XUNTA DE GALICIA
CONSELLERÍA DE EDUCACIÓN
E ORDENACIÓN UNIVERSITARIA
Dirección Xeral de Educación, Formación
Profesional e Innovación Educativa

Educación secundaria
para personas adultas

Ámbito científico tecnológico

Educación a distancia semipresencial

Módulo 3

Unidad didáctica 4

Estadística

Índice

1. Introducción

1.1 Descripción de la unidad didáctica

En esta unidad estudiaremos la estadística. En la estadística se trabaja con una gran, grandísima, cantidad de datos. En esta unidad veremos cómo organizar esos datos para proceder a su posterior estudio y ser capaces de sacar conclusiones sobre los datos con los que trabajamos.

Veremos también las representaciones gráficas más usuales en estadística y su interpretación.

1.2 Conocimientos previos

Para trabajar con esta unidad es necesario recordar los conceptos sobre estadística estudiados en el módulo anterior, en especial:

- Conceptos de población, muestra e individuo.
- Concepto y tipos de variables estadísticas.
- Frecuencia absoluta y frecuencia relativa. Organización de los datos en tablas.
- Representaciones gráficas: diagramas de barras, diagramas de sectores y polígonos de frecuencias.
- Medidas de centralización: media, moda y mediana.

1.3 Criterios de evaluación

- Elaborar informaciones estadísticas para describir un conjunto de datos mediante tablas y gráficas adecuadas a la situación analizada y justificar si las conclusiones son representativas para la población estudiada.
- Calcular e interpretar los parámetros de posición y de dispersión de una variable estadística para resumir los datos y comparar distribuciones estadísticas.

2. Secuencia de contenidos y actividades

2.1 La estadística. Fases y tareas de un estudio estadístico

La necesidad de recontar datos sobre las diferentes poblaciones surgió al aparecer sociedades organizadas como, por ejemplo, en China, Egipto o Roma.

Con el transcurrir del tiempo se crearon nuevas necesidades que no se limitaban solo al campo social. Aparece la conveniencia de hacer estudios de tipo estadístico en las distintas ramas del saber, como pueden ser la economía o la biología.

De este modo, a partir del siglo XIX se aplicó la estadística a distintos campos, dando lugar a la aparición de nuevas especialidades y subdivisiones de ellas: la econometría, la bioestadística (también llamada biometría)...

En la actualidad las técnicas estadísticas se utilizan en casi todos los aspectos de la vida. Se diseñan encuestas para recoger información previa a las consultas electorales, se realizan muestreos sobre las características de calidad de un producto, se sondea a los consumidores para indagar si un producto que puede salir a la venta tendrá éxito o no, se realizan experimentos y mediciones para determinar el modo más apropiado de evitar el deterioro del medio ambiente, etcétera.

Hoy en día, la estadística es una ciencia que se encarga de estudiar una determinada característica de un conjunto de individuos recogiendo, analizando e interpretando datos.

La estadística se divide en dos ramas principales:

- **Estadística descriptiva o deductiva**: trata del recuento, la ordenación y la clasificación de los datos obtenidos en las observaciones. Se construyen tablas y se representan en gráficos, que permiten simplificar en gran medida la complejidad de los datos que intervienen en la distribución. A partir de los datos se obtienen los parámetros estadísticos que caracterizan la distribución. Esta parte de la estadística se limita a realizar deducciones directamente a partir de los datos y los parámetros obtenidos.
- **Estadística inferencial o inductiva**: formula y resuelve el problema de establecer previsiones y deducciones generales sobre una población a partir de resultados obtenidos de una muestra. Utiliza resultados obtenidos mediante la estadística descriptiva y se apoya fuertemente en el cálculo de probabilidades.

En un típico trabajo estadístico, se procede a la recogida de datos que se representan gráficamente para facilitar su visión global y se calculan ciertos parámetros para caracterizar su tendencia central (media, moda, mediana...) y otros para indicar su dispersión (varianza, desvío típico...). Posteriormente, se intenta ver si los datos se pueden vincular con algún modelo estadístico para poder obtener más información sobre el objeto del estudio.

La información estadística llega a nosotros mediante gráficas o tablas muy bien construidas, con las que resulta fácil entender la información dada. Pero para llegar a ellas, hay que realizar un largo proceso que se inicia ahora.

- *¿Qué queremos estudiar?* Necesitamos saber lo que pretendemos estudiar, por ejemplo, qué aficiones deportivas tienen los alumnos y las alumnas de un centro.
- *Selección de las variables que se van a analizar*. Debe ser evidente cuál es la variable y cuáles sus posibles valores.
- *Recogida de datos*. Se efectúan las medidas o se realizan las encuestas.
- *Organización de datos*. Se ordenan, se pasan a papel, o mejor, se introducen en el computador.
- Los pasos siguientes son la elaboración de tablas y gráficas y el cálculo de parámetros, a los que dedicaremos el resto de la unidad.

Actividades propuestas

S1. Se quiere hacer una encuesta para estudiar la afición de la gente joven a la lectura. Diga, justificadamente, cuáles de las preguntas siguientes le parecen razonables y cuáles no.

- Di cuáles son tus lecturas favoritas.
- De los géneros literarios siguientes, señala los que leíste más de una hora el último mes.

☐ Novela ☐ Historia ☐ Poesía ☐ Teatro ☐ Filosofía ☐ Biografía ☐ Cómic

- ¿Lees periódicos? Si es así, ¿de qué tipo?
- ¿A cuál o cuáles de las siguientes publicaciones periódicas dedicas más de dos horas semanales?:

☐ Periódicos de información general ☐ Periódicos deportivos ☐ Revistas científicas

☐ Revistas del corazón ☐ Otros (indíquese cuáles)

2.2 Población y muestra

2.2.1 Conceptos básicos

Como acabamos de ver en el apartado anterior, la estadística estudia una o más características de un grupo de individuos, al que se llama población, para obtener resultados y conclusiones y hacer predicciones sobre el comportamiento de la población con respecto a las características estudiadas.

Se llama **población** al conjunto de personas, objetos o individuos sobre los que se quiere estudiar una determinada característica. Cada miembro de la población se llama **individuo**.

Si, por ejemplo, queremos estudiar el número de piezas defectuosas fabricadas por un taller en un día, la población serían las piezas fabricadas en ese taller. Cada una de las piezas fabricadas sería un individuo.

A veces resulta prácticamente imposible observar todos los individuos de la población, en especial cuando son muy numerosos. En lugar de examinar toda la población, lo que se hace es examinar una pequeña parte de ella a la que llamamos **muestra**.

Las muestras deben ser representativas de la población, por lo que tienen que cumplir dos requisitos: tienen que ser aleatorias y tienen que ser representativas de la población que representan, esto es, tienen que presentar las mismas características.

2.2.2 Métodos de selección de muestras. Representatividad

A veces, cuando queremos estudiar una característica de una población, resulta muy costoso o no es posible estudiar todos los miembros de esa población. Así, por ejemplo, si queremos conocer el peso medio de las merluzas del Cantábrico, no parece recomendable la estrategia de capturarlas todas y pesarlas. En situaciones de este tipo y de otras semejantes, la mejor solución es hacer el estudio sobre una parte de los individuos de la población, es decir, seleccionar una muestra.

Analicemos la siguiente situación. En un pueblo hay 50 000 viviendas habitadas y el ayuntamiento quiere saber cuál es el porcentaje de viviendas que tienen lavavajillas. Para eso, escoge al azar 400 viviendas y un empleado del ayuntamiento va casa por casa preguntando si hay o no lavavajillas en la casa. El entrevistador establece que 252 de las viviendas escogidas tenían lavavajillas y 148 no lo tenían.
En este supuesto, la población está formada por 50 000 individuos, que son las viviendas habitadas. La muestra es el conjunto de viviendas escogidas, tiene un tamaño de 400 individuos.

Como en la muestra encontraron 252 viviendas con lavavajillas sobre 400, para obtener la proporción, en tanto por ciento, de las viviendas que tienen lavavajillas, hacemos simplemente:

$$\frac{252}{400} \times 100 = 63\ \%$$

Por lo que el ayuntamiento tiene un indicio razonable de que el porcentaje de viviendas que dispone de lavavajillas será, aproximadamente, de un 63 %.

Pero resulta evidente que estos pequeños cálculos no resuelven de modo inapelable el problema; surgen preguntas del tipo de: "¿hasta qué punto podemos suponer que un porcentaje obtenido en la muestra se puede generalizar a toda la población?"

La búsqueda de respuestas a este tipo de preguntas lleva a analizar los distintos tipos de muestreos posibles y su representatividad.

Ahora vamos a describir resumidamente los tipos de muestras más usados:

Muestreo aleatorio simple. Todas las muestras del mismo tamaño que se pueden obtener sobre la población tienen la misma probabilidad de salir. Para llevar a cabo un muestreo aleatorio simple, lo más habitual es elaborar un listado con todos los individuos de la población, asignando un número a cada uno de ellos. A continuación se escoge la muestra por sorteo o usando unas tablas especiales en las que los números aparecen al azar, sin seguir ninguna tendencia, llamadas tablas de números aleatorios.

<u>Ejemplo</u>. Una empresa de 500 trabajadores desea saber cuántos de ellos vienen caminando a trabajar. Después de elaborar una lista numerada de los trabajadores, introduce en un bombo 500 bolas numeradas del 1 al 500 y extrae 25. A continuación entrevista a los 25 trabajadores escogidos para saber si van caminando al trabajo. Esto es un muestreo aleatorio simple.

Muestreo aleatorio con reposición. Consiste en obtener una muestra aleatoria simple, pero aunque un individuo resultara escogido una vez para formar parte de la muestra, puede ser escogido otras veces. El resultado final será una muestra en la que cada individuo puede aparecer más de una vez. Para poblaciones grandes, mucho más grandes que el tamaño de la muestra final, no existen diferencias significativas entre este método y el anterior, porque la probabilidad de obtener un individuo repetido es prácticamente nula.

Ejemplo. Para saber la proporción de machos y hembras en la población de ranas de un lago, un grupo de biólogos sigue el siguiente procedimiento: captura una rana, comprueba su sexo y la devuelve al lago; realiza este experimento 100 veces. Esto es una muestreo aleatorio con reposición.

Muestreo estratificado. Se divide la población en unos pocos grupos grandes (estratos) y se hace muestreo aleatorio simple en cada uno de ellos.

Ejemplo:. En un ayuntamiento hay tres institutos y el número de alumnos en cada uno de ellos es similar. El ayuntamiento estudia la posibilidad de organizar una actividad conjunta para los tres centros y quiere conocer el grado de aceptación de esta iniciativa entre el alumnado. Para eso, escoge al azar 50 alumnos en cada uno de los institutos y entrevista a los 150 estudiantes. Esto es un muestreo estratificado.

Muestreo por conglomerados. Se divide a la población en muchos grupos pequeños (conglomerados). A continuación, se elabora un listado de esos grupos y, con ellos, se hace muestreo aleatorio simple para obtener una muestra de conglomerados y se estudian todas las unidades de los conglomerados escogidos.

Ejemplo. Supongamos que, en el ejemplo anterior, el ayuntamiento adopta una estrategia de sondeo totalmente distinta: elabora un mapa de distritos de su territorio de modo que en cada distrito haya, aproximadamente, 25 estudiantes cursando estudios en algún instituto. A continuación, escoge por sorteo 6 de esos distritos y entrevista a todos los alumnos que se encuentran en las zonas escogidas. Esto es un muestreo por conglomerados.

Muestreo sistemático. Se escoge una unidad en particular, puede ser al azar, y todas las demás unidades de la muestra son obtenidas mediante un método preestablecido.

Ejemplo. Se desea saber cuántos usuarios de una autopista son turistas. El procedimiento de sondeo, desarrollado en un día cualquiera, consiste en escoger al primero de los usuarios que pase por el peaje después de las ocho de la mañana y preguntarle si es turista. Después, se le hace la misma pregunta a uno de cada 10 usuarios hasta las ocho de la mañana del día siguiente. Esto es un muestreo sistemático.

Muestreo oportunista o intencional. Es un tipo de muestreo no científico y de nula fiabilidad en el que la muestra es confeccionada de modo que los resultados presentan una tendencia definida en sesgo. Por ejemplo, no es fiable una encuesta sobre calidad de la enseñanza en un determinado nivel cuando los alumnos que integran la muestra fueron escogidos entre el colectivo de los que aprobaron todas las áreas dese nivel.

El tipo de muestreo que da mayor fiabilidad a la hora de generalizar a la población los resultados obtenidos en la muestra es el muestreo aleatorio, simple o con reposición. Es la mejor forma para conservar en la muestra las características de la población.

A la hora de generalizar los resultados a la población, siempre vienen especificados unos márgenes de error. Esos márgenes de error se pueden observar, por ejemplo, en cualquier sondeo electoral que publican los medios de comunicación en vísperas de elecciones.

Actividades propuestas

S2. Un profesor de matemáticas hace una prueba de control, sin influencia en la nota, en una clase de 30 estudiantes, pero no examina a todos. El día del control, el profesor lleva a clase un bombo con diez bolas numeradas del 0 al 9 y extrae una bola por sorteo; si sale el 0 deben hacer el control las personas 10, 20 y 30; si sale el 3 deben hacerlo las que tienen los números 3, 13 y 23, etcétera. ¿Qué tipo de muestreo usa el profesor para escoger a las personas que se deben examinar? ¿Cuál es la población? ¿Cuáles son los tamaños de la población y de la muestra?

S3. Un banco decide sortear 5 coches entre todos los clientes que tengan la nómina domiciliada en alguna de sus oficinas. Interprétese el proceso de sorteo como un muestreo indicando cuál es la población y cuál es el tamaño de la muestra. ¿Qué tipo de muestreo debe llevar a cabo el banco para garantizar la equidad en el sorteo?

S4. En un mercado hay 15 puestos de venta de pimientos de Padrón. Los vendedores desean saber qué porcentaje de pimientos tiene sabor picante y deciden coger 10 pimientos de cada puesto y probarlos. ¿Qué tipo de muestreo están haciendo?

S5. Una compañía aérea pretende conocer la opinión de sus clientes sobre la puntualidad de los vuelos. Para eso, contrata entrevistadores que preguntan a los viajeros de los aviones que aterrizaron a la hora prevista. ¿Considera adecuado este método de sondeo? ¿Cómo calificaría la técnica de muestreo utilizada?

S6. Una granja experimental estudia una determinada planta. El invernadero en el que se lleva a cabo el cultivo tiene 1000 plantas y los científicos quieren estimar el número de frutos que es esperable obtener en una plantación de estas características. Para eso, escogen al azar 100 plantas y cuentan el número de frutos de cada una de ellas, después multiplican el número total de frutos contados por 10. ¿Qué tipo de muestreo han hecho?

S7. Cinco depósitos cubren el suministro de agua de una ciudad. El control semanal de salubridad se hace del siguiente modo: cada día de la semana es analizada, por sorteo, el agua de uno de estos depósitos. ¿Qué tipo de muestreo se usa en el control semanal del agua de los depósitos?

S8. Un mes antes de las elecciones municipales, un partido político encarga un sondeo sobre intención de voto a una empresa. En el municipio viven 60 000 personas con derecho a voto que se distribuyen del siguiente modo:

	Hombres	Mujeres
Menores de 40 años	13 500	16 500
De 40 años o más	13 800	16 200

La empresa va a seleccionar una muestra de 2000 personas para hacer el estudio. ¿Cuántas personas serán menores de 40 años? ¿Cuántas serán mujeres? Cuántos serán hombres de 40 años o más?

S9. De cada uno de los siguientes estudios estadísticos, indique cuál es la población y si considera necesario elegir una muestra.

Horas diarias de sueño de los habitantes de una provincia.	Preferencias literarias de las personas mayores de edad que viven en un edificio.

2.3 Variables estadísticas. Frecuencias absolutas y relativas. Representaciones gráficas

2.3.1 Variables estadísticas

Se llama **variable estadística** a cada una de las propiedades que se pretendan estudiar.

Las variables estadísticas pueden ser cualitativas o cuantitativas. Las **variables estadísticas cualitativas** son las que no se pueden expresar como valores numéricos y las **variables estadísticas cuantitativas** son las que pueden tomar valores numéricos.

Las variables estadísticas cuantitativas pueden ser **discretas**, cuando no son posibles todos los valores numéricos, y **continuas** cuando son posibles todos los valores numéricos.

Actividades propuestas

S10. Indique si las siguientes variables son cualitativas o cuantitativas: la música preferida, el peso, el color del pelo, la edad, las calificaciones de un examen.

S11. Está usted estudiando la edad (en años) de los 25 estudiantes de una clase. Esta variable estadística, ¿es discreta o es continua?

S12. Está usted estudiando el peso de las manzanas de un manzano de su propiedad. Esta variable estadística, ¿es discreta o es continua?

2.3.2 Frecuencias absolutas y relativas

Cuando se recogen datos sobre un estudio estadístico, estos suelen ser muchos. Si nos limitamos a presentarlos como una lista de valores, serán poco explicativos y de difícil comprensión, por eso se utilizan métodos gráficos y numéricos para describir una variable estadística y sus valores.

Para representar los datos recogidos, lo primero es construir una tabla de frecuencias. En ella aparecen todos los valores que puede tomar la variable y el número de veces que se obtiene cada valor de la misma.

Para cualquier valor de la variable que estemos estudiando, se llama **frecuencia absoluta** de cada valor al número de veces que aparece dicho valor. La frecuencia absoluta se designa por f_i.

Actividad resuelta

Le preguntamos a 50 personas el número de veces que van al cine en un mes y obtenemos las siguientes respuestas:

0	1	2	1	0	2	0	1	1	1	1	0	0	1	0	3	1	1	1	1	0	1	0	1	1
1	1	2	1	0	2	1	1	0	1	1	1	1	1	0	1	1	1	1	2	1	1	1	1	1

Calcule las frecuencias absolutas de esta variable.

Efectuaremos un recuento de los datos ordenándolos en una tabla que muestre la frecuencia absoluta (número de veces que aparece ese dato), que llamaremos f_i.

Nos fijamos que, en este caso, los valores máximos y mínimo que toma nuestra variable estadística (número de veces que asiste al cine en un mes) son 3 y 0 respectivamente. Así, en la tabla de recogida de datos debemos poner los valores de la variable que hay entre 0 y 3.

x_i = número de veces que asiste al cine en un mes.	Recuento	f_i = frecuencia absoluta
0	||||| ||||| |	11
1	||||| ||||| ||||| ||||| ||||| ||||| |||	33
2	|||||	5
3	|	1
TOTAL		N = 50

Es decir, hay 11 personas que no van al cine, 33 que van una vez por mes, 5 que van dos veces y 1 que va tres veces al mes. Como preguntamos a 50 personas, la suma de las frecuencias absolutas tiene que ser 50.

A veces, las variables no son discretas, sino continuas, por lo que pueden tomar cualquier valor. En estos casos se establecen intervalos que llamaremos **intervalos de clase** y a cada uno de ellos le asociaremos como valor el punto medio del intervalo, que se llama **marca de clase**.

También es posible que la variable sea discreta, pero, ante la dificultad de hacer un recuento de cada valor de la variable o que haya un número muy grande de valores diferentes, se puede hacer un recuento de los datos agrupándolos en intervalos de clase.

Las clases no tienen por qué ser todas de la misma amplitud, pero es aconsejable hacerlas de la misma amplitud para evitar problemas en alguna de las representaciones gráficas que veremos en esta unidad.

Una vez hecho esto, se procede del mismo modo que en el ejemplo anterior.

Actividad resuelta

Se quiere realizar un estudio sobre la longitud de un tipo de tornillos que se hacen en una fábrica. Se elige al azar una muestra de 32 y se obtienen los siguientes resultados en milímetros.

161	171	167	172	170	170	165	169	170	169	172	162	169	166	174	178
167	169	168	176	169	162	168	167	175	168	164	179	172	167	170	173

Agrupe los datos en clases, establezca las marcas de clase y calcule cada frecuencia absoluta.

Efectuaremos un recuento de los datos ordenándolos en una tabla que muestre la frecuencia absoluta (número de veces que aparece ese dato), que llamaremos f_i.

Reparemos en que, en este caso, los valores máximo y mínimo que toma nuestra variable estadística (longitud de los tornillos) son 179 y 161, respectivamente. Así, en la tabla de recogida de datos debemos poner clases que incluyan estos valores.

Comenzaremos en el 160 y terminaremos en el 180 haciendo clases de longitud 5.

Intervalos de clase	x_i = marca de clase	Recuento	f_i = frecuencia absoluta
$160 \le x < 165$	162,5	||||	4
$165 \le x < 170$	167,5	||||| ||||| ||||	14
$170 \le x < 175$	172,5	||||| |||||	10
$175 \le x < 180$	177,5	||||	4
TOTAL			N = 32

Como analizamos 32 tornillos, la suma de las frecuencias absolutas tiene que ser 32.

Pero la frecuencia absoluta, como solo indica el número de veces que ha aparecido un valor, no da mucha información. Veamos un ejemplo: el primer día de clase, el profesor del ámbito científico tecnológico del módulo 3 en la presentación de la materia indica que el curso pasado suspendieron 5 personas. Esta información por sí misma no permite a las personas matriculadas hacerse una idea de la posible dificultad de la materia, pues si las personas matriculadas eran 10, suspendieron la mitad (el 50 %); pero si las personas matriculadas eran 50, entonces suspende una de cada 10 (un 10 %). Por lo tanto, para poder tener una idea mejor, debemos poner en relación la frecuencia absoluta con el número total de datos.

El número total de datos se representa habitualmente por **N**, como puede ver en las anteriores actividades resueltas.

Se llama **frecuencia relativa**, de un valor cualquiera de la variable estadística, a la frecuencia absoluta de ese valor dividida entre el número total de datos. La frecuencia relativa se designa por h_i.

Actividad resuelta

Calcule las frecuencias relativas en las anteriores actividades resueltas.

En la actividad del número de veces que van al cine en un mes, ya habíamos hecho el recuento y calculado las frecuencias absolutas.

Ahora, en la tabla que teníamos, añadimos una columna para calcular las frecuencias relativas.

x_i	Recuento	f_i	h_i = frecuencia relativa
0	||||| |||||		11	$\frac{11}{50} = 0{,}22$
1	||||| ||||| ||||| ||||| ||||| ||||| |||	33	$\frac{33}{50} = 0{,}66$
2	|||||	5	$\frac{5}{50} = 0{,}1$
3			1	$\frac{1}{50} = 0{,}02$
TOTAL		N = 50	1,00

Se debe señalar que la suma de las frecuencias relativas siempre vale 1.

En la actividad de la longitud de los tornillos, ya habíamos realizado el recuento y calculado las frecuencias absolutas.

Ahora, en la tabla que teníamos, añadimos una columna para calcular las frecuencias relativas.

Clases	x_i = marca de clase	Recuento	f_i	h_i = frecuencia relativa
$160 \leq x < 165$	162,5	||||	4	$\frac{4}{32} = 0{,}125$
$165 \leq x < 170$	167,5	||||| ||||| ||||	14	$\frac{14}{32} = 0{,}4375$
$170 \leq x < 175$	172,5	||||| |||||	10	$\frac{10}{32} = 0{,}3125$
$175 \leq x < 180$	177,5	||||	4	$\frac{4}{32} = 0{,}125$
TOTAL			N = 32	1,00

Es necesario señalar que la suma de las frecuencias relativas siempre vale 1.

La frecuencia relativa es el "*tanto por uno*" de veces que aparece cada valor de la variable. Es decir, si multiplicamos por 100 la frecuencia relativa obtenemos el tanto por ciento de veces que aparece cada uno de los valores.

Así, en el primero de los ejemplos, un 22 % de las personas a las que les hemos preguntado no va al cine, un 66 % va una vez al mes, un 10 % va dos veces al mes y un 2 % va al cine tres veces al mes.

2.3.3 Frecuencias absolutas y relativas acumuladas

Las frecuencias absolutas y las relativas estudiadas en el punto anterior nos dan una visión concreta de cada uno de los valores de la variable estadística, pero, a veces, queremos ver también la evolución de la variable desde el primero de los valores hasta un valor concreto. Para poder ver esto, se definen las frecuencias acumuladas, tanto las absolutas como las relativas.

Las frecuencias acumuladas de las actividades resueltas anteriores permitirán responder, con solo mirar la tabla, preguntas del tipo *"¿cuántas personas van al cine menos de dos veces al mes?"* o *"¿cuántos tornillos miden menos de 170 milímetros?"*

La **frecuencia absoluta acumulada** de un valor x_i (la representaremos por F_i) es el número de veces que aparece el valor x_i, o cualquiera de los anteriores a él. Es decir, para calcular la frecuencia absoluta del valor x_i, hay que sumar las frecuencias absolutas de todos los valores anteriores a x_i más la frecuencia absoluta de x_i.

Se llama **frecuencia relativa acumulada**, de un valor cualquiera de la variable estadística, a la frecuencia absoluta acumulada de ese valor dividida entre el número total de datos. La frecuencia relativa acumulada se designa por H_i.

Actividad resuelta

Calcule las frecuencias absolutas acumuladas y las frecuencias relativas acumuladas en las anteriores actividades resueltas.

En la actividad del número de veces que van al cine en un mes, ya habíamos hecho el recuento y calculado las frecuencias absolutas y relativas.

Ahora, en la tabla que teníamos, añadimos dos columnas, una para las frecuencias absolutas acumuladas y otra para las frecuencias relativas acumuladas. Por cuestiones de espacio, suprimiremos la columna del recuento.

x_i	f_i	h_i	F_i = frecuencia absoluta acumulada	H_i = frecuencia relativa acumulada
0	11	$\frac{11}{50} = 0{,}22$	11	$\frac{11}{50} = 0{,}22$
1	33	$\frac{33}{50} = 0{,}66$	44	$\frac{44}{50} = 0{,}88$
2	5	$\frac{5}{50} = 0{,}1$	49	$\frac{49}{50} = 0{,}98$
3	1	$\frac{1}{50} = 0{,}02$	50	$\frac{50}{50} = 1$
TOTAL	N = 50	1,00		

Es importante darse cuenta de que la frecuencia absoluta acumulada del mayor valor de la variable tiene que ser igual al número de datos (N), pues todos los valores obtenidos son menores o iguales que el mayor valor de la variable.

Del mismo modo, y por la misma razón, la frecuencia relativa acumulada del mayor valor de la variable tiene que ser igual a 1.

En la actividad de la longitud de los tornillos, ya habíamos realizado el recuento y calculado las frecuencias absolutas y relativas.

Ahora, en la tabla que teníamos, añadimos dos columnas, una para las frecuencias absolutas acumuladas y otra para las frecuencias relativas acumuladas. Por cuestiones de espacio suprimiremos la columna del recuento.

Clases	x_i	f_i	h_i	F_i = frecuencia absoluta acumulada	H_i = frecuencia relativa acumulada
160 ≤ x < 165	162,5	4	$\frac{4}{32} = 0{,}125$	4	$\frac{4}{32} = 0{,}125$
165 ≤ x < 170	167,5	14	$\frac{14}{32} = 0{,}4375$	18	$\frac{18}{32} = 0{,}5625$
170 ≤ x < 175	172,5	10	$\frac{10}{32} = 0{,}3125$	28	$\frac{28}{32} = 0{,}875$
175 ≤ x < 180	177,5	4	$\frac{4}{32} = 0{,}125$	32	$\frac{32}{32} = 1$
TOTAL		N = 32	1,00		

Observe que la frecuencia absoluta acumulada del mayor valor de la variable es igual al número de datos (32) y que su frecuencia relativa acumulada también vale 1.

De todas las formas, conviene señalar que para calcular las frecuencias acumuladas en la tabla podemos proceder de la siguiente forma:

El primer valor de las frecuencias acumuladas coincide con el primer valor de las frecuencias.

x_i	f_i	h_i	F_i = frecuencia absoluta acumulada	H_i = frecuencia relativa acumulada
0	11	$\frac{11}{50} = 0{,}22$	11	$\frac{11}{50} = 0{,}22$
1	33	$\frac{33}{50} = 0{,}66$	44	$\frac{44}{50} = 0{,}88$
2	5	$\frac{5}{50} = 0{,}1$	49	$\frac{49}{50} = 0{,}98$
3	1	$\frac{1}{50} = 0{,}02$	50	$\frac{50}{50} = 1$
TOTAL	N = 50	1,00		

El resto de los valores puede obtenerse de la siguiente forma:

- En la frecuencia absoluta acumulada, el valor siguiente se obtiene sumando la frecuencia absoluta acumulada del valor anterior con la frecuencia absoluta de este valor.

x_i	f_i	h_i	F_i = frecuencia absoluta acumulada	H_i = frecuencia relativa acumulada
0	11	$\frac{11}{50} = 0{,}22$	11	$\frac{11}{50} = 0{,}22$
1	33	$\frac{33}{50} = 0{,}66$	44	$\frac{44}{50} = 0{,}88$
2	5	$\frac{5}{50} = 0{,}1$	49	$\frac{49}{50} = 0{,}98$
3	1	$\frac{1}{50} = 0{,}02$	50	$\frac{50}{50} = 1$
TOTAL	N = 50	1,00		

- En la frecuencia relativa acumulada, el valor siguiente se obtiene sumando la frecuencia relativa acumulada del valor anterior con la frecuencia relativa de este valor.

x_i	f_i	h_i	F_i = frecuencia absoluta acumulada	H_i = frecuencia relativa acumulada
0	11	$\frac{11}{50} = 0{,}22$	11	$\frac{11}{50} = 0{,}22$
1	33	$\frac{33}{50} = 0{,}66$	44	$\frac{44}{50} = 0{,}88$
2	5	$\frac{5}{50} = 0{,}1$	49	$\frac{49}{50} = 0{,}98$
3	1	$\frac{1}{50} = 0{,}02$	50	$\frac{50}{50} = 1$
TOTAL	N = 50	1,00		

Actividades propuestas

S13. Un ayuntamiento revisa 50 pisos para saber cuánta gente vive en cada uno de ellos obteniendo los siguientes resultados:

0	2	4	1	0	2	3	3	1	0	4	2	3	0	1	4	2	4	1	0	5	2	1	3	0
4	2	4	3	5	1	2	3	4	0	1	2	3	2	1	3	2	0	1	4	2	3	1	2	0

Elabore la tabla de frecuencias incluyendo las relativas, las absolutas, las relativas acumuladas y las absolutas acumuladas.

S14. Las posibles respuestas a una encuesta son: MB (muy bueno), B (bueno), R (regular), M (malo) y MM (muy malo). Las respuestas de 50 personas fueron las siguientes:

R	B	MB	M	R	R	MM	MB	M	R
R	MM	R	B	R	MB	R	R	MB	R
M	R	B	R	MB	R	R	B	R	R
M	R	B	R	MB	R	R	B	R	MM
R	R	B	R	R	M	R	B	B	R

Elabore la tabla de frecuencias y responda a las siguientes preguntas:

a) ¿Cuántas personas responden mal o muy mal?

b) ¿Qué porcentaje de personas responde bien o muy bien?

S15. La siguiente tabla representa la puntuación obtenida por 100 estudiantes en un test que constaba de ocho preguntas.

Puntos	0	1	2	3	4	5	6	7	8
N.º personas	0	2	6	9	18	22	24	12	7

Elabore la tabla de frecuencias y responda a las siguientes preguntas:

a) ¿Cuántas personas obtienen 5 puntos? ¿Qué porcentaje representan?

b) ¿Cuántas personas tienen 6 o más puntos? ¿Qué porcentaje representan?

2.3.4 Representaciones gráficas

Para representar gráficamente las frecuencias, tanto absolutas como relativas, disponemos de varias alternativas:

- Diagrama de barras.
- Polígono de frecuencias.
- Diagrama de sectores.
- Histogramas.

Diagrama de barras

Se utiliza para representar tablas de frecuencias correspondientes a variables cuantitativas discretas. Por eso, las barras son estrechas y se sitúan sobre los valores puntuales de la variable. También se puede utilizar para representar variables cualitativas.

Para dibujar un diagrama de barras, representamos en el eje horizontal los diferentes valores de la variable (x_i) y en el eje vertical la frecuencia que deseamos representar; sobre cada valor de la variable dibujamos una barra hasta la altura de la frecuencia que queremos representar.

Actividad resuelta

Dibuje los diagramas de barras para las frecuencias absolutas y las absolutas acumuladas correspondientes a la variable estadística "nota obtenida en el examen de matemáticas" en una clase de 20 estudiantes donde las notas alcanzadas son las siguientes:

10	1	3	9	7	5	4	5	6	2	5	5	3	4	5	9	5	4	6	8

Calculamos las frecuencias absolutas y las absolutas acumuladas que aparecen en la tabla siguiente.

x_i	f_i	F_i
1	1	1
2	1	2
3	2	4
4	3	7
5	6	13
6	2	15
7	1	16
8	1	17
9	2	19
10	1	20
TOTAL	N = 20	

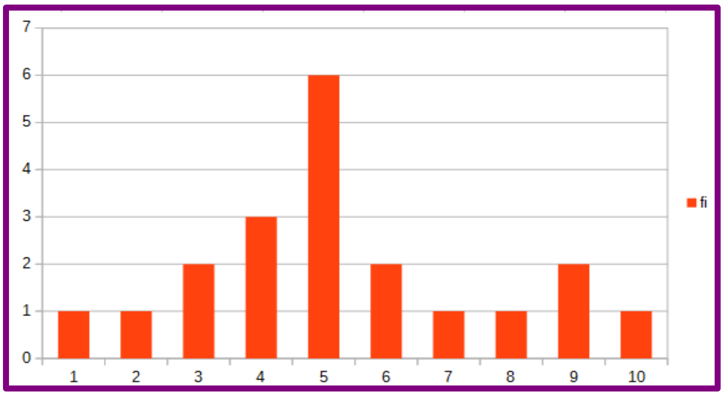


Diagrama de barras de las frecuencias absolutas

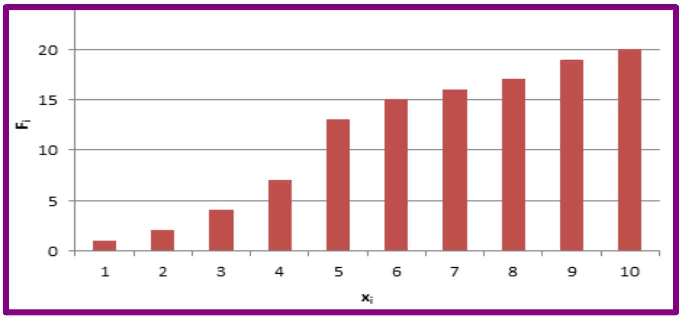


Diagrama de barras de las frecuencias absolutas acumuladas

En el caso de que sea una variable continua agrupada en clases, tomaremos como valores de la variable las marcas de clase.

Polígono de frecuencias

Para dibujar un polígono de frecuencias representamos en el eje horizontal los diferentes valores de la variable (x_i) y en el eje vertical la frecuencia que deseamos representar. Sobre cada valor de la variable dibujamos un punto a la altura de la frecuencia que queremos representar y posteriormente unimos dichos puntos.

Actividad resuelta

Construya los polígonos de frecuencias para las frecuencias relativas y las relativas acumuladas correspondientes a la variable estadística de la anterior actividad resuelta.

10	1	3	9	7	5	4	5	6	2	5	5	3	4	5	9	5	4	6	8

Calculamos las frecuencias relativas y las relativas acumuladas que aparecen en la tabla siguiente.

x_i	f_i	h_i	H_i
1	1	0,05	0,05
2	1	0,05	0,1
3	2	0,1	0,2
4	3	0,15	0,35
5	6	0,3	0,65
6	2	0,1	0,75
7	1	0,05	0,8
8	1	0,05	0,85
9	2	0,1	0,95
10	1	0,05	1
TOTAL	N = 20	1	

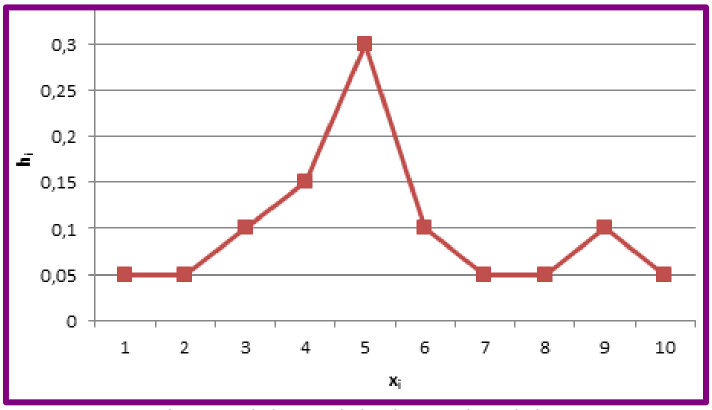


Diagrama de barras de las frecuencias relativas

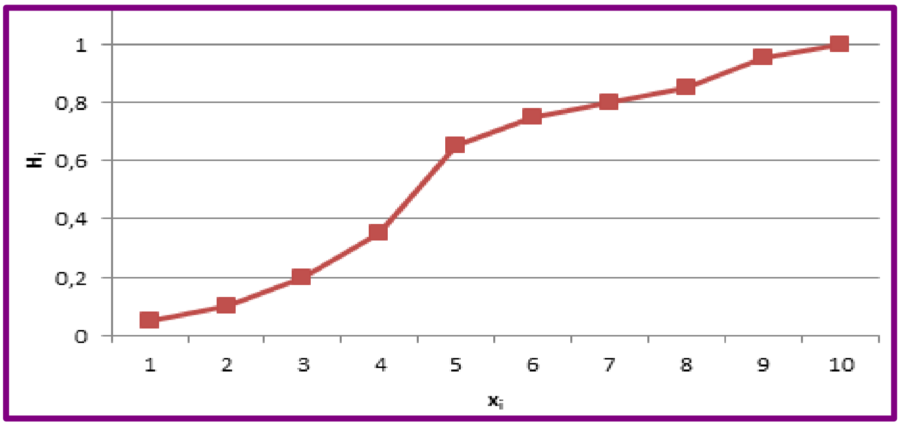


Diagrama de barras de las frecuencias relativas acumuladas

Del mismo modo que para el diagrama de barras, en el caso de que sea una variable continua agrupada en clases, tomaremos como valores de la variable las marcas de clase.

Diagramas de sectores

Se usan para representar frecuencias absolutas y frecuencias relativas. No se usa para representar las frecuencias acumuladas, ni las absolutas ni las relativas.

Para dibujar un diagrama de sectores dibujamos un círculo y, a modo de tartas de colores, representamos proporcionalmente la frecuencia en el ángulo de cada sector. Se puede utilizar para todo tipo de variables, pero se usa frecuentemente para las variables cualitativas.

Actividad resuelta

Dibuje los diagramas de sectores para las frecuencias absolutas de la variable estadística de la actividad resuelta anterior "nota obtenida en el examen de matemáticas en una clase de 20 estudiantes".

Ya teníamos las frecuencias absolutas que aparecen en la tabla siguiente.

x_i	f_i
1	1
2	1
3	2
4	3
5	6
6	2
7	1
8	1
9	2
10	1
TOTAL	N = 20

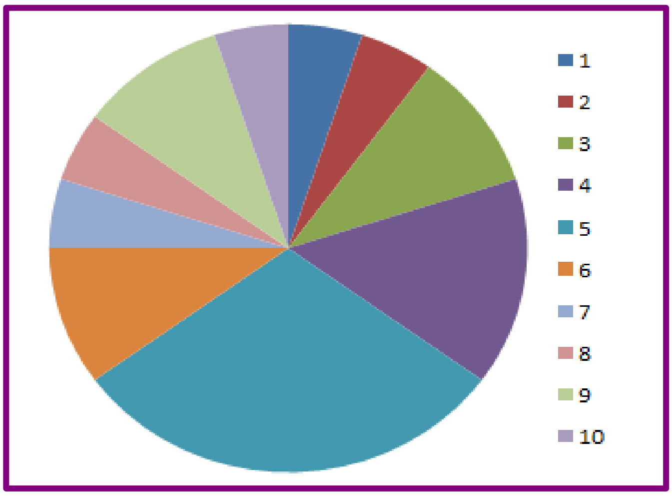


Diagrama de sectores de las frecuencias absolutas

En el caso de que sea una variable continua agrupada en clases, tomaremos como valores de la variable las marcas de clase.

Histogramas

Se usan para representas frecuencias absolutas y frecuencias relativas para variables continuas.

Para dibujar un histograma dibujamos en el eje horizontal las clases. Sobre cada clase dibujamos un rectángulo con el área proporcional a la frecuencia que queremos representar. Si todas las clases son de la misma amplitud, el histograma se construye como el diagrama de barras, pero en este caso el ancho de la barra es el del intervalo de clase.

Actividad resuelta

En el apartado 2.3.2 trabajamos en una actividad resuelta que trataba sobre la longitud de un tipo de tornillos que se hacen en una fábrica. Teníamos una muestra de 32 tornillos con los siguientes resultados en milímetros.

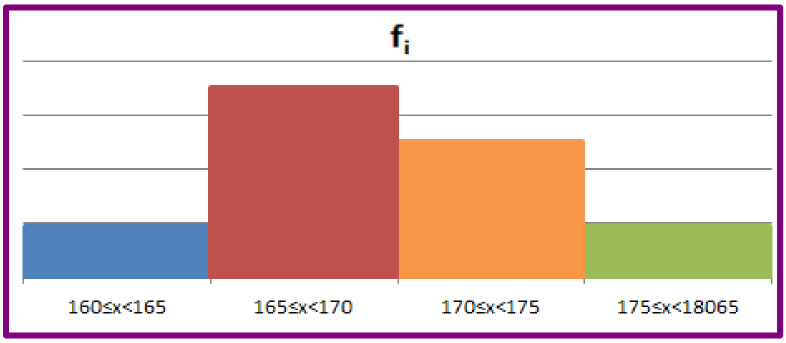
161	171	167	172	170	170	165	169	170	169	172	162	169	166	174	178
167	169	168	176	169	162	168	167	175	168	164	179	172	167	170	173

Agrupábamos los datos en clases y establecíamos las marcas de clase.

Dibuje el histograma correspondiente a esta variable.

Ya teníamos la tabla siguiente:

Intervalos	x_i	f_i
$160 \leq x < 165$	162,5	4
$165 \leq x < 170$	167,5	14
$170 \leq x < 175$	172,5	10
$175 \leq x < 180$	177,5	4
TOTAL		N = 32



Actividad resuelta

En un teatro miden la duración de los aplausos: 40 veces los aplausos duraron menos de 10 segundos, 30 veces duraron entre 10 y 15 segundos, 20 veces duraron entre 15 y 20 segundos y 36 veces duraron entre 20 y 30 segundos. Dibuje el histograma correspondiente.

Como las clases tienen diferente anchura, debemos ver la altura que tiene que tener cada rectángulo, ya que la frecuencia de cada intervalo viene dada por el área del mismo y no por su altura.

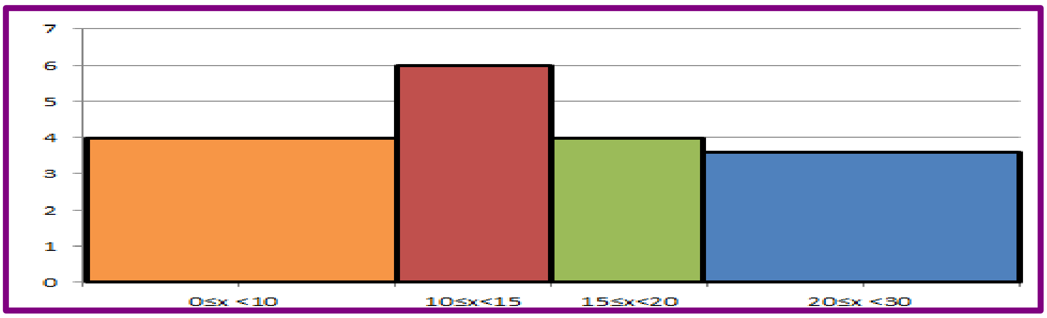
Como los intervalos 1.º y 4.º tienen el doble de anchura que los otros, para que el área sea 40 tienen que tener como altura la mitad de la frecuencia.

La tabla es:

Intervalos	f_i
$0 \leq x < 10$	40
$10 \leq x < 15$	30
$15 \leq x < 20$	20
$20 \leq x < 30$	36
TOTAL	N = 126

Para calcular la altura de cada intervalo dividimos la frecuencia absoluta entre la anchura del intervalo y obtenemos:

Intervalos	f_i	Altura del histograma en este intervalo f_i/ancho intervalo
$0 \leq x < 10$	40	40 / 10 = 4
$10 \leq x < 15$	30	30 / 5 = 6
$15 \leq x < 20$	20	20 / 5 = 4
$20 \leq x < 30$	36	36 / 10 = 3,6
TOTAL	N = 126	



Y lo que indica la cantidad de datos que hay en cada uno de los intervalos es el área del rectángulo y no su altura.

Actividades propuestas

S16. La frecuencia con que va por semana a la biblioteca el alumnado de un centro escolar se puede observar en la tabla siguiente. Realice un diagrama de barras, uno de sectores y un polígono de frecuencias.

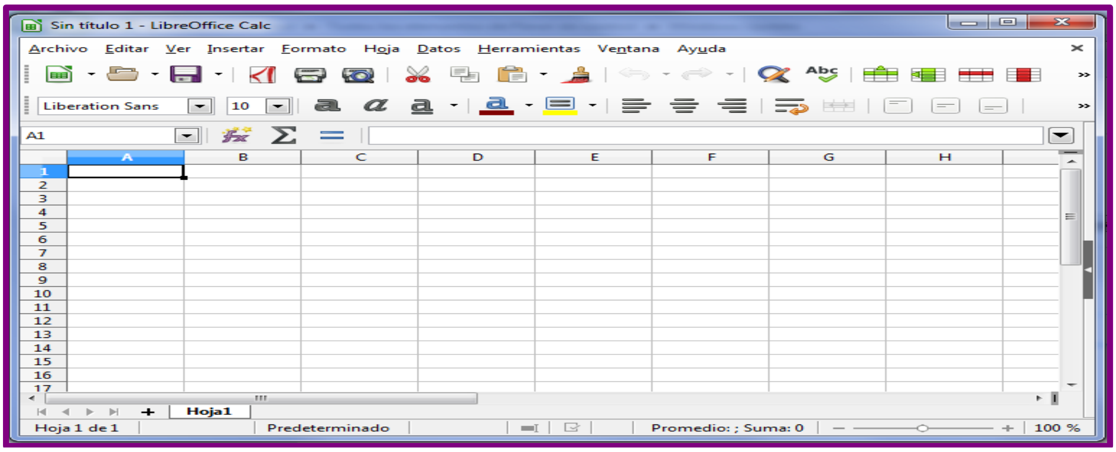
Frecuencia	Nunca	Una vez	Dos veces	Más de dos veces
N.º personas	480	540	120	60

S17. En un concurso de rompecabezas participan 54 personas: 8 personas tardan menos de 5 minutos en hacerlo, 12 tardan entre 5 y 10 minutos, 15 tardan entre 10 y 15 minutos, 10 tardan entre 15 y 20 minutos y 9 tardan entre 20 y 25 minutos. Dibuje el histograma que representa esta distribución.

2.4 Aplicaciones informáticas para la representación gráfica de datos estadísticos

2.4.1 Diagramas de barras con LibreOffice Calc

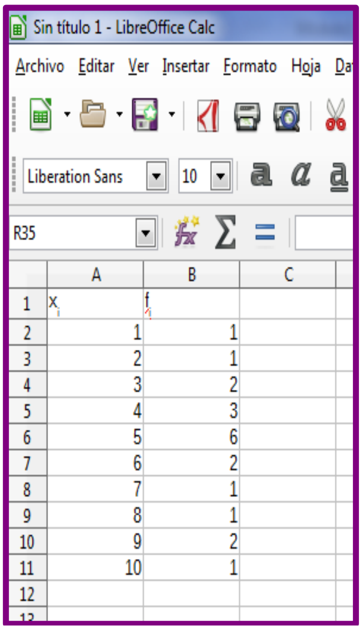
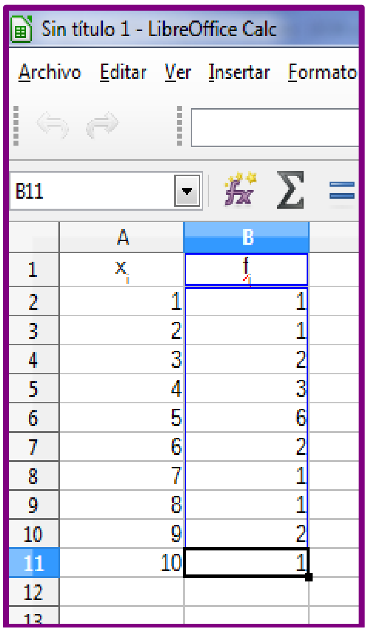
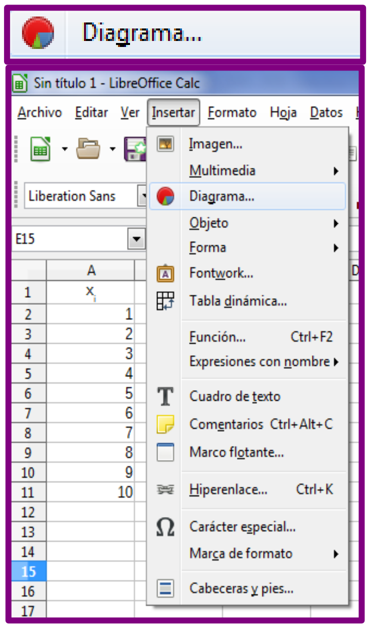
Para realizar estas representaciones gráficas utilizaremos una hoja de cálculo. Una hoja de cálculo es un cuadro formado por celdas en el que se pueden colocar números, textos o fórmulas. Cada celda se identifica con una letra, que indica la columna, y un número, que indica la fila.



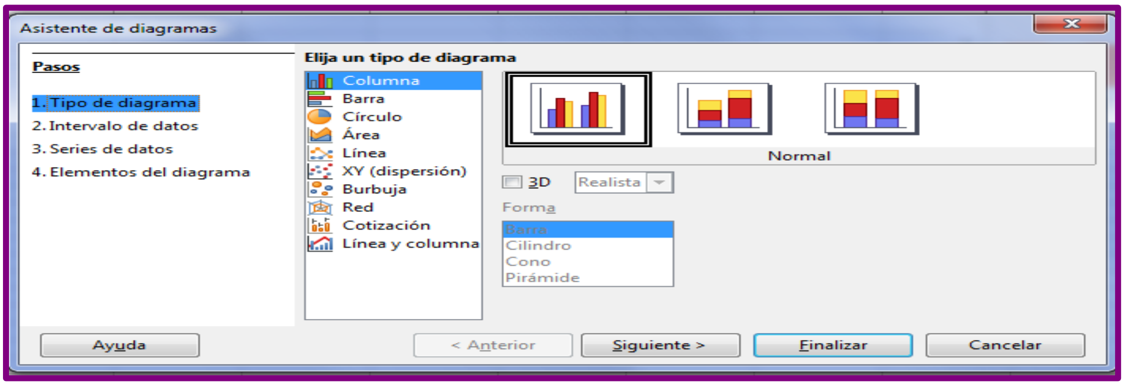
Dibujaremos el diagrama de barras para las frecuencias absolutas correspondientes a la variable estadística "nota obtenida en el examen de matemáticas" en una clase de 20 estudiantes, ya realizada en el apartado anterior, donde las notas alcanzadas son las siguientes:

10	1	3	9	7	5	4	5	6	2	5	5	3	4	5	9	5	4	6	8

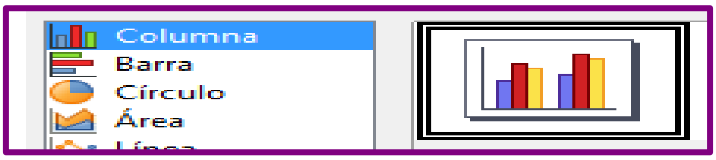
Para hacerlo, introducimos en las celdas de la cuadrícula el valor de la variable en la primera columna y su frecuencia absoluta en la segunda.

	Seleccionamos los datos de la columna B	A continuación, seleccionamos en el menú la opción "Insertar"
		

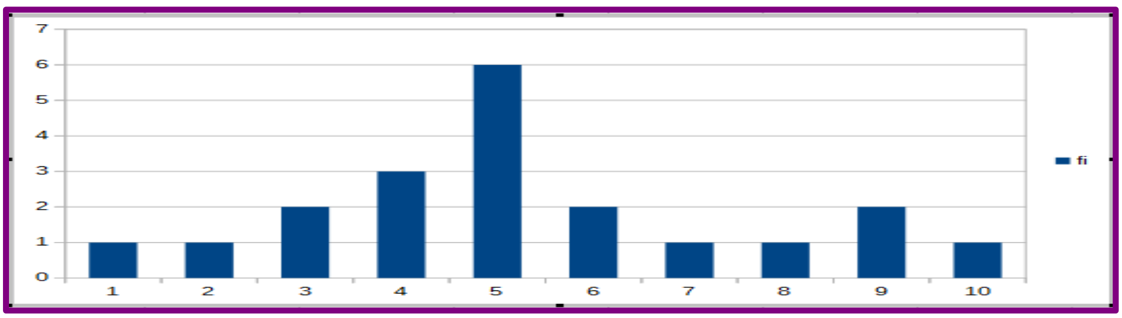
Aparece, entonces, el siguiente cuadro de diálogo:



Como queremos hacer un diagrama de barras, escogemos la opción columna.

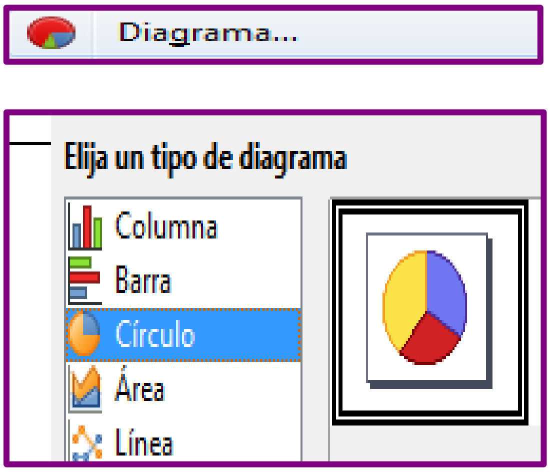


Pulsamos en "Finalizar" y nos aparece el gráfico siguiente:

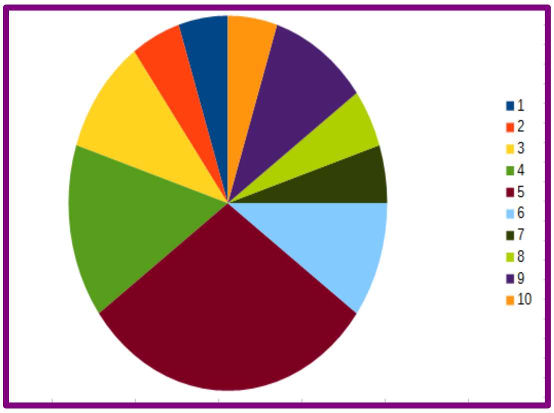


2.4.2 Diagramas de sectores con LibreOffice Calc

El proceso es el mismo que para hacer los diagramas de barras, pero cuando seleccionamos “Insertar” escogemos:



Pulsamos en “Finalizar” y nos aparece el gráfico siguiente:



2.5 Parámetros estadísticos

Después de obtener los datos de una distribución, necesitamos sintetizar la información para su posterior análisis. Para eso, obtendremos los parámetros estadísticos que serán de tres tipos: de centralización, de dispersión y de posición.

- Los **parámetros de centralización** nos indican alrededor de qué valor se distribuyen los datos.
- Los **parámetros de dispersión** nos informan sobre cuánto se separan del centro los valores de la distribución.
- Los **parámetros de posición** dividen un conjunto de datos en grupos con el mismo número de individuos. Para calcularlos, es necesario que los **datos** estén ordenados de **menor a mayor**.

2.5.1 Parámetros de centralización

Dentro de los parámetros de centralización tenemos la media, la mediana y la moda.

Media

La **media** es el valor central de la distribución. Las diferencias que les faltan a los valores que no alcanzan la media se compensan con lo que les sobra a los valores que son mayores que la media.

La media se denota por $\bar{x}$.

Para obtenerla, sumamos todos los valores que toma la variable (el número de veces que aparece cada uno de ellos) y dividimos entre el número total de datos.

Otra forma más fácil de obtenerla es multiplicar cada valor de la variable (x_i) por su frecuencia absoluta (f_i), sumar esos productos y el resultado dividirlo entre el número total de datos (N).

Es decir:

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \cdots + x_n \cdot f_n}{N}$$

Pero la suma de varios sumandos con la misma estructura, como pasa en este caso, se representa utilizando el signo Σ, que indica desde qué valor hasta qué valor va la suma. Por lo tanto, tenemos que:

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \cdots + x_n \cdot f_n}{N} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{N}$$

Actividad resuelta

Calcule la media de la variable estadística “nota obtenida en el examen de matemáticas” en una clase de 20 estudiantes donde las notas alcanzadas son las siguientes:

10	1	3	9	7	5	4	5	6	2	5	5	3	4	5	9	5	4	6	8

Una vez que tenemos las columnas de los valores de la variable (x_i) y de las frecuencias absolutas (f_i), debemos calcular una nueva columna que va a contener los productos de los números de las dos columnas anteriores ($x_i \cdot f_i$).

x_i	f_i	$x_i \cdot f_i$
1	1	1
2	1	2
3	2	6
4	3	12
5	6	30
6	2	12
7	1	7
8	1	8
9	2	18
10	1	10
TOTAL	N = 20	$\sum_{i=1}^{10} x_i \cdot f_i = 106$

Así, tenemos que la media es:

$$\bar{x} = \frac{106}{20} = 5{,}3$$

Mediana

Si ordenamos los datos de la distribución de menor a mayor, la **mediana** es el valor que se encuentra en el medio; es decir, es el valor que deja tantos individuos antes como después. Si el número de datos fuese par, a la mediana se le asigna el valor medio de los dos términos centrales.

La mediana se denota por Me.

Para obtenerla deberíamos ordenar todos los valores de la variable y ver cuál ocupa el lugar central. Pero esto puede llevar mucho tiempo si el número de datos es muy grande, por lo que debemos encontrar una manera más eficaz para calcularla.

Calcularemos la mediana en la actividad resuelta anterior. Para eso nos van hacer falta las frecuencias absolutas acumuladas.

Actividad resuelta

Calcule la mediana de la variable estadística "nota obtenida en el examen de matemáticas" en una clase de 20 estudiantes de la anterior actividad resuelta.

Una vez que tenemos las columnas de los valores de la variable (x_i) y de las frecuencias absolutas (f_i), debemos tener también la columna de la frecuencia absoluta acumulada (F_i).

x_i	f_i	F_i
1	1	1
2	1	2
3	2	4
4	3	7
5	6	13
6	2	15
7	1	16
8	1	17
9	2	19
10	1	20
TOTAL	N = 20	

Si interpretamos la columna de las frecuencias absolutas acumuladas, tenemos que el valor más pequeño de la variable es 1, el segundo es 2, el tercero y el cuarto valen 3, los tres siguientes (es decir, el 5.º, el 6.º y el 7.º) valen 4, los seis siguientes valen 5, y así sucesivamente.

Es decir, la columna de las frecuencias absolutas acumuladas nos ordenó los datos.

Solo falta saber cuál es el lugar del medio:

- Si el número de datos es impar, hay un único lugar en el medio.
- Si el número de datos es par, hay dos lugares en el medio.

En este caso, el número de datos es 20, que es par, por lo que hay dos valores en los lugares del medio.

Los lugares del medio son la mitad de N y el siguiente, es decir, 10 y 11.

Como la frecuencia absoluta acumulada del 4 es 7 y la del 5 es 13, tenemos que los lugares 8.º, 9.º, 10.º, 11.º, 12.º y 13.º valen 5. Por lo tanto, los lugares 10.º y 11.º están ocupados por dos cincos.

Así, tenemos que la mediana es:

$$Me = \frac{5+5}{2} = 5$$

Moda

La **moda** es el valor que tiene la mayor frecuencia absoluta.

La moda se denota por Mo.

La moda no tiene por qué tener un único valor. En el caso de que haya varios valores que tengan la misma frecuencia absoluta y que esta sea la mayor que aparece, tendríamos una variable con varias modas.

Actividad resuelta

Calcule la moda de la variable estadística "nota obtenida en el examen de matemáticas" en una clase de 20 estudiantes de la actividad resuelta anterior.

x_i	f_i
1	1
2	1
3	2
4	3
5	6
6	2
7	1
8	1
9	2
10	1
TOTAL	N = 20

Una vez que tenemos las columnas de los valores de la variable (x_i) y de las frecuencias absolutas (f_i), no necesitamos nada más para calcular la moda, ya que solo tenemos que buscar cuál es el valor que tiene la frecuencia absoluta mayor.

Como el valor que tiene la frecuencia absoluta mayor es $x_i = 5$, ya que su $f_i = 6$, tenemos que la moda es:

$$Mo = 5$$

A

Actividad resuelta

Las edades de los asistentes a la fiesta de cumpleaños de su hija son:

12	10	11	13	12	11	13	12	13	13	12	13	11	12	13
13	11	12	10	12	11	14	12	14	12	10	14	13	11	13

Calcule la media, la mediana y la moda.

- Para calcular la media necesitamos tres columnas de la tabla: x_i, f_i y $x_i \cdot f_i$.
- Para calcular la mediana necesitamos tres columnas: x_i, f_i y F_i (en realidad f_i solo es necesaria para calcular F_i, no para calcular la mediana).
- Para calcular la moda necesitamos dos columnas: x_i y f_i.

Por lo tanto, para calcular estos tres parámetros de centralización necesitamos cuatro columnas de la tabla, que son: x_i, f_i, F_i y $x_i \cdot f_i$.

x_i	f_i	F_i	$x_i \cdot f_i$
10	3	3	30
11	6	9	66
12	9	18	108
13	9	27	117
14	3	30	42
TOTAL	N = 30		363

Así tenemos:

- Media:

$$\bar{x} = \frac{363}{30} = 12{,}1$$

- Mediana: los lugares del medio (15.º y 16.º) están ocupados polo valor $x_i = 12$

$$Me = \frac{12 + 12}{2} = 12$$

- Moda: hay dos modas. Se dice que es una distribución bimodal. Las modas son:

$$Mo = 12 \text{ y } Mo = 13$$

Esta información que ofrecen los parámetros de centralización nos dice que estos datos están todos alrededor del valor 12. Surge, entonces, la siguiente pregunta: si todos están alrededor del valor 12, ¿son todos próximos a 12?

Esta pregunta tiene sentido si pensamos que para obtener 12 de media, podemos partir de 2 y 22 o bien de 10 y 14. En ambos casos la media es 12, pero los datos de partida son muy diferentes. Esto hace necesario conocer más sobre los datos de la distribución, para eso tenemos los parámetros de dispersión, que nos informarán de cómo están de aproximados los datos de la tabla.

Actividades propuestas

S18. Las notas obtenidas en un examen de matemáticas de la ESA con sus frecuencias absolutas son las siguientes:

x_i	0	1	2	3	4	5	6	7	8	9	10
f_i	0	5	4	2	2	1	1	2	3	4	8

Calcule la media, la moda y la mediana.

S19. Las estaturas de varias personas (ya agrupadas en clases con las marcas de clase correspondientes) y sus frecuencias absolutas son las siguientes:

x_i	151	156	161	166	171	176
f_i	2	5	11	14	5	3

Calcule la media, la moda y la mediana.

2.5.2 Parámetros de dispersión

Dentro de los parámetros de dispersión tenemos la desviación media, la varianza y la desviación típica.

Lo que calculan los parámetros de dispersión es la media de las distancias de los datos con respecto a la media. Pero no se puede tener en cuenta el signo de estas distancias (negativo si el dato es menor que la media y positivo cuando el dato es mayor que la media). Recuerde que cuando definimos la media dijimos que "*las diferencias que les faltan a los valores que no alcanzan la media se compensan con lo que les sobra a los valores que son mayores que la media*." Por lo que, si consideramos el signo de las diferencias, su suma siempre será cero.

Por lo tanto, para calcular la media de las distancias de los datos respecto a la media de la distribución debemos transformar esas distancias en positivas. Eso podemos hacerlo de dos formas.

- Mediante el valor absoluto, y así calculamos la desviación media.
- Elevando las distancias al cuadrado, y así calculamos la varianza.

Pero antes de ver como se calculan la desviación media y la varianza, veremos qué es el rango o recorrido de una variable estadística.

Rango o recorrido

El **rango** o **recorrido** es la diferencia entre los datos mayor y menor de la variable. Viene siendo la longitud del tramo dentro del cual están los datos.

Actividad resuelta

Las edades de los asistentes a la fiesta de cumpleaños de su hija son:

12	10	11	13	12	11	13	12	13	13	12	13	11	12	13
13	11	12	10	12	11	14	12	14	12	10	14	13	11	13

Calcule el recorrido de esta variable.

Para calcular el recorrido, necesitamos localizar los valores máximo y mínimo que toma esta variable.

- El valor mínimo es 10.
- El valor máximo es 14.
- Por lo tanto, el rango es:

$$\text{rango} = 14 - 10 = 4$$

Desviación media

La **desviación media** es la media de las distancias de los datos a la media. Se calcula mediante la media de las diferencias en valor absoluto. La desviación media se denota por DM.

$$DM = \frac{|x_1 - \bar{x}| \cdot f_1 + |x_2 - \bar{x}| \cdot f_2 + \cdots + |x_n - \bar{x}| \cdot f_n}{N} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}| \cdot f_i}{N}$$

Actividad resuelta

En la misma actividad resuelta anterior, las edades de los asistentes a la fiesta de cumpleaños de su hija eran:

12	10	11	13	12	11	13	12	13	13	12	13	11	12	13
13	11	12	10	12	11	14	12	14	12	10	14	13	11	13

Calcule la desviación media

Para calcular la desviación media necesitamos:

- Calcular la media, por lo que necesitamos tres columnas de la tabla: x_i, f_i y $x_i \cdot f_i$.
- Calcular la media de $|x_i - \bar{x}|$, por lo que necesitamos tres columnas: $|x_i - \bar{x}|$, f_i y $|x_i - \bar{x}| \cdot f_i$.

Por lo tanto, para calcular la desviación media necesitamos cinco columnas de la tabla que son:

x_i, f_i, $x_i \cdot f_i$, $|x_i - \bar{x}|$, y $|x_i - \bar{x}| \cdot f_i$.

x_i	f_i	$x_i \cdot f_i$
10	3	30
11	6	66
12	9	108
13	9	117
14	3	42
TOTAL	N = 30	363

Así tenemos.

- Media:

$$\bar{x} = \frac{363}{30} = 12{,}1$$

Como ya habíamos calculado anteriormente (ver el apartado de medidas de centralización).

Una vez calculada la media, ya podemos calcular la desviación media.

x_i	f_i	$x_i \cdot f_i$	$\lvert x_i - \bar{x}\rvert$	$\boldsymbol{\lvert x_i - \bar{x}\rvert \cdot f_i}$
10	3	30	2,1	6,3
11	6	66	1,1	6,6
12	9	108	0,1	0,9
13	9	117	0,9	8,1
14	3	42	1,9	5,7
TOTAL	N = 30	363		27,6

Así, tenemos:

- Desviación media:

$$DM = \frac{27{,}6}{30} = 0{,}92$$

Varianza y desviación típica

La **varianza** es la media de los cuadrados de las distancias de los datos a la media. La varianza se denota por σ^2.

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 \cdot f_1 + (x_2 - \bar{x})^2 \cdot f_2 + \cdots + (x_n - \bar{x})^2 \cdot f_n}{N} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot f_i}{N}$$

La varianza tiene el problema de que las unidades en que se expresa, al estar elevadas al cuadrado, desvirtúan las medidas. Así, por ejemplo, si estudiamos la estatura en cm, al elevar al cuadrado, las unidades serían cm^2 y esto no representa una longitud, sino una superficie. Por eso extraemos su raíz cuadrada, que es lo que llamamos desviación típica.

La **desviación típica** es la raíz cuadrada de la varianza. La desviación típica se denota por σ.

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

Actividad resuelta

En la misma actividad resuelta anterior, las edades de los asistentes a la fiesta de cumpleaños de su hija eran:

12	10	11	13	12	11	13	12	13	13	12	13	11	12	13
13	11	12	10	12	11	14	12	14	12	10	14	13	11	13

Calcule la desviación típica.

Para calcular la desviación típica necesitamos calcular la varianza y para calcular la varianza necesitamos:

- Calcular la media, por lo que necesitamos tres columnas de la tabla: x_i, f_i y $x_i \cdot f_i$.
- Calcular la media de $(x_i - \bar{x})^2$, por lo que necesitamos cuatro columnas: $(x_i - \bar{x})$, $(x_i - \bar{x})^2$, f_i y $(x_i - \bar{x})^2 \cdot f_i$.

Por lo tanto, para calcular la desviación típica necesitamos seis columnas de la tabla, que son:

x_i, f_i, $x_i \cdot f_i$, $(x_i - \bar{x})$, $(x_i - \bar{x})^2$, y $(x_i - \bar{x})^2 \cdot f_i$.

x_i	f_i	$x_i \cdot f_i$
10	3	30
11	6	66
12	9	108
13	9	117
14	3	42
TOTAL	N = 30	363

Así, tenemos:

- Media:

$$\bar{x} = \frac{363}{30} = 12{,}1$$

Como ya habíamos calculado anteriormente (ver el apartado de medidas de centralización).

Una vez calculada la media, ya podemos calcular la varianza y la desviación típica.

x_i	f_i	$x_i \cdot f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
10	3	30	-2,1	4,41	13,23
11	6	66	-1,1	1,21	7,26
12	9	108	-0,1	0,01	0,09
13	9	117	0,9	0,81	7,23
14	3	42	1,9	3,61	10,83
TOTAL	N = 30	363			38,64

Así, tenemos:

- Varianza:

$$\sigma^2 = \frac{38{,}64}{30} = 1{,}288$$

Y también:

- Desviación típica:

$$\sigma = \sqrt{\sigma^2} = \sqrt{1{,}288} = 1{,}1349$$

Actividad resuelta

Calcule los parámetros de dispersión de la variable estadística “nota obtenida en el examen de matemáticas” en una clase de 20 estudiantes donde las notas alcanzadas son las siguientes:

10	1	3	9	7	5	4	5	6	2	5	5	3	4	5	9	5	4	6	8

Para calcular los parámetros de dispersión necesitamos:

- Para el rango, determinar los valores máximo y mínimo de la variable.
- Para la desviación media, x_i, f_i y $x_i \cdot f_i$ para calcular la media y $|x_i - \bar{x}|$, f_i y $|x_i - \bar{x}| \cdot f_i$ para la desviación media.
- Para la desviación típica, x_i, f_i y $x_i \cdot f_i$ para calcular la media y $(x_i - \bar{x})$, $(x_i - \bar{x})^2$, f_i y $(x_i - \bar{x})^2 \cdot f_i$ para la varianza.

Por lo tanto, para calcular los parámetros de dispersión necesitamos ocho columnas de la tabla que son:

x_i, f_i, $x_i \cdot f_i$, $|x_i - \bar{x}|$, $|x_i - \bar{x}| \cdot f_i$, $(x_i - \bar{x})$, $(x_i - \bar{x})^2$, y $(x_i - \bar{x})^2 \cdot f_i$.

Las columnas $|x_i - \bar{x}|$ y $(x_i - \bar{x})$ son la misma excepto el signo, por lo que solo calcularemos la columna de $|x_i - \bar{x}|$. En definitiva, son siete columnas.

x_i	f_i	$x_i \cdot f_i$
1	1	1
2	1	2
3	2	6
4	3	12
5	6	30
6	2	12
7	1	7
8	1	8
9	2	18
10	1	10
TOTAL	N = 20	106

Así, tenemos:

- Media:

$$\bar{x} = \frac{106}{20} = 5{,}3$$

Ahora ya podemos calcular los parámetros de dispersión:

- Rango:

$$\text{rango} = 10 - 1 = 9$$

x_i	f_i	$x_i \cdot f_i$	$\boldsymbol{|x_i - \bar{x}|}$	$\boldsymbol{|x_i - \bar{x}| \cdot f_i}$	$\boldsymbol{(x_i - \bar{x})^2}$	$\boldsymbol{(x_i - \bar{x})^2 \cdot f_i}$
1	1	1	4,3	4,3	18,49	18,49
2	1	2	3,3	3,3	10,89	10,89
3	2	6	2,3	4,6	5,29	10,58
4	3	12	1,3	3,9	1,69	5,07
5	6	30	0,3	1,8	0,09	0,54
6	2	12	0,7	1,4	0,49	0,98
7	1	7	1,7	1,7	2,89	2,89
8	1	8	2,7	2,7	7.29	7,29
9	2	18	3,7	7,4	13,69	27,38
10	1	10	4,7	4,7	22.09	22,09
TOTAL	N = 20	106		35,8		106,2

- Desviación media:

$$DM = \frac{35{,}8}{20} = 1{,}79$$

- Varianza:

$$\sigma^2 = \frac{106{,}2}{20} = 5{,}31$$

- Desviación típica:

$$\sigma = \sqrt{\sigma^2} = \sqrt{5{,}31} = 2{,}3043$$

Otra forma de calcular la varianza es utilizando la expresión:

$$\sigma^2 = \frac{{x_1}^2 \cdot f_1 + {x_2}^2 \cdot f_2 + \cdots + {x_n}^2 \cdot f_n}{N} - \bar{x}^2 = \frac{\sum_{i=1}^{n} {x_i}^2 \cdot f_i}{N} - \bar{x}^2$$

Actividad resuelta

Calcule la varianza y la desviación típica de la actividad resuelta anterior utilizando esta última expresión.

Para calcular la varianza de este modo, en vez de utilizar las columnas $(x_i - \bar{x})$, $(x_i - \bar{x})^2$, f_i y $(x_i - \bar{x})^2 \cdot f_i$ debemos utilizar ${x_i}^2$ y ${x_i}^2 \cdot f_i$.

Ya habíamos calculado la media que era:

$$\bar{x} = \frac{106}{20} = 5{,}3$$

x_i	f_i	$x_i \cdot f_i$	$\boldsymbol{{x_i}^2}$	$\boldsymbol{{x_i}^2 \cdot f_i}$
1	1	1	1	1
2	1	2	4	4
3	2	6	9	18
4	3	12	16	48
5	6	30	25	150
6	2	12	36	72
7	1	7	49	49
8	1	8	64	64
9	2	18	81	162
10	1	10	100	100
TOTAL	N = 20	106		668

- Varianza:

$$\sigma^2 = \frac{668}{20} - 5{,}3^2 = 33{,}4 - 28{,}09 = 5{,}31$$

- Desviación típica:

$$\sigma = \sqrt{\sigma^2} = \sqrt{5{,}31} = 2{,}3043$$

Que, naturalmente, coincide con lo calculado anteriormente.

Actividades propuestas

S20. Calcule el rango, la desviación media, la varianza y la desviación típica de la variable de la actividad propuesta número 18.

S21. Calcule el rango, la desviación media y la desviación típica de la variable de la actividad propuesta número 19.

2.5.3 Parámetros de posición. Los cuartiles

Dentro de los parámetros de posición solo veremos los cuartiles.

Los **cuartiles** de una serie estadística son Q_1, Q_2, y Q_3, de tal modo que dividen la distribución de los datos ya ordenados en cuatro partes iguales:

- Q_1 deja a su izquierda el 25 % de los datos.

- Q_2 deja a su izquierda el 50 % de los datos y, por lo tanto, coincide con la mediana.
- Q_3 deja a su izquierda el 75 % de los datos.

La diferencia Q_3 - Q_1 se llama **recorrido intercuartílico**.

Actividad resuelta

Las edades de los asistentes a la fiesta de cumpleaños de su hija son:

12	10	11	13	12	11	13	12	13	13	12	13	11	12	13
13	11	12	10	12	11	14	12	14	12	10	14	13	11	13

Calcule los cuartiles

Para calcular los cuartiles necesitamos la frecuencia absoluta acumulada.

x_i	f_i	F_i
10	3	3
11	6	9
12	9	18
13	9	27
14	3	30
TOTAL	N = 30	

Así, tenemos:

- Q_1 deja tras él la cuarta parte de la distribución. Como N = 30, tenemos que $\frac{N}{4} = 7{,}5$. Así, Q_1 debe dejar 7,5 elementos a su izquierda, por lo que debe ser el octavo elemento. Por lo tanto, $Q_1 = 11$.
- Q_2 es la mediana que, como N es par, será la media de los elementos que ocupan las posiciones centrales, que son las posiciones 15ª y 16ª. Por lo tanto, $Q_2 = \frac{12+12}{2} = 12$.
- Q_3 deja tras él las tres cuartas partes de la distribución. Como N = 30, tenemos que $\frac{3N}{4} = 22{,}5$. Así, Q_3 debe dejar 22,5 elementos a su izquierda, por lo que debe ser el 23.º elemento. Por lo tanto, $Q_3 = 13$.

Actividades propuestas

S22. Calcule los cuartiles de la variable de la actividad propuesta número 18.

S23. Calcule los cuartiles de la variable de la actividad propuesta número 19.

2.5.4 Diagrama de caja y bigotes

El **diagrama de caja y bigotes** es una forma muy clara de representar los cuartiles y el recorrido de la variable. Veámoslo en las siguientes actividades resueltas:

Actividad resuelta

Calcule el recorrido y los cuartiles y haga el diagrama de caja y bigotes correspondiente a la variable estadística "nota obtenida en el examen de matemáticas" en una clase de 20 estudiantes donde las notas alcanzadas son las siguientes:

10	1	3	9	7	5	4	5	6	2	5	5	3	4	5	9	5	4	6	8

Para calcular los cuartiles, necesitamos ordenar los valores y calcular la frecuencia absoluta acumulada.

x_i	f_i	F_i
1	1	1
2	1	2
3	2	4
4	3	7
5	6	13
6	2	15
7	1	16
8	1	17
9	2	19
10	1	20
TOTAL	N = 20	

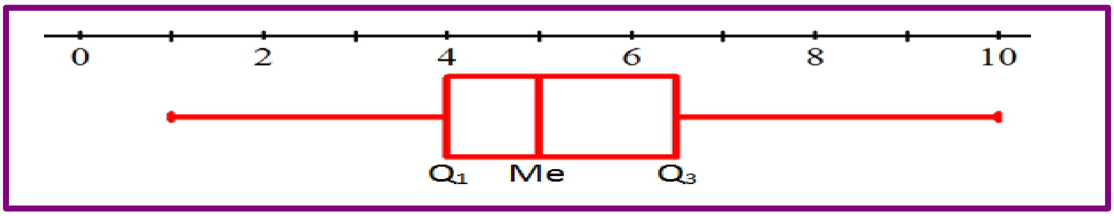
Así, tenemos que:
Recorrido: 10 - 1= 9.

Q_1 deja tras él la cuarta parte de la distribución. Como N = 20, tenemos que $\frac{N}{4} = 5$. Así, Q_1 debe dejar 5 elementos a su izquierda, por lo que debe ser la media del 5.º y 6.º elemento. Por lo tanto, $Q_1 = \frac{4+4}{2} = 4$.

Q_2 es la mediana que, como N es par, será la media de los elementos que ocupan las posiciones centrales, que son las posiciones 10.ª y 11.ª. Por lo tanto, $Q_2 = \frac{5+5}{2} = 5$.

Q_3 deja tras él las tres cuartas partes de la distribución. Como N = 20, tenemos que $\frac{3N}{4} = 15$. Así, Q_3 debe dejar 15 elementos a su izquierda, por lo que debe ser la media entre los elementos 15.º y el 16.º. Por lo tanto, $Q_3 = \frac{6+7}{2} = 6{,}5$.

El diagrama de caja y bigotes es el siguiente:

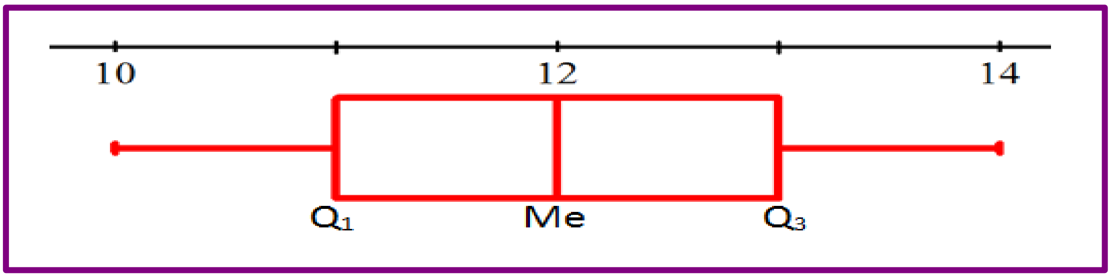


La caja de este diagrama comprende el recorrido intercuartílico en el que se marca la mediana y los bigotes alcanzan los valores máximos y mínimos de la distribución.

Actividad resuelta

Dibuje el diagrama de caja y bigotes de la actividad resuelta del apartado anterior (2.5.3)

Teníamos calculado que $Q_1 = 11$, $Q_2 = 12$ y $Q_3 = 13$. El valor mínimo era 10 y el máximo era 14.



Actividades propuestas

S24. Dibuje el diagrama de caja y bigotes correspondiente a la variable de la actividad propuesta número 18.

S25. Dibuje el diagrama de caja y bigotes correspondiente a la variable de la actividad propuesta número 19.

2.5.5 Interpretación conjunta de la media y de la desviación típica

El **coeficiente de variación** es el cociente entre la desviación típica y la media.

$$CV = \frac{\sigma}{\bar{x}}$$

Se expresa en porcentaje y sirve para medir la dispersión relativa. En general, es menos dispersa la variable que presenta menor coeficiente de variación, no la de menor desviación típica, pues la dispersión depende también de la media. No es lo mismo una desviación típica de 5 cuando la media vale 10 que una desviación típica de 5 cuando la media vale 1000. Las desviaciones típicas son iguales, pero en el caso de que la media sea 1000, hay menos dispersión relativa que cuando la media es 10.

Actividad resuelta

Las anotaciones de los jugadores de dos equipos en la liga ACB de baloncesto en la última temporada fueron los siguientes:

	Jug1	Jug2	Jug3	Jug4	Jug5	Jug6	Jug7	Jug8	Jug9	Jug10	Jug11	Jug12
Equipo 1	570	433	110	266	269	316	204	162	73	36	31	24
Equipo 2	352	388	304	194	197	293	267	182	186	188	148	145

Qué equipo tiene menor dispersión relativa.

Equipo 1:

Su media es:

$$\overline{x_1} = \frac{570+433+110+266+269+316+204+162+73+36+31+24}{12} = \frac{2494}{12} = 206{,}83$$

Su desviación típica es (antes calcularemos la varianza):

$${\sigma_1}^2 = \frac{570^2+433^2+110^2+266^2+269^2+316^2+204^2+162^2+73^2+36^2+31^2+24^2}{12} - 206{,}83^2 =$$

$$= \frac{843484}{12} - 206{,}83^2 = 70290{,}33 - 42778{,}65 = 27511{,}68 \Rightarrow \sigma_1 = 165{,}87$$

Equipo 2:

Su media es:

$$\overline{x_2} = \frac{352+388+304+194+197+293+267+182+186+188+148+145}{12} = \frac{2844}{12} = 237$$

Su desviación típica es (antes calcularemos la varianza):

$${\sigma_2}^2 = \frac{352^2+388^2+304^2+194^2+197^2+293^2+267^2+182^2+186^2+188^2+148^2+145^2}{12} - 237^2 =$$

$$= \frac{746440}{12} - 237^2 = 62228{,}33 - 56169 = 6059{,}33 \Rightarrow \sigma_2 = 77{,}84$$

Los coeficientes de variación son:

$CV_1 = \frac{\sigma_1}{\overline{x_1}} = \frac{165{,}87}{206{,}83} = 0{,}80$ para el equipo 1 y $CV_2 = \frac{\sigma_2}{\overline{x_2}} = \frac{77{,}84}{237} = 0{,}33$ para el equipo 2.

Por lo que, podemos concluir que la dispersión relativa es muchísimo menor en el equipo 2.

Actividades propuestas

S26. Calcule el coeficiente de variación de la variable de la actividad propuesta 18.

S27. Calcule el coeficiente de variación de la variable de la actividad propuesta 19.

3. Actividades finales

S28. Para cada una de las siguientes variables aleatorias, indique si es cualitativa, cuantitativa discreta o cuantitativa continua.

Longitud de una uña.	Tiempo que dura una película.
Puntos obtenidos en total en un partido de baloncesto.	Edad de un árbol.
Color de una flor.	Olor de un perfume.
Sabor de una fruta.	Número de trenes diarios entre A Coruña y Vigo.
Temperatura de un horno.	Peso de un paquete de azúcar.

S29. Las notas de 20 personas en una prueba de inglés son las siguientes: 5, 3, 4, 8, 5, 2, 1, 9, 6, 5, 4, 3, 8, 4, 2, 6, 9, 7, 8 y 5. Calcule la media, mediana, moda, rango, desviación media, desviación típica y coeficiente de variación; represente las frecuencias absolutas en un diagramas de sectores y las absolutas acumuladas en un diagrama de barras. Construya también el correspondiente diagrama de caja y bigotes.

S30. Las notas de un grupo de personas en una prueba de matemáticas son las siguientes:

Puntos	3	4	5	6	7	8	9
N.º personas	20	48	77	45	24	17	19

Calcule la media, mediana, moda, rango y desviación típica.

S31. Las notas de un grupo de personas en una prueba de lengua gallega son las siguientes:

Puntos	3	4	5	6	7	8	9
N.º personas	5	12	116	53	29	17	18

Calcule la media, mediana, moda, rango y desviación típica.

S32. El precio del gasóleo en los diferentes países de la Unión Europea el 3 de junio de 2017 eran:

País	Precio	País	Precio
Alemania	1,15 €/l	Grecia	1,329 €/l
Austria	1,018 €/l	Hungría	1,117 €/l
Bélgica	1,313 €/l	Reino Unido	1,359 €/l
Bulgaria	1,045 €/l	Irlanda	1,229 €/l
Chipre	1,166 €/l	Italia	1,404 €/l
Croacia	1,165 €/l	Letonia	1,01 €/l
Dinamarca	1,314 €/l	Lituania	0,942 €/l
Eslovaquia	1,035 €/l	Luxemburgo	0,972 €/l

Eslovenia	1,146 €/l	Malta	1,18 €/l
España	1,101 €/l	Polonia	0,987 €/l
Estonia	1,159 €/l	Portugal	1,289 €/l
Finlandia	1,26 €/l	República Checa	1,115 €/l
Francia	1,219 €/l	Romania	0,996 €/l
Holanda	1,285 €/l	Suecia	1,33 €/l

Agrupe los datos en clases de amplitud 0,05 siendo la primera de ellas la que va de 0,925 €/l hasta 0,975 €/l. Después dibuje el histograma correspondiente.

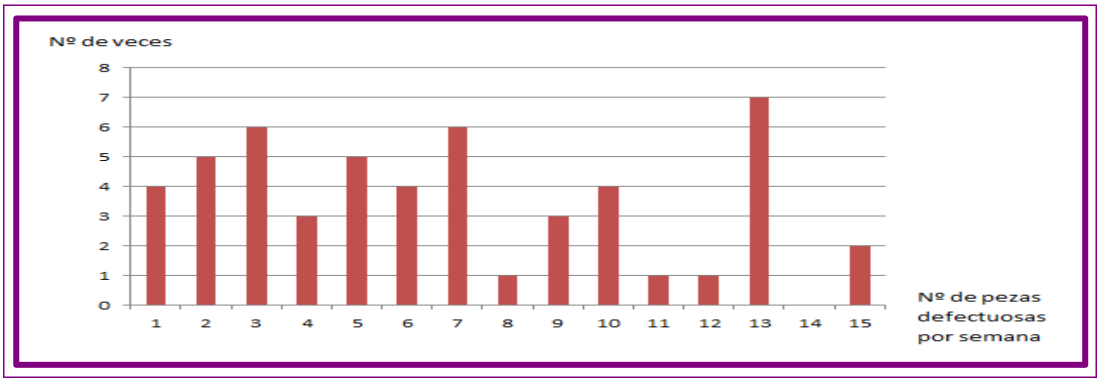
S33. Con los datos del ejercicio anterior agrupados en las clases que se indicaron, calcule el coeficiente de variación y dibuje el diagrama de caja y bigotes.

S34. En los doce meses de un año, en un centro meteorológico se obtuvieron los siguientes valores de temperatura máxima y mínima:

	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sept	Oct	Nov	Dic
Temp. min.	4,6	7,8	5,8	6,2	10,8	12,2	13,6	14,0	11,8	10,4	5,0	1,8
Temp. máx.	15,8	23	25,6	24	24,8	23	31,6	28,0	28,2	26,2	20,4	16,2

Indique cuál de estas dos variables tiene menor dispersión relativa.

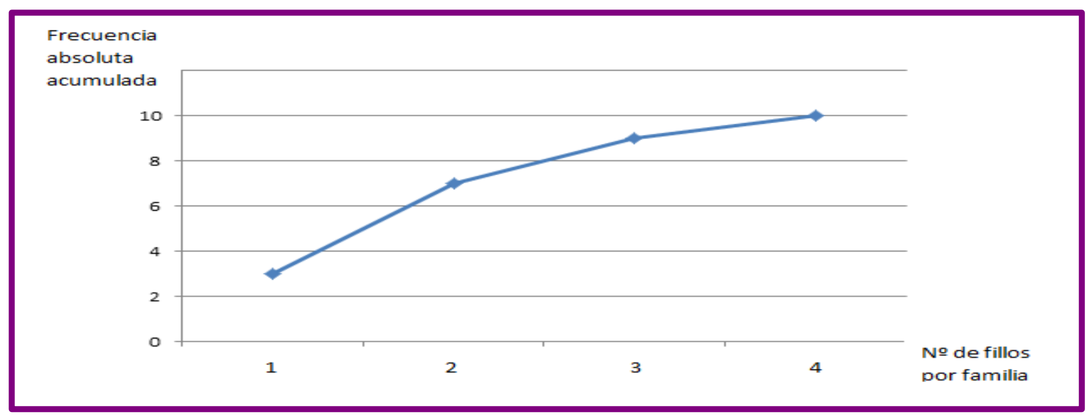
S35. En una fábrica de piezas para motores de coches se ha controlado durante un año el número de piezas defectuosas que se fabricaban cada semana. Los datos obtenidos aparecen recogidos en el siguiente diagrama de barras:



Calcule los parámetros de centralización, los de dispersión de esta variable y construya su diagrama de caja y bigotes.

S36. Se ha contabilizado el número de materias suspensas en la segunda evaluación por el alumnado de un curso de 2.º de la ESO. Los resultados fueron: 1, 3, 1, 0, 4, 0, 1, 1, 2, 3, 1, 1, 2, 1, 2, 4, 1, 0, 2, 3, 2, 3, 1, 1, 6, 0, 0, 2, 1, 4. Calcule el rango, la moda, la mediana, la media y la desviación media.

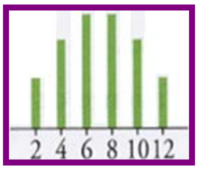
S37. Se ha preguntado a diez familias por el número de hijos e hijas que tenían y el resultado se muestra en el siguiente polígono de frecuencias en el que se representan las frecuencias absolutas acumuladas:



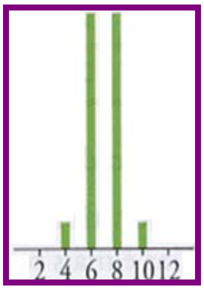
Calcule el coeficiente de variación de esta variable.

S38. Las tres distribuciones siguientes tienen la misma media, ¿cuál es?

La



B



C

Sus desviaciones típicas son 3’8, 1’3 y 2’9. Observando las gráficas, ¿a cuál corresponde cada una?

4. Solucionario

4.1 Soluciones de las actividades propuestas

S1. Las preguntas 1.ª y 3.ª no son apropiadas para la recogida de datos y la posterior elaboración de un estudio estadístico debido a la gran variedad de respuestas que podemos obtener. Las preguntas 2.ª y 4.ª, al estar más dirigidas, van a facilitar el tratamiento de los datos obtenidos con las encuestas.

S2. Es un muestreo por conglomerados. La población son las 30 personas de la clase que tiene 30 individuos. La muestra está formada por 3 individuos.

S3. Es un muestreo de tamaño 5 sobre la población "clientes que tienen la nómina domiciliada en alguna oficina del banco". El banco debe elaborar un muestreo aleatorio simple.

S4. Es una muestreo estratificado.

S5. Se trata de una muestreo oportunista y no fiable; la opinión de los viajeros entrevistados será, casi seguro, mejor que la de los que tomaron vuelos que sufrieron demora.

S6. Es un muestreo aleatorio simple.

S7. Es una muestreo aleatorio con reposición de tamaño 7 sobre una población de tamaño 5.

S8. Personas menores de 40 años: 1000; mujeres: 1090; hombres de 40 años o más: 460.

S9.

Horas diarias de sueño de los habitantes de una provincia. LA POBLACIÓN SON LOS HABITANTES DE LA PROVINCIA Y ES NECESARIO OBTENER UNA MUESTRA PARA HACER EL ESTUDIO, DEBIDO AL GRAN TAMAÑO DE LA POBLACIÓN.	Preferencias literarias de las personas mayores de edad que viven en un edificio. LA POBLACIÓN SON LAS PERSONAS MAYORES DE EDAD QUE VIVEN EN UN EDIFICIO Y NO ES NECESARIO OBTENER UNA MUESTRA PARA HACER EL ESTUDIO, YA QUE NADA IMPIDE PREGUNTAR A TODOS LOS INDIVIDUOS.

S10. Cualitativas: la música preferida, el color del pelo. Cuantitativas: el peso, la edad, las calificaciones de un examen.

S11. Discreta, ya que solo puede tomar valores naturales.

S12. Continua, ya que la manzana puede tomar cualquier valor dentro de un intervalo.

S13.

x_i	f_i	F_i	h_i	H_i
0	9	9	0,18	0,18
1	10	19	0,20	0,38
2	12	31	0,24	0,62
3	9	40	0,18	0,80
4	8	48	0,16	0,96
5	2	50	0,04	1
TOTAL	50		1	

S14.

x_i	f_i	h_i
MB	6	0,12
B	9	0,18
R	27	0,54
M	5	0,10
MM	3	0,06
TOTAL	50	1

No calculamos las frecuencias acumuladas por no ser datos numéricos.

a) 8 personas.

b) 30 %.

S15.

x_i	f_i	F_i	h_i	H_i
0	0	0	0	0
1	2	2	0,02	0,02
2	6	8	0,06	0,08
3	9	17	0,09	0,17
4	18	35	0,18	0,35
5	22	57	0,22	0,57
6	24	81	0,24	0,81
7	12	93	0,12	0,93
8	7	100	0,07	1
TOTAL	100		1	

a) 22 personas que representan el 22 %.

b) 43 personas que representan el 43 %.

S16.

Diagrama de barras

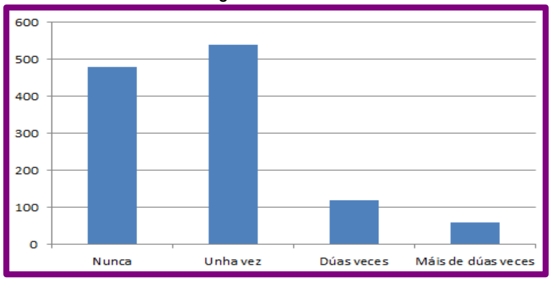
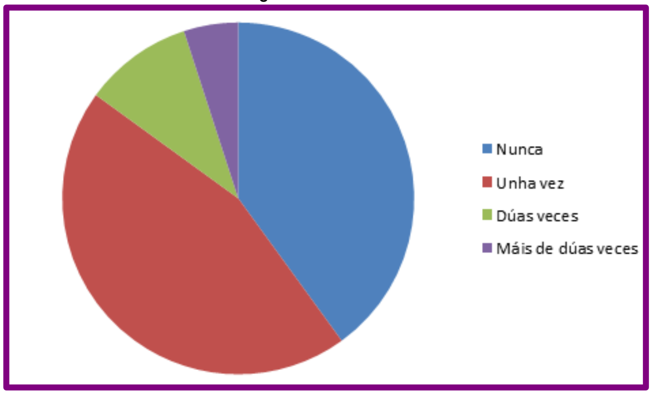
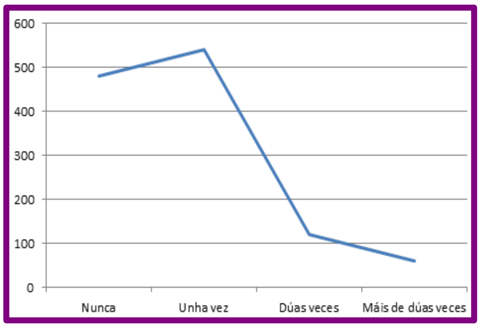


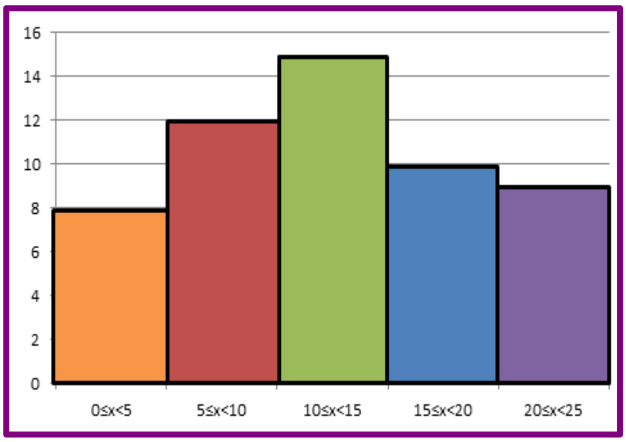
Diagrama de sectores



Polígono de frecuencias



S17. Como todos los intervalos tienen la misma longitud, es igual que hacer un diagrama de barras.



S18.

- Para calcular la media necesitamos tres columnas de la tabla: x_i, f_i y $x_i \cdot f_i$.
- Para calcular la mediana necesitamos tres columnas: x_i, f_i y F_i (en realidad f_i solo es necesaria para calcular F_i, no para calcular la mediana).
- Para calcular la moda necesitamos dos columnas: x_i y f_i.

Por lo tanto, para calcular estos tres parámetros de centralización necesitamos cuatro columnas de la tabla que son: x_i, f_i, F_i y $x_i \cdot f_i$.

x_i	f_i	F_i	$x_i \cdot f_i$
0	0	0	0
1	5	5	5
2	4	9	8
3	2	11	6
4	2	13	8
5	1	14	5
6	1	15	6
7	2	17	14
8	3	20	24
9	4	24	36
10	8	32	80
TOTAL	32		192

Así, tenemos:

- **Media:**
$$\bar{x} = \frac{192}{32} = 6$$
- **Mediana**: los lugares del medio (16.º y 17.º) están ocupados por el valor $x_i = 7$.
$$Me = \frac{7+7}{2} = 7$$
- **Moda**: hay una única moda que es:
$$Mo = 10$$

S19.

- Para calcular la media necesitamos tres columnas de la tabla: x_i, f_i y $x_i \cdot f_i$.
- Para calcular la mediana necesitamos tres columnas: x_i, f_i y F_i (en realidad f_i solo es necesaria para calcular F_i, no para calcular la mediana).
- Para calcular la moda necesitamos dos columnas: x_i y f_i.

Por lo tanto, para calcular estos tres parámetros de centralización necesitamos cuatro columnas de la tabla que son: x_i, f_i, F_i y $x_i \cdot f_i$.

x_i	f_i	F_i	$x_i \cdot f_i$
151	2	2	302
156	5	7	780
161	11	18	1771
166	14	32	2324
171	5	37	855
176	3	40	528
TOTAL	40		6560

Así, tenemos:

- **Media:**
$$\bar{x} = \frac{6560}{40} = 164$$
- **Mediana**: los lugares del medio (20.º y 21.º) están ocupados po-lo valor $x_i = 166$.
$$Me = \frac{166+166}{2} = 166$$
- **Moda**: hay una única moda que es:
$$Mo = 166$$

S20.

Para calcular los parámetros de dispersión necesitamos:

- Para el rango, determinar los valores máximo y mínimo de la variable.
- Para la desviación media, x_i, f_i y $x_i \cdot f_i$ para calcular la media y $|x_i - \bar{x}|$, f_i y $|x_i - \bar{x}| \cdot f_i$ para DM.
- Para la desviación típica, x_i, f_i y $x_i \cdot f_i$ para la media y $(x_i - \bar{x})$, $(x_i - \bar{x})^2$, f_i y $(x_i - \bar{x})^2 \cdot f_i$. para la varianza.

Por lo tanto, para calcular los parámetros de dispersión necesitamos ocho columnas de la tabla que son:

x_i, f_i, $x_i \cdot f_i$, $|x_i - \bar{x}|$, $|x_i - \bar{x}| \cdot f_i$, $(x_i - \bar{x})$, $(x_i - \bar{x})^2$, y $(x_i - \bar{x})^2 \cdot f_i$.

Las columnas $|x_i - \bar{x}|$ y $(x_i - \bar{x})$ son la misma excepto el signo, por lo que solo calcularemos la columna de $|x_i - \bar{x}|$.

En definitiva son siete columnas. Sabemos, por la actividad 18, que $\bar{x} = 6$.

x_i	f_i	F_i	$x_i \cdot f_i$	$\|x_i - \bar{x}\|$	$\|x_i - \bar{x}\| \cdot f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
0	0	0	0	6	0	36	0
1	5	5	5	5	25	25	125
2	4	9	8	4	16	16	64
3	2	11	6	3	6	9	18
4	2	13	8	2	4	4	8
5	1	14	5	1	1	1	1
6	1	15	6	0	0	0	0
7	2	17	14	1	2	1	2
8	3	20	24	2	6	4	12
9	4	24	36	3	12	9	36
10	8	32	80	4	32	16	128
TOTAL	32		192		104		394

- Rango: $\text{rango} = 10 - 0 = 10$
- Desviación media: $DM = \frac{104}{32} = 3{,}25$
- Varianza: $\sigma^2 = \frac{394}{32} = 12{,}31$
- Desviación típica: $\sigma = \sqrt{12{,}31} = 3{,}51$

S21.

Necesitamos las mismas columnas de la tabla que en la actividad anterior y sabemos, por la actividad 19, que $\bar{x} = 164$.

x_i	f_i	F_i	$x_i \cdot f_i$	$\|x_i - \bar{x}\|$	$\|x_i - \bar{x}\| \cdot f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
151	2	2	302	13	26	169	338
156	5	7	780	8	40	64	320
161	11	18	1771	3	33	9	99
166	14	32	2324	2	28	4	56
171	5	37	855	7	35	49	245
176	3	40	528	12	36	144	432
TOTAL	40		6560		198		1490

- Rango: $\text{rango} = 176 - 151 = 25$
- Desviación media: $DM = \frac{198}{40} = 4{,}95$
- Varianza: $\sigma^2 = \frac{1490}{40} = 37{,}25$
- Desviación típica: $\sigma = \sqrt{37{,}25} = 6{,}10$

S22.

Para calcular los cuartiles necesitamos la frecuencia absoluta acumulada.

x_i	f_i	F_i
0	0	0
1	5	5
2	4	9
3	2	11
4	2	13
5	1	14
6	1	15
7	2	17
8	3	20
9	4	24
10	8	32
TOTAL	32	

Así, tenemos:

- Q_1 deja tras él la cuarta parte de la distribución. Como N = 32, tenemos que $\frac{N}{4} = 8$. Así, Q_1 debe dejar 8 elementos a su izquierda, por lo que debe estar entre el 8.º y el 9.º elemento. Por lo tanto, $Q_1 = \frac{2+2}{2} = 2$.
- Q_2 es la mediana y, como N es par, será la media de los elementos que ocupan las posiciones centrales, que son las posiciones 16.ª y 17.ª. Por lo tanto, $Q_2 = \frac{7+7}{2} = 7$.
- Q_3 deja tras él las tres cuartas partes de la distribución. Como N = 32, tenemos que $\frac{3N}{4} = 24$. Así, Q_3 debe dejar 24 elementos a su izquierda, por lo que debe estar entre el 24.º y el 25.º elemento. Por lo tanto, $Q_3 = \frac{9+10}{2} = 9{,}5$.

S23.

Para calcular los cuartiles necesitamos la frecuencia absoluta acumulada.

x_i	f_i	F_i
151	2	2
156	5	7
161	11	18
166	14	32
171	5	37
176	3	40
TOTAL	40	

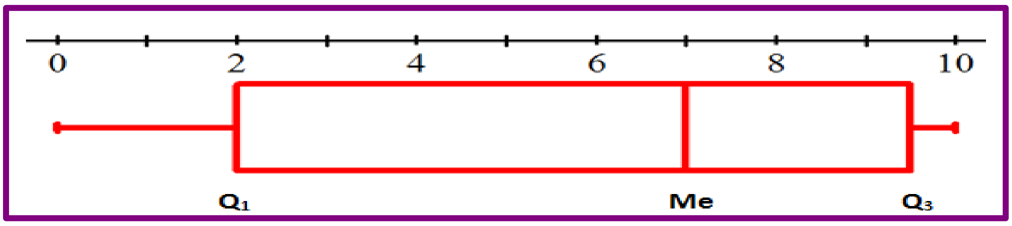
Así, tenemos:

- Q1 deja tras él la cuarta parte de la distribución. Como N = 40, tenemos que $\frac{N}{4} = 10$. Así, Q1 debe dejar 10 elementos a su izquierda, por lo que debe estar entre el 10.º y el 11.º elemento. Por lo tanto, Q1 = $\frac{161+161}{2} = 161$.
- Q2 es la mediana y, como N es par, será la media de los elementos que ocupan las posiciones centrales, que son las posiciones 20.ª y 21.ª. Por lo tanto, Q2 = $\frac{166+166}{2} = 166$.
- Q3 deja tras él las tres cuartas partes de la distribución. Como N = 40, tenemos que $\frac{3N}{4} = 30$. Así, Q3 debe dejar 30 elementos a su izquierda, por lo que debe estar entre el 30.º y el 31.º elemento. Por lo tanto, Q3 = $\frac{166+166}{2} = 166$.

S24.

Por la actividad 22 sabemos que $Q_1 = 2$, $Q_2 = 7$ y $Q_3 = 9{,}5$. También sabemos que el valor mínimo de la variable es 0 y el mayor es 10. Así:

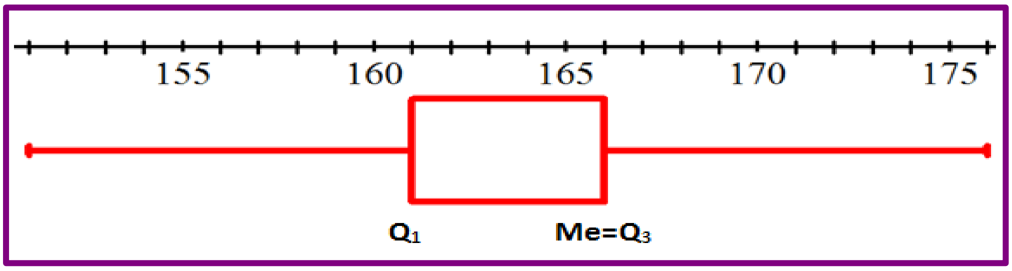
El diagrama de caja y bigotes es el siguiente:



S25.

Por la actividad 23 sabemos que Q_1 = 161, Q_2 = 166 y Q_3 = 166. También sabemos que el valor mínimo de la variable es 151 y el mayor es 176. Así:

El diagrama de caja y bigotes es el siguiente:



S26. Por la actividad 18 sabemos que $\bar{x} = 6$ y por la actividad 20 sabemos que $\sigma = 3{,}51$. Por lo tanto, el coeficiente de variación es $CV = \frac{3{,}51}{6} = 0{,}59$.

S27. Por la actividad 19 sabemos que $\bar{x} = 164$ y por la actividad 21 sabemos que $\sigma = 6{,}10$. Por lo tanto, el coeficiente de variación es $CV = \frac{6{,}10}{164} = 0{,}04$.

4.2 Soluciones de las actividades finales

S28.

Longitud de una uña. *Cuantitativa continua.*	Tiempo que dura una película. *Cuantitativa continua.*
Puntos obtenidos en total en un partido de baloncesto. *Cuantitativa discreta.*	Edad de un árbol. *Cuantitativa continua*, aunque lo normal es tratarla como discreta.
Color de una flor. *Cualitativa.*	Olor de un perfume. *Cualitativa.*
Sabor de una fruta. *Cualitativa*	Número de trenes diarios entre A Coruña y Vigo. *Cuantitativa discreta.*
Temperatura de un horno. *Cuantitativa continua,* aunque lo normal es tratarla como discreta.	Peso de un paquete de azúcar *Cuantitativa continua.*

S29.

x_i	f_i	F_i	$x_i \cdot f_i$	$\lvert x_i - \bar{x}\rvert$	$\lvert x_i - \bar{x}\rvert \cdot f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
1	1	1	1	4,2	4,2	17,64	17,64
2	2	3	4	3,2	6,4	10,24	20,48
3	2	5	6	2,2	4,4	4,84	9,68
4	3	8	12	1,2	3,6	1,44	4,32
5	4	12	20	0,2	0,8	0,04	0,16
6	2	14	12	0,8	1,6	0,64	1,28
7	1	15	7	1,8	1,8	3,24	3,24
8	3	18	24	2,8	8,4	7,84	23,52
9	2	20	18	3,8	7,6	14,44	28,88
TOTAL	20		104		38,8		109,2

- Media: $\bar{x} = \frac{104}{20} = 5{,}2$
- Mediana: $Me = \frac{5+5}{2} = 5$
- Moda: $Mo = 5$
- Desviación media: $DM = \frac{38{,}8}{20} = 1{,}94$
- Varianza: $\sigma^2 = \frac{109{,}2}{20} = 5{,}46$
- Desviación típica: $\sigma = \sqrt{5{,}46} = 2{,}34$
- Rango: $rango = 9 - 1 = 8$
- Coeficiente de variación: $CV = \frac{2{,}34}{5{,}2} = 0{,}45$
- 1.er cuartil: $Q_1 = 3{,}5$
- 3.er cuartil: $Q_3 = 7{,}5$
- Recorrido intercualtílico: $7{,}5 - 3{,}5 = 4$

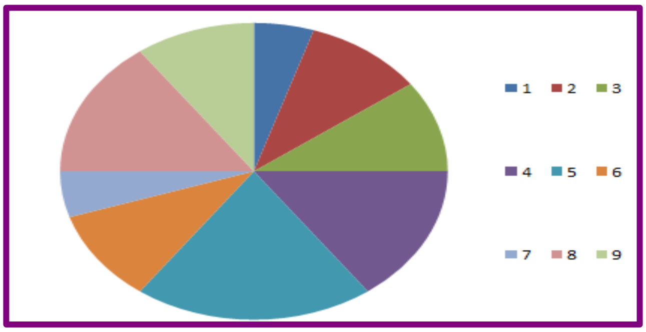


Diagrama de sectores de las frecuencias absolutas

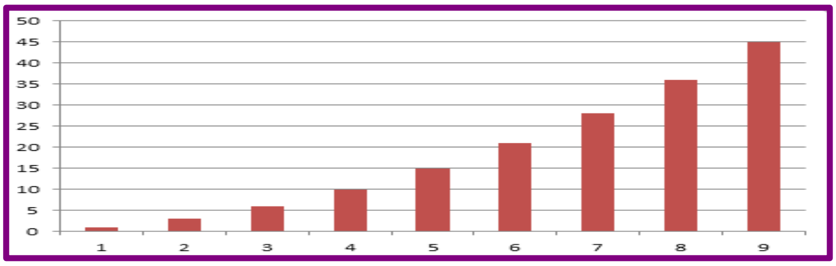
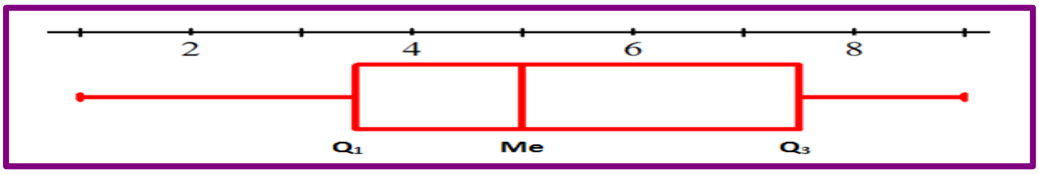


Diagrama de barras de las frecuencias absolutas acumuladas

Diagrama de caja y bigotes



S30.

x_i	f_i	F_i	$x_i \cdot f_i$	x_i^2	$x_i^2 \cdot f_i$
3	20	20	60	9	180
4	48	68	192	16	768
5	77	145	385	25	1925
6	45	190	270	36	1620
7	24	214	168	49	1176
8	17	231	136	64	1088
9	19	250	171	81	1539
TOTAL	250		1382		8296

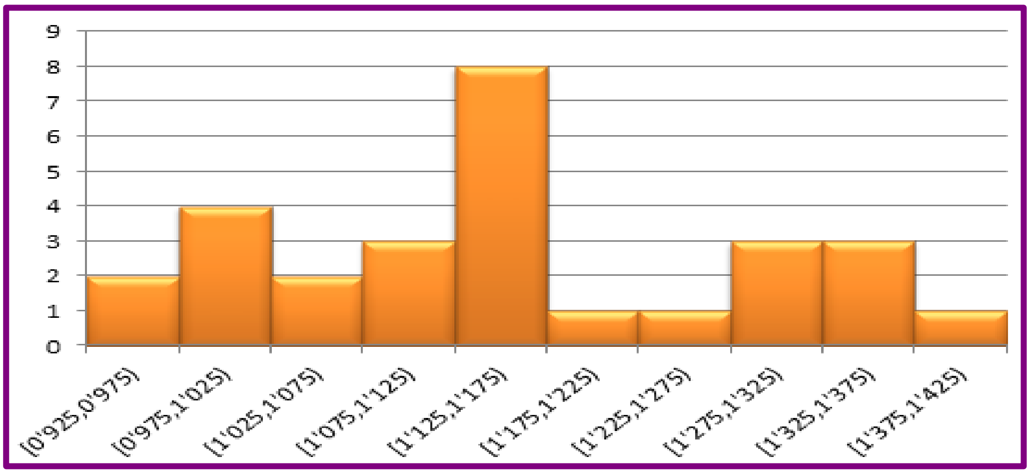
- Media: $\bar{x} = \frac{1382}{250} = 5{,}53$
- Mediana: $Me = \frac{5+5}{2} = 5$
- Moda: $Mo = 6$
- Varianza: $\sigma^2 = \frac{8296}{250} - 5{,}53^2 = 2{,}63$
- Desviación típica: $\sigma = \sqrt{5{,}46} = 1{,}62$
- Rango: $rango = 9 - 3 = 6$

S31.

x_i	f_i	F_i	$x_i \cdot f_i$	$\boldsymbol{x_i^2}$	$\boldsymbol{x_i^2 \cdot f_i}$
3	5	5	15	9	45
4	12	17	48	16	192
5	116	133	580	25	2900
6	53	186	318	36	1908
7	29	215	203	49	1421
8	17	232	136	64	1088
9	18	250	162	81	1458
TOTAL	250		1462		9012

- Media: $\bar{x} = \frac{1462}{250} = 5{,}85$
- Mediana: $Me = \frac{5+5}{2} = 5$
- Moda: $Mo = 5$
- Varianza: $\sigma^2 = \frac{9012}{250} - 5{,}85^2 = 1{,}85$
- Desviación típica: $\sigma = \sqrt{5{,}46} = 1{,}36$
- Rango: $rango = 9 - 3 = 6$

S32.



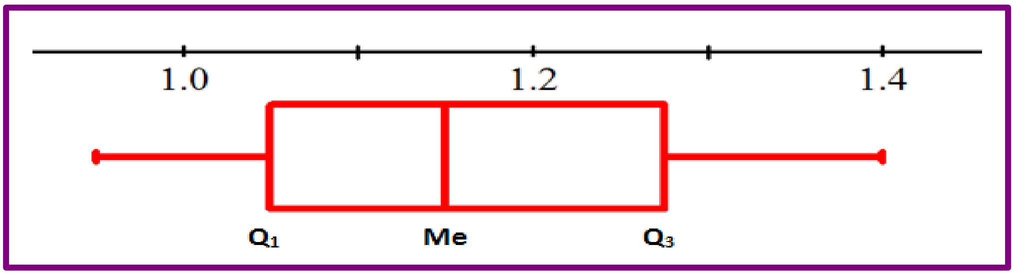
S33.

Clase	x_i	f_i	F_i	$x_i \cdot f_i$	$x_i{}^2$	$x_i{}^2 \cdot f_i$
[0'925,0'975)	0,95	2	2	1,9	0,9025	1,805
[0'975,1'025)	1	4	6	4	1	4
[1'025,1'075)	1,05	2	8	2,1	1,1025	2,205
[1'075,1'125)	1,1	3	11	3,3	1,21	3,63
[1'125,1'175)	1,15	8	19	9,2	1,3225	10,58
[1'175,1'225)	1,2	1	20	1,2	1,44	1,44
[1'225,1'275)	1,25	1	21	1,25	1,5625	1,5625
[1'275,1'325)	1,3	3	24	3,9	1,69	5,07
[1'325,1'375)	1,35	3	27	4,05	1,8225	5,4675
[1'375,1'425)	1,4	1	28	1,4	1,96	1,96
	TOTAL	28		32,3		35,915

- Media:
  $\bar{x} = \frac{32,3}{28} = 1,15$
- Mediana:
  $Me = \frac{1,15 + 1,15}{2} = 1,15$
- Rango:
  $rango = 1,4 - 0,95 = 0,45$

- Varianza:
  $\sigma^2 = \frac{35,915}{28} - 1,15^2 = 0,0164$
- Desviación típica:
  $\sigma = \sqrt{5,46} = 0,13$

- Coeficiente de variación:
  $CV = \frac{0,13}{1,15} = 0,11$
- 1.er cuartil:
  $Q_1 = \frac{1,05 + 1,05}{2} = 1,05$
- 3.er cuartil:
  $\bar{x} = \frac{1,25 + 1,3}{2} = 1,275$



S34.

x_i	f_i	$x_i \cdot f_i$	x_i^2	$x_i^2 \cdot f_i$
4,6	1	4,6	21,16	21,16
7,8	1	7,8	60,84	60,84
5,8	1	5,8	33,64	33,64
6,2	1	6,2	38,44	38,44
10,8	1	10,8	116,64	116,64
12,2	1	12,2	148,84	148,84
13,6	1	13,6	184,96	184,96
14	1	14	196	196,00
11,8	1	11,8	139,24	139,24
10,4	1	10,4	108,16	108,16
5	1	5	25	25,00
1,8	1	1,8	3,24	3,24
TOTAL	12	104		1076,16

TEMPERATURAS MÍNIMAS

- Media:
$$\bar{x} = \frac{104}{12} = 8,67$$
- Varianza:
$$\sigma^2 = \frac{1076,16}{12} - 8,67^2 = = 14,57$$
- Desviación típica:
$$\sigma = \sqrt{14,57} = 3,82$$
- Coeficiente de variación:
$$CV = \frac{3,82}{8,67} = 0,44$$

x_i	f_i	$x_i \cdot f_i$	x_i^2	$x_i^2 \cdot f_i$
15,8	1	15,8	249,64	249,64
23	1	23	529	529,00
25,6	1	25,6	655,36	655,36
24	1	24	576	576,00
24,8	1	24,8	615,04	615,04
23	1	23	529	529,00
31,6	1	31,6	998,56	998,56
28	1	28	784	784,00
28,2	1	28,2	795,24	795,24
26,2	1	26,2	686,44	686,44
20,4	1	20,4	416,16	416,16
16,2	1	16,2	262,44	262,44
TOTAL	12	286,8		7096,88

TEMPERATURAS MÁXIMAS

- Media:
$$\bar{x} = \frac{286,8}{12} = 23,9$$
- Varianza:
$$\sigma^2 = \frac{7096,88}{12} - 23,9^2 = = 20,20$$
- Desviación típica:
$$\sigma = \sqrt{20,20} = 4,49$$
- Coeficiente de variación:
$$CV = \frac{4,49}{23,9} = 0,19$$

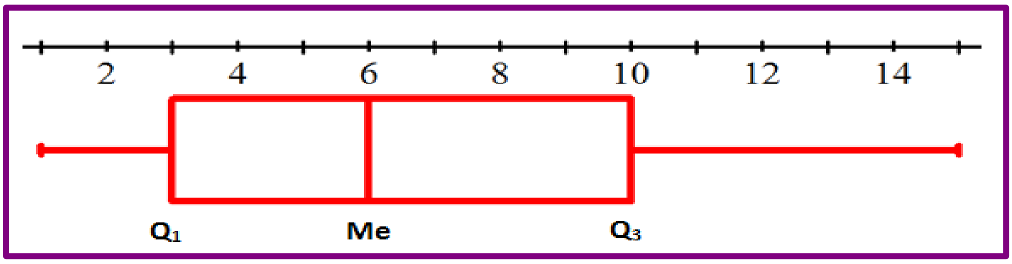
Tiene menos dispersión relativa la de las temperaturas mínimas por tener el coeficiente de variación más pequeño.

S35.

x_i	f_i	F_i	$x_i \cdot f_i$	$\|x_i - \bar{x}\|$	$\|x_i - \bar{x}\| \cdot f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
1	4	4	4	5,81	23,24	33,7561	135,02
2	5	9	10	4,81	24,05	23,1361	115,68
3	6	15	18	3,81	22,86	14,5161	87,10
4	3	18	12	2,81	8,43	7,8961	23,69
5	5	23	25	1,81	9,05	3,2761	16,38
6	4	27	24	0,81	3,24	0,6561	2,62
7	6	33	42	0,19	1,14	0,0361	0,22
8	1	34	8	1,19	1,19	1,4161	1,42
9	3	37	27	2,19	6,57	4,7961	14,39
10	4	41	40	3,19	12,76	10,1761	40,70
11	1	42	11	4,19	4,19	17,5561	17,56
12	1	43	12	5,19	5,19	26,9361	26,94
13	7	50	91	6,19	43,33	38,3161	268,21
14	0	50	0	7,19	0	51,6961	0
15	2	52	30	8,19	16,38	67,0761	134,15
TOTAL	52		354		181,62		884,08

- Media: $\bar{x} = \frac{354}{52} = 6{,}81$
- Mediana: $Me = \frac{6+6}{2} = 6$
- Moda: $Mo = 13$
- Desviación media: $DM = \frac{181{,}62}{52} = 3{,}49$
- Varianza: $\sigma^2 = \frac{884{,}08}{52} = 17{,}00$
- Desviación típica: $\sigma = \sqrt{17{,}00} = 4{,}12$
- 1.er cuartil: $Q_1 = 3$
- 3.er cuartil: $\mathbf{Q_3 = 10}$
- Rango: $rango = 15 - 1 = 14$

Diagrama de caja y bigotes



S36.

x_i	f_i	F_i	$x_i \cdot f_i$	$\|x_i - \bar{x}\|$	$\|x_i - \bar{x}\| \cdot f_i$
0	5	5	0	6,81	34,05
1	11	16	11	5,81	63,91
2	6	22	12	4,81	28,86
3	4	26	12	3,81	15,24
4	3	29	12	2,81	8,43
5	0	29	0	1,81	0
6	1	30	6	0,81	0,81
TOTAL	30		53		151,3

- Media: $\bar{x} = \frac{53}{30} = 1{,}77$
- Mediana: $\mathrm{Me} = \frac{1+1}{2} = 1$
- Moda: $\mathrm{Mo} = 1$
- Desviación media: $\mathrm{DM} = \frac{151{,}3}{30} = 5{,}04$
- Rango: $\mathrm{rango} = 6 - 0 = 6$

S37.

x_i	f_i	$x_i \cdot f_i$	x_i^2	$x_i^2 \cdot f_i$
0	3	0	0	0
1	4	4	1	4
2	2	4	4	8
3	1	3	9	9
TOTAL	10	11		21

- Media:

$$\bar{x} = \frac{11}{10} = 1{,}1$$

- Varianza:

$$\sigma^2 = \frac{21}{10} - 1{,}1^2 = 0{,}89$$

- Desviación típica:

$$\sigma = \sqrt{0{,}89} = 0{,}94$$

- Coeficiente de variación:

$$CV = \frac{0{,}94}{1{,}1} = 85{,}45$$

S38. $\bar{x} = 7$. La desviación típica de 3,8 corresponde a la gráfica C, la de 2,9 es la de la gráfica A y la de 1,3 es la de la gráfica B.

5. Glosario

C	▪ Coeficiente de variación	El cociente entre la desviación típica y la media.
	▪ Cuartiles	Números que dividen los datos de la distribución ya ordenados en cuatro partes iguales.
D	▪ Desviación media	Es la media aritmética de las distancias de los datos a la media.
	▪ Desviación típica	Es la raíz cuadrada de la varianza.
	▪ Diagrama de caja y bigotes	Forma muy clara de representar los cuartiles y el recorrido de la variable.
E	▪ Estadística descriptiva o deductiva	Parte de la estadística que trata del recuento, la ordenación y la clasificación de los datos obtenidos en las observaciones.
	▪ Estadística inferencial o inductiva	Parte de la estadística que formula y resuelve el problema de establecer previsiones y deducciones generales sobre una población a partir resultados obtenidos de una muestra.
F	▪ Frecuencia absoluta	Número de veces que aparece un valor.
	▪ Frecuencia absoluta acumulada de un valor x_i	Número de veces que aparece el valor x_i o cualquiera de los anteriores a él.
	▪ Frecuencia relativa	Frecuencia absoluta de un valor determinado dividida polo número total de datos.
	▪ Frecuencia relativa acumulada	Frecuencia absoluta acumulada dividida entre el número total de datos.
I	▪ Individuo	Cada uno de los miembros de una población.
	▪ Intervalos de clase	Cada uno de los intervalos en los que se agrupan los datos, normalmente de una variable continua.
M	▪ Marca de clase	Punto medio del intervalo de clase.
	▪ Media	Es el valor central de la distribución.
	▪ Mediana	Es el valor que se encuentra en el medio, es decir, es el valor que deja tantos individuos antes como después.
	▪ Moda	Es el valor que tiene la mayor frecuencia absoluta.
	▪ Muestra	Parte de una población.
P	▪ Parámetros de centralización	Son los que indican alrededor de qué valor se distribuyen los datos.
	▪ Parámetros de dispersión	Son los que informan sobre cuánto se separan del centro los valores de la distribución.
	▪ Parámetros de posición	Son los que dividen un conjunto de datos en grupos con el mismo número de individuos.
	▪ Recorrido o rango	Es la diferencia entre los datos mayor y menor de la variable.
	▪ Recorrido intercuartílico	Distancia entre los cuartiles 3.º y 1.º.
	▪ Población	Conjunto de personas, objetos o individuos sobre los que se quiere estudiar una determinada característica.

R	▪ Rango o recorrido	Es la diferencia entre los datos mayor y menor de la variable.
V	▪ Variable estadística	Cada una de las propiedades a las que se pretenda aplicar un estudio estadístico.
	▪ Variables estadísticas cualitativas	Variables estadísticas que no se pueden expresar como valores numéricos.
	▪ Variables estadísticas cuantitativas	Variables estadísticas que pueden tomar valores numéricos.
	▪ Varianza	Es la media aritmética de los cuadrados de las distancias de los datos a la media.

6. Bibliografía y recursos

Bibliografía

- *Libros para la educación secundaria a distancia de adultos. Ámbito tecnológico-matemático.* Consellería de Educación e Ordenación Universitaria.
- *Matemáticas ESO 1.* Ed. Anaya, 2016.
- *Matemáticas ESO 2.* Ed. Anaya, 2016.
- *Matemáticas orientadas a las enseñanzas académicas ESO 3.* Ed. Anaya, 2016.
- *Matemáticas orientadas a las enseñanzas aplicadas ESO 3.* Ed. Anaya, 2016.
- *Matemáticas enseñanzas académicas ESO 3.* Ed. Santillana.
- *Matemáticas enseñanzas aplicadas ESO 3.* Ed. Santillana.

Enlaces de Internet

En estos enlaces encontrará trucos e información que puede consultar para mejorar su práctica.

- https://matematicasiesoja.wordpress.com/estadistica/
- http://www.vitutor.com/estadistica/descriptiva/estadistica.html
- http://recursostic.educacion.es/descartes/web/materiales_didacticos/estadistica_1_ciclo/indice.htm
- https://campusvirtual.ull.es/ocw/pluginfile.php/6104/mod_resource/content/1/tema6/PR6-estdescriptiva.pdf
- http://www3.uji.es/~mateu/t1-alumnos.pdf
- https://matematicasiesoja.files.wordpress.com/2013/10/ejerciciosresueltos1.pdf
- http://www.est.uc3m.es/esp/nueva_docencia/getafe/ciencias_empresariales/metodestad_empresa/doc_generica/archivos/Ejercicios%20resueltos%20Tema%201.pdf
- http://www.est.uc3m.es/esp/nueva_docencia/getafe/ciencias_empresariales/metodestad_empresa/doc_generica/archivos/Ejercicios%20resueltos%20Tema%202.pdf