CBS

Colegio Bautista Shalom

Estadística II
Quinto PCOC
Primer Bimestre

Contenidos

MEDIDAS DE VARIABILIDAD O DISPERSIÓN

- ✓ RANGO, AMPLITUD TOTAL O RECORRIDO.
- ✓ RANGO INTERCUARTÍLICO.
- ✓ CONSTRUCCIÓN DEL DIAGRAMA DE CAJA.
- ✓ RANGO PERCENTIL.
- ✓ RANGO PERCENTILAR (10 – 90).
- ✓ DESVIACIÓN CUARTÍLICA.
- ✓ DESVIACIÓN MEDIA.
- ✓ VARIANZA.
- ✓ DESVIACIÓN TÍPICA O ESTÁNDAR.
- ✓ PROPIEDADES DE LA VARIANZA Y DESVIACIÓN TÍPICA.

NOTA: conforme vayas avanzando en el aprendizaje de cada uno de los temas desarrollados, encontrarás ejercicios a resolver. Sigue las instrucciones de tu catedrático(a).

MEDIDAS DE VARIABILIDAD O DISPERSIÓN

Los estadísticos de tendencia central o de posición nos indican en qué lugar se sitúa un grupo de puntuaciones. Los de variabilidad o dispersión nos indican si esas puntuaciones o valores están próximas entre sí o si por el contrario están o muy dispersas.

RANGO, AMPLITUD TOTAL O RECORRIDO

El Rango es una medida razonable de la variabilidad podría ser la amplitud o rango, que se obtiene restando el valor más bajo de un conjunto de observaciones del valor más alto.

Propiedades del rango:

- ✓ Es fácil de calcular y sus unidades son las mismas que las de la variable.
- ✓ No utiliza todas las observaciones (sólo dos de ellas);
- ✓ Se puede ver muy afectada por alguna observación extrema;
- ✓ El rango aumenta con el número de observaciones, o bien se queda igual. En cualquier caso, nunca disminuye.

El rango se suele definir como la diferencia entre los dos valores extremos que toma la variable. Es la medida de dispersión más sencilla y también, por tanto, la que proporciona menos información. Además, esta información puede ser errónea, pues el hecho de que no influyan más de dos valores del total de la serie puede provocar una deformación de la realidad.

Comparemos, por ejemplo, estas dos series:

Serie 1: 1 5 7 7 8 9 9 10 17
Serie 2: 2 4 6 8 10 12 14 16 18

Ambas series tienen rango 16, pero están desigualmente agrupadas, pues mientras la primera tiene una mayor concentración en el centro, la segunda se distribuye uniformemente a lo largo de todo el recorrido.

El uso de esta medida de dispersión, será pues, bastante restringido. En la estadística descriptiva se denomina rango estadístico (R) o recorrido estadístico al intervalo entre el valor máximo y el valor mínimo; por ello, comparte unidades con los datos. Permite obtener una idea de la dispersión de los datos, cuanto mayor es el rango, más dispersos están los datos de un conjunto. Por ejemplo, para una serie de datos de carácter cuantitativo, como lo es la estatura medida en centímetros, tendríamos:

Tenemos…

$$x_1 = 185, x_2 = 165, x_3 = 170, x_4 = 182, x_5 = 155$$

es posible ordenar los datos como sigue:

$$x_{(1)} = 155, x_{(2)} = 165, x_{(3)} = 170, x_{(4)} = 182, x_{(5)} = 185$$

donde la notación $x_{(i)}$ indica que se trata del elemento *i*-ésimo de la serie de datos. De este modo, el rango sería la diferencia entre el valor máximo (k) y el mínimo; o, lo que es lo mismo:

$$R = x_{(k)} - x_{(1)}$$

En nuestro ejemplo, con cinco valores, nos da que R = 185-155 = 30.

RANGO INTERCUARTÍLICO

El rango intercuartílico IQR (o rango intercuartil) es una estimación estadística de la dispersión de una distribución de datos. Consiste en la diferencia entre el tercer y el primer cuartil. Mediante esta medida se eliminan los valores extremadamente alejados.

El rango intercuartílico es altamente recomendable cuando la medida de tendencia central utilizada es la mediana (ya que este estadístico es insensible a posibles irregularidades en los extremos).

$$IQR = Q_3 - Q_1$$

Con el IQR podremos elaborar los diagramas de caja Diagramas de Caja, que es un instrumento muy visual para evaluar la dispersión de una distribución.

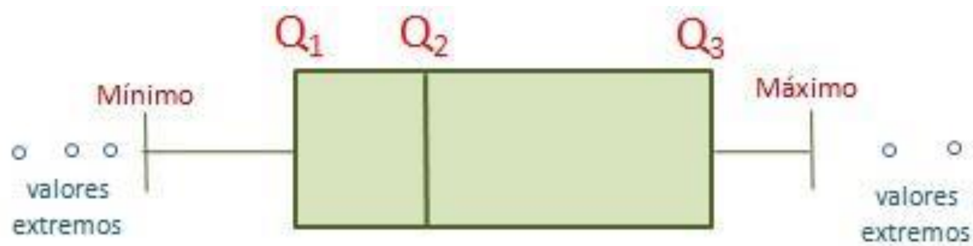
El Diagrama de Caja es un gráfico utilizado para representar una variable cuantitativa (variable numérica). El gráfico nos sirve como una herramienta; la cual, nos permite visualizar por medio de los cuartiles: cómo es la distribución, su grado de asimetría, los valores extremos, la posición de la mediana, y más.

El *Diagrama de Caja* se compone de:

1. Un rectángulo (*caja*) delimitado por el **primer y tercer cuartil** (Q_1 y Q_3). Dentro de la caja una línea indica dónde se encuentra la mediana (segundo cuartil Q_2).

2. Dos *brazos*, uno que empieza en el primer cuartil y acaba en el **mínimo**, y otro que empieza en el tercer cuartil y acaba en el **máximo**.

3. Los **datos atípicos** (o valores extremos) que son los valores distintos que no cumplen ciertos requisitos de heterogeneidad de los datos.



Los diagramas de caja son muy útiles para comprar una variable en diferentes grupos.

CONSTRUCCIÓN DEL DIAGRAMA DE CAJA

Para construirlo, debemos seguir los siguientes pasos:

1. Ordenar los datos.

2. Calcular los tres cuartiles (Q_1, Q_2 y Q_3). Después, dibujamos el rectángulo (*caja*) delimitado por el primer y tercer cuartil, dibujando entre los dos cuartiles una línea para indicar dónde está la mediana (segundo cuartil).

3. Calcular el rango intercuartílico, que es el tercer cuartil menos el primero.

$$IQR = Q_3 - Q_1$$

4. Se calculan los **límites** admisibles inferior y superior (*LI* y *LS*) para identificar los valores extremos.

$$\begin{aligned} LI = &\ Q_1 - 1,5 \cdot IQR \\ LS = &\ Q_3 + 1,5 \cdot IQR \end{aligned}$$

   Los límites marcarán los datos atípicos de la variable. Todos aquellos puntos que sean menores que *LI* ($x < LI$) o mayores que *LS* ($x > LS$) son **valores extremos**. Es decir, son todos aquellos valores que no están en el intervalo [*LI,LS*].

5. El **mínimo** es el menor valor del conjunto que sea mayor o igual que *LI*. El **máximo** es el mayor valor del conjunto que es menor o igual que *LS*.

   Dibujamos los dos ***brazos***. El primero va desde el primer cuartil hasta el mínimo. El segundo, desde el tercer cuartil hasta el máximo.

6. Se dibujan los **valores extremos**, representados por puntos o círculos pequeños.

Por ejemplo. En un bosque plantaron veinte (*N*=20) árboles y, al cabo de unos años, se mide la altura para ver su evolución. Un muy buen método para ver cómo han crecido y comprobar si existen valores extremos es el diagrama de caja**.**

Mediante esta representación gráfica podemos ver si hay árboles que han crecido más o menos de lo habitual.

Altura de los árboles (metros)									
7,13	5,35	10,14	4,15	5,85	4,69	6,84	0,94	3,45	5,79
5,64	4,92	5,32	6,56	6,17	4,33	6,18	6,50	3,74	2,98

Se ordenan los datos y se calculan los tres cuartiles.

Altura de los árboles (metros) - Ordenada									
0,94	2,98	3,45	3,74	4,15	4,33	4,69	4,92	5,32	5,35 Q_2
Q_2 5,64	5,79	5,85	6,17	6,18 Q_1	6,5	6,56	6,84	7,13	10,14

Q_3

A partir del conjunto ordenado calculamos los cuartiles:

$$Q_1 = X_{(N+1)/4} = X_{5,25} = X_5 + 0,25 \cdot (X_6 - X_5)$$

$$Q_1 = 4,15 + 0,25 \cdot (4,33 - 4,15) = 4,20$$

$$Q_2 = Mediana(X) = X_{(N+1)/2} = X_{10,5}$$

$$Q_2 = \frac{X_{10} + X_{11}}{2} = \frac{5,35 + 5,64}{2} = 5,50$$

$$Q_3 = X_{3(N+1)/4} = X_{15,75} = X_{15} + 0,75 \cdot (X_{16} - X_{15})$$

$$Q_3 = 6,18 + 0,75 \cdot (6,50 - 6,18) = 6,42$$

Los tres cuartiles son: Q_1 = 4,20, Q_2 = 5,50 y Q_3 = 6,42.

Se calculan los límites admisibles inferior y superior (*LI* y *LS*) para determinar los valores extremos.

El rango intercuartílico es:

$$IQR = Q_3 - Q_1 = 6,42 - 4,20 = 2,22$$

A partir del rango calculamos los límites:

$$LI = Q_1 - 1,5 \cdot IQR = 4,20 - 1,5 \cdot 2,22 = 0,96$$

$$LS = Q_3 + 1,5 \cdot IQR = 6,42 + 1,5 \cdot 2,22 = 9,59$$

Los valores extremos serán todos los árboles que midan menos de 0,96m o más de 9,59m. Tenemos dos árboles, uno de 0,94m y otro de 10,14m que serán valores extremos.

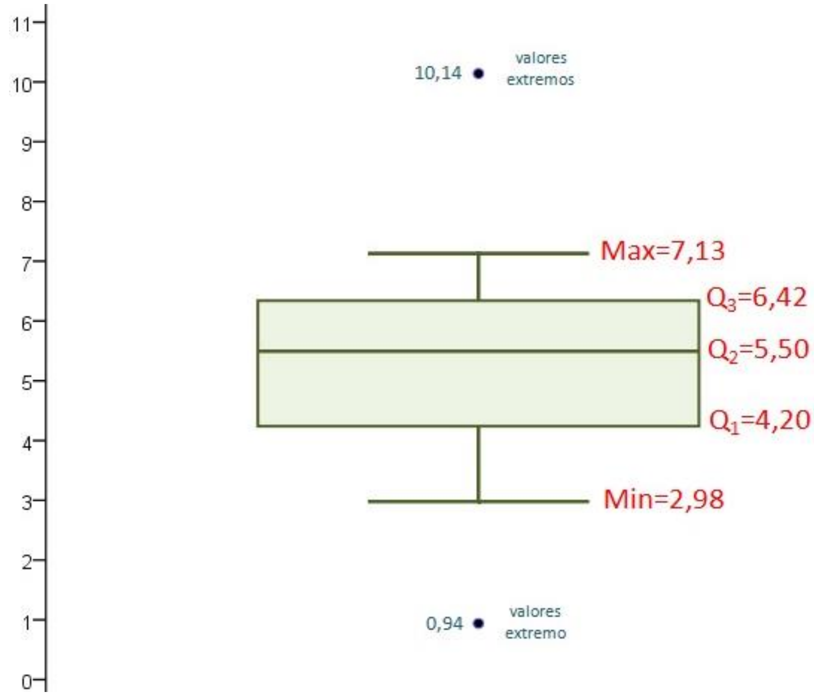
Estos valores los representamos con puntos en el diagrama de caja.

1. El *mínimo* es el menor elemento del conjunto que sea mayor o igual al límite inferior. El *máximo* es el mayor elemento que sea menor o igual al límite superior. En este caso, el mínimo es 2,98 y el máximo 7,13.

2. Se dibujan los *brazos* del diagrama de caja. El brazo inferior irá desde el primer cuartil hasta el mínimo (desde el 4,20 a 2,98). El brazo superior abarcará desde el tercer cuartil hasta el máximo (desde el 6,42 hasta el 7,13).

3. Los dos puntos extremos se representan mediante un punto o círculo.

El diagrama de caja del conjunto de la altura de estos veinte árboles es:

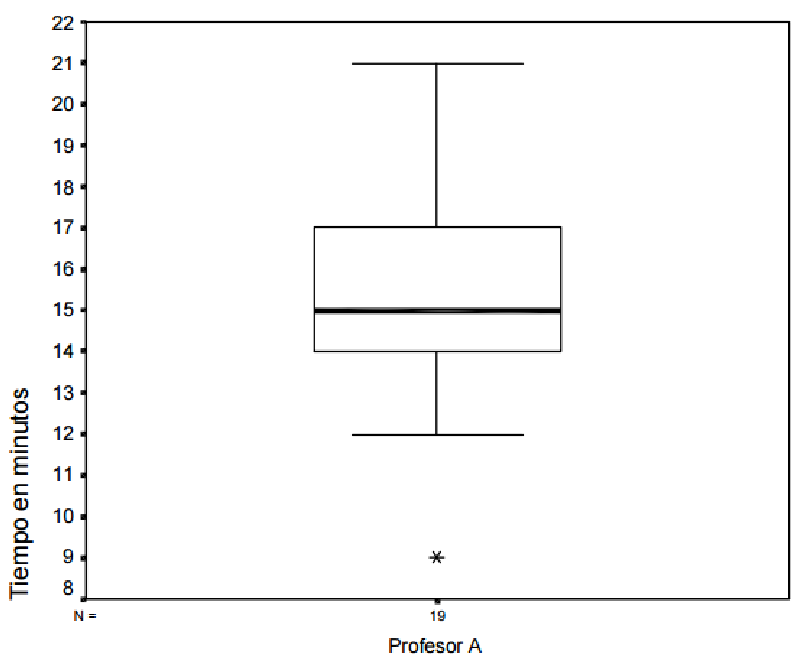


Esta representación proporciona una visión rápida de la distribución, apreciándose una asimetría al no estar Q_2 en el centro, en este caso porque hay árboles más altos que la mediana cuya altura está más separada de la mediana que los que tienen una altura inferior a ella, que están más agrupados. También se puede apreciar la existencia de valores extremos.

EJERCICIO 01: realiza lo que se te indique en los incisos.

Dos profesores (A y B) están interesados en estudiar los hábitos de sueño de los estudiantes en sus clases. Ambos profesores registran el tiempo (en minutos) que demoran en quedarse dormidos sus alumnos desde que empieza la clase.

El siguiente diagrama de caja muestra los tiempos que demoran en quedarse dormidos los alumnos del profesor A.



Resuelve los siguientes incisos:

a) ¿Cuál es el valor aproximado de las medidas de dispersión del tiempo del Profesor A?
b) ¿Qué porcentaje de alumnos se queda dormido antes de los 14 minutos con el Profesor A?
c) Los datos del Profesor B son los siguientes:

10,5	**11,3**	**11,9**	**12**	**12,3**	**12,3**	**12,5**	**12,7**	**13,4**	**13,7**
13,8	**14,2**	**14,8**	**15,1**	**15,3**	**16,7**	**16,8**	**18,8**	**20,8**	

Construye un diagrama de caja correspondiente a los tiempos en que se quedan dormidos los alumnos en la clase del Profesor B.

Utiliza los siguientes espacios para el desarrollo de tus respuestas:

Inciso a)

Inciso b)

Inciso c)

Para dibujar el gráfico de caja necesitamos calcular los cuartiles, y verificar si existen valores extremos:

1°	2°	3°	4°	**5°**	6°	7°	8°	9°	**10°**
10,5	11,3	11,9	12	**12,3**	12,3	12,5	12,7	13,4	**13,7**
				Q_1					**Q_2**
11°	12°	13°	14°	**15°**	16°	17°	18°	19°	
13,8	14,2	14,8	15,1	**15,3**	16,7	16,8	18,8	20,8	
				Q_3					

Cálculo de la mediana:

Cálculo del Cuartil 1:

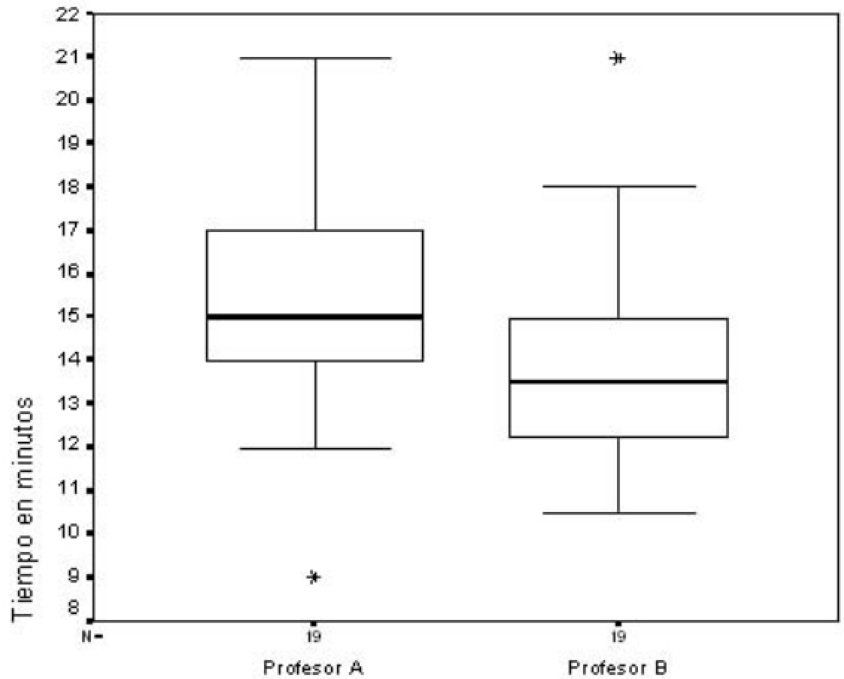
Cálculo del Cuartil 3:

Cálculo de valores extremos:

Se ha verificado si el máximo es un valor extremo si 20,8 es mayor que:

Se ha verificado si el siguiente número es valor extremo:

Con estos elementos, finalmente se construye el diagrama de caja:



EJERCICIO 02: a continuación, se te presenta una serie de datos. Debes ordenarlos, calcular los cuartiles correspondientes, realizar el gráfico de caja, realiza una breve explicación del gráfico de caja y determina el rango intercuartílico.

36 25 37 24 39 20 36 45 31 31
39 24 29 23 41 40 33 24 34 40

Utiliza los siguientes espacios para el desarrollo de tus respuestas:

Ordena los datos

20 23 24 24 24 25 29 31 31 33 34 36 36 37 39 39 40 40 41 45

Calcula los cuartiles

$Q_1 =$

$Q_2 =$

$Q_3 =$

Realiza el diagrama de caja

Explica el diagrama:

Determina el Rango Intercuartílico

El *rango intercuartílico* =

EJERCICIO 03: tu catedrático(a) te proporcionará 3 ejercicios más que debes desarrollar en hojas aparte. Debes desarrollar cada uno de los problemas y entregarlos con su respectiva identificación, dentro de folder con gancho, según te haya indicado.

RANGO PERCENTIL

El rango de percentil compara un dato específico de una variable con el conjunto de datos de donde se tomó. El resultado significa el porcentaje de casos que recibió puntuaciones más bajas que quien tuvo el dato del que se calcula el Rango Percentil.

El rango percentil de un dato depende del grupo de donde se haya tomado.

Para calcular el Rango Percentil de un dato debemos emplear la siguiente formula:

$$RP_{¿?} = \%aa + \left[\frac{(¿\,? - \text{Liv})}{\text{i}} \cdot (\%)\right]$$

En donde:

$¿\,? = \text{Dato del que se calculará el Rango Percentil.}$
$\%aa = \text{Porcentaje acumulado anterior a la clase donde se localiza el valor } ¿\,?.$
$Liv = \text{Límite Inferior Verdadero.}$
$\% = \text{porcentaje de la clase donde se localiza el } ¿\,?.$
$i = \text{Tamaño del intervalo.}$

Observa el siguiente ejemplo:

Clases	x	f	fa	%	%aa
90-99	94.5	6	49	12.24	100.00
80-89	84.5	8	43	16.33	87.76
70-79	74.5	12	35	24.49	71.43
60-69	64.5	10	23	20.41	46.94
50-59	54.5	7	13	14.29	26.53
40-49	44.5	6	6	12.24	12.24
N=49					

Ahora, calculemos el Rango Percentil para una calificación de 92.

Deducimos:

- ✓ El límite inferior del intervalo crítico es: 89.5
- ✓ El tamaño del intervalo crítico es: 10

Encontramos el porcentaje dentro del intervalo crítico:

$$\% = (100)\left(\frac{f}{N}\right) \qquad \% = (100)\left(\frac{6}{49}\right) \qquad \% = (100)0.1224 \qquad \% = 12.24$$

El porcentaje acumulado asociado al intervalo de clase anterior 80 – 89 es: 87.76.

Calculamos el Rango Percentil para una calificación de 92.

Datos:

¿? = 92
%aa = 87.76
Liv = 89.5
% = 12.24
i= 10

Sustitución de los datos en la fórmula:

$$RP_{92} = 87.76 + \left[\frac{92 - 89.5}{10} \cdot (12.24)\right]$$

$$RP_{92} = 87.76 + 3.06$$

$$RP_{92} = 90.82\%$$

Interpretación 1: Una persona cuya calificación fue de 92, supera al 90.82% del grupo en su calificación de matemáticas.

Interpretación 2: el 90.82% de las calificaciones es menor que un 92, sólo el 9.18% de las calificaciones es superior a 92.

RANGO PERCENTILAR (10 – 90)

El rango percentilar 10 -90 de un conjunto de datos se define como:

$$Rango\ Percentilar\ 10 - 90{:}\ P_{90} - P_{10}$$

Donde $P_{90} - P_{10}$ son los percentiles 10º y 90º de los datos.

Por ejemplo:

NOTAS DEL EXÁMEN BIMESTRAL DE ESTADÍSTICA									
5	95	35	85	75	15	25	15	55	90
10	85	20	90	60	5	35	45	70	35
25	25	5	95	55	20	40	35	80	15
10	65	50	45	75	40	50	40	55	35
25	45	15	10	5	70	5	30	40	90

Clases	x	f	fa	%	%a	f ▪ x	x^2	f ▪ x^2
5 – 22	13.5	14	14	28	28	189	182.25	2551.5
23 – 40	31.5	14	28	28	56	441	992.25	13891.5
41 – 58	49.5	8	36	16	72	396	2450.25	19602
59 – 76	67.5	6	42	12	84	405	4556.25	27337.5
77 – 94	85.5	6	48	12	96	513	7310.25	43861.5
95 – 112	103.5	2	50	4	100	207	10712.25	21424.5
		Σ=50		100		2151		

$$RP_{10} = 72 + \left[\frac{(10 - 4.5)}{18} \cdot (28)\right] = \quad 72 + \frac{5.5}{18} \cdot (28) = \quad 72 + 0.3055 \ldots (28) = 72 + 8.555 \ldots = 80.60$$

$$RP_{90} = 84 \ + \left[\frac{(90 - 77.5)}{18} \cdot (12)\right] = \quad 84 + \frac{12.5}{18} \cdot (28) = 84 + 0.69.44 \ldots = 84.70$$

$$Rango\ Percentilar\ 10 - 90 = \ 84.69444 - \ 80.5555 \ = 4.1\%$$

EJERCICIO 04: con los datos de la siguiente tabla:

Clases	X	F	Fa	%	%aa
90 – 99	94.5	6	49	12.24	100.00
80 – 89	84.5	8	43	16.33	87.76
70 – 79	74.5	12	35	24.49	71.43
60 – 69	64.5	10	23	20.41	46.94
50 – 59	54.5	7	13	14.29	26.53
40 – 49	44.5	6	6	12.24	12.24
N = 49					

Calcula:

a. El rango percentilar para una calificación de 92.
b. Realiza la interpretación correspondiente.

Datos:

¿? = ________

%aa = ________

Liv = ________

% = ________

i = _______

Sustitución de valores:

$$RP_{92} = \quad + \left[\frac{(\quad - \quad)}{\quad} \cdot (\quad)\right] = \quad + \quad = \quad \%$$

Expone tu interpretación en clase.

DESVIACIÓN CUARTÍLICA

Se calcula dividiendo entre dos la diferencia del tercer menos el primer cuartil.

$$DQ = \frac{Q_3 - Q_1}{2} =$$

Para ejemplificarte su cálculo, utilizaremos los datos de la siguiente tabla de distribución de frecuencias:

Clases		Xi	fi	Fi	ni	Ni
5	**18**	12	5	5	0.10	0.10
18	**31**	25	6	11	0.12	0.22
31	**44**	38	6	17	0.12	0.34
44	**57**	51	23	40	0.46	0.80
57	**70**	64	3	43	0.06	0.86
70	**83**	77	3	46	0.06	0.92
83	**96**	90	4	50	0.08	1
			n = 50		1	

Calculamos el primer cuartil…

El primer cuartil (= Q1) se calcula determinando primero su clase, para ello se busca el valor 1/4 × n, en nuestro ejemplo tenemos que: 1/4 × 50 = 12.5 = 13 y este resultado se busca por la frecuencia absoluta acumulada y vemos que está comprendido en la tercera clase: **31 – 44.**

Una vez determinada la clase que corresponde al primer cuartíl, aplicamos la fórmula:

$$Q_1 = L_i + \frac{1/4 * n - F_{i-1}}{f_i} * Ac = 31 + \frac{0.25*50 - 11}{6} * 13$$

$$= 34.25$$

Calculamos el tercer cuartil…

De manera parecida calculamos el tercer cuartil (= Q3). Determinamos su clase buscando el valor 3/4 × n = 0.75 × 50 = 38 y este resultado se busca en la columna de la frecuencia absoluta acumulada y vemos que se localiza en la cuarta clase: **44 – 57.**

$$Q_3 = L_i + \frac{3/4 * n - F_{i-1}}{f_i} * Ac = 44 + \frac{0.75*50 - 17}{23} * 13$$

$$Q_3 = 55.58$$

Ahora, podremos calcular la *Desviación Cuartílica (DQ)…*

Se calcula dividiendo entre dos la diferencia del tercer menos el primer cuartil.

$$DQ = \frac{Q_3 - Q_1}{2} = \frac{55.58 - 34.25}{2} = \frac{21.33}{2} = 10.6 = 11$$

EJERCICIO 05: se te presenta una tabla de distribución de frecuencias:

Clases	Xi	fi	Fi	ni	Ni	fi * Xi	(Xi - 47.42)2	fi * (Xi - 47.42)2
5 – 18	12	5	5	0.10	0.10	58	1255	6275
18 – 31	25	6	11	0.12	0.22	147	503	3018
31 – 44	38	6	17	0.12	0.34	225	89	534
44 – 57	51	23	40	0.46	0.80	1162	13	299
57 – 70	64	3	43	0.06	0.86	191	275	825
70 – 83	77	3	46	0.06	0.92	230	875	2625
83 – 96	90	4	50	0.08	1	358	1813	7252
		n = 50		1		2371		20828

En donde debes calcular lo siguiente:

- **a.** El primer cuartil.
- **b.** El tercer cuartil.
- **c.** La desviación cuartílica.

Una vez determinada la clase que corresponde al primer cuartíl, aplicamos la fórmula:

$$Q_1 = L_i + \frac{1/4 * n - F_{i-1}}{f_i} * Ac =$$

Una vez determinada la clase que corresponde al tercer cuartíl, aplicamos la fórmula:

$$Q_3 = L_i + \frac{3/4 * n - F_{i-1}}{f_i} * Ac =$$

Ahora, calculemos la desviación cuartílica. Se calcula dividiendo entre dos la diferencia del tercer menos el primer cuartil.

$$DQ = \frac{Q_3 - Q_1}{2} =$$

Expone tu interpretación en clase.

DESVIACIÓN MEDIA

La desviación puede referirse a cada una de las medidas de tendencia central: media, mediana o moda; pero el interés se suele centrar en la medida de la desviación con respecto a la media, que llamaremos desviación media. Puede definirse como la media aritmética de las desviaciones de cada uno de los valores con respecto a la media aritmética de la distribución, y de indica así:

$$\boldsymbol{DM} \ \ equivale\ a \ \ \boldsymbol{D_{\bar{x}}} \ \ en\ las\ fórmulas$$

Empleamos la siguiente fórmula para encontrar la *Desviación Media*…

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \ldots + |x_n - \bar{x}|}{N}$$

Abreviando dicha fórmula, nos queda…

$$\boxed{DM = \frac{\sum |x - \bar{x}|}{N}}$$

A continuación, tenemos la siguiente distribución:

9, 3, 8, 8, 9, 8, 9, 18

Ahora, procedemos a encontrar la $\bar{X}$...

$$\bar{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

Con este dato, ya podremos encontrar la *DM o* $D_{\bar{x}}$ ...

$$D_{\bar{x}} = \frac{|9-9| + |3-9| + |8-9| + |8-9| + |9-9| + |8-9| + |9-9| + |18-9|}{8} = 2.25$$

Observa que se toman las desviaciones en valor absoluto, es decir, que la fórmula no distingue si la diferencia de cada valor de la variable con la media es en más o en menos. Ya se habrá advertido que esta expresión sirve para calcular la desviación media en el caso de datos sin agrupar.

Veamos otro ejemplo:

Se tiene los valores:

2, 2, 4, 4, 4, 5, 6, 7, 8, 8.

Averiguar la desviación media de estos valores.

Encontrando a $\bar{X}$...

$$\bar{X} = \frac{2+2+4+4+4+5+6+7+8+8}{10} = \frac{50}{10} = 5$$

x	$x - \bar{x}$	$\lvert x \rvert$
2	-3	3
2	-3	3
4	-1	1
4	-1	1
4	-1	1
5	0	0
6	1	1
7	2	2
8	3	3
8	3	3
Σ=50		

Ahora, calculemos la DM o $D_{\bar{x}}$ ...

$$DM = \frac{|2-5|+|2-5|+|4-5|+|4-5|+|4-5|+|5-5|+|6-5|+|7-5|+|8-5|+|8-5|}{10} =$$

$$DM = \frac{|-3|+|-3|+|-1|+|-1|+|-1|+|0|+|1|+|2|+|3|+|3|}{10} =$$

$$DM = \frac{3+3+1+1+1+0+1+2+3+3}{10} =$$

$$DM = \frac{18}{10} = 1{,}8$$

Veamos ahora cómo se calcula la desviación media en el caso de datos agrupados en intervalos.

$$\boxed{DM = \frac{\sum \left|(x-\bar{x}) f_i\right|}{N}}$$

...donde observamos que ahora las desviaciones van multiplicadas por las frecuencias de los intervalos correspondientes.

Además, las desviaciones son de cada centro, o marca de clase, a la media aritmética. Es decir:

$$\boxed{\bar{x} = \frac{\sum \left|f_i \cdot x\right|}{N}}$$

Ejemplo:

Para hallar la desviación media de la siguiente tabla referida a las edades de los 100 empleados de una cierta empresa:

Clase	**n_i**
16-20	2
20-24	8
24-28	8
28-32	18
32-36	20
36-40	18
40-44	15
44-48	8
48-52	3

Observa cómo se procede:

Clase	x_i	fi	$f_i \cdot x_i$	$\|x - \overline{x}\|$	$f_i \cdot \|x - \overline{x}\|$
16-20	18	2	36	16.93	33.86
20-24	22	8	176	12.93	103.44
24-28	26	8	208	8.93	71.44
28-32	30	18	540	4.93	88.74
32-36	34	20	680	0.93	18.6
36-40	38	18	684	3.07	55.26
40-44	42	18	756	7.07	127.26
44-48	46	8	368	11.07	88.56
48-52	50	3	150	15.07	45.21
		103	3598		632.37

Encontrando a $\bar{X}$...

$$\overline{x} = \frac{3598}{103} = 34.93$$

Con este dato, ya podremos encontrar la *DM o* $D_{\bar{x}}$ ...

$$DM = \frac{632.37}{34.93} = 18.1$$

La desviación media viene a indicar el grado de concentración o de dispersión de los valores de la variable. Si es muy alta, indica gran dispersión; si es muy baja refleja un buen agrupamiento y que los valores son parecidos entre sí.

La desviación media se puede utilizar como medida de dispersión en todas aquellas distribuciones en las que la medida de tendencia central más significativas haya sido la media. Sin embargo, para las mismas distribuciones es mucho más significativa la desviación típica, que estudiaremos a continuación, y eso hace que el uso de la desviación media sea cada vez más restringido.

VARIANZA

La varianza, S^2 o empleando el símbolo σ (letra griega sigma), se define como la media de las diferencias cuadráticas de n puntuaciones con respecto a su media aritmética, es decir:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad \sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{N}$$

Esta medida es siempre una cantidad positiva, con propiedades interesante para la realización de inferencia estadística. Como sus unidades son las del cuadrado de la variable, es más sencillo usar su raíz cuadrada, que es la que vemos en la siguiente sección.

Propiedades de la varianza:

- ✓ La varianza será siempre un valor positivo o cero, en el caso de que las puntuaciones sean iguales.
- ✓ Si a todos los valores de la variable se les suma un número la varianza no varía.
- ✓ Si todos los valores de la variable se multiplican por un número la varianza queda multiplicada por el cuadrado de dicho número.

- ✓ Si tenemos varias distribuciones con la misma media y conocemos sus respectivas varianzas se puede calcular la varianza total.

En caso las muestras tienen el mismo tamaña:

$$\sigma^2 = \frac{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2}{n}$$

Si las muestras tienen distinto tamaño:

$$\sigma^2 = \frac{k_1 \cdot \sigma_1^2 + k_2 \cdot \sigma_2^2 + \ldots + k_n \cdot \sigma_n^2}{k_1 + k_2 + \ldots + k_n}$$

DESVIACIÓN TÍPICA O ESTÁNDAR:

La varianza no tiene la misma magnitud que las observaciones (por ejemplo: si las observaciones se miden en metros, la varianza lo hace en metros cuadrados.

Si queremos que la medida de dispersión sea de la misma dimensionalidad que las observaciones bastaran con tomar su raíz cuadrada. Por ello se define la desviación típica, S, como:

$$S = \sqrt{S^2}$$

Ejemplo de cálculo de medidas de dispersión Calcular el rango, varianza y desviación típica de las siguientes cantidades medidas en metros:

3, 3, 4, 4, 5

Solución:

El rango de esas observaciones es la diferencia entre la mayor y menor de ellas; es decir: 5 − 3 = 2.

Para calcular las restantes medidas de dispersión es necesario calcular previamente el valor con respecto al cual vamos a medir las diferencias.

Esta es la media:

$$\overline{x} = \frac{3+3+4+4+5}{5} = 3{,}8\ metros.$$

La varianza es:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 = \frac{1}{5}\left(3^2 + 3^2 + 4^2 + 4^2 + 5^2\right) - 3,8^2 = 0,56 \text{ metros}^2$$

Siendo la desviación típica su raíz cuadrada:

$$S = \sqrt{S^2} = \sqrt{0,56} = 0,748 \text{ metros}$$

PROPIEDADES DE LA VARIANZA Y DESVIACIÓN TÍPICA:

Ambas son sensibles a la variación de cada una de las puntuaciones, es decir, si una puntuación cambia, cambia con ella la varianza.

La razón es que, si miramos su definición, la varianza es función de cada una de las puntuaciones.

La desviación típica tiene la propiedad de que en el intervalo

$$(\overline{x} - 2\mathcal{S}, \overline{x} + 2\mathcal{S}) \overset{\text{def}}{\sim} \overline{x} \pm 2\mathcal{S}$$

...se encuentra, al menos, el 75 % de las observaciones Incluso si tenemos muchos datos y estos provienen de una distribución normal (se definirá este concepto más adelante), podremos llegar al 95 %.

No es recomendable el uso de ellas, cuando tampoco lo sea el de la media como medida de tendencia central.

Ejemplo:

A continuación, se presenta una serie de datos a la que debemos calcularle la varianza y desviación típica, como aprendimos en clase:

12, 6, 7, 3, 15, 10, 18, 5

Lo primero que debemos realizar es, encontrar la media:

$$\overline{X} = \frac{12+6+7+3+15+10+18+5}{8} = 9.5$$

Ahora, calculemos la varianza:

$$\sigma^2 = \frac{12^2+6^2+7^2+3^2+15^2+10^2+18^2+5^2}{8} - 9.5^2 = 23.75$$

Encontremos la desviación típica:

$$\sigma = \sqrt{23.75} = 4.87$$

¡IMPORTANTE!

Como puedes observar; se ha empleado la letra de nuestro alfabeto "*S*", como el símbolo (letra griega) " σ " para representar el cálculo de la *Varianza* y *Desviación Típica* en sus respectivas fórmulas.

EJERCICIO 06: encuentra la Desviación Típica, Varianza y Desviación Media de las siguientes agrupaciones de datos o distribuciones. Realiza tus procedimientos en los espacios en blanco y utiliza el subrayado para colocar tus respuestas. Debes de entregar el ejercicio finalizado a tu catedrático/a, según te haya indicado.

1) 9, 3, 8, 8, 9, 8, 9, 18
$\overline{x} =$
$D_{\bar{x}} =$
$\sigma^2 =$
$\sigma =$

2) 10, 2, 8, 7, 7, 10, 8, 8, 7, 10, 9
$\overline{x} =$
$D_{\bar{x}} =$
$\sigma^2 =$
$\sigma =$
3) 2, 3, 6, 8, 11.
$\overline{X} =$
$D_{\bar{x}} =$
$\sigma^2 =$
$\sigma =$
4) 12, 6, 7, 3, 15, 10, 18, 5.
$\overline{X} =$
$D_{\bar{x}} =$
$\sigma^2 =$
$\sigma =$
5) 10, 12, 16, 18, 22, 15, 18, 18, 15, 16, 13 + 15
$\overline{x} =$

$D_{\bar{x}} =$
$\sigma^2 =$
$\sigma =$
6) 8, 7, 6, 5, 5, 4, 3, 5, 4, 3
$\overline{x} =$
$D_{\bar{x}} =$
$\sigma^2 =$
$\sigma =$
7) 1, 3, 5, 2, 4, 3, 2, 4
$\overline{x} =$
$D_{\bar{x}} =$
$\sigma^2 =$
$\sigma =$
8) 11, 13, 11, 15, 11, 14
$\overline{x} =$
$D_{\bar{x}} =$

$\sigma^2 =$
$\sigma =$
9) 1, 2, 3, 4, 5, 5, 4, 2, 3, 2, 3, 2, 1, 2, 3, 5, 4, 3, 2, 1
$\overline{x} =$
$D_{\bar{x}} =$
$\sigma^2 =$
$\sigma =$
10) 1, 4, 3, 5, 1, 2, 5
$\overline{x} =$
$D_{\bar{x}} =$
$\sigma^2 =$
$\sigma =$

EJERCICIO 07: debes completar la siguiente tabla; luego, encuentra la Varianza y Desviación Típica de la siguiente distribución de datos:

	x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[10, 20)	15	1		
[20, 30)	25	8		
[30,40)	35	10		
[40, 50)	45	9		
[50, 60	55	8		
[60,70)	65	4		
[70, 80)	75	2		
		42	1 820	88 050

$\bar{x}$ =	σ^2 =	σ =

EJERCICIO 08: 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:

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

Debes calcular la Desviación Típica de los datos expresados en la tabla. Para ello, crea la tabla de distribución de frecuencias correspondiente.

Ahora, puedes calcular la Desviación Típica (recuerda que antes debes realizar otros cálculos).

$\bar{X} =$	$\sigma^2 =$	$\sigma =$

EJERCICIO 09: el resultado de lanzar dos dados 120 veces viene dado por la tabla:

Sumas	2	3	4	5	6	7	8	9	10	11	12
Veces	**3**	**8**	**9**	**11**	**20**	**19**	**16**	**13**	**11**	**6**	**4**

Construye la tabla de distribución de frecuencias correspondiente.

Calcular la Desviación Estándar (recuerda que antes debes realizar otros cálculos).

$\bar{X} =$	$\sigma^2 =$	$\sigma =$

EJERCICIO 10: calcular la Desviación Típica de una distribución estadística que viene dada por la siguiente tabla:

	[10, 15)	[15, 20)	[20, 25)	[25, 30)	[30, 35)
f_i	3	5	7	4	2

Construye la tabla de distribución de frecuencias correspondiente.

Calcular la Desviación Estándar (recuerda que antes debes realizar otros cálculos).

$\bar{X} =$	$\sigma^2 =$	$\sigma =$

EJERCICIO 11: calcular la Desviación Típica de la distribución de la tabla:

	x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[10, 20)	15	1	15	225
[20, 30)	25	8	200	5000
[30,40)	35	10	350	12 250
[40, 50)	45	9	405	18 225
[50, 60)	55	8	440	24 200
[60,70)	65	4	260	16 900
[70, 80)	75	2	150	11 250
		42	1 820	88 050

Calcular la Desviación Estándar (recuerda que antes debes realizar otros cálculos).

$\bar{X} =$	$\sigma^2 =$	$\sigma =$

INFORMACIÓN (INCLUÍDA EN ESTE DOCUMENTO EDUCATIVO) TOMADA DE:

Sitios web:

http://www.vitutor.com/estadistica/descriptiva/b_13.html

http://www.vitutor.com/estadistica/descriptiva/a_15.html

http://www.sectormatematica.cl/media/NM4/NM4_medidas_de_dispersion.doc

http://www.universoformulas.com/estadistica/descriptiva/diagrama-caja/

http://dta.utalca.cl/estadistica/ejercicios/obtener/descriptiva/EjerciciosResueltosEstadisticaDescriptiva.pdf

http://www.estadisticaparatodos.es/taller/graficas/cajas.html

https://sites.google.com/site/ulaeconomia/ejercicios-estadistica-i

http://www.vitutor.com/estadistica/descriptiva/a_14.html

http://aymaraolivo95.blogspot.com/

http://www.monografias.com/trabajos89/desviacion-estandar/desviacion-estandar.shtml

http://berumenii.blogspot.com/2011/08/rango-percentil-y-percentil.html

http://fcm.ens.uabc.mx/~chelo/estadistica/doc-pdf/lec-02-4.pdf