 = \frac{n}{2} + 0,5 = \frac{54}{2} + 0,5 = 27,5$ y cómo la muestra es par, es la semisuma de los dos valores centrales:

$$\tilde{X} = 2$$

Es decir que el 50% de las familias consume carne vacuna 2 días por semana o menos, o bien el 50% de las familias consume carne vacuna 2 días por semana o más.

- *Moda* (\hat{X}): dicho valor lo podemos obtener directamente de la tabla de frecuencia.

$$\hat{X} = 1 \text{ vez por semana}$$

El mayor consumo de carne vacuna es una vez por semana.

Medidas de Dispersión

- *Rango o Recorrido* (R)

$$R = x_{\max} - x_{\min} = 7 - 0 = 7$$

- *Varianza* (S^2)

$$S^2 = \frac{(2 - 2,04)^2 + (2 - 2,04)^2 + (1 - 2,04)^2 + \dots + (1 - 2,04)^2 + (7 - 2,04)^2}{53} = 3,39$$

- *Desviación estándar* (S)

$$S = \sqrt{S^2} = \sqrt{3,39} = 1,84$$

- *MAD*

$$MAD = \text{mediana}\{|2 - 2|, |2 - 2|, \dots, |1 - 2|, |7 - 2|\} = \text{mediana}\{0, 0, \dots, 1, 5\} = 1$$

- *Coefficiente de variación* (CV)

$$CV = \frac{S}{|\tilde{X}|} \times 100 = \frac{1,84}{|2,04|} \times 100 = 0,9020 \times 100 = 90,20\%$$

Como el $CV = 90,20\% > 20\%$, el promedio de días de consumo de carne vacuna no es representativa.

b) Como ya tenemos, al calcular la mediana, ordenados los datos:

- $(P_{25})_0 = \frac{25 \times 54}{100} + 0,5 = 14 \Rightarrow P_{25} = 1$

Esto nos dice que el 25% que menos consume carne vacuna, lo hace uno o ningún día.

- $(P_{75})_0 = \frac{75 \times 54}{100} + 0,5 = 41 \Rightarrow P_{75} = 3$

Esto nos dice que el 75% que menos días consume carne vacuna lo hace 3 días o menos, o bien que el 25% que más días consume carne lo hace 3 días o más.

- c) Antes de calcular los coeficientes de asimetría y curtosis, vamos a generar la siguiente tabla complementaria, recordando que $\bar{X} = 2,04$:

Días (x_i)	$(x_i - \bar{X})$	$(x_i - \bar{X})^3$	$(x_i - \bar{X})^4$
2	-0,04	-0,000064	0,00000256
2	-0,04	-0,000064	0,00000256
1	-1,04	-1,124864	1,16985856
1	-1,04	-1,124864	1,16985856
3	0,96	0,884736	0,84934656
4	1,96	7,529536	14,75789056
...
2	-0,04	-0,000064	0,00000256
2	-0,04	-0,000064	0,00000256
1	-1,04	-1,124864	1,16985856
1	-1,04	-1,124864	1,16985856
7	4,96	122,023936	605,2387226
Total		346,406144	2030,847626

De la tabla obtenemos: $m_3 = \frac{346,406144}{54} = 6,414929$ y $m_4 = \frac{2030,847626}{54} = 37,608289$

- Asimetría (A_S)

$$A_S = \frac{m_3}{s^3} = \frac{6,414929}{(1,84)^3} = \frac{6,414929}{6,229504} \cong 1,0298$$

Como la asimetría es positiva y supera el valor 0,3, tenemos mayor concentración de datos por debajo de la media, es decir que la mayoría consume carne dos días por semana o menos.

- Curtosis (C_U)

$$C_U = \frac{m_4}{s^4} - 3 = \frac{37,608289}{(1,84)^4} - 3 = \frac{37,608289}{11,462287} - 3 \cong 0,281$$

Como la $C_U \cong 0,281 \in (-0,3; 0,3)$, el comportamiento de la distribución es mesocúrtica. Es decir que los datos se comportan de manera aproximadamente normal.

Los siguientes datos corresponden al rendimiento de trigo (en toneladas por hectárea) de 42 campos de la región pampeana.

17,2 13,5 17,7 20,1 13,7 16,6 18,0 18,9 14,4 13,5 19,3 15,4 17,6 21,6
 19,0 12,8 15,5 17,9 19,7 9,9 17,2 15,7 16,0 18,6 12,1 10,2 18,8 17,8
 14,2 13,9 9,9 17,1 18,4 14,2 15,7 13,9 22,8 19,3 13,9 18,8 14,8 20,0

- Determine las medidas de tendencia central y dispersión que considere oportunas e interprete los resultados obtenidos en relación a la variable medida.
- Calcule el P_{10} y P_{90} e interprete los resultados obtenidos.
- Calcule los coeficientes de asimetría y curtosis e interprete, de acuerdo a la variable, los resultados obtenidos.

Solución:

- Al estar en presencia de una variable cuantitativa continua (toneladas por hectárea), podemos calcular todas las medidas de tendencia central y de dispersión.

Medidas de Tendencia Central

- *Media aritmética* (\bar{X}):

$$\bar{X} = \frac{17,2 + 13,5 + 17,7 + \dots + 13,9 + 18,8 + 14,8 + 20,0}{42} = 16,32$$

Es decir que campo, en promedio, produce 16,32 toneladas por hectárea.

- *Mediana* (\tilde{X}): en primer lugar, tenemos que ordenar los datos:

9,9; 9,9; 10,2; 12,1; 12,8; 13,5; 13,5; 13,7; ...; 19,3; 19,3; 19,7; 20,0; 20,1; 21,6; 22,8

Como el tamaño de la muestra es par, la mediana es la semisuma de los dos valores centrales.

$$\tilde{X} = 16,85$$

Es decir que el 50% de los campos produce 16,85 ton/ha o menos; o bien que el 50% de los campos produce 16,85 ton/ha o más.

- *Moda* (\hat{X}): dicho valor lo podemos obtener directamente de los datos ordenados.

$$\hat{X} = 13,9$$

La producción que más se repite es de 13,9 ton/ha.

Medidas de Dispersión

- *Rango o Recorrido* (R)

$$R = x_{\max} - x_{\min} = 22,8 - 9,9 = 12,9$$

- *Varianza* (S^2)

$$S^2 = \frac{(17,2 - 16,32)^2 + (13,5 - 16,32)^2 + \dots + (14,8 - 16,32)^2 + (20,0 - 16,32)^2}{41} = 9,357$$

- *Desviación estándar* (S)

$$S = \sqrt{S^2} = \sqrt{9,357} \cong 3,059$$

- *MAD*

$$\begin{aligned} \text{MAD} &= \text{mediana}\{|17,2 - 16,85|, |13,5 - 16,85|, \dots, |14,8 - 16,85|, |20,0 - 16,85|\} = \\ &= \text{mediana}\{0,35; 3,35; 0,85; \dots, 1,95; 2,05; 3,15\} = 2,30 \end{aligned}$$

- *Coefficiente de variación* (CV)

$$CV = \frac{3,059}{16,32} \times 100 = 0,1874 \times 100 = 18,74\%$$

Como el $CV = 18,74\% < 20\%$, el promedio de producción de los campos es representativo.

b) Como ya tenemos, al calcular la mediana, ordenados los datos:

$$(P_{10})_0 = \frac{10 \times 42}{100} + 0,5 = 4,7 \Rightarrow P_{10} = 12,45$$

Esto nos dice que el 10% que menos rinde produce 12,45 ton/ha o menos, o bien que el 90% que más rinde, produce 12,45 ton/ha o más.

- $(P_{90})_0 = \frac{90 \times 42}{100} + 0,5 = 38,3 \Rightarrow P_{90} = 19,85$

Esto nos dice que el 90% que menos rinde produce 19,85 ton/ha o menos, o bien que el 10% que más rinde, produce 19,85 ton/ha o más.

- a) Antes de calcular los coeficientes de asimetría y curtosis, vamos a generar la siguiente tabla complementaria, recordando que $\bar{X} = 16,32$:

Ton/ha (x_i)	$(x_i - \bar{X})$	$(x_i - \bar{X})^3$	$(x_i - \bar{X})^4$
17,2	0,88	0,681472	0,599695
13,5	-2,82	-22,425768	63,240666
17,7	1,38	2,628072	3,626739
20,1	3,78	54,010152	204,158375
13,7	-2,62	-17,984728	47,119987
16,6	0,28	0,021952	0,006147
...
19,3	2,98	26,463592	78,861504
13,9	-2,42	-14,172488	34,297421
18,8	2,48	15,252992	37,827420
14,8	-1,52	-3,511808	5,337948
20,0	3,68	49,836032	183,396598
Total		-293,068736	9068,738159

De la tabla obtenemos: $m_3 = \frac{-293,068736}{42} = -6,977827$ y $m_4 = \frac{9068,738159}{42} = 215,922337$

- **Asimetría (A_s)**

$$A_s = \frac{m_3}{s^3} = \frac{-6,977827}{(3,059)^3} = \frac{-6,977827}{28,624534} \cong -0,2438$$

La asimetría es negativa pero como $-0,2438 \in (-0,3; 0,3)$, tenemos una concentración uniforme (o levemente asimétrica negativa) alrededor de la media.

- **Curtosis (C_U)**

$$C_U = \frac{m_4}{s^4} - 3 = \frac{215,922337}{(3,059)^4} - 3 = \frac{215,922337}{87,562451} - 3 \cong -0,5341$$

Como $C_U \cong -0,5341 < -0,3$ el comportamiento de la distribución es platicúrtica. Es decir que los datos están muy dispersos alrededor de la media.

Nota: todos los ejercicios resueltos, presentados anteriormente, pueden hacerse utilizando el software estadístico InfoStat.

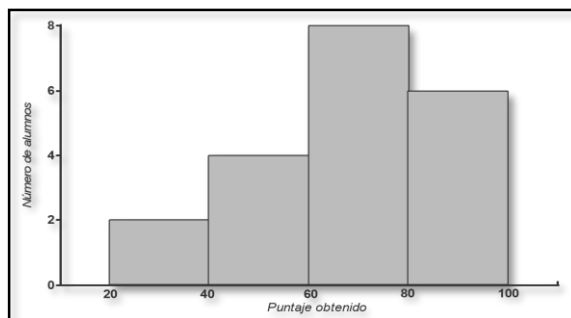
Ejercicios propuestos

1. Responda, de manera breve y concisa, las siguientes preguntas:

- Si tenemos una distribución de puntajes de ansiedad donde, a mayor puntaje mayor ansiedad, ¿qué tipo de distribución es más deseable en relación a la simetría/asimetría? ¿Y si fuera una distribución con puntajes de capacidad para resolver problemas?
- Dé un ejemplo, de una variable, en el cual lo óptimo sea una distribución simétrica y fundamente su elección.
- ¿Qué significado tiene un valor de desviación estándar cada vez más próximo a 0?
- ¿Por qué la moda no es la medida de tendencia central más confiable?

- e) Si calculamos el CV de una distribución y resulta ser igual al 69%, ¿qué decisión debiera tomar para poder explicar el comportamiento de la muestra de manera fiable?
- f) Si Ud. rindió 6 materias y su promedio fue de 8,50 puntos, ¿cuál será su promedio si rinde una materia más y obtiene como nota un 9?

2. El siguiente histograma corresponde a los puntajes finales por estudiantes de un curso de Estadística Básica.



¿Hubo alguna calificación menor a 20 puntos?	
¿Qué porcentaje de estudiantes obtuvo entre 60 y 80 puntos?	
¿Qué proporción de estudiantes obtuvo entre 40 y 60 puntos?	
¿Qué porcentaje de estudiantes obtuvo más de 60 puntos?	
¿Qué proporción de estudiantes obtuvo menos de 40 puntos o más de 80 puntos?	
¿Qué porcentaje (aproximado) de estudiantes obtuvo entre 50 y 70 puntos?	

3. Ejercicio n°22 (Organización de la Información, pág.10), calcule e interprete:
 - a) Media aritmética, mediana y moda.
 - b) La media aritmética, calculada en el punto a), ¿es representativa?
4. Veinte personas en una habitación tienen una altura promedio de 179 cm.
 - a) Si entra una persona que mide 150 cm, ¿cuál es el promedio de las 21 personas?
 - b) ¿Cuánto debe medir la persona 21 para que el promedio aumente 1 cm?
5. Los pesos netos en gramos del contenido de cinco envases, seleccionados en forma aleatoria de la línea de producción, son: 85,4 – 85,3 – 84,9 – 85,4 – 85,0 ¿Cuál es la media y la mediana de las observaciones de la muestra? Interprete los resultados obtenidos.
6. Hace dos años la empresa XX pagó salarios anuales de u\$d 30360 a cada uno de los 5 empleados administrativos, u\$d 47784 a cada uno de los 2 contadores y de u\$d 89868 al gerente.
 - a) ¿Cuál fue el salario promedio?
 - b) ¿Cuántos de los empleados ganaron menos que la media?
 - c) ¿Cuál es la mediana del salario? Interprete la respuesta en función de la variable.

El año pasado, dicha empresa, pagó los mismos salarios a los cinco empleados administrativos y a los dos contadores, pero el gerente cobró u\$d 95772.

 - d) ¿Cuál es el nuevo valor de la media y diga por qué se ve afectada respecto al valor de la media obtenido en el punto a)?
 - e) ¿Dicho cambio afecta a la mediana?, ¿por qué?

7. Los tiempos que necesitaron varias empresas de seguro para revisar solicitudes para servicios de coberturas semejantes son (en minutos): 50, 47, 52, 58, 51, 56 y 50. ¿Cuál es la media y la mediana del tiempo necesario para revisar una solicitud? Explique los resultados obtenidos.
8. En un estudio de datos de los archivos personales de las empleadas de un hospital, el coeficiente de asimetría para la edad de las empleadas es -2,25. Interprete este resultado en función a las edades de las empleadas.
9. De una muestra de 20 operarios de una compañía de accesorios eléctricos se obtuvieron los siguientes montos netos de pago (en \$), tras deducciones, en una semana determinada.

2520	2520	2520	2520	2520	2520	2520	2520	2677,5	2677,5
2782,5	2782,5	2940	2940	3045	3150	3202,5	3412,5	3465	3570

Calcule la media aritmética, la mediana, la moda, la MAD y el coeficiente de variación. Interprete los resultados obtenidos.

10. Los siguientes datos corresponden a los tiempos de reacción, frente a un estímulo sensorial, de una muestra de 33 sujetos, medidos en centésimas de segundos (cs):

55	51	60	56	64	56	63	63	61	57	62	50	49	70	72	67	63	52
54	48	53	58	66	68	45	74	65	58	61	62	59	64	57			

- a) Calcule la media aritmética e interprete el resultado obtenido.
- b) Calcule la mediana e interprete el resultado obtenido.
- c) Determine el valor máximo del 25% de sujetos que menos tiempo tarda en reaccionar e interprete el resultado obtenido.
- d) Determine el valor mínimo del 25% de sujetos que más tiempo tarda en reaccionar e interprete el resultado obtenido.
11. Ejercicio nº6 (Organización de la Información, pág.6), calcule e interprete:
 - a) ¿Cuánto dinero, en promedio, se gasta diariamente en comestibles?
 - b) ¿Cuál es el gasto que se registra con mayor frecuencia en estos hogares?
 - c) ¿Cuál es la variabilidad absoluta del gasto semanal en relación a la media?
12. Ejercicio nº7 (Organización de la Información, pág.6), calcule e interprete:
 - a) El saldo máximo que posee la mitad de los ahorristas que menos ahorros tiene.
 - b) ¿Cuál es la brecha entre el cliente que más ahorros tiene y el que menos ahorros tiene?
 - c) ¿Cuál es la variabilidad relativa en el conjunto total de cuentas bancarias?
 - d) A partir del resultado del punto a), determine la variabilidad respecto a la mediana.
 - e) Determine el valor de la curtosis.
13. Ejercicio nº8 (Organización de la Información, pag.6), calcule e interprete:
 - a) Determine el promedio de edad de los clientes.
 - b) ¿Cuál es la brecha entre el cliente que más edad tiene y el de menor edad?
 - c) ¿Cuál es la variabilidad respecto a la media en el conjunto total de viajeros?
 - d) ¿Y la variabilidad respecto a la mediana en el conjunto total de viajeros?
 - e) ¿Qué puede decir acerca de la simetría o asimetría del grupo total?
14. Ejercicio nº10 (Organización de la Información, pág.7), calcule e interprete:

- a) La moda de la distribución.
 - b) La razón entre el salario más alto y el salario más bajo.
 - c) Si hubiese 900 empleados en lugar de 70, ¿cuántos de ellos ganarían entre \$7000 y \$8000? y ¿cuántos ganarían más de \$9000?
- 15.** Ejercicio nº11 (Organización de la Información, pág.7), calcule e interprete:
- a) La moda del conjunto de puntajes.
 - b) Si hubiese 1500 locales en alquiler, ¿cuántos de ellos costarían entre 440 y 470 u\$d?
 - c) ¿Cuál es la razón entre los locales más caros y los más baratos?
- 16.** Ejercicio nº12 (Organización de la Información, pág.7), calcule e interprete:
- a) En el año 2008 se creó una ley que regulaba los precios de este tipo de ventas. ¿Cuál fue el promedio anual de ventas desde el 2003 al 2008 por una parte y de 2009 a 2014 por la otra?
 - b) A partir de la información anterior, ¿qué opina acerca de los efectos de esta ley?
 - c) ¿Cómo es la forma de la distribución de las ventas si evaluamos la curtosis?
- 17.** Ejercicio nº13 (Organización de la Información, pág.7), calcule e interprete:
- a) ¿Cuál es la patología que se presenta con mayor frecuencia?
 - b) En 400 pacientes, ¿cuántos tendrían “disartria” y cuántos “dispraxia verbal”?
 - c) De acuerdo a la información obtenida, si ingresa un paciente con “deglución atípica”, ¿cuántos se espera que ingresen con “agrafia”?
- 18.** Ejercicio nº14 (Organización de la Información, pág.8), calcule e interprete:
- a) ¿La razón de turistas que viajan a Córdoba por cada uno que viaja a Cataratas?
 - b) Determine la moda.
 - c) ¿La razón entre la cantidad de turistas que ha viajado a Cataratas y los que han viajado a Mar del Plata?
 - d) De cada 8000 paquetes turísticos, ¿cuántos se han vendido para Córdoba?
- 19.** Ejercicio nº15 (Organización de la Información, pág.8), calcule e interprete:
- a) ¿El valor de la mediana de las horas de estudio del grupo total?, ¿y la MAD?
 - b) ¿Cuál media es más representativa en este grupo, ¿la de horas de estudio o la de edad?
 - c) La razón entre los rendimientos: Bajo y Medio; Medio y Alto; y Alto y Bajo.
 - d) Si fuesen 3500 estudiantes en lugar de 20, ¿cuál sería la proporción de estudiantes que dedican 7 h./semana, o más, al estudio?, ¿y cuál la proporción de mujeres y hombres?
 - e) La moda del rendimiento de los hombres y de las mujeres.
 - f) El promedio de edad de las mujeres y el de los hombres.
 - g) ¿Es más representativo el número de horas de estudio de los hombres o de las mujeres?
 - h) La mayor cantidad de horas de estudio para el 25% que menos estudia, tanto para el grupo de hombres y como para el grupo de mujeres.
 - i) La menor cantidad de horas de estudio para el 25% que más estudia, tanto para el grupo de mujeres y como para el grupo de hombres.
- 20.** Ejercicio nº17 (Organización de la Información, pág.9), calcule e interprete:
- a) Los estudiantes que tardan hasta media hora en llegar, ¿cuánto tardan en promedio?,

¿y los que tardan más de media hora? ¿Cuál de los dos grupos es más heterogéneo y por qué?

- b) ¿Cuántos minutos tarda como máximo en llegar el 55% más rápido del grupo?, ¿y el 85%? ¿Cuánto tarda, como mínimo, el 30% más lento o que más demora?
- c) Los coeficientes de asimetría y curtosis del grupo total en estudio.

21. Ejercicio nº18 (Organización de la Información, pág.9), calcule e interprete:

- a) El peso promedio de los neonatos.
- b) La DE del peso de los neonatos.
- c) La razón entre los neonatos con *bajo peso* respecto a los con *sobre peso*. ⁽¹⁾
- d) El P₅ y el P₉₅ del grupo total de neonatos.

22. Ejercicio nº20 (Organización de la Información, pág.9), calcule e interprete:

- a) El promedio de pesos.
- b) Diga si la media calculada en el punto a) es representativa del grupo.
- c) La mediana.
- d) ¿Cuál es la variabilidad respecto a la mediana?

23. En el punto de partida de un laberinto hay cuatro puertas idénticas A, B, C y D. Si un sujeto elige la puerta A vuelve al punto de partida después de recorrer dos metros; si elige la puerta B recorre cinco metros y vuelve al mismo punto; si elige la puerta C camina tres metros y vuelve al lugar de partida. Si elige la puerta D sale al exterior recorriendo un metro. ¿Qué distancia recorre en promedio, el sujeto antes de salir, si siempre elige una puerta distinta de las seleccionadas como punto de partida?

24. Los siguientes datos se refieren al tiempo de entrega (en minutos) de comida a domicilio del Restaurante “Shima Uta”, especialista en comida japonesa, ubicado en el centro de una Ciudad.

30	55	38	34	30	24	45	28	51	51	62	21	33	49	57	49	27
22	47	42	36	39	65	26	37	44	33	47	19	43	27	51	36	54
21	14	25	36	61	46	48	35	40	36	67	56	45	35	54	34	

- a) Calcule la media aritmética, la mediana y la moda e interprete los resultados obtenidos.
- b) ¿Es representativa la media aritmética obtenida en el punto a)?
- c) ¿Cuántos minutos tarda como máximo, en entregar el pedido, el 25% más rápido?
- d) ¿Cuánto tarda como mínimo, en entregar el pedido, el 25% más lento o que más demora?

25. Las cifras de la siguiente tabla corresponden a miligramos de hidroxiprolina absorbidos por 1 miligramo de masa intestinal, analizados en ocho pacientes:

Hidroxiprolina (mg)	77,3	61,2	82,4	75,9	61,0	70,2	65,0	80,0
Paciente nº	3	10	15	13	8	5	2	0

Calcule el promedio de hidroxiprolina por paciente e interprete el resultado obtenido. ¿La media obtenida, en dicho grupo, es representativa? Justifique su respuesta.

⁽¹⁾ *Bajo peso*: menos de 2500 g y *sobre peso*: más de 4500 g.

ALGUNAS DUDAS FRECUENTES

Un trabajo en el cual sólo aparece la media, sin ninguna medida de dispersión, ¿es un trabajo con rigor científico?

Una medida de tendencia central debe ir siempre acompañada de una medida de dispersión que nos permita evaluar su representatividad. En el caso de la media, la medida de dispersión más utilizada es la desviación estándar, pero dependiendo de los objetivos del trabajo, puede ser otra (recorrido, coeficiente de variación, etc.).

¿En qué casos debemos utilizar la media y en cuáles no?

Para calcular la media es necesario sumar sus valores, luego los datos deben ser cuantitativos. No obstante, aunque pueda calcularse, si hay datos muy grandes o muy pequeños, puede no ser la medida de tendencia central más adecuada ya que es muy sensible a valores extremos (outliers). La medida de dispersión que utilicemos nos ayudará a decidir si la media es adecuada o no, para sintetizar esos datos.

¿Puede ser más útil la mediana que la media?

Sí, puede que la mediana sea más útil, si la media no representa bien a los datos.

Si una medida de dispersión nos dice que la media no es adecuada, ¿debemos incluir en la publicación la media en lugar de la mediana, siempre?

No siempre. Depende de lo que se vaya a hacer en el estudio. Por ejemplo, si después utilizamos un test que compara medianas, podríamos incluir la mediana en la parte descriptiva y la MAD.

Algunas publicaciones presentan media \pm desviación estándar, y otras, media \pm error estándar. ¿Cuál debo utilizar? ¿Cuál es mejor?

Depende del objetivo del trabajo. Si lo que se pretende es saber si la media representa bien a sus datos, utilizará media \pm desviación estándar. Si lo que quiere es saber si se espera que la media sea estable en el muestreo, o sea, cuán parecida será su media a la de otros investigadores que hagan estudios similares, utilizará media \pm error estándar.

En algunas publicaciones aparece media \pm 2 errores estándar, ¿qué significa?

Esta expresión se corresponde con el cálculo del intervalo de confianza para la media.

Aclaremos esto: cuando se tienen datos procedentes de una muestra se calcula la media muestral. Obviamente cambiando un solo dato, cambia la media. Por lo tanto, esa media muestral es una estimación de la verdadera media que es la que obtendríamos estudiando toda la población. Pero generalmente no se dispone de la población.

Lo que se hace es, a partir de la media calculada en la muestra, estimar con una alta fiabilidad, entre qué valores cabe esperar que esté la verdadera media. Eso es lo que hacemos al calcular media \pm 2 error estándar. Esto significa que, si sumamos y restamos a la media muestral, 2 veces el valor del error estándar, se obtienen los extremos de un intervalo que con una alta

probabilidad (aproximadamente el 95%) contendrá a la verdadera media (la media poblacional).

En otras publicaciones aparece media ± 3 errores estándar, ¿qué significa?

Es la misma idea, pero en este caso se obtienen los extremos de un intervalo que contendrá a la verdadera media (la media poblacional), con una probabilidad superior al 99%.

¿Y no podría entenderse media \pm error estándar de una forma similar?

Sí, en ese caso, se obtienen los extremos de un intervalo que contendrá a la verdadera media (la media poblacional), con una probabilidad del 68%.

¿Sería correcto escribir en una publicación la mediana acompañada de la desviación estándar o la varianza?

No. Ya que la desviación estándar nos indica si la media es representativa o no. Si no se utiliza la media, esa información no es de interés. Lo apropiado sería escribir la mediana acompañada con la MAD.

Documento de Circulación Interna