

Estadística Aplicada a la Educación

Tema 4

Tutor.

UNED Madrid-Sur (A.U. Parla)

Miguel Ángel Daza

migdaza@madridsur.uned.es

1

- La Estadística en el proceso de investigación pedagógica empírica.

2

- Problema, hipótesis / objetivos, variables y datos. Niveles de medida

4

- Organización de los datos. análisis exploratorio de datos.

5

- Reducción de datos. Medidas descriptivas básicas y representaciones gráficas.

6

- Medidas individuales.

7

- Relación entre variables. Las correlaciones. La regresión.

8

- Aplicaciones de la correlación: fiabilidad y validez de las medida.

9

- Modelos estadísticos y probabilidad. La curva normal de probabilidades.

10

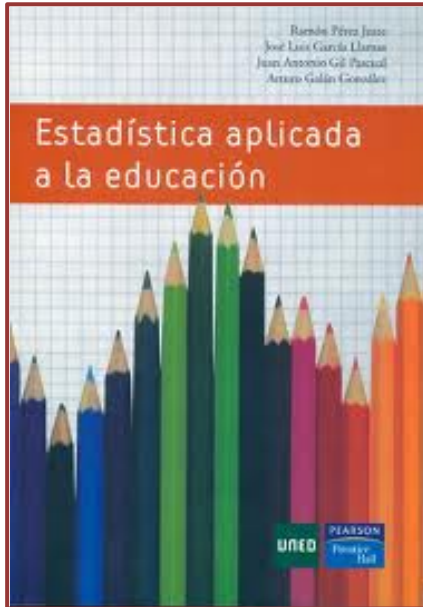
- Los baremos o normas. Muestreo. Aplicaciones.

11

- Estimación de parámetros. Errores de estimación.

12

- Introducción al contraste de hipótesis: la prueba t para el contraste de medias en los diseños de dos grupos.



- 1) Introducción.
- 2) De la definición del problema y las variables a la observación y recogida de datos.
- 3) Permisos y ética en la investigación y recogida de datos.
- 4) De los instrumentos a los datos: elección del programa, la matriz de datos y el libro de códigos.
- 5) Organización de los datos: depuración de datos y distribuciones de frecuencia.
 - 1) Distribuciones de frecuencia y aproximación al concepto de percentil.
- 6) Aproximación intuitiva a las representaciones gráficas y la curva normal.

4. ORGANIZACIÓN DE LOS DATOS. ANÁLISIS EXPLORATORIO DE DATOS.

Registro y codificación de datos.

- La matriz de datos
 - Columnas: variables
 - Filas: sujetos
- Los valores perdidos

ejercicios habilitación.sav - Editor de datos SPSS

	rendimiento	grupo	var	var	var	var
1	10.00	1.00				
2	9.00	1.00				
3	9.00	1.00				
4	8.00	1.00				
5	8.00	1.00				
6	8.00	1.00				
7	7.00	1.00				
8	6.00	1.00				
9	6.00	1.00				
10	4.00	1.00				
11	10.00	1.00				
12	9.00	1.00				
13	9.00	1.00				
14	8.00	1.00				
15	8.00	1.00				
16	8.00	1.00				
17	7.00	1.00				
18	6.00	1.00				
19	6.00	1.00				
20	4.00	1.00				
21	10.00	2.00				
22	8.00	2.00				
23	7.00	2.00				
24	7.00	2.00				
25	7.00	2.00				
26	6.00	2.00				
27	4.00	2.00				
28	4.00	2.00				
29	3.00	2.00				
30	6.00	2.00				
31	4.00	2.00				
32	3.00	2.00				
33	6.00	2.00				
34	4.00	2.00				
35	3.00	2.00				
36	6.00	2.00				
37	4.00	2.00				
38	3.00	2.00				
39	6.00	2.00				
40	4.00	2.00				
41	3.00	2.00				
42	6.00	2.00				
43	4.00	2.00				
44	3.00	2.00				
45	6.00	2.00				
46	4.00	2.00				
47	3.00	2.00				
48	6.00	2.00				
49	4.00	2.00				
50	3.00	2.00				
51	6.00	2.00				
52	4.00	2.00				
53	3.00	2.00				
54	6.00	2.00				
55	4.00	2.00				
56	3.00	2.00				
57	6.00	2.00				
58	4.00	2.00				
59	3.00	2.00				
60	6.00	2.00				
61	4.00	2.00				
62	3.00	2.00				
63	6.00	2.00				
64	4.00	2.00				
65	3.00	2.00				
66	6.00	2.00				
67	4.00	2.00				
68	3.00	2.00				
69	6.00	2.00				
70	4.00	2.00				
71	3.00	2.00				
72	6.00	2.00				
73	4.00	2.00				
74	3.00	2.00				
75	6.00	2.00				
76	4.00	2.00				
77	3.00	2.00				
78	6.00	2.00				
79	4.00	2.00				
80	3.00	2.00				
81	6.00	2.00				
82	4.00	2.00				
83	3.00	2.00				
84	6.00	2.00				
85	4.00	2.00				
86	3.00	2.00				
87	6.00	2.00				
88	4.00	2.00				
89	3.00	2.00				
90	6.00	2.00				
91	4.00	2.00				
92	3.00	2.00				
93	6.00	2.00				
94	4.00	2.00				
95	3.00	2.00				
96	6.00	2.00				
97	4.00	2.00				
98	3.00	2.00				
99	6.00	2.00				
100	4.00	2.00				



VIDEOCLASE: Ordenación y Categorización de Datos

http://www.intecca.uned.es/portavip/grabacion.php?ID_Grabacion=56793&ID_Sala=60353&hashData=e6bc01b9135499b30233c5eaa6779d95¶msToCheck=SURfR3JhYmFjaW9uLEIEX1NhGEs

4.1 Introducción.

En este capítulo, con el que comienza el **segundo bloque temático**, vamos a pasar de los grandes principios conceptuales a la práctica, a la parte empírica: **la recogida de datos y su primer ordenación.**

La **Estadística Descriptiva** comienza con un conjunto de datos: el investigador intenta trasladar las características esenciales de los datos a formas más interpretables (**distribuciones de frecuencias, gráficos**) y calcula **índices numéricos** como **promedios, percentiles y medidas de variabilidad.**

Primero organización y después análisis exploratorio. Nos referiremos al **análisis exploratorio** de los datos cuando el investigador comienza por la observación de la realidad, después estudia las observaciones y finalmente encuentra un patrón o modelo.

4.2 De la definición del problema y las variables a la observación y recogida de datos.

Uno de los errores más frecuentes de los investigadores en formación es recoger datos sin seguir adecuadamente los pasos del **proceso de investigación**.

- Centrado el tema de investigación.
- Revisada la literatura al respecto.
- Identificadas las variables a utilizar para dar respuesta al problema.

Recogida de la información: atribuir valores a las variables que son objeto del estudio. Para ello se necesita algún procedimiento que permita asignar dichos valores, número, rangos, categoría, a las variables conceptuales

Ejemplo: *Estudiar si un programa de mejora de la convivencia consigue disminuir el número de partes de amonestación en institutos de educación secundaria y si mejora el clima escolar.*

- *Variable independiente* → el programa.
- *Variables dependientes* → los partes de amonestación y el clima escolar.

4.2 De la definición del problema y las variables a la observación y recogida de datos.

Ejemplo: *Estudiar si un programa de mejora de la convivencia consigue disminuir el número de partes de amonestación en institutos de educación secundaria y si mejora el clima escolar.*

- **Variable independiente** → *el programa.*
 - 0/1 no se aplica /se aplica.
- **Variables dependientes** → *los partes de amonestación y el clima escolar.*
 - Partes → variable cuantitativa discreta de razón, de 0 a N
 - Clima escolar → cuestionarios, escalas estandarizadas, etc.

En la **investigación empírica** siempre hay que trabajar con datos empíricos, es decir, transformar en números la realidad observada. Para ello se necesitan instrumentos de medida **válidos** (que midan lo que deben medir) y **fiables** (precisos).

Una vez realizado el diseño de la investigación y elegidos los instrumentos de medida, se puede realizar:

el **trabajo de campo** → aplicar los instrumentos, recoger los datos.

4.3 Permisos y ética en la investigación y recogida de datos.

Principios éticos que deben seguirse en la investigación educativa:

1. **Participación voluntaria** de los sujetos
2. **Informe consentido** = los sujetos conocen la naturaleza de la investigación
3. **Evitación de riesgo de daño físico o psíquico**
4. **Confidencialidad y anonimato**
5. **Respetar el lugar**, causando las menos alteraciones posibles.
6. **No manipular los datos obtenidos** y reflejar fielmente los resultados aunque contradigan su hipótesis inicial.
7. **Citar adecuadamente los autores y las fuentes** de información durante la revisión del estado de la cuestión.

PRÁCTICAS ÉTICAS = DEONTOLOGÍA PROFESIONAL

En educación no se da prácticamente la **investigación experimental** (muestreo aleatorio y control estricto que chocaría con algunos aspectos éticos), llegando sólo al nivel "**cuasi-experimental**" (muestreo incidental y control relativo)

4.4 De los instrumentos a los datos: elección del programa, la matriz de datos y el libro de códigos.

Una vez que hemos aplicado los instrumentos, debemos trasladar los datos a una hoja de cálculo a algún programa que nos permita posteriormente trabajar con ellos: **INTRODUCIR** o **TABULAR** los datos.

Las puntuaciones que introducimos se llaman **PUNTUACIONES DIRECTAS (Xi)**: es decir, la puntuación o puntuaciones que obtiene un sujeto tras aplicarle un instrumento de medida.

La **CODIFICACIÓN** de datos consiste en la asignación de números o caracteres a los valores de la variable. Antes de empezar a tabular introducir datos, es conveniente escribir el **LIBRO DE CÓDIGOS**:

documento en el que se especifican todas las variables del estudio en el orden en el que serán introducidas en la matriz de datos.

4.4 De los instrumentos a los datos: elección del programa, la matriz de datos y el libro de códigos.

Ítem	Variable	Etiqueta variable	Código	Etiqueta valores
C.1	C1Ident	Identificación	Asignar un valor numérico a cada sujeto (por ejemplo, 01; 02; 03, ...).	—
C.2	C2Sexo	Sexo	0 1	Hombre Mujer
C.3	C3C_aut	Comunidad Autónoma	1 2 3 4 5	Galicia Extremadura Andalucía Madrid Castilla y León
C.4	C4CI	Cociente intelectual	Cualquier valor entre 50-150	—
C.5	C5Rmat	Rendimiento matemático	Cualquier valor entre 0-100	—
C.6	C6Satis	Satisfacción con el curso	1 2 3 4 5	Muy insatisfecho Muy satisfecho

4.4 De los instrumentos a los datos: elección del programa, la matriz de datos y el libro de códigos.

Es importante definir el valor que asignaremos a los **DATOS PERDIDOS** o **"missing data"**: hay dos opciones:

1. Dejar en blanco en la matriz de datos (**"missing de sistema"**, de modo que no se contabilice a la hora de calcular los estadísticos (**un error muy común es poner ceros para representar los valores perdidos**))
2. Definir un "valor perdido por el usuario" (**"missing de usuario"**, (siempre que no sea un valor admisible dentro de la escala de medida de la variable))

No contesta → **"missing de sistema"** **No sabe** → **"missing de usuario"**



<http://www.r-project.org/>



<https://www.gnu.org/software/pspp/>



<http://www-01.ibm.com/software/es/analytics/spss/>



<http://office.microsoft.com/es-es/excel/>

4.4 De los instrumentos a los datos: elección del programa, la matriz de datos y el libro de códigos.

	A	B	C	D	E	F	G
1	ID	EF1	EF2	EF3	PARCIAL	PRACTICO 1	
2	1	37,04	3				
3	2	48,15	4				
4	3	22,22	6		28,33	1	
5	4	44,44	4		35	1	
6	5	37,04	5	44,12	41,67	4,1	
7	6	74,07	7	67,65	81,67	5,5	
8	7	33,33		26,47	75	6,5	
9	8		5				
10	9	37,04	6	38,24	38,33	2,1	
11	10	37,04	6	44,12	53,33	6,9	
12	11	55,56	9		78,33	7,1	
13	12	70,37	8	64,71	88,33	7	
14	13	74,07	6		76,67	5,6	
15	14	51,65	4	35,88	60	6	
16	15	62,96	10	97,06	98,33	9	
17	16	59,26	6	70,59	86,67	7,7	
18	17	40,74	8	82,35	88,33	9,6	
19	18	44,44	2		35	1	
20	19	74,07	4				
21	20				61,67	3,1	
22	21	62,96	8	73,53	63,33	6,8	
23	22	40,74	4				
24	23	62,96	6		41,67	1	

La forma habitual de introducir los datos consiste en conformar lo que se denomina una **MATRIZ DE DATOS**: una tabla de doble entrada donde:

- las **filas** representan a los **sujetos**
- las **columnas** a las distintas **variables** medidas.

Figura 4.1. Matriz de datos.

4.5 Organización de los datos: depuración de datos y distribuciones de frecuencia.

Antes de empezar a realizar los análisis estadísticos, debemos hacer la **DEPURACIÓN DE DATOS**, que tiene 2 fases:

1. El **CONTROL DE CALIDAD** de la tabulación: comprobar la fidelidad de la tabulación.
2. La **DEPURACIÓN DE DATOS**: consiste en verificar si hay valores “fuera de rango” según se habían definido en el libro de código. Para ello, lo más sencillo es hacer un análisis solicitando los **valores mínimo y máximo** para cada variable.

Tabla 4.1. Estadísticos descriptivos.

	N	Mínimo	Máximo
city	119	0,00	1,00
p1 questionair no	119	1	69
p2 code	116	1	3
p3 relation	118	1	3
p4 father lives with the child	117	0	1
p5 education level father	116	1	5
p6 employment father	106	0	1
p6.1 work type father	78	1	3
N válido (según lista)	69		

4.5 Organización de los datos: depuración de datos y distribuciones de frecuencia.

Una primera forma de organización de los datos es la **DISTRIBUCIÓN DE FRECUENCIAS**:

En una **distribución de frecuencias** aparecen todas las puntuaciones obtenidas en una variable (puntuaciones directas: X_i) y el número de veces que se repite cada puntuación (**frecuencia absoluta: f_i**)

También son muy utilizadas:

- la **frecuencia relativa** (frecuencia absoluta dividida por el número total de puntuaciones: f_i/N), que multiplicada por 100, indica el **porcentaje** de aparición de una puntuación respecto del total;
- y la **frecuencia acumulada**, que muestra el número de sujetos a los que supera una puntuación determinada.

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Father	28	23,5	23,7	23,7
	Mother	76	63,9	64,4	88,1
	Other	14	11,8	11,9	100,0
	Total	118	99,2	100,0	
Perdidos	Sistema	1	0,8		
Total		119	100,0		

4.5 Organización de los datos: depuración de datos y distribuciones de frecuencia.

Una **distribución de frecuencias** suele incluir:

1ª Columna: el número de sujetos o casos que corresponde a cada puntuación directa: **frecuencia absoluta**.

2ª Columna: el porcentaje relativo (o **frecuencia relativa** multiplicada por 100) para cada código, incluidos los valores perdidos.

3ª Columna: el porcentaje válido, que es el **porcentaje relativo** de cada valor excluidos los valores perdidos.

4ª Columna: el **porcentaje acumulado**, indica el porcentaje de casos que deja por debajo de sí el límite superior de cada puntuación.

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Father	28	23,5	23,7	23,7
	Mother	76	63,9	64,4	88,1
	Other	14	11,8	11,9	100,0
	Total	118	99,2	100,0	
Perdidos	Sistema	1	0,8		
Total		119	100,0		

Frecuencia absoluta $\rightarrow n_i$

$$\sum_{i=1}^k n_i = n_1 + \dots + n_k = n$$

Frecuencia relativa $\rightarrow f_i$

$$f_i = \frac{n_i}{n}$$

$$\sum_{i=1}^k f_i = f_1 + \dots + f_k = 1$$

Frecuencia absoluta acumulada $\rightarrow N_i$

$$N_i = n_1 + \dots + n_i = \sum_{j=1}^i n_j$$

Frecuencia relativa acumulada $\rightarrow F_i$

$$F_i = \frac{N_i}{n}$$

$$F_i = f_1 + \dots + f_i = \sum_{j=1}^i f_j$$

4.5.1 Distribuciones de frecuencia y aproximación al concepto de percentil.

Ejemplo: un sujeto que haya puntuado con un "5" supera casi a un 61% de los sujetos de su grupo. Los porcentajes acumulados (sin decimales) se utilizan mucho en la **construcción de baremos** para interpretar las puntuaciones en los tests, donde reciben el nombre de *percentiles*.

El percentil indica el porcentaje de sujetos que deja por debajo de sí una puntuación determinada.

Tabla 4.4. ATCL_01.

Puntuación directa	Frecuencia absoluta	Frecuencia relativa $\times 100$	Frecuencia acumulada $\times 100$ (porcentaje acumulado)
Válidos 1	179	17,7	17,7
2	93	9,2	26,9
3	107	10,6	37,5
4	128	12,6	50,1
5	108	10,7	60,8
6	131	12,9	73,7
7	266	26,3	100,0
Total	1.012	100,0	

4.5.1 Distribuciones de frecuencia y aproximación al concepto de percentil.

Ejemplo:

Distribución de Alumnos EAE (Parla) por la Inicial del Primer Apellido

Inicial	Frecuencia
A	4
B	3
C	7
D	5
F	3
G	10
H	1
L	2
M	2
P	4
R	6
S	4
	51

Inicial	Frecuencia
A	3
B	2
C	3
D	1
E	1
F	3
G	2
H	1
L	4
M	1
N	2
P	1
Q	1
R	4
S	3
V	1
	33



4.5.1 Distribuciones de frecuencia y aproximación al concepto de percentil.

Ejemplo:

Distribución de Alumnos EAE (Parla) por la Inicial del Primer Apellido

Inicial	Frecuencia	Frecuencia Relativa	Porcentaje
A	4	0,078	7,84%
B	3	0,059	5,88%
C	7	0,137	13,73%
D	5	0,098	9,80%
F	3	0,059	5,88%
G	10	0,196	19,61%
H	1	0,020	1,96%
L	2	0,039	3,92%
M	2	0,039	3,92%
P	4	0,078	7,84%
R	6	0,118	11,76%
S	4	0,078	7,84%
	51	1	100%

Inicial	Frecuencia	Frecuencia Relativa	Porcentaje
A	3	0,091	9,09%
B	2	0,061	6,06%
C	3	0,091	9,09%
D	1	0,030	3,03%
E	1	0,030	3,03%
F	3	0,091	9,09%
G	2	0,061	6,06%
H	1	0,030	3,03%
L	4	0,121	12,12%
M	1	0,030	3,03%
N	2	0,061	6,06%
P	1	0,030	3,03%
Q	1	0,030	3,03%
R	4	0,121	12,12%
S	3	0,091	9,09%
V	1	0,030	3,03%
	33	1	100%

4.5.1 Distribuciones de frecuencia y aproximación al concepto de percentil.

Ejemplo:

Distribución de Alumnos EAE (Parla) por la Inicial del Primer Apellido

Inicial	Frecuencia	Frecuencia Relativa	Porcentaje	Frecuencia Acumulada	Frecuencia Relativa Acumulada	Porcentaje Acumulado
A	4	0,078	7,84%	4	0,078	7,84%
B	3	0,059	5,88%	7	0,137	13,73%
C	7	0,137	13,73%	14	0,275	27,45%
D	5	0,098	9,80%	19	0,373	37,25%
F	3	0,059	5,88%	22	0,431	43,14%
G	10	0,196	19,61%	32	0,627	62,75%
H	1	0,020	1,96%	33	0,647	64,71%
L	2	0,039	3,92%	35	0,686	68,63%
M	2	0,039	3,92%	37	0,725	72,55%
P	4	0,078	7,84%	41	0,804	80,39%
R	6	0,118	11,76%	47	0,922	92,16%
S	4	0,078	7,84%	51	1,000	100,00%
	51	1	100%			



4.5.1 Distribuciones de frecuencia y aproximación al concepto de percentil.

Ejemplo:

Distribución de Alumnos EAE (Parla) por la Inicial del Primer Apellido

Inicial	Frecuencia	Frecuencia Relativa	Porcentaje	Frecuencia Acumulada	Frecuencia Relativa Acumulada	Porcentaje Acumulado
A	3	0,091	9,09%	3	0,091	9,09%
B	2	0,061	6,06%	5	0,152	15,15%
C	3	0,091	9,09%	8	0,242	24,24%
D	1	0,030	3,03%	9	0,273	27,27%
E	1	0,030	3,03%	10	0,303	30,30%
F	3	0,091	9,09%	13	0,394	39,39%
G	2	0,061	6,06%	15	0,455	45,45%
H	1	0,030	3,03%	16	0,485	48,48%
L	4	0,121	12,12%	20	0,606	60,61%
M	1	0,030	3,03%	21	0,636	63,64%
N	2	0,061	6,06%	23	0,697	69,70%
P	1	0,030	3,03%	24	0,727	72,73%
Q	1	0,030	3,03%	25	0,758	75,76%
R	4	0,121	12,12%	29	0,879	87,88%
S	3	0,091	9,09%	32	0,970	96,97%
V	1	0,030	3,03%	33	1,000	100,00%
	33	1	100%			

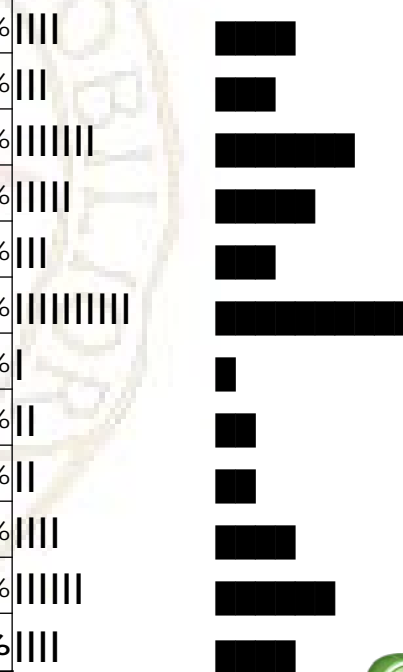
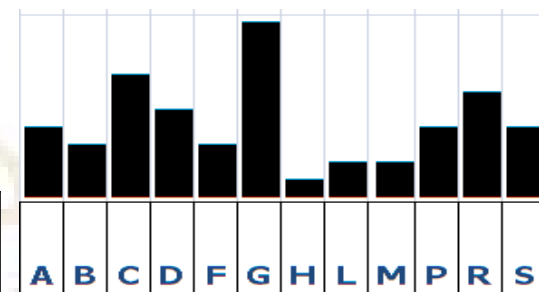


4.5.1 Distribuciones de frecuencia y aproximación al concepto de percentil.

Ejemplo:

Distribución de Alumnos EAE (Parla) por la Inicial del Primer Apellido

Inicial	Fecuencia	Frecuencia Relativa	Porcentaje	Frecuencia Acumulada	Porcentaje Acumulado
A	4	0,078	7,84%	4	7,84%
B	3	0,059	5,88%	7	13,73%
C	7	0,137	13,73%	14	27,45%
D	5	0,098	9,80%	19	37,25%
F	3	0,059	5,88%	22	43,14%
G	10	0,196	19,61%	32	62,75%
H	1	0,020	1,96%	33	64,71%
L	2	0,039	3,92%	35	68,63%
M	2	0,039	3,92%	37	72,55%
P	4	0,078	7,84%	41	80,39%
R	6	0,118	11,76%	47	92,16%
S	4	0,078	7,84%	51	100,00%
	51	1	100%		

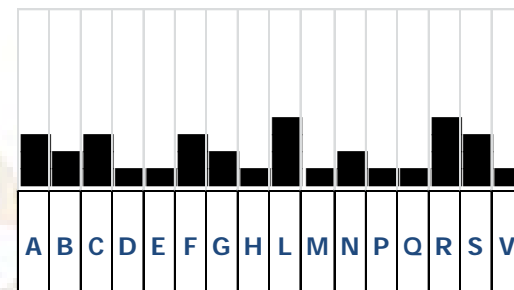


4.5.1 Distribuciones de frecuencia y aproximación al concepto de percentil.

Ejemplo:

Distribución de Alumnos EAE (Parla) por la Inicial del Primer Apellido

Inicial	Frecuencia	Frecuencia Relativa	Porcentaje	Frecuencia Acumulada	Frecuencia Relativa Acumulada	Porcentaje Acumulado
A	3	0,091	9,09%	3	0,091	9,09%
B	2	0,061	6,06%	5	0,152	15,15%
C	3	0,091	9,09%	8	0,242	24,24%
D	1	0,030	3,03%	9	0,273	27,27%
E	1	0,030	3,03%	10	0,303	30,30%
F	3	0,091	9,09%	13	0,394	39,39%
G	2	0,061	6,06%	15	0,455	45,45%
H	1	0,030	3,03%	16	0,485	48,48%
L	4	0,121	12,12%	20	0,606	60,61%
M	1	0,030	3,03%	21	0,636	63,64%
N	2	0,061	6,06%	23	0,697	69,70%
P	1	0,030	3,03%	24	0,727	72,73%
Q	1	0,030	3,03%	25	0,758	75,76%
R	4	0,121	12,12%	29	0,879	87,88%
S	3	0,091	9,09%	32	0,970	96,97%
V	1	0,030	3,03%	33	1,000	100,00%
	33	1	100%			

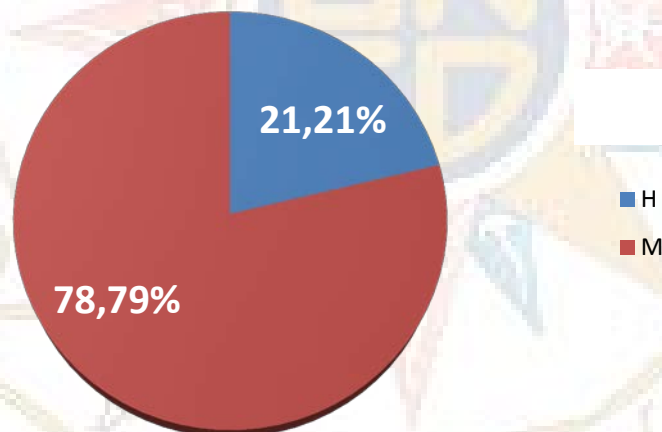


4.5.1 Distribuciones de frecuencia y aproximación al concepto de percentil.

Ejemplo:

Distribución de Alumnos EAE (Parla) por sexo

Inicial	Frecuencia	Frecuencia Relativa	Porcentaje	Frecuencia Acumulada	Frecuencia Relativa Acumulada	Porcentaje Acumulado
H	7	0,212	21,21%	7	0,212	21,21%
M	26	0,788	78,79%	33	1,000	100,00%
	33	1	100,0%			



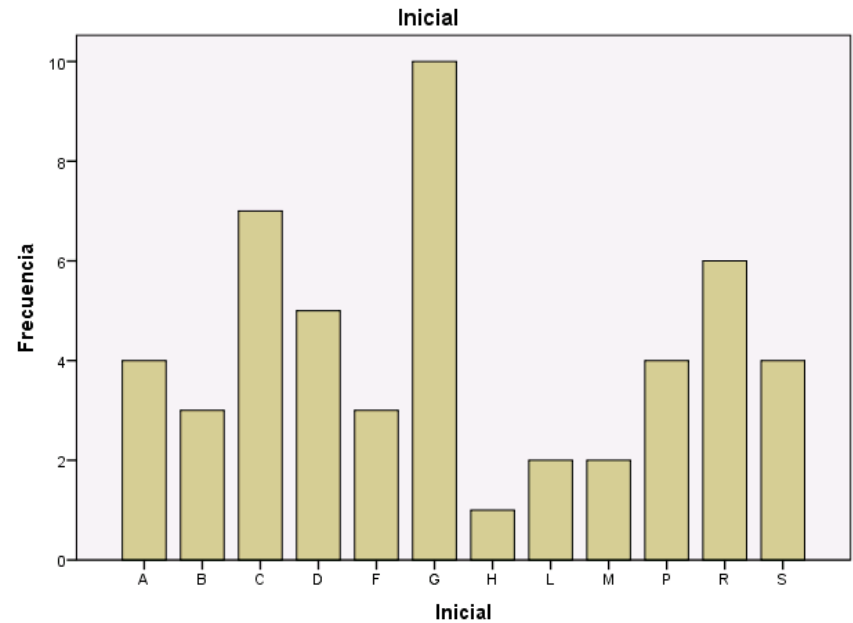
4.5.1 Distribuciones de frecuencia y aproximación al concepto de percentil.

*Ejemplo_EAE.sav [Conjunto_de_datos0] - IBM SPSS Statistics Editor de datos

Archivo Edición Ver Datos Transformar Analizar Marketing directo Gráficos Utilidades Ventana Ayuda

28: Inicial

	Inicial	casos	var	var	var	var	var	var
1	A	4,00						
2	B	3,00						
3	C	7,00						
4	D	5,00						
5	F	3,00						
6	G	10,00						
7	H	1,00						
8	L	2,00						
9	M	2,00						
10	P	4,00						
11								
12								
13								
14								



Inicial

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	A	4	7,8	7,8	7,8
	B	3	5,9	5,9	13,7
	C	7	13,7	13,7	27,5
	D	5	9,8	9,8	37,3
	F	3	5,9	5,9	43,1
	G	10	19,6	19,6	62,7
	H	1	2,0	2,0	64,7
	L	2	3,9	3,9	68,6
	M	2	3,9	3,9	72,5
	P	4	7,8	7,8	80,4
	R	6	11,8	11,8	92,2
	S	4	7,8	7,8	100,0
	Total	51	100,0	100,0	

Casos ponderados por casos

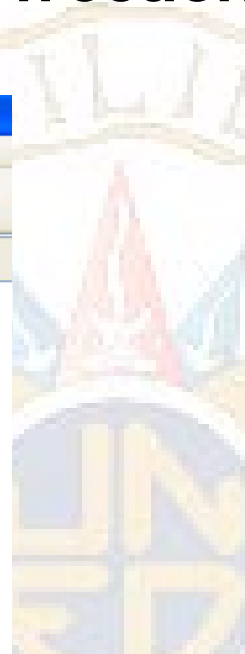


4.5.1 Distribuciones de frecuencia y aproximación al concepto de percentil.

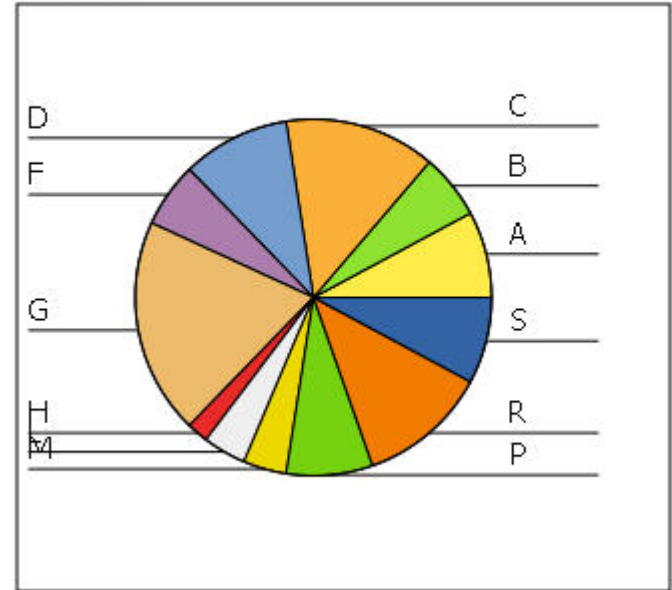
*[Ejemplo_EAE] – PSPP/IRE Editor de Datos

Archivo Editar Vista Datos Transformar Analizar Utilidades Ventanas Ayuda

Caso	inicial	CASOS
1	A	4.00
2	B	3.00
3	C	7.00
4	D	5.00
5	F	3.00
6	G	10.00
7	H	1.00
8	L	2.00
9	M	2.00
10	P	4.00
11	R	6.00
12	S	4.00
13		



inicial



inicial

Etiqueta de Valor	Valor	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
A		4.00	7.84	7.84	7.84
B		3.00	5.88	5.88	13.73
C		7.00	13.73	13.73	27.45
D		5.00	9.80	9.80	37.25
F		3.00	5.88	5.88	43.14
G		10.00	19.61	19.61	62.75
H		1.00	1.96	1.96	64.71
L		2.00	3.92	3.92	68.63
M		2.00	3.92	3.92	72.55
P		4.00	7.84	7.84	80.39
R		6.00	11.76	11.76	92.16
S		4.00	7.84	7.84	100.00
<i>Total</i>		51.00	100.0	100.0	



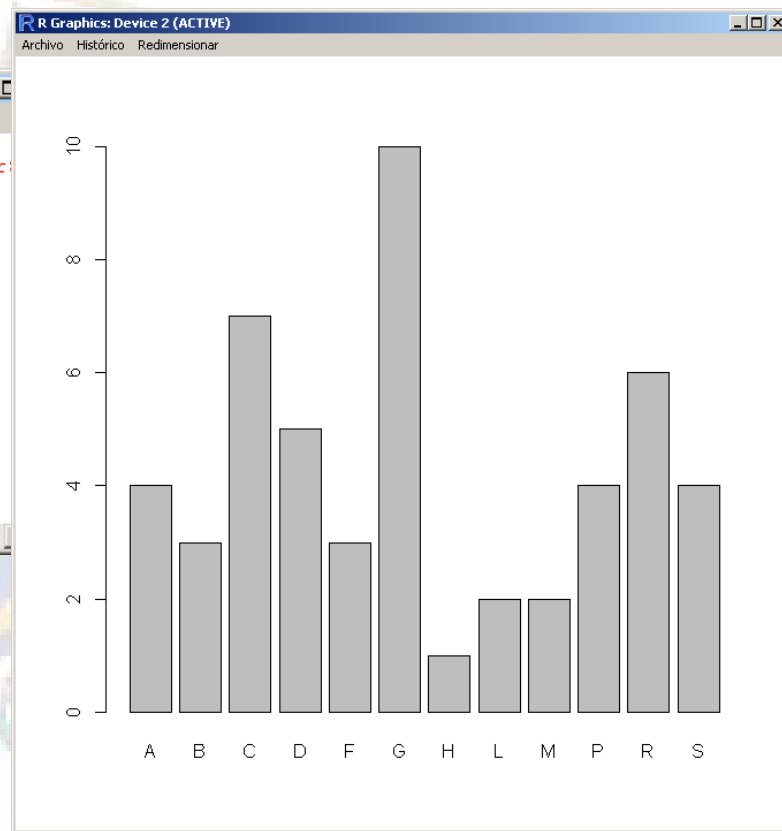
4.5.1 Distribuciones de frecuencia y aproximación al concepto de percentil.

```
R Console (32-bit)
Archivo Editar Misc Paquetes Ventanas Ayuda

> Inicial<-c(rep("A",4),rep("B",3),rep("C",7),rep("D",5),rep("F",3),rep("G",10),rep("H",1),rep("L",2),rep("M",2),rep("P",4),rep("R",6),rep("S",4))
> Inicial
 [1] "A" "A" "A" "A" "B" "B" "B" "C" "C" "C" "C" "C" "C" "C" "D" "D" "D" "D" "D"
 [20] "F" "F" "F" "G" "G" "G" "G" "G" "G" "G" "G" "G" "G" "H" "L" "L" "M" "M" "P"
 [39] "P" "P" "P" "R" "R" "R" "R" "R" "R" "R" "R" "S" "S" "S" "S"
> summary(Inicial)
  Length      Class      Mode
    51 character character
> table(Inicial)
Inicial
 A  B  C  D  F  G  H  L  M  P  R  S
 4  3  7  5  3 10  1  2  2  4  6  4
> barplot(table(Inicial))
```

```
R Console (32-bit)
Archivo Editar Misc Paquetes Ventanas Ayuda

> Inicial<-c("A","B","C","D","F","G","H","L","M","P","R","S")
> ni<-c(4,3,7,5,3,10,1,2,2,4,6,4) #frecuencia absoluta
> N<-sum(ni)
> fi<-ni/N #frecuencia relativa
> pi<-fi*100 #porcentaje
> nai<-cumsum(ni) #frecuencia absoluta acumulada
> fai<-cumsum(fi) #frecuencia relativa acumulada
> pai<-fai*100 #porcentaje acumulado
> data.frame(Inicial,ni,fi,pi,nai,fai,pai)
  Inicial ni      fi      pi nai      fai      pai
1      A  4 0.07843137  7.843137  4 0.07843137  7.843137
2      B  3 0.05882353  5.882353  7 0.13725490 13.725490
3      C  7 0.13725490 13.725490 14 0.27450980 27.450980
4      D  5 0.09803922  9.803922 19 0.37254902 37.254902
5      F  3 0.05882353  5.882353 22 0.43137255 43.137255
6      G 10 0.19607843 19.607843 32 0.62745098 62.745098
7      H  1 0.01960784  1.960784 33 0.64705882 64.705882
8      L  2 0.03921569  3.921569 35 0.68627451 68.627451
9      M  2 0.03921569  3.921569 37 0.72549020 72.549020
10     P  4 0.07843137  7.843137 41 0.80392157 80.392157
11     R  6 0.11764706 11.764706 47 0.92156863 92.156863
12     S  4 0.07843137  7.843137 51 1.00000000 100.000000
> |
```



```
Inicial<-c("A","B","C","D","F","G","H","L","M","P","R","S")
ni<-c(4,3,7,5,3,10,1,2,2,4,6,4) #frecuencia absoluta
N<-sum(ni)
fi<-ni/N #frecuencia relativa
pi<-fi*100 #porcentaje
nai<-cumsum(ni) #frecuencia absoluta acumulada
fai<-cumsum(fi) #frecuencia relativa acumulada
pai<-fai*100 #porcentaje acumulado
data.frame(Inicial,ni,fi,pi,nai,fai,pai)
```



Ejemplo:

Distribución de Páginas libro EAE por Temas

Tema	Página Inicial	Página final	Fecuencia (Num. Pag.)	Frecuencia Relativa	Porcentaje	Frecuencia Acumulada	Frecuencia Relativa Acumulada	Porcentaje Acumulado
T1	1	14	14	0,069	6,93%	14	0,069	6,93%
T2	15	40	26	0,129	12,87%	40	0,198	19,80%
T4	81	94	14	0,069	6,93%	54	0,267	26,73%
T5	95	114	20	0,099	9,90%	74	0,366	36,63%
T6	115	128	14	0,069	6,93%	88	0,436	43,56%
T7	129	150	22	0,109	10,89%	110	0,545	54,46%
T8	151	176	26	0,129	12,87%	136	0,673	67,33%
T9	177	200	24	0,119	11,88%	160	0,792	79,21%
T10	201	222	22	0,109	10,89%	182	0,901	90,10%
T11	223	242	20	0,099	9,90%	202	1,000	100,00%
			202	1	100%			

4.6 Aproximación intuitiva a las representaciones gráficas y la curva normal.

Una **representación gráfica** es una forma atractiva de ordenar la información disponible en la matriz de datos y comprenderla con un simple golpe de vista. Para realizar un gráfico necesitamos simplemente una distribución de frecuencias.

Hay distintas formas de representar los datos, dependiendo de la naturaleza de las variables: *sectores, barras, polígono de frecuencias, tallo y hojas, etc.*

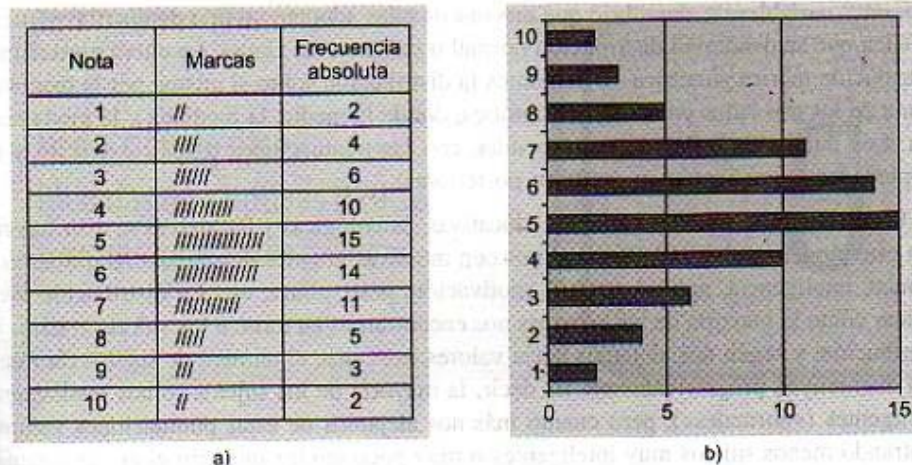


Figura 4.3

4.6 Aproximación intuitiva a las representaciones gráficas y la curva normal.

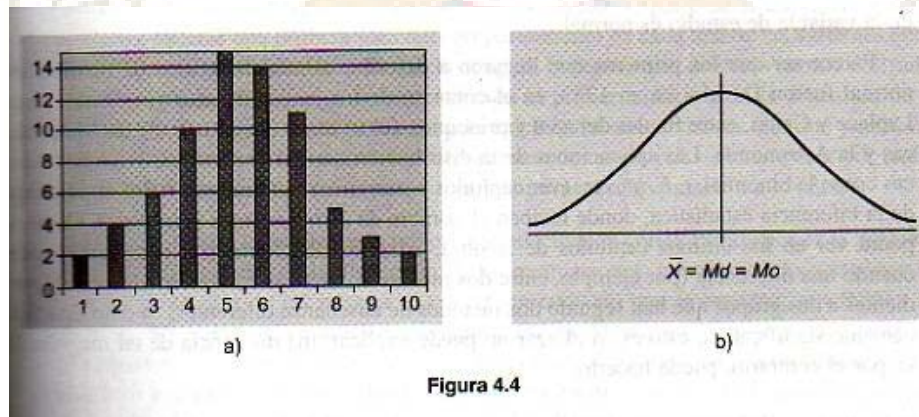
Muchas variables educativas, psicológicas y biológicas se distribuyen según la llamada

distribución normal = curva normal = campana de Gauss

(rendimiento académico, inteligencia, aptitud, motivación, peso, altura)

La distribución viene a indicar cómo la mayoría de las personas nos encontramos en torno a los valores medios de la distribución y según nos alejamos hacia valores extremos el número de sujetos disminuye progresivamente

La **curva normal** es una distribución teórica, **simétrica** y **asintótica**, donde la **media**, la **mediana** y la **moda coinciden**



Resumen

Estadística Descriptiva

Análisis exploratorio

Método confirmatorio

Proceso de investigación

Recogida de la información

Elección de los instrumentos de medida

Trabajo de campo

Ética de la investigación

Principios éticos que deben seguirse

Libro de códigos

Introducir o tabular los datos

Puntuaciones directas

La codificación de datos

Datos perdidos o "*missing data*"

No SABE --> "*missing de usuario*"

No CONTESTA --> "*missing de sistema*"

Matriz de datos

Filas-> *Sujetos* y Columnas-> *Variables*

Ordenar y categorizar los datos

Depuración de datos -> valores fuera de rango

Control de calidad de la tabulación

Distribución de frecuencias

Puntuaciones directas

Frecuencia absoluta

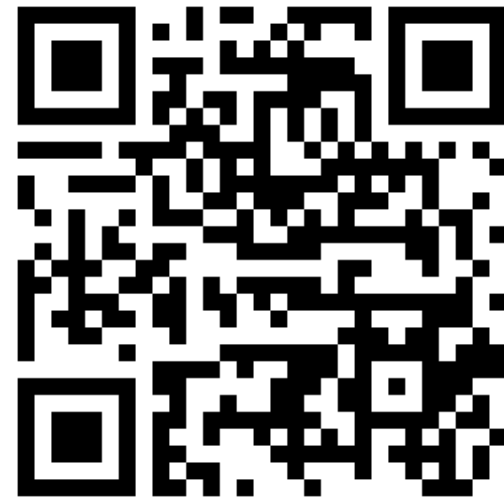
Frecuencia relativa

Porcentaje relativo

Frecuencia acumulada

PREGUNTAS

Exámenes
anteriores



Trasladar los datos recogidos a una hoja de cálculo o programa estadístico que nos permita trabajar con ellos, se denomina:

Seleccione una:

- a. Etiquetar datos.
- b. Codificar datos.
- c. Tabular datos.

Antes de darles tratamiento estadístico, que es necesario hacer con los datos que aparecen en la matriz?

Seleccione una:

- a. Completarlos.
- b. Depurarlos.
- c. Ordenarlos.

Trasladar los datos recogidos a una hoja de cálculo o programa estadístico que nos permita trabajar con ellos, se denomina:

Seleccione una:

- a. Etiquetar datos.
- b. Codificar datos.
- c. Tabular datos. ✓

Antes de darles tratamiento estadístico, que es necesario hacer con los datos que aparecen en la matriz?

Seleccione una:

- a. Completarlos.
- b. Depurarlos. ✓
- c. Ordenarlos.

Para trasladar la variable "religión" con tres categorías (católico, musulmán y judío) a una matriz de datos, la variable ocupará:

Seleccione una:

- a. Tres filas, una por cada nivel o categoría.
- b. Tres columnas, una por cada nivel o categoría.
- c. Una única columna.

En una matriz de datos, según la forma habitual de componerla:

Seleccione una:

- a. El tipo de análisis nos condicionará si el sujeto ocupa una fila o una columna.
- b. Los datos de un sujeto nos ocupará una columna.
- c. Los datos de un sujeto nos ocupará una fila

Para trasladar la variable "religión" con tres categorías (católico, musulmán y judío) a una matriz de datos, la variable ocupará:

Seleccione una:

- a. Tres filas, una por cada nivel o categoría.
- b. Tres columnas, una por cada nivel o categoría.
- c. Una única columna. ✓

En una matriz de datos, según la forma habitual de componerla:

Seleccione una:

- a. El tipo de análisis nos condicionará si el sujeto ocupa una fila o una columna.
- b. Los datos de un sujeto nos ocupará una columna.
- c. Los datos de un sujeto nos ocupará una fila ✓

La comprobación de errores de mecanización de los datos recogidos se produce en la fase de:

Seleccione una:

- a. Análisis descriptivo de datos.
- b. Organización de datos.
- c. Análisis inferencial de datos.

Los aspectos éticos de la investigación educativa, según Creswell, deberán ser tenidos en cuenta:

Seleccione una:

- a. Durante el proceso de investigación.
- b. Como una reflexión a posteriori.
- c. Al comienzo de una investigación.

La comprobación de errores de mecanización de los datos recogidos se produce en la fase de:

Seleccione una:

- a. Análisis descriptivo de datos.
- b. Organización de datos. ✓
- c. Análisis inferencial de datos.

Los aspectos éticos de la investigación educativa, según Creswell, deberán ser tenidos en cuenta:

Seleccione una:

- a. Durante el proceso de investigación.
- b. Como una reflexión a posteriori.
- c. Al comienzo de una investigación. ✓

Una matriz de datos es. ..

Seleccione una:

- a. Un fichero aún no depurado.
- b. El fichero sin datos.
- c. Un fichero que está aislado.

La depuración de datos nos permite:

Seleccione una:

- a. Depurar las responsabilidades de los errores de codificación.
- b. Identificar los valores fuera de rango de las variables del estudio.
- c. Realizar el contraste estadístico de hipótesis.

Una matriz de datos es. ...

Seleccione una:

- a. Un fichero aún no depurado. ✓
- b. El fichero sin datos.
- c. Un fichero que está aislado.

La depuración de datos nos permite:

Seleccione una:

- a. Depurar las responsabilidades de los errores de codificación.
- b. Identificar los valores fuera de rango de las variables del estudio. ✓
- c. Realizar el contraste estadístico de hipótesis.

La curva normal identifica a los sujetos “normales o frecuentes”:

Seleccione una:

- a. En el extremo izquierdo de la distribución.
- b. En la parte central de la distribución.
- c. En el extremo derecho de la distribución.

El análisis exploratorio de datos pretende:

Seleccione una:

- a. Confirmar la validez de una hipótesis.
- b. Describir lo que sucede en la realidad.
- c. Explorar si se confirma la hipótesis del investigador.

La curva normal identifica a los sujetos "normales o frecuentes":

Seleccione una:

- a. En el extremo izquierdo de la distribución.
- b. En la parte central de la distribución. ✓
- c. En el extremo derecho de la distribución.

El análisis exploratorio de datos pretende:

Seleccione una:

- a. Confirmar la validez de una hipótesis.
- b. Describir lo que sucede en la realidad. ✓
- c. Explorar si se confirma la hipótesis del investigador.



Estadística aplicada a la Educación

EJERCICIOS

TEMA : 4.

ACTIVIDADES RECOMENDADAS TEMA 4

Construya un pequeño cuestionario de unas 10 preguntas o ítems sobre cualquier temática cuyas respuestas exijan una respuesta numérica (edad, ingresos, grado de acuerdo ante algunas afirmaciones, etc.) al menos en 5 de ellas.

Aplíquelo a una muestra mínima de 10 sujetos.

Construya el libro de códigos y utilice una hoja de cálculo para introducir los datos.

Construya una distribución de frecuencias y valore si los datos fueron introducidos correctamente. Puede hacerlo manualmente, pero es preferible utilizar algún programa a su alcance. Por ejemplo, Excel tiene una aplicación gratuita de Análisis de datos que se autoinstala como un complemento del programa. Hay otros programas de software libre por la red.

