open green road

Guía Matemática

Medidas de posición para datos no agrupados

tutor: Juan José Muñoz

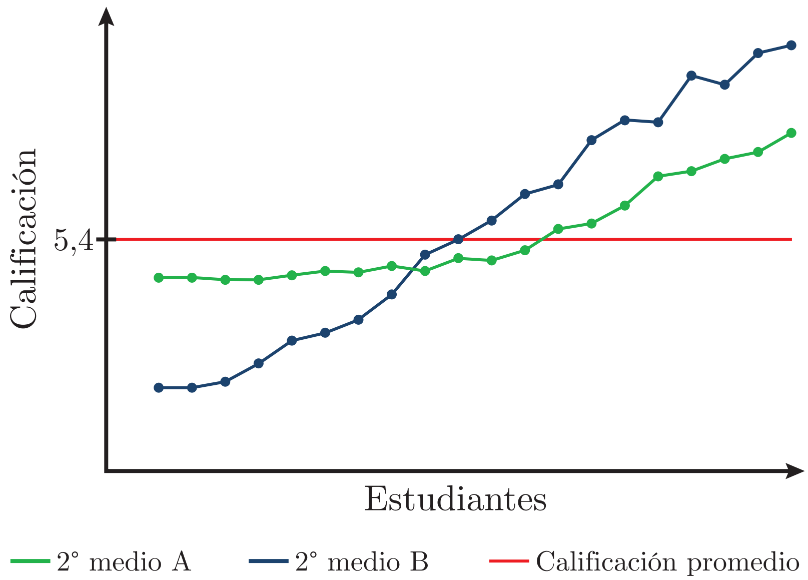
puntaje nacional.cl

1. Cuantiles

Los segundos medios A y B, ambos con igual número de estudiantes, rindieron una prueba de conocimientos múltiples. La calificación promedio de ambos cursos es la misma e igual a 5,4, y la desviación estándar es mayor para el segundo medio B, con $\sigma_A = 0,51$ y $\sigma_B = 1,38$.

¿A qué conclusión llegamos?

Con dichos datos solo podemos afirmar que la dispersión de las calificaciones obtenidas es mayor para el segundo medio B, o equivalentemente, el nivel académico de sus estudiantes es más dispar. Pero, ¿qué pasa con la distribución de las calificaciones para cada uno de los segundos medios? ¿Qué forma adopta la distribución? ¿Existe algún indicador que haga referencia al comportamiento de las calificaciones dentro del conjunto de datos? Existen indicadores que nos entregan una idea global acerca de la distribución de los datos del conjunto. Dichos indicadores irán esbozando una gráfica como el de la siguiente figura, la cual ilustra las calificaciones obtenidas por los segundos medios.



Cuando se desea analizar un conjunto de un determinado número de datos, no siempre las medidas de tendencia central o las medidas de dispersión son suficientes para poder extraer información relevante para el análisis deseado.

Cuando un conjunto de datos está ordenado por magnitud, el **valor central o mediana** divide al conjunto en dos partes iguales. Podemos extender esta idea pensando en aquellos valores que dividen al conjunto en cuatro partes iguales, en diez partes iguales o incluso en cien partes iguales. A estos valores los llamamos **cuantiles**.

En estadística descriptiva, los cuantiles corresponden a medidas de posición no central y permiten conocer la forma en que se distribuyen los datos dentro del conjunto. Los cuantiles más usados son los cuartiles, los quintiles, los deciles y los percentiles.

1.1. Cuartiles

En un conjunto de datos **ordenados de forma creciente**, se llama cuartiles a los tres valores C_1, C_2 y C_3 que dividen al conjunto en cuatro partes iguales. Como consecuencia, el segundo cuartil corresponde a la mediana del conjunto, pues lo divide en dos partes porcentualmente iguales. El primer cuartil corresponde a la mediana de la primera mitad de valores, mientras que el tercer cuartil corresponde a la mediana de la segunda mitad de valores.

Dado un conjunto de n datos, la **posición** P_k del cuartil k-ésimo la obtenemos mediante la siguiente relación:

$$P_k = \frac{kn}{4}, \qquad \text{con } k = 1, 2, 3 \tag{1}$$

Una vez encontrado el valor que corresponde a P_k, el primer valor cuya frecuencia absoluta acumulada lo supera es C_k.

Cuando el número de elementos del conjunto es par (n es par), el cálculo de los cuartiles considera el promedio los dos valores centrales, esto es, el valor que define el cuartil $k-$ésimo y el valor inmediatamente mayor.

✎ ***Ejemplo***

Los siguientes ejemplos ilustran 4 casos que debe tener presente en el cálulo de los cuartiles.

1. Sea el conjunto de números $\{3,\ 7,\ 1,\ 8,\ 6,\ 5,\ 2,\ 4\}$. Calcule el valor de cada cuartil.

   Solución:
   Ordenamos los números de forma creciente:

   $$1 - 2 - 3 - 4 - 5 - 6 - 7 - 8$$

   Usamos $P_k = \dfrac{kn}{4}$, con $k = 1, 2, 3$ para obtener la posición (o equivalentemente, la frecuencia absoluta acumulada) de cada cuartil:

   $$P_1 = \frac{1 \cdot 8}{4} = 2$$
   $$P_2 = \frac{2 \cdot 8}{4} = 4$$
   $$P_3 = \frac{3 \cdot 8}{4} = 6$$

   Observe que el conjunto está conformado por un número par de elementos. Si lo dividimos a la mitad se obtienen dos subconjuntos pares. Luego, el primer cuartil queda definido por el promedio entre los elementos que ocupan la segunda y tercera posición, el segundo cuartil queda definido por el promedio entre los elementos que ocupan la cuarta y quinta posición y el tercer cuartil queda definido por el promedio entre los elementos que ocupan la sexta y séptima posición:

   $$C_1 = \frac{2 + 3}{2} = 2,5$$

   $$C_2 = \frac{4 + 5}{2} = 4,5$$

$$C_3 = \frac{6+7}{2} = 6,5$$

La siguiente solución es una visión más gráfica del problema. Dado que el conjunto está conformado por 8 elementos, al dividirlo a la mitad se forman 2 subconjuntos de 4 elementos cada uno. Los cuartiles C_1 y C_3, destacados en negrita, corresponden al promedio de los elementos centrales de dichos subconjuntos:

$$1 - \mathbf{2} - \mathbf{3} - 4 \qquad\qquad 5 - \mathbf{6} - \mathbf{7} - 8$$

De la misma forma, C_2 corresponde al promedio de los elementos centrales del conjunto completo:

$$1 - 2 - 3 - \mathbf{4} - \mathbf{5} - 6 - 7 - 8$$

2. Sea el conjunto ordenado de números $\{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$. Calcule el valor de cada cuartil.

   Solución:
   Dado que el conjunto ya se encuentra ordendo, usamos $P_k = \frac{kn}{4}$, con $k = 1, 2, 3$ para obtener la posición (o equivalentemente, la frecuencia absoluta acumulada) de cada cuartil:

$$P_1 = \frac{1 \cdot 6}{4} = 1,5$$

$$P_2 = \frac{2 \cdot 6}{4} = 3$$

$$P_3 = \frac{3 \cdot 6}{4} = 4,5$$

   En este caso, el conjunto está conformado por un número par de elementos, de modo que en el cálculo de C_2 se considera el promedio sus dos valores centrales. Así, el segundo cuartil corresponde al promedio de los elementos que ocupan la tercera y cuarta posicón, y cuyos valores son 3 y 4.

   Al dividir el conjunto a la mitad se forman dos subconjuntos impares, de modo que para el cálculo de C_1 y C_3 se considera el valor proporcionado por P_k. Así, el primer cuartil corresponde al elemento que ocupa la segunda posición y cuyo valor es 2, mientras que el segundo cuartil corresponde al elemento que ocupa la quinta posición y cuyo valor es 5.

   Gráficamente, los cuartiles C_1 y C_3 corresponden a los elementos destacados en negrita:

$$1 - \mathbf{2} - 3 \qquad\qquad 4 - \mathbf{5} - 6$$

   De la misma forma, C_2 corresponde al promedio de los elementos centrales del conjunto completo:

$$1 - 2 - \mathbf{3} - \mathbf{4} - 5 - 6$$

3. Sea el conjunto de números $\{3,\ 7,\ 1,\ 4,\ 6,\ 5,\ 2\}$. Calcule el valor de cada cuartil.

   Solución:
   Primero ordenamos los números de forma creciente:

$$1 - 2 - 3 - 4 - 5 - 6 - 7$$

Usamos $\mathrm{P}_k = \dfrac{kn}{4}$, con $k = 1, 2, 3$ para obtener la posición (o equivalentemente, la frecuencia absoluta acumulada) de cada cuartil:

$$\mathrm{P}_1 = \frac{1 \cdot 7}{4} = 1,75$$

$$\mathrm{P}_2 = \frac{2 \cdot 7}{4} = 3,5$$

$$\mathrm{P}_3 = \frac{3 \cdot 7}{4} = 5,25$$

Observe que el conjunto está conformado por un número impar de elementos. Si lo dividimos a la mitad se obtienen dos subconjuntos impares. Luego, el primer cuartil corresponde al elemento que ocupa la segunda posición y cuyo valor es 2, el segundo cuartil corresponde al elemento que ocupa la cuarta posición y cuyo valor es 4, y por último, el tercer cuartil corresponde al elemento que ocupa la sexta posición y cuyo valor es 6.

Advierta que cada una de las posiciones que definen los cuartiles se han aproximado al entero consecutivo (o como se enunció anteriormente, "una vez encontrado el valor que corresponde a P_k, el primer valor cuya frecuencia absoluta acumulada lo supera es C_k").

La siguiente solución es una visión más gráfica del problema. Dado que el conjunto está conformado por 7 elementos, al dividirlo a la mitad se forman 2 subconjuntos de 3 elementos cada uno, unidos por un valor central correspondiente a C_2. Los cuartiles C_1 y C_3, destacados en negrita, corresponden a los elementos centrales de dichos subconjuntos:

$$1 - \mathbf{2} - 3 \qquad\qquad 4 \qquad\qquad 5 - \mathbf{6} - 7$$

4. Sea el conjunto ordenado de números $\{1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9\}$. Calcule el valor de cada cuartil.

Solución:
Dado que el conjunto ya se encuentra ordendo, usamos $\mathrm{P}_k = \dfrac{kn}{4}$, con $k = 1, 2, 3$ para obtener la posición (o equivalentemente, la frecuencia absoluta acumulada) de cada cuartil:

$$\mathrm{P}_1 = \frac{1 \cdot 9}{4} = 2,25$$

$$\mathrm{P}_2 = \frac{2 \cdot 9}{4} = 4,5$$

$$\mathrm{P}_3 = \frac{3 \cdot 9}{4} = 6,75$$

En este caso en particular el conjunto está conformado por un número impar de elementos, de modo que en el cálculo de C_2 se considera su valor central. Así, el segundo cuartil corresponde al valor que ocupa la quinta posición dentro del conjunto y cuyo valor es 5.

Al dividir el conjunto a la mitad se forman dos subconjuntos pares, de modo que en el cálculo de C_1 y C_2 se considera el promedio de sus dos valores centrales. Así, el primer cuartil corresponde al promedio de los elementos que ocupan la segunda y tercera posición, y cuyos valores son 2 y 3 respectivamente, mientras que el tercer cuartil corresponde al promedio del los elementos que ocupan la sexta y séptima posición, y cuyos valores son 6 y 7 respectivamente.

Gráficamente, los cuartiles C_1 y C_3 corresponden al promedio de los elementos destacados en negrita:

$$1 - \mathbf{2} - \mathbf{3} - 4 \qquad\qquad 5 \qquad\qquad 6 - \mathbf{7} - \mathbf{8} - 9$$

De la misma forma, C_2 corresponde al valor central del conjunto:

$$1 - 2 - 3 - 4 \qquad\qquad \mathbf{5} \qquad\qquad 6 - 7 - 8 - 9$$

Bajo el primer cuartil se encuentra el 25 % de los datos del conjunto. Bajo el segundo cuartil (mediana) se encuentra el 50 %, y bajo el tercer cuartil se encuentra el 75 % de los datos. Visto de otra forma, decimos que sobre el primer cuartil se encuentra el 75 % de los datos del conjunto, sobre el segundo cuartil se encuentra el 50 % y sobre el tercer cuartil se encuentra el 25 % de los datos.

25%	25%	25%	25%

C_1 $\qquad$ C_2 $\qquad$ C_3

Considere el conjunto ordenado $S = \{1,\ 2,\ \mathbf{2},\ 2,\ 2,\ \mathbf{2},\ 3,\ 3,\ \mathbf{3},\ 3,\ 4\}$, donde $C_1 = C_2 = 2$ y $C_3 = 3$ (en negrita). ¡Compruébelo!

Por definición, se afirma que "el 75 % de los datos es mayor que $C_1 = 2$". No obstante, en rigor esto **no siempre es cierto**, ¿en qué caso siempre se cumple?. La razón de esto se haya en que no hemos considerado los valores inmediatamente adyacentes a C_1. En este caso particular, dichos valores son iguales al del primer cuartil. Así, la afirmación correcta sería "el 90,9 % de los datos es **mayor o igual** que $C_1 = 2$". Advierta que la diferencia porcentual entre ambas afirmaciones es significativa.

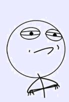
Para evitar realizar cálculos inncesarios y puesto que no siempre se cuenta con el conjunto de datos, las afirmaciones se generalizan a fin que sean matemáticamente correctas. Así, para el conjunto S se puede afirmar que **al menos** el 75 % de los datos es **mayor o igual** que $C_1 = 2$.

Por la misma razón expuesta, se afirma que **al menos** el 50 % de los datos es **mayor o igual** que $C_2 = 2$ y que **al menos** el 25 % de los datos es **mayor o igual** que $C_3 = 3$.

La expresión para el caso complementario, es decir, cuando el porcentaje de los datos se encuentra por debajo del cuartil $k-$ésimo, se deja propuesto.

Finalmente, el mismo razonamiento anterior aplica para cada uno de los cuantiles.

Desafío I



¿Qué sucede con los cuartiles de un conjunto ordenado de forma decreciente?
Respuesta

✎ ***Ejemplo***

1. En una universidad hay dos grupos de porristas: uno conformado exclusivamente por hombres (grupo H) y el otro exclusivamente por mujeres (grupo M), ambos con la misma cantidad de integrantes. Se sabe que el primer cuartil de estaturas para H y M es 1,58 m y 1,63 m respectivamente, el segundo cuartil para ambos grupos es 1,70 m, y el tercer cuartil para H y M es 1,82 m y 1,75 m respectivamente. Al respecto, ¿cuál(es) de las siguientes afirmaciones es (son) verdadera(s)?

   I) Al menos el 75 % de los porristas del grupo de varones mide 1,82 m o menos

   II) Al menos el 25 % de las porristas del grupo de damas mide 1,75 m o más

   III) Se espera que las estaturas de los hombres superen gradualmente las estaturas de las mujeres

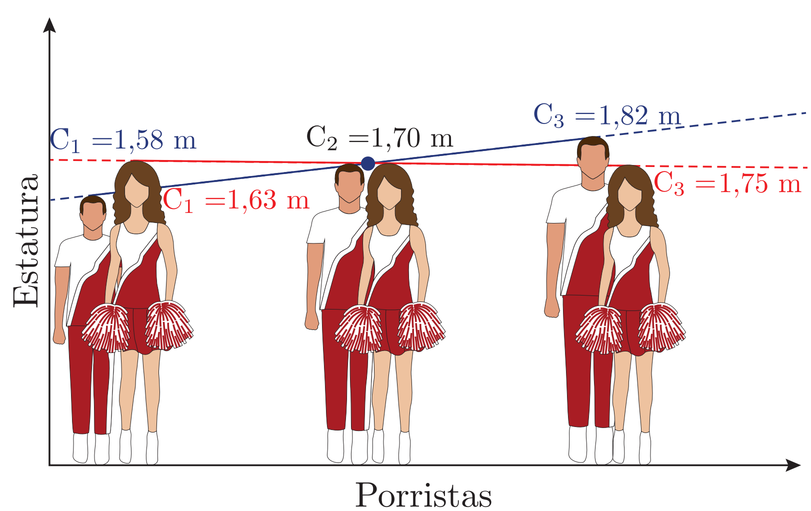
   Solución:
   Veamos la veracidad de cada una de las afirmaciones.

   • *Al menos el 75 % de los porristas del grupo de varones mide 1,82 m o menos.*
   Como se mencionó anteriormente, bajo el tercer cuartil se encuentra el 75 % de los datos. Puesto que el tercer cuartil de las estaturas de H es 1,82 m, se deduce que **al menos** el 75 % de los porristas del grupo de varones mide 1,82 m o menos. Destacamos "al menos" porque no conocemos el valor de las estaturas próximas a C_3, como se explicó previamente.

   • *Al menos el 25 % de las porristas del grupo de damas mide 1,75 m o más.*
   Como se mencionó anteriormente, sobre el tercer cuartil se encuentra el 25 % de los datos. Puesto que el tercer cuartil de las estaturas de M es 1,75 m, se deduce que **al menos** el 25 % de las porristas del grupo de damas mide 1,75 m o más. Nuevamente, advierta el uso de "al menos".

   • *Se espera que las estaturas de los hombres superen gradualmente las estaturas de las mujeres.*
   La siguiente gráfica se ha elaborado a partir de los datos proporcionados por el enunciado del problema. A partir de ella se puede ver un crecimiento gradual en las estaturas de ambos grupos, superando los hombres a las mujeres a partir del segundo cuartil.

   Tenga presente que la gráfica es una **proyección** de los datos del enunciado, y como tal no es necesariamente cierta. Sin embargo, resulta lógica a la naturaleza del problema.



De esta manera, se puede concluir que todas las afirmaciones son verdaderas.

2. La siguiente tabla muestra la cantidad de jugadores de un equipo de fútbol según su edad en años:

Edad (en años)	**Cantidad de jugadores**
17	1
18	2
19	1
20	3
21	3
22	2
23	4
24	2
25	1
26	2
27	1

Si al 25 % de los jugadores de mayor edad del equipo se le entrega un bono motivacional de \$35.000 por partido y se juegan 4 partidos al mes, ¿cuánto dinero invierte mensualmente en bonos el club?

Solución:
El 25 % de mayor edad del equipo corresponde a aquellos jugadores cuya edad se encuentra sobre el tercer cuartil. Usamos $\mathrm{P}_k = \frac{kn}{4}$, con $k = 1, 2, 3$ para encontrar la posición de este valor, sabiendo que en total hay 22 jugadores:

$$\mathrm{P}_3 = \frac{3 \cdot 22}{4} = \frac{66}{4} = 16,5$$

El tercer cuartil corresponde a la edad donde la frecuencia absoluta acumulada es mayor a $16,5$. Esto significa que el tercer cuartil es 24 años. Por lo tanto, el 25 % mayor del equipo son aquellos jugadores que tienen 25, 26 y 27 años, más un jugador con 24 años. Dentro del equipo hay 5 jugadores que satisfacen esta condición, por lo tanto, la inversión mensual I del club es:

$$\mathrm{I} = 5 \cdot 4 \cdot \$35.000 = \$700.000$$

Nota: Cuando un cuartil recae en un valor que se repite más de una vez, la medida de posición no central corresponde a una de las repeticiones.

La diferencia entre el tercer y primer cuartil se conoce como **rango intercuartil** *y es una estimación estadística de la dispersión de la distribución de datos. Mediante esta medida se eliminan los datos extremadamente alejados y su uso es altamente recomendable cuando la medida de tendencia central utilizada es la mediana.*

1.2. Quintiles

Los quintiles, que llamamos Q_1, Q_2, Q_3, y Q_4, corresponden a los cuatro valores que dividen a un conjunto de datos en cinco partes porcentualmente iguales. Esta medida es ampliamente usada para clasificar a la población por niveles de ingresos económicos. Las personas que se encuentran bajo el primer quintil corresponden al 20 % de la población con menores ingresos, mientras que las personas que se encuentran sobre el cuarto quintil corresponden al 20 % con mayores ingresos.

En Chile, los quintiles están distribuidos de la siguiente manera[1]:

Quintil	Desde	Hasta
1	\$0	\$70.966
2	\$70.967	\$118.854
3	\$118.855	\$182.793
4	\$182.794	\$333.909
5	\$333.910	-

Nota: Los quintiles corresponden a los cuatro valores que vemos en la tercera columna, pero comúnmente se denomina quintil a cualquiera de los cinco grupos que podemos apreciar en la tabla.

La forma para calcular el quintil al que una persona pertenece es la siguiente:

- Se suman los ingresos de los integrantes del grupo familiar, menos los descuentos legales (salud, AFP e impuestos).
- Se divide el monto total por el número de integrantes del grupo familiar.
- El resultado corresponde al ingreso per cápita del grupo familiar.

Si el ingreso per cápita de cierto grupo familiar es \$200.000, entonces dicha familia pertenece al cuarto quintil.

✎ ***Ejemplo***

Pertenecer a cualquiera de los cuatro primeros quintiles es uno de los requisitos para participar en el proceso de admisión de una destacada universidad. Si Verónica desea postular y su grupo familiar está compuesto por 5 personas, ¿cuánto deben sumar, como máximo, los ingresos de todos los integrantes de su familia?

Solución: Verónica pertenecerá a cualquiera de los primeros cuatro quintiles si el ingreso per cápita de su grupo familiar no supera los \$333.909. Por lo tanto, debemos planter la siguiente desigualdad:

$$\frac{x}{5} \leq \$333.909$$

Donde x corresponde a la suma de los ingresos de los integrantes de su grupo familiar. Para resolver esta inecuación, multiplicamos por 5 a ambos lados y obtenemos que:

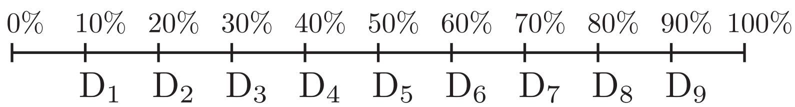
$$x \leq \$1.669.545$$

Por lo tanto, la suma de los ingresos de todos los integrantes de la familia de Verónica no puede superar los \$1.669.545, de lo contrario no cumplirá con uno de los requisitos para postular a dicha universidad.

[1] Fuente: Becas y créditos, Ministerio de Educación.

1.3. Deciles

Los deciles, que llamamos D_1, D_2, $\ldots$, D_9, corresponden a los nueve valores que dividen un grupo de datos en diez partes porcentualmente iguales. Cada una de estas partes representa un décimo del total. De esta forma el primer decil corresponde al valor donde está el 10 % de los datos, el segundo decil donde está el 20 % de los datos, y así sucesivamente.



Note que el quinto decil corresponde a la mediana del grupo de datos, pues divide al conjunto exactamente en dos partes iguales. En un grupo de datos, el quinto decil es equivalente al segundo cuartil.

La **posición** P_k del decil k-ésimo de una muestra de n datos es:

$$P_k = \frac{kn}{10}, \qquad \text{con } k = 1, 2, 3, \ldots, 9 \tag{2}$$

1.4. Percentiles

Los percentiles o centiles, que llamaremos[2] S_1, S_2, $\ldots$, S_{99}, corresponden a los 99 valores que dividen a un conjunto en cien subconjuntos de igual tamaño. Al igual que el resto de los cuantiles, por debajo del percentil k-ésimo queda el k % de los datos y por sobre él queda el $(100 - k)$ % de la muestra.

En una muestra de datos, el percentil 50 coincide con el quinto decil y con el segundo cuartil (mediana). De esta misma forma, el percentil 25 coincide con el primer cuartil y el percentil 75 con el tercer cuartil.

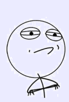
La expresión que nos permite obtener la posición P_k del percentil k-ésimo está dada por:

$$P_k = \frac{kn}{100}, \qquad \text{con } k = 1, 2, 3, \ldots, 99 \tag{3}$$

La forma en que se calculan los percentiles es igual a como se calculan los porcentajes, pero la gran ventaja de los percentiles es que éstos entregan información sobre la ubicación de cada dato con respecto al resto de la muestra.

Cuando los datos se ordenan en una tabla, los percentiles están ubicados en la columna de porcentajes acumulados.

Desafío II



El percentil 25 es equivalente al primer cuartil, el percentil 50 es equivalente al segundo cuartil y al quinto decil. ¿Existen más relaciones de este tipo entre las medidas de posición? Respuesta

[2] Generalmente se denotan como P_k, pero usaremos esta notación para no confundirnos con la posición que ocupan dentro de un conjunto de datos.

✎ ***Ejemplo***

En la PSU de Matemática, Natalia obtuvo 770 puntos y es parte del percentil 94 del grupo formado por todos los estudiantes que rindieron la prueba el año 2013. Esto significa que al menos el 6 % de los postulantes que rindieron la prueba obtuvo un puntaje mayor o igual que el de Natalia, o equivalentemente, al menos el 94 % obtuvo un puntaje menor o igual.

Desafío III

Una encuesta realizada aleatoriamente a 300 personas de Santiago busca recoger información acerca de la cantidad de monedas que traen en el bolsillo en cierto instante del día. Los resultados se resumen en la siguiente tabla:

Moneda	**Número de personas**
\$1	1
\$5	3
\$10	70
\$50	128
\$100	512
\$500	91

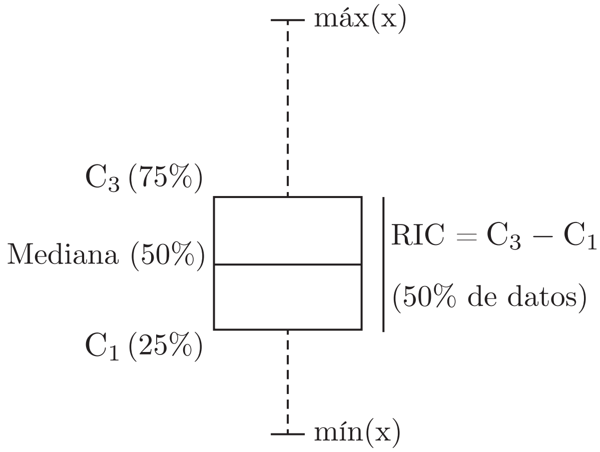
¿Cuál es el valor de cada cuartil, de los deciles 7 y 9, y de los percentiles 40 y 62? Sin realizar cálculos, ¿cuál es el valor esperado para cada una de estas medidas de posición? Respuesta

Se recurre a las medidas de posición cuando la medida de tendencia central usada para estudiar la muestra de datos corresponde a la mediana, y la elección de la medida de posición más adecuada depende de la forma en que deseamos clasificar los datos y del número de observaciones, principalmente. Por ejemplo, si un conjunto de datos tiene 10 elementos no es recomendable usar deciles, pues no entregarán información relevante, en este caso sería más pertinente usar los cuartiles.

1.5. Representación gráfica

La forma de representar gráficamente la información que entrega un conjunto de datos es variada; puede ser mediante polígonos de frecuencias, gráficos de barra, entre otros. Las medidas de posición no central pueden ser representadas de la misma forma, ubicando en el eje horizontal cada cuantil y en el eje vertical su valor.

Para representar los cuartiles de un conjunto de datos ordenados, se utilizan los llamados **diagramas de caja**, como el que se muestra a continuación:



Donde RIC corresponde al rango intercuartil. A través de este gráfico es posible visualizar el conjunto de datos que se está estudiando y su distribución según las medidas de posición. En los extremos del gráfico se encuentran los valores mínimo y máximo del conjunto y los tres segmentos paralelos de la caja representan a cada cuartil. Cuando la distribución es simétrica, la mediana (segmento central en la caja) se ubica justo en el medio de la figura.

Podemos apreciar que dentro de la caja (desde C_1 hasta C_3) se encuentra el 50 % de la información.

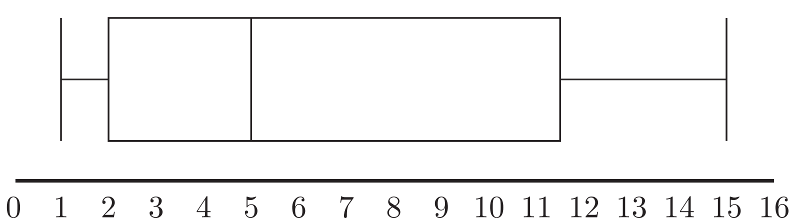
✎ ***Ejemplo***

1. Elabora un diagrama de caja y bigotes para el conjunto $\{15, 2, 3, 1, 2, 7, 9, 14\}$.

   Solución:
   Ordenamos los elementos del conjunto de forma creciente:

   $$1 - 2 - 2 - 3 - 7 - 9 - 14 - 15$$

   El valor mínimo del conjunto es 1, mientras que el máximo es 15. Los valores de cada cuartil son 2, 5 y $11,5$ respectivamente. Por lo tanto, el diagrama queda de la siguiente manera:

2. Un conjunto está formado por 100 números naturales desde el 1 hasta el 5. El primer cuartil es 2, el primer decil es 1 y el percentil 40 es 4. Representa esta información en un polígono de frecuencias.

Solución:
Usamos (1) para encontrar la posición del primer cuartil dentro de la muestra de datos ordenada:

$$P_1 = \frac{1 \cdot 100}{4} = \frac{100}{4} = 25$$

Luego, el primer cuartil queda determinado por la posición 25 de la muestra de datos ordenada de forma creciente, pero el número de datos es par, por lo tanto, debemos considerar la media aritmética entre los datos 25 y 26 para obtener el valor del cuartil y debe cumplirse que:

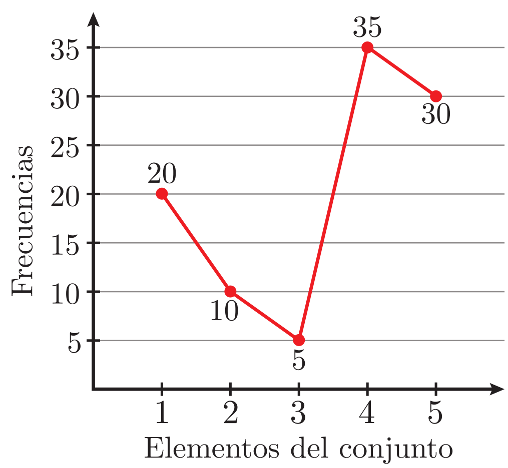
$$\frac{P_{25} + P_{26}}{2} = 2 \Longrightarrow P_{25} + P_{26} = 4$$

Es decir, la suma de los valores ubicados en la posición 25 y 26 debe ser igual a 4. Para saber cómo podemos construir un gráfico con esta información es recomendable ir anotando los valores en una tabla de frecuencias acumuladas. Ahora, la posición del primer decil la obtenemos usando (2):

$$P_1 = \frac{1 \cdot 100}{10} = 10$$

El primer decil está en la posición 10 y es igual a 1. De forma análoga, usando (3) obtenemos que el percentil 40 ocupa la posición 40 y es igual a 4.

Todos estos datos los representamos en una gráfica de frecuencias, de manera tal que la frecuencia acumulada en el puesto 10 sea igual a 1, entre los puestos 25 y 26 sea igual a 2, y en el puesto 40 sea igual a 4. Existen muchos conjuntos de números que satisfacen esta condición, uno de ellos es el que se muestra en la figura:



Esta es solo una representación de la forma en que podrían estar distribuidos los datos del conjunto, pero no es la única.

✍ *Ejercicios*

1. Los cuartiles C_1, C_2 y C_3 del conjunto $\{33, 5, 54, 12, 8, 8, 33, 4, 6, 5\}$ son respectivamente:

   A) 33, 8 y 5
   B) 5, 8 y 33
   C) 4, 8 y 54
   D) 6, 12 y 54
   E) 4, 8 y 33

2. La siguiente tabla muestra las estaturas, en metros, de un grupo de niñas:

Estatura (metros)	**Número de personas**
$1,52$	2
$1,53$	4
$1,54$	4
$1,55$	7
$1,56$	3
$1,58$	6
$1,60$	4
$1,61$	4
$1,63$	2
$1,66$	3
$1,67$	3
$1,73$	2

   Respecto a la tabla anterior, ¿cuál(es) de las siguientes afirmaciones es (son) verdadera(s)?

   I) El primer cuartil es $1,54$ m
   II) La diferencia entre el segundo y el primer cuartil es 3 cm
   III) Exactamente la mitad del curso mide entre $1,55$ y $1,63$ m

   A) Solo II
   B) Solo I y II
   C) Solo I y III
   D) Solo II y III
   E) I, II y III

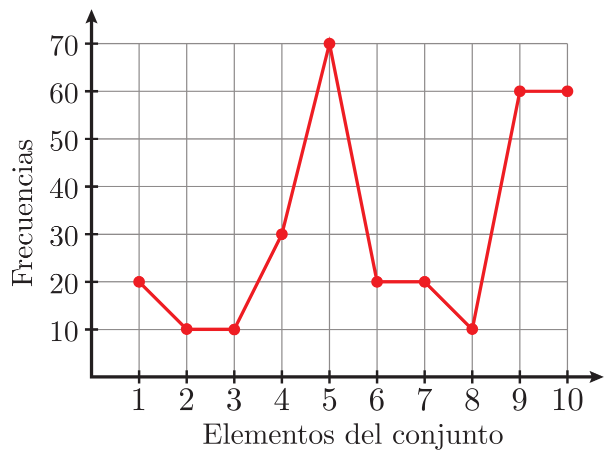
3. En un conjunto de datos ordenados de menor a mayor, el segundo quintil es equivalente al

   A) primer cuartil.
   B) segundo decil.
   C) cuarto decil.
   D) segundo percentil.
   E) cuarto percentil.

4. El sueldo bruto del padre de Jorge es $900.000 y su sueldo líquido es el 88 % del sueldo bruto. Si su grupo familiar está compuesto de tres personas y solo el padre trabaja, ¿a qué quintil pertenece la familia de Jorge?

   A) Primero
   B) Segundo
   C) Tercero
   D) Cuarto
   E) Quinto

5. Un conjunto está formado por números enteros de 1 al 10. El siguiente gráfico muestra la distribución de los números:



   ¿Cuál es la suma de los cuartiles de la distribución anterior?

   A) 20
   B) 30
   C) 60
   D) 90
   E) 150

6. La edad de los participantes de un concurso de talentos oscila entre los 6 y los 55 años. Si los cuartiles de esta población son 18, 34 y 43 años, ¿en qué tramo etario se concentra el 50 % central de los participantes?

   A) Entre los 18 y 34 años
   B) Entre los 18 y 43 años
   C) Entre los 34 y 43 años
   D) Entre los 9 y 16 años
   E) Entre los 16 y 25 años

Claves de los ejercicios propuestos.

Desafíos resueltos

✓ Desafío I:
Al ordenar los datos de un conjunto de forma decreciente, el valor de la mediana se mantiene constante. La diferencia es que la ubicación de los cuartiles se invierte de manera simétrica respecto al segundo cuartil, es decir, existe una simetría en los valores con respecto a la mediana. La elección de la forma en que se ordenan los datos depende exclusivamente del análisis que se desee hacer, pero en general se utiliza el orden creciente de los datos. Volver

✓ Desafío II:
En una muestra de datos ordenados el tercer cuartil es equivalente al percentil 75, el primer quintil es equivalente al segundo decil y al percentil 20, el segundo quintil es equivalente al cuarto decil y al percentil 40, el tercer quintil es equivalente al sexto decil y al percentil 60, y el cuarto quintil es equivalente al octavo decil y al percentil 80.

También podemos observar que el primer decil es equivalente al percentil 10, el segundo decil es equivalente al percentil 20 y así sucesivamente. Volver

✓ Desafío III:
Que la encuesta se haya realizado a 300 personas es solo un distractor, pues son los datos de la tabla los que necesitamos para determinar el valor de cada medida de posición. Para el desarrollo del problema es recomendable adjuntar la columna de frecuencias acumuladas a la tabla:

Moneda	Número de personas	Frecuencia acumulada
\$1	1	1
\$5	3	4
\$10	70	74
\$50	128	202
\$100	512	714
\$500	91	805

Usamos (1) para calcular la posición de cada cuartil en el conjunto de datos:

$$\mathrm{P}_1 = \frac{1 \cdot 805}{4} = \frac{805}{4} = 201,25$$

$$\mathrm{P}_2 = \frac{2 \cdot 805}{4} = \frac{1.610}{4} = 402,5$$

$$\mathrm{P}_3 = \frac{3 \cdot 805}{4} = \frac{2.415}{4} = 603,75$$

El primer cuartil es \$50, mientras que el segundo y tercer cuartil es \$100. ¿Te sorprende este resultado? Ahora usamos (2) para calcular la posición de los deciles 7 y 9, nos queda:

$$\mathrm{P}_7 = \frac{7 \cdot 805}{10} = \frac{5.635}{10} = 563,5$$

$$\mathrm{P}_9 = \frac{9 \cdot 805}{10} = \frac{7.245}{10} = 724,5$$

Luego, los deciles 7 y 9 son \$100 y \$500 respectivamente. Por último, usamos (3) para calcular la posición de los percentiles 40 y 62:

$$\mathrm{P}_{40} = \frac{40 \cdot 805}{100} = \frac{32.200}{100} = 322$$

$$\mathrm{P}_{62} = \frac{62 \cdot 805}{100} = \frac{49.910}{100} = 499,1$$

Ambos percentiles son iguales a \$100. Note que el dato de mayor frecuencia es \$100 y corresponde aproximadamente al 64 % del total de los datos. Este valor a la vez corresponde a la mediana de la muestra y gran parte de la distribución está concentrada en este valor, siendo el valor esperado al calcular cada medida de posición. Volver

✓ Claves de los ejercicios propuestos:

1	2	3	4	5	6
B	A	C	D	A	B

Volver

Bibliografía

[1] Introducción a la Estadística, *Primera edición*, JC Sáez Editor (2011)
Nancy Lacourly

[2] Probabilidad y Estadística para ingeniería y ciencias, *cuarta edición*, Prentice-Hall Hispanoamericana, S.A. (1997)
William Mendenhall, Terry Sincich