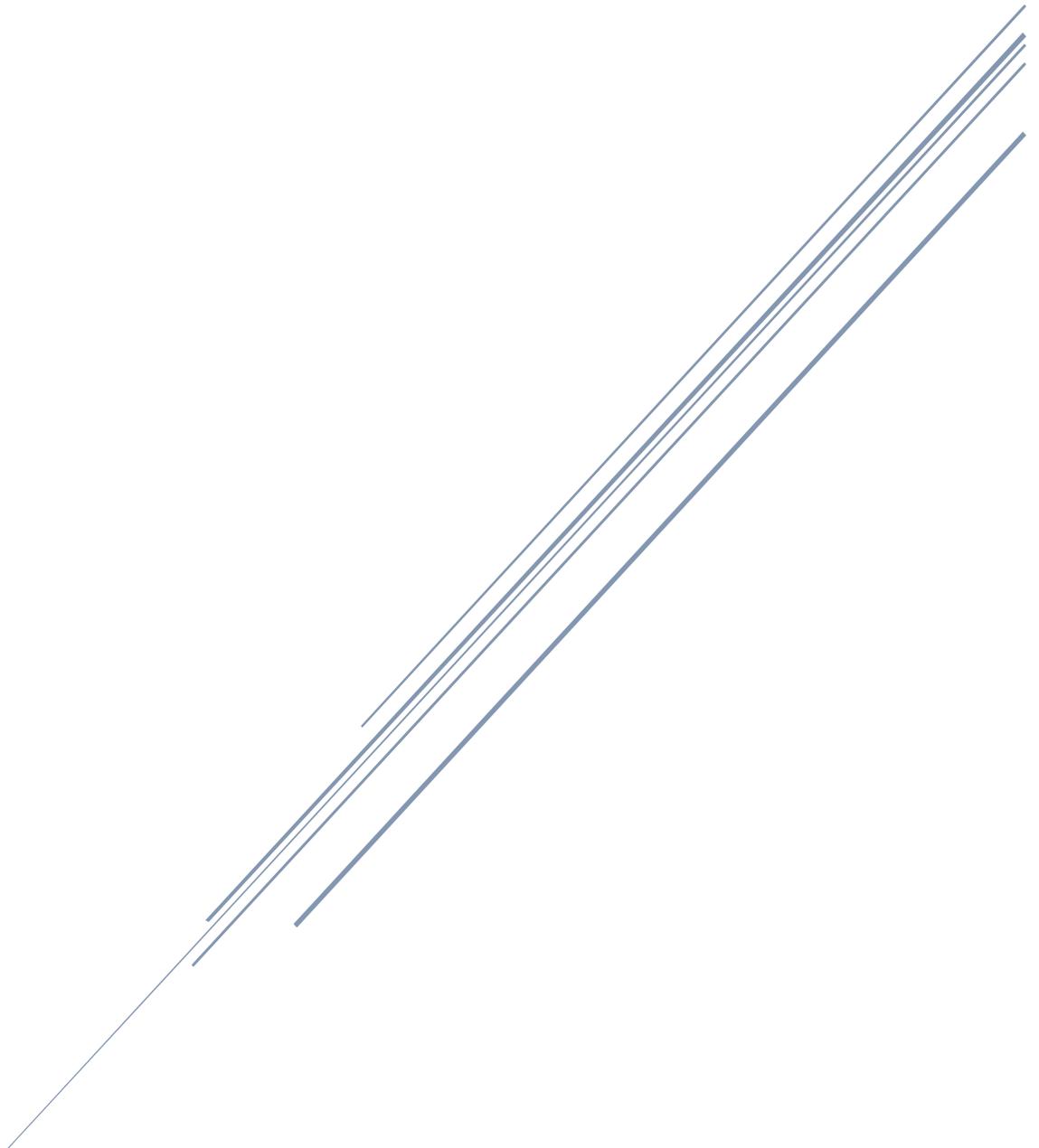


TEMA 3. TABULACIÓN Y REPRESENTACIÓN GRÁFICA DE DATOS UNIVARIANTES

Curso OCW de “Estadística Descriptiva con Excel para Grados de Ciencias Sociales”



1. TABULACIÓN

Suele ocurrir que, una vez obtenidos los datos objeto de estudio, éstos son lo suficientemente numerosos como para que una simple inspección visual de ellos no nos aporte prácticamente ninguna información. Es por ello que se necesita resumir todos esos datos y agruparlos si es preciso.

Denominamos *tabulación* al proceso de ordenación y agrupamiento de un conjunto de datos. Como resultado del mismo se obtienen las *distribuciones de frecuencias* de las observaciones.

En lo que sigue, emplearemos la siguiente notación:

- Tamaño de la población o muestra: N
- Número de modalidades: k
- Valores de la variable: $\{x_1, x_2, \dots, x_k\}$
- Supuesto (cuando tenga sentido): $x_1 < x_2 < \dots < x_k$

1.1 Frecuencias

Sea X una variable o característica medida en una población de tamaño N que toma los valores $\{x_i; i = 1 \dots k\}$.

La **Frecuencia Absoluta** se define como el número de veces que se observa o repite el valor x_i , y la denotamos por n_i . Se cumple que:

$$N = \sum_{i=1}^k n_i$$

La **Frecuencia Relativa** se define como el porcentaje de veces que se observa el valor x_i , y la denotamos por f_i : $f_i = n_i/N$. Se cumple que:

$$\sum_{i=1}^k f_i = 1$$

Las frecuencias absolutas y relativas se pueden obtener siempre, independientemente del tipo de variable que estemos manejando. Para aquellos datos que presentan alguna ordenación, es decir, todos excepto los de tipo nominal, podemos obtener las **frecuencias acumuladas**.

La **Frecuencia Absoluta Acumulada** se define como el número total de observaciones menores o iguales que x_i , y la denotamos por N_i :

$$N_i = \sum_{j=1}^i n_j$$

Se cumple que $N_k = N$

La **Frecuencia Relativa Acumulada** se define como el porcentaje de observaciones que son menores o iguales que x_i , y la denotamos por F_i :

$$F_i = \sum_{j=1}^i f_j \quad \text{o} \quad F_i = \frac{N_i}{N}$$

Se cumple que $F_k = 1$.

1.2. Distribución de frecuencias

Una variable X queda perfectamente definida por su denominada **distribución de frecuencias**, que es el conjunto formado por todos los valores de la variable X y sus correspondientes frecuencias. Para ello emplearemos una representación tabular.

La obtención de dicha distribución dependerá del tipo de datos con los que trabajamos:

- a) **Variables cualitativas nominales.** Sólo tiene sentido calcular las frecuencias absolutas y relativas ordinarias (no acumuladas):

X	n_i	f_i
x_1	n_1	$f_1 = n_1/N$
x_2	n_2	$f_2 = n_2/N$
\vdots	\vdots	\vdots
x_k	n_k	$f_k = n_k/N$
Σ	N	1

b) **Variables cualitativas ordinales y variables cuantitativas discretas** (no agrupadas). En este caso también se pueden calcular las frecuencias acumuladas:

X	n_i	f_i	N_i	F_i
x_1	n_1	$f_1 = \frac{n_1}{N}$	$N_1 = n_1$	$F_1 = f_1$ ó $F_1 = \frac{N_1}{N}$
x_2	n_2	$f_2 = \frac{n_2}{N}$	$N_2 = n_1 + n_2$	$F_2 = f_1 + f_2$ ó $F_2 = \frac{N_2}{N}$
\vdots	\vdots	\vdots	\vdots	\vdots
x_k	n_k	$f_k = \frac{n_k}{N}$	$N_k = n_1 + \dots + n_k = N$	$F_k = f_1 + \dots + f_k = 1$ ó $F_k = \frac{N_k}{N} = 1$
Σ	N	1		

c) **Variables cuantitativas continuas y variable cuantitativas discretas agrupadas.** En el caso de que la variable X esté constituida por una gran cantidad de valores, éstos se agrupan en intervalos. En ese caso hablamos de **distribución de frecuencias agrupadas**. Para ello es preciso calcular tanto los intervalos como un valor puntual que represente a cada uno de los intervalos.

Los **intervalos** se representan mediante la expresión $(L_{i-1}, L_i]$ donde L_{i-1} es el extremo inferior del intervalo y L_i es el extremo superior. A continuación, se determina un valor representativo de cada uno de ellos, habitualmente su punto medio, y que se denomina **marca de clase**:

$$x_i = \frac{L_{i-1} + L_i}{2}$$

Una variable con datos agrupados recibirá el mismo tratamiento que una con datos sin agrupar, considerando la marca de clase como el valor puntual de la variable para los individuos de ese grupo.

En el caso de que los intervalos no sean todos de la misma amplitud, se debe proceder a calcular una nueva característica de la población, su **densidad de frecuencia**, que puede ser diferente para cada intervalo, y se define como el cociente entre la frecuencia absoluta de cada intervalo y su correspondiente **amplitud**:

$$d_i = \frac{n_i}{a_i}$$

donde a_i representa la amplitud del intervalo y se calcula como $a_i = L_i - L_{i-1}$.

También es posible calcular la densidad empleando la frecuencia relativa:

$$d_i = \frac{f_i}{a_i}$$

Con todo ello podemos obtener la **distribución de frecuencias agrupadas**:

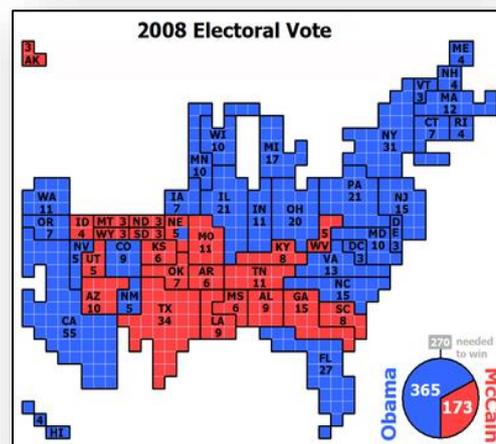
Intervalo	Marca clase	Amplitud	Densidad	Frecuencias			
				n_i	f_i	N_i	F_i
$(L_{i-1}, L_i]$	x_i	a_i	d_i	n_i	f_i	N_i	F_i
$(L_0, L_1]$	x_1	a_1	d_1	n_1	f_1	N_1	F_1
$(L_1, L_2]$	x_2	a_2	d_2	n_2	f_2	N_2	F_2
...
$(L_{k-1}, L_k]$	x_k	a_k	d_k	n_k	f_k	N_k	F_k
<i>Totales</i>				N		1	

2. REPRESENTACIÓN GRÁFICA DE DATOS

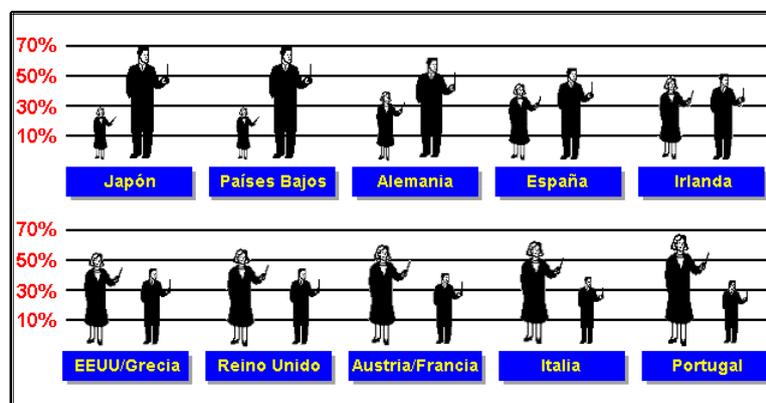
La representación gráfica de los datos dependerá, de nuevo, del tipo de variable con que estamos trabajando.

- Para Variables Cualitativas:
 - Cartogramas, Pictogramas
 - Diagramas de Sectores
 - Diagrama de Barras
- Para Variables Cuantitativas Discretas no agrupadas:
 - Diagrama de Barras
- Para Variables Cuantitativas Continuas y Discretas Agrupadas:
 - Histograma
 - Polígonos de Frecuencias

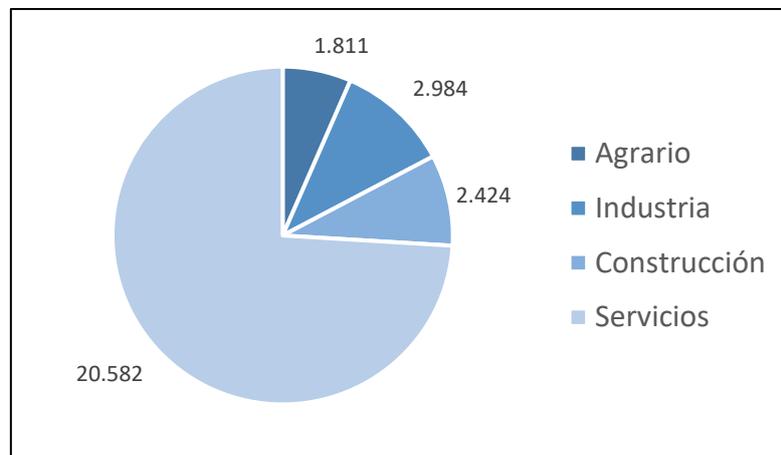
2.1. Cartogramas



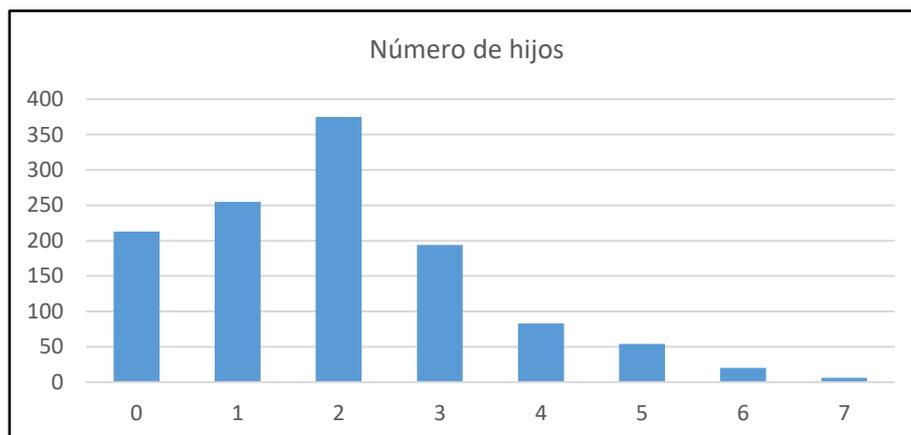
2.2. Pictogramas



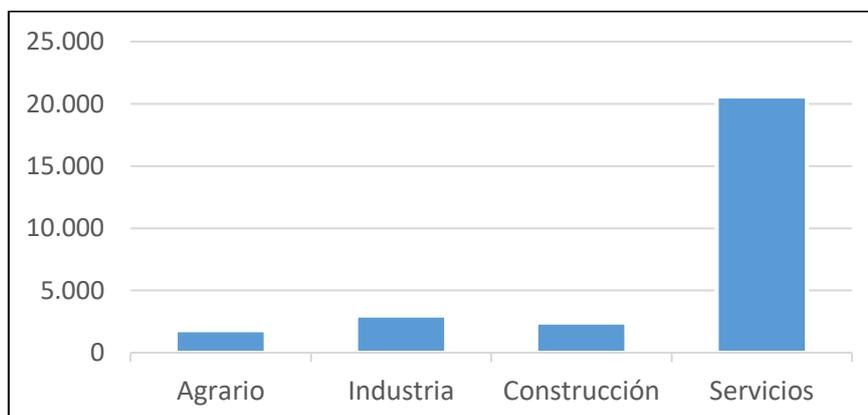
2.3. Diagrama de sectores. Especialmente indicados para variables cualitativas nominales.



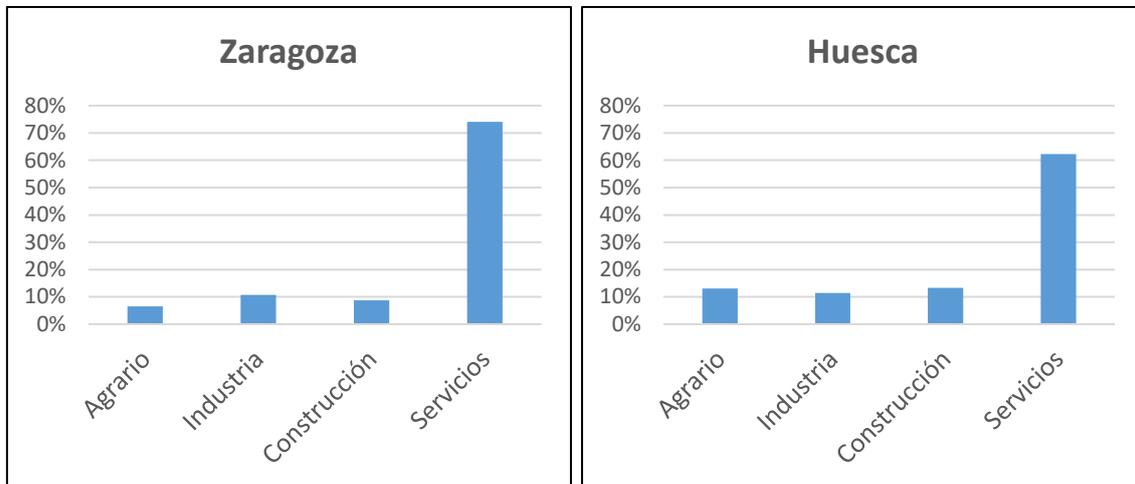
2.4. Diagrama de barras. Sobre un par de ejes se representan en el eje de las abscisas las diferentes modalidades de la variable y se levanta, para cada una de ellas, una barra con una altura proporcional a la frecuencia (absoluta o relativa) de dicha modalidad.



En el caso de variables **nominales**, el eje de abscisas representa cada una de las modalidades, sin importar el orden seguido.



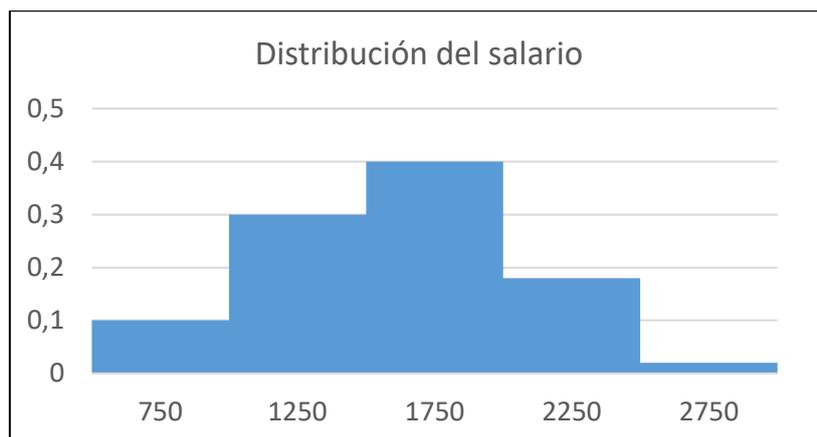
En el eje vertical podemos representar las frecuencias absolutas o relativas, pero si vamos a comparar gráficos debemos utilizar las frecuencias relativas:



2.5. Histogramas

Un **histograma** es una representación gráfica de una variable en forma de barras contiguas, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados. En el eje vertical se representan las densidades de frecuencia (o las frecuencias en caso de ser intervalos de la misma amplitud) y en el eje horizontal los intervalos de las variables, señalando las marcas de clase.

Este tipo de gráfico está especialmente indicado para el caso de variables con distribuciones de frecuencias agrupadas.



3.2.5 Polígonos de frecuencias

El **polígono de frecuencias** es una representación gráfica de la distribución de frecuencias que resulta esencialmente equivalente al histograma y se obtiene uniendo mediante segmentos los centros de las bases superiores de los rectángulos que corresponden a las marcas de clase.

