T. 3 – Estadísticos de posición grupal

1. **Estadísticos de tendencia central**
    - **1.1. Variables categóricas: la moda**
    - **1.2. Variables ordinales: la mediana**
    - **1.3. Variables cuantitativas: la media y sus alternativas robustas**
2. **Otros estadísticos de posición grupal**
    - **2.1. El mínimo y el máximo**
    - **2.2. Los cuantiles**

• Tanto en este tema como en los dos siguientes se van a revisar una serie de <u>índices estadísticos</u> que nos van a permitir resumir la información contenida en la distribución de frecuencias de una variable. Estos índices proporcionan valores numéricos que, de forma sintética, reflejan diferentes características de una distribución de frecuencias, tales como su posición, su variabilidad y su asimetría.

• Este tema, en concreto, se centra en una serie de índices estadísticos que permiten describir numéricamente cuál es la localización o <u>posición</u> de la distribución de frecuencias de una variable respecto a la escala de medida utilizada en la recogida de datos de esa variable.

• Se les denomina aquí con el calificativo de <u>grupal</u> porque sintetizan la información de todos los datos de una variable, esto es, para todo el grupo de casos a partir de los que esos datos han sido recogidos. Frente a éstos, en el tema 6 se tratarán algunos procedimientos orientados a obtener información sobre la posición de un caso <u>individual</u> en relación al resto de individuos dentro de su grupo.

1. Estadísticos de tendencia central

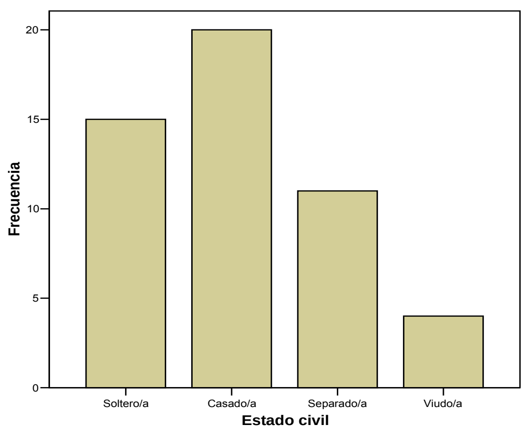
• Los índices que se van a revisar en este apartado son indicativos de la tendencia central de los datos y, por tanto, proporcionan un valor que expresa la posición entorno a la que se *centra* la distribución de frecuencias de una variable, esto es, un valor que va a ejercer de *representante* de todos los datos recogidos para esa variable.

1.1. Variables categóricas: la moda

• La moda de una variable *X* (Mo_X) es un estadístico de tendencia central que se obtiene como el valor que más se repite en el conjunto de datos correspondiente a una variable, esto es, el que tiene la frecuencia absoluta (n_i) más alta en su distribución de frecuencias.

• La moda, aparte de en variables categóricas, se puede obtener también en variables ordinales y cuantitativas, y lo mismo se puede decir de otros índices estadísticos que serán presentados para variables categóricas en temas sucesivos. En cualquier caso, destacar también que no son los que mejor captan la información contenida en las variables ordinales y cuantitativas.

Ejemplo de cálculo de la moda para los datos de la variable "Estado civil" obtenidos en una muestra de 50 personas de la ciudad de Castellón (n = 50):
Dado que el valor que más se repite es "Casado/a" (n_i = 20), éste sería la moda de esta variable para los datos considerados.

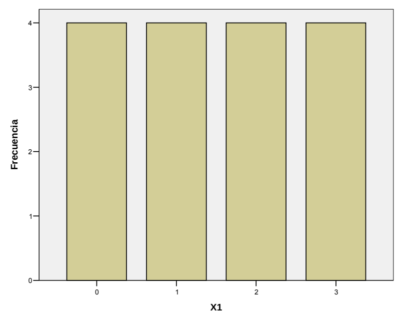


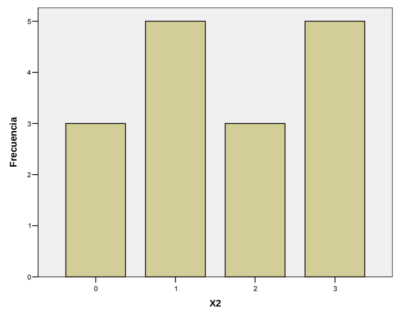
Ejemplo de obtención de la moda en algunos casos particulares: sean los datos de la variable "Lugar de nacimiento" [0: Fuera de la CV; 1: Prov. de Alicante; 2: Prov. de Valencia; 3: Prov. de Castellón] en dos muestras distintas de 16 sujetos residentes en la Comunidad Valenciana:

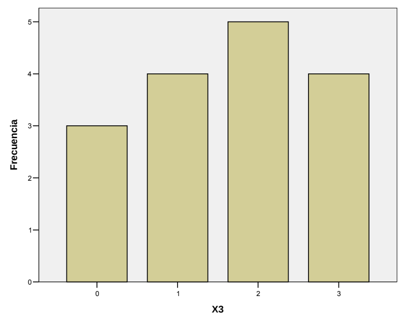
X1: {0, 0, 1, 2, 1, 0, 1, 3, 2, 3, 1, 3, 0, 2, 3, 2 => $Mo_{X1} = \varnothing$ => Distribución amodal

X2: {1, 2, 0, 3, 1, 1, 0, 3, 3, 3, 1, 2, 0, 2, 1, 3} => Mo_{X2} = 1 y 3 => Distribución bimodal

• En base al número de modas de una distribución de frecuencias, se ha derivado una clasificación de las mismas en que diferencian los siguientes tipos: amodal o uniforme, unimodal, bimodal y multimodal. Los siguientes diagramas de barras muestran tres de estos tipos:







1.2. Variables ordinales: la mediana.

• Análogamente a lo que se señalaba para las variables categóricas, la mediana puede ser aplicada también en variables cuantitativas, si bien, no es el índice estadístico que mejor resume la tendencia central de este tipo de variables.

• La mediana de una variable *X* (Mdn_X) es el valor del dato que, tras ordenar todos los datos de la variable, deja al mismo número de ellos por debajo que por encima. En el caso de un número par de datos, se obtiene el promedio de los valores de los datos centrales.

Ejemplo:

X3: {6, 8, 1, 4, 2, 5, 6} => $Mdn_{X3} = Mdn$ {1,2,4,5,6,6,8} = 5

X4: {9, 6, 8, 1, 4, 2, 5, 6} => $Mdn_{X4} = Mdn$ {1,2,4,5,6,6,8,9} = (5+6) / 2 = 5,5

Ejemplo de obtención de la mediana de una variable con los datos recogidos con la siguiente pregunta de un test: “Ansiedad que siente cuando se encuentra con mucha gente alrededor”. Las alternativas de respuesta a esta cuestión eran: 1: Nada; 2: Algo; 3: Bastante; 4: Mucha.

X_i	Frec. absoluta (n_i)	Porcentaje ($\%_i$)	Frec. absoluta acumulada (n_a)	Porcentaje acumulado ($\%_a$)
1	23	19,0	23	19
2	36	29,7	59	48,7
3	47	38,9	106	87,6
4	15	12,4	121	100
	121	100		

La mediana sería el valor del dato que, tras ordenar los 121 datos, ocupe la posición central, en este caso, la posición 61. En consecuencia, la mediana en este ejemplo será igual a 3 (Bastante).

• En el caso de contar con una distribución de frecuencias, lo más sencillo, antes que desagregar la distribución de frecuencias para ver cuál es el valor central, es fijarse en la columna de porcentajes acumulados: la mediana será el valor de la variable cuyo $\%_a$ sea igual a 50 o, en su caso, el superior a 50 más pequeño. Así, en el ejemplo anterior de la pregunta de un test, la mediana sería igual a 3 (Bastante), ya que su $\%_a$ (87,6) es el superior a 50 más bajo.

1.3. Variables cuantitativas: la media y sus alternativas robustas

1.3.1. La media.

• La media [aritmética] o promedio de una variable X ($\bar{X}$ o μ_X) es un índice estadístico de tendencia central consistente en sumar los valores de los datos de esa variable y dividir por el número de ellos:

$$\bar{X} = \frac{\sum X_i}{n}$$

Ejemplo: si X:{2, 3, 2, 7}, entonces $\bar{X}$ = (2+3+2+7) / 4 = 3,5

• Si tenemos los datos en forma de distribución de frecuencias, el cálculo de la media supone sumar el producto de los valores por la correspondiente frecuencia absoluta, y dividir el resultado por el nº de casos:

$$\bar{X} = \frac{\sum X_i \cdot n_i}{n}$$

• Una <u>fórmula alternativa</u> en el cálculo de la media a partir la información de una distribución de frecuencias consiste en sumar el producto de cada valor de la variable por su frecuencia relativa (proporción):

$$\overline{X} = \sum X_i \cdot p_i$$

Ejemplo de la variable "Tiempo empleado en completar un laberinto" por una muestra de 20 ratas (n = 20).

Tiempo (seg.)	n_i	p_i
9	3	0,15
10	8	0,4
11	6	0,3
12	2	0,1
13	1	0,05

$$\overline{X} = \frac{\sum X_i \cdot n_i}{n} = \frac{9\cdot 3 + 10\cdot 8 + 11\cdot 6 + 12\cdot 2 + 13\cdot 1}{20} = 10,5$$

o, también, $\overline{X} = \sum X_i \cdot p_i = 9\cdot 0,15 + 10\cdot 0,4 + 11\cdot 0,3 + 12\cdot 0,1 + 13\cdot 0,05 = 10,5$

• Si lo que tenemos es una distribución de frecuencias con los <u>valores agrupada en intervalos</u>, entonces podemos calcular la media a partir de los valores medios de cada intervalo, esto es, las conocidas como marcas de clase.

Ejemplo de la variable "Peso" a partir de una distribución de frecuencias en que han sido agrupados en intervalos los valores de la variable:

Peso (Kg.)	Marca de clase	n_i	n_a	p_i	p_a
40 – 50	45	5	5	0,086	0,086
50 – 60	55	10	15	0,172	0,258
60 – 70	65	21	36	0,362	0,620
70 – 80	75	11	47	0,190	0,810
80 – 90	85	5	52	0,086	0.896
90 – 100	95	3	55	0,052	0,948
100 – 110	105	3	58	0,052	1

$$\overline{X} = \frac{\sum X_i \cdot n_i}{n} = \frac{45\cdot 5 + 55\cdot 10 + 65\cdot 21 + 75\cdot 11 + 85\cdot 5 + 95\cdot 3 + 105\cdot 3}{58} = 68,8$$

o también:

$$\overline{X} = \sum X_i \cdot p_i = 45 \cdot 0,086 + 55 \cdot 0,172 + 65 \cdot 0,362 + 75 \cdot 0,190 + 85 \cdot 0,086 + 95 \cdot 0,052 + 115 \cdot 0,052 = 68,8$$

Ejercicio 1: Obtener la media, la mediana y la moda de la siguiente distribución de frecuencias de la variable “Nº de hijos/as” obtenida para una muestra de 200 familias.

X_i	n_i
0	40
1	80
2	60
3	20

Ejercicio 2: Inventa un conjunto de 7 datos que tengan media 9, mediana 10 y moda 11.

• Si para una misma variable disponemos de medias obtenidas en diferentes grupos de sujetos, se puede obtener la media que resultaría de juntar los datos de esos diferentes grupos a través de la que se conoce como fórmula de la media ponderada:

$$\overline{X}_T = \frac{n_1 \cdot \overline{X}_1 + n_2 \cdot \overline{X}_2 + \cdots + n_k \cdot \overline{X}_k}{n_1 + n_2 + \cdots + n_k}$$

Ejemplo: Si la nota media en Evaluación Psicológica del grupo A (n = 160) ha sido de 7,8, la del grupo B (n = 110) de 5,7 y la del grupo C (n = 148) de 6,7, ¿cuál sería la nota media si se juntasen los 3 grupos?

$$\overline{X} = \frac{160 \cdot 7,8 + 110 \cdot 5,7 + 148 \cdot 6,7}{418} = 6,86$$

Ejercicio 3: Un grupo de 50 estudiantes de Estadística está compuesto por un 64% de estudiantes de 1ª matrícula, un 20% de 2ª matrícula y un 16% de 3ª matrícula o posterior. Sus medias en la evaluación final de la asignatura fueron 7,2, 6,3 y 5,9, respectivamente. ¿Cuál es la media del grupo total?

1.3.2. Algunas anotaciones sobre la aplicación del estadístico de la media

(1) La media aritmética también es aplicada en la práctica a variables ordinales –hecho que puede resultar cuestionado por razones teóricas sobre las que no vamos a entrar ahora aquí. Por otra parte, la obtención de la media de una variable ordinal supondrá obtener, en no pocas ocasiones, valores que no se encuentren en el rango de valores de la escala de medida de la variable. En cualquier caso, su aplicación es bastante frecuente en la práctica pues tiene alguna ventaja como la que se comentará en el siguiente epígrafe y, por otra parte, el que obtengamos un valor que no coincida con ninguno de los valores originales de la variable puede no representar un inconveniente importante en muchas ocasiones a la hora de comunicar los resultados.

A modo de **ejemplo**, el cálculo de la media sobre la variable ordinal "Ansiedad que siente cuando se encuentra con mucha gente alrededor" presentada en un ejemplo anterior, da como resultado 2,45, que no es ninguno de los valores que podía tomar esa variable (1: Nada; 2: Algo; 3: Bastante; 4: Mucha), sin embargo, es de esperar que ello sea entendido por la mayoría de las audiencias a las que nos podamos dirigir como un nivel de ansiedad a medio camino entre "Algo" y "Bastante".

Como contrapartida a este posible inconveniente, una ventaja de la aplicación de la media sobre las variables ordinales es que ésta es capaz de captar de un modo más preciso la información relativa a la tendencia central de los datos. Ello se debe a que tiene en cuenta a todos los datos, y no sólo el valor del dato que ocupa la posición central, como ocurre con la mediana.

Por **ejemplo**, sean X e Y dos variables ordinales: X:{2, 3, 4, 4, 5, 5, 5, 6, 6} e Y:{2, 3, 4, 4, 5, 5, 5, 6, 7} que, como puede observarse, sólo se diferencian, mínimamente, en un dato. Si se calcula la media y la mediana en ambas variables se observará como la mediana es la misma para las dos (Mdn = 5), no captando esa diferencia entre ambas variables; en cambio, la media sí que es más sensible y capta la diferencia existente entre ambos conjuntos de datos ($\overline{X}$ = 4,44; $\overline{Y}$ = 4,55).

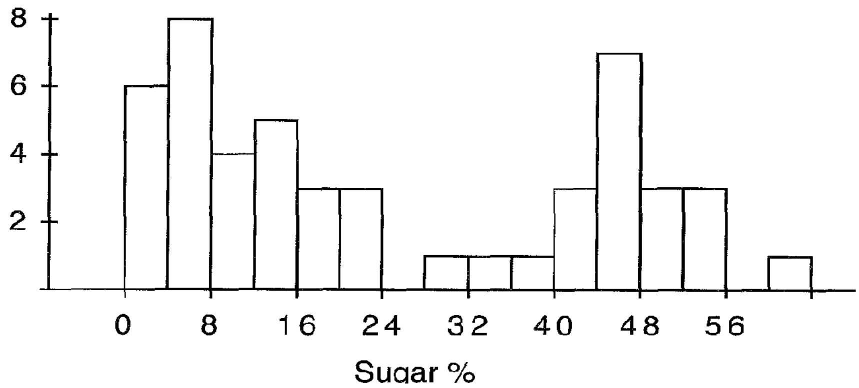
(2) El problema señalado de la media y las variables ordinales de que nos dé un valor que no se encuentre entre los valores posibles de la variable, puede darse también con <u>variables cuantitativas discretas</u>.

Ejemplo: al obtener la media de la variable cuantitativa discreta "Nº de hijos/as" a partir de la distribución de frecuencias que se muestra a continuación da igual a 1,65.
=> ¿es comprensible este valor para presentarlo en un informe de resultados? –En caso afirmativo, utilícese la media; en caso contrario, tal vez mejor utilizar la mediana o la moda.

X_i	**Frec. absoluta (n_i)**	**Frec. relativa (p_i)**	**Porcentaje ($\%_i$)**
0	3	0,15	15
1	6	0,30	30
2	7	0,35	35
3	3	0,15	15
4	1	0,05	5
	20	1	100

(3) Cuando la forma de la distribución de una variable cuantitativa continua sea bimodal o multimodal, la obtención de la media puede no tener mucho sentido -y lo mismo se podría decir de la mediana. En estos casos, la obtención de la moda representa la opción más conveniente.

Ejercicio 4: Dada el histograma de la distribución de la variable "contenido de azúcar en %" medida para diversas marcas de cereales, reconstruir la distribución de frecuencias correspondiente y calcular la media y la mediana de la misma. ¿Son ambos estadísticos un buen resumen de la tendencia central de esta distribución?



1.3.3. Alternativas robustas al cálculo de la media

• La media es muy sensible a datos anómalos, datos cuyos valores son de magnitud bastante diferente a la mayoría de los valores (i.e., distribuciones de frecuencias muy asimétricas).

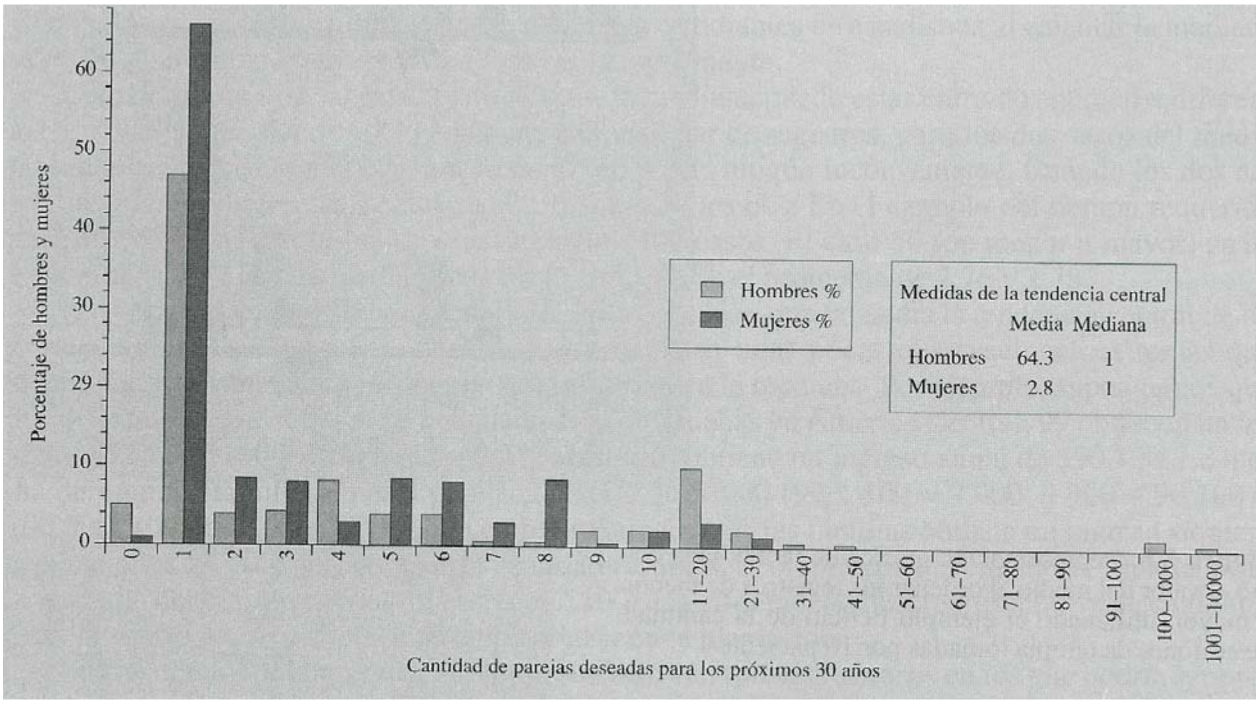
Ejemplo:

$X5$:{8, 8, 9, 10, 10, 12, 14} $\rightarrow$ $\bar{X} = 10{,}14$; $Mdn = 10$

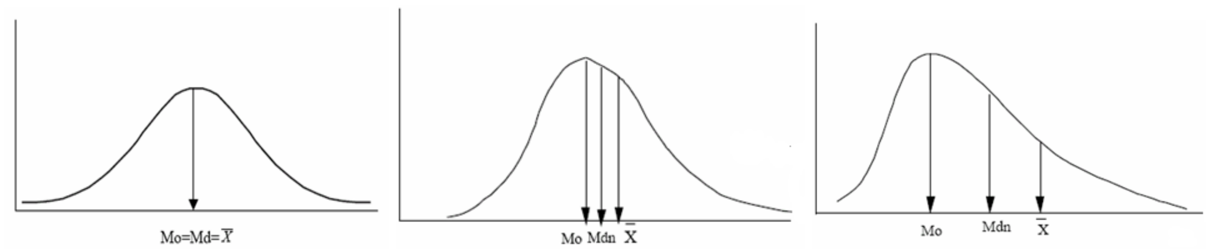
$X6$:{8, 8, 9, 10, 10, 12, 50} $\rightarrow$ $\bar{X} = 15{,}28$; $Mdn = 10$

La variable *X6* contiene un dato anómalo (tal vez motivado por un error en la recogida o en el proceso de los datos) y, como consecuencia del mismo, su media se ve muy afectada al alza, resultando poco representativo del grueso de los datos. La mediana, en cambio, no resulta afectada por ese dato atípico, siendo su valor el mismo en ambas variables.

Ejemplo del estudio de Millar y Fishkin (1997) acerca de la base evolutiva de la elección de la pareja humana, siendo la hipótesis que los hombres prefieren tener más parejas que las mujeres a lo largo de la vida. Ante la pregunta de cuál sería el número ideal de parejas a lo largo de un periodo de 30 años, se obtuvieron como resultados una media de 64,3 parejas para los hombres y de 2,8 para las mujeres, resultados que evocan de inmediato conclusiones dignas de una primera página en la prensa sensacionalista.
Sin embargo, si observamos más detenidamente un diagrama de barras de la distribución de frecuencias de ambos grupos...



• Dada la sensibilidad de la media a los datos anómalos o extremos, únicamente coincidirán los valores de la media, la mediana y la moda en una distribución simétrica unimodal, sin embargo, a medida que la distribución deje de ser simétrica, la coincidencia entre estos índices tenderá a disiparse, y el valor de la media se desplazará hacia los valores extremos o anómalos de la distribución.



• En variables con distribuciones que contengan datos anómalos (o distribuciones asimétricas) es más conveniente aplicar índices que no se vean tan afectados por este tipo de datos. Entre estos índices, conocidos también como estadísticos robustos, se encuentra la mediana (ya se vio anteriormente cómo este estadístico no se ve afectado por la presencia de valores anómalos) y la media recortada.

• La media recortada consiste en la obtención de la media aritmética excluyendo del cálculo un porcentaje de los casos situados en los extremos superior e inferior de la distribución, es decir, recortando en un porcentaje determinado las colas de la distribución. Por ejemplo, la media recortada al 10% excluye del cálculo al 10% inferior y al 10% superior de los valores obtenidos.

Por **ejemplo**, los valores de *X* obtenidos en un grupo de 10 sujetos son:

$$X: \{7, 8, 9, 10, 10, 12, 14, 16, 19, 57\} \qquad \rightarrow \qquad \overline{X} = 16{,}2$$

La presencia de un dato anómalo en la distribución (el valor 57) hace poco recomendable la obtención de la media aritmética. En este caso, es más aconsejable obtener la mediana (11), o bien, la media recortada. La media recortada al 10% se obtiene eliminando del cálculo el 10% inferior y el 10% superior de los valores, esto es, un valor de cada cola de la distribución (7 y 57, respectivamente) y, de esta manera, se obtendría un valor de la media igual a 12,25. ¿Cuál sería el resultado de la media recortada al 20% para los datos anteriores?

2. Otros estadísticos de posición grupal

• A continuación se presenta otro conjunto de estadísticos que permiten describir la posición o localización de los datos de una variable pero que, a diferencia de los del apartado anterior, no tienen por objeto proporcionar un valor que represente el *centro* de la distribución. Una característica común de éstos es que podrán ser obtenidos con variables ordinales y cuantitativas, pero no con categóricas.

2.1 . El mínimo y el máximo

• El <u>mínimo</u> es, de las modalidades que adopta una variable, aquélla cuyo valor sea más bajo en la escala de medida de la variable; complementariamente, el <u>máximo</u> será aquélla cuyo valor sea más alto. Ambos valores permiten hacerse una idea de entre qué valores se localizan los datos de una variable.

> **Ejemplo**: Para la variable con los datos recogidos con la pregunta: "Ansiedad que siente cuando se encuentra con mucha gente alrededor" (ver distribución de frecuencias en sección anterior), el mínimo sería "Nada" y el máximo "Mucha".

Ejercicio 5: obtener el mínimo y el máximo de las variables *X1*, *X2*, *X3* y *X4* de los ejemplos vistos en las secciones precedentes.

2.2 . Los cuantiles

• Se define como <u>cuantil</u> k (C_k) al valor de la variable tal que un *k%* de los sujetos tienen un valor inferior o igual a ese valor (k puede ser cualquier número entre 0 y 100, entero o decimal).

• La <u>mediana</u> es un caso particular de cuantil, en concreto, el cuantil 50 (C_{50}).

• El cálculo de un determinado cuantil k de una variable resulta relativamente sencillo de obtener a partir de la distribución de frecuencias de dicha variable si nos fijamos en la columna de los porcentajes acumulados: el C_k corresponderá al valor de la variable cuyo porcentaje acumulado sea igual a k o, en su caso, el porcentaje acumulado mayor a k que sea más bajo.

Ejemplo con la variable obtenida a partir de los datos recogidos con la siguiente pregunta de un test de cultura organizacional en empresas: "Se valora en los empleados la creatividad y la capacidad de creación". Al test contestó una muestra de 200 empleados de diferentes empresas. Las alternativas de respuesta a esta cuestión eran las siguientes: 1: Muy en desacuerdo; 2: Bastante en desacuerdo; 3: Algo en desacuerdo; 4: Ni en desacuerdo ni de acuerdo; 5: Algo de acuerdo; 6: Bastante de acuerdo; 7: Muy de acuerdo.

La distribución de frecuencias obtenida fue la siguiente:

X_i	n_i	$\%_i$	$\%_a$
2	21	10,5	10,5
3	31	15,5	26
4	36	18	44
5	47	23,5	67,5
6	38	19	86,5
7	27	13,5	100
	200	100	

En este ejemplo:

- el cuantil 67,5 de esta variable será igual a 5 => $C_{67,5}$ = 5, (o sea, que un 67,5% de sujetos contestaron que "Algo de acuerdo" o menos).

- el cuantil 40 de esta variable será igual a 4 => C_{40} = 4 (o sea, que un 40% de sujetos contestaron que "Ni en desacuerdo ni de acuerdo" o menos).

• Interpretación de los cuantiles:

- El valor de un determinado cuantil nos dice el porcentaje de casos que son iguales o inferiores a ese valor en la variable (su porcentaje acumulado). Por ejemplo, dados los datos de la variable "Peso" para una muestra de ratones, ¿para qué valor de esa variable hay un 90% de ratones que tienen un peso inferior o igual a ese valor? => El C_{90}.

- Complementariamente, el valor de un determinado cuantil nos dice también el porcentaje de casos que son superiores a ese valor en la variable. Así, para el ejemplo anterior, ¿para qué valor de la variable hay un 30% de ratones que tienen un peso superior a ese valor? => El C_{70}.

- También es posible plantearse dar respuesta a cuestiones relativas a intervalos de valores de la distribución de frecuencias. Por ejemplo, ¿entre qué valores se encuentra el 50% central de los ratones en la distribución de la variable Peso? => Entre el C_{25} y el C_{75}.

Ejercicio 6: a partir de los datos de la pregunta de un test "Ansiedad que siente cuando se encuentra con mucha gente alrededor" presentados en un ejemplo anterior, obtener: el cuantil 48,7 ($C_{48,7}$); el cuantil 50 (C_{50}); el cuantil 19 (C_{19}); el cuantil 20 (C_{20}); y el cuantil 75 (C_{75}). ¿Qué porcentaje de sujetos están por encima del C_{75}? ¿Y entre el C_{20} y el C_{80}?

Ejercicio 7: a partir de los datos recogidos con la pregunta de un test "Se valora en los empleados la creatividad y la capacidad de creación" (ver distribución de frecuencias más arriba), obtener el cuantil 5 (C_5), el cuantil 25 (C_{25}), el cuantil 45 (C_{45}), el cuantil 65 (C_{65}).

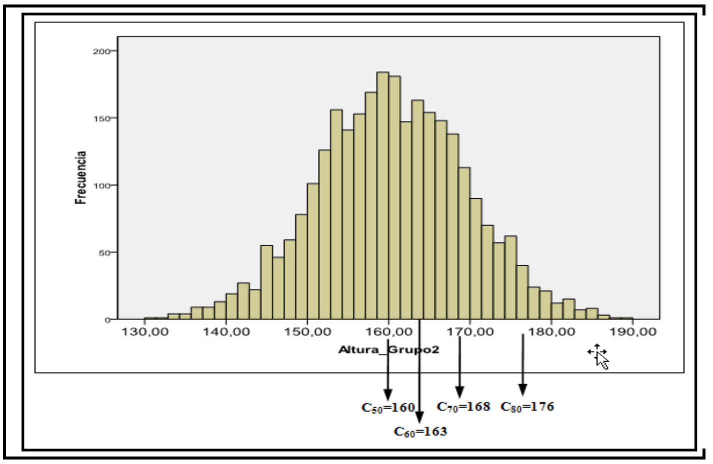
• A tener en cuenta acerca de la distancia entre los cuantiles: entre pares de cuantiles equidistantes existe el mismo porcentaje de sujetos. Por ejemplo, entre el C_{10} y el C_{20} existe un 10% de casos, lo mismo que entre el C_{20} y el C_{30}, y esto es así en la distribución de cualquier variable. Sin embargo, las distancias entre cuantiles no tienen por qué ser constantes en términos de distancias entre los valores de la variable.

Tomando como **ejemplo** la distribución de frecuencias de la pregunta del test de cultura organizacional, donde:

cuantil 5 (C_5)	→	2 (Bastante en desacuerdo)
cuantil 25 (C_{25})	→	3 (Algo en desacuerdo)
cuantil 45 (C_{45})	→	5 (Algo de acuerdo)
cuantil 65 (C_{65})	→	5 (Algo de acuerdo)

Se puede observar que entre los valores que corresponden al C_5 y el C_{25} la distancia es de 1 unidad, entre los del C_{25} y el C_{45} la distancia es de 2 unidades, y entre los del C_{45} y el C_{65} de 0. En cambio, entre los 3 pares de cuantiles hay, respectivamente, un 20% de sujetos.

Otro **ejemplo**: distribución de los valores de la variable "Altura (cm.)" medida en una muestra de niños de 14 años.



• <u>Tipos específicos de cuantiles</u> muy utilizados en la práctica en la difusión de resultados:

- Los centiles o <u>percentiles</u> (P_k): hacen referencia a los cuantiles 1 a 99, esto es, el valor de *k* es un número entero comprendido entre 1 y 99. Percentiles posibles: P_1, P_2, P_3, P_4 … P_{99} => dividen la distribución de la variable en 100 partes, cada una conteniendo el 1% de los casos.
- Los <u>cuartiles</u> (Q_k): hacen referencia a los cuantiles 25 (Q_1), 50 (Q_2) y 75 (Q_3). => dividen la distribución de la variable en 4 partes, cada una conteniendo el 25% de los casos.
- Los <u>deciles</u> (D_k): hacen referencia a los cuantiles 10 (D_1), 20 (D_2), 30 (D_3) … a 90 (D_9) => dividen la distribución de la variable en 10 partes, cada una conteniendo el 10% de los casos.

Tabla de equivalencias entre cuantiles

P_{10}	→	D_1	→	→	→	C_{10}
P_{20}	→	D_2	→	→	→	C_{20}
P_{25}	→	→	→	Q_1	→	C_{25}
P_{30}	→	D_3	→	→	→	C_{30}
P_{40}	→	D_4	→	→	→	C_{40}
P_{50}	→	D_5	→	Q_2	→	C_{50}
P_{60}	→	D_6	→	→	→	C_{60}
P_{70}	→	D_7	→	→	→	C_{70}
P_{75}	→	→	→	Q_3	→	C_{75}
P_{80}	→	D_8	→	→	→	C_{80}
P_{90}	→	D_9	→	→	→	C_{90}

Ejercicio 8: a partir de los datos recogidos con la pregunta de un test “Se valora en los empleados la creatividad y la capacidad de creación” (ver distribución de frecuencias más arriba), obtener el mínimo y el máximo, la moda, la mediana, el Q_3, el $C_{10,5}$, el P_3 y el D_9. ¿Qué porcentaje de sujetos

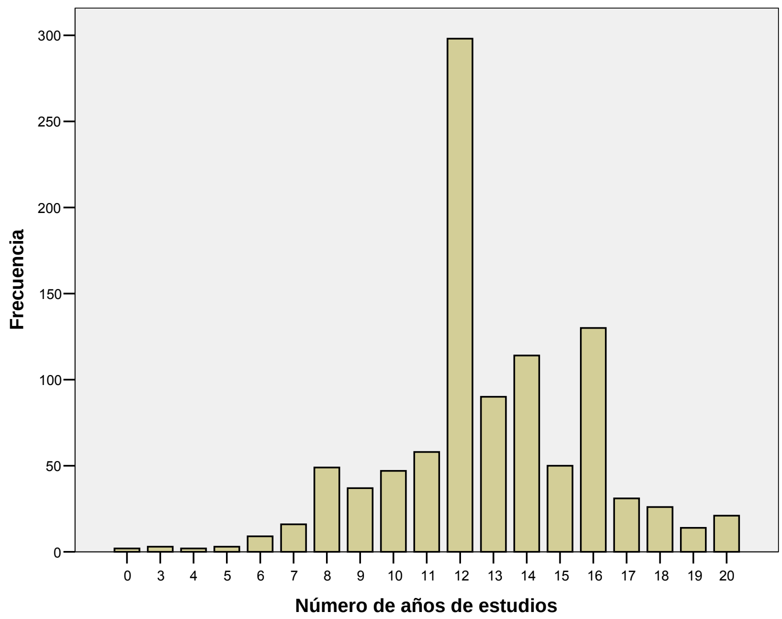
tienen un valor igual o inferior al C_{44}?; ¿cuántos sujetos se corresponden a ese porcentaje?; ¿qué porcentaje de sujetos tienen un valor superior al C_{44}?; ¿y, en concreto, cuántos sujetos son ese porcentaje?

Ejercicio 9: De una determinada variable X de la que se ha recogido información para 500 sujetos, ¿cuántos sujetos estarán entre el Q_1 y el Q_3?, ¿cuántos entre el C_{10} y el C_{90}?, y ¿cuántos entre el D_4 y el P_{60}?

Ejercicio 10: A partir de la distribución de frecuencias de la variable "Nº de años de estudios", obtenida a partir de una muestra aleatoria de 1000 sujetos adultos de la ciudad de Elche (n = 1000): (1) obtener los siguientes estadísticos de posición: la moda, el mínimo y el máximo, la media aritmética, la mediana, y los cuantiles P_{10}, D_2, P_{30}, D_9, y Q_3; (2) en función del tipo de variable y de la forma de su distribución de frecuencias (ver gráfico abajo), razona qué estadístico de tendencia central resultará más conveniente aplicar para esta variable.

Número de años de escolarización

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0	2	,2	,2	,2
	3	3	,3	,3	,5
	4	2	,2	,2	,7
	5	3	,3	,3	1,0
	6	9	,9	,9	1,9
	7	16	1,6	1,6	3,5
	8	49	4,9	4,9	8,4
	9	37	3,7	3,7	12,1
	10	47	4,7	4,7	16,8
	11	58	5,8	5,8	22,6
	12	298	29,8	29,8	52,4
	13	90	9,0	9,0	61,4
	14	114	11,4	11,4	72,8
	15	50	5,0	5,0	77,8
	16	130	13,0	13,0	90,8
	17	31	3,1	3,1	93,9
	18	26	2,6	2,6	96,5
	19	14	1,4	1,4	97,9
	20	21	2,1	2,1	100,0
	Total	1000	100,0	100,0	



• Acerca de la obtención de cuantiles en variables cuantitativas continuas: En este caso se suele tener en cuenta el carácter continuo de este tipo de variables y, en consecuencia, se suele aplicar una fórmula que permite obtener de un modo preciso cuál sería el valor concreto que deja por debajo el *k%* de observaciones correspondiente al cuantil que se quiera obtener (C_k). La fórmula la podemos encontrar en el texto de Botella et al. (2001, p. 70), si bien, se puede estimar con bastante precisión por interpolación a partir de la columna de los valores de la variable (X) y la de los porcentajes acumulados ($\%_a$).

Ejemplo: sea la variable "Tiempo empleado en completar un laberinto" por una muestra de 20 ratas (n = 20). Viendo que C_{85}=11 y que el C_{95}=12, el valor del C_{90} sería aproximadamente igual a 11,5. Estimar los valores de los cuantiles 35, 25, 50, 70, 75 y 96.

Tiempo (seg.)	n_i	p_i	p_a	$\%_a$
9	3	0,15	0,15	15
10	8	0,4	0,55	55
11	6	0,3	0,85	85
12	2	0,1	0.95	95
13	1	0,05	1	100

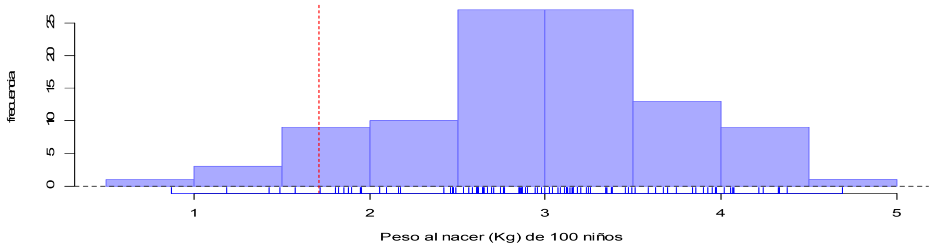
Ejercicio 11: A continuación se muestra la distribución de frecuencias de una variable ("distancia") que se obtuvo, tras medir en una muestra de 108 sujetos, la distancia (mm.) del centro de la pituitaria a la fisura ptérigo-maxilar. Calcular a partir de la misma los valores de los siguientes cuantiles: $C_{68,5}$, P_4, D_6, Q_2, P_{87}, C_{90}, $C_{2,8}$ y P_{99}.

distancia

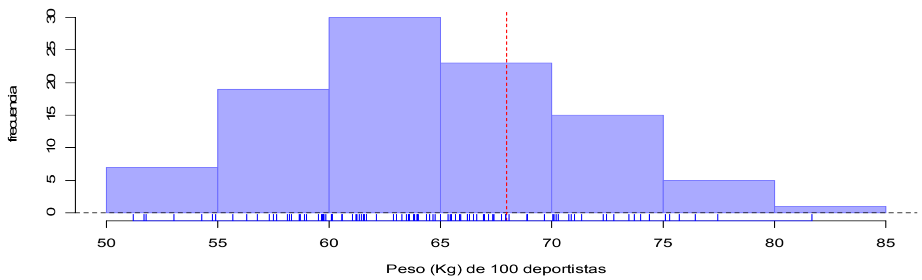
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	16,5	1	,9	,9	,9
	17,0	1	,9	,9	1,9
	19,0	2	1,9	1,9	3,7
	19,5	1	,9	,9	4,6
	20,0	4	3,7	3,7	8,3
	20,5	2	1,9	1,9	10,2
	21,0	5	4,6	4,6	14,8
	21,5	9	8,3	8,3	23,1
	22,0	4	3,7	3,7	26,9
	22,5	7	6,5	6,5	33,3
	23,0	11	10,2	10,2	43,5
	23,5	7	6,5	6,5	50,0
	24,0	6	5,6	5,6	55,6
	24,5	8	7,4	7,4	63,0
	25,0	6	5,6	5,6	68,5
	25,5	6	5,6	5,6	74,1
	26,0	7	6,5	6,5	80,6
	26,5	4	3,7	3,7	84,3
	27,0	2	1,9	1,9	86,1
	27,5	3	2,8	2,8	88,9
	28,0	4	3,7	3,7	92,6
	28,5	1	,9	,9	93,5
	29,0	1	,9	,9	94,4
	29,5	1	,9	,9	95,4
	30,0	1	,9	,9	96,3
	31,0	3	2,8	2,8	99,1
	31,5	1	,9	,9	100,0
	Total	108	100,0	100,0	

Ejercicio 12: A continuación se muestran, en forma de histograma, las distribuciones de frecuencias de 3 variables cuantitativas continuas (Barón-López, 2005). En cada caso aparecen una o varias líneas verticales discontinuas que marcan un determinado cuantil de la distribución. Se trata, en cada caso, de contestar a las cuestiones que a continuación se plantean.

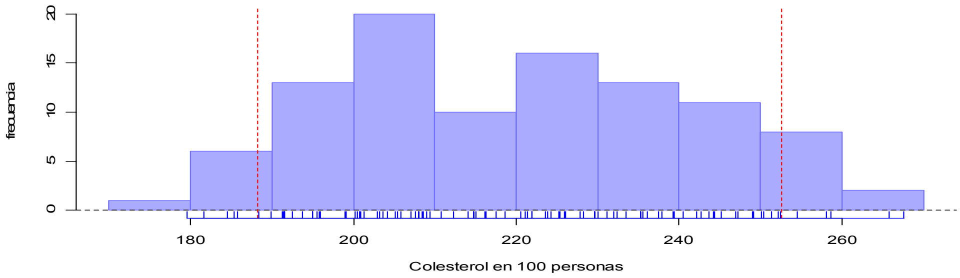
– En el siguiente histograma de la variable "Peso al nacer (Kg)" en una muestra de 100 niños ¿cuál es aproximadamente el valor del percentil 5 (línea discontinua)?, ¿cómo se interpreta?



- En el siguiente histograma de la variable "Peso" en una muestra de 100 deportistas ¿qué percentil nos indica el valor de peso que supera el 25% de los deportistas más pesados (ver línea discontinua)?, ¿cuál es, aproximadamente, el valor de este percentil?, ¿cuántos sujetos superan este peso?



- En un estudio en que se ha obtenido la distribución de frecuencias de la tasa de colesterol en sangre de una muestra de 100 sujetos de la población española, se afirma que el rango de valores de colesterol saludable corresponde al 90% de valores centrales de esa distribución.
¿Entre qué cuantiles se encuentra el 90% central de la distribución?; ¿a qué valores concretos de colesterol corresponden, aproximadamente, esos cuantiles (ver líneas discontinuas)?; ¿cuántos sujetos se encuentran fuera de ese intervalo?



Ejercicio 13: A continuación se muestran las tablas de crecimiento de la OMS para la variable altura de chicos y chicas de 5 a 19 años. A partir de ellas realiza las siguientes actividades:

a) A continuación se muestran los datos de altura de Ester entre sus 5 y 19 años. Ubica en la tabla correspondiente (con puntos) los datos de Ester y comenta cuál ha sido su evolución.

Años	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Altura (cm)	105	110	113	120	130	134	145	150	160	162	167	167	168	170	170

b) Carlos tiene 17 años y mide 175 cm., ¿qué percentil ocupa?, ¿qué percentil ocupaba cuando tenía 8 años?

c) Isabel tiene 15 años y mide 160 cm., ¿qué percentil ocupa?

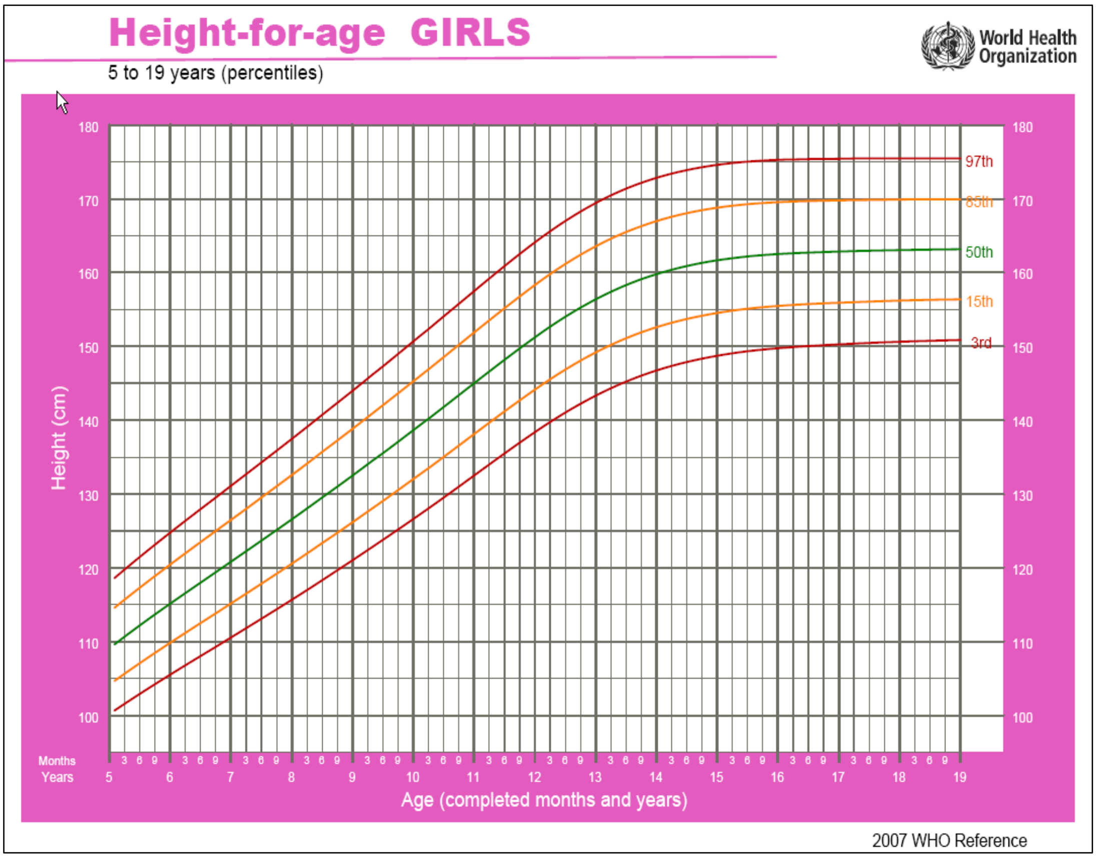
d) Juan tiene 18 años y su altura corresponde al P_{85}, ¿cuánto mide?

e) Una persona de 19 años que mide 175 cm., ¿qué percentil ocuparía si es una chica?, ¿y si es un chico?

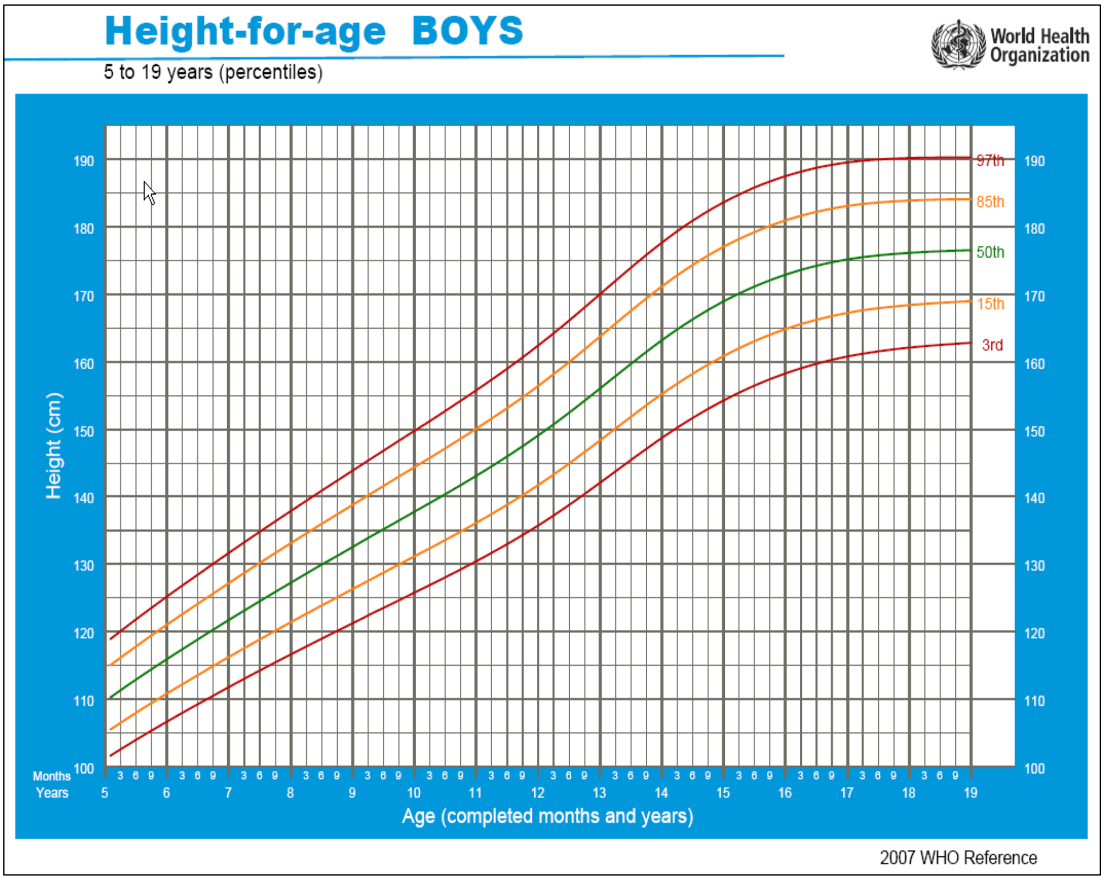
f) ¿Cuál es la mediana de altura de la población de chicas con 9 años de edad?

g) ¿Entre qué valores de altura se encuentra el 70% central de la población de chicos de 16 años?

h) Si se considera que una puntuación inferior al P_{15} puede indicar retraso en el crecimiento, se diagnosticaría que un niño de 10 años puede tener un problema de crecimiento si su altura es inferior a ……. cm.



Height-for-age GIRLS
5 to 19 years (percentiles)
World Health Organization
Height (cm)
97th
85th
50th
15th
3rd
Months
Years
Age (completed months and years)
2007 WHO Reference



Height-for-age BOYS
5 to 19 years (percentiles)
World Health Organization
Height (cm)
97th
85th
50th
15th
3rd
Months
Years
Age (completed months and years)
2007 WHO Reference

Referencias:

Barón-López, J. (2005). *Bioestadística: métodos y aplicaciones.* Apuntes y material disponible en http://www.bioestadistica.uma.es/baron/apuntes/

Botella, J., León, O. G., San Martín, R. y Barriopedro, M. I. (2001). *Análisis de datos en psicología I: teoría y ejercicios.* Madrid: Pirámide.

Aron, A. y Aron, E. N. (2001). *Estadística para psicólogos.* Buenos Aires: Pearson Education.

Millar, L. C. y Fishkin, S. A. (1997). Sobre la dinámica del éxito humano y el éxito reproductivo. En J. A. Simpson y D. T. Kendrick (Eds.): *Psicología Social Evolutiva.* Mahwah, NJ: LEA.

Solanas, A., Salafranca, L., Fauquet, J. y Núñez, M. I. (2005). *Estadística descriptiva en Ciencias del Comportamiento.* Madrid: Thompson.