Origen de la Estadística. Introducción

El origen de la Estadística no es conocido con precisión. El historiador romano Tácito afirma en sus anales que Augusto llevó a cabo una gran encuesta sobre los bienes de su Imperio. Dicha encuesta incluía el número de soldados, de naves y de todo tipo de bienes públicos. Precisamente el nacimiento de Cristo está ligado a un acto "estadístico" del Imperio: un empadronamiento realizado en una de sus provincias.

La palabra **Estadística** proviene de que fueron los Estados los impulsores de las primeras elaboraciones de datos, sobre todo, en el siglo XVIII.

En la Europa de esa época, aparecieron las compañías de seguros, que basaron sus ganancias en el conocimiento de la esperanza de vida de los individuos, circunstancia extraída de las tablas de mortalidad que, si bien defectuosas, demostraron ser eficaces.

Pero la Estadística no se ocupa sólo de encuestas. El Diccionario de la Lengua Española define el término **estadística** como *censo o recuento de la población, de los recursos naturales e industriales, del Estado, provincia, pueblo, clase, etc.* Como segunda acepción afirma *estudio de los hechos morales o físicos del mundo que se presentan a numeración o recuento y a comparación de cifras a ellos referentes.* Más concretamente, y referido a las matemáticas, dice:

Ciencia que utiliza conjuntos de datos numéricos para obtener a partir de ellos inferencias basadas en el cálculo de probabilidades.

Así pues, la estadística estudia fenómenos que, de alguna manera, se pueden cuantificar; que generan conjuntos de datos. Trabajando con estos datos el estadístico, utilizando las técnicas apropiadas, tratará de simplificar al máximo la información disponible, a fin de que pueda ser clara y útil. Además, si el fenómeno lo requiere, tratará de deducir, de inferir, las leyes que expliquen el comportamiento de ese fenómeno.

La idea de **inferencia** es la de deducción arriesgada y, por tanto, con posibilidad de error (error que debe ser conocido), pues la estadística se mueve en el campo del azar, de lo probable, y es ahí donde pretende buscar las leyes que determinan su comportamiento a fin de poder tomar las decisiones oportunas, conociendo con antelación la significación de esos resultados.

Veamos tres situaciones:

I. En un regimiento de 900 soldados, deseamos conocer la estatura de todos ellos.

II. Se desea conocer la talla de los varones españoles que tengan una edad comprendida entre 18 y 30 años.

III. Un fabricante de bombillas se encuentra con una producción de 10.000 unidades que, por algún fallo en el proceso de fabricación, comprobado posteriormente, teme que sean defectuosas. Desea, pues, conocer sus características antes de lanzarlas al mercado, porque, en caso de ser defectuosas, considera preferible destruirlas.

En el primer caso es muy fácil tallar a los 900 soldados. Los resultados obtenidos se tabularán y, si se quiere, se representarán gráficamente.
Sin embargo, es prácticamente imposible tallar a los millones del caso II. No quedará más remedio que tomar una *muestra* (un pequeño conjunto de ellos) y efectuar la medida en éstos. Las conclusiones se harán extensivas a la totalidad.

Las características que el fabricante de bombillas desea comprobar son la luminosidad y la duración de cada una de ellas. La primera, la luminosidad, la puede comprobar exahustivamente (en cada una de las 10.000 bombillas), pero sería un proceso demasiado largo. La duración, sin embargo, no puede medirla en todas, pues la única forma de averiguar la vida de una lámpara es dejarla encendida hasta que se funda y cronometrar su duración (es un proceso destructivo). No le quedará más remedio que proceder al estudio de una muestra y extender las conclusiones a la totalidad.

La estadística se puede dividir en dos partes:

Estadística descriptiva La estadística descriptiva o deductiva, trata del recuento, ordenación y clasificación de los datos obtenidos por las observaciones. Se construyen tablas y se representan gráficos que permiten simplificar, en gran medida, la complejidad de todos los datos que intervienen en la distribución. Asimismo se calculan parámetros estadísticos que caracterizan la distribución. ***En esta parte de la estadística no se hace uso del cálculo de probabilidades, y únicamente se limita a realizar deducciones directamente a partir de los datos y parámetros obtenidos*** **(caso I).**

Estadística inferencial La estadística inferencial o inductiva plantea y resuelve el problema de establecer previsiones y conclusiones válidas generales sobre una población a partir de los resultados obtenidos de una muestra. ***Utiliza resultados obtenidos mediante la estadística descriptiva y se apoya fuertemente en el cálculo de probabilidades*** **(casos II y III).**

Población y Muestra

Se denomina **población** al conjunto de todos los elementos que cumplen una determinada característica, que deseamos medir o estudiar. Los elementos de la población se llaman **individuos** (debido a su origen demográfico).

Se llama **muestra** a cualquier subconjunto de la población (precisamente aquel subconjunto sobre el que se va a realizar el estudio). El número de elementos de la muestra se llama **tamaño** de la misma. Cuando la muestra coincide con la población se llama **censo**.

Ejemplo: En los problemas que planteamos antes, las *poblaciones* estaban formadas por:

- 900 soldados de un regimiento.
- Todos los varones españoles de edades entre 18 y 30 años.

- 10.000 bombillas.

Los *individuos* eran:

- cada uno de los soldados.
- cada uno de los varones españoles de 18 a 30 años.
- cada una de las 10.000 bombillas.

Un mismo conjunto puede ser población o muestra según los casos. Por ejemplo, el conjunto de seres vivos encontrados en una parcela de una Hectárea, será población si nuestro interés se limita a esa parcela; será muestra si, pretendiendo conocer la ecología de una región, acotamos esa parcela para sacar de su estudio conclusiones extensivas a toda la región.

Encuestas. Muestreo

La encuesta

Se considera encuesta a una modalidad de investigación destinada a la recogida de información por medio de preguntas formuladas directamente a los sujetos. El objetivo de la encuesta es obtener información de una población específica (votantes, amas de casa, jóvenes...), valiéndose para ello de los datos obtenidos en una muestra; por esta razón se le suele llamar **encuestas por muestreo**. Esta información se conseguirá mediante las estimaciones oportunas.

La información que habitualmente se persigue suele estar relacionada con intereses, intenciones, deseos o actitudes de grupos de población. Todos estos aspectos suelen ser cambiantes a lo largo del tiempo, ya que dependen de aspectos personales o, a veces, de modas; y siempre de los individuos encuestados.

Así pues, una encuesta viene a ser una instantánea fotográfica de una población (más exactamente de una parte de la población, de los encuestados). Si esa fotografía está bien enfocada, y enmarcada convenientemente, describirá acertadamente la realidad; en caso contrario, creará confusión.

El enmarcado de una encuesta puede asociarse con el acierto en la elección de la muestra. El enfocado, con el cuestionario apropiado y con la eficaz aplicación por parte de los encuestadores. Siguiendo con la fotografía, cabría hablar del revelado, que se correspondería con un buen tratamiento de los datos, con una interpretación correcta y con una presentación clara.

El cuestionario

A este instrumento de investigación se le exige que sea completo, concreto y claro: que no induzca a errores; que tenga en cuenta todas las posibilidades de respuesta.

Ha de permitir aflorar la información deseada: evitando preguntas complejas, comprometedoras o demasiado íntimas.

La muestra

Por lo que respecta a la muestra, primero hay que tener definido el **marco muestral**, esto es, el conjunto de individuos susceptibles de poseer la información buscada. En segundo lugar hay que determinar el método de muestreo más apropiado para la investigación. **Muestreo** es el proceso seguido en la extracción de la muestra.

El muestreo puede ser **probabilístico** o no. En las muestras no probabilísticas no se usa el azar, sino el criterio del investigador. Este tipo de muestreo se usa con frecuencia en el mundo periodístico para conocer rápidamente la opinión de oyentes o lectores sobre una cuestión de actualidad. Suele presentar grandes sesgos y es poco fiables.

El denominado **aleatorio** puede ser muy diverso; a continuación, señalamos algunos tipos:

Muestreo aleatorio simple

Es el método más fiable para la elección de una muestra. Consiste en extraer, mediante un sorteo riguroso, una serie de individuos de una población. Todos los individuos han de tener la misma probabilidad de ser elegidos. Es el método más puro, pero, sin embargo, no se usa siempre. Es de fácil aplicación cuando la población es pequeña y se complica para poblaciones grandes. La razón estriba, en primer lugar, en el alto costo que supone: imagínese lo que costaría entrevistar a 3000 personas distribuidas aleatoriamente por todo el territorio español. En otros casos no es aconsejable porque, paradójicamente, puede viciar la muestra. Esto ocurre cuando la población a entrevistar se presenta muy claramente estratificada. En este caso se recurre a otro tipo de selección.

Muestreo sistemático

Para realizar el muestreo sistemático se ordenan previamente los individuos de la población, después se elige a uno de ellos al azar y a continuación, a intervalos constantes, se elige a todos los demás hasta completar la muestra. Por ejemplo, para diseñar una muestra de 30 individuos entre 300, los numeramos de 1 a 300, sorteamos uno de ellos (supongamos que sale el 57); a partir de aquí tomamos el 67, 77, 87,....., de 10 en 10, hasta completar los 30.

Muestreo aleatorio estratificado (proporcional)

Este tipo de muestreo divide la totalidad de los individuos en clases homogéneas, según criterios lógicos; por ejemplo, por grupos de edades, o por número de habitantes de las distintas poblaciones.

Supongamos que se pretende hacer una encuesta entre los 800 alumnos de un Instituto. Si tomamos una muestra aleatoria simple de 80 individuos, el azar podría dar una representación de los diferentes cursos distinta de la real. Se recurre entonces a seleccionar a los individuos según el estrato al que pertenezcan.

Si por ejemplo, hay 350 alumnos de Primero, 200 de Segundo, 150 de Tercero y 100 de COU, una muestra de 80 individuos adecuada se obtendría tomando al azar 35 alumnos de Primero, 20 alumnos de Segundo, 15 de Tercero y 10 de COU, ya que la edad y el nivel son condicionantes muy importantes en los sondeos de opinión.

Muestreo no aleatorio

La elección deliberada de individuos (aunque no se los conozca) puede dar lugar a errores graves, pues el que elige la muestra puede introducir un criterio de selección sin pretenderlo.

Si, por ejemplo, se toma la guía telefónica como fuente de selección de nombres, estamos tomando individuos de un determinado nivel económico (dado por "tener teléfono") o eliminando los individuos que menos tiempo están en casa, etc. Precisamente un error de este tipo cometió en 1936 la revista norteamericana *Literary Digest* que se cita a continuación: la fuente de la encuesta fue su lista de suscriptores y el listín telefónico. Esa fue la principal causa de que no acertara en sus predicciones.

El tamaño de la muestra

El tamaño de la muestra dependerá del fenómeno analizado y de la precisión y fiabilidad que quiera exigirse a los resultados, pero no necesariamente una muestra mayor da resultados más fiables (hay que tener en cuenta el tipo de muestreo, el buen hacer de los encuestadores, el cuestionario...). En nuestro país, la mayoría de las encuestas se hacen sobre muestras de un tamaño entre 800 y 2000 personas.

Tratamiento de los datos

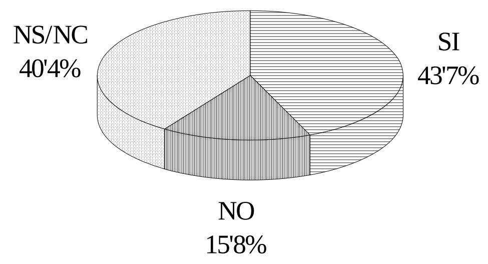
Un buen tratamiento de los datos requiere conocimientos estadísticos, experiencia, trabajo, espíritu crítico y objetividad..., aspectos que, sin duda alguna, alcanzarán todos los pacientes lectores de estos apéndices.

La ficha técnica

Toda encuesta debe ir acompañada de una ficha técnica, que debe indicar: el tamaño de la muestra, el método de muestreo, la fiabilidad de los resultados, y las fechas en que se realizó.

La siguiente ficha técnica corresponde a una encuesta realizada sobre el "Tratado de Maastricht" en la que se hacía la pregunta: ¿Cree usted que España debería aceptar el Tratado de Maastricht?.

- **Universo:** Personas mayores de 18 años.
- **Ámbito:** Nacional.
- **Muestra:** Ochocientas entrevistas, con un error posible de $\pm 3'5\%$ para un nivel de confianza del 95%; y p/q=50/50.
- **Selección:** Aleatoria, a partir del sistema de cuotas por sexo, edad y ocupación.
- **Entrevista:** Telefónica.
- **Fecha de trabajo de campo:** 6 y 8 de Junio de 1992.
- **Realización:** Sigma Dos



Origen y auge de los sondeos

El auge de los sondeos y encuestas comenzó en 1936 en Estados Unidos. Con motivo de las elecciones presidenciales, una revista norteamericana, *Literary Digest*, se lanzó con todo lujo de medios a realizar una consulta popular en gran escala.

Se registraron más de cuatro millones de entrevistas. Con todas estas respuestas, la revista anticipó el nombre del vencedor: el republicano Landon. Pero sus previsiones no se cumplieron.

En ese mismo año, un estadístico, entonces casi desconocido, **G. Gallup**, efectuó un sondeo en el que sólo fueron interrogadas 4500 personas. Gallup, a partir de estos datos, pronosticó el triunfo de Roosvelt y, con un pequeño error acertó incluso en el porcentaje de votos. Con ello, Gallup alcanzó una gran popularidad y el **Instituto Gallup** es una de las organizaciones de mayor fiabilidad en la realización de sondeos.

Sondeos previos y resultados en la elecciones al Parlamento Europeo (12-6-87)

	Sigma Dos (Diario 16)	**Demoscopia** (El País)	**Gallup** (Época)	**Iope-Etmar** (El Periódico)	**CIS**	**La Ser**	**Escaños**	**%**
PSOE	25-27	23-26	29-30	28	30	28-29	**28**	46'66
AP	15-16	15-17	16	17	17	16-17	**17**	28'33
CDS	7-9	9-10	6	6	6	6-7	**7**	11'66
IU	3	1-2	3	3	2	3-4	**3**	5

CiU	2-3	2-3	2-3	3	2	3	**3**	5
HB	1	0-1	0	1	0	1	**1**	1'66
Otros	1	0-1	1	1	0	0-1	**1**	1'66

Caracteres y modalidades

Los **caracteres estadísticos** de una población son las propiedades o cualidades de los individuos que nos interesa estudiar. Nos fijamos en un carácter estadístico e ignoramos, en esta encuesta el resto. Por ejemplo, si hacemos un estudio estadístico sobre el nivel de estudio de los españoles, es claro que dejaremos de lado las cuestiones como la estatura, peso, etc.

Se distinguen dos tipos de caracteres estadísticos: cuantitativos y cualitativos.

Caracteres estadísticos cuantitativos son aquellos que se pueden medir: Ejemplos:

- La talla de un individuo.
- El diámetro de una pieza de precisión.
- Número de acciones vendidas en la bolsa de Madrid.
- El cociente intelectual de un alumno.
- La renta per cápita en cada una de las comunidades autónomas.

Caracteres estadísticos cualitativos son aquellos que no se pueden medir. Ejemplos:

- La profesión de una persona.
- El estado civil.
- El color de los ojos.
- La carrera que piensa estudiar un alumno de COU.
- El idioma elegido en el bachillerato.

Se llama **modalidades** de un carácter estadístico a cada una de las diferencias que se pueden establecer dentro de un mismo carácter estadístico cualitativo. Son modalidades del carácter profesión las siguientes: economista, psicólogo, sociólogo, ingeniero, matemático, biólogo, etc.

Variables estadísticas

Se llama **variable estadística** al conjunto de valores que toma un carácter estadístico. Dependiendo del carácter, una variable estadística puede ser *cuantitativa* o *cualitativa*.

Variable estadística discreta

Una variable estadística se llama discreta cuando puede tomar un número finito de valores o infinito numerable.

Variables estadísticas discretas son por ejemplo:

- Número de empleados de una fábrica.
- Número de hijos de 20 familias.
- Número de goles marcados por la Selección Nacional de Fútbol en cada una de las últimas 30 temporadas.

Variable estadística continua

Una variable estadística es continua cuando puede tomar, al menos teóricamente, todos los valores posibles dentro de un cierto intervalo de la recta real.

Variables estadísticas continuas son por ejemplo:

- La altura de un individuo.
- El peso de un individuo.
- El tamaño de los objetos.
- El tiempo que tarda en caer un objeto.

En la práctica muchas medidas de carácter continuo se hacen discretas, como, por ejemplo, la talla de los individuos: la estatura suele redondearse en cm.. Otras veces, fundamentalmente para obtener resultados teóricos, una variable discreta puede tratarse como continua.

Los valores de las variables estadísticas se acostumbran a representar por $\mathbf{x_1, x_2, x_3, \cdots\cdots\cdots, x_n}$

Frecuencias absolutas y relativas

<u>Frecuencia absoluta</u>

Se llama frecuencia absoluta del valor x_i, y lo representamos por $\mathbf{f_i}$, ***al número de veces que se repite dicho valor.***

<u>Frecuencia absoluta acumulada</u>

Se llama frecuencia absoluta acumulada del valor x_i, y lo representamos por $\mathbf{F_i}$, ***a la suma de las frecuencias absolutas de todos los valores anteriores a*** $\mathbf{x_i}$ ***más la frecuencia absoluta de*** $\mathbf{x_i}$***:***

$$\boxed{\mathbf{F_i = f_1 + f_2 + f_3 + \cdots\cdots\cdots + f_i}}$$

La frecuencia absoluta no es suficiente para reflejar la intensidad con que se repite un determinado valor de la variable estadística. Por ejemplo, no es lo mismo obtener tres veces un cinco en diez lanzamientos de un dado que obtenerlo en 1000 lanzamientos del mismo.

<u>Frecuencia relativa</u>

Se llama frecuencia relativa de un valor x_i y lo representaremos por $\mathbf{f_r}$, ***al cociente entre la frecuencia absoluta de*** $\mathbf{x_i}$ ***y el número total de datos que intervienen en la distribución***:

$$\boxed{\mathbf{f_r = \frac{f_i}{\sum f_i}}}$$

siendo $\sum \mathbf{f_i}$ el número total de datos.

Si cada frecuencia relativa se multiplica por 100 se obtiene el tanto por ciento correspondiente a cada valor.

<u>Frecuencia relativa acumulada</u>

Se llama frecuencia relativa acumulada del valor x_i, y la representaremos por $\mathbf{F_r}$, **a la suma de las frecuencias relativas de todos los valores anteriores a $\mathbf{x_i}$ más la frecuencia relativa de $\mathbf{x_i}$.**

En todos los casos, la suma de las frecuencias absolutas debe ser igual al total de elementos, la suma de las frecuencias relativas será igual a la unidad, y la suma de los porcentajes deberá ser igual a 100.

Tratamiento de la información. Tablas estadísticas

A continuación vamos a estudiar cómo debemos proceder ordenadamente para analizar la muestra:

1. **Recogida de datos**: Consiste en la toma de datos numéricos procedentes de la muestra.

2. **Ordenación de los datos**: Una vez recogidos los datos los colocaremos en orden creciente o decreciente.

3. **Recuento de frecuencias**: Efectuaremos el recuento de los datos obtenidos.

4. **Agrupación de los datos**: En caso de que la variable sea continua o bien discreta pero con un número de datos muy grande es muy aconsejable agrupar los datos.

   Se llama **intervalos de clase** ***a cada uno de los intervalos en que pueden agruparse los datos de una variable estadística***. Ahora bien, ¿Cuál es el número idóneo de intervalos que debemos escoger a la hora de agrupar? No existe una contestación tajante a esta pregunta; existen incluso varios criterios para dar respuesta a esta cuestión.

   Con carácter muy general podemos enunciar como uno de los criterios muy sencillos el de **Norcliffe** que establece que ***el número de intervalos debe ser aproximadamente igual a la raíz cuadrada positiva del número de datos***. Para obtener la amplitud de dichos intervalos se procede del siguiente modo:

   - *Se localizan los valores menor y mayor de la distribución.*

   - *Se restan. Si la diferencia es divisible entre el número de intervalos, el cociente nos da su amplitud. En caso contrario se busca el primer número entero por exceso de la diferencia que sea divisible por el número de intervalos. El cociente de esta división nos da la amplitud de los mismos.*

   - *Se forman los intervalos de modo que contengan todos los datos*.

   Con el fin de que la clasificación esté bien hecha se deben construir de tal manera que ***el límite superior de un intervalo coincida con el límite inferior del siguiente***. *Se debe adoptar el criterio de que los intervalos sean cerrados por la izquierda y abiertos por la derecha.*

   El punto medio entre los extremos de cada intervalo se llama **marca de clase**.

5. **Construcción de la tabla estadística**: En la tabla deberán figurar los valores de la variable (y en caso de que se encuentre agrupada en intervalos, los límites superior e inferior, así como las marcas de clase), frecuencias absolutas y frecuencias relativas. A veces es conveniente incluir las frecuencias absolutas acumuladas y las frecuencias relativas acumuladas.

   En muchas ocasiones es interesante trabajar con ***porcentajes****; éstos se obtienen multiplicando las frecuencias relativas por 100.*

Tabla estadística para una variable estadística discreta

Un profesor tiene anotadas en su cuaderno las notas de los 30 alumnos de una clase. Construir la tabla sabiendo que son las siguientes:

5 3 4 1 2 8 9 8 7 6

6 7 8 9 7 7 1 0 1 5

9 9 8 0 8 8 8 9 5 7

x_i	f_i	f_r	%	F_i	F_r
0	2	0'0666	6'66	2	0'0666
1	3	0'1000	10	5	0'1666
2	1	0'0333	3'33	6	0'1999
3	1	0'0333	3'33	7	0'2332
4	1	0'0333	3'33	8	0'2665
5	3	0'1000	10	11	0'3665
6	2	0'0666	6'66	13	0'4331
7	5	0'1666	16'66	18	0'5997
8	7	0'2333	23'33	25	0'8333
9	5	0'1666	16'66	30	1
	30	**1**	**100**		

Tabla estadística para una variable estadística continua

Construir la tabla estadística de las edades de las personas que acuden a un logopeda a lo largo de un mes, sabiendo que son las de la tabla adjunta.

3 2 11 13 4 3 2 4 5 6 7 3

4 5 3 2 5 6 27 15 4 21 12 4

3 6 29 13 6 17 6 13 6 5 12 26

Hay 36 datos, por tanto el número de intervalos que debemos formar es $\sqrt{36} = 6$.

Por otra parte, como el valor menor de la distribución es 2 y el mayor es 29 la diferencia es 27. El número entero más próximo a 27 por exceso que es divisible entre 6 es 30 por tanto los intervalos deben tener de amplitud $\frac{30}{6} = 5$.

Intervalos	**Marcas x_i**	f_i	f_r	**%**	F_i	F_r
$[0,5)$	2'5	13	0'3611	36'11	13	0'3611
$[5,10)$	7'5	11	0'3055	30'55	24	0'6666
$[10,15)$	12'5	6	0'1666	16'66	30	0'8332
$[15,20)$	17'5	2	0'0555	5'55	32	0'8887
$[20,25)$	22'5	1	0'0277	2'77	33	0'9164
$[25,30)$	27'5	3	0'0833	8'33	36	1
		36	**1**	**100**		

Gráficos estadísticos

Las tablas estadísticas presentan toda la información posible de modo esquemático y están preparadas para el cálculo posterior, si bien esta información puede transmitirse de una manera más expresiva a través de los gráficos estadísticos, que aunque en algunas ocasiones sea menos exacta y puntual que en el caso de las tablas, las magnitudes y frecuencias se recuerdan con más facilidad si se representan gráficamente. Los gráficos tienen un mayor poder de comunicación de los resultados al utilizar formas visuales de fácil comprensión (*la información entra por los ojos)*.

La claridad que requiere un gráfico hace que no sea indiferente el modelo utilizado. Según como sea la naturaleza del carácter estudiado, utilizaremos uno u otro tipo de representación gráfica.

Diagrama de barras

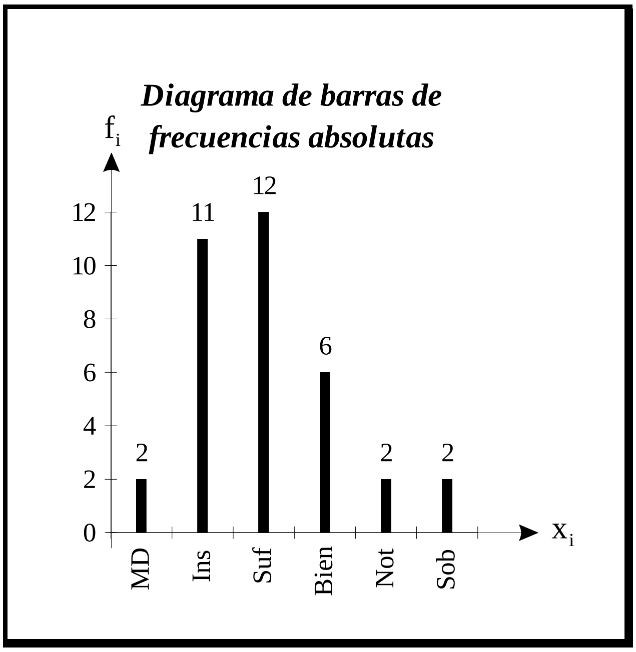
Recibe el nombre de **diagrama de barras** el gráfico que ***asocia a cada valor de la variable una barra estrecha, generalmente vertical, de longitud proporcional a la frecuencia (o a la cantidad) con que se presenta.*** Suele usarse cuando la variable es discreta.

Para trazarlos se representan sobre el eje de abscisas los valores de la variable y sobre el eje de ordenadas las frecuencias absolutas o relativas, según proceda. A continuación, por los puntos marcados en el eje de abscisas se levantan trazos gruesos o barras, de longitud igual a la frecuencia correspondiente.

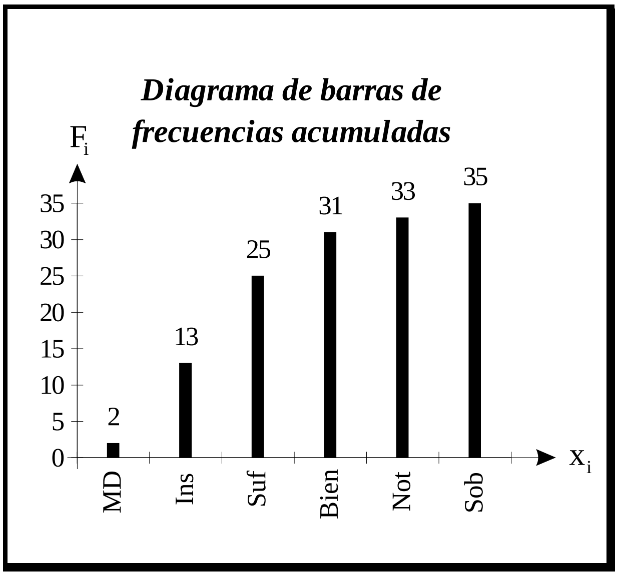
Las barras pueden representarse también *horizontalmente*. La ventaja de la horizontalidad estriba en que es más fácil añadir leyendas.

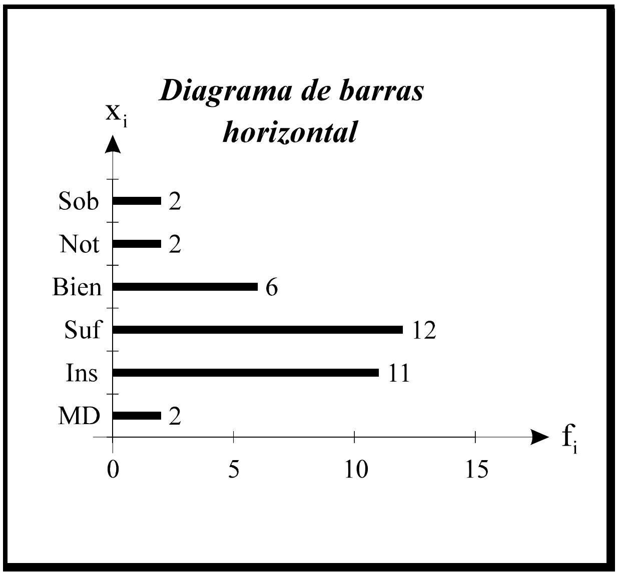
Las notas de matemáticas de los 35 alumnos de una clase vienen dadas por la siguiente tabla:

Calificaciones	Nº de alumnos
MD	2
Ins	11
Suf	12
Bien	6
Not	2
Sob	2



A continuación se representa el diagrama de barras de frecuencias acumuladas y el diagrama de barras horizontal.

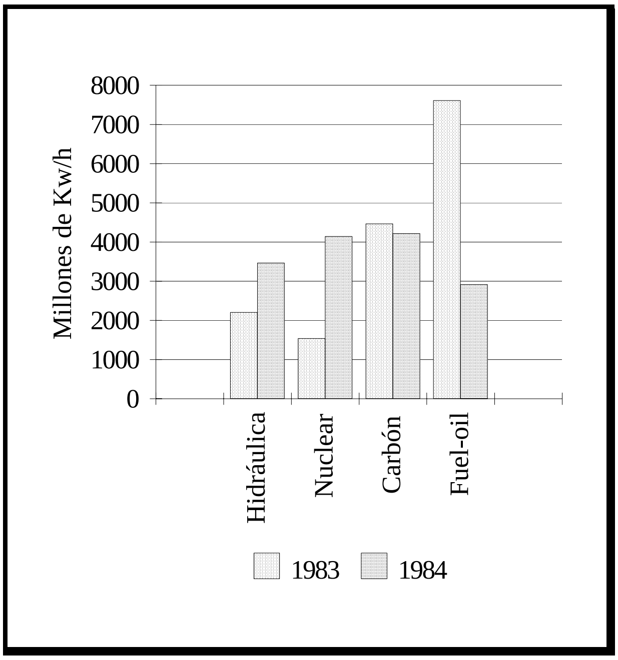


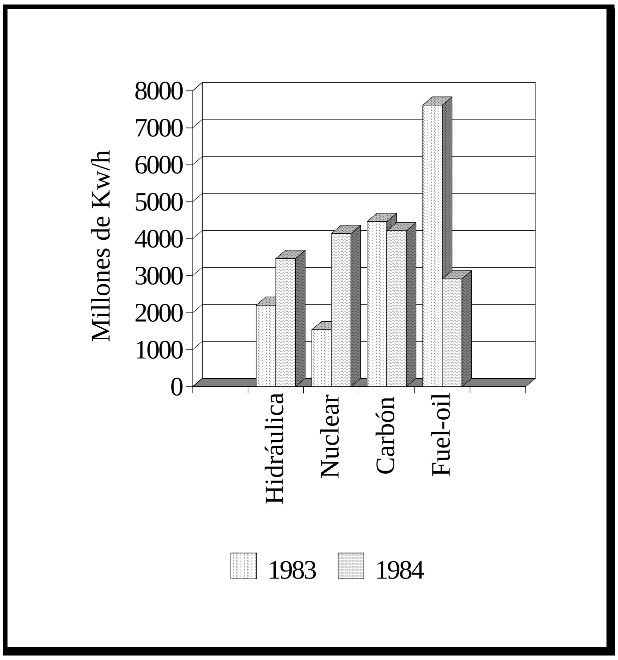


Los diagramas de barras pueden usarse **para comparar frecuencias** de las modalidades de una variable cualitativa en dos conjuntos de datos distintos.

La tabla del margen representa la producción eléctrica de Iberdrola II en dos años consecutivos 1983 y 1984

	Año	
Medio de generación	1983	1984
Hidraúlica	2204	3464
Nuclear	1540	4141
Carbón	4464	4216
Fuel – Oil	7612	2913
Total	15820	14734

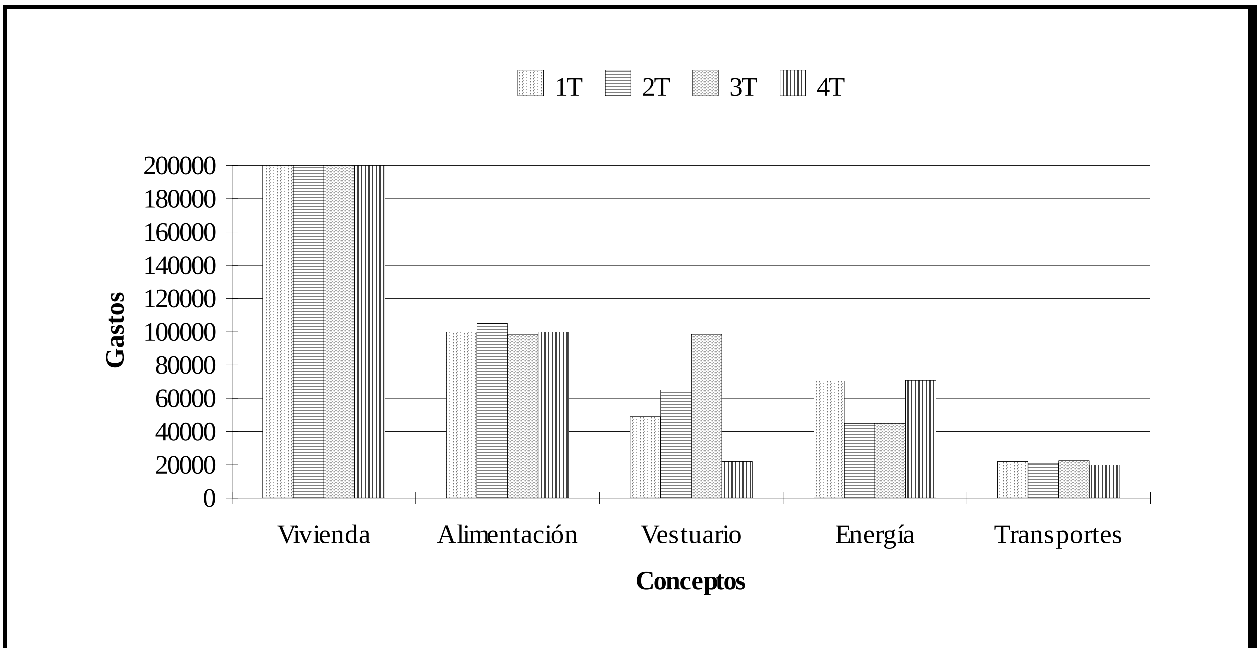


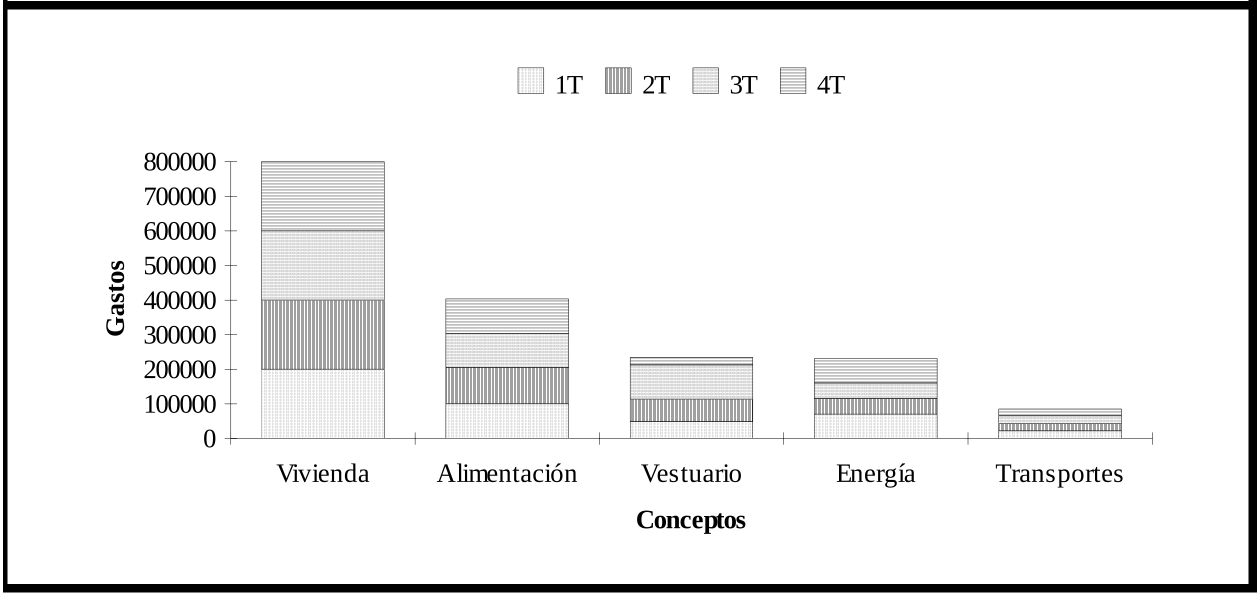


Veamos a través de un ejemplo todas las posibilidades que nos ofrecen los gráficos de columnas. Supongamos la tabla que sigue donde se reflejan los gastos anuales de una familia.

Gastos	1er T	2º T	3er T	4º T	Totales
Vivienda	200.000	200.000	200.000	200.000	800.000
Alimentación	100.000	105.000	98.500	99.999	403.499
Vestuario	49.000	65.000	98.500	22.000	234.500
Energía	70.500	45.000	45.000	70.800	231.300
Transportes	22.000	21.000	22.500	20.000	85.500
Total	441.500	436.000	464.500	412.799	1.754.799

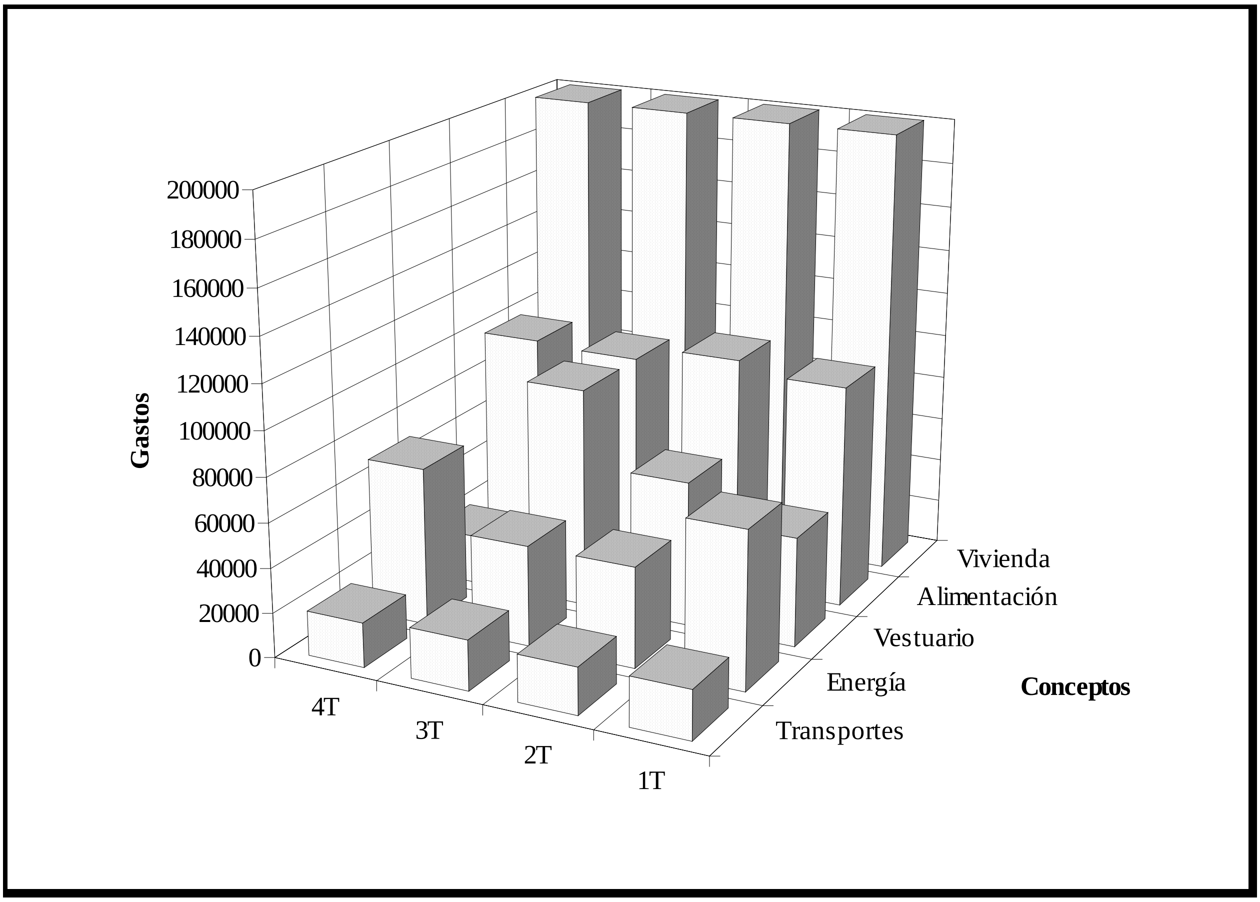
El siguiente gráfico de columnas permite comparar fácilmente los gastos trimestrales en los distintos conceptos.





Los gráficos de columnas apiladas como el anterior permiten ver el gasto totalizado por trimestre.

El siguiente gráfico es más atractivo a la vista por ser tridimensional.



Histogramas

Los histogramas son diagramas de barras que se utilizan específicamente para distribuciones de variable estadística continua, o bien para distribuciones de variable estadística discreta con gran número de datos y que se han agrupado por intervalos (clases). Generalmente se acostumbra a agrupar los datos obtenidos en intervalos de igual amplitud.

Estos gráficos asocian a cada intervalo un rectángulo de superficie proporcional a la frecuencia correspondiente a dicho intervalo.

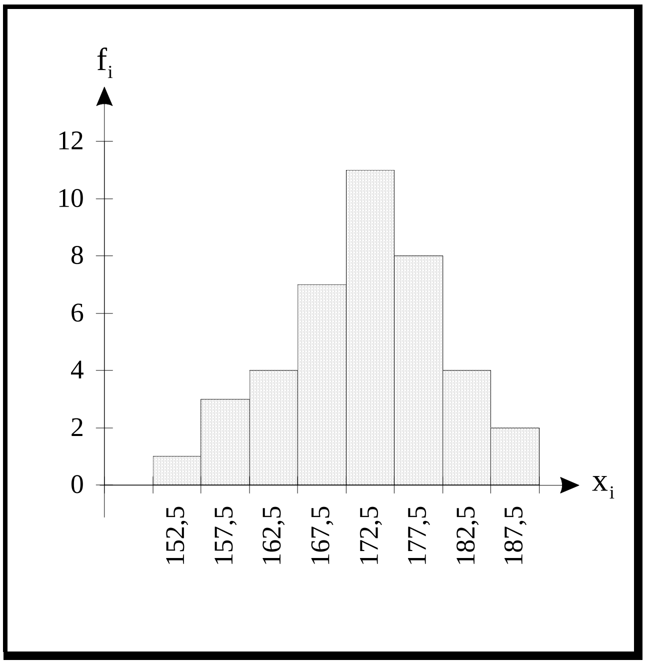
Para construir el histograma se representa sobre el eje de abscisas los límites de los intervalos. Sobre dicho eje se construyen unos ***rectángulos que tienen por base la amplitud del intervalo y por alturas los cocientes entre las frecuencias absolutas y las longitudes de los intervalos correspondientes (*****densidad de frecuencia*****).*** De esta manera, el área del rectángulo coincide con la frecuencia del intervalo.

Intervalos de la misma amplitud

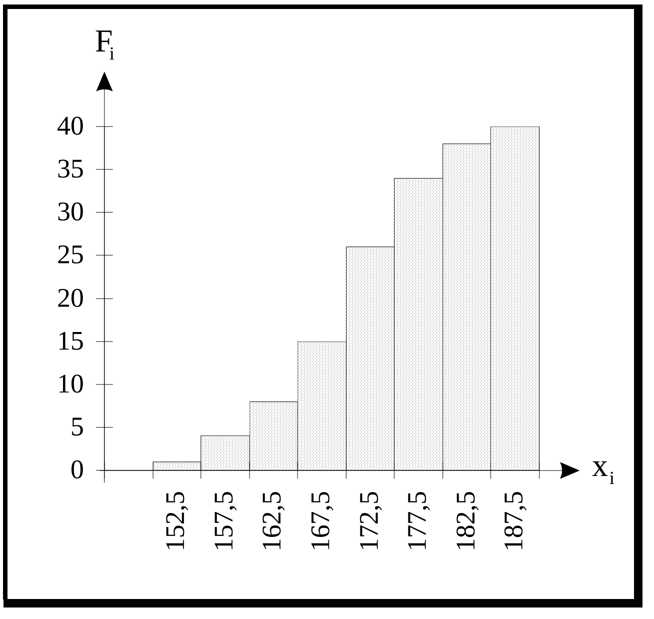
En este caso, las alturas son directamente proporcionales a las frecuencias.

En un grupo de 40 alumnos se ha estudiado su talla, obteniéndose los resultados de la tabla adjunta.

Intervalos en cm.	x_i	f_i
$[150,155)$	152'5	1
$[155,160)$	157'5	3
$[160,165)$	162'5	4
$[165,170)$	167'5	7
$[170,175)$	172'5	11
$[175,180)$	177'5	8
$[180,185)$	182'5	4
$[185,190)$	187'5	2
		40



En la figura adjunta se representa el histograma acumulativo correspondiente a las estaturas de los alumnos recogidas en el ejemplo anterior.



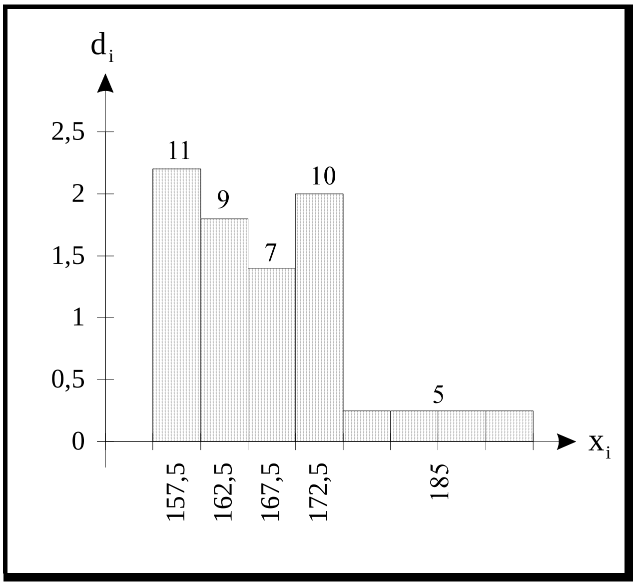
Como en estos dos gráficos la amplitud de los intervalos es la misma, hemos cogido directamente las frecuencias en el eje de ordenadas en vez de las densidades de frecuencia, lo que haremos en gráficos posteriores por comodidad.

Intervalos de distinta amplitud

Cuando los intervalos sean de distinta amplitud, las alturas de los rectángulos deben ser siempre igual a la densidad de frecuencia. En el ejemplo siguiente se observa que no todos los intervalos tienen la misma amplitud. El último intervalo tiene una amplitud de 20 y por tanto la altura del rectángulo es de 0'25 que es el resultado de dividir la frecuencia del intervalo, en este caso 5, entre la anchura del mismo, 20. Observa que el producto de la amplitud por la altura del rectángulo da la frecuencia indicada en la figura.

Ejemplo: **En un grupo de 40 alumnos se ha estudiado su talla, obteniéndose los resultados de la tabla adjunta.**

Intervalos en cm.	f_i
$[155,160)$	11
$[160,165)$	9
$[165,170)$	7
$[170,175)$	10
$[175,195)$	5
	42

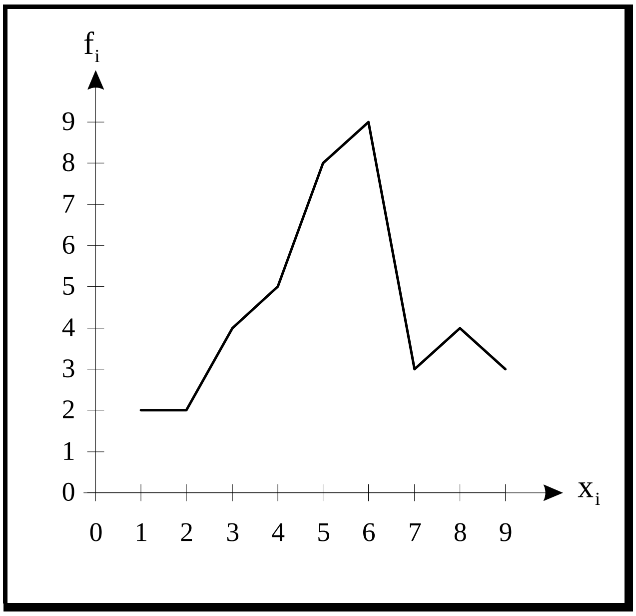


Polígono de frecuencias

Se considera polígono de frecuencias a la línea poligonal (línea quebrada) que une los puntos correspondientes a las frecuencias de cada valor de la variable estadística, o de los extremos superiores de las barras.

Ejemplo: **La tabla siguiente representa las calificaciones en la asignatura de Historia del arte de los 40 alumnos de una clase:**

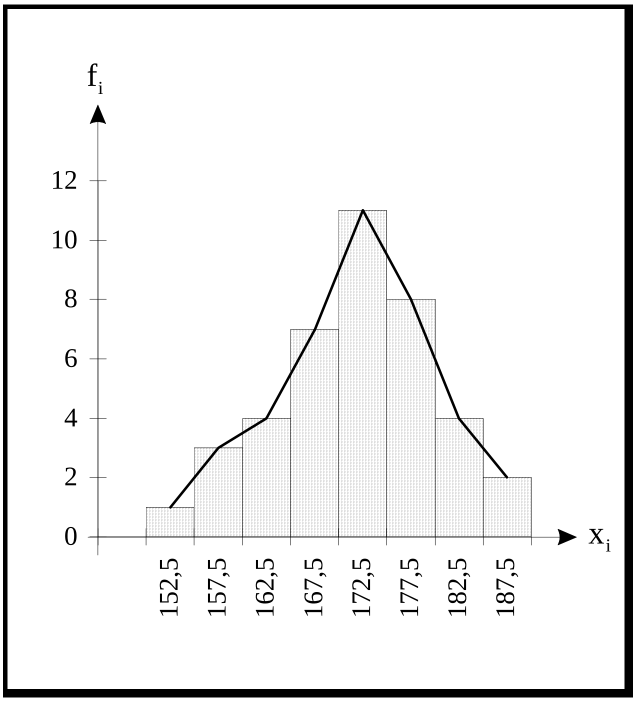
x_i	f_i
1	2
2	2
3	4
4	5
5	8
6	9
7	3
8	4
9	3
	40



Si los datos vienen dados en intervalos, unirá los puntos correspondientes a las marcas de clase.

Ejemplo: **En un grupo de 40 alumnos se ha estudiado su talla, obteniéndose los resultados de la tabla adjunta.**

Intervalos	x_i	f_i
$[150,155)$	152,5	1
$[155,160)$	157,5	3
$[160,165)$	162,5	4
$[165,170)$	167,5	7
$[170,175)$	172,5	11
$[175,180)$	177,5	8
$[180,185)$	182,5	4
$[185,190)$	187,5	2
		40



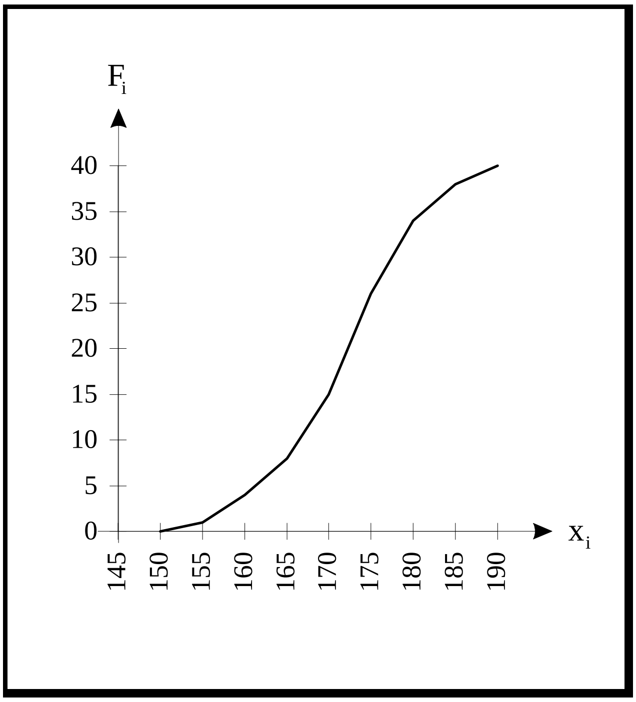
<u>Polígono de frecuencias acumuladas</u>

Los diagramas de frecuencias acumuladas, también llamados "diagramas en ojiva", expresan en abscisas los diversos valores de la variable estadística y en ordenadas las frecuencias acumuladas. Son útiles cuando se compara la forma de la curva resultante con la que representaría el crecimiento uniforme que sería una recta.

En el caso de un polígono de frecuencias acumuladas consideramos todo el intervalo, y no sólo la marca de clase.

Ejemplo: **En un grupo de 40 alumnos se ha estudiado su talla, obteniéndose los resultados de la tabla adjunta.**

Intervalos	x_i	f_i	F_i
$[150,155)$	152,5	1	1
$[155,160)$	157,5	3	4
$[160,165)$	162,5	4	8
$[165,170)$	167,5	7	15
$[170,175)$	172,5	11	26
$[175,180)$	177,5	8	34
$[180,185)$	182,5	4	38
$[185,190)$	187,5	2	40
		40	



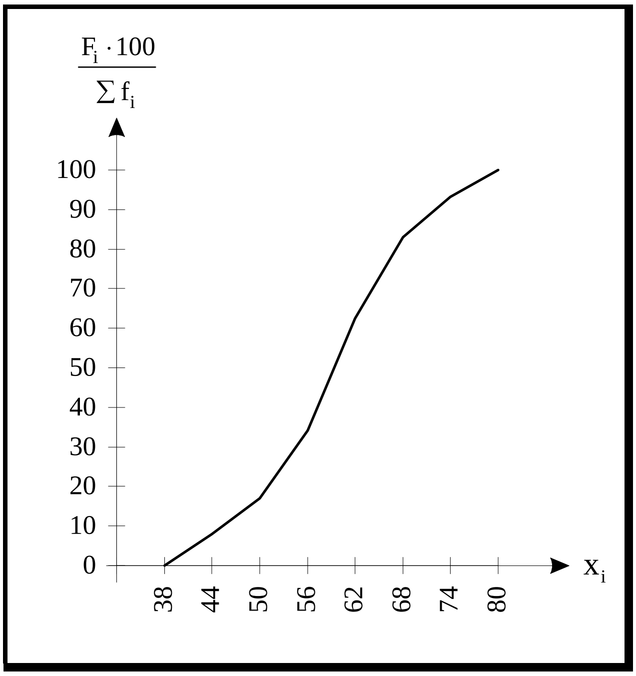
Polígono de porcentajes acumulados

Son útiles para calcular parámetros estadísticos como los percentiles. Los porcentajes acumulados se obtienen al multiplicar la frecuencia absoluta acumulada por el cociente entre 100 y la suma de todas las frecuencias absolutas.

$$\textbf{Porcentaje de frecuencias absolutas acumuladas} = F_i \cdot \frac{100}{\sum f_i}$$

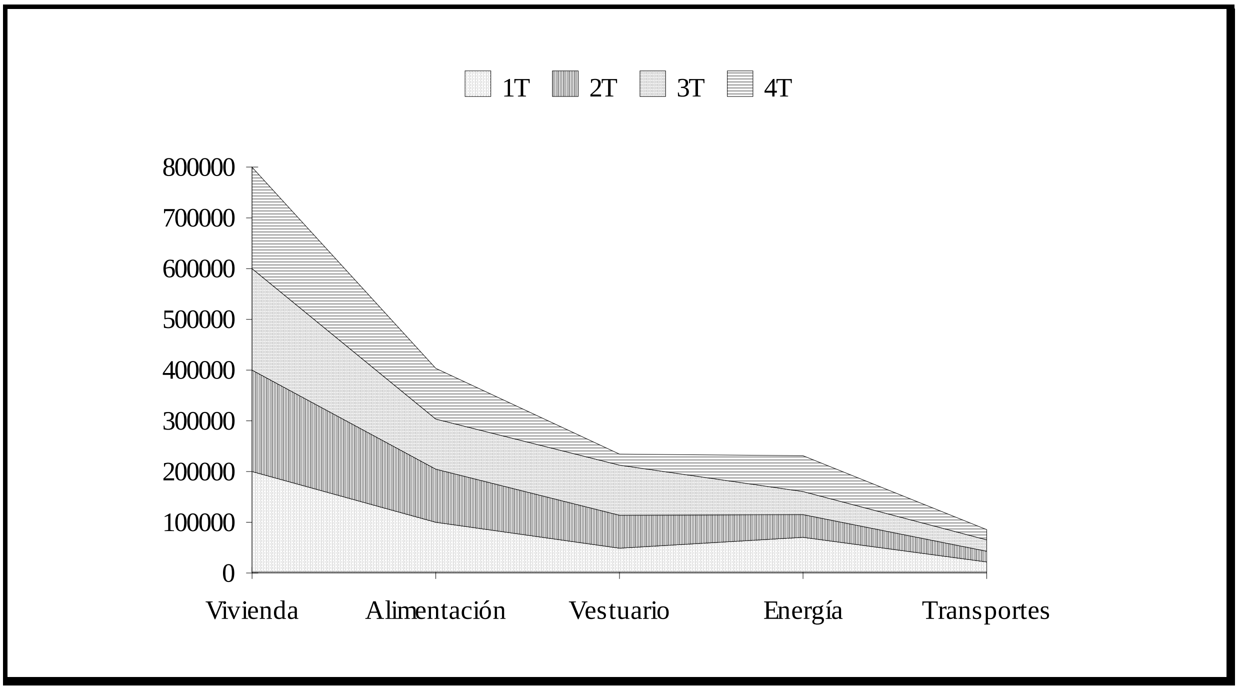
Ejemplo: **Se ha aplicado un test sobre satisfación en el trabajo a 88 empleados de una fábrica, obteniéndose los resultados de la tabla:**

Clases	f_i	F_i	Porcentajes $F_i \cdot \frac{100}{\sum f_i}$
$[38,44)$	7	7	7'95
$[44,50)$	8	15	17'04
$[50,56)$	15	30	34'09
$[56,62)$	25	55	62'5
$[62,68)$	18	73	82'95
$[68,74)$	9	82	93'18
$[74,80)$	6	88	100
	88		

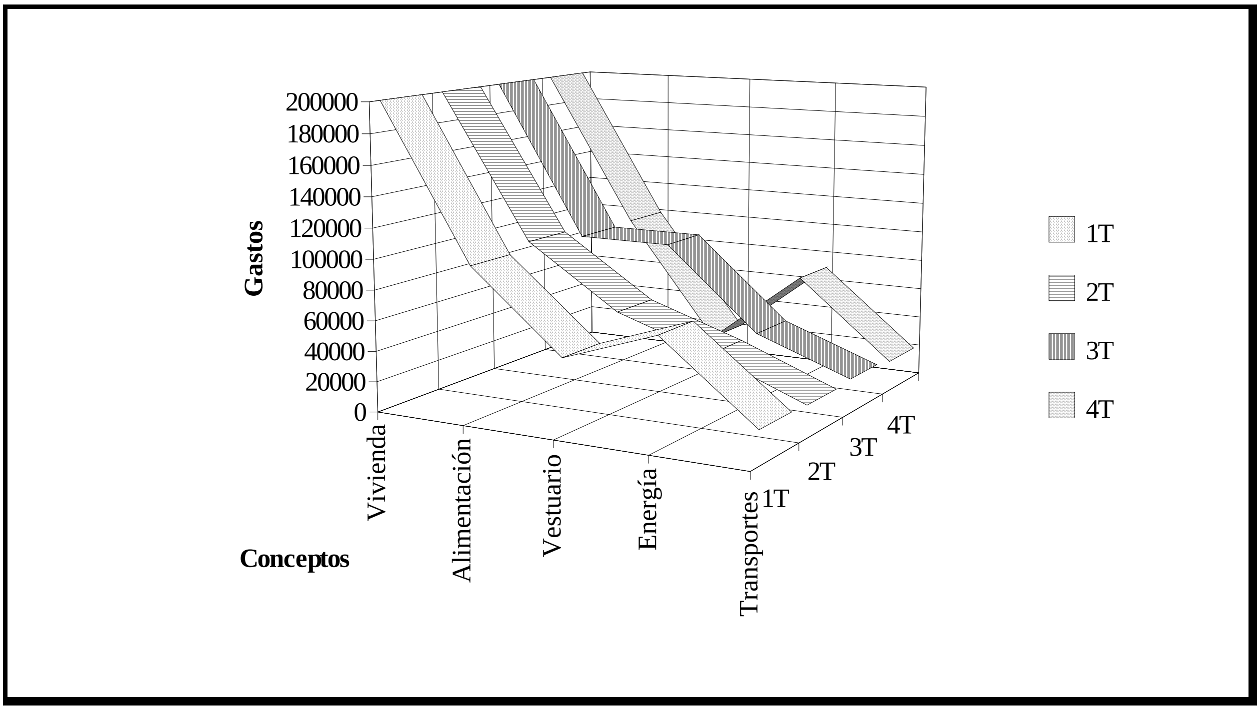


De forma análoga se pueden representar polígonos de *frecuencias relativas y relativas acumuladas*. **Algunas veces son interesantes los** gráficos de áreas**, que son gráficos construidos a partir de los gráficos de líneas, rellenando el espacio entre la línea y el eje de abscisas.** Representemos gráficamente los datos de la tabla siguiente

Gastos	1^{er} T	2^{o} T	3^{er} T	4^{o} T	Totales
Vivienda	200.000	200.000	200.000	200.000	800.000
Alimentación	100.000	105.000	98.500	99.999	403.499
Vestuario	49.000	65.000	98.500	22.000	234.500
Energía	70.500	45.000	45.000	70.800	231.300
Transportes	22.000	21.000	22.500	20.000	85.500
Total	441.500	436.000	464.500	412.799	1.754.799



Otras veces los gráficos tridimensionales de líneas ayudan a visualizar mejor la información.



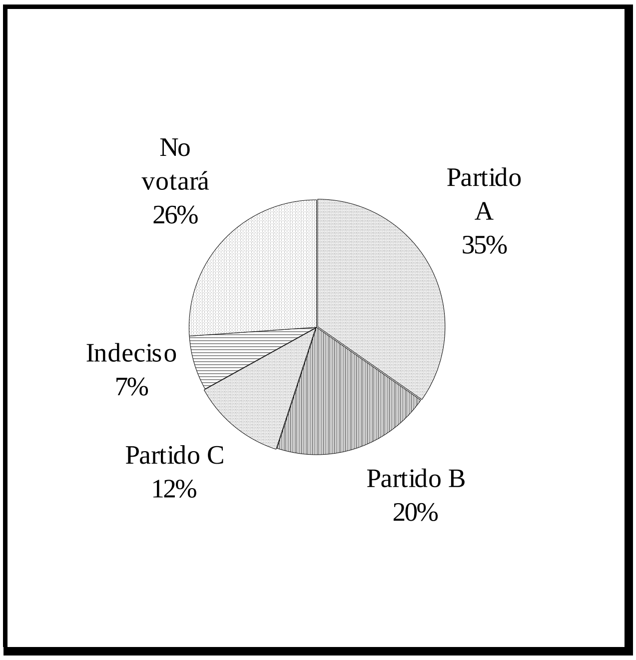
Diagramas de sectores

En los gráficos de diagramas de sectores ***cada suceso viene representado por un sector circular de amplitud proporcional a su frecuencia absoluta***. La amplitud de cada sector se obtiene mediante una sencilla regla de tres

Los diagramas de sectores son especialmente adecuados para representar las distintas partes de un todo y para representar varias situaciones similares y poder establecer comparaciones.

Ejemplo: **En un cierto país, la intención de voto en unas elecciones de sus habitantes se refleja en la siguiente tabla.**

Voto	Porcentaje
Partido A	**35%**
Partido B	**20%**
Partido C	**12%**
Indeciso	**7%**
No votará	**26%**

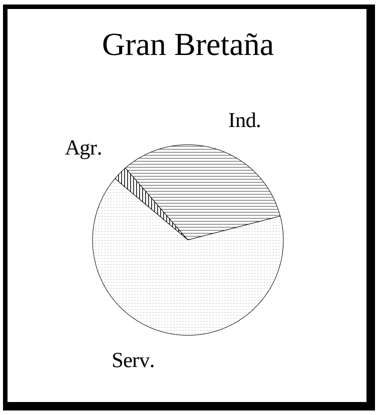


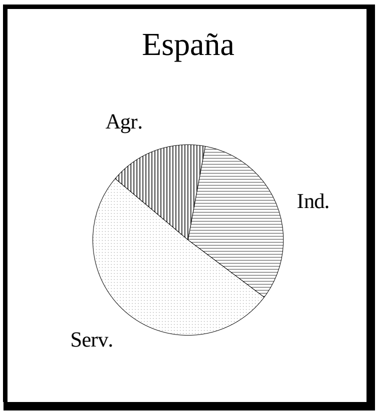
Frecuentemente los datos de una tabla se suelen dar en términos absolutos y con frecuencia se interpretan mejor en términos relativos. Por ejemplo, los datos que aparecen al margen se interpretan mejor si cada uno de los tres de abajo se expresa como porcentaje del de arriba.

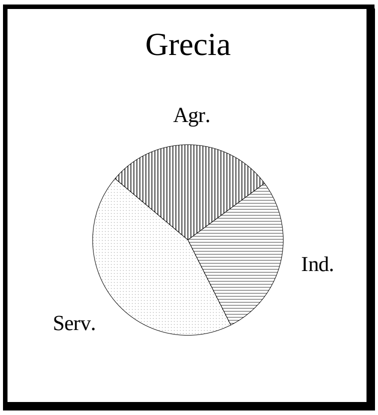
	Gran Bretaña	**España**	**Grecia**
Empleo civil	**24'07**	**10'42**	**3'59**
Agricultura	0'62	1'77	1'04
Industria	7'80	3'35	0'98
Servicios	15'64	5'31	1'57

Así, para el caso de Gran Bretaña tenemos:

$$\text{Agri.} \rightarrow \frac{0'62\cdot 100}{24'07} \cong 2'6\% \qquad \text{Ind.} \rightarrow \frac{7'80\cdot 100}{24'07} = 32'40 \qquad \text{Serv.} \rightarrow \frac{15'64\cdot 100}{24'07} = 64'97$$







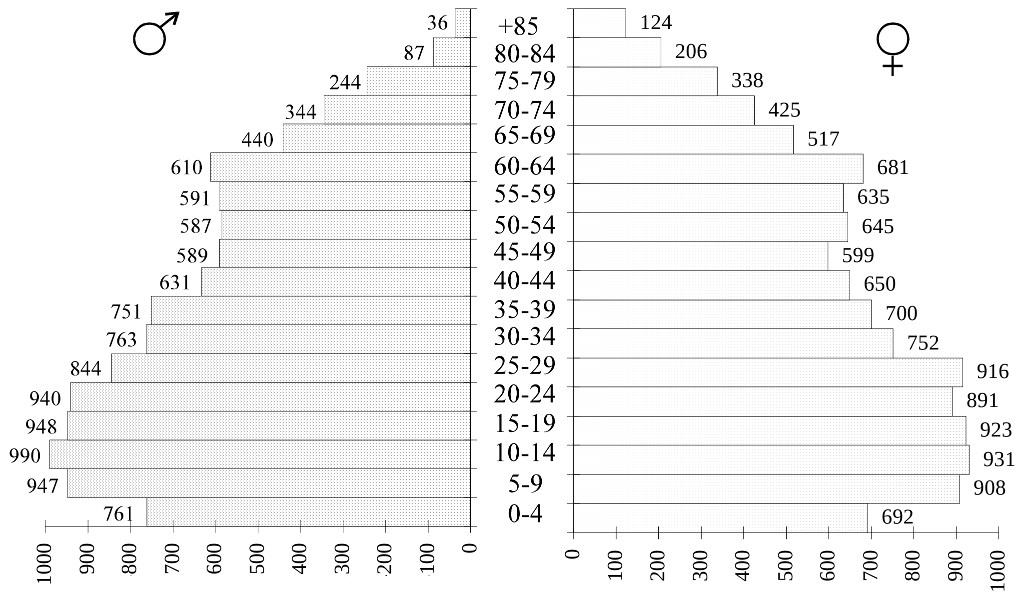
Pirámides de población

Son un caso particularmente importante de histogramas. En realidad ***se trata de dos histogramas de distribución de edades, uno para hombres y otro para mujeres. Se utilizan para estudiar conjuntamente la variable edad y el atributo sexo.*** En el eje de ordenadas se representa los intervalos de edades cuya anchura puede ser anual, quinquenal o decenal, dependiendo del detalle necesario. En el eje de abscisas se representa el sexo. Para la modalidad mujer se toma el semieje positivo, y para la modalidad hombre el semieje negativo.

En la pirámide de población es posible analizar la distribución de la población por edad y sexo, deducir las vicisitudes sufridas (guerras, catástrofes, etc.), conocer el comportamiento demográfico (control o no de la natalidad), conocer el desarrollo de la población, etc. Asimismo, permite realizar previsiones sobre el futuro.

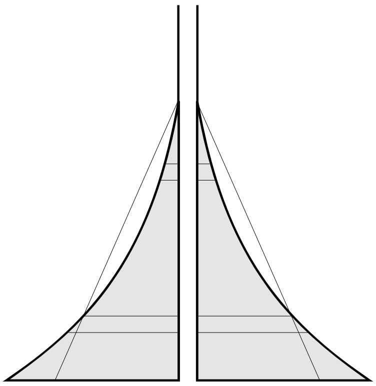
A continuación se representa la pirámide de población de la ciudad de Denia correspondiente al año 1988.

Nacidos en Denia. Datos de 1988



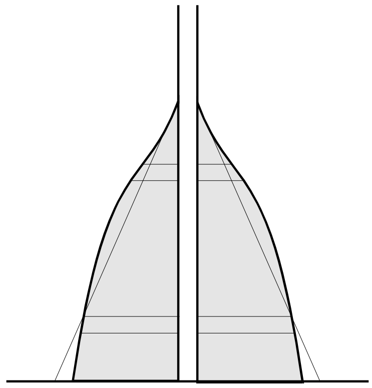
Según la forma de la pirámide de población, se puede deducir si se trata de una población eminentemente joven (perfil de pagoda), madura (perfil de bulbo) o vieja (perfil de as de pique). A continuación se representan tres ejemplos de pirámides de población que responden a cada uno de estos casos.

Pirámide de población joven



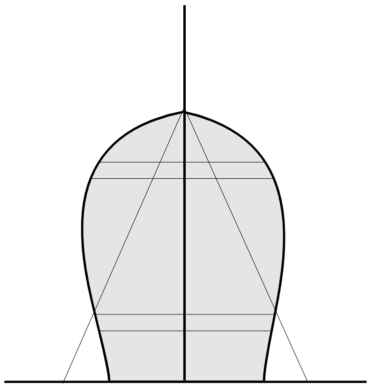
Tipo Pagoda

Pirámide de población madura



Tipo Bulbo

Pirámide de población vieja



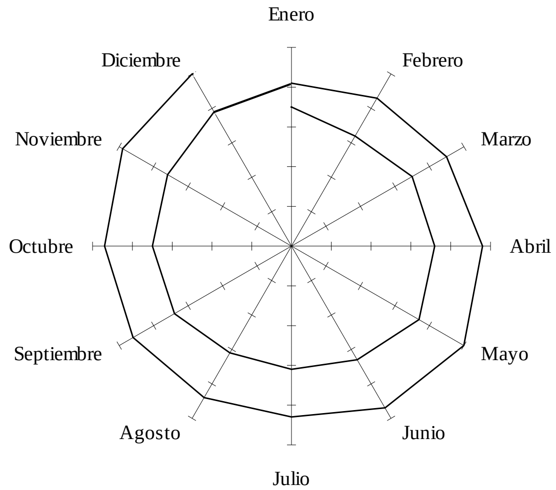
Tipo As de Pique

Gráficos en espiral o de radar

Se utiliza para representar series de tiempo con fuerte tendencia a la expansión.

Ejemplo: Aquí aparece representada la evolución del precio (en pts.) del kilo de miel, mes a mes, durante dos años.

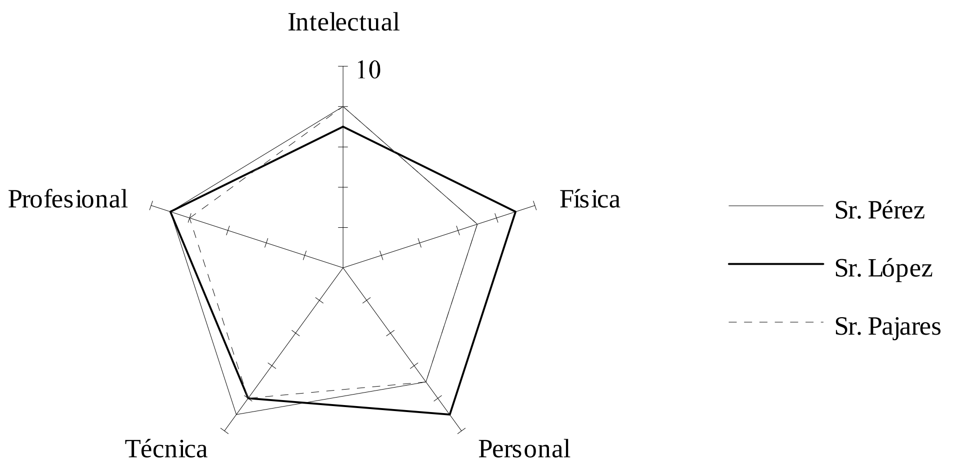
Este tipo de gráfico pone muy claramente de manifiesto la fuerza expansiva del fenómeno, así como las fluctuaciones periódicas. (Por ejemplo, el precio de la miel tiende a subir en algunos meses como diciembre y mayo, y tiende a bajar en junio, inmediatamente después de la recolección).



Ejemplo: En la tabla de abajo se muestran los resultados de las pruebas efectuadas por una consultora para elegir el gerente de una empresa. Las pruebas efectuadas por los tres candidatos son de distintos tipos (pruebas intelectuales, físicas, de personalidad, etc.).

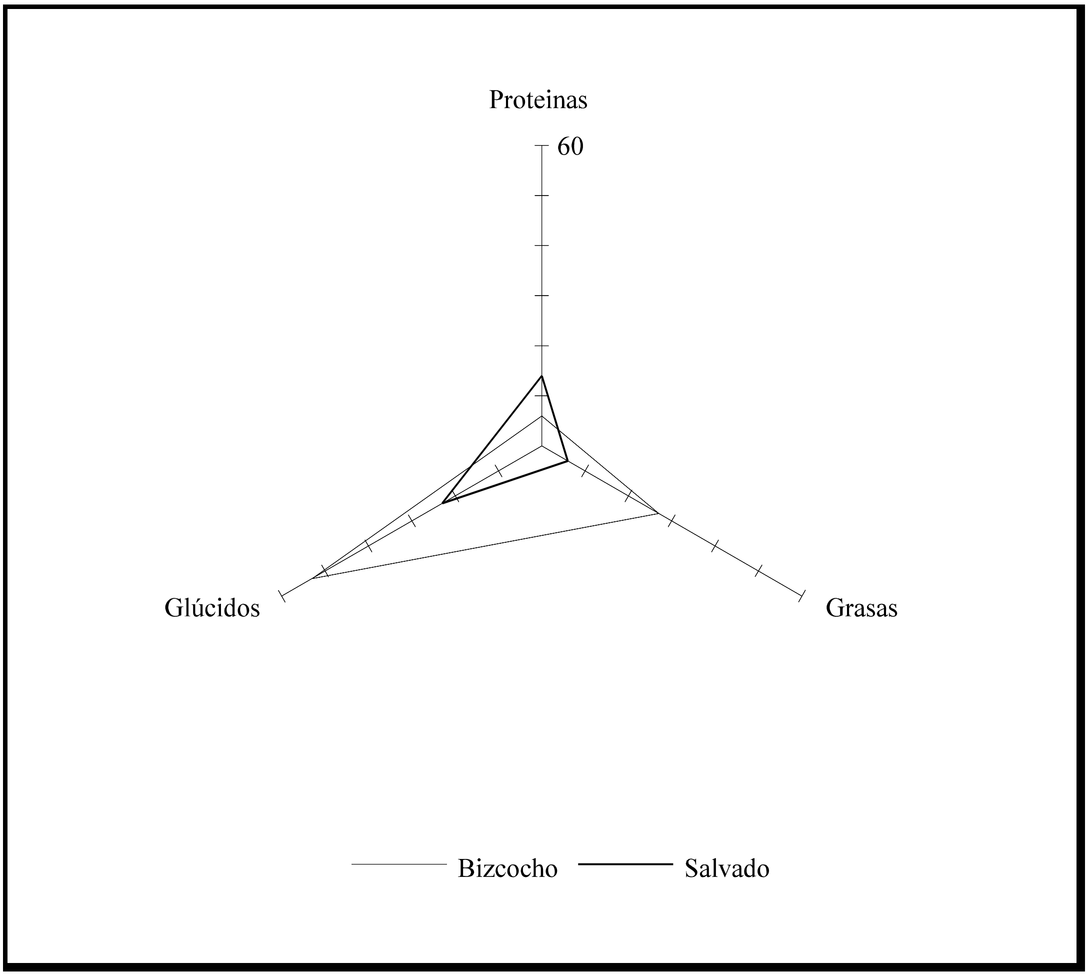
En el gráfico de radar representado se aprecia que el Sr. López destaca sobre los otros dos candidatos por sus aptitudes físicas y personales. Puede apreciarse también que en los aspectos intelectual, técnico y profesional, los tres candidatos están prácticamente al mismo nivel.

	Intelectual	Física	Personal	Técnica	Profesional	Promedio
Sr. Pérez	8	7	7	9	9	8
Sr. López	7	9	9	8	9	8'4
Sr. Pajares	8	7	7	8	8	7'6
Promedio	7'7	7'7	7'7	8'3	8'7	8



La tabla adjunta pone de relieve los principios inmediatos existentes en 100 gr. de bizcocho y del salvado de trigo. Vamos a estudiar a través de un gráfico de radar el valor nutritivo del bizcocho y del salvado.

	Proteinas	Grasas	Glúcidos
Bizcocho	6	27	53
Salvado de trigo	14	6	23



Pictogramas

Los **pictogramas** ***son gráficos en los que utilizan dibujos alusivos a la distribución que se pretende estudiar, y que mediante su forma, tamaño, etc., ofrecen una descripción lo más expresiva posible de la distribución estadística***

Cartogramas

Se llama **cartograma** ***a los gráficos que se realizan sobre un mapa, señalando sobre determinadas zonas con distintos colores o rayados lo que se trata de poner de manifiesto***.

Por ejemplo, se suelen utilizar este tipo de diagramas para representar la densidad demográfica de una nación, la renta per cápita, las horas de sol anuales sobre una determinada parte de la tierra, los índices de lluvias de una nación, etc.

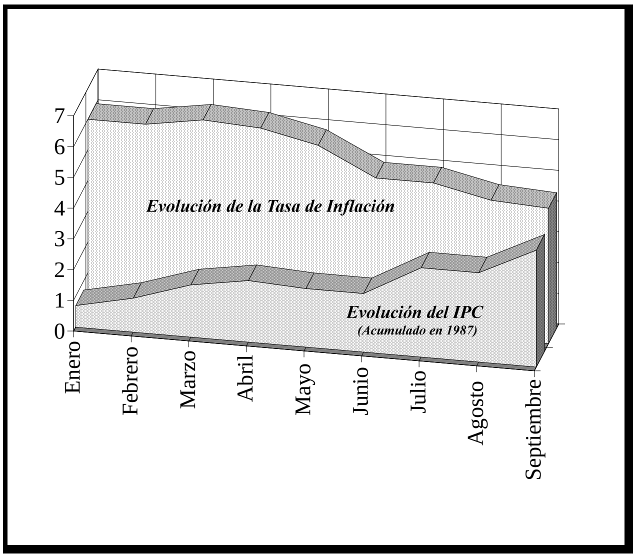
Series cronológicas

Se llama **serie cronológica** a un conjunto de observaciones hechas a lo largo del tiempo, normalmente a intervalos iguales. Es, por tanto, una serie estadística en la cual la variable independiente es el tiempo

Ejemplo de series cronológicas son: La producción de acero en España durante los últimos veinte años, las temperaturas diarias medidas a la misma hora y en el mismo lugar todos los días de un año y el gasto mensual de energía eléctrica para el alumbrado de las calles de una ciudad, durante cinco años.

La variable cuya evolución temporal se estudia puede ser un **flujo** o un **nivel**. ***Flujo, si la cantidad es acumulativa***; por ejemplo, producciones, gastos, etc. Las observaciones se refieren a un período: producción durante cada año mensual..., cantidad de agua de lluvia por metro cuadrado en cada semana, etc. ***Nivel, si la observación se refiere a un instante***; temperatura, número de obreros que se encuentran en paro el primer día de cada mes, cantidad de agua almacenada en un pantano en ciertos instantes.

Aquí se reproduce el diagrama que expresa la evolución del IPC y la Tasa de Inflación durante los nueve primeros meses del año 1987.



Números índices

Para la descripción de las series de tiempo se utilizan, muy frecuentemente, sobre todo en economía, los **números índices**. Que el índice de la bolsa ha alcanzado el valor 143 significa que, globalmente, el valor de las acciones ha sido 143/100, es decir 1'43 veces el valor que tenían en un cierto instante que se toma como referencia, y que suele ser el día primero del año en cuestión. Esto supone hacer unos reajustes anuales por los cuales cada primer año el índice de la bolsa vuelve a valer 100.

La comparación de cifras en series de datos paralelos se hace tarea tediosa en múltiples ocasiones. Se recurre con frecuencia a los números índices en Estadística para facilitar esta tarea.

La siguiente tabla nos proporciona, en kg/persona, el diferente consumo de seis alimentos básicos en España en el período 1965-1980.

	1965	1970	1975	1980
Carnes	28'1	45	61'2	69'3
Cereales panificables	92'4	76'2	79'7	75'8
Huevos	10'2	11'6	16'2	17'3
Leche y productos lácteos	64'4	86'4	104'2	113'4
Leguminosas	9'9	7	7'4	5'6
Patatas	104'8	110	111'3	113'1

La tabla siguiente resalta mejor las comparaciones relativas en la evolución del consumo de cada alimento. Se fija el índice 100 para cada producto en el primer año considerado y se alternan los restantes datos de la tabla en la misma proporción.

	1965	1970	1975	1980
Carnes	**100**	160	218	247
Cereales panificables	**100**	82	86	82
Huevos	**100**	114	159	170
Leche y productos lácteos	**100**	134	161	176
Leguminosas	**100**	71	75	57
Patatas	**100**	105	106	108

Por ejemplo: $160=\dfrac{45}{28'1}\cdot 100;\qquad 82=\dfrac{76'2}{92'4}\cdot 100;\qquad 108=\dfrac{113'1}{104'8}\cdot 100;\qquad 105=\dfrac{110}{104'8}\cdot 100$

Problema 1

Un grupo de alumnos discutían sobre la altura de los chicos y las chicas que comenzaban a estudiar primero de BUP en el Instituto. Decidieron elegir una muestra de 40 alumnos para su encuesta. Estaría compuesta por 20 chicos y 20 chicas elegidos por sorteo. Después de preguntar a cada uno de ellos su estatura se obtuvieron los siguientes datos:

Chicas: 154, 161, 158, 170, 155, 153, 156, 160, 159, 162, 160, 163, 165, 166, 159, 158, 166, 170, 157, 155.

Chicos: 162, 160, 163, 158, 166, 159, 161, 160, 167, 170, 155, 158, 169, 154, 161, 166, 162, 166, 155, 167.

a) Agrupa los datos de ambos grupos en intervalos de clase y realiza una tabla de frecuencias para los mismos.

b) Realiza a partir de las tablas de frecuencias los histogramas correspondientes a las alturas para los chicos y las chicas del estudio. Deduce de los histogramas correspondientes las características de la distribución de alturas en ambos sexos y, si es posible, concluye quiénes son más altos, los chicos o las chicas de ésta edad.

a) Como los datos son 20 tanto para las chicas como para los chicos el número de intervalos que formaremos será de $\sqrt{20} = 4'58$, es decir de 5.

Se observa que la chica más alta mide 170 cm. y la más baja 153 cm. La diferencia es 17 por tanto el entero más próximo a 17 que es divisible por 5 es 20. Como $\frac{20}{5} = 4$, la amplitud de los intervalos será de 4. Por otro lado, el chico más alto mide 170 cm. y el más bajo 154 cm, por tanto la diferencia es 16 y el entero más próximo que es divisible entre 5 es 20. Como $\frac{20}{5} = 4$, la amplitud de los intervalos será de 4.

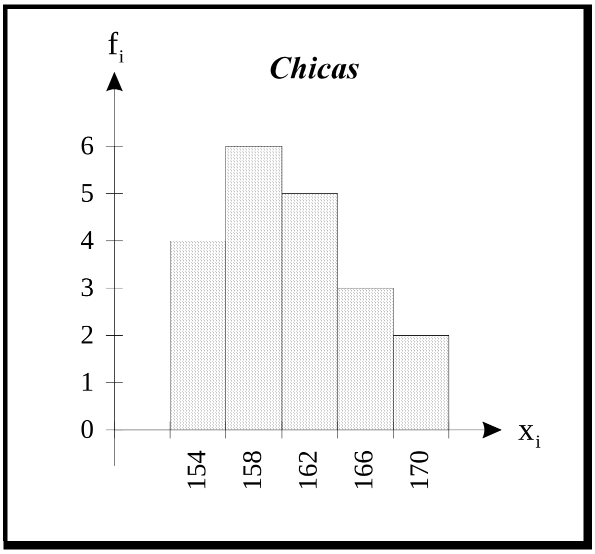
CHICAS

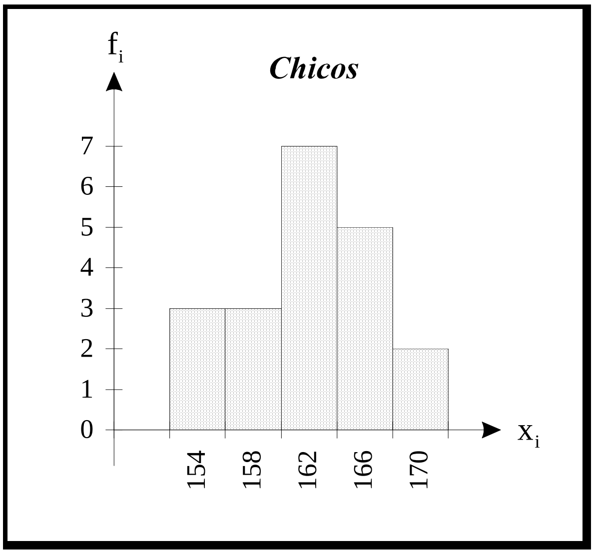
Intervalos	**Marcas**	**f_i**	**f_r**	**%**	**F_i**	**F_r**
152 – 156	154	4	0'2	20	4	0'2
156 – 160	158	6	0'3	30	10	0'5
160 – 164	162	5	0'25	25	15	0'75
164 – 168	166	3	0'15	15	18	0'90
168 – 172	170	2	0'10	10	20	1
		20	1	100		

CHICOS

Intervalos	Marcas	f_i	f_r	%	F_i	F_r
152 – 156	154	3	0'15	15	3	0'15
156 – 160	158	3	0'15	15	6	0'30
160 – 164	162	7	0'35	35	13	0'65
164 – 168	166	5	0'25	25	18	0'90
168 – 172	170	2	0'10	10	20	1
		20	1	100		

b)



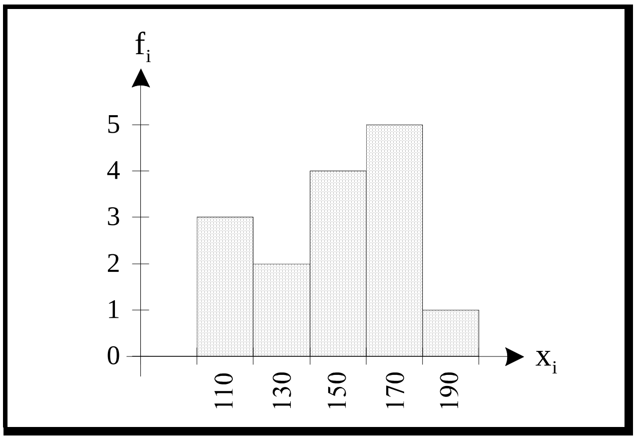


Problema 2

El siguiente histograma corresponde al salario, en miles de pesetas, de los empleados de una empresa

a) Calcula la tabla de frecuencias.

b) Dibuja el diagrama de sectores correspondientes.



a) La anchura de los intervalos es constante e igual a 20. Así pues, dependiendo de la altura de la columna se obtendrán las frecuencias para cada intervalo.

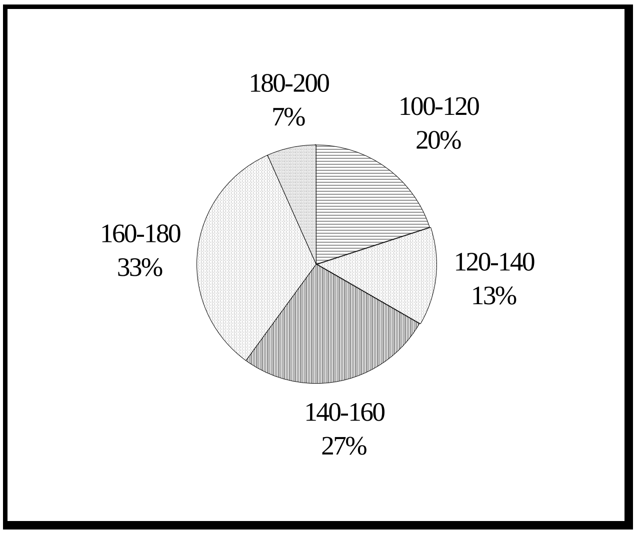
El cálculo correspondiente se encuentra en tabla adjunta.

Intervalos	Alturas	Frecuencias
$100-120$	3	$3\cdot 20 = 60$
$120-140$	2	$2\cdot 20 = 40$
$140-160$	4	$4\cdot 20 = 80$
$160-180$	5	$5\cdot 20 = 100$
$180-200$	1	$1\cdot 20 = 20$
		300

.

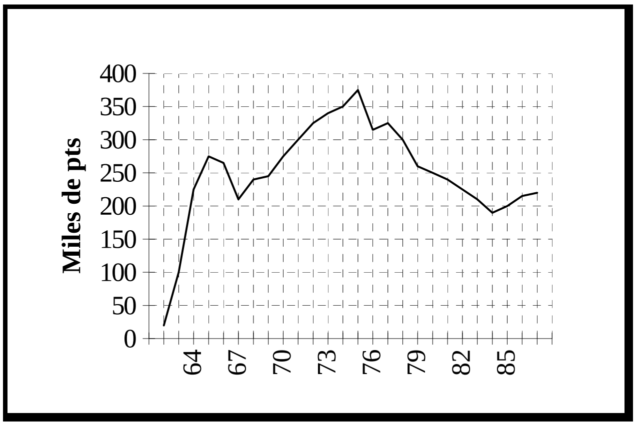
b)

Intervalos	f_i	f_r	%
$100-120$	60	0'2	20
$120-140$	40	0'1333	13'33
$140-160$	80	0'2666	26'66
$160-180$	100	0'3333	33'33
$180-200$	20	0'0666	6'66
	300	1	100



Problema 3

Comenta brevemente la figura adjunta que representa el número de viviendas construidas en España en el período 1961-1988.



Para este período se observa:

- Un crecimiento desde 1961 a 1965, superando en este último año las 275.000 viviendas construidas.
- Una bajada desde 1965 hasta 1967, año en el que se construyen algo más de 200.000 viviendas; 75.000 menos que dos años antes.

- Desde 1967 y hasta 1975 se observa una subida sostenida, alcanzándose para ese último año la cifra récord del período estudiado, ya que se construyen 375.000 viviendas.

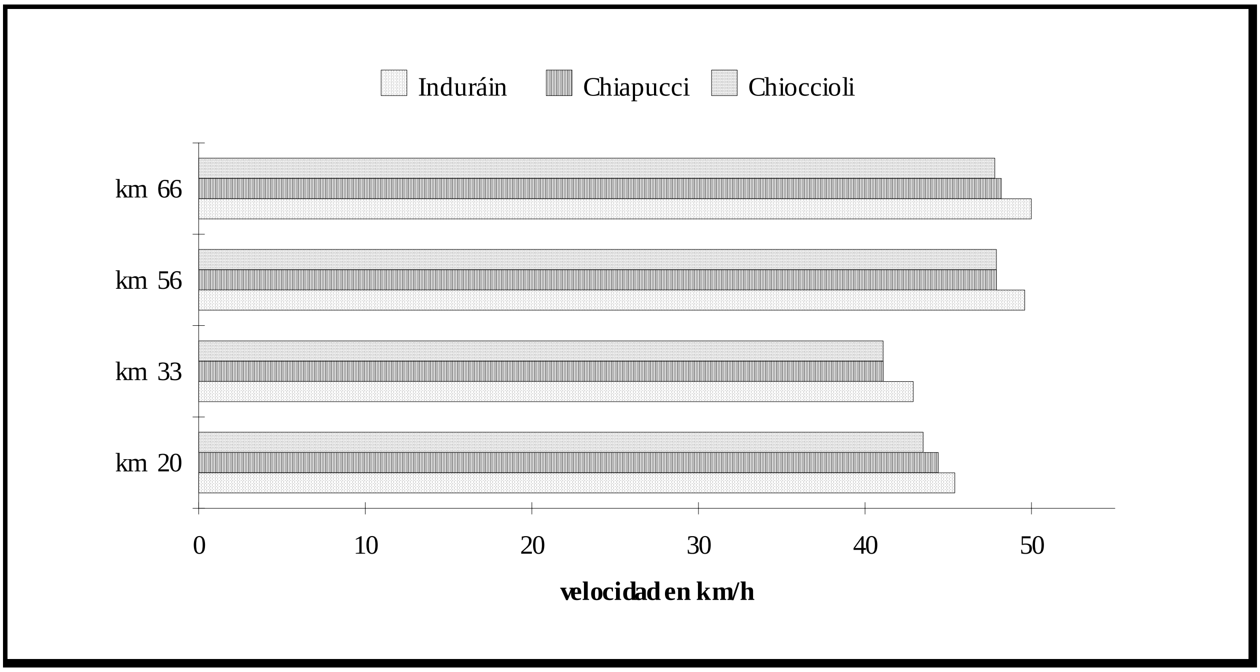
- Desde 1975 se acusa una bajada continua, hasta que en 1984 aparecen síntomas de recuperación.

Problema 4

Representa, gráficamente, mediante un diagrama de barras horizontales, las velocidades medias de la carrera contrarreloj de el *Giro* de Italia. Compara las velocidades de los tres ciclistas.

Velocidad media (km./h)

	km 20	**km 33**	**km 56**	**km 66**
Induráin	45'4	42'9	49'6	50'0
Chiapucci	44'4	41'1	47'9	48'2
Chioccioli	43'5	41'1	47'9	47'8



Problema 5

Se ha medido el perímetro craneal a niños de edades comprendidas entre 2 y 3 años, y se han obtenido los siguientes valores (en cm.):

41, 39'5, 43'2, 40'5, 42'3, 44'5, 38'5, 42'5, 40'3
45'6, 40'1, 43'5, 40'2, 42'7, 45, 45'2, 47'9, 44'2.

a) Agrupar los datos en intervalos de longitud 2 cm.

b) Obtén la tabla de frecuencias para los intervalos definidos.

c) Construye el histograma de frecuencias absolutas y el polígono de frecuencias acumuladas.

a) El perímetro craneal varía entre 38'5 y 47'9, por tanto la diferencia es 9'4. Como la longitud debe ser 2, comenzamos en 38 y acabamos en 48.

b)

Intervalo	Marcas	f_i	F_i
38 – 40	39	2	2
40 – 42	41	5	7
42 – 44	43	5	12
44 – 46	45	5	17
46 – 48	47	1	18
		18