



# INSTITUCIÓN EDUCATIVA NUESTRA SEÑORA DEL PALMAR

Código; FR 202 GA  
Versión: 001  
Emisión: 2020-08-6  
Actualización:

## GUÍA DE APRENDIZAJE

GUÍA No: 5	ÁREA: MATEMÁTICAS	ASIGNATURA: ESTADÍSTICA
PERIODO DE COBERTURA DESDE: 13 DE JUNIO		HASTA: 5 DE AGOSTO
FECHA DE RECEPCIÓN DEL ENTREGABLE: SEMANA DEL 1 AL 5 DE AGOSTO		
DOCENTE: MARIA ISLANDIA ESPINOSA SANCHEZ Y SUBLEYMAN IVONNE USMAN NARVÁEZ		
ESTUDIANTE:		GRUPO: DECIMO

### ¿QUÉ VOY A APRENDER?

A lo largo de esta guía se logrará:

- Calcular e interpretar medidas de posición para un conjunto de datos agrupados
- Indicar el lugar o posición donde se ubica un valor de la variable dentro de un conjunto de datos ordenado de valores.
- Representar mediante un diagrama de caja las medidas de posición de un conjunto de datos no agrupados.

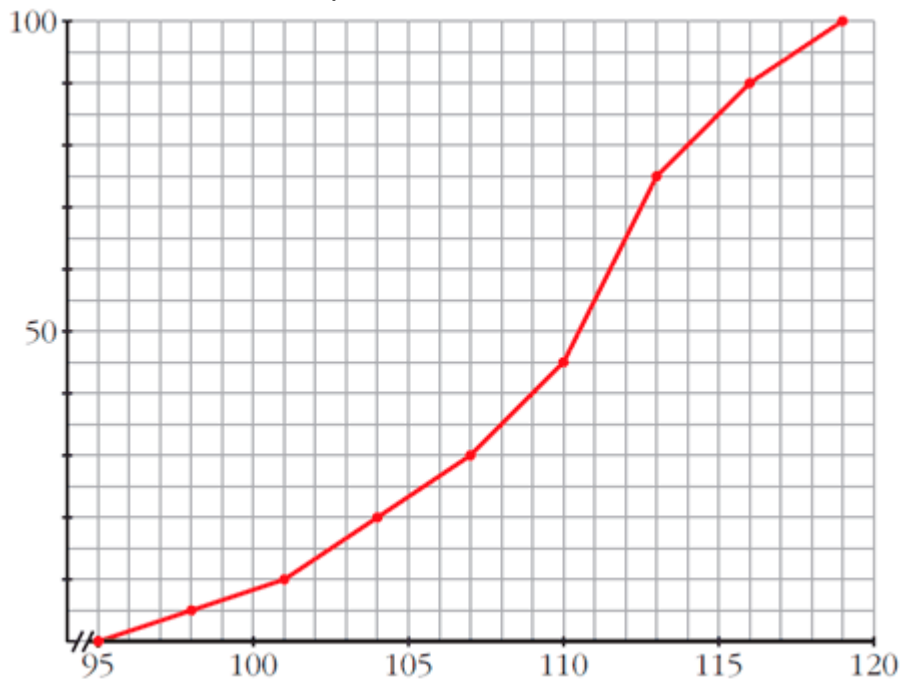
**Reflexionemos sobre esta frase: “El pesimismo lleva a la debilidad. El optimismo, al poder”**

William James

### LO QUE ESTOY APRENDIENDO

#### ACTIVO MIS CONOCIMIENTOS PREVIOS:

Este es el polígono de porcentajes acumulados de la distribución del CI (coeficiente intelectual) de un colectivo de 200 personas.



- a) Trabajando sobre el gráfico, asigna, aproximadamente, los valores de:  
 $Q_1$ , Me,  $Q_3$ ,  $P_5$ ,  $P_{10}$ ,  $P_{40}$ ,  $P_{80}$ ,  $P_{90}$ ,  $P_{95}$
- b) ¿Cuántas personas (aproximadamente) de este colectivo tienen un CI comprendido entre 101 y 110? ¿Y inferior a 107?
- c) ¿Qué percentil tiene una persona con un CI de 116?

## MEDIDAS DE POSICION PARA DATOS AGRUPADOS

### CUANTILES PARA DATOS AGRUPADOS

#### Cuartiles

En esta guía de aprendizaje estudiaremos los cuartiles para datos agrupados. Recordemos que las medidas de posición indican el lugar o posición donde se ubica un valor de la variable dentro de un conjunto ordenado de valores.

**Los Percentiles** son los 99 valores de la variable que dividen a un conjunto de datos ordenados en 100 partes iguales (de 1% cada parte). De manera que para resolver un problema sobre percentiles solamente tenemos que hallar  $P_1$ ;  $P_2$ ;  $P_3$ ;  $P_4$ ; ... ;  $P_{99}$  NOTA: Para hallar los Percentiles se sigue el mismo procedimiento que los Cuartiles. Para la obtención de las medidas de posición en datos agrupados, es decir los percentiles, cuartiles y deciles, se hace uso de las siguientes formulas

**Los cuartiles** son una medida de posición que divide un conjunto de datos, ordenados de menor a mayor, en cuatro partes iguales. Cada uno de estos valores se puede interpretar como el valor que acumula por debajo de él, el 25%, el 50% y el 75% de valores iguales o inferiores a él.

**Los Deciles** son los nueve valores de la variable que dividen a un conjunto de datos ordenados en 10 partes iguales (de 10% cada parte). De manera que para resolver un problema sobre deciles solamente tenemos que hallar  $D_1$ ;  $D_2$ ;  $D_3$ ;  $D_4$ ; ...  $D_9$

Percentiles

$$P_k = L_{i-1} + C \left[ \frac{\frac{kn}{100} - F_{i-1}}{f_i} \right]$$

$$k = 1, 2, \dots, 99$$

Cuartiles

$$Q_k = L_{i-1} + C \left[ \frac{\frac{kn}{4} - F_{i-1}}{f_i} \right]$$

$$k = 1, 2, 3$$

Deciles

$$D_k = L_{i-1} + C \left[ \frac{\frac{kn}{10} - F_{i-1}}{f_i} \right]$$

$$k = 1, 2, 3, \dots, 9$$

P: es el número que identifica al percentil dado

Q: es el número que identifica al cuartil dado

D<sub>j</sub>: es el número que identifica al decil dado

n: es el número de datos

L<sub>i</sub>: es el límite real inferior del intervalo donde está el cuartil dado

Fi-1: es la diferencia acumulada antes del intervalo de la clase al cuartil dado

f: es la frecuencia absoluta del intervalo donde está el cuartil dado

c: es el ancho de clase

**Ejemplo:** de acuerdo con la información de la siguiente tabla determina:

Salarios	Número de empleados fi	Fi
200 – 299	85	85
300 – 399	90	175
400 – 499	120	295
500 – 599	70	365
600 – 699	62	427
700 – 799	36	463

Calcular:

- Q1, Q2
- D1, D5
- P30, P50
- Realiza conclusiones

Solución

- Q1, Q2

En primer lugar, buscamos la **clase** donde se encuentra el cuartil solicitado, en la **tabla de las frecuencias acumuladas**

**Para Q1:** Se ubica en la frecuencia acumulada el valor igual o superior a 115.75, es decir 175 posición que corresponde al intervalo de 300 a 399

$$Q_k = L_{i-1} + c \left[ \frac{\frac{kn}{4} - F_{i-1}}{f_i} \right]$$

$k = 1, 2, 3$

$$\text{Para el } Q1 = \frac{463}{4} = 115,75$$

$$L_i = 300$$

$$F_{i-1} = 85$$

$$C = 100$$

$$f_i = 90$$

$$Q1 = 300 + \left( \frac{115,75 - 85}{90} \right) * 100$$

$$Q1 = 300 + \left( \frac{30,75}{90} \right) * 100$$

$$Q1 = 300 + 34,16$$

$$Q1 = 334$$

se concluye que: Este resultado indica que el 25% de los empleados ganan salarios por debajo de 334

**Para Q2:** Se ubica en la frecuencia acumulada el valor igual o superior a 231,5, es decir 295 posición que corresponde al intervalo de 400 a 499

Para el  $Q_2 = \frac{2 \cdot 463}{4} = 231,5$

$L_i = 400$

$F_{i-1} = 175$

$C = 100$

$f_i = 120$

$$Q_k = L_{i-1} + C \left[ \frac{\frac{kn}{4} - F_{i-1}}{f_i} \right]$$

$k = 1, 2, 3$

$$Q_2 = 400 + \left( \frac{231,5 - 175}{120} \right) * 100$$

$$Q_2 = 400 + \left( \frac{56,5}{120} \right) * 100$$

$$Q_2 = 400 + 47,086$$

$$Q_2 = 447,08$$

se concluye que: Este resultado indica que el 50% de los empleados ganan salarios por debajo de 447,08 y el 50% de los empleados ganan salarios por encima de 447,08

**b. D5 y D7**

Para  $D_5 = \frac{5 \cdot 463}{10} = 231,5$

$L_i = 400$

$F_{i-1} = 175$

$C = 100$

$f_i = 120$

$$D_k = L_{i-1} + C \left[ \frac{\frac{kn}{10} - F_{i-1}}{f_i} \right]$$

$k = 1, 2, 3, \dots, 9$

$$D_5 = 400 + \left( \frac{231,5 - 175}{120} \right) * 100$$

$$D_5 = 400 + \left( \frac{56,5}{120} \right) * 100$$

$$D_5 = 400 + 47,086$$

$$D_5 = 447,08$$

Se concluye que: Este resultado indica que el 50% de los empleados ganan salarios por debajo de 447,08 y el 50% de los empleados ganan salarios por encima de 447,08. D5 y Q2 es la mediana de los datos.

Para el D7=  $\frac{7 \cdot 463}{10} = 324,1$

$$D_k = L_{i-1} + C \left[ \frac{\frac{kn}{10} - F_{i-1}}{f_i} \right]$$

$k = 1, 2, 3, \dots, 9$

$L_i = 500$

$F_{i-1} = 295$

$C = 100$

$f_i = 70$

$$D7 = 500 + \left( \frac{324,1 - 295}{70} \right) * 100$$

$$D7 = 500 + \left( \frac{29,1}{70} \right) * 100$$

$$D7 = 500 + 41,57$$

$$D7 = 541,57$$

Se concluye que estos resultados indican que el 70% de los empleados ganan salarios por debajo de 541,57

**c. P30 y P50**

Para el P30=  $\frac{30 \cdot 463}{100} = 138,9$

$$P_k = L_{i-1} + C \left[ \frac{\frac{kn}{100} - F_{i-1}}{f_i} \right]$$

$k = 1, 2, \dots, 99$

$L_i = 300$

$F_{i-1} = 85$

$C = 100$

$f_i = 90$

$$P30 = 300 + \left( \frac{138,9 - 85}{90} \right) * 100$$

$$P30 = 300 + \left( \frac{53,9}{90} \right) * 100$$

$$P30 = 300 + 59,88$$

$$P30 = 359,88$$

Se concluye que estos resultados indican que el 70% de los empleados ganan salarios por encima de 359,88

Para el P50=  $\frac{50 \cdot 463}{100} = 231,5$

$L_i = 400$

$F_{i-1} = 175$

$C = 100$

$f_i = 120$

$$P_k = L_{i-1} + C \left[ \frac{\frac{kn}{100} - F_{i-1}}{f_i} \right]$$

$k = 1, 2, \dots, 99$

$$P50 = 400 + \left( \frac{231,5 - 175}{120} \right) * 100$$

$$P50 = 400 + \left( \frac{56,5}{120} \right) * 100$$

$$P50 = 400 + 47,086$$

$$P50 = 447,08$$

Se concluye que: Este resultado indica que el 50% de los empleados ganan salarios por debajo de 447,08 y el 50% de los empleados ganan salarios por encima de 447,08.  $D5 = Q2 = P50$ , es la mediana de los datos

### Diagrama de cajas y bigotes

Un diagrama de caja es un gráfico, basado en cuartiles, mediante el cual se visualiza un conjunto de datos.

Está compuesto por un rectángulo, la "caja", y dos brazos, los "bigotes". Es un gráfico representativo de las distribuciones de un conjunto de datos en cuya construcción se usan cinco medidas descriptivas de los mismos, a saber: (Me) mediana, (Q1) primer cuartil, (Q3) tercer cuartil, valor máximo y valor mínimo. Esta presentación visual, asocia las cinco medidas que suelen trabajarse de forma individual. Presenta al mismo tiempo, información sobre la tendencia central, dispersión y simetría de los datos de estudio. Además, permite identificar con claridad y de forma individual, observaciones que se alejan de manera poco usual del resto de los datos. A estas observaciones se les conoce como valores atípicos.

### Construcción:

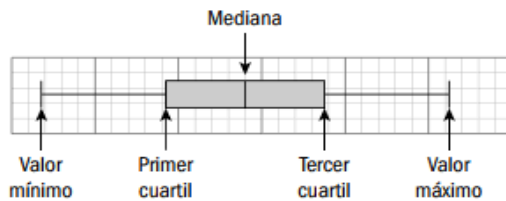
Una gráfica de este tipo consiste en una **caja** rectangular, donde los lados más largos muestran el **recorrido intercuartílico**. Este rectángulo está dividido por un segmento vertical que indica donde se posiciona la mediana y por lo tanto su relación con los cuartiles primero y tercero (recordemos que el segundo cuartil coincide con la mediana).

Esta caja se ubica a escala sobre un segmento que tiene como extremos los valores mínimo y máximo de la variable. Las líneas que sobresalen de la caja se llaman **bigotes**. Estos bigotes tienen un límite de prolongación, de modo que cualquier dato o caso que no se encuentre dentro de este rango es marcado e identificado individualmente

Es un gráfico que suministra información sobre los valores mínimo y máximo, los cuartiles Q1, Q2 o mediana y Q3, y sobre la existencia de valores atípicos y la simetría de la distribución.

- 1.-Límite superior: Es el extremo superior del bigote. Los valores por encima de este límite se consideran atípicas.
- 2.-Tercer cuartil (Q3): Por debajo de este valor se encuentran como máximo el 75% de los datos.
- 3.-Mediana: Coincide con el segundo cuartil. Divide a la distribución en dos partes iguales. De este modo, 50% de las observaciones están por debajo de la mediana y 50% está por encima.
- 4.-Primer cuartil (Q1): Por debajo de este valor se encuentra como máximo el 25% de los datos.
- 5.-Límite inferior: Es el extremo inferior del bigote. Las opiniones por debajo de este valor se consideran atípicas.
- 6.-Valores atípicos: valores que están apartados del cuerpo principal de datos. Pueden representar efectos de causas extrañas, opiniones extremas o en el caso de la tabulación manual, errores de medición o registro.

## ELEMENTOS DE DIAGRAMA DE CAJA Y BIGOTES



## Representación gráfica de los cuartiles. Diagramas de Caja

**EJEMPLO DISTRIBUCIÓN DE EDADES:** con la siguiente información, que representa las edades de un grupo de personas: 36, 25, 37, 24, 39, 20, 36, 45, 31, 31, 39, 24, 29, 23, 41, 40, 33, 24, 34, 40

**1 paso ordenar los datos:** 20 23 24 24 24 25 29 31 31 33 34 36 36 37 39 39 40 40 41 45

## CALCULO DE CUARTILES

**Q1**, el cuartil Primero es el valor mayor que el 25% de los valores de la distribución. Como  $N = 20$  resulta que  $N/4 = 5$ ; el primer cuartil es la media aritmética de dicho valor y el siguiente:

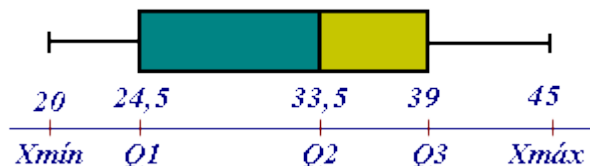
$$Q1 = (24 + 25) / 2 = 24,5$$

**Q2**, el Segundo Cuartil es, evidentemente, la mediana de la distribución, es el valor de la variable que ocupa el lugar central en un conjunto de datos ordenados. Como  $N/2 = 10$ ; la mediana es la media aritmética de dicho valor y el siguiente:  $me = Q2 = (33 + 34) / 2 = 33,5$

**Q3**, el Tercer Cuartil, es el valor que sobrepasa al 75% de los valores de la distribución. En nuestro caso, como  $3N / 4 = 15$ , resulta;  $Q3 = (39 + 39) / 2 = 39$

RIC: Rango intercuartílico:  $Q3 - Q1 = 39 - 24,5 = 14,5$

## DIBUJAR LA CAJA Y LOS BIGOTES



El *bigote* de la izquierda representa al colectivo de edades (**Xmín, Q1**)

La primera parte de la caja a (**Q1, Q2**),

La segunda parte de la caja a (**Q2, Q3**)

El *bigote* de la derecha viene dado por (**Q3, Xmáx**).

hay que calcular los límites superior e inferior, Ls y Lm, que identifiquen a los valores atípicos. Para ello se calcula cuándo se consideran atípicos los valores. Son aquellos inferiores a  $Q1 - 1.5 \times RIC$  o superiores a  $Q3 + 1.5 \times RIC$ .

En el ejemplo:

- inferior:  $24,5 - (1.5 \times 14,5) = 2,25$  • superior:  $39 + (1.5 \times 14,5) = 60,75$  Ahora se buscan los últimos valores que NO son atípicos, que serán los extremos de los bigotes.
- En el ejemplo: 20 y 45
- Marcar como atípicos todos los datos que están fuera del intervalo (X min, X max). En el ejemplo: 2,25 y 60,75

**Conclusiones:** Podemos obtener abundante información de una distribución a partir de estas Representaciones. Veamos alguna:

La parte izquierda de la caja es mayor que la de la derecha; ello quiere decir que las edades comprendidas entre el 25% y el 50% de la población está más dispersa que entre el 50% y el 75%.

- El bigote de la izquierda (Xmín, Q1) es más corto que el de la derecha; por ello el 25% de los más jóvenes están más concentrados que el 25% de los mayores.
- El rango intercuartílico =  $Q3 - Q1 = 14,5$ ; es decir, el 50% de la población está comprendido en 14,5 años.

**Veamos otro ejemplo:** Se ha preguntado a varias personas aficionadas al fútbol, el número de balones de fútbol que hay en su casa. Los datos obtenidos han sido los siguientes. 0 0 0 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 5, los cuales ya están ordenados.

Procedemos a identificar; El valor mínimo, el valor máximo, y los cuartiles Q1, Q2 y Q3 y el rango intercuartílico (RIC).

- Valor mínimo = 0
- Valor máximo = 5

Mediana: (50%) como el número de datos es par ( $N = 38$ ) dividimos  $\frac{n+1}{2}$

Mediana =  $\frac{38+1}{2} = \frac{39}{2} = 19,5$  corresponde entre la posición 19 y 20, en este conjunto de datos es

Me = 2. **0 0 0 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 5**

**Hallaremos el Q1 (25%):** 0 0 0 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2, nuevamente hallaremos la mediana de la primera mitad.

$Q1 = \frac{19+1}{2} = 10$ , (es la posición que corresponde al dato **0 0 0 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2**)

**Hallaremos el Q3 (75%):** 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 5, la mediana de la segunda mitad de los datos

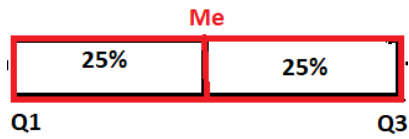
$Q3 = \frac{19+1}{2} = 10$ , que es la posición que corresponde al dato **2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 5**

El rango intercuartílico:  $Q3 - Q1 = 3 - 1 = 2$

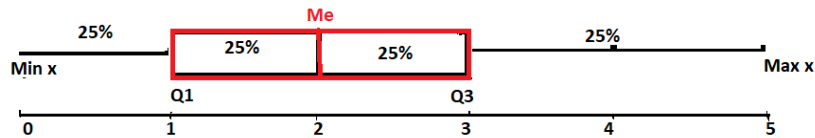
El rango  $V_{max} - V_{min} = 5 - 0 = 5$



Procederemos a construir el diagrama: Dibujar un rectángulo con Q1 y Q3 como extremos e indicar la posición de la mediana (Q2) mediante una línea



• Para dibujar los bigotes, las líneas que se extienden desde la caja, hay que calcular los límites superior e inferior, LI y LS, que identifiquen a los valores atípicos.



Para ello se calcula cuándo se consideran atípicos los valores. Para este ejemplo no se considerarán.

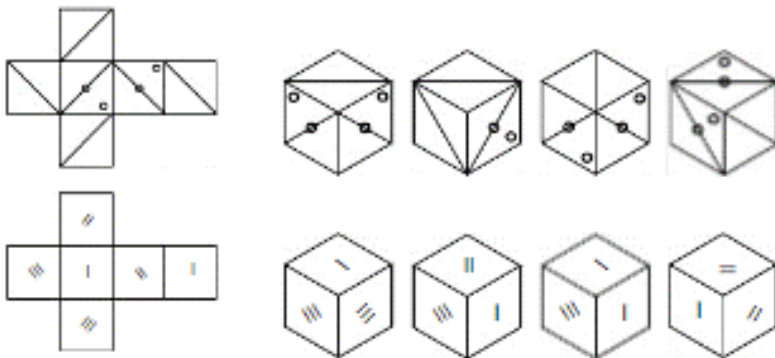
Escribir las conclusiones:

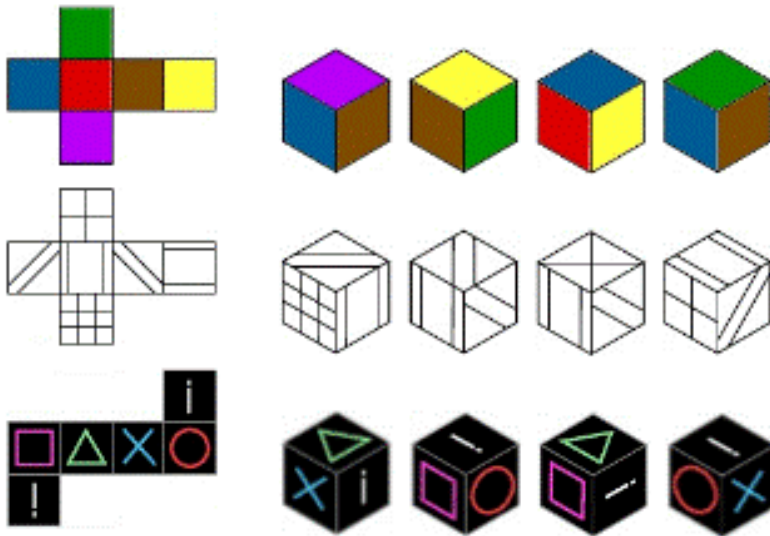
1. \_\_\_\_\_
2. \_\_\_\_\_

Observa este otro ejemplo: <https://youtu.be/GBNpyyApgdA>

**Dando cumplimiento al proyecto de competencia ciudadana se propone la siguiente actividad.**

En las siguientes figuras identifique el sólido que se puede armar, constrúyalos en cartulina y compruebe la respuesta dada.





Tomado de; [es.scribd.com/doc/65446275/Ejercicios-Análisis-imágenes](https://es.scribd.com/doc/65446275/Ejercicios-Análisis-imágenes)

**PRACTICO LO QUE APRENDI:**

**En el cuaderno de estadística desarrollar con el acompañamiento de tu docente la siguiente actividad**

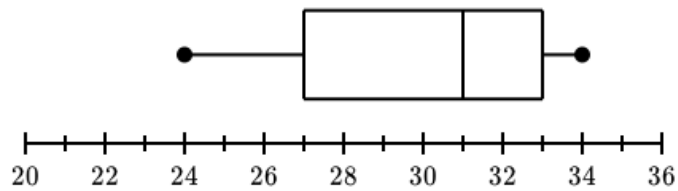
1. Encuentra el P30, P50 y P70 de la siguiente tabla de frecuencias que muestra el tiempo de servicio de los trabajadores de una empresa

Años de servicio	fi
1 – 8	9
8 – 15	7
15 – 22	12
22 – 29	16
29 – 36	8
36 – 93	3

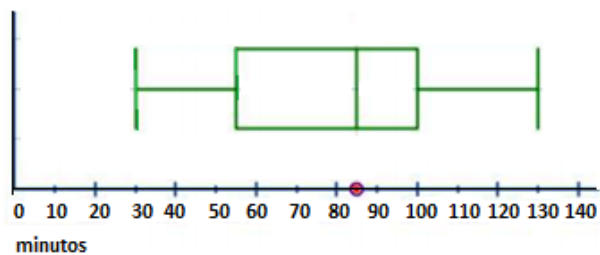
2. Calcula el Q1, Q2 y Q3 en la siguiente distribución de frecuencias que corresponde a los puntajes obtenidos por 60 estudiantes.

Puntajes	f
43 – 51	9
52 – 60	7
61 -69	10
70 – 78	14
79 – 87	12
88 – 96	8
Total	60

3. ¿Cuál es el conjunto de datos que puede representarse por los diagramas de caja que se muestran abajo?



- a. 24, 25, 29, 31, 31, 31, 32, 34, 35
  - b. 24, 25, 29, 30, 31, 31, 32, 34, 34
  - c. 24, 27, 29, 30, 30, 31, 32, 34, 34
  - d. 24, 25, 29, 30, 30, 31, 34, 34, 34
4. Analiza el siguiente diagrama de caja y bigotes que muestra los minutos que ha tardado en hacer efecto un medicamento en una población. Interpreta la información que presenta y responde a las preguntas:



- a. ¿A qué porcentaje de la población había hecho efecto al cabo de 30 minutos?
- b. ¿Al cabo de cuántos minutos había hecho efecto al 50 % de la población?
- c. ¿Cuántos minutos tardó en hacer efecto al 100% de la población?
- d. ¿A qué porcentaje había hecho efecto a los 55 minutos?
- e. ¿Cuánto tardó en hacer efecto a las tres cuartas partes de la población?

Ejercicios tomados de <https://app.box.com/s/wigci72lb77009x47fs4szycr7z2fbcw>

### ¿CÓMO SÉ QUE APRENDÍ?

**ACTIVIDAD PARA ENTREGAR:** se realiza en el cuaderno de estadística y se entrega en la fecha estimada por la docente de la asignatura

1. Calcula los D3, D6 y D8 de la siguiente distribución de frecuencias, edades de 80 estudiantes de un colegio.

EDADES	6 – 8	9 – 11	12 – 14	15 – 17	18 – 20	Total
f	19	15	17	20	9	80

## CONTESTA LAS PREGUNTAS 2 Y 3 CON LA SIGUIENTE INFORMACIÓN

La distribución dada en la tabla de la derecha presenta las estaturas, en centímetros, de 40 estudiantes de un colegio, ahora responde

Estaturas	f
140 – 143	7
144 – 147	5
148 – 151	3
152 – 155	8
156 – 159	4
160 – 163	10
164 - 167	3

2. La estatura que corresponde al Q1 es:
  - a) 145,5
  - b) 147,5
  - c) 144,1
  - d) 145,9
3. De la anterior distribución de frecuencias se puede decir que el D8 es:
  - a) 163,5
  - b) 161,5
  - c) 161,9
  - d) 162,3

Ejercicios tomados de la página

<https://app.box.com/s/wigci72lb77009x47fs4szycr7z2fbcw>

**¿QUÉ APRENDÍ?** Estas preguntas te servirán de auto evaluación. Responde en tu cuaderno

1. ¿Aprendiste a calcular los cuantiles dados en una tabla de frecuencias?
2. ¿aprendiste a sacar las conclusiones de acuerdo a los cuantiles calculados?
3. ¿aprendiste a construir el diagrama de caja y a interpretar la información dada en el?
4. ¿Te quedó claro las equivalencias entre algunas medidas de posición?
5. De lo visto en la guía que fue lo que más se te dificultó. ¿por qué? ¿Cómo crees que puedes mejorar en tu proceso académico?

QUE SIGNIFICA PARA TI LA SIGUIENTE FRASE:

"LA EDUCACIÓN ES EL PASAPORTE HACIA EL FUTURO, EL MAÑANA PERTENECE A AQUELLOS QUE SE PREPARAN PARA ÉL EN EL DÍA DE HOY"

MALCOM X