
Tema 1. Tabulación y representación gráfica de los datos

Resumen del tema

1.1. Introducción a la Estadística

Estadística: ciencia que se ocupa de recoger, clasificar, representar y resumir los datos de muestras, y de hacer inferencias (extraer conclusiones) acerca de las poblaciones de las que éstas proceden.

1. *Estadística descriptiva*: parte de la estadística que se ocupa de recoger, clasificar, representar y resumir los datos de las muestras.
2. *Estadística inferencial*: parte de la estadística que se ocupa de llegar a conclusiones (inferencias) acerca de las poblaciones a partir de los datos de las muestras extraídas de ellas.

CONCEPTOS GENERALES:

- *Población*: conjunto de individuos con propiedades comunes sobre los que se realiza una investigación de tipo estadístico.
- *Muestra*: subconjunto de la población.
- *Tamaño muestral*: número de individuos que forman la muestra.
- *Muestreo*: proceso de obtención de muestras representativas de la población.
- *Variable*: propiedad o cualidad que puede manifestarse bajo dos o más formas distintas en un individuo de una población.
- *Modalidades, categorías o clases*: distintas formas en que se manifiesta una variable.
- Las variables se clasifican en:
 1. *Cuantitativas*: se expresan numéricamente. Se clasifican en:
 - a) *Discretas*: toman valores numéricos aislados, por lo que, fijados dos consecutivos, no pueden tomar ningún valor intermedio.
 - b) *Continuas*: pueden tomar cualquier valor dentro de unos límites, por lo que entre dos valores cualesquiera, por próximos que sean, siempre pueden encontrarse valores intermedios.
 2. *Cualitativas*: no se expresan numéricamente. Se clasifican en:
 - a) *Ordinales*: admiten una ordenación de menor a mayor aunque sus resultados no son numéricos.
 - b) *Nominales*: no admiten una ordenación de menor a mayor.

1.2. Tabulación de los datos

Los datos se agrupan en clases si son cualitativos o discretos, o en intervalos de clase (de igual longitud, generalmente) si son continuos (o discretos con muchos valores distintos).

- Número adecuado de intervalos: $k = 1 + 3'322 \log n$, siendo n el número total de datos. Si los datos no están agrupados en intervalos, también denotaremos por k al número de datos (o de categorías) diferentes.
- *Amplitud* del intervalo de clase $(\ell_i, \ell_{i+1}]$: $d_i = \ell_{i+1} - \ell_i$.
- *Marca de clase* del intervalo $(\ell_i, \ell_{i+1}]$: $x_i = \frac{\ell_i + \ell_{i+1}}{2}$.
- *Frecuencia absoluta* de la clase i -ésima: f_i =número de observaciones contenidas dentro de ella.
- *Frecuencia relativa* de la clase i -ésima: $h_i = \frac{f_i}{n}$.
- *Porcentaje* de la clase i -ésima: $\%_i = 100 h_i$.
- *Frecuencia absoluta acumulada* de la clase i -ésima: $F_i = f_1 + f_2 + \dots + f_i$.
- *Frecuencia relativa acumulada* de la clase i -ésima: $H_i = h_1 + h_2 + \dots + h_i = \frac{F_i}{n}$.
- *Distribución de frecuencias*: tabla conteniendo las distintas clases y las frecuencias correspondientes a cada una de ellas.

1.3. Representaciones gráficas

1. Variables cualitativas

- a) *Diagrama de barras*: se sitúan en el eje horizontal las clases y sobre cada una de ellas se levanta un segmento rectilíneo (o un rectángulo) de altura igual a la frecuencia (absoluta o relativa) o al porcentaje de cada clase.
- b) *Gráfico de sectores*: se divide el área de un círculo en sectores circulares de ángulos proporcionales a las frecuencias absolutas de las clases.

2. Variables cuantitativas con datos no agrupados en intervalos

- a) *Diagrama de barras*: se sitúan en el eje horizontal los diferentes resultados de la variable y sobre cada uno de ellos se levanta un segmento rectilíneo de altura igual a la frecuencia (absoluta o relativa) o al porcentaje de cada resultado.
- b) *Polígono de frecuencias*: se sitúan los puntos que resultan de tomar en el eje horizontal los distintos valores de la variable y en el eje vertical sus correspondientes frecuencias (no acumuladas), uniendo después los puntos mediante segmentos rectilíneos.
- c) *Gráfico de frecuencias acumuladas*: es la representación gráfica de las frecuencias acumuladas, para todo valor numérico. Siempre es una gráfica en forma de *escalera*.

3. Variables cuantitativas con datos agrupados en intervalos

- a) *Histograma*: se sitúan en el eje horizontal los intervalos de clase y sobre cada uno se levanta un rectángulo de área igual o proporcional a la frecuencia absoluta.
- b) *Polígono de frecuencias*: se sitúan los puntos que resultan de tomar en el eje horizontal las marcas de clase de los intervalos y en el eje vertical sus correspondientes frecuencias (no acumuladas), uniendo después los puntos mediante segmentos rectilíneos.

- c) *Gráfico de frecuencias acumuladas*: es la representación gráfica de las frecuencias acumuladas para todo valor numérico, teniendo en cuenta que dentro de cada intervalo de clase se supone que el número de observaciones se distribuye uniformemente. Siempre es un polígono.

Ejemplos que se van a resolver en clase

Ejemplo 1.1. En la tabla siguiente aparece el idioma en que están escritos 40 libros de una biblioteca.

francés	inglés	francés	inglés	francés	alemán	ruso	español	francés	inglés
francés	inglés	español	francés	español	francés	alemán	inglés	español	inglés
inglés	español	inglés	francés	español	ruso	alemán	francés	inglés	español
alemán	inglés	español	francés	alemán	inglés	inglés	inglés	español	francés

- Determinar la distribución de frecuencias absolutas.
- Determinar la distribución de frecuencias relativas.
- Determinar la distribución de porcentajes.

Ejemplo 1.2. La tabla siguiente muestra el número de palabras por línea de 48 líneas escogidas al azar entre todas las de un libro.

14	13	3	13	7	12	13	11	13	12	11	13
7	10	12	13	14	11	13	12	4	12	10	13
9	12	13	11	13	14	10	12	11	13	15	9
12	11	13	10	13	11	12	5	9	12	13	15

- Determinar la distribución de frecuencias absolutas, frecuencias relativas y porcentajes.
- Determinar la distribución de frecuencias acumuladas absolutas.
- Determinar la distribución de frecuencias acumuladas relativas.
- Determinar la distribución de porcentajes acumulados.

Ejemplo 1.3. La tabla siguiente recoge los datos correspondientes a la puntuación (de 0 a 10) de una muestra de 25 alumnos en un test de catalogación.

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

- Agrupar los datos en intervalos de la misma amplitud.
- A partir de la agrupación anterior determinar la distribución de frecuencias absolutas, relativas, acumuladas absolutas y acumuladas relativas.

Ejemplo 1.4. Dibujar el diagrama de barras de frecuencias absolutas de los datos de la Tabla 1.1.

Ejemplo 1.5. La siguiente tabla muestra el país de procedencia de los documentos primarios de los resúmenes contenidos en un determinado volumen de las tres revistas siguientes: Computer Abstracts, Lead Abstracts y Sociological Abstracts. Dibujar el diagrama de barras conjunto de frecuencias absolutas.

Tabla 1.4			
país de procedencia	Computer Abstracts	Lead Abstracts	Sociological Abstracts
Países Bajos	42	34	22
Francia	55	7	76
Alemania	162	37	14
Gran Bretaña	310	147	24
EEUU	966	265	552
Rusia	191	37	42
Otros	265	79	239
suma	1.991	606	969

Ejemplo 1.6. Dibujar el gráfico de sectores de los datos de la Tabla 1.1.

Ejemplo 1.7. Dibujar el diagrama de barras de frecuencias absolutas de los datos de la Tabla 1.2.

Ejemplo 1.8. Dibujar el polígono de frecuencias relativas de los datos de la Tabla 1.2.

Ejemplo 1.9. Dibujar el gráfico de frecuencias acumuladas absolutas de los datos de la Tabla 1.2.

Ejemplo 1.10. Dibujar el histograma de los datos de la Tabla 1.3 agrupados en intervalos de la misma amplitud.

Ejemplo 1.11. Dibujar el polígono de frecuencias absolutas de los datos de la Tabla 1.3 agrupados en intervalos de la misma amplitud.

Ejemplo 1.12. Dibujar el gráfico de frecuencias acumuladas absolutas de los datos de la Tabla 1.3 agrupados en intervalos de la misma amplitud.

Problemas propuestos

Problema 1.1. El gasto de una biblioteca, en euros, durante un año determinado, es:

Gasto en personal	6.570
Gasto en libros	3.450
Otros gastos	2.380

Hacer un diagrama de barras de frecuencias absolutas y un gráfico de sectores.

Problema 1.2. Una biblioteca contiene una cantidad de estantes de libros en varios idiomas tal como muestra la siguiente tabla:

Idioma	Nº de estantes
Francés	78
Alemán	47
Ruso	20
Español	30

Determinar la distribución de frecuencias relativas. Hacer un diagrama de barras de frecuencias relativas y un gráfico de sectores.

Problema 1.3. La estadística de fotocopias de una biblioteca, durante un año determinado, es la siguiente:

Reproducción de catálogos	16.110
Trabajo del personal de la biblioteca	63.350
Préstamo interbibliotecario	2.600
Copias para usuarios de la biblioteca	43.540

Determinar la distribución de porcentajes. Hacer un diagrama de barras de porcentajes y un gráfico de sectores.

Problema 1.4. La estadística de fotocopias de 4 bibliotecas (A, B, C y D), durante un año, está recogida en la siguiente tabla:

	A	B	C	D
Reproducción de catálogos	16.110	3.640	0	3.400
Trabajo del personal de la biblioteca	63.350	11.360	3.080	5.500
Préstamo interbibliotecario	2.600	1.090	560	250
Copias para usuarios de la biblioteca	43.540	58.040	1.980	0

Hacer un diagrama de barras conjunto de frecuencias absolutas.

Problema 1.5. El número de citas en diferentes campos de investigación y en distintos años viene dado en la tabla siguiente:

	1970	1980	1990
Sociología	330	414	547
Economía	299	393	295
Política	115	357	137
Psicología	329	452	258

Hacer un diagrama de barras conjunto de frecuencias relativas.

Problema 1.6. El número de descriptores (*keywords*) de 72 artículos de investigación viene dado por:

Nº de descriptores	3	4	5	6	7	8	9	10	11	12	13	14
Nº de artículos	5	8	12	7	9	9	10	5	3	2	1	1

Hacer un diagrama de barras de frecuencias absolutas.

Problema 1.7. La altura, en centímetros, de una colección de libros es la siguiente:

Altura	15	16	17	18	19	20	21	22	23	24	25	26	27
Nº de libros	1	0	3	4	4	2	4	5	2	2	2	1	1

Determinar la distribución de frecuencias relativas y hacer un polígono de frecuencias relativas.

Problema 1.8. El número de palabras por línea de una página de un libro viene dado por:

Nº de palabras	4	5	8	9	10	11	12	13	14	15	16	17
Nº de líneas	1	1	2	3	2	7	11	14	3	2	1	1

Determinar la distribución de frecuencias acumuladas absolutas y hacer el gráfico de frecuencias acumuladas absolutas.

Problema 1.9. Los siguientes datos corresponden al número de palabras por resumen de los artículos científicos de autores españoles que han publicado en una determinada revista de investigación durante un año concreto:

10	15	16	20	17	19	21	14	13	19
11	14	17	19	20	20	22	15	13	12
12	15	17	19	18	23	22	17	21	20
15	18	16	18	12	17	14	15	17	15

Determinar la distribución de frecuencias absolutas, relativas, acumuladas absolutas y acumuladas relativas. Hacer un diagrama de barras de frecuencias absolutas, un polígono de frecuencias relativas y un gráfico de frecuencias acumuladas relativas.

Problema 1.10. Los siguientes datos agrupados en intervalos se refieren al número de llamadas telefónicas recibidas en el servicio de información de una biblioteca pública durante 45 días elegidos al azar:

Nº de llamadas	(9,15]	(15,21]	(21,27]	(27,33]	(33,39]	(39,45]	(45,51]
Nº de días	2	4	8	14	10	6	1

Dibujar el histograma, el polígono de frecuencias y el gráfico de frecuencias acumuladas absolutas.

Problema 1.11. El número de socios de 84 bibliotecas públicas viene dado por:

1.995	1.050	2.500	3.000	3.000	1.500	2.500
995	995	3.000	3.000	1.200	1.450	2.500
2.750	3.000	1.600	3.000	2.250	2.750	1.800
1.250	3.250	1.800	1.750	3.250	2.100	4.500
2.100	995	3.500	2.500	1.700	2.100	1.250
3.500	3.250	1.200	950	3.250	1.700	3.000
1.500	3.500	1.500	995	2.750	3.500	2.150
1.750	2.000	2.200	1.750	2.800	750	2.000
1.500	3.500	4.500	1.950	3.000	2.200	1.600
1.200	2.400	750	1.850	2.400	1.250	3.000
800	2.750	4.000	2.050	5.500	3.750	950
995	3.750	1.500	1.800	1.200	2.500	1.250

Aunque la variable es cuantitativa discreta, se desea agrupar los datos en intervalos de la misma amplitud. A partir de esta agrupación, determinar la distribución de frecuencias y dibujar el histograma, el polígono de frecuencias y el gráfico de frecuencias acumuladas relativas.

Soluciones de los problemas propuestos

Solución del problema 1.1. La variable estadística es el *tipo o modalidad de gasto*. Es cualitativa nominal. Tiene 3 categorías, clases o modalidades. Cada vez que se realiza un gasto en la biblioteca se observa dicha variable (cada individuo es cada gasto que se hace).

Categorías (Tipos de gasto)	f_i	ángulos
Gasto en personal	6570	190'74°
Gasto en libros	3450	100'16°
Otros gastos	2380	69'10°
suma	12400	360'00°

- *Diagrama de barras de frecuencias absolutas*: se sitúan en el eje horizontal las categorías y sobre cada una de ellas se levanta un rectángulo de altura igual a la frecuencia absoluta, f_i .
- *Gráfico de sectores*: se divide el área de un círculo en sectores circulares de ángulos iguales a los que aparecen en la última columna de la tabla anterior.

Solución del problema 1.2. La variable estadística es el *idioma*. Es cualitativa nominal. Tiene 4 categorías, clases o modalidades. Los individuos a los que se les observa dicha variable son los estantes (se supone que en cada estante sólo hay libros en el mismo idioma; es decir, en un estante no se mezclan dos idiomas).

Categorías (Idiomas)	f_i	h_i	ángulos
Francés	78	0'4457	160'452°
Alemán	47	0'2686	96'696°
Ruso	20	0'1143	41'148°
Español	30	0'1714	61'704°
suma	175	1'0000	360'000°

- *Diagrama de barras de frecuencias relativas*: se sitúan en el eje horizontal las categorías y sobre cada una de ellas se levanta un rectángulo de altura igual a la frecuencia relativa, h_i .
- *Gráfico de sectores*: se divide el área de un círculo en sectores circulares de ángulos iguales a los que aparecen en la última columna de la tabla anterior.

Solución del problema 1.3. La variable estadística es el *tipo de fotocopia (¿con qué fin está hecha?)*. Es cualitativa nominal. Tiene 4 categorías, clases o modalidades. Los individuos a los que se les observa dicha variable son todas y cada una de las fotocopias que se realizan en la mencionada biblioteca durante el determinado año.

Categorías (Tipos de fotocopia)	f_i	$\%_i$	ángulos
Reproducción de catálogos	16110	12'83	46'188°
Trabajo del personal de la biblioteca	63350	50'44	181'584°
Préstamo interbibliotecario	2600	2'07	7'452°
Copias para usuarios de la biblioteca	43540	34'67	124'812°
suma	125600	100'00	360'000°

- *Diagrama de barras de porcentajes*: se sitúan en el eje horizontal las categorías y sobre cada una de ellas se levanta un rectángulo de altura igual al porcentaje, $\%_i$.

- *Gráfico de sectores*: se divide el área de un círculo en sectores circulares de ángulos iguales a los que aparecen en la última columna de la tabla anterior.

Solución del problema 1.4. Tenemos 4 variables estadísticas cualitativas nominales cuyas categorías son las mismas (Reproducción de catálogos, Trabajo del personal de la biblioteca, Préstamo interbibliotecario y Copias para usuarios de la biblioteca). Cada una de estas cuatro variables es totalmente análoga a la variable definida en el problema anterior.

Categorías (Tipos de fotocopia)	A	B	C	D
	f_i	f_i	f_i	f_i
Reproducción de catálogos	16 110	3 640	0	3 400
Trabajo del personal de la biblioteca	63 350	11 360	3 080	5 500
Préstamo interbibliotecario	2 600	1 090	560	250
Copias para usuarios de la biblioteca	43 540	58 040	1 980	0

Diagrama de barras conjunto de frecuencias absolutas: se sitúan en el eje horizontal las cuatro categorías y sobre cada una de ellas se levanta un rectángulo de altura igual a la frecuencia absoluta, f_i , con distinto color o trama de relleno para cada una de las cuatro bibliotecas.

Solución del problema 1.5. Tenemos 3 variables estadísticas cualitativas nominales cuyas categorías son las mismas (sociología, economía, política y psicología). Por ejemplo, la primera de las variables es *área de investigación de las citas que aparecen en los artículos publicados en 1970*. Los individuos a los que se les observa dicha variable son todas y cada una de las citas que aparecen en los artículos publicados en 1970. Las otras dos variables se definen de forma análoga (... 1980 y ... 1990).

Categorías (Áreas de investigación)	1970		1980		1990	
	f_i	h_i	f_i	h_i	f_i	h_i
Sociología	330	0'3075	414	0'2562	547	0'4422
Economía	299	0'2787	393	0'2432	295	0'2385
Política	115	0'1072	357	0'2209	137	0'1108
Psicología	329	0'3066	452	0'2797	258	0'2086
suma	1 073	1'0000	1 616	1'0000	1 237	1'0000

Diagrama de barras conjunto de frecuencias relativas: se sitúan en el eje horizontal las cuatro categorías y sobre cada una de ellas se levanta un rectángulo de altura igual a la frecuencia relativa, h_i , con distinto color o trama de relleno para cada uno de los tres años.

Solución del problema 1.6. La variable estadística es el *número de descriptores o keywords*. Es cuantitativa discreta. Los individuos a los que se les observa la variable son todos y cada uno de los 72 artículos de investigación de la muestra.

x_i	3	4	5	6	7	8	9	10	11	12	13	14
f_i	5	8	12	7	9	9	10	5	3	2	1	1

Diagrama de barras de frecuencias absolutas: se sitúan en el eje horizontal los x_i y sobre cada uno de ellos se levanta un segmento rectilíneo de altura igual a la correspondiente frecuencia absoluta, f_i .

Solución del problema 1.7. La variable estadística es la *altura de los libros*. Es cuantitativa continua. Los individuos a los que se les observa la variable son los 31 libros de la muestra.

x_i	15	16	17	18	19	20	21	22	23	24	25	26	27
f_i	1	0	3	4	4	2	4	5	2	2	2	1	1
h_i	0'032	0'000	0'097	0'129	0'129	0'065	0'129	0'161	0'065	0'065	0'065	0'032	0'032

Polígono de frecuencias relativas: se sitúan los puntos que resultan de tomar en el eje horizontal los distintos valores de la variable, x_i , y en el eje vertical sus correspondientes frecuencias relativas, h_i , uniendo después los puntos mediante segmentos rectilíneos.

Solución del problema 1.8. La variable estadística es el *número de palabras por línea*. Es cuantitativa discreta. Los individuos a los que se les observa la variable son todas y cada una de las 48 líneas de la página del libro.

x_i	4	5	8	9	10	11	12	13	14	15	16	17
f_i	1	1	2	3	2	7	11	14	3	2	1	1
F_i	1	2	4	7	9	16	27	41	44	46	47	48

Gráfico de frecuencias acumuladas absolutas: es la representación gráfica de las frecuencias acumuladas absolutas, F , para todo valor numérico, x . Es una gráfica en forma de “escalera”.

Solución del problema 1.9. La variable estadística es el *número de palabras por resumen*. Es cuantitativa discreta. Los individuos a los que se les observa la variable son los artículos científicos de autores españoles que han publicado en la determinada revista de investigación durante el determinado año.

x_i	f_i	h_i	F_i	H_i
10	1	0'025	1	0'025
11	1	0'025	2	0'050
12	3	0'075	5	0'125
13	2	0'050	7	0'175
14	3	0'075	10	0'250
15	6	0'150	16	0'400
16	2	0'050	18	0'450
17	6	0'150	24	0'600
18	3	0'075	27	0'675
19	4	0'100	31	0'775
20	4	0'100	35	0'875
21	2	0'050	37	0'925
22	2	0'050	39	0'975
23	1	0'025	40	1'000

- *Diagrama de barras de frecuencias absolutas*: se sitúan en el eje horizontal los x_i , y sobre cada uno de ellos se levanta un segmento rectilíneo de altura igual a la correspondiente frecuencia absoluta, f_i .
- *Polígono de frecuencias relativas*: se sitúan los puntos que resultan de tomar en el eje horizontal los distintos valores de la variable, x_i , y en el eje vertical sus correspondientes frecuencias relativas, h_i , uniendo después los puntos mediante segmentos rectilíneos.
- *Gráfico de frecuencias acumuladas relativas*: es la representación gráfica de las frecuencias acumuladas relativas, H , para todo valor numérico, x . Es una gráfica en forma de “escalera”.

Solución del problema 1.10. La variable estadística es el *número de llamadas telefónicas recibidas en el servicio de información de una biblioteca pública*. Es cuantitativa discreta. Los individuos a los que se les observa la variable son los días.

$(\ell_i, \ell_{i+1}]$	(9,15]	(15,21]	(21,27]	(27,33]	(33,39]	(39,45]	(45,51]
f_i	2	4	8	14	10	6	1
x_i	12	18	24	30	36	42	48
F_i	2	6	14	28	38	44	45

- *Histograma*: se sitúan en el eje horizontal los intervalos de clase, $(\ell_i, \ell_{i+1}]$, y sobre cada uno se levanta un rectángulo de área proporcional a la frecuencia absoluta. Como los intervalos tienen la misma amplitud, basta con hacer las alturas de los rectángulos iguales a las frecuencias absolutas, f_i .
- *Polígono de frecuencias*: se sitúan los puntos que resultan de tomar en el eje horizontal las marcas de clase, x_i , y en el eje vertical sus correspondientes frecuencias absolutas, f_i , uniendo después los puntos mediante segmentos rectilíneos.
- *Gráfico de frecuencias acumuladas absolutas*: se sitúan los puntos que resultan de tomar en el eje horizontal los extremos superiores de los intervalos de clase, ℓ_{i+1} , y en el eje vertical sus correspondientes frecuencias acumuladas absolutas, F_i , uniendo después dichos puntos mediante segmentos rectilíneos.

Solución del problema 1.11. La variable estadística es el *número de socios de la biblioteca*. Es cuantitativa discreta. Los individuos a los que se les observa la variable son las bibliotecas públicas.

$(\ell_i, \ell_{i+1}]$	f_i	x_i	H_i
(675,1 375]	19	1 025	0'2262
(1 375,2 075]	22	1 725	0'4881
(2 075,2 775]	18	2 425	0'7024
(2 775,3 475]	14	3 125	0'8690
(3 475,4 175]	8	3 825	0'9643
(4 175,4 875]	2	4 525	0'9881
(4 875,5 575]	1	5 225	1'0000

- *Histograma*: se sitúan en el eje horizontal los intervalos de clase, $(\ell_i, \ell_{i+1}]$, y sobre cada uno se levanta un rectángulo de área proporcional a la frecuencia absoluta. Como los intervalos tienen la misma amplitud, basta con hacer las alturas de los rectángulos iguales a las frecuencias absolutas, f_i .
- *Polígono de frecuencias*: se sitúan los puntos que resultan de tomar en el eje horizontal las marcas de clase, x_i , y en el eje vertical sus correspondientes frecuencias absolutas, f_i , uniendo después los puntos mediante segmentos rectilíneos.
- *Gráfico de frecuencias acumuladas relativas*: se sitúan los puntos que resultan de tomar en el eje horizontal los extremos superiores de los intervalos de clase, ℓ_{i+1} , y en el eje vertical sus correspondientes frecuencias acumuladas relativas, H_i , uniendo después dichos puntos mediante segmentos rectilíneos.