INDICE

ANALISIS DESCRPTIVO DE UNA VARIABLE

ANALISIS DESCRPTIVO CONJUNTO DE DOS O MÁS VARIABLES

6. Relación lineal entre dos variables. Correlaciones.

    6.1. Distribución conjunta de dos variables. Diagrama de dispersión.
    6.2. Cálculo del índice de correlación por el método de Bravais – Pearson, con los datos de su variable y del alumno contiguo.
    6.3. Valoración de la significación e interpretación de la correlación obtenida.

7. Correlaciones con más de dos variables. (Utilizarán todas las variables asignadas al grupo).

    7.1. Cálculo de la correlación parcial entre cada una de las variables y las otras variables asignadas. Valoración, significación e interpretación.
    7.2. Cálculo del coeficiente de correlación múltiple con todas las variables. Significación e interpretación.

TIPIFICACIÓN Y ELABORACIÓN DE BAREMOS

8. Elaboración de escalas de medida. Baremos.

    8.1. Elaboración de escalas en normas ordinales: deciles y centiles.
    8.2. Elaboración de escalas en normas típicas: puntuaciones z, puntuaciones D, puntuaciones T, puntuaciones normalizadas y eneatipos.

ANALISIS DESCRIPTIVO DE UNA VARIABLE

1. Acopio, ordenación y organización de los datos de la variable.

1B			
15	15	16	21
21	16	17	16
16	17	15	18
16	16	17	16
12	21	16	16
14	16	13	21
16	16	19	14
16	17	12	14
18	20	14	11
16	18	19	15
14	19	17	15
14	21	19	18
14	16	11	15
12	20	18	16
14	16	15	14
15	17	19	15
14	16	13	15
16	17	14	13
15	18	18	17
16	17	12	14
16	18	15	14
16	17	15	14
13	18	15	16
14	20	15	14
15	20	13	16

1.1. Determinación de la amplitud total o rango de la distribución de la variable. Definición de la amplitud y número de intervalos para la distribución de los datos.

La amplitud total se define como, la resta entre el valor superior y el inferior más la unidad, es decir, el incremento entre el valor superior y el inferior más la unidad.

$$A_T = (X_s - X_i) + 1$$

En nuestro caso, la amplitud total es:

$$A_T = (21 - 11) + 1 = 11$$

Para definir la amplitud de cada intervalo, utilizamos la fórmula de la amplitud por intervalo:

$$\text{Amplitud i} = \frac{A_T}{N^o \text{ de } i} = \frac{11}{11} = 1$$

1.2. Distribución de frecuencias en los intervalos definidos.

En nuestro caso particular, la distribución de frecuencias hace referencia a la siguiente tabla, donde se muestran los intervalos que se desprenden de los datos recogidos y la distribución que hemos creído oportuna.

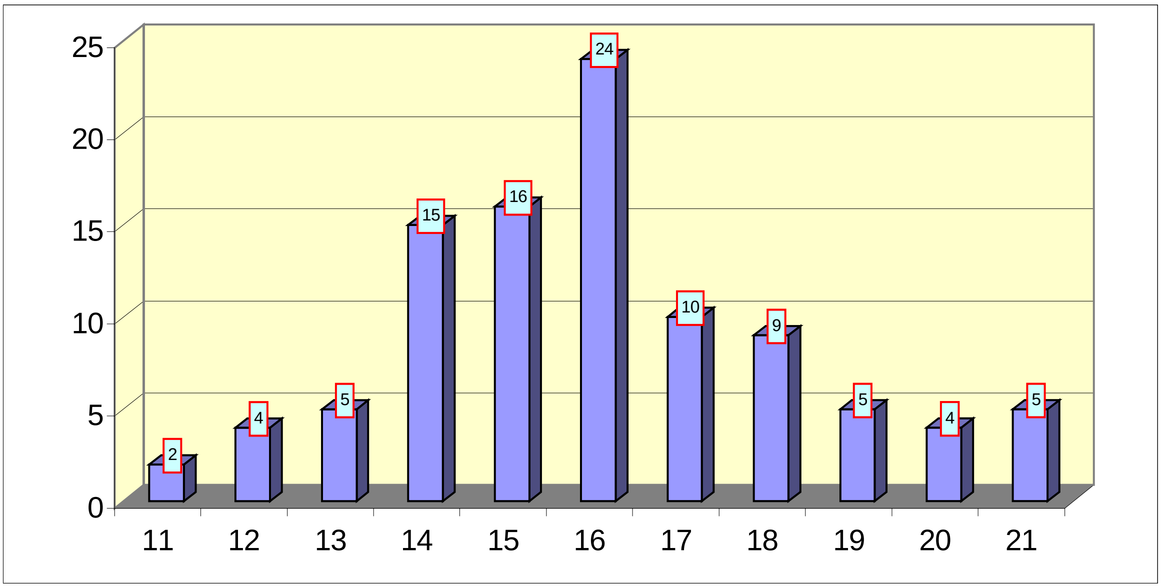
X (Intervalos)	**F (frecuencias)**
21	5
20	4
19	5
18	9
17	10
16	24
15	16
14	16
13	5
12	4
11	2

N = Σ f = 5 + 4 + 5 + 9 + 10 + 24 + 16 + 16 + 5 + 4 + 2 = 100

1.3. Representación gráfica del histograma y polígono de frecuencias.

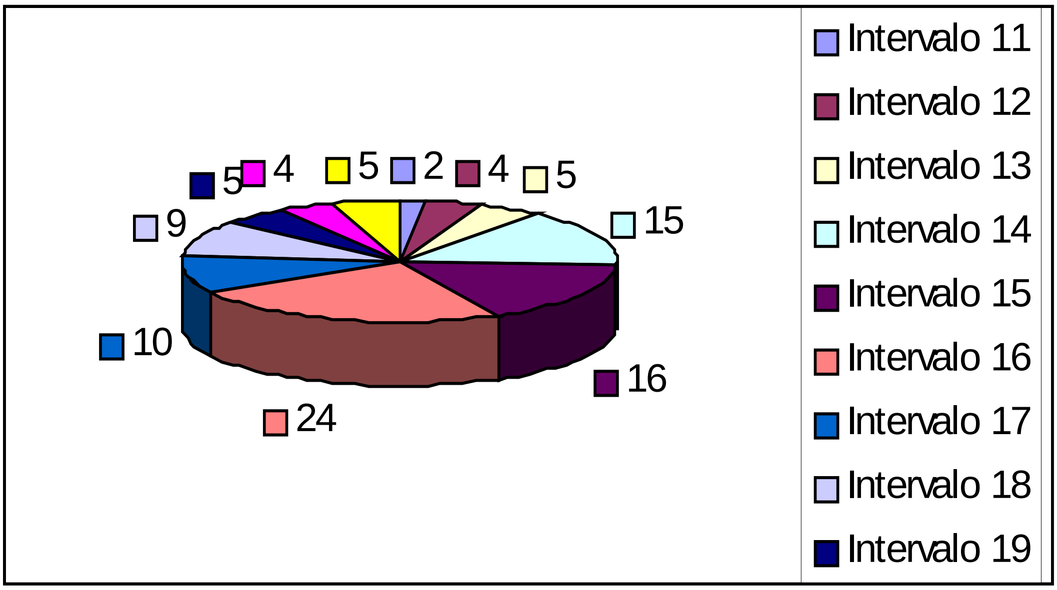
Lo que pretende mostrar en el histograma es una superficie que represente la distribución de la variable, pero en vez de puntos se utilizan barras. Con nuestros datos hemos elaborado el siguiente:

HISTOGRAMA

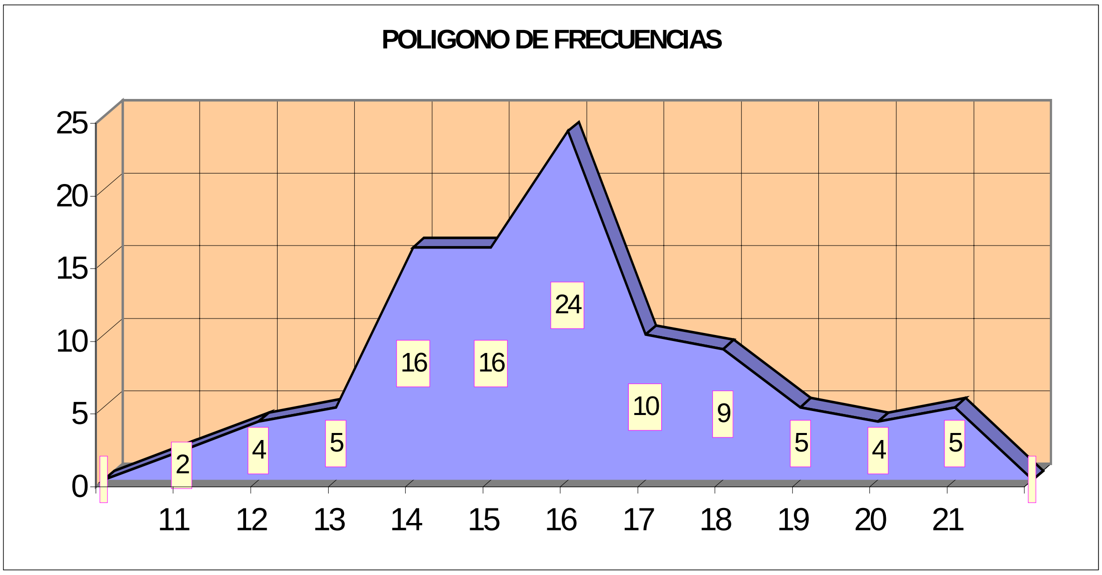


Cuando el total de casos suma el 100 %, es decir estamos tratando con porcentajes o el número de casos es de 100, podemos utilizar otro sistema de representación que es denominado en estadística, Pictograma. Con nuestros datos resulta el siguiente gráfico (Utilizamos diferentes colores para diferenciar las porciones del total).

PICTOGRAMA



Como hemos dicho el diagrama de sectores se utiliza para representar partes de un centenar, es decir tantos por cien de un total de casos. El introducirlo aquí ha sido con el único objetivo de representar la idea de pictograma. El siguiente gráfico es el polígono de frecuencias, el cual hace referencia a una superficie que está delimitada por los puntos que nosotros introducimos en la base de datos y que como es de suponer cuanto más altos sean los valores mayor será la superficie.



1.4. Representación gráfica del polígono de frecuencias suavizado. Métodos aritmético y gráfico.

La suavización del polígono de frecuencias se hace para que la toma de intervalos no influya en la interpretación del gráfico. Existen dos métodos, aunque realmente son el mismo, pero uno se realiza sobre el papel y otro a través de cálculos. A continuación describimos el método aritmético y posteriormente junto con el gráfico, expondremos el método sobre el papel.

Método Aritmético:

Se realiza a través de una fórmula que permite el cálculo de un punto intermedio que limite el posible error producido por la toma de intervalos. La frecuencia suavizada, según la fórmula, es igual a la frecuencia anterior más dos veces la frecuencia que queremos suavizar más la frecuencia posterior y todo dividido por cuatro:

$$f_s = \frac{f_a + 2f_s + f_p}{4}$$

La siguiente tabla muestra la aplicación de esta fórmula a los datos que nos han sido consignados.

$$Fs\ 1 = \frac{2 + 2 \times 4 + 5}{4} = \frac{15}{4} = 3.75$$

$$Fs\ 2 = \frac{4 + 2 \times 5 + 16}{4} = \frac{30}{4} = 7.5$$

$$Fs\ 3 = \frac{5 + 2 \times 16 + 16}{4} = \frac{53}{4} = 13.25$$

$$Fs\ 4 = \frac{16 + 2 \times 16 + 24}{4} = \frac{72}{4} = 18$$

$$Fs\ 5 = \frac{16 + 2 \times 24 + 10}{4} = \frac{74}{4} = 18.5$$

$$Fs\ 6 = \frac{24 + 2 \times 10 + 9}{4} = \frac{53}{4} = 13.25$$

$$Fs\ 7 = \frac{10 + 2 \times 9 + 5}{4} = \frac{33}{4} = 8.25$$

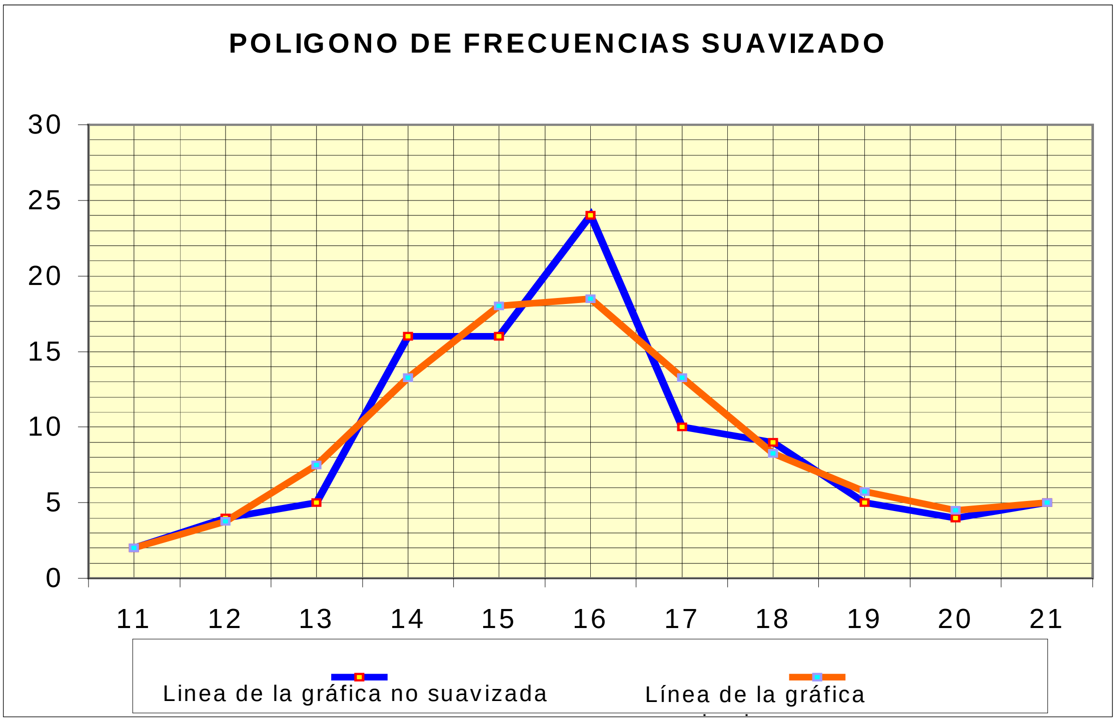
$$Fs\ 8 = \frac{9 + 2 \times 5 + 4}{4} = \frac{23}{4} = 5.75$$

$$Fs\ 9 = \frac{5 + 2 \times 4 + 5}{4} = \frac{18}{4} = 4.5$$

X	f	f suavizada
21	5	
20	4	3.75
19	5	7.5
18	9	13.25
17	10	18
16	24	18.5
15	16	13.25
14	16	8.25
13	5	5.75
12	4	4.5
11	2	

Método Gráfico:

De igual forma la suavización se puede hacer sobre el papel milimetrado, debiendo a parecer el mismo valor por los dos métodos. En la siguiente gráfica se muestra la representación de ambos métodos. El método gráfico consiste en unir la frecuencia anterior y la posterior mediante una línea. La prolongación vertical de la frecuencia al chocar con la línea que acabamos de trazar, nos define un segmento que se debe encontrar entre la línea trazada y el punto que corresponde con la frecuencia que queremos suavizar. El punto medio de ese segmento será el valor suavizado para esa frecuencia. En el caso de la primera y la última frecuencia, se debe suponer que hay una frecuencia anterior que es cero (en el caso de la 10 frecuencia) y otra posterior a la ultima con igual valor.

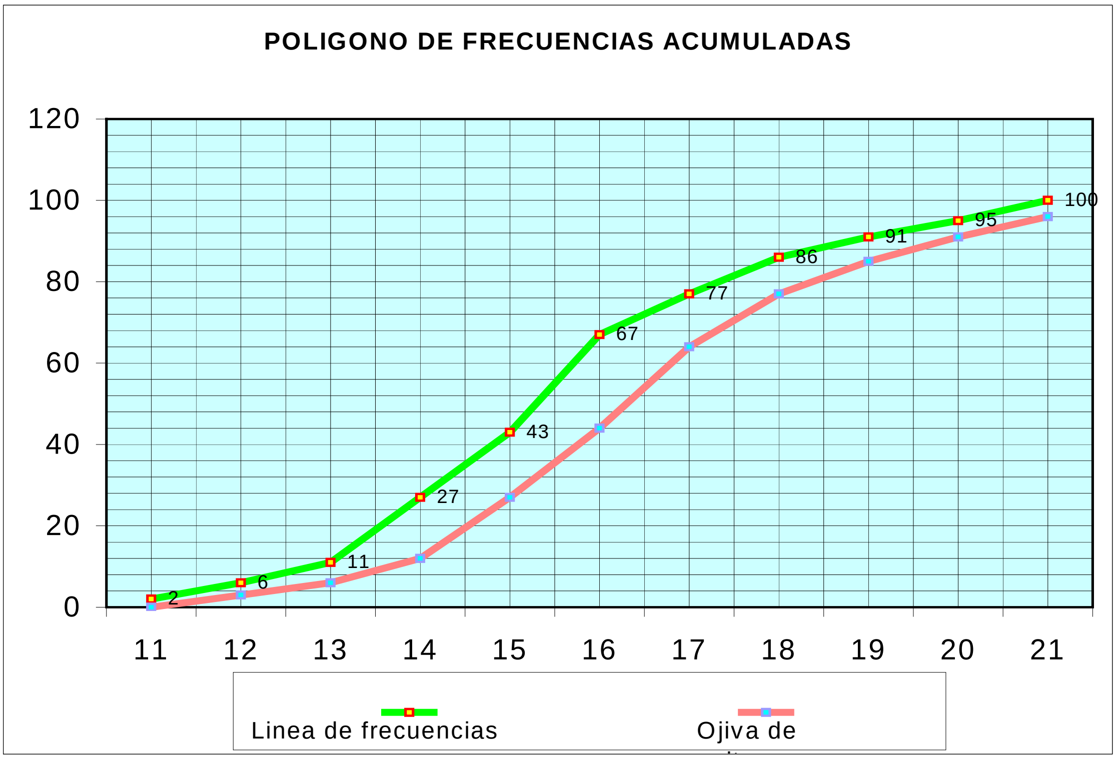


1.5. Distribución de frecuencias acumuladas en intervalos. Representación gráfica.

La frecuencia acumulada es aquella que es el resultado de la suma de las anteriores frecuencias, es decir, es la suma de todas las anteriores. La siguiente tabla muestra la relación de los intervalos con las frecuencias convencionales y las acumuladas.

X	f	facumulada
21	5	100
20	4	95
19	5	91
18	9	86
17	10	77
16	24	67
15	16	43
14	16	27
13	5	11
12	4	6
11	2	2

La representación gráfica de esta tabla nos debe dar una clásica forma en Ojiva (Galton). En la siguiente gráfica se muestra la función matemática que resulta de nuestros datos comparada con la mencionada curva de Galton.



2.1. Cálculo de la media aritmética. Método general con intervalos de clase y método reducido o de la media supuesta.

Se supone que cuando hablamos de media, estamos hablando de media aritmética, aunque conozcamos la existencia de otros tipos de medias que no lo sean y que después ejemplificaremos. De cualquier forma la existencia de la media es trascendental para el conocimiento del carácter de la distribución, ya que es el valor que mejor la representa. Cuando hablamos de media aritmética, no referimos a encontrar el valor medio de una variable que represente de forma genérica la tendencia media hacia una variable concreta de esa población, existiendo tres formas de localizarla:

El ***primer método*** se utiliza cuando la muestra o la población no son muy numerosas y son pocos valores con los que se tiene que operar. El procedimiento para encontrar exige que se sumen los valores obtenidos en la muestra y se divida por el número de casos. En el caso de nuestros datos hemos obtenido los siguientes resultados:

$$\bar{X} = \frac{\Sigma X}{N} = \frac{1595}{100} = 15,95$$

Esto quiere decir que el valor que mejor representa nuestra muestra es este, 15,95 y que debe repetirse en los dos procedimientos restantes para hallar la media.

El ***segundo método,*** consiste en hallar el punto medio de cada intervalo. Una vez hecho esto se multiplica cada uno por la frecuencia de su intervalo y dividido por el número de casos también debe darnos la media. En la siguiente tabla se muestra cual es el procedimiento de obtención:

X intervalos	**Frecuencias**	**X_m**	**f • X_m**
21	5	21	105
20	4	20	80
19	5	19	95
18	9	18	162
17	10	17	170
16	24	16	384
15	16	15	240
14	16	14	224
13	5	13	65
12	4	12	48
11	2	11	22
N = 1f = 100		Σf • Xm = 1595	

La fórmula con la que hallamos la media en este caso es la siguiente:

$$\bar{X} = \frac{\Sigma f \cdot Xm}{N} = \frac{1595}{100} = 15{,}95$$

Solo nos queda el ***tercer método*** por descubrir para hallar la media aritmética, para lo cual hay que puntualizar que igual que dijimos que el primer método solo nos servía para pocos datos y no agrupados, decimos que el segundo y tercer método se utilizan para gran cantidad de datos y en intervalos agrupados. Este método recibe el nombre de método de la media supuesta debido a que suponemos un valor que nosotros estimamos (puede ser cualquiera) y afirmamos lo siguiente.

$$\bar{X} = \bar{X}_s + C$$

Es decir que el valor de la media real, será igual a la media que nosotros suponemos más un factor corrector que será el que corregirá la media supuesta para hacerla real. El problema está entonces en encontrar este factor de corrección. El valor medio supuesto que nosotros estimemos (puede ser cualquiera) estará comprendido en un intervalo, debiendo ser X_s el valor medio de dicho intervalo. En definitiva la fórmula global queda de la siguiente forma.

$$\bar{X} = \bar{X}_s + C = X_s + \frac{\Sigma f \cdot X'}{N} . \; i$$

Donde la i es la amplitud de cada período y la X= es un nuevo valor que se define por el número de intervalos de distancia que hay hasta el que nosotros hemos considerado como media. Ponemos como ejemplo los datos que nosotros utilizamos y en el que hemos supuesto 16 como media.

X intervalos	Frecuencias	X′	F • X′
21	5	5	25
20	4	4	16
19	5	3	15
18	9	2	18
17	10	1	10
16	24	0	0
15	16	-1	-16
14	16	-2	-32
13	5	-3	-15
12	4	-4	-16
11	2	-5	-10
N = 1f = 100		Σf • X′= = - 5	

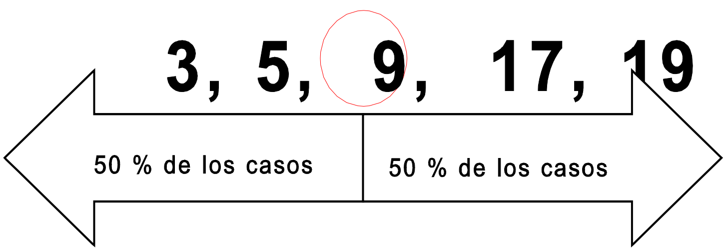
Sustituyendo en la fórmula que habíamos indicado anteriormente obtenemos lo que esperábamos, es decir que los valores obtenidos por los tres procedimientos coinciden:

$$\bar{X} = \bar{X}_s + C = \bar{X}_s + \frac{\Sigma f \cdot X'}{N} \; . \; i = 16 + \frac{-5}{100} \cdot 1 = 15{,}95$$

2.2. Cálculo de la mediana.

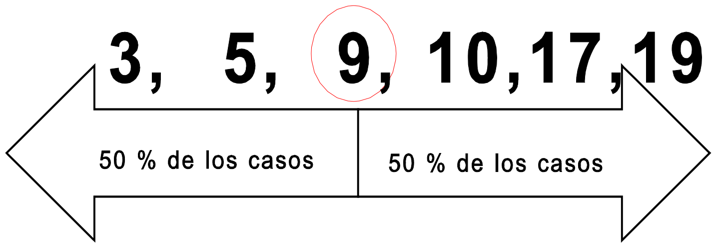
La mediana es un punto que deja por encima y por debajo de sí el 50 % de los casos, es decir, es el valor que corresponde a ese punto. Los siguientes ejemplos pueden ilustrar la complejidad que suscita este término en diferentes distribuciones:

∗ ***Cuando una serie es impar:***



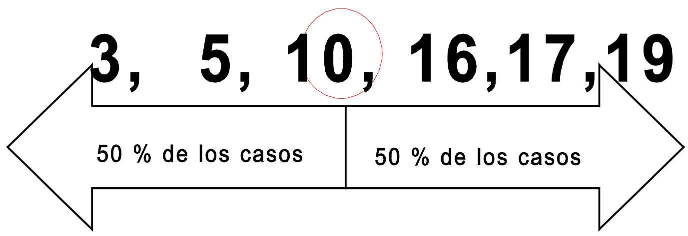
Es el más claro de los casos, la serie tiene un número de casos impar, por lo que siempre habrá un caso central cuyo valor será la mediana.

∗ ***Cuando la serie es par y continua:***



En este segundo caso lo que se hace es coger el valor del punto medio entre los dos valores centrales, es decir, en este ejemplo, sería 9'5.

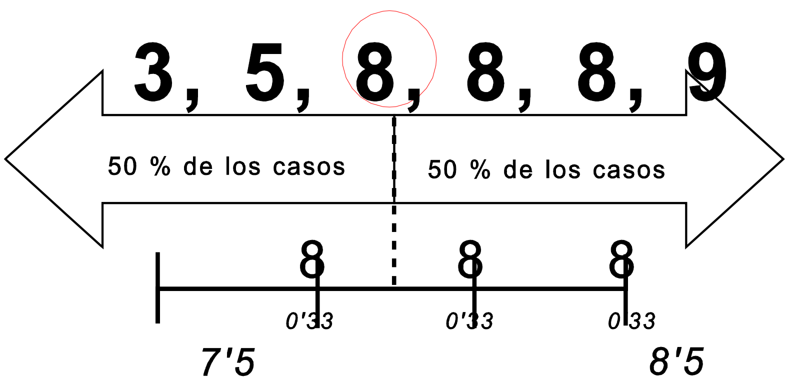
**** Cuando la serie es par y discontinua:***



Este es el tercer caso, y en el se observa que hay un salto en la serie justo en los valores en los que debe encontrarse la mediana. Lo que se hace es tomar el punto medio de los valores centrales, es decir en nuestro ejemplo sería el valor correspondiente al punto 13 de la serie.

∗ ***Cuando la mediana se encuentra entre dos valores de la serie que son iguales:***

El último de los casos puede ser el más complicado, debido a que se ha de realizar una serie de cálculos matemáticos que pueden parecer complicados pero que las ilustraciones nos aclararán perfectamente:



En esta serie observamos que para el intervalo 7'5 a 8'5 (una unidad) existen tres valores que son candidatos a ser mediana. Para que quede el 50 % de datos a cada lado, debemoa partir la serie , dejando el primer 8 a la izquierda y los otros a la derecha. Los cálculos a desarrollar, son que la mediana será igual al límite inferior más lo que tengo que avanzar dentro del intervalo para que se alcance el 50 %, como lo que hay que avanzar es 1/3 de 1, es decir 0'33 el resultado final será:

$$Mdn = 7{,}5 + 0{,}33 = 7{,}83$$

Pero todos estos datos son necesarios cuando los datos no están agrupados, ya que el método a utilizar con los datos agrupados es el siguiente:

1º: Dividir el N:2, lo cual en mi caso sería N = 100, N/2 = 50

2º: Nos dirigimos a la frecuencia acumulada y buscamos al sujeto 50, anotando el intervalo en el que se encuentra este sujeto, puesto que dentro de este intervalo será donde se encuentre la mediana.

3º: Aplicamos la fórmula siguiente para localizar exactamente el punto:

$$Mdn = Lim\ i + (N/2 - fa_{ant}) \cdot i / f_i$$

Lo que quiere decir, que la media es igual al límite real inferior del intervalo más el número de casos partido de 2, menos la frecuencia acumulada anterior al intervalo que estamos utilizando, multiplicando el paréntesis por el resultado de la fracción entre la amplitud y la frecuencia del intervalo actual. Halamos la mediana por este método y obtenemos el siguiente resultado:

X intervalos	**Frecuencias**	**Frecuencia Acumulada**
21	5	100
20	4	95
19	5	91
18	9	86
17	10	77
16	24	67
15	16	43
14	16	27
13	5	11
12	4	6
11	2	2
N = Σf = 100		

$$\textbf{Mdn} = 15.5 + (100 / 2 - 43) \cdot 1 / 24 = 15.5 + 7 \cdot 1 / 24 = 15.79$$

2.3. Cálculo de la Moda.

La moda hace referencia a aquel valor de la distribución que más se repite, lo cual en mi caso es el 16, pero hay varias formas de hallarla. Si miramos las frecuencias de la tabla de arriba nos damos cuenta de que los valores que más se repiten son los del intervalo 16, pero cual de ellos debemos coger, el 15.5, el 16 o el 16.5 o cual. Si tenemos la posibilidad de contar los datos hay que decir que la moda es el valor más repetido en la distribución. Si no contamos con los datos hay que decir que el valor es Mo=16, es decir el valor medio del intervalo. Pero existe otra posibilidad. Si conocemos la mediana y la media, se puede hallar la moda con la siguiente fórmula:

$$M_o = 3\ Mdn - 2\ \overline{X}$$

Sustituyendo los valores anteriormente obtenidos en los procesos para hallar la media y la mediana, se obtiene el siguiente valor de moda:

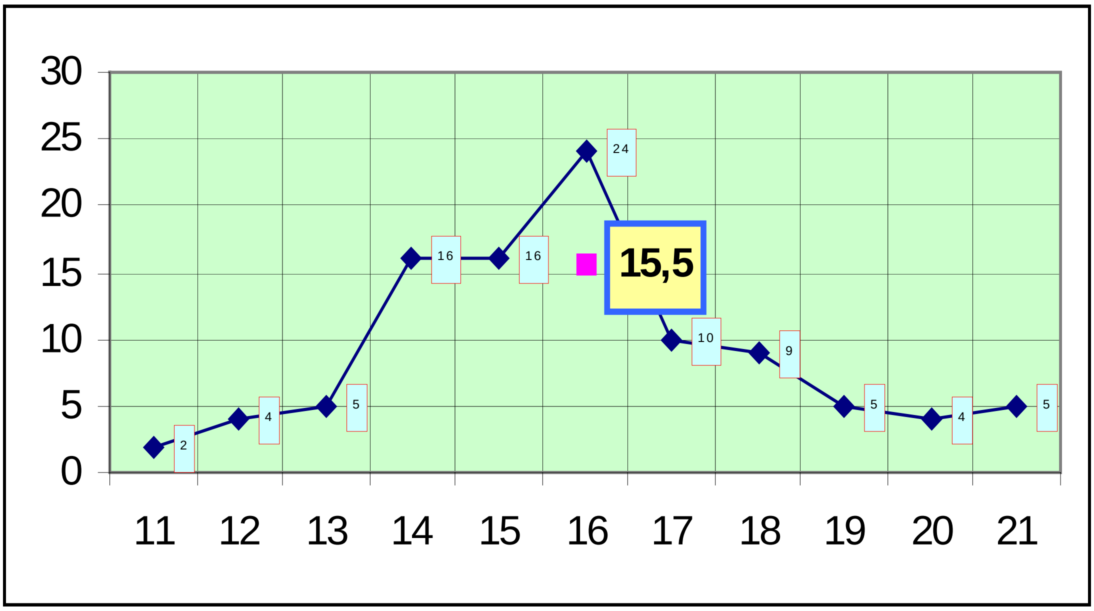
$$\mathrm{M_o} = 3\ (15.79) - 2\ (15.95) = 15.47$$

Los valores obtenidos con los distintos métodos varían pero la moda real es aquella que se deriva de los datos reales, pero a veces por tener los datos agrupados o por ser un proceso demasiado largo se utilizan estas aproximaciones. De cualquier forma las fluctuaciones son mínimas.

También existe la posibilidad de que el valor más repetido no sea único, es decir que haya dos valores que sean los que más se repiten significativamente en relación a los demás, cuando esto ocurre se dice que la distribución no es unimodal.

2.4. Localización de estos tres estadísticos en el gráfico de la distribución y elaboración de un juicio comparativo sobre los mismos.

La localización de los tres puntos en la gráfica se realiza por medio de la representación que utilizamos en la primera parte de este cuaderno para representar los valores de la distribución.



El punto 15,5 representa aproximadamente la Media, Mediana y Moda.

Como se puede observar perfectamente en el gráfico, los tres valores coinciden muy puntualmente, dentro del intervalo 15,5 – 16,5 y muy próximos entre sí. En principio esto es una muestra de que la muestra es homogénea, pero habrá que esperar a otros resultados del análisis para poder afirmar que ciertamente lo es.

Existen otras tipos de medias que deben ser tenidas en cuenta, tales como:

1. La media ponderada: Es la media de las medias, es decir es hacer la media aritmética de una serie de medias obtenidas anteriormente. Su fórmula es el sumatorio de las respectivas medias partido por el número de medias consideradas y se expresa de la siguiente manera:

$$\bar{X}_{tt} = \frac{\Sigma \ \bar{X}}{N^{o} \ \bar{X}}$$

Esta fórmula sólo se puede utilizar cuando el N es para todas las medias igual, cuando no es así la fórmula a utilizar es otra

$$\bar{X}_{tt} = \frac{\Sigma\ N \cdot \bar{X}}{N^o\ \bar{X}} = \frac{N_1 \cdot \bar{X} + N_2 \cdot \bar{Y} + N_3 \cdot \bar{Z} + N_n \cdot \bar{X}_n}{N_t}$$

Con un caso práctico lo entenderemos más fácilmente. Se han realizado tres estudios diferentes, donde se han obtenidos medias de tres poblaciones diferentes y que se quieren comparar hallando la media ponderal entre las mismas, los datos se muestran en la siguiente tabla:

Nombre de la media	**Valor de la media**	**Asignación del número de casos**	**Número de casos**
X	15	N_1	200
Y	17	N_2	80
Z	17	N_3	40

Utilizando la fórmula arriba descrita llegamos a los valores de la media ponderal de este ejemplo:

$$X\ t = \frac{15 \cdot 200 + 17 \cdot 80 + 17 \cdot 40}{320} = 15{,}75$$

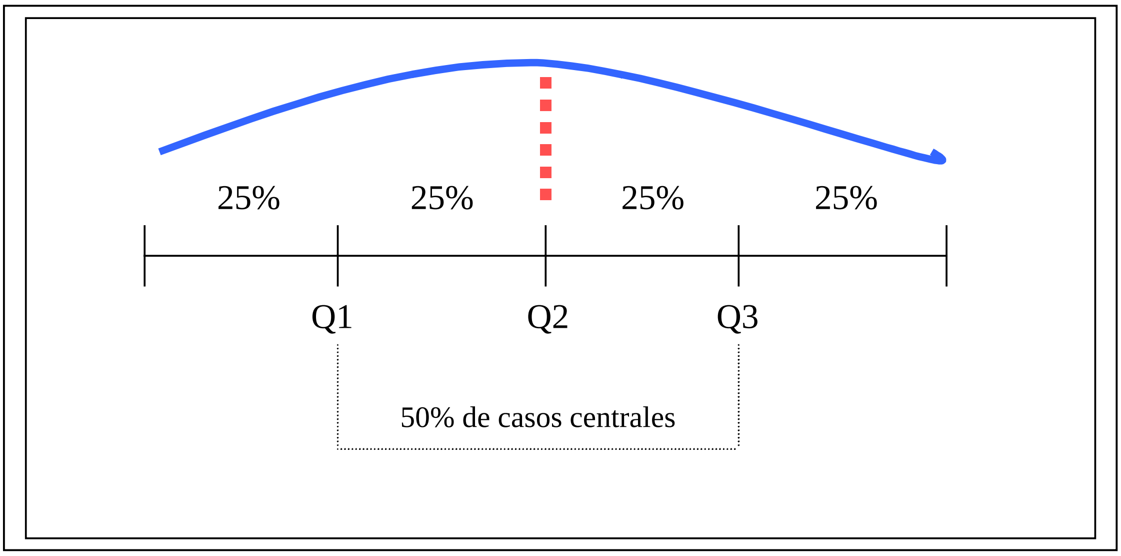
3.1. Cálculo de los puntos cuartiles (Q_1,Q_2,Q_3) de la distribución.

En primer lugar debemos decir que en los estadísticos de variabilidad encontramos unos que denominamos de recorrido, que son por un lado la amplitud total que ya conocemos y por otro la desviación o amplitud intercuartil que definiremos posteriormente.

Hay que hacer una distinción entre el cuartil y el cuantil. El cuantil es una cantidad en la que se divide la muestra proporcionalmente, es decir son unos puntos que determinan distintas proporciones de área o de frecuencia. Es una división proporcional que si se realiza a través de cuartos se denomina cuartil, si la dividimos en diez fracciones se denominan decil y si la dividimos en cien serán centiles.

El cuartil 1 (Q_1) es aquel que deja por debajo de si el 25 % de los casos de la muestra. Q_2 deja por debajo de sí el 50 % de los casos de la muestra (coincide con la mediana) y Q_3 deja el 75 % de los casos por debajo de sí.

La amplitud intercuatil es el recorrido entre Q_1 y Q_3 y siempre debe encontrarse el 50 % de los casos centrales.



La amplitud intercuartil se utiliza para evitar el influjo de los valores extremos y así poder comparar muestras a través de este 50 % de datos centrales. A continuación se describe la forma de cálculo de estos cuartiles.

⇒ **Pasos a seguir:**

1º Encontramos el número de casos que corresponde al cuartil que pretendemos encontrar utilizando la siguiente fórmula:

$$\frac{\text{Nº de casos}}{Q_n} = \frac{N}{4} \cdot n$$

2º El proceso anterior nos dará un número de casos concreto que debemos localizar en la frecuencia acumulada para encontrar en qué intervalo se encuentra el Q y después, aunque se puede utilizar un razonamiento lógico deductivo abogamos por la utilización de la siguiente fórmula:

$$Q_n = \lim_{inf} + (N/4 \,.\, n – f\,a_{ant})\, i / f_i$$

Donde n pequeña indica el número de cuartil que corresponda, que es igual al límite exacto inferior del intervalo donde caiga el cuartil, más el número de casos totales de la muestra dividido entre cuatro y multiplicado por el número de cuartil correspondiente menos la frecuencia acumulada anterior y todo el paréntesis multiplicado por la amplitud que a su vez está dividida de la frecuencia propia del intervalo pero sin acumular. Parece un lío de mucho cuidado pero con los siguientes ejemplos se entenderá perfectamente:

∗ ***Hallando el Q*** $_1$***...***

1. Hallamos el n1 de casos por cuartil:

$$\frac{\text{Nº de casos}}{Q_n} = \frac{N}{4} \cdot n = \frac{100}{4} \cdot 1 = 25 \text{ casos}$$

2. Hallamos el intervalo donde se encuentra el Q_1:

Intervalos	**Frecuencias**	**F. acumuladas**	**Q**
21	5	100	
20	4	95	
19	5	91	
18	9	86	
17	10	77	Q_3 75 casos
16	24	67	Q_2 50 casos
15	16	43	
14	16	27	Q_1 25 casos
13	5	11	
12	4	6	
11	2	2	
N = 1f = 100			

3. Por último debemos encontrar cual es el valor real de Q_1 en la distribución, para lo cual utilizamos la fórmula:

$$Q_1 = \lim_{inf} + (N/4 \,.\, n - f\,a_{ant})\; i / f_i = 13{,}5 + (100 / 4 \cdot 1 - 11)\; 1/ 16 = 14.375$$

Para calcular los siguientes cuartiles no seremos tan descriptivos, puesto que el procedimiento es similar y no es necesario parafrasear nuestros propias palabras. Los resultados obtenidos para Q_2 y Q_3 son los que expresan las siguientes fórmulas:

∗ ***Hallando el Q*** $_2$***...***

$$Q_2 = \lim_{inf} + (N/4 \,.\, n - f\,a_{ant})\; i / f_i = 15.5 + (100 / 4 \cdot 2 - 43)\; 1/ 24 = 15.79$$

∗ ***Hallando el Q*** $_3$***...***

$$Q_3 = \lim_{inf} + (N/4 \,.\, n - f\,a_{ant})\; i / f_i = 16.5 + (100 / 4 \cdot 3 - 67)\; 1/ 10 = 17.3$$

Pero lo que nos ha movido a realizar esta cantidad de cálculo es única y exclusivamente halar la amplitud intercuartil, la cual se halla restando Q_3 - Q_1:

$$Q_3 - Q_1 = 17.3 - 14.375 = 2.925$$

Este resultado tiene que hacernos pensar, encontramos que en menos de dos valores se encuentran el 50 % de los valores de la distribución.

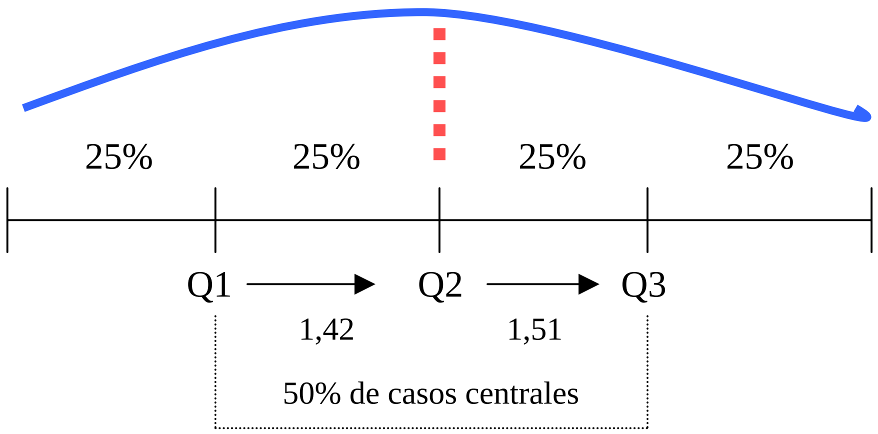
3.2. Cálculo de la amplitud semi – intercuartil.

El siguiente proceso va encaminado a estimar si podemos decir que hay una simetría de la distribución en la amplitud intercuartil con respecto a la ordenada cero, es decir como oscila la gráfica en este tramo central, para ello debemos encontrar primero la amplitud semi - intercuartil, la cual se representa por Q y su fórmula es la siguiente:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{2.925}{2} = 1.4625$$

Para saber si existe simetría en la distribución comparamos la amplitud semi-intercuartil con las diferencias entre los cuartiles, si son similares la distribución del 50 % de los casos centrales será simétrica, sino es similar la distribución estará apuntada, es decir, tendrá curtosis.

Si observamos la siguiente figura, donde se han hallado las diferencias entre Q_3 - Q_2 y entre Q_2 -Q_1, se puede observar los resultados de la simetría. Aunque la diferencia es mínima (3 centésimas) se puede decir que la distribución en sus 50 % de casos centrales es prácticamente simétrica, aunque como se puede observar en los resultado hay una cierta acumulación de datos entre Q3 -Q2, ya que en menos espacio hay el mismo n1 de casos que entre Q_2 -Q_1.



3.3. Cálculo de la desviación media (D.M).

Para calcular la desviación media se utiliza una fórmula que relaciona, el sumatorio de los diferentes valores que toma la muestra con la media aritmética y partido del número total de casos:

$$D.M = \frac{\sum |(X - \bar{X})|}{N}$$

Esto bajo el supuesto de que los datos no estén *agrupados* en intervalos, en cuyo caso la fórmula a utilizar sería (si están agrupados) la siguiente:

$$D.M = \frac{\sum f \cdot |(Xm - \bar{X})|}{N}$$

En la siguiente tabla procedemos a calcular la desviación media que hacer honor a nuestra distribución:

$$\bar{X} = \frac{\sum f \cdot Xm}{N} = \frac{1595}{100} = 15.95$$

X intervalos	Frecuencias	Xm	$Xm - \bar{X}$	$f \cdot \lvert (Xm - \bar{X}) \rvert$
21	5	21	5.05	25.25
20	4	20	4.05	16.2
19	5	19	3.05	15.25
18	9	18	2.05	18.45
17	10	17	1.05	10.5
16	24	16	0.05	1.2
15	16	15	- 0.95	15.2
14	16	14	- 1.95	31.2
13	5	13	- 2.95	14.75
12	4	12	- 3.95	15.8
11	2	11	- 4.95	9.9
N = 1f = 100		$\sum f \cdot \lvert (Xm - X) \rvert = 173.7$		

$$D.M = \frac{\sum f \cdot \lvert (Xm - \bar{X}) \rvert}{N} = \frac{173.7}{100} = 1.737$$

3.4. Desviación Típica o Estándar (S).

La desviación típica o estándar se puede definir de varias formas, una de ellas es que la Desviación típica es la raíz cuadrada de la varianza, como la varianza es S^2, queda que la raíz cuadrada de una cuadrado es S. Otra forma de definir la desviación típica es la raíz cuadrada del sumatorio de las desviaciones al cuadrado dividido por el número de casos. Se simboliza por S en una muestra y por σ. Esto nos aclara que el cálculo de la desviación típica no tiene sentido sin el de la **varianza**, la cual es la media de las desviaciones al cuadrado de los valores a la media aritmética. Las siguientes fórmulas se utilizan para datos no agrupados y se muestra tanto la desviación típica como la varianza:

$$S^2 = \frac{\sum (X - \bar{X})^2}{N}$$

$$\mathbf{S} = \sqrt{\frac{\boxed{\sum (X - X)^2}}{N}}$$

Pero solo es posible utilizar estas fórmulas cuando son pocos datos y no están agrupados, sino es así, las fórmulas a utilizar son un método que ya hemos visto que se denomina **desviación típica por el método reducido** y atiende a siguiente fórmula:

$$\boxed{\mathbf{S} = i \cdot} \sqrt{\boxed{\sum f x'^2 / N - (\sum f x' / N)^2}}$$

Donde i es la amplitud del intervalo, f la frecuencia, x= el valor de la diferencia entre la supuesta desviación que nosotros pensamos que podrá ser. El ejemplo nos aclarará aún más:

$X_{intervalos}$	Frecuencias	X´	f • X´	f. · x´· x´ = f. · x´^{2}
21	5	5	25	125
20	4	4	16	64
19	5	3	15	45
18	9	2	18	36
17	10	1	10	10
16	24	0	0	0
15	16	-1	-16	16
14	16	-2	-32	64
13	5	-3	-15	45
12	4	-4	-16	64
11	2	-5	-10	50
N = 1f = 100		∑f • X´^{2} = 519 ∑f . X´ = - 5		

Si utilizamos la fórmula para hallar la desviación típica con datos agrupados de nuestra distribución obtenemos los siguientes resultados:

$$\boxed{\mathbf{S} = i \cdot} \sqrt{\boxed{\sum f x'^2 / N - \left(\sum f x' / N \right)^2}}$$

$$\mathbf{S} = 1 \cdot \sqrt{519/ 100 - (- 5 / 100)^2} = 1. \sqrt{5.1875} = 2.28$$

Por tanto la varianza debe ser la siguiente:

$$\textit{Varianza} = S^2 = 2.28^2 = 5.1875$$

3.5. Coeficiente de variabilidad.

Este coeficiente trata de expresar la variación que hay de los casos respecto de la media, es decir, relaciona o expresa l relación entre el estadístico de variabilidad y el promedio. Para calcularlo debemos de disponer del 0 absoluto, es decir, que se produzca una constancia en las razones. Este estadístico se calcularía utilizando escalas de razones, las cuales pueden ser nominales y ordinales. Se expresa por la fórmula:

$$CV = \frac{S}{\bar{X}} \cdot 100$$

En nuestro caso el coeficiente de variabilidad es el siguiente:

$$CV = \frac{S}{\bar{X}} \cdot 100 = \frac{2.28}{15.95} \cdot 100 = 14.30\ \%$$

3.6. Características, propiedades y valoración de estos estadísticos.

Para establecer los criterios de normalidad se puede decir que una distribución es apropiada siempre y cuando el coeficiente de variabilidad no es superior al 30 %. En nuestro caso, se puede observar que la distribución cumple este requisito, por lo que se puede considerar una curva de distribución normal. Para establecer los valores de normalidad, es necesario, sumar y restar a la media un valor de la desviación típica, entre los valores obtenidos deberían estar comprendidos el 68,26% de los casos. Al sumar y restar a la media do valores de la desviación típica, entre los valores obtenidos deberían encontrarse el 95 % de los casos. Al sumar y restar a la media 2'5 valores de la desviación típica, entre los valores obtenidos deberían estar comprendidos el 99% de los casos.

A continuación vamos ha hacer un resumen de las características y propiedades de los estadísticos de dispersión.

∗ ESTADÍSTICOS DE RECORRIDO:

⇒ Amplitud total: Influye en los valores extremos de la distribución, su fórmula es la siguiente:

$$A_T = (X_s - X_i) + 1$$

⇒ Desviación intercuartil: Los cuantiles son puntos que determinan diferentes proporciones de área. Los más significativos son los Cuartiles que dividen la muestra en cuatro partes, de las que se desprenden 3 puntos intercuartiles:

- Q_1 es el cuartil que deja por debajo de sí el 25 % de los casos.
- Q_2 es el cuartil que deja por debajo de sí el 50 % de los casos.
- Q_3 es el cuartil que deja por debajo de sí el 75 % de los casos.

Ver puntos anteriores para recordar los estadísticos de recorrido.

4.1. Cálculo de la asimetría mediante coeficientes de sesgo de la variable.

Cálculo de los coeficientes de sesgo según diferentes fórmulas y autores:

⇒ *Pearson*: La fórmula que utiliza es la siguiente:

$$S' = \frac{\bar{X} - Mo}{S}$$

$$S' = \frac{\bar{X} - Mo}{S} = \frac{15{,}95 - 15{,}47}{2{,}28} = 0{,}21$$

Utilizando la segunda fórmula de Pearson, obtenemos resultados similares:

$$S'' = \frac{3\,(\bar{X} - Mdn)}{S} = \frac{3\,(15{,}95 - 15{,}79)}{2{,}28} = 0{,}21$$

⇒ Bowley : Utiliza la siguiente fórmula, que aplicada a nuestros datos se obtienen los siguientes resultados:

$$S''' = \frac{Q_3 - 2\,Mdn + Q_1}{Q_3 - Q_1} = \frac{17{,}3 - 2 \cdot 15{,}79 + 14{,}375}{2{,}925} = 0{,}0325$$

⇒ Kelly: Es necesario encontrar unos puntos antes de realizar los cálculos. La fórmula que utiliza es la siguiente:

$$Sk = \frac{P_{90} + P_{10}}{2} - P_{50}$$

Para hallar estos puntos vamos a realizar el siguiente **proceso**:

1. Hallar el P_{90} y el P_{10} , para lo que utilizaremos la siguiente fórmula:

$$P_n = \frac{n \cdot N}{100}$$

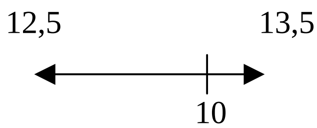
Como nuestra muestra es de 100 sujetos, es decir N = 100, el P_{90} = 90 y el P_{10} =10.

2. Buscar en la frecuencia acumulada el valor 90 y el 10.

X	**f**	**facumulada**
21	5	100
20	4	95
19	5	91⇐
18	9	86
17	10	77
16	24	67
15	16	43
14	16	27
13	5	11⇐
12	4	6
11	2	2

Para hallar el P_{10}, seguimos el siguiente razonamiento:

I = 1



Si el intervalo 13 que tiene una amplitud de uno (i=1), está ocupado por 5 sujetos (frecuencia) 4 sujetos ocuparán X. El 4 sale de restar el n, es decir 10 a la Frecuencia acumulada anterior.

1 --------------------- 5 (frecuencia del intervalo)
X -------------------- 10 - 6 (el punto - faa)

X =4/5= 0,8

Como el valor real del límite inferior del intervalo es 12,5, si a esto le sumamos el 0'8, obtenemos un P_{10} = 13,3.

Para hallar el P_{90}, lo haremos de forma rápida:

1 ----------------------- 5 (frecuencia del intervalo)
X ----------------------- 90 - 86 (el punto - faa)

La X en este caso es 0,8, como el límite inferior del intervalo es 18,5, obtenemos que el punto P_{90} = 19,3.

También conocemos el P_{50}, ya que este punto coincide con Q_2 y con la mediana, por lo que su valor es: = 15,79.

Si sustituimos estos valores en la fórmula de Kelly se obtienen los siguientes resultados:

$$Sk= \frac{P_{90} + P_{10}}{2} - P_{50} = \frac{19,3 + 13,3}{2} - 15,79 = 0,51$$

Todos los coeficientes de sesgo que hemos calculado nos dan un resultado muy similar. Todos están en torno del -0'03. Sabiendo que se permite como normal un coeficiente de sesgo de más o menos 0'350, debemos concluir que en este parámetro se videncia que nuestra distribución es normal. El signo negativo o positivo muestra una cierta tendencia a la asimetría positiva o negativa, que es de la siguiente forma:

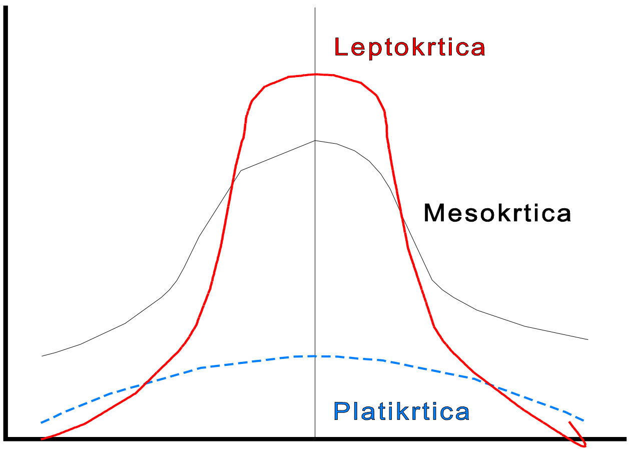
La cola más larga se desplaza hacia la izquierda

La cola más larga se desplaza a la derecha

En nuestro caso, como los valores entre media, mediana y moda son tan similares, no se aprecian desviaciones significativas, es por estoque en algunas formular el resultado nos da positivo y en otras negativo. De cualquier forma a la vista de la gráfica que se obseva en el siguiente apartado, debemos decantarnos por la asimetría positiva.

4.2. Cálculo de la Kurtosis, mediante los coeficiente de apuntamiento.

La curtosis nos habla de la forma de la curva, diciendonos si está más o menos apuntada, es decir, si es más o menos vertical. A la distribución normal se la denomina, distribución mesokúrtica y para que esto ocurra, el valor resultante de la kurtosis debe ser alrededor de 0'263, de no ser así ocurre que la curva no es mesokúrtica sino:



∗ Si el valor obtenido del cálculo de la kurtosis es > de 0'263 la curva es platikúrtica, su forma es achatada, es decir predomina el eje horizontal.

* Si el valor obtenido es < de 0'263 la curva es leptokúrtica, es decir tiene una forma alta y delgada.

La kurtosis se hala a través de una fórmula que exponemos a continuación, de cuyo resultado debemos obtener el dato para compararlo con estas referencias. Si aplicamos la fórmula a nuestros datos, el resultado es el siguiente:

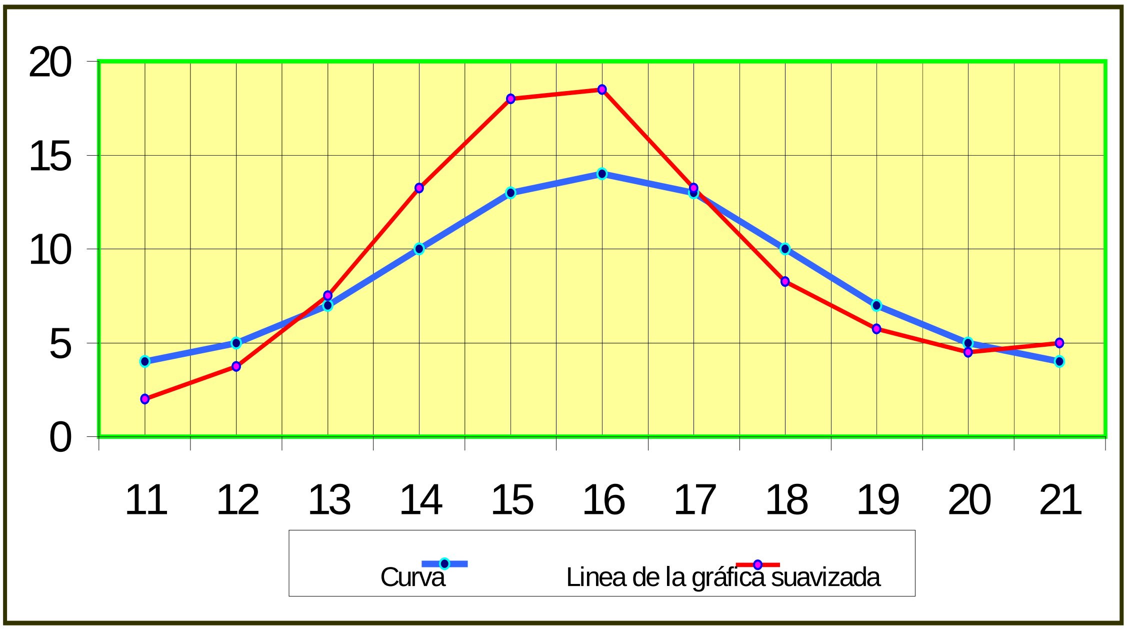
$$Ku = \frac{Q_3 - Q_1}{2\,(P_{90} - P_{10})} = \frac{17{,}3 - 14{,}375}{2\,(19{,}3 - 13{,}3)} = 0{,}244$$

De los datos que obtengamos, no tendremos en cuenta el signo (aunque casi siempre nos saldrá positivo), si obtenemos algún valor negativo lo consideraremos en su valor absoluto.

De los datos obtenidos en nuestro caso particular, podemos deducir que nuestra curva es leptokúrtica.

4.3. Determinar gráfica y aritméticamente el tipo de curva de la distribución de la variable asignada.

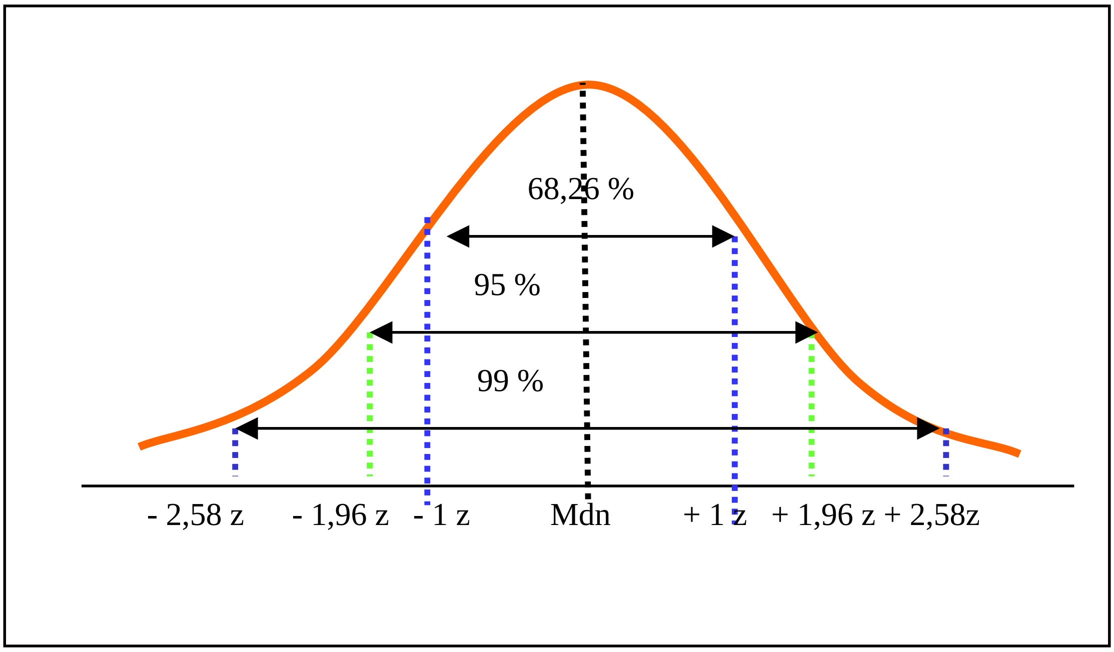
La línea en azul muestra que nuestra gráfica está ligeramente desplazada hacia la izquierda , es decir muestra una cierta asimetría positiva. De igual forma si nos fijamos en los puntos rojos, que hacen referencia a los puntos sin suavizar, podemos observar que la distribución es claramente leptokúrtica, aunque si es cierto que se ajusta bastante bien a la curva de normalidad.



En nuestro caso los valores de la media y moda son muy similares, es decir, se diferencian en dos centésimas, por lo que la distribución parece bastante normal. De otra parte los cálculos de los coeficientes de sesgo nos hacen seguir considerando nuestra distribución como normal, ya que los valores obtenidos son muy inferiores a 0'360 y en la mayoría de los casos positivos, con lo que se ratifica nuestra apreciación.

5.1. Manejo de las tablas de áreas de la curva normal.

Esta gráfica pretende ilustrar, los % de casos a los que se hace referencia cuando se utiliza una desviación típica o una Z, cuando se utilizan casi 2 (1'96)y el caso donde más fiabilidad se tiene, en el caso de utilizar 2'58 z, con lo que se obtiene una representatividad del 99 % de los sujetos. Esto recibe el nombre de confianza o en su lado contrario, margen de error.



Las tablas de puntuación típica determinan las áreas de curva normal desde la ordenada 0 (la media) hasta el punto que se quiera, medido en una escala de puntuaciones típicas. Así pues transformaremos las puntuaciones directas, en puntuaciones típicas, a través de una fórmula de conversión y llevaremos el resultado obtenido a la tabla de áreas para conocer el porcentaje de los casos.

5.2. Transformación de las puntuaciones directas en diferenciales y típicas.

La transformación en puntuaciones diferenciales se realiza a través de la siguiente fórmula, y en la siguiente tabla:

$$X = \text{Xm} - \bar{X}$$

X intervalos	Xm	X= Xm - X
21	21	5,05
20	20	4,05
19	19	3,05
18	18	2,05
17	17	1,05
16	16	0,05
15	15	-0,95
14	14	-1,95
13	13	-2,95
12	12	-3,95
11	11	-4,95

Ahora pretendemos realizar la transformación a puntuaciones típicas, para lo que utilizamos la siguiente fórmula:

$$Z = \frac{Xm - \bar{X}}{S}$$

Xm	X= Xm - X		Z
21	5,05	Z1	2,215
20	4,05	Z2	1,776
19	3,05	Z3	1,338
18	2,05	Z4	0,899
17	1,05	Z5	0,461
16	0,05	Z6	0,022
15	-0,95	Z7	-0,417
14	-1,95	Z8	-0,885
13	-2,95	Z9	-1,294
12	-3,95	Z10	-1,732
11	-4,95	Z11	-2,171

5.3. Juicio valorativo sobre la normalidad de la distribución teniendo en cuenta los estadísticos y coeficientes calculados, así como la variabilidad de los datos.

Es interesante volver a repasar las valoraciones que se introdujeron sobre la moda, mediana y media en los apartados 2.4 y 4.3.

⇒ Los cuartiles: Son tres valores de la variable que dividen la distribución en 4 partes iguales. Cada uno de ellos será el 25 % del total. Siempre que la distribución sea simétrica. Las distancias entre los cuartiles serán iguales. En nuestro caso, las distancias se muestran en las gráfica siguiente. Pero lo que nos ha movido a realizar esta cantidad de cálculo es única y exclusivamente halar la amplitud intercuartil, la cual se halla restando $Q_3 - Q_1$:

$$Q_3 - Q_1 = 17{,}3 - 14{,}375 = 2{,}925$$

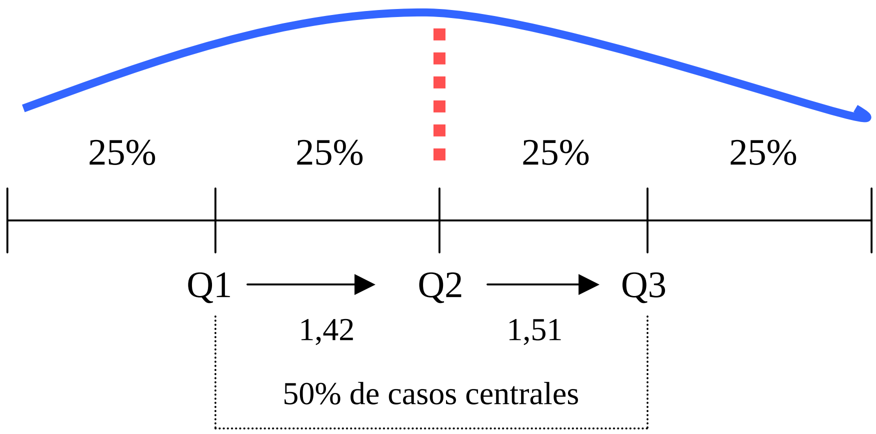
Este resultado tiene que hacernos pensar, encontramos que en menos de tres valores se encuentran el 50 % de los valores de la distribución.

⇒ Amplitud intercuartil: El siguiente proceso va encaminado a estimar si podemos decir que hay una simetría de la distribución en la amplitud intercuartil con respecto a la ordenada cero, es decir como oscila la gráfica en este tramo central, para ello debemos encontrar primero la amplitud semi-intercuartil, la cual se representa por Q y su fórmula es la siguiente:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{2{,}925}{2} = 1{,}4625$$

Para saber si existe simetría en la distribución comparamos la amplitud semi-intercuartil con las diferencias entre los cuartiles, si son similares la distribución del 50 % de los casos centrales será simétrica, sino es similar la distribución estará apuntada, es decir, tendrá curtosis.

Si observamos la siguiente figura, donde se han hallado las diferencias entre Q_3 - Q_2 y entre Q_2 -Q_1, se puede observar los resultados de la simetría. Aunque la diferencia es mínima (9 centésimas) se puede decir que la distribución en sus 50 % de casos centrales es prácticamente simétrica, aunque como se puede observar en los resultados hay una cierta acumulación de datos entre Q_3 -Q_2, ya que en menos espacio hay el mismo n1 de casos que entre Q_2 -Q_1.



⇒ **La desviación media:** Si a la media le sumo y le resto la Dm, deberíamos obtener el 58 % de los casos, si los sujetos están distribuidos normalmente.

En nuestros datos la media es 15.95, por lo que quedaría de la siguiente forma:

15.95 + 1Dm (1'737) = 17.687

Entre estos dos valores deben estar 58% casos por ser N=100

15.95 - 1Dm (1'737) = 14.213

Sabemos que el intervalo en el que se encuentra el valor 17.687 tiene una frecuencia de 9 individuos. Si una unidad está ocupada por 9 individuos, la diferencia entre el valor obtenido y el límite inferior del intervalo obtendrá X:

1 ------------------ 9

17.687 – 17.5 ------- X = 1.683

Para el valor inferior realizamos el mismo proceso pero restamos al límite superior (14.5), el valor de 14.213 y los resultados son:

1 ----------------- 16

14,5-14,213 ---------- X= 4.592

Como lo que pretendemos es hallar el número de sujetos que se encuentran entre estos dos intervalos, tenemos que sumar las frecuencias de los intervalos que están en medio. El resultado es 50 individuos.

⇒ La desviación típica: Está relacionada con la varianza y se denota por s. Se define como la raíz cuadrada de la varianza. Es de gran utilidad en distribuciones en las que se conocen los porcentajes aproximados de puntuaciones situadas a una, dos, tres o más desviaciones típicas, respecto de la media.

Caso 1: Sumamos y restamos a la media una desviación típica (2.28) y debemos obtener que el 68 % de los casos están comprendidos entre +- 1 s.

Media + 1s = 15.95 + 2.28 = 18.23
Media - 1 s = 15.95 – 2.28 = 13.67

El intervalo en el que se encuentra el punto 18.23 es el intervalo 18, en el que hay una frecuencia de 9 individuos y la amplitud del intervalo es 1 unidad. Realizamos la siguiente regla de tres. Si en una unidad caben 9 individuos, cuantos cabrán en la diferencia entre el valor 18.23 y el límite inferior del intervalo (17.5).

1---------------------- 9

18.23 – 17.5 --------------X = 6.57

El otro cálculo lo realizamos de la misma pero, restando al límite superior del intervalo el valor que hemos obtenido:

1------------------- 16

14.5 – 13.67 ----------------- X = 13.28

Como se observa hemos calculado los sujetos que se encuentran en los extremos superior e inferior.

La suma total de individuos que se encuentran en esos intervalos es, del intervalo 17 entran 6'57 y del intervalo inferior 14 solo 13.28 de 16 posibles a lo que hay que sumar los 50 individuos anteriormente hallados:

Total: 6,57 + 13.28 + 50 = 69,85

Esto denota lo que veníamos diciendo de la gran normalidad de nuestra distribución. Es decir en +- 1 s donde deberían estar el 68 % de los sujetos, nosotros tenemos 69,85 sujetos.

Caso 2: Se utiliza los mismos procedimientos anteriores pero con +- 2 s, y debe cumplirse si la distribución es normal, que se encuentre el 95 % de los casos.

Media + 2s = 15,95 + 2.28 x 2 = 20,51
Media - 2s = 15,95 – 2.28 x 2 = 11.39

Si una unidad está ocupada por 5 individuos, cuanto individuos ocuparán la diferencia entre el valor y el límite inferior del intervalo:

1-----------------5

20,51- 20,50 ---------- X = 0,05

El límite inferior es algo similar pero como siempre en vez de restar el valor al límite inferior, restamos al valor superior el que nosotros hemos obtenido:

1---------------- 2

11.5 – 11,39 --------------X = 0,22

La suma total por orden será la suma de los individuos contenidos entre los extremos de los intervalos, es decir:

Total: 0,22 + 50 + 0,05 + 18 de (9 + 5 + 4) + 25 de (16 + 5 + 4) = 93.27.

También éste es un signo de normalidad debido a que lo que deberíamos haber obtenido es un 95 %.

Caso 3: Este es el último caso que debemos comprobar, con +- 3 s. Si la distribución es normal, es necesario que se encuentre el 99% de los casos.

Media + 3s = 15,95 + 2,28 x 3 = 22.79
Media - 3s = 15,95 – 2,28 x 3 = 9.11

Si una unidad está ocupada por 0 individuos, cuanto individuos ocuparán la diferencia entre el valor y el límite inferior del intervalo:

1------------------ 0

22.79 – *-------------------- X = 0

El límite inferior es algo similar pero como siempre en vez de restar el valor al límite inferior, restamos al valor superior el que nosotros hemos obtenido:

1 -------------------- 0

9.11 - * ---------------------X = 0

La suma total por orden será la suma de los individuos contenidos entre los extremos de los intervalos, es decir:

Total: 0 + 0 + 25 + 18 + 50 + 5 + 2 = 100 individuos.

También éste es un signo de normalidad debido a que lo que deberíamos haber obtenido es un 99 %.

Aunque los valores dan muy próximo a los valores teóricos debemos concluir que pensamos que nuestra gráfica presenta una asimetría positiva, pero muy ligera.

⇒ <u>Coeficiente de Variación</u>: Es el último parámetro que nos queda para analizar la normalidad de nuestra distribución. Es el cociente entre la desviación típica y la media aritmética multiplicado por 100. Permite comparar dos desviaciones típicas de muestras distintas o de variables distintas. Es un tipo de variabilidad relativa. Mide el grado de representatividad de la media aritmética, cuanto mayor es el C.V menos representativa es la media ya que aumenta la dispersión de los datos. En nuestro caso la C.V es la siguiente:

$$CV = \frac{S}{\bar{X}} \cdot 100 = \frac{2.28}{15.95} \cdot 100 = 14.30\ \%$$

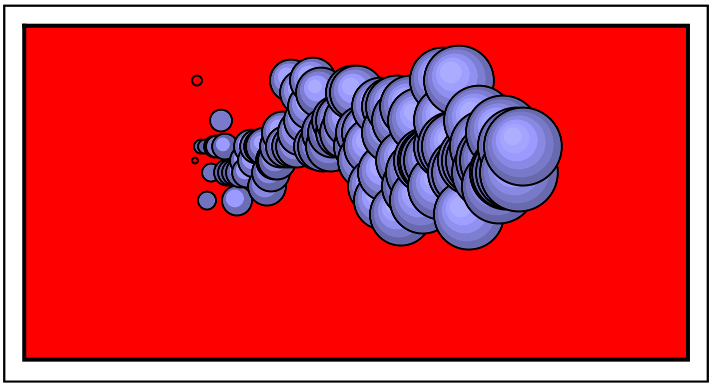
Esta es la prueba definitiva en la argumentación de la normalidad de nuestra distribución de datos. El C.V es muy inferior al 30% permitido como normal. Aunque como dijimos tiene una ligera asimetría positiva.

6.1. Distribución conjunta de dos variables. Diagrama de dispersión.

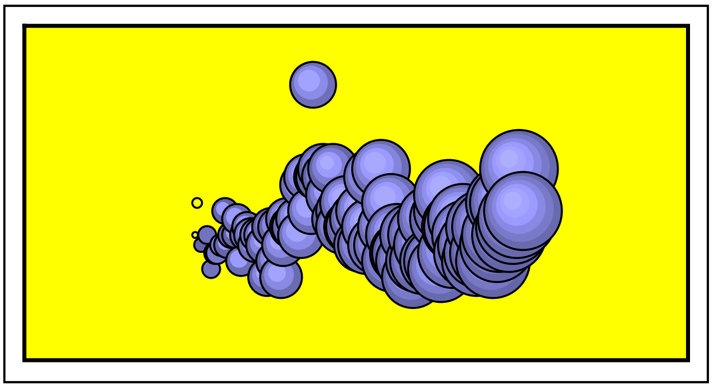
ACOPIO DE DATOS DE LAS DOS VARIABLES 200 Datos																			
1 B	2 B	1 B	2 B	1 B	2 B	1 B	2 B	1 B	2 B	1 B	2 B	1 B	2 B	1 B	2 B	1 B	2 B	1 B	2 B
15	15	14	14	16	15	16	16	17	22	16	14	17	12	15	14	21	16	15	12
21	19	14	15	16	14	16	17	16	20	17	18	19	15	15	17	14	17	15	14
16	14	14	15	13	10	17	15	17	23	15	14	11	13	15	12	14	15	13	15
16	15	12	17	14	16	20	21	18	17	17	16	18	13	15	12	11	13	17	19
12	15	14	12	15	12	18	22	17	18	16	22	15	12	13	11	15	14	14	19
14	11	15	15	15	16	19	18	18	16	13	14	19	15	21	17	15	12	14	15
16	13	14	16	16	10	21	33	17	19	19	23	13	10	16	18	18	13	14	16
16	13	16	15	17	14	16	22	18	16	12	14	14	12	18	20	15	16	16	17
18	14	15	14	16	17	20	22	20	16	14	16	18	15	16	13	16	13	14	23
16	18	16	15	21	16	16	23	20	17	19	19	12	12	16	16	14	16	16	18

Con la siguiente gráfica pretendemos dar una idea de la relación que existe entre las dos variables, esto se consigue con una gráfica de dispersión, cómo la que mostramos:

VARIABLE 1B



VARIABLE 2B



6.2. Cálculo del índice de correlación por el método de Bravais - Pearson, con los datos de nuestra variable y los del alumno contiguo, es decir Variables 1B y 2B.

De cada uno de los factores de la siguiente fórmula se obtendrá un resultado que después aplicaremos a ella para obtener el resultado de la correlación.

$$\mathbf{r}_{xy} = \frac{\dfrac{\sum f\,x'\cdot y'}{N} - \dfrac{\sum f\,x'}{N} - \dfrac{\sum f\,y'}{N}}{\sqrt{\left\{\dfrac{\sum f\,x'^2}{N} - \left(\dfrac{\sum f\,x'}{N}\right)^2\right\}\cdot\left\{\dfrac{\sum f\,y'^2}{N} - \left(\dfrac{\sum f\,y'}{N}\right)^2\right\}}}$$

Iremos hallando uno por uno los factores que son necesario para sustituirlos al final en esta fórmula. Los datos se obtienen de la tabla de registro de datos que hemos adosado en la página anterior, de la que tras un más que complejo procedimiento manual hemos obtenido los siguientes datos:

$$\Rightarrow \frac{\mathbf{\sum fx'\cdot y'}}{\mathbf{N}} = \frac{135}{100} = 1{,}35$$

$$\Rightarrow \frac{\sum \mathbf{fx'}}{\mathbf{N}} = \frac{-140}{100} = -1{,}40$$

$$\Rightarrow \frac{\sum \mathbf{f\,y'}}{\mathbf{N}} = \frac{-5}{100} = -0{,}05$$

Ahora pasamos a realizar los cálculos del divisor de la ecuación:

$$\Rightarrow \mathbf{(\sum fx' / N)^2} = (-140 / 100)^2 = 1{,}96$$

$$\Rightarrow \mathbf{\sum fx'^2 / N} = 322 / 100 = 3{,}22$$

$$\Rightarrow \mathbf{(\sum fy' / N)^2} = (-5 / 100)^2 = 0{,}0025$$

$$\Rightarrow \mathbf{\sum fy'^2 / N} = 519 / 100 = 5{,}19$$

Sustituyendo todos estos valores en la fórmula inicial obtenemos los siguientes resultados:

$$\mathbf{r}_{xy} = \frac{1{,}35 - (-1{,}40) \cdot (-0{,}005)}{\sqrt{\ (3{,}22 - 1{,}96) \cdot (5{,}19 - 0{,}005)}}$$

$$\mathbf{r}_{xy} = 0{,}525431807$$

6.3. Valoración de la significación e interpretación de la correlación obtenida.

Para valorar la significación de la correlación debemos procurarle antes un proceso matemático para hacerla entendible, es decir la valoramos en porcentaje, elevando el valor de r_{xy} y multiplicando por 100. Si realizamos esta operación con nuestros valores obtenemos los siguiente resultados:

$$\mathbf{r}_{xy} = 0{,}525431807 = (0{,}525431807)^2 \cdot 100 = 27{,}60\ \%$$

Nuestras dos variables tienen una varianza común del 27,60 %, esto quiere decir que la relación de la una con la otra es así. Si nuestras dos variables fueran la talla y el peso, querría decir que la relación que pueden tener la talla y el peso es del 28 %. Pero para la interpretación hemos visto en clase la siguiente clasificación:

ϕ Correlaciones entre **0 y 0'2**: No hay correlación o ha de considerarse **nula**.

ϕ Correlaciones entre **0'2 y 0'4**: Correlación **baja**, ya que la relación efectiva entre la dos variables no supera en ningún caso el 16 %.

ϕ Correlaciones entre **0'4 y 0'7**: Este es nuestro caso y son correlaciones **moderadas**, pero significativas ya que se pueden obtener valores máximos de varianza común del 50 %. En las correlaciones de los test, 0'5 es lo mínimo imprescindible para que éste tenga validez.

ϕ Correlaciones entre **0'7 y 0'9**: La correlación es **alta** y puede llegar al 81 % de varianza común.

ϕ Correlaciones de más de **0'9**: La correlación es **muy alta** ya que son valores por encima del 81 % de varianza común. La herramienta de medida en los test debe tener una fiabilidad de 0'9 como mínimo.

Nuestras variables presentan entonces, un 28 % de varianza común y un 72 % de independencia, lo que según la escala anterior debe presentarse como una correlación moderada.

8.1. Elaboración de escalas en normas ordinales: Deciles y centiles.

8.2. Elaboración de escalas en normas típicas: puntuaciones z, puntuaciones D, puntuaciones T, puntuaciones normalizadas, eneatipos.

A partir de los datos de los intervalos, de las frecuencias y de las frecuencias acumuladas obtenemos las puntuaciones típicas y las típicas derivadas como muestra la siguiente tabla. La X tomada es el valor inferior del intervalo.

			Puntuaciones típicas			**Típicas derivadas**		
Intervalos	**f**	**fa**	**X**	**Z**		**X**	**D**	
21	5	100	21	Z_{11}	2,215	21	D_{11}	94,3
20	4	95	20	Z_{10}	1,776	20	D_{10}	85,52
19	5	91	19	Z_9	1,338	19	D_9	76,76
18	9	86	18	Z_8	0,899	18	D_8	67,98
17	10	77	17	Z_7	0,461	17	D_7	59,22
16	24	67	16	Z_6	0,022	16	D_6	50,44
15	16	43	15	Z_5	-0,417	15	D_5	41,66
14	16	27	14	Z_4	-0,885	14	D_4	32,3
13	5	11	13	Z_3	-1,294	13	D_3	24,12
12	4	6	12	Z_2	-1,732	12	D_2	15,36
11	2	2	11	Z_1	-2,171	11	D_1	6,58
Fórmulas para operar				$Z=\frac{X-\bar{X}}{S}$			D=50+20Z	

A continuación pasamos a calcular. Los percentiles. Se utiliza una fórmula bastantes compleja que exponemos a continuación y que hemos desglosado en términos en la tabla siguiente para facilitar los cálculos:

PERCENTILES	**Porcentaje de la frecuencia acumulada**	**Límite inferior**	**más**	**Amplitud / frecuencia del intervalo**	**Por**	**Frecuencia acumulada del decil - frecuencia acumulada del intervalo anterior**	**Total**
Pc	Pfa	L_{inf}	+	i/fi	X	fac-fa i.a.	Redondeado
99	99	19,5	+	¼	X	99-91	21,5= 22
95	95	19,5	+	¼	X	95-91	19,5=20
90	90	17,5	+	1/9	X	90- 77	18,94=19
80	80	16,5	+	1/10	X	80-67	17,8= 18
75	75	15,5	+	1/24	X	75-43	16,83=17
70	70	15,5	+	1/24	X	70-43	16,66=17
60	60	14,5	+	1/16	X	60-27	16,56=17
50	50	14,5	+	1/16	X	50-27	15,93=16
40	40	13,5	+	1/16	X	40-11	15,31=15
30	30	13,5	+	1/16	X	30-11	14,68=15
25	25	12,5	+	1/5	X	25-6	16,3=16
20	20	12,5	+	1/5	X	20-6	15,3=15
10	10	11,5	+	¼	X	10-2	13,5=14
5	5	10,5	+	½	X	0-0	11
1	1	10,5	+	½	X	0-0	11
Pfa = Pc·N/100		Pc= L inf + i/fi (fac – fa int.ant)					