

Palabras clave

- Medidas de posición, polígonos de frecuencia acumulada.
- Diagramas de caja.

Repasa

Porcentaje

Para calcular el P % de N utiliza la siguiente expresión:

$$x = \frac{P}{100} \cdot N$$

Ejemplo: Calcular el 5% de 40.

$$x = \frac{5}{100} \cdot 40$$

$$x = \frac{5}{100} \cdot \frac{40}{1} = \frac{200}{100} = 2$$

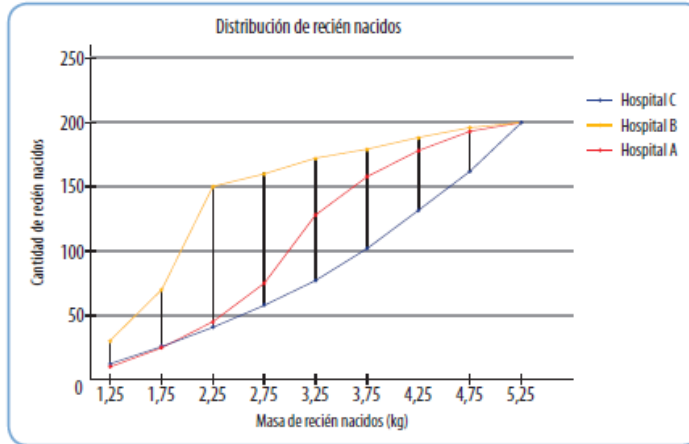
Es decir, el 5% de 40 es 2.

¿Cómo se interpretan las medidas de posición?

- Es posible que hayas conocido a recién nacidos que han nacido con una estatura por sobre el normal de los niños y lo más probable es que el doctor haya mencionado que está en el percentil 90. ¿Cómo interpretas esta información?

Interpretación de medidas de posición a partir de polígonos de frecuencias acumuladas

Milena es la nutricionista encargada de estudiar las masas de los recién nacidos en tres hospitales. Para realizar su estudio analiza el polígono de frecuencias acumuladas en el que se ha utilizado la marca de clase.



Milena determina que se debe proporcionar una incubadora al hospital donde el P_{10} es menor que 2,25 kg, ¿cuál o cuáles hospitales deben recibirla?

Además, Milena está diseñando un programa de seguimiento a todos los niños y niñas que al momento de nacer, lo hicieron con una masa hasta el primer cuartil de cada hospital. ¿Bajo qué masa deben estar los recién nacidos para pertenecer al programa?

- Paso 1** ➔ Calcular los porcentajes respectivos a las medidas de posición requeridas. P_{10} corresponde al valor bajo el cual está el 10% de los recién nacidos. $P_{25} = Q_1$ corresponde al valor bajo el cual está el 25% de los recién nacidos. Luego, calculamos estos porcentajes:

Total: 200 10% de 200 es 20 25% de 200 es 50

- Paso 2** ➔ Estimar el valor de las medidas de posición respectivas a los porcentajes anteriores.

Veamos para el hospital A, para determinar el valor de las medidas de posición debemos ubicar en la frecuencia absoluta los porcentajes encontrados. Las estimaciones requeridas corresponderán a los valores de la variable correspondientes a las frecuencias encontradas.

- Al ubicar la frecuencia 20, la curva de color rojo pasa por 1,8 kg, por lo tanto, el P_{10} corresponde, aproximadamente, a ese valor.
- Al ubicar la frecuencia 50, la curva de color rojo pasa por la masa 2,3 kg, por lo tanto, el primer cuartil corresponde, aproximadamente, a ese valor.

Repetiendo el procedimiento para los otros hospitales se tienen las siguientes medidas de posición.

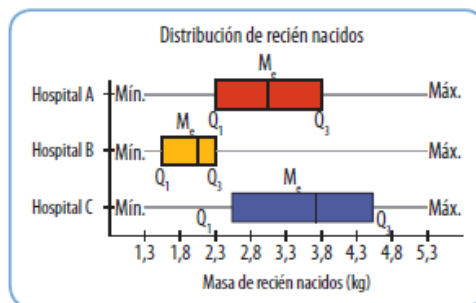
	Hospital A	Hospital B	Hospital C
P_{10}	1,8 kg	1,25 kg	1,8 kg
$P_{25} = Q_1$ (1er cuartil)	2,3 kg	1,5 kg	2,5 kg

Al determinar P_{10} se observa que en todos los hospitales este es menor que 2,25 kg, por lo tanto, los tres deberían recibir la incubadora.

Por otra parte, los recién nacidos que pertenecerán al programa del hospital A deben estar bajo los 2,3 kg. Los del hospital B deben estar bajo el 1,5 kg y los del hospital C, bajo los 2,5 kg, todos en forma aproximada.

Interpretación de medidas de posición a partir de diagramas de cajas o Boxplot

Para su estudio, Milena necesita comparar la mediana de las masas de los recién nacidos y determinar en qué hospital el rango intercuartil es menor. Para esto consulta el siguiente boxplot o diagrama de cajas donde se muestra la distribución de los datos según el hospital.



Paso 1 Comparar las medianas en los diagramas de caja.

Al comparar las medianas se concluye que el Hospital B es el que tiene un valor menor (aproximadamente 1,75 kg) y el Hospital C el que tiene un mayor valor de (aprox. 3,75 kg).

Paso 2 Comparar el rango intercuartil para cada boxplot.

Al comparar el tamaño de las cajas se observa que el hospital B presenta un rango intercuartil menor que los otros hospitales.

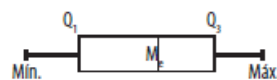
Razona y comenta

- ¿Cuál es la utilidad de un diagrama de caja? ¿Qué se puede inferir sobre un conjunto de datos al analizar este tipo de gráfico?
- ¿Cuál es la diferencia entre rango y rango intercuartil? ¿Cuál de las dos medidas dice más sobre la concentración de los datos?

En resumen

Podemos **interpretar** las medidas de posición a partir de un polígono de frecuencia acumulada o de un **diagrama de cajas o Boxplot** que es un gráfico que muestra la distribución de los datos, dividiendo estos en cuatro partes iguales mediante los cuartiles.

Para construir un **Boxplot** se dibuja una caja que va desde Q_1 hasta Q_3 . Dentro de ella se traza una línea vertical en la mediana. Luego, se trazan líneas desde la caja a los valores mínimo y máximo.



Se llama **rango intercuartil** a la diferencia entre el tercer cuartil (Q_3) y el primer cuartil (Q_1).

Repaso

1. **Analiza** la información presentada en la tabla y luego responde.

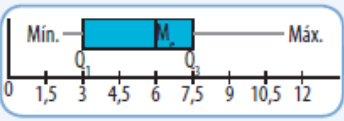
Andrea realiza un estudio sobre la cantidad de hermanos que tienen los alumnos de tres cursos, y resume la información en la siguiente tabla.

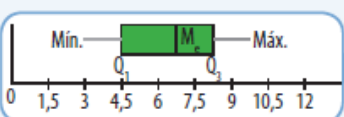
Número de hermanos	Curso A	Curso B	Curso C
0	4	4	66
1	15	5	9
2	25	7	7
3	29	9	7
4	15	24	5
5	10	49	4

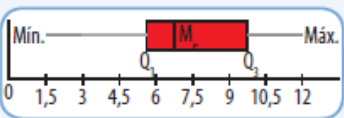
- Construye el gráfico de frecuencia acumulada de la tabla de datos.
- ¿Cuál es el primer cuartil de los datos de cada curso?
- ¿Cuál es el segundo cuartil de los datos de cada curso?
- ¿Cuál es el tercer cuartil de los datos de cada curso?

Práctica guiada

2. **Relaciona** el diagrama de caja con el conjunto de datos representados en él.

- 

Mín: 3
 Q_1 : 4,5
 M_e : 6,75
 Q_3 : 8,25
 Máx: 9,75
- 

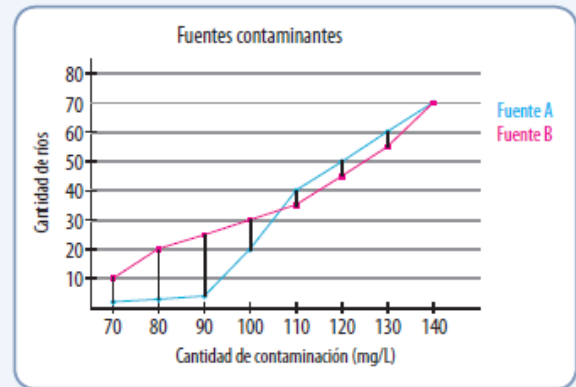
Mín: 1,5
 Q_1 : 6
 M_e : 6,75
 Q_3 : 9,75
 Máx: 12
- 

Mín: 2,25
 Q_1 : 3
 M_e : 6
 Q_3 : 7,5
 Máx: 11,25

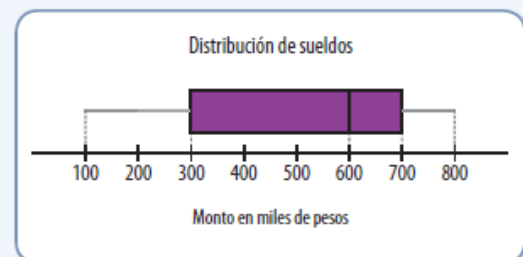
Aplico

3. **Analiza** la información presentada en el gráfico y luego responde.

- Patricio se encuentra estudiando la contaminación que producen dos fuentes: A y B a los ríos de una región. La cantidad de contaminación producida por cada fuente se encuentra representada en el siguiente gráfico.

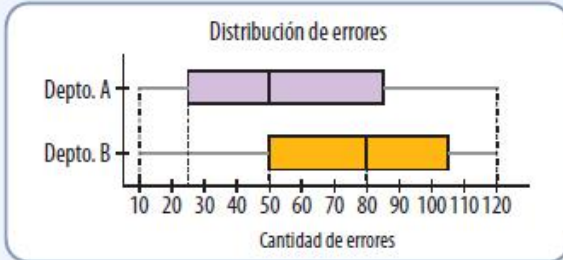


- Si Patricio necesita realizar un informe que describa la distribución de los datos y la comparación entre las fuentes A y B, ¿qué información debería contener el documento? Descríbela.
- Antonia debe determinar las remuneraciones de una empresa. Para esto utiliza el diagrama de caja que aparece a continuación.



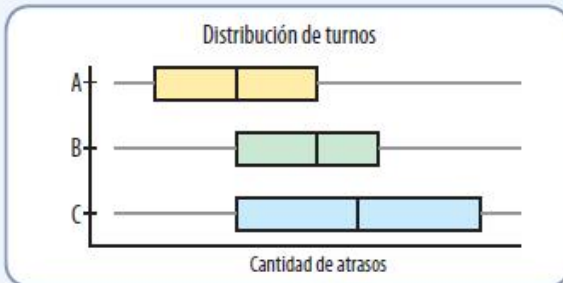
- ¿Cuál es el sueldo máximo que paga la empresa?
- Si el número de trabajadores de la empresa es de 2000, ¿cuántos de ellos reciben un sueldo igual o inferior a \$600 000?

- c) Ramón es el encargado de realizar un estudio sobre los errores en la ortografía cometidos en dos departamentos de una editorial. Para esto construye el siguiente gráfico en el que se exhibe la distribución de los errores para una muestra de 100 trabajos para el departamento A y B.



- Ramón realizará una capacitación a los trabajadores del departamento editorial cuyo Q_2 supere los 40 errores. ¿En cuál departamento se realizará la capacitación?
- Ramón debe realizar un informe comparando los cuartiles y la media aritmética entre los departamentos. ¿Cuál sería dicho informe?
- Si los departamentos A y B trabajan con la misma cantidad de personal y en las mismas condiciones, ¿cuál es el departamento más eficiente?

- d) Romeo y Elizabeth perdieron los datos de tres variables A, B y C correspondiente a los atrasos de tres turnos de una empresa. Solamente poseen el siguiente diagrama de caja:



Si deben confirmar un informe que contiene las afirmaciones que aparecen a continuación, ¿cuál de ellas es la correcta? Corrige aquellas que sean erróneas para que puedan ser incluidas en el informe.

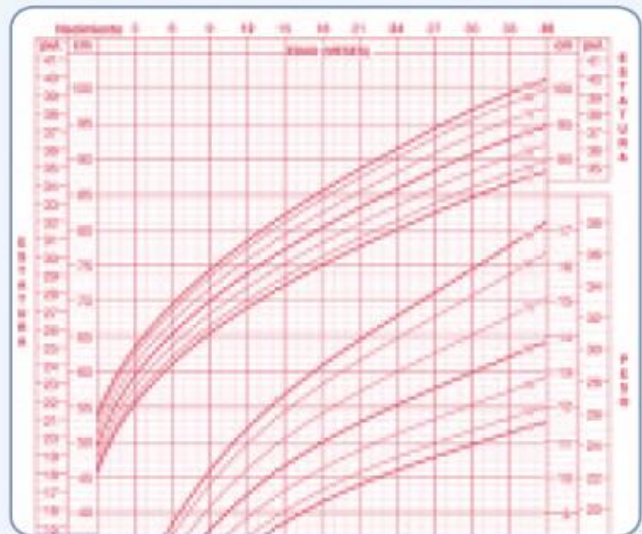
Reflexión

- ¿Qué ventajas y desventajas tiene analizar un conjunto de datos a partir de las medidas de posición, respecto a las medidas de tendencia central?

- El primer cuartil del turno A es mayor que el primer cuartil de turno B.
- El segundo cuartil del turno C es mayor que la mediana de los datos de los turnos A y B por separado.
- El segundo cuartil de A es mayor que el segundo cuartil de C.
- La mediana de B es igual a la mediana de C.
- El rango intercuartil del turno A es el mayor.

4. **Conecta.** Los pediatras utilizan ciertas tablas que permiten determinar la relación entre el peso y talla de un bebé, uno de estos gráficos corresponde a la curva de crecimiento fetal.

A partir del gráfico extrae 3 conclusiones relacionadas con medidas de posición.



5. **Crea.** Utiliza los datos del INE para analizar la distribución de algunos que sean de tu interés. ¿Consideras que es más sencillo analizarlos con medidas de posición?

Refuerzo

En la empresa donde trabaja Antonia (pregunta 3 b.), trabajan 2000 personas, ¿cuántos de ellos están dentro del rango intercuartil?