Estadística
I$^{\text{o}}$ Medio

Profesor
Alberto Alvaradejo Ojeda

Índice

1. Estadística

Métodos y procedimientos que implican recopilación, presentación, ordenación y análisis de datos, con el objetivo que a partir de estos puedan inferirse conclusiones.
Pueden distinguirse dos ramas diferentes en Estadística

1.1. Estadística Descriptiva

Es la que se utiliza en la descripción y análisis de conjuntos de datos o población.

1.2. Inferencia Estadística

Es la que hace posible la estimación de una característica de una población, o la toma de una decisión con respecto a una población, con base únicamente en resultados muestrales.

2. Glosario

2.1. Población o Universo

Conjunto completo de individuos, objetos, o medidas los cuales poseen una característica común.

2.2. Muestra

Es un subconjunto o una porción de la población.

2.3. Variable

Característica de una población o muestra que será estudiada, la que puede tomar diferentes valores.

2.4. Variables Cuantitativas

Se pueden expresar en forma numérica. Se dividen en discretas y continuas.

2.4.1. Variables Cuantitativas Discretas

Son respuestas numéricas que surgen de un proceso de conteo, siendo siempre un número entero.

Ejemplo 2.1 *:*

- *Número de asignaturas del curso*
- *Número de integrantes del grupo familiar.*
- *Número de salas de clases del colegio.*

2.4.2. Variables Cuantitativas Continuas

Son respuestas numéricas que surgen de un proceso de medición, las cuales pueden tomar valores entre dos números enteros.

Ejemplo 2.2 *:*

- *Estatura.*
- *Temperatura.*
- *Peso.*

2.5. Datos

Números o medidas que han sido recopiladas como resultado de la observación.

2.6. Estadístico

Es una medida, un valor que se calcula para describir una característica a partir de una sola muestra.

2.7. Parámetro

Es una característica cuantificable de una población.

3. Tabulación de los datos

La tabulación consiste en presentar los datos estadísticos en forma de tablas o cuadros. En los experimentos estadísticos los datos recolectados pueden corresponder a una población o muestra. En ambos casos los procedimientos de resumen de datos son análogos y se designan por:

$$N = \text{Tamaño de la población estudiada}$$
$$n = \text{Tamaño de la muestra}$$

3.1. Tabla de distribución de frecuencias

Es una tabla en la que se disponen los datos divididos en grupos ordenados numéricamente y que se denominan clases o categorías.

3.1.1. Tabulación de datos de variables cualitativas

Para construir una tabla de distribución de frecuencias con datos cualitativos debemos enumerar los atributos con su frecuencia absoluta f_i

Frecuencia absoluta f_i: indica el número de veces que se repite el atributo de la variable.

Ejemplo 3.1 *En una muestra de 200 trabajadores de una empresa se obtuvieron los siguientes resultados al preguntarles por su situación civil.*

Estado Civil	f_i
Soltero	35
Casado	100
Viudo	5
Separado	60
Total	200

Cuadro 1: Estado civil de 200 personas

3.1.2. Tabulación de datos de variables cuantitativas discretas

Las tablas de distribución de frecuencia de esta variable llevan, al menos, cinco columnas.

a) **Frecuencia absoluta** f_i: Número de veces que se repite la variable.

b) **Tamaño de la muestra** (n): Indica la cantidad de elementos que conforman la muestra, se obtiene sumando todas las frecuencias absolutas.

$$n = f_1 + f_2 + f_3 + \cdots + f_m \tag{3.1}$$

m = número de intervalos

c) **Frecuencia relativa** h_i: Se obtiene dividiendo la frecuencia absoluta de cada intervalo por el tamaño de la muestra.

$$h_i = \frac{f_i}{n} \tag{3.2}$$

La suma se estas frecuencia debe ser igual a 1

$$0 \leqslant h_i \leqslant 1 \tag{3.3}$$

d) **Frecuencia absoluta acumulada** F_i: Es la suma de las frecuencias absolutas de los valores menores o iguales al valor de la variable en cuestión. El último valor de esta debe ser igual al número total de datos.

e) **Frecuencia relativa acumulada** H_i: Es la suma de las frecuencias relativas de los valores menores o iguales al valor de la variable en cuestión. El último valor de esta debe ser igual a 1.

Ejemplo 3.2 *Cien familias se han clasificado según el número de hijos, resultando los siguientes datos:*

Número de hijos	Frecuencia Absoluta f_i	Frec. Abs. Acumulada F_i	Frecuencia Relativa h_i	Frec. Rel. Acumulada H_i
0	11	11	0,11	0,11
1	13	24	0,13	0,24
2	20	44	0,2	0,44
3	25	69	0,25	0,69
4	14	83	0,14	0,83
5	10	93	0,1	0,93
6	4	97	0,04	0,97
7	2	99	0,02	0,99
8	1	100	0,01	1
	N=100		1	

Cuadro 2: Tabulación de datos de variables cuantitativas discretas

Análisis

- *El último valor de la distribución de frecuencias absolutas acumuladas coincide con N.*
- *El último valor de la distribución de frecuencias relativas acumuladas es 1 (salvo error de redondeo).*
- *La distribución de frecuencias acumulada nos permite conocer la proporción (o el número) de observaciones por debajo de cierto valor, entre dos valores o por encima de una cantidad.*
- *¿Qué proporción de familias tiene menos de 2 hijos? R. 0,24*
- *¿Cuántas familias tienen menos de 4 hijos? R. 69*
- *¿Qué proporción de familias tiene más de 6 hijos?*
  $0,03 = 1 - 0,97 = 0,01 + 0,02$
- *¿Qué proporción de familias tiene más de 3 hijos pero menos de 7?*
  $0,28 = 0,14 + 0,1 + 0,04 = 0,97 - 0,69$

3.1.3. Tabulación de datos de variables cuantitativas continuas

Para tabular este tipo de variables es necesario determinar:

a) **Rango o recorrido**: es la diferencia que toma el valor máximo y el mínimo de la variable.

$$R = x_{Max} - x_{Min} \tag{3.4}$$

b) **Número de intervalos o clases (m)**: Es el número de grupos en que es posible dividir los valores de la variable.
Un número pequeño de clases puede ocultar la naturaleza general de los datos y un número muy grande puede ser demasiado detallado como para revelar alguna información útil. Se recomienda que el número de clases esté entre cinco y veinte. Hay una regla llamada Regla de Sturges que puede dar una aproximación razonable para el número de clases:

$$m = 1 + 3,3\log_{10}(n) \tag{3.5}$$

n número de datos de la muestra

c) **Amplitud del intervalo o amplitud de la clase (α)**:

$$\alpha = \frac{\text{Recorrido}}{\text{N}^o \text{ de clases}} = \frac{R}{m} \tag{3.6}$$

d) **Límites de un intervalo**: Son los valores extremos de una clase. El menor valor es considerado como el límite inferior y el valor que se obtiene sumando al límite inferior la amplitud del intervalo es el límite inferior de la segunda clase.

e) **Límites reales de un intervalo**: Se obtienen calculando el promedio entre el límite superior de una clase y el límite inferior de la clase siguiente.

f) **Marca de clase (x_i)**: Punto medio de un intervalo.

g) **Frecuencia absoluta (f_i)**: Es el número de observaciones que pertenece a un intervalo dado.

h) **Frecuencia relativa (h_i)**: Es la proporción de datos que se encuentra en un intervalo, se determina dividiendo la frecuencia absoluta del intervalo por el tamaño de la muestra.

$$h_i = \frac{f_i}{n} \tag{3.7}$$

i) **Frecuencia absoluta acumulada** F_i: Indica el número de datos de la muestra menores o iguales al límite real superior del intervalo i.

j) **Frecuencia relativa acumulada** H_i: Indica la proporción de datos de la muestra menores o iguales al límite real superior del intervalo (i).

Ejemplo 3.3 *Los siguientes datos se recopilaron con el fin de determinar la edad de 50 estudiantes de II^o Medio del Colegio Especial Montaña*

15	16	17	18	19	15	20	18	20	17
15	16	15	19	25	15	30	42	15	20
15	16	19	20	16	15	16	20	20	42
16	17	17	20	19	18	19	60	42	22
19	19	25	17	25	31	20	25	30	42

Solución

a) Rango

$$R = X_{Max} - X_{Min}$$

$$R = 60–15 = 45$$

b) Número de datos

$$n = 50$$

c) m, número de intervalos

$$m = 1 + 3,3\log_{10}(50) = 6,64 \approx 7$$

d) Amplitud de los intervalos

$$\alpha = \frac{\textit{Recorrido}}{\textit{N}^o \textit{ de clases}} = \frac{45}{7} = 6,4$$

Clase	Intervalo	Marca de Clase	Frecuencia absoluta	Frec. abs. acumulada	Frecuencia relativa	Frec. rel. acumulada
Nro.	Nro.	x_i	f_i	F_i	h_i	H_i
1	15 - 22	18,5	37	37	0,74	0,74
2	22 - 29	25,5	5	42	0,10	0,84
3	29 - 36	32,5	3	45	0,06	0,90
4	36 - 43	39,5	4	49	0,08	0,98
5	43 - 50	46,5	0	49	0,00	0,98
6	50 - 57	53,5	0	49	0,00	100
7	57 - 64	60,5	1	50	0,02	
	Total		50		1,00	

Cuadro 3: Edad de 50 estudiantes de IIº Medio

Al elaborar la columna de las frecuencias absolutas, un valor muestral coincide con uno de los límites del intervalo, convenimos en tomar ese valor en aquella clase donde aparece como límite inferior del intervalo. Es decir, son intervalos cerrados – abiertos. Por ejemplo, el valor 22 que aparece como límite superior del primer intervalo pertenece a la segunda clase.
El último intervalo lo tomamos cerrado para que el x_{Max} y los valores que coinciden con él no queden fuera de la tabla.
Un análisis de la tabla de distribución de frecuencias nos permite afirmar:

- *37 estudiantes de los IIº Medios tienen unan edad entre 15 y 22 años correspondientes al 74 % de la muestra.*
- *De los 50 estudiantes 49 son menores de 43 años, lo cual corresponde al 98 % de la muestra tomada.*
- *Las edades más frecuentes están entre los 15 y 22 años, por tener esta clase la máxima frecuencia absoluta.*

3.2. Ejercicios

1. En una industria es necesario realizar un estudio respecto al peso de engranajes de gran tamaño. Los siguientes datos corresponden al peso, en kilogramos, de 30 de estas piezas, que poseen las mismas dimensiones, pero distinta aleación.

58	52	50	52	40	50	38	52	50	45
36	45	55	42	42	52	50	45	42	38
42	38	40	46	45	45	55	42	45	40

   a) Construir una tabla de frecuencias de amplitud 5 comenzando desde 36.
   b) ¿Cuántos engranajes pesan entre 46 y 55 Kg.?
   c) ¿Qué porcentaje representa a aquellos engranajes cuyo peso es inferior a 51 Kg.?
   d) ¿Cuál es la frecuencia relativa para aquel intervalo cuya marca de clase es 48?
   e) ¿Qué porcentaje representa a aquellas piezas que pesan más de 50 Kg.?

2. En una industria automotriz es necesario realizar un estudio debido a una partida defectuosa de discos de embrague. Para ello se ha recopilado la siguiente información referente a la duración en horas de 50 de ellos.

285	300	286	302	313	314	289	292	321	327
293	289	292	289	308	326	303	287	293	322
304	329	295	307	297	302	294	301	285	313
308	307	304	291	288	297	316	322	317	308
321	324	323	316	292	286	299	294	328	296

   a) Construir una tabla de frecuencia de amplitud cinco comenzando desde 285.
   b) ¿Cuántos discos duraron entre 290 y 299 horas?.
   c) ¿Cuántos discos no alcanzaron a durar 300 horas?.
   d) ¿Qué porcentaje representan los discos que duraron entre 310 y 314 horas?.
   e) ¿Qué porcentaje representan los discos que duraron menos de 305 horas?.
   f) ¿Cuántos discos duraron más de 309 horas?.
   g) ¿Cuántos discos duraron menos de 305 horas?.
   h) ¿Qué porcentaje representan los discos que duraron entre 285 y 294 horas?.
   i) ¿Cuál es el intervalo de mayor frecuencia absoluta?.

3. En un conjunto habitacional se pretende hacer un estudio del número de personas que consumen productos enlatados. Los datos que han sido obtenidos de 50 bloques del conjunto habitacional son:

63	69	83	85	93	73	81	94	104	125
64	90	75	138	133	110	60	91	87	136
137	134	129	96	99	72	104	97	84	98

   a) Construir una tabla de frecuencia de amplitud 10 partiendo desde 60.
   b) ¿Cuántas personas consumen entre 100 y 129 productos enlatados?
   c) ¿Qué porcentaje representa a las personas que consumen menos de 90 productos enlatados?
   d) ¿Qué cantidad de personas consumen más de 80 productos enlatados?

4. Las ganancias por acción de 40 compañías de la industria de la construcción son

4,6	0,3	1,1	5,7	0,1	1,3	2,5	1,6
1,3	2,1	2,1	1,4	7,3	5,4	3,5	1,9
6,0	0,8	1,9	2,1	3,2	0,2	7,1	2,8
9,6	3,7	5,1	3,6	4,9	2,3	1,8	0,4
4,2	2,1	0,9	3,2	3,7	1,1	0,5	1,9

a) Construya una distribución de frecuencias que comience en 0,1 y tenga una amplitud de 2,0

b) ¿Cuál es la frecuencia absoluta del tercer intervalo?. Interprete el resultado.

c) Qué porcentaje de las compañías tienen a lo más una ganancia de 6,0?

d) ¿Cuántas compañías tienen una ganancia a lo menos de 4,1?

e) Interprete la frecuencia acumulada del segundo intervalo.

f) Interprete la frecuencia relativa acumulada del cuarto intervalo.

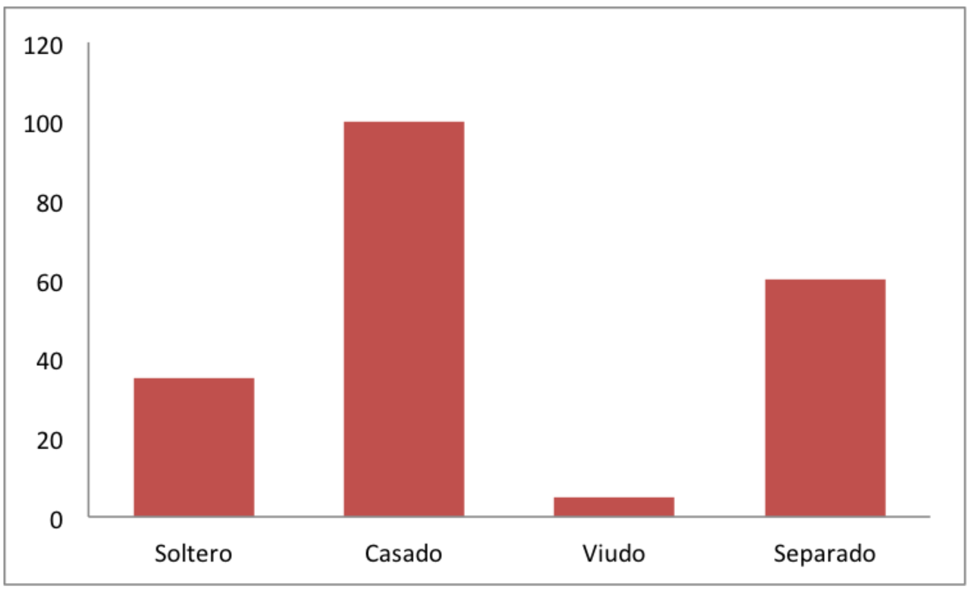
4. Gráficos

Un gráfico estadístico es una representación visual de una serie de datos. Es una herramienta muy eficaz, ya que permite:

- captar la atención del lector;
- presentar la información de forma sencilla, clara y precisa;
- facilitar la comparación de datos y destaca las tendencias y las diferencias;
- ilustrar el mensaje, tema o trama del texto al que acompaña.

4.1. Gráfico de barras

Un gráfico de barras es una representación gráfica en un eje cartesiano de las frecuencias de una variable cualitativa o discreta.



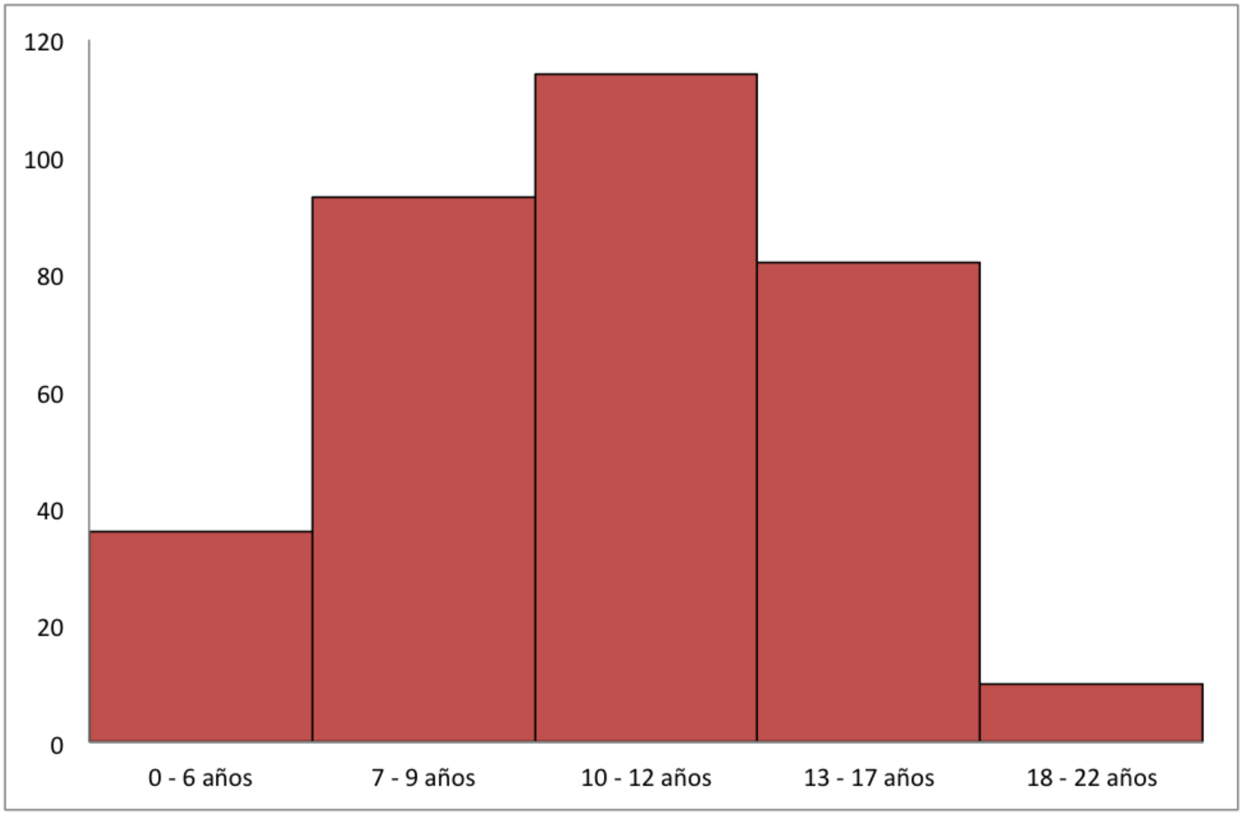
Estado Civil	f_i
Soltero	35
Casado	100
Viudo	5
Separado	60
Total	200

Cuadro 4: Estado civil

En uno de los ejes se posicionan las distintas categorías o modalidades de la variable cualitativa o discreta (estado civil) y en el otro el valor o frecuencia de cada categoría en un determinada escala (número de casos).

4.2. Histograma

Se usa para representar las frecuencias de una variable cuantitativa continua. En uno de los ejes se posicionan las clases de la variable continua (los intervalos o las marcas de clase que son los puntos medios de cada intervalo) y en el otro eje las frecuencias. No existe separación entre las barras.



Clases	f_i	F_i	h_i	H_i
0 - 6 años	36	36	0,10	0.10
7 - 9 años	93	129	0,27	0.38
10 - 12 años	114	243	0,34	0.72
13 - 17 años	82	325	0,24	0.97
18 - 22 años	10	335	0,03	1
Total	335		1	

Cuadro 5: Años de escolaridad

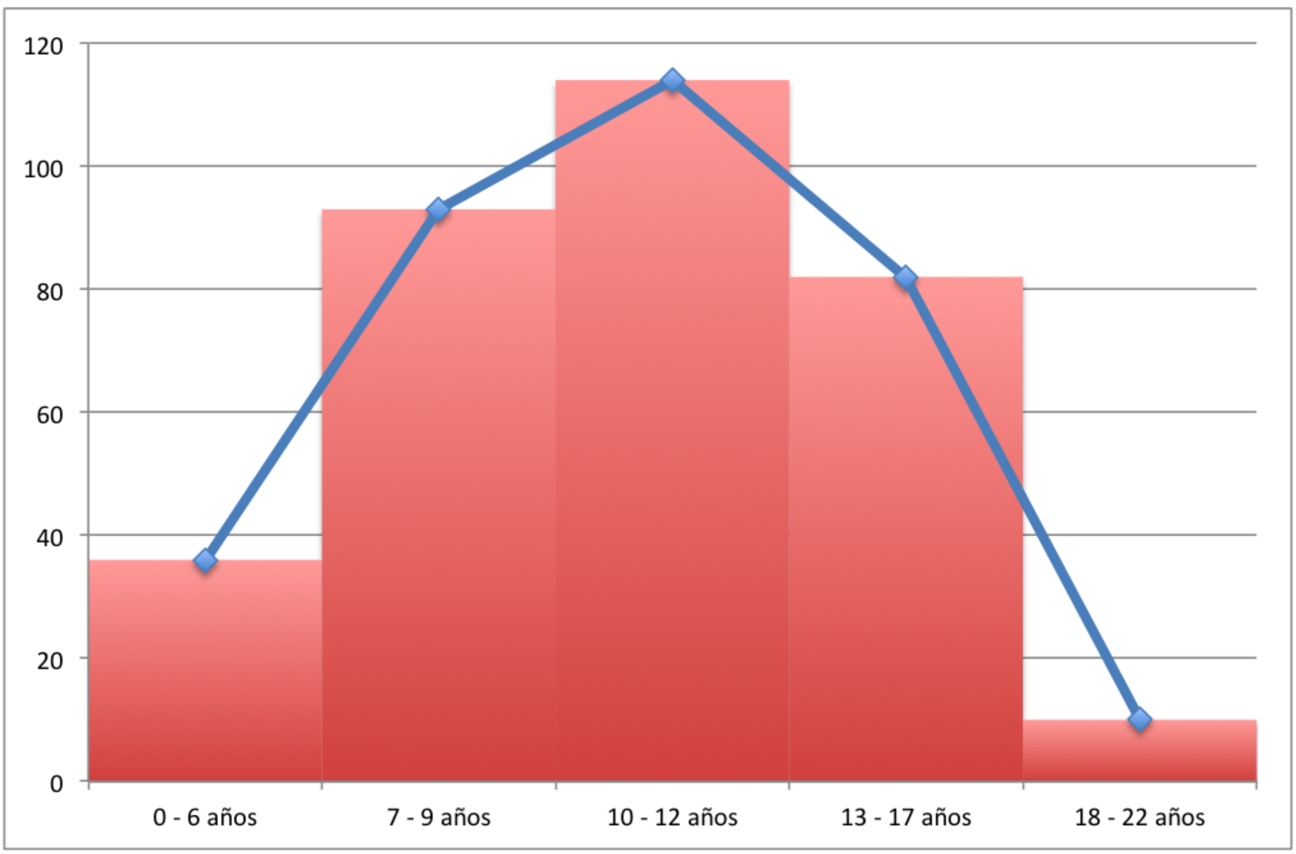
4.3. Polígono de frecuencias

Se crea a partir de un histograma de frecuencia. Se forma a partir de la unión de las marcas de clases columnas que configuran lo que es un histograma de frecuencia.

Los polígonos de frecuencia para datos agrupados son aquellos que se desarrollan mediante la marca de clase que tiene coincidencia con el punto medio de las distintas columnas del histograma.
Vamos a utilizar los datos del ejemplo anterior y graficaremos la marca de clase.

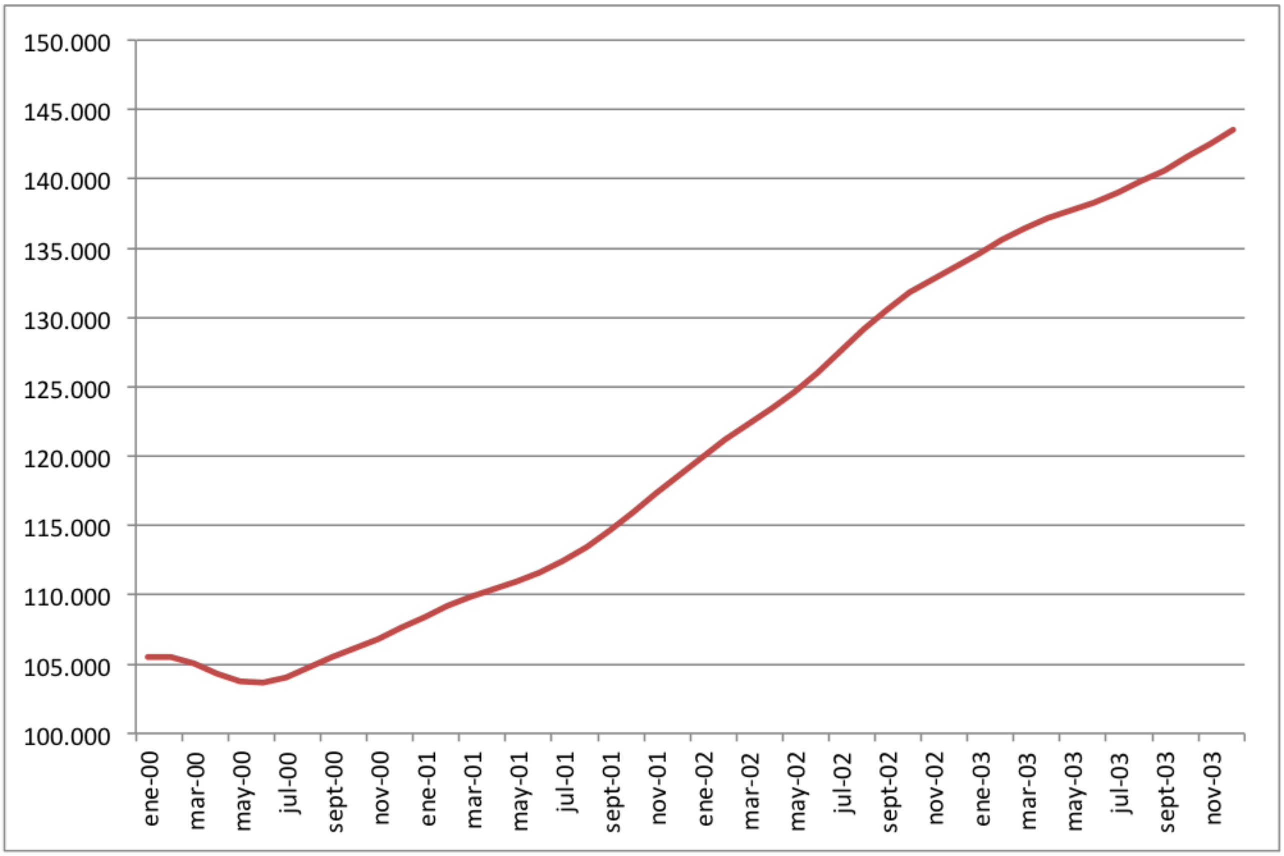
Clases	f_i	F_i	h_i	H_i
0 - 6 años	36	36	0,10	0.10
7 - 9 años	93	129	0,27	0.38
10 - 12 años	114	243	0,34	0.72
13 - 17 años	82	325	0,24	0.97
18 - 22 años	10	335	0,03	1
Total	335		1	

Cuadro 6: Años de escolaridad



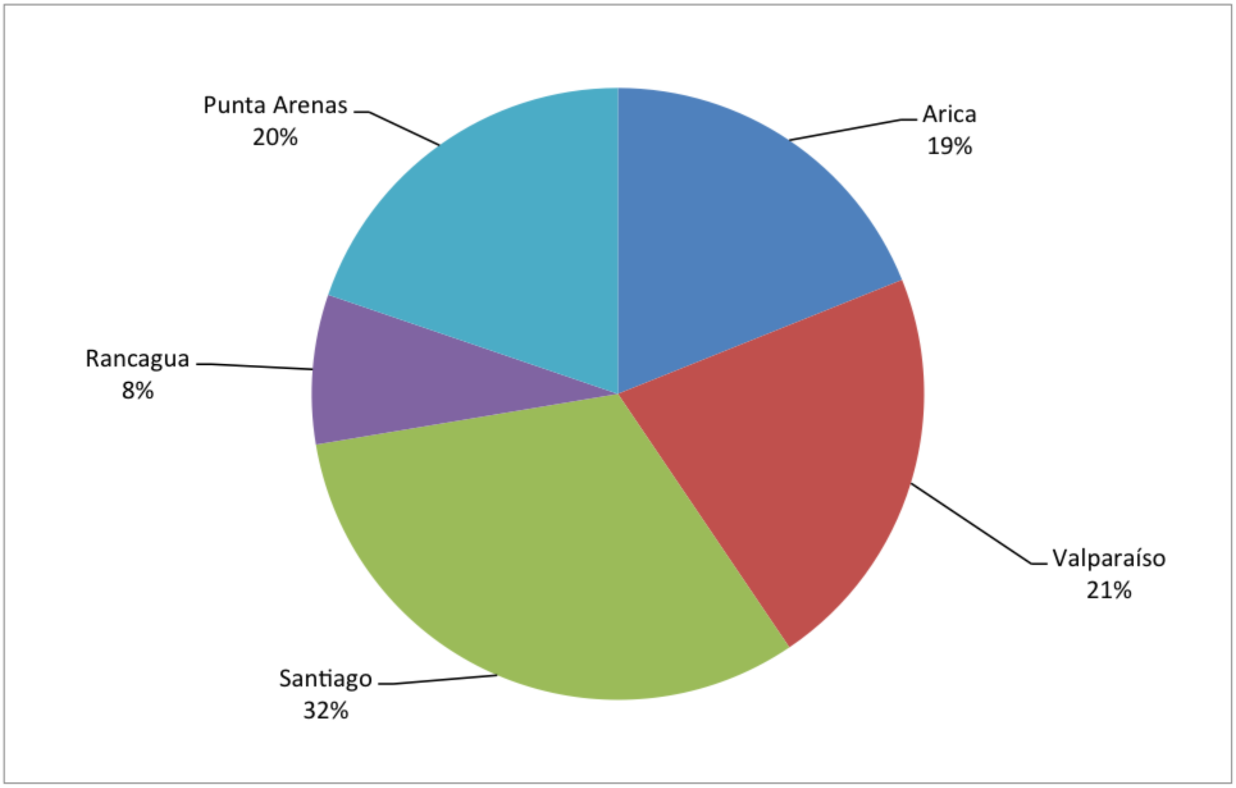
4.4. Gráfico de líneas

Es una representación gráfica en un eje cartesiano de la relación que existe entre dos variables reflejando con claridad los cambios producidos.



4.5. Gráfico de sectores

Un gráfico de sectores es una representación circular de las frecuencias relativas de una variable cualitativa o discreta que permite, de una manera sencilla y rápida, su comparación.



Tienda	Total ventas
Arica	91
Valparaíso	104
Santiago	153
Rancagua	38
Punta Arenas	95
TOTAL	481

Cuadro 7: Volumen de ventas

El círculo representa la totalidad que se quiere observar (total de ventas) y cada porción, llamadas sectores, representan la proporción de cada categoría de la variable (Ventas por ciudad) respecto el total. Suele expresarse en porcentajes.

Obtención de los ángulos de cada sector:

$$\text{Angulo}=\text{frecuencia relativa} \cdot 360^{o} \tag{4.1}$$

Son útiles cuando las categorías son pocas. Si el gráfico tuviera muchas variables, no aportaría casi información y sería prácticamente incomprensible.

4.6. Ejercicios

1. Las puntuaciones obtenidas por un grupo de alumnos en una prueba han sido: 15, 20, 15, 18, 22, 13, 13, 16, 15, 19, 18, 15, 16, 20, 16, 15, 18, 16, 14, 13.
   Construir la tabla de distribución de frecuencias y dibuja el polígono de frecuencias.

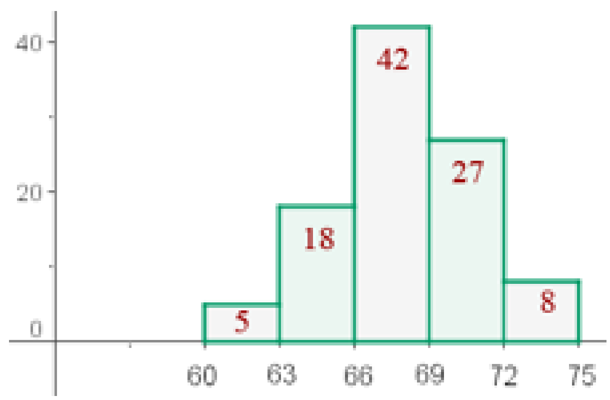
2. Los pesos de los 65 empleados de una fábrica vienen dados por la siguiente tabla:

Peso	50-60	60-70	70-80	80-90	90-100	100-100	110-120
f_i	8	10	16	14	10	5	2

   Construir la tabla de frecuencias y representar el histograma y el polígono de frecuencias.

3. Los 40 alumnos de una clase han obtenido las siguientes puntuaciones, sobre 50, en un examen de Física.
   3, 15, 24, 28, 33, 35, 38, 42, 23, 38, 36, 34, 29, 25, 17, 7, 34, 36, 39, 44, 31, 26, 20, 11, 13, 22, 27, 47, 39, 37, 34, 32, 35, 28, 38, 41, 48, 15, 32, 13.
   Construir la tabla de frecuencias. Dibujar el histograma y el polígono de frecuencias.

4. El histograma de la distribución corresponde al peso de 100 alumnos de IV^o Medio:



   Elabora la tabla de frecuencias que corresponde a este gráfico.

5. En un penal de Mujeres a 44 internas se le solicito su estado civil señalando que su estado actual es el siguiente: casada 12, soltera 20, divorciada 8, viuda 4. Construir un gráfico circular que represente la situación anterior.

6. Resumen estadístico Enero a Agosto 2014-2015 a nivel nacional para delitos con mayor connotación social:

Delito	**Total de casos**
Robo con violencia	13.553
Robo con intimidación	27.619
Robo con sorpresa	26.045
Lesiones	57.387
Homicidio	345
Violación	1.304

   Construir el gráfico de barras que represente la situación anterior.

5. Medidas de tendencia central

La mayor parte de los conjuntos de datos muestran una tendencia a agruparse alrededor de un punto central y por lo general es posible elegir algún valor que describa todo un conjunto de datos. Un valor descriptivo como ese es una medida de tendencia central o posición.

5.1. Media aritmética

Es la medida de tendencia central de mayor uso. Se calcula sumando todas las frecuencias absolutas de un conjunto de datos, dividiendo después ese total entre el número total de elementos.

5.1.1. Datos no agrupados

Media aritmética de una muestra:

$$\overline{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \sum_{i=1}^{n} \frac{x_i}{n} \tag{5.1}$$

n =tamaño de la muestra

Ejemplo 5.1 *Calcular la media aritmética de los notas obtenidas por un alumno en la signatura de matemática.*

$$4,5; 3,6; 5,8; 6,5; 6,3; 6,7$$

$$\overline{x} = \frac{4,5 + 3,6 + 5,8 + 6,5 + 6,3 + 6,7}{6} = \frac{33,4}{6} = 5,56 \approx 5,6$$

5.1.2. Datos agrupados

Cuando los datos están ordenado en una tabla de distribución de frecuencia utilizamos:

$$\overline{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + ... + x_m f_m}{f_1 + f_2 + f_3 + ... + f_m} = \sum_{i=1}^{m} \frac{x_i f_i}{n} \tag{5.2}$$

Donde:

- x_i es la marca de clase del intervalo i-ésimo
- f_i es la frecuencia del intervalo i-ésimo
- n es el número de datos de la muestra
- m es el número de intervalos

Ejemplo 5.2 *Calcular la media aritmética para el peso de 40 personas, como lo muestra la tabla:*

Peso (kg.)	x_i	f_i	$x_i f_i$
55 - 62	58,5	5	292,5
63 - 70	66,5	15	997,5
71 - 78	74,5	12	894,0
79 - 86	82,5	5	412,5
87 - 94	90,5	3	271,5
Total		40	2.868,0

Cuadro 8: Peso de 40 personas

$$\overline{x} = \sum_{i=1}^{5} \frac{x_i f_i}{n} = \frac{2{,}868}{40} = 71,7 \approx 72$$

5.2. Mediana

La mediana es el valor que se encuentra en el centro de una secuencia ordenada de datos. La mediana no se ve afectada por observaciones extremas en un conjunto de datos. Por ello, cuando se presenta alguna información extrema, resulta apropiado utilizar la mediana, y no la media, para describir el conjunto de datos. Su símbolo es M_e

5.2.1. Datos no agrupados

Se ordenan los datos de forma creciente o decreciente. Para muestras con un número par de observaciones, la mediana es el dato que queda en el centro de la muestra y para muestras con número impar de observaciones la mediana es el promedio de los dos datos centrales.

a) Muestra con un número impar de datos

$$M_e = \frac{X_{n+1}}{2} \tag{5.3}$$

Datos = 4,7,5,6,3,2,7
Datos ordenados = 2,3,4,**5**,6,7,7

$$M_e = \frac{X_{7+1}}{2} = x_4 = 5$$

b) Muestra con un número par de datos

$$M_e = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} \tag{5.4}$$

Datos = 12,15,14,16,11,10,10,13
Datos ordenados = 16,15,14,**13**,**12**,11,10,10

$$M_e = \frac{X_{\frac{8}{2}} + X_{\frac{8}{2}+1}}{2} = \frac{X_4 + X_5}{2} = \frac{13+12}{2} = 12,5$$

5.2.2. Datos agrupados

$$M_e = L_i + \left(\frac{\frac{n}{2} - F_{i-1}}{f_i} \right) \cdot a \tag{5.5}$$

Donde:

- i es el primer intervalo cuya frecuencia acumulada supera a $\frac{n}{2}$
- L_i es el límite real inferior del intervalo de la mediana.
- n es el número de datos.
- F_{i-1} es la frecuencia acumulada anterior al intervalo de la mediana.
- f_i es la frecuencia absoluta del intervalo de la mediana.
- a es la amplitud del intervalo.

Ejemplo 5.3 *La tabla muestra la distribución de frecuencias de la duración, en horas, de uso continuo de 212 dispositivos electrónicos iguales, sometidos a un cierto control. El intervalo donde se encuentra la Mediana*

Nro.	Duración	f_i	F_i
1	350 - 399	4	4
2	400 - 449	6	10
3	450 - 499	9	19
4	500 - 549	20	39
5	550 - 599	31	**70**
6	**600 - 649**	80	150
7	650 - 699	42	192
8	700 - 749	10	202
9	750 - 799	8	210
10	800 - 849	2	212
	Total	212	

Cuadro 9: Duración en horas de dispositivos electrónicos

es el primer intervalo en el cual $\frac{n}{2} \leqslant F_i$

$$\frac{n}{2} = \frac{212}{2}$$

$$106 \leqslant F_i$$

$$106 \leqslant 150$$

$$en\ el\ 6^o\ intervalo\ i = 6$$

$$a = 50$$

$$F_5 = 70$$

$$f_6 = 80$$

$$L_6 = 599,5$$

$$M_e = 599,5 + \left(\frac{106 - 70}{80} \right) \cdot 50 = 622\ horas$$

5.3. Moda

La moda es el valor de un conjunto de datos que aparece con mayor frecuencia. Se le obtiene fácilmente a partir de un arreglo ordenado. A diferencia de la media aritmética, la moda no se ve afectada ante la ocurrencia de valores extremos.

Sin embargo, sólo se utiliza la moda para propósitos descriptivos porque es más variable, para distintas muestras, que las demás medidas de tendencia central. Un conjunto de datos puede tener más de una moda o ninguna. Su símbolo es M_o

5.3.1. Datos no agrupados

- Datos 2,4,5,6,7,7,8,7,6

$$M_o = 7$$

- Datos 1,1,3,1,1,2,2,4,2,3,2,5,6

$$M_o = 1 \; y \; 2$$

- 0,0,2,3,4,5

$$M_o = 0$$

- Datos 0,1,2,3,4,5

$$M_o = \text{No existe}$$

5.3.2. Datos agrupados

$$M_o = L_i + \left(\frac{f_{i+1}}{f_{i-1} + f_{i+1}} \right) \cdot a \tag{5.6}$$

Donde:

- i es el intervalo de mayor frecuencia absoluta.
- a es la amplitud del intervalo

Ejemplo 5.4 *De la tabla anterior.*

$$i = 6$$
$$f_{i+1} = f_{6+1} = f_7 = 42$$
$$f_{i-1} = f_{6-1} = f_5 = 31$$
$$L_i = L_6 = 599,5$$
$$a = 50$$
$$M_o = 599,5 + \left(\frac{42}{31 + 42} \right) \cdot 50 \approx 628,26 \; horas$$

5.4. Ejercicios

1. En una industria dos operarios en siete días de trabajo, son capaces de producir, por día, y en forma individual la siguiente cantidad de árboles para fresa de 250 mm de longitud por 300 mm de diámetro.

Operario A	105	106	104	102	103	100	101
Operario B	103	102	107	101	105	102	103

   Determine:

   a) Producción media de cada operario.

   b) Moda del operario A.

   c) Mediana del operario B.

2. Se hace una encuesta entre 100 personas acerca del número de horas diarias que se dedican a ver televisión, obteniéndose la siguiente información:

Nº de Horas	f_i
0 - 1	30
2 - 3	20
4 - 5	15
6 - 7	32
8 - 9	1
10 - 11	2
Total	100

   Calcular la media, la mediana y la moda.

3. De un total de 100 datos, 20 son 4, 40 son 5, 30 son 6 y el resto 7. Hallar la media y la moda.

4. Cuatro grupos de estudiantes, consistentes en 15, 20, 10 y18 individuos, dieron pesos de 60, 72, 55 y 65 kilos. Hallar el peso medio de los estudiantes.

5. Los puntajes de un estudiante en sus pruebas han sido 84, 91, 72, 68, 87 y 78. Hallar la media, la mediana y la moda.

6. La siguiente tabla corresponde a la estatura de 80 estudiantes de una determinada carrera.

Estatura mts.	f_i
1,65 - 1,69	6
1,70 - 1,74	12
1,75 - 1,79	30
1,80 - 1,84	22
1,85 - 1,89	8
1,90 - 1,94	2
Total	80

   Hallar la media, mediana y moda de la estatura.

7. La oficina de Censo, proporcionó las edades de hombres y mujeres divorciados (en miles de personas de 15 años de edad o más).

Edad	Hombre	Mujer
15 - 19	2	2
20 - 24	80	210
25 - 29	174	303
30 - 34	210	315
35 - 39	385	656
40 - 44	450	656
45 - 49	295	409
50 - 54	174	200
Total	1.770	2.751

Obtener las medidas de tendencia central

8. Se toma una muestra de 12 estudiantes matriculados en estadística y se les pregunta por el número de horas que emplearon en estudiar la asignatura en la semana anterior al examen final:

12	7	4	16	21	5
9	3	11	14	10	6

Determinar:

a) Media muestral

b) Mediana muestral

6. Medidas de posición

Son medidas que dan cuenta de una determinada posición dentro de la distribución. La **mediana** es, además de una medida de tendencia central, una medida de posición. Asume valores que informan sobre una posición en la distribución de valores de la variable. Se ubica en el "medio" de la distribución dejando la mitad de las unidades por debajo y la otra mitad por encima.

Puede ser interesante saber cuánto gana como máximo el 25 % de los trabajadores peor remunerados de una empresa, o cuánto gana como mínimo el 25 % de los mejor remunerados.

Las medidas de posición son ideales para obtener información adicional a partir de datos resumidos, es decir, que presentan perdida de información por agrupamiento en intervalos de clase.

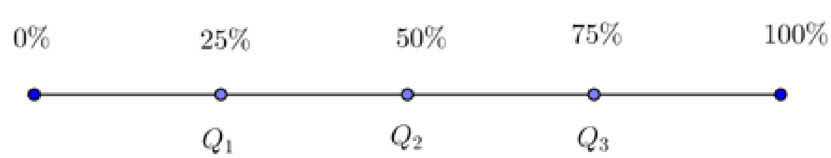
Podemos definir entonces como medidas de posición a los indicadores estadísticos que muestran la frecuencia acumulada hasta un valor k cualquiera.

6.1. Cuantiles

Constituyen una generalización del concepto de mediana. Así como la mediana divide a la serie estudiada en dos partes con el mismo número de elementos cada una, si la división se hace en cuatro partes, o en diez partes, o en cien partes, llegamos al concepto de cuantil.
Hay, principalmente, tres cuantiles importantes: cuartiles, deciles y percentiles.

6.2. Cuartiles

Los cuartiles son los tres valores de la variable que dividen a un conjunto de datos ordenados en cuatro partes iguales. Q_1, Q_2 y Q_3 determinan los valores correspondientes al 25 %, al 50 % y al 75 % de los datos. Q_2 coincide con la mediana.



Lo cuartiles son representados como Q_i con $i = 1, 2, 3, 4$.

- El primer cuartil (Q_1) considera el 25 % de la información a su izquierda y el 75 % a la derecha.
- El segundo cuartil (Q_2) considera el 50 % de la información tanto a la derecha como a la izquierda, este coincide con la mediana.
- El tercer cuartil (Q_3) considera el 75 % de la información a la izquierda y el 25 % a la derecha.
- El cuarto cuartil (Q_4), que no fue indicado en el gráfico considera a toda la información.

Por lo anterior podemos afirmar que los cuartiles son tres valores que dividen a la serie de datos en cuatro partes iguales, como se puede apreciar en la figura.

6.3. Percentiles

Valor que divide un conjunto ordenado de datos numéricos, de forma que un porcentaje de tales datos sea inferior a dicho valor. Así, un individuo en el percentil 80 está por encima del 80 % del grupo al que pertenece.

La mediana se ubica en la posición central de un conjunto de datos ordenados, puede interpretarse como el segundo cuartil y el percentil 50. De la misma forma, también se utilizan como medida de posición los quintiles, que separan los datos en cinco grupos iguales, los deciles, que los separan en diez grupos iguales, etcétera.

6.4. Datos agrupados

Para calcular los percentiles en datos agrupados, utilizamos:

$$P_k = L_k + \frac{k \cdot (\frac{n}{100}) - N_k}{n_k} \cdot C \tag{6.1}$$

Donde:

L_k menor valor de la clase donde se encuentra el k-ésimo percentil.

n número de observaciones

N_k frecuencia acumulada de la clase que precede a la clase del k-ésimo percentil.

n_k frecuencia absoluta de la clase del k-ésimo percentil.

C amplitud de la clase del k-ésimo percentil.

Ejemplo 6.1 *Dada la siguiente distribución de frecuencias de 212 puntajes obtenidos en la P.S.U., determinar el percentil 45.*

Clase	f_i	F_i	h_i	H_i
[400, 450[	10	10	4.716	4.716
[450, 500[	9	19	4.245	8.962
[500, 550[	20	39	9.433	0.183
[550, 600[	31	70	0.146	0.330
[600, 650[	80	150	0.377	0.707
[650, 700[	42	192	0.198	0.905
[700, 750[	10	202	4.716	0.952
[750, 800[	10	212	4.716	1
Total	212			

Cuadro 10: 212 puntajes PSU

Calculamos en 45% de 212

$$212 \cdot 0,45 = 95,4$$

Verificamos en la frecuencia acumulada que clase supera este valor, en este caso corresponde a la clase $[600, 650[$.
Completamos en la fórmula:

$$P_{45} = 600 + \frac{45 \cdot \left(\frac{212}{100}\right) - 70}{80} \cdot 50 = 615,9$$

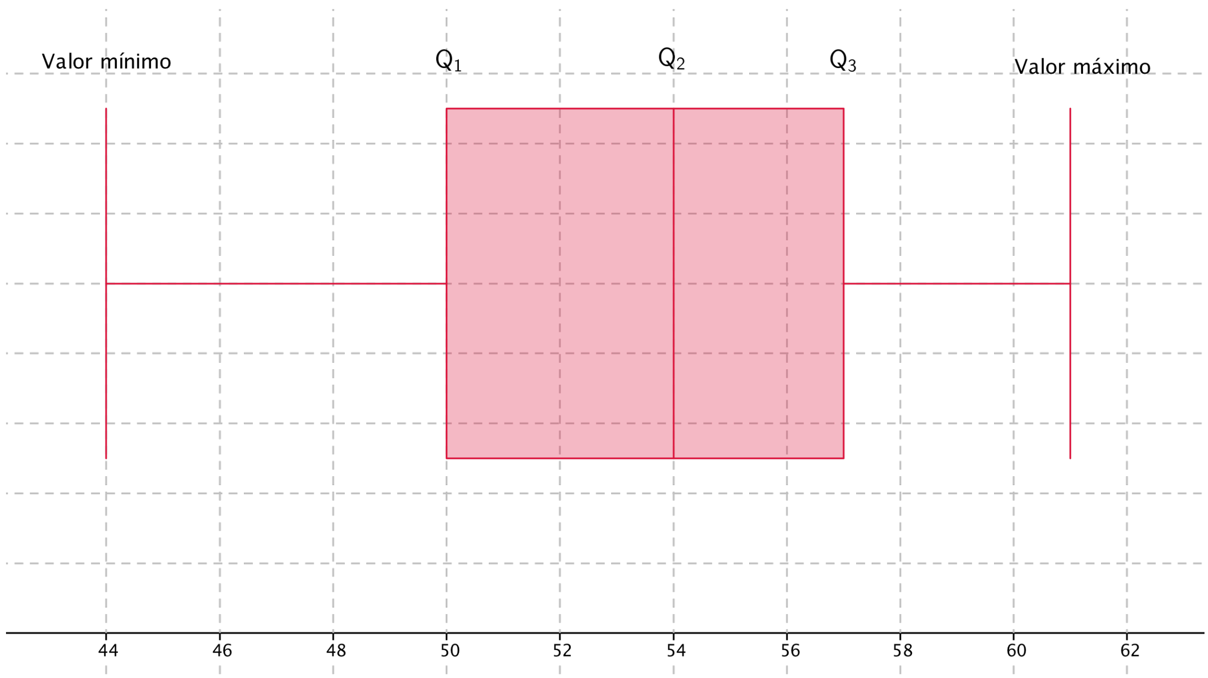
El resultado nos indica que el 45% de los estudiantes obtuvo puntajes menores o iguales a 615,9.

6.5. Diagrama de caja

Es una representación gráfica de los datos que permite analizar conjuntamente una serie de medidas numéricas, tales como el mínimo, el máximo, la mediana y los cuartiles.

Ejemplo 6.2 *Los siguientes datos corresponden a la masa (en kg.) de 24 mujeres de 17 años. Se obtuvieron los siguientes datos:*

$$n = 24$$
$$M_e = 54$$
$$Q_1 = 50$$
$$Q_2 = M_e = 54$$
$$Q_3 = 57$$
$$\textit{Mínimo}{=}44$$
$$\textit{Máximo}{=}61$$
$$\textit{Rango}{=}17$$



En el gráfico, los extremos del rectángulo indican los cuartiles Q_1 y Q_3, mientras que la línea que divide a este verticalmente indica la mediana (Q_2).
Las líneas que sobresalen del rectángulo indican el valor mínimo y máximo de la distribución.

Se observa que:
$Q_1 = 50$ *indica que el 25% pesó menos o igual que 50 kilogramos;*
Q_2*, que el 50% pesó menos o igual que 54 kilogramos, y*
Q_3*, que el 75% pesó menos de 57 kilogramos o igual.*

Resumiendo: Los percentiles se conocen como medidas de posición y hacen referencia al lugar que ocupa un dato cuando todos estos están ordenados de menor a mayor.
Puede ocurrir que un mismo valor represente más de un percentil, esto ocurre generalmente para un número pequeño de observaciones.
Los cuartiles son un caso particular de los percentiles y reciben este nombre porque dividen un conjunto de datos en cuatro partes iguales. Otros casos particulares son los quintiles y los deciles

6.6. Diagrama de tallo y hojas

Diagrama que permite resumir u ordenar un conjunto de datos, de modo de conocer intuitivamente la forma de su distribución. Se utiliza para estudiar la dispersión de los valores de una muestra.
Es un método usado para organizar datos. El valor mayor común de los datos se utiliza como tallo y el siguiente valor mayor de posición común se usa para formar las hojas.

Ejemplo 6.3 *Los siguientes datos corresponden a las edades de 20 personas de un curso para adultos en un liceo nocturno:*

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

Comenzamos seleccionando los tallos que en nuestro caso son las cifras de las decenas, es decir 3, 2, 4, que reordenadas son 2, 3 y 4.
A continuación contamos y vamos «añadiendo» cada hoja a su tallo de forma ordenada.

2	0 3 4 4 4 5 9
3	1 1 3 4 6 6 7 9 9
4	0 0 1 5

Se observa la mayor cantidad de personas se encuentra entre los 30 y 40 años, además la edad menor es 20 y la mayor es 45 años.
La moda es 24 años El tallo representa la cifra de las decenas, y las hojas, las unidades.

Ejemplo 6.4 *Podemos comparar, dos distribuciones. Supongamos una segunda distribución:*

$$35, 38, 32, 28, 30, 29, 27, 19, 48, 40$$

$$39, 24, 24, 34, 26, 41, 29, 48, 28, 22$$

De ella podemos elaborar sus diagrama de Tallos y Hojas y compararla con la anterior.

(N = 20) Hojas	Tallos	Hojas (N = 20)
9	1	
9 9 8 8 7 6 4 4 2	2	0 3 4 4 4 5 9
9 8 5 4 2 0	3	1 1 3 4 6 6 7 9 9
8 8 1 0	4	0 0 1 5

6.7. Ejercicios

1. Se tiene el siguiente conjunto de 26 datos:
   10, 13, 4, 7, 8, 11 10, 16, 18, 12, 3, 6, 9, 9, 4, 13, 20, 7, 5, 10, 17, 10, 16, 14, 8, 18
   Obtener la mediana ,los cuartiles y la media.

2. Un pediatra obtuvo la siguiente tabla sobre los meses de edad de 50 niños de su consulta en el momento de caminar por primera vez:

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

   a) Dibujar el polígono de frecuencias.
   b) Calcular la moda, la mediana, la media y los cuartiles.
   c) Representa los datos en un diagrama de caja.

3. Un siquiatra de la ciudad ha tomado una muestra aleatoria de 20 niños con desórdenes de conducta, anotando el tiempo necesario (en horas) que requirió para lograr un plan integral de tratamiento con cada uno de ellos:
   6, 7, 7, 8, 8, 8, 8, 9, 9, 9, 9, 9, 9, 9, 10, 10, 10, 10, 10, 11.

   a) Calcular las medidas de tendencia central
   b) Calcular los cuartiles
   c) construir un diagrama de caja para estos datos.

4. El número de estudiantes en cada curso de un colegio hasta 8 Básico son: 40 31 23 12 14 23 31 41 24 31 42 25 32 33 33 34
   Construir el diagrama de tallo y hoja y determinar la moda.

5. Un estudio en una muestra de clientes de una empresa del número de llamadas recibidas reveló la siguiente información.
   52, 43, 30, 38, 30, 42, 12, 46, 39, 37, 34, 46, 32, 18, 41, 5

   a) Desarrolle un diagrama de tallo y hojas.
   b) ¿Cuántas llamadas recibe un cliente típico?
   c) ¿Cuáles fueron el mayor y el menor número de llamadas recibidas?

6. Aloha Banking Co. está estudiando el número de veces por día que es utilizado su cajero automático localizado en Loblaws Supermarket. A continuación se muestra el número de veces que fue usado durante los últimos 30 días.
   83, 63, 95, 64, 80, 36, 84, 84, 78, 76, 73, 61, 84, 68, 59, 54, 52, 84, 75, 65, 95, 59, 90, 47, 70, 52, 87, 61, 77, 60

   a) Construye un diagrama de tallo y hojas.
   b) Resume los datos que indican el número de veces que el cajero automático fue usado.
   c) ¿Cuáles son el menor y el mayor número de veces que fue usado?
   d) ¿Alrededor de qué números se agrupa la cantidad de veces que fue usado el cajero?

7. Construcción utilizando Software

Se puede construir una tabla de distribución de frecuencias y su análisis utilizando software computacional. Vamos a construir tablas de distribución de frecuencias utilizando una planilla de calculo y geogebra.

7.1. Planilla de calculo

En el siguiente ejemplo utilizaremos Excel 2010 para construir la tabla de distribución de frecuencias y después encontremos las medidas de tendencia central.

Ejemplo 7.1 *Los siguientes datos corresponden al sueldo (en miles de pesos) de 40 trabajadores de una empresa:*

119	135	138	144	146	150	156	164
125	135	140	144	147	150	157	165
126	135	140	145	147	152	158	168
128	136	142	145	148	153	161	173
132	138	142	146	149	154	163	176

7.1.1. Tabla de distribución de frecuencias

1. Copiamos los datos en la planilla de cálculo:

	A	B	C	D	E	F	G	H	I
1	119	135	138	144	146	150	156	164	
2	125	135	140	144	147	150	157	165	
3	126	135	140	145	147	152	158	168	
4	128	136	142	145	148	153	161	173	
5	132	138	142	146	149	154	163	176	
6									

2. calcular los datos necesarios para construir las clases:

	A	B	
1	119	135	138
2	125	135	140
3	126	135	140
4	128	136	142
5	132	138	142
6			
7	Máximo	=MAX(A1:H5)	
8	Mínimo	=MIN(A1:H5)	
9	Rango	=B7-B8	
10	Intervalos(m)	=ENTERO(1+3,3*LOG(40))	
11	Amplitud (a)	=ENTERO(B9/B10)	
12			
13			

Resulta:

	A	B
1	119	135
2	125	135
3	126	135
4	128	136
5	132	138
6		
7	Máximo	176
8	Mínimo	119
9	Rango	57
10	Intervalos(m)	6
11	Amplitud (a)	9
12		

3. construimos los intervalos o clases Comenzamos con el mínimo (119) y sumamos la amplitud (9) para el primer intervalo. Luego continuamos con el resto:

	A	B	C	D
4	128	136	142	145
5	132	138	142	146
6				
7	Máximo	176		Sueldos
8	Mínimo	119		119 - 128
9	Rango	57		129 - 138
10	Intervalos(m)	6		139 - 148
11	Amplitud (a)	9		149 - 158
12				159 -168
13				169 - 178
14				

creamos una columna donde indicamos el límite superior (L_{sup})

	A	B	C	D	E
4	128	136	142	145	148
5	132	138	142	146	149
6					
7	Máximo	176		Sueldos	Ls
8	Mínimo	119		119 - 128	128
9	Rango	57		129 - 138	138
10	Intervalos(m)	6		139 - 148	148
11	Amplitud (a)	9		149 - 158	158
12				159 -168	168
13				169 - 178	178
14					

creamos una columna donde vamos a registrar la frecuencia absoluta. Es importante **selecionar el rango** y después escribir la función, en este caso:

=FRECUENCIA(A1:H5;E8:E13)

una vez escrita presionar **CTRL+SHIFT+ENTER** y Excel encuentra las frecuencias absolutas automáticamente, como lo muestra la siguiente figura:

F8 *fx* {=FRECUENCIA(A1:H5;E8:E13)}

	A	B	C	D	E	F
4	128	136	142	145	148	153
5	132	138	142	146	149	154
6						
7	Máximo	176		Sueldos	Ls	fi
8	Mínimo	119		119 - 128	128	4
9	Rango	57		129 - 138	138	7
10	Intervalos(m)	6		139 - 148	148	13
11	Amplitud (a)	9		149 - 158	158	9
12				159 -168	168	5
13				169 - 178	178	2
14						
15						

Los demás calculos de la tabla se calculan aplicando las siguientes fórmulas:

Sueldos	Ls	fi	Fi	hi	Hi
119 - 128	128	=FRECUENCIA(A1:H5;E8:E13)	=F8	=F8/$F$14	=H8
129 - 138	138	=FRECUENCIA(A1:H5;E8:E13)	=G8+F9	=F9/$F$14	=I8+H9
139 - 148	148	=FRECUENCIA(A1:H5;E8:E13)	=G9+F10	=F10/$F$14	=I9+H10
149 - 158	158	=FRECUENCIA(A1:H5;E8:E13)	=G10+F11	=F11/$F$14	=I10+H11
159 - 168	168	=FRECUENCIA(A1:H5;E8:E13)	=G11+F12	=F12/$F$14	=I11+H12
169 - 178	178	=FRECUENCIA(A1:H5;E8:E13)	=G12+F13	=F13/$F$14	=I12+H13
		=SUMA(F8:F13)		=SUMA(H8:H13)	

para obtener:

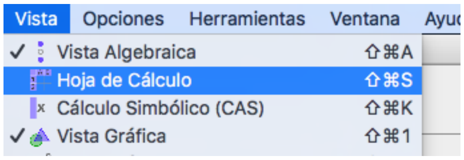
D7 | fx Sueldos

	A	B	C	D	E	F	G	H	I
1	119	135	138	144	146	150	156	164	
2	125	135	140	144	147	150	157	165	
3	126	135	140	145	147	152	158	168	
4	128	136	142	145	148	153	161	173	
5	132	138	142	146	149	154	163	176	
6									
7	Máximo	176		Sueldos	Ls	fi	Fi	hi	Hi
8	Mínimo	119		119 - 128	128	4	4	0,10	0,10
9	Rango	57		129 - 138	138	7	11	0,18	0,28
10	Intervalos(m)	6		139 - 148	148	13	24	0,33	0,60
11	Amplitud (a)	9		149 - 158	158	9	33	0,23	0,83
12				159 - 168	168	5	38	0,13	0,95
13				169 - 178	178	2	40	0,05	1,00
14						40		1,00	
15									

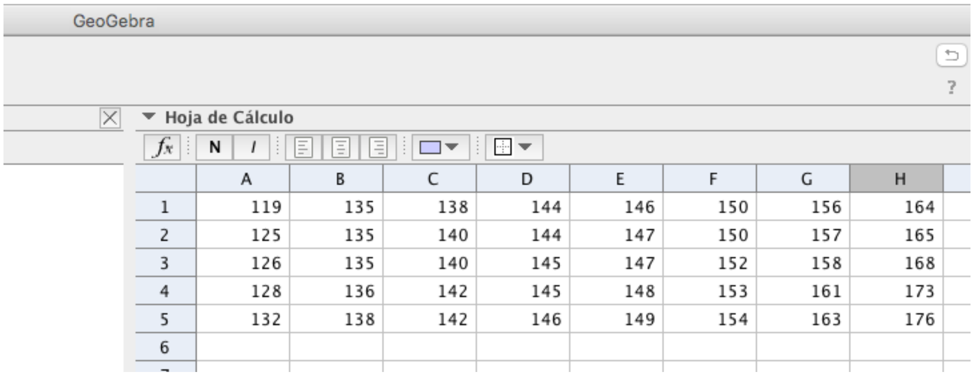
7.2. Geogebra

Utilizamos el mismo ejemplo anterior para esta demostración.

1. Abrimos geogebra y en 'Vista' seleccionamos 'Hoja de cálculo'

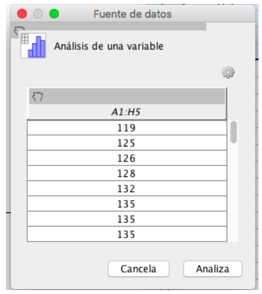


2. Ingresamos los datos del ejemplo anterior:

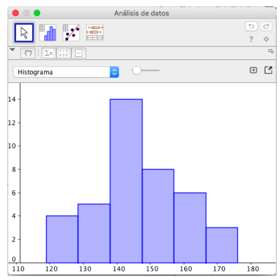


3. Seleccionamos el rango de datos y hacemos click en el botón 'Análisis de una variable'

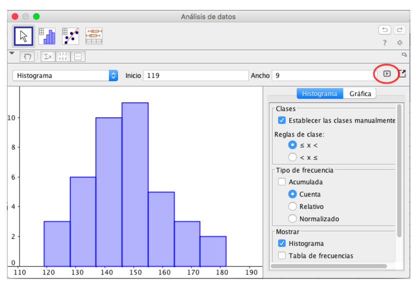
4. Aparece una ventana emergente y hacemos click en el botón 'Analiza'



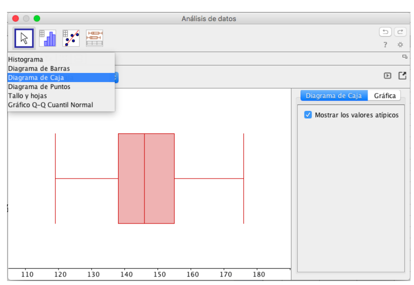
5. Geogebra nos muestra en la ventana 'Análisis de datos' un Histograma por defecto de los datos que seleccionamos.



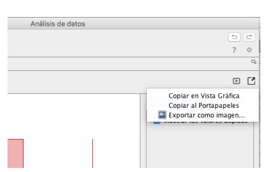
6. Hacemos click en opciones y fijamos inicio en 119 y ancho en 9.



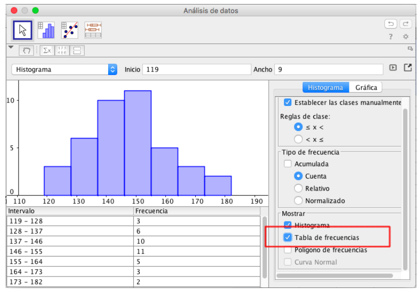
7. Para obtener un gráfico de cajas seleccionamos esta opción



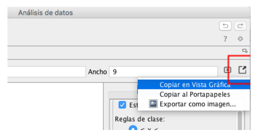
8. Para poder imprimir o exportar a otro documento debemos copiar el gráfico a 'vista gráfica'



9. Volviendo al Histograma, podemos mostrar la tabla de frecuencias



10. Exportamos a 'Vista gráfica'



11. Quedando lista para impresión o exportar a otros formatos.