Tema 1

1 Conceptos

- **Fenómenos determinísticos.** Aquellos que dan lugar al mismo resultado siempre que se realicen bajo idénticas condiciones.
- **Fenómenos aleatorios.** Se caracterizan porque sus resultados puede variar, incluso si el experimento se realiza bajo idénticas condiciones iniciales.
- **Población, colectivo o universo.** Conjunto de unidades o elementos con alguna/s característica/s en común, sobre el que se desea obtener cierta información.
- **Muestra.** Subconjunto de la población elegido en términos de representatividad.
- **Carácter o característica estadística.** Toda propiedad que se desee estudiar en la población, que debe poder observarse sobre todos y cada uno de los individuos que la componen.
  - **Modalidad.** Formas posibles en las que se puede manifestar el carácter, y cada individuo o unidad de la población debe presentar una y solo una de ellas.
  - **Carácter cualitativo (atributo).** Aquel cuyas modalidades no son medibles o cuantificables por números.
  - **Carácter cuantitativo.** Aquel cuyas modalidades son numéricamente medibles o cuantificables.
- **Escalas de medida.** La realización de cualquier estudio estadístico requiere, como primer paso, la identificación precisa de las modalidades del carácter bajo estudio y la asignación de símbolos o números a las distintas modalidades; esto es lo que se denomina *medición del carácter*.

  Si denotamos por X al carácter, y A y B son dos individuos cuyas medidas de X son x_A y x_B, se distinguen cuatro tipos de escala:
  - **Escala nominal.** Solo se puede decir que $x_A = x_B$ o que $x_A \neq x_B$.
  - **Escala ordinal.** No solo se puede decir que $x_A = x_B$ o $x_A \neq x_B$, sino que $x_A < x_B$ o $x_A > x_B$.
  - **Escala de intervalo.** Se puede decir que $x_A = x_B$, $x_A \neq x_B$, $x_A < x_B$, $x_A > x_B$ y que A es $x_A - x_B$ unidades diferente (superior o inferior) que B.
  - **Escala de razón.** Se puede decir que A es x_A/x_B veces superior a B.
- **Variable.** Una variable es un símbolo que representa a distintos valores numéricos. Cuando estos valores son el resultado de mediciones y observaciones estadística, hablaremos de variable estadística.

  Así, un carácter cuantitativo irá representado por una variable estadística, y sus diversas modalidades serán los valores que toma dicha variable.
  - **Variables discretas.** El paso de un valor de la variable al siguiente representa un salto (el conjunto de números reales que soporta la variable está formado por puntos aislados).
  - **Variables continuas.** La variable puede tomar todos los valores comprendidos en un intervalo de la recta real.

2 Definiciones y fórmulas

- Frecuencia absoluta: número total de individuos en la población que presenta dicho valor (modalidad), n_i.

- Frecuencia relativa: p roporción del número de individuos que presenta dicha modalidad, $f_i = \dfrac{n_i}{n}$, $i = 1, \ldots, k$
$$\sum_{i=1}^{k} n_i = n; \quad \sum_{i=1}^{k} f_i = 1$$

- Frecuencia absoluta acumulada: número de individuos que presentan un valor (modalidad) menor o igual que x_i
$$N_i = \sum_{j=1}^{i} n_j, i = 1, \ldots, k$$

- Frecuencia relativa acumulada: proporción de individuos que presentan un valor (modalidad) menor o igual que x_i.
$$F_i = \frac{N_i}{n} = \sum_{j=1}^{i} f_j, i = 1, \ldots, k$$

- Distribución de frecuencias: conjunto formado por cada uno de los valores (modalidades) junto con sus frecuencias
$$\{(x_i, n_i) : i = 1, \ldots, k\}$$

3 Tablas estadísticas

3.1 Variables discretas y atributos

Modalidades	Frec. Abs	Frec. Rel.	Frec. Abs. Acum.	Frec. Rel. Acum.
x_1	n_1	f_1	$N_1 = n_1$	$F_1 = f_1$
x_2	n_2	f_2	$N_2 = n_1 + n_2$	$F_2 = f_1 + f_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	n_i	f_i	$N_i = n_1 + n_2 + \cdots + n_i$	$F_i = f_1 + f_2 + \cdots + f_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_k	f_k	$N_k = n_1 + n_2 + \cdots + n_k = n$	$F_k = f_1 + f_2 + \cdots + f_k = 1$

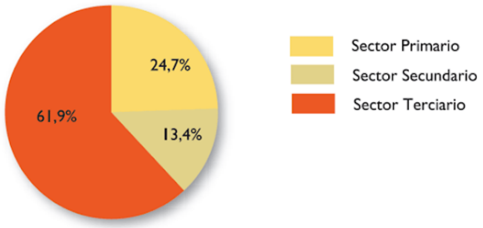
3.2 Variables continuas

Intervalos	Marcas	Amplitud	Frec. Abs.	Frec. Rel.	Frec. Acum.
$(e_0, e_1]$	$c_1 = \frac{e_0+e_1}{2}$	$a_1 = e_1 - e_0$	n_1	f_1	$N_1 = n_1$
$(e_1, e_2]$	$c_2 = \frac{e_1+e_2}{2}$	$a_2 = e_2 - e_1$	n_2	f_2	$N_2 = n_1 + n_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$(e_{i-1}, e_i]$	c_i	a_i	n_i	f_i	$N_i = n_1 + n_2 + \cdots + n_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$(e_{k-1}, e_k]$	c_l	a_k	n_k	f_k	$N_k = n_1 + n_2 + \cdots + n_k = n$

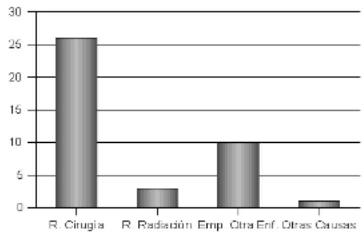
4 Representaciones gráficas

4.1 Atributos

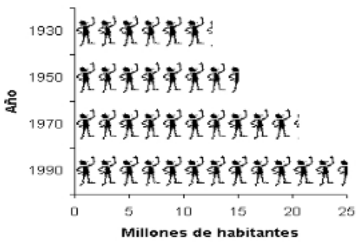
- **Diagrama de sectores.** Es un círculo dividido en tantos sectores circulares como modalidades tenga el carácter, siendo el área de cada un proporcional a la frecuencia absoluta o relativa de la modalidad.

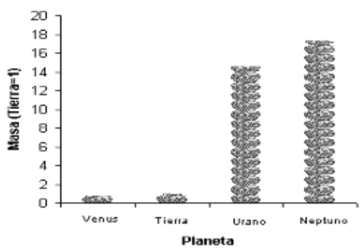


- **Diagrama de rectángulos o barras.** Consiste en varios rectángulos (uno por modalidad) de base constante y alturas proporcionales a las frecuencias (absolutas o relativas) de cada modalidad.



- **Pictograma.** Se dibujan figuras, normalmente alusivas al carácter que se está estudiando, bien una para cada modalidad con tamaño proporcional a su frecuencia, o bien repitiendo la figura tantas veces como requieran las frecuencias.

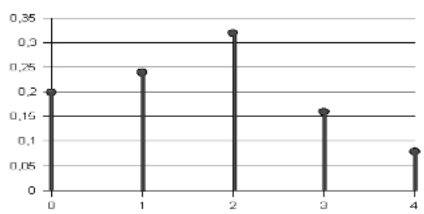




4.2 Variables discretas

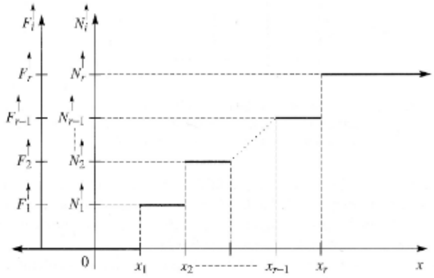
- **Diagrama de barras.** Sistema de ejes cartesianos en el que se representa en el eje de abcisas los valores de la variable, y se trazan barras verticales con longitudes proporcionales a sus frecuencias.

x_i	n_i	$f_i = n_i/100$
0	20	$0,2$
1	24	$0,24$
2	32	$0,32$
3	16	$0,16$
4	8	$0,08$



- **Curva de distribución**. Representación de la función de distribución, que es una función definida para cada número real x como la proporción de datos menores o iguales que x. Así, si $x_1 < x_2 < \cdots < x_k$ son los valores de la variable ordenados,

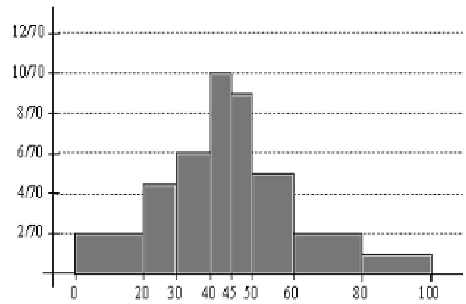
$$F(x) = \begin{cases} 0 \ \forall x < x_1 \\[2ex] \dfrac{\sum_{j=1}^{i} n_j}{n} = \sum_{j=1}^{i} f_j = \dfrac{N_i}{n} = F_i \ \forall x : x_i \leq x < x_{i+1} \\[2ex] 1 \ \forall x \geq x_k \end{cases}$$



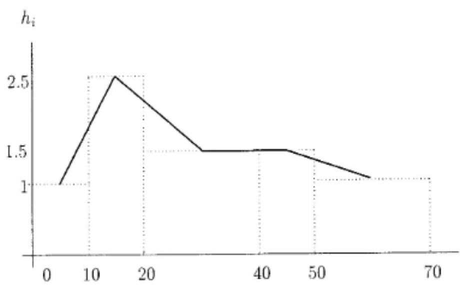
4.3 Variables continuas

- **Histograma.** Está formado por rectángulos yuxtapuestos, cuyas bases son las diferentes clases o intervalos de definición de la variable, y cuyas alturas son las frecuencias medias $h_i = f_i/a_i$ por unidad de amplitud (densidades de frecuencia).

I_i	n_i	f_i	h_i	Altura$_i = 5h_i$
$(0, 20]$	8	8/70	4/700	2/70
$(20, 30]$	9	9/70	9/700	4.5/70
$(30, 40]$	12	12/70	12/700	6/70
$(40, 45]$	10	10/70	10/700	10/70
$(45, 50]$	9	9/70	1.8/700	9/70
$(50, 60]$	10	10/70	10/700	5/70
$(60, 80]$	8	8/70	4/700	2/70
$(80, 100]$	4	4/70	2/700	1/70



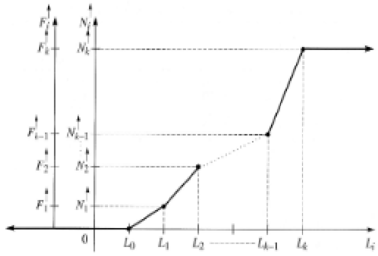
- **Poligonal de frecuencias.** Es la poligonal que resulta de unir los puntos correspondientes a los techos de las marcas de clase de los intervalos en el histograma.



- **Curva de distribución**.

$$F(e_i) = \sum_{j=1}^{i} f_j$$

$$F(-\infty) = 0, \ F(+\infty) = 1$$



5 Características unidimensionales

Propiedades deseable (propiedades de Yule)

1. Deben definirse de manera objetiva, de forma que dos personas diferentes deben dar iguales resultados.
2. Deben usar todas las observaciones y no algunas solamente.
3. Deben tener un significado concreto, para que sean rápidas y fácilmente interpretables.
4. Deben ser sencillas de calcular.
5. Deben prestarse fácilmente al cálculo algebraico.
6. Deben ser poco sensibles a fluctuaciones muestrales, de forma que si cambiasen valores extremos de los datos, no cambiasen en gran medida las características del conjunto.

5.1 Medidas de posición

- **Media aritmética**. Es la suma de todos los valores de la variable dividida por el número total de observaciones.
  - Si consideramos una variable estadística discreta en una población de tamaño n con distribución ode frecuencias $\{(x_i, n_i(f_i)); i = 1, \ldots, k\}$, la media aritmética es

    $$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} f_i x_i$$

  - Si la variable es continua y los datos están agrupados en intervalos de clase:

    $$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i c_i = \sum_{i=1}^{k} f_i c_i$$

  - La media aritmética está acotada por los valores extremos de la variables: $x_1 \leq \overline{x} \leq x_k$.
  - La media aritmética de las desviaciones de los datos respecto de la media aritmética es igual a cero:

    $$\sum_{i=1}^{k} f_i(x_i - \overline{x}) = 0$$

  - Si se somete una variable X a una transformación lineal afín, la media aritmética de la nueva variable es la imagen de la media de X por la misma transformación:

    $$Y = aX + b \Longrightarrow \overline{y} = a\overline{x} + b$$

  - La media aritmética de los cuadrados de las desviaciones respecto a la media aritmética es mínima:

    $$\sum_{i=1}^{k} f_i(x_i - \overline{x})^2 < \sum_{i=1}^{k} f_i(x_i - a)^2, \qquad \forall a \neq \overline{x}$$

- **Media geométrica**. Es la raíz n-ésima del producto de los n valores (o marcas de clase) de la distribución:

  $$G = \sqrt[n]{\prod_{i=1}^{k} x_i^{n_i}}$$

- **Media armónica**. Se define como la inversa de la media aritmética de los valores inversos de la variable:

  $$H = \frac{n}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \cdots + \frac{n_k}{x_k}}$$

- **Media cuadrática**. Raíz cuadrada de la media aritmética de los cuadrados de los valores de la variables:

  $$Q = \sqrt{\sum_{i=1}^{k} f_i x_i^2}$$

- **Mediana**.