Análisis de datos I

Grado en Psicología UAM

Apuntes

1. CONCEPTOS GENERALES

Estadística: es la ciencia que se ocupa de la ordenación y análisis de datos procedentes de muestras y de la realización de inferencias sobre las poblaciones de las que éstas proceden

Estadística descriptiva:

Estadística inferencial:

Población: conjunto de todos los individuos o unidades que son objeto principal del estudio.
Muestra: subconjunto de unidades o individuos de la población de estudio.
Individuo: cada uno de los elementos que componen una población o muestra, ya sean personas, objetos...
Parámetro: aquel valor que resume una determinada información referente a la población.
Estadístico: aquel valor que expresa una determinada información referente a una muestra.

Variable: Característica observable que cambia entre los elementos de una población. Característica que puede ser medida o contada y que puede adoptar más de un valor.

Variables nominales: son aquellas variables que tienen un conjunto de categorías que no tienen ninguna tipo de jerarquía.

Variables ordinales: son aquellas variables donde el conjunto de categorías tienen una jerarquía u orden.

Variables cuantitativas o numéricas: Aquellas variables que recogen como información una medida (una cantidad numérica) de lo que se está observando. Ej.: Edad, peso, tensión arterial, número de hijos, hermanos, etc. (datos métricos).

- **Discretas:** Son aquellas variables que tienen un conjunto finito de valores y en caso infinito solamente toman los números enteros

- **Continuas:** Son aquellas variables donde el conjunto de posibles valores entre dos números fijos es infinito.

2. ORGANIZACIÓN Y REPRESENTACIÓN DE DATOS. MEDIDAS DE POSICIÓN.

DISTRIBUCIÓN DE FRECUENCIAS

Frecuencias absolutas (n_i): es el número de veces que aparece cada observación

Frecuencias absolutas acumuladas (n_a): Suma de las frec. absolutas menores o iguales al dato xi.

Frecuencia relativa (p_i): Cociente de la frecuencia absoluta entre el total de observaciones

Frecuencia relativa acumulada (p_a): Suma de las frecuencias relativas menores o iguales al dato xi

Ejemplo: clase a la que va cada alumno: B, B, A, C, B, B, B, C, B, A.

x_i	n_i	n_a	p_i	p_a
A				
B				
C				

Ejemplo SPSS:

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Aprobado	23	65,7	65,7	65,7
	Suspendido	12	34,3	34,3	100,0
	Total	35	100,0	100,0	

REPRESENTACIÓN GRÀFICA

Gráfico de líneas

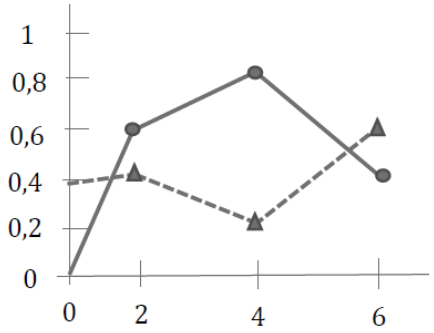


Diagrama del pastel/sectores

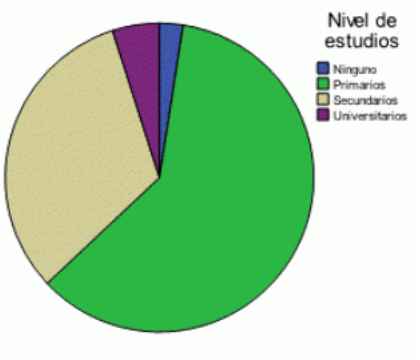
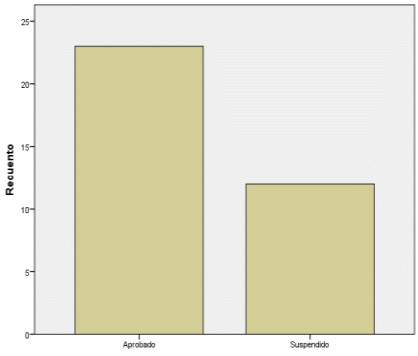
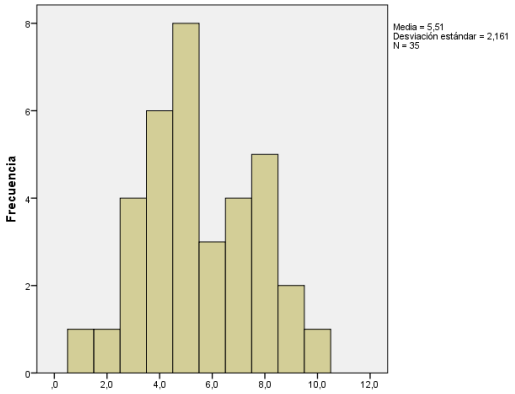


Diagrama de barras



Histograma

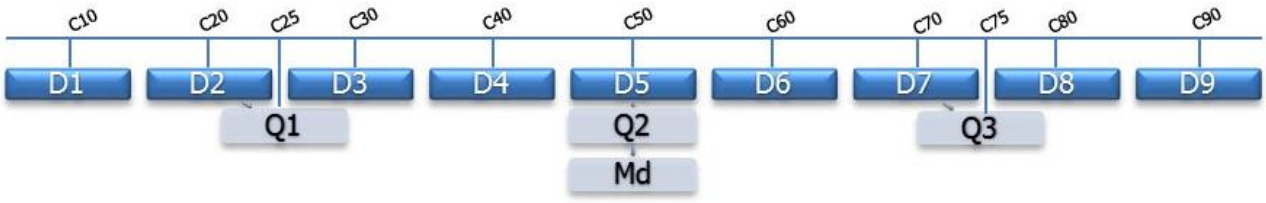


INDICADORES DE POSICIÓN O CUANTILES

Centiles: puntuación (Xi) que deja debajo de sí un porcentaje acumulado (k) de sujetos de la distribución.

Deciles: los 9 valores que dividen la distribución en 10 partes iguales.

Cuartiles: los 3 valores que dividen la distribución en 4 partes iguales.



3. ESTADÍSTICOS UNIVARIADOS

MEDIDAS DE TENDENCIA CENTRAL

Media aritmética:

$$\bar{x} = \frac{\sum x_i}{N}$$

Puntuación diferencial=puntuación-media

Mediana:

$$Pos_{Md} = \frac{n+1}{2}$$

Moda:

Ejemplo: calcula los indicadores de tendencia central con los datos 3,7,5,3,6,7,8,6,7,9

MEDIDAS DE VARIACIÓN

Varianza:

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{N} = \frac{\sum {x_i}^2}{N} - \bar{x}^2$$

Si la varianza la divdimos entre n-1 tendremos la varianza insesgada muestral o casi-varianza.

Desviación típica:

$$S = \sqrt{S^2}$$

Coeficiente de variación:

$$CV = \frac{S}{\bar{x}} \cdot 100$$

Amplitud total y amplitud semi-intercuartil:

$$A_T = X_{max} - X_{min} \qquad Q = \frac{Q3 - Q1}{2}$$

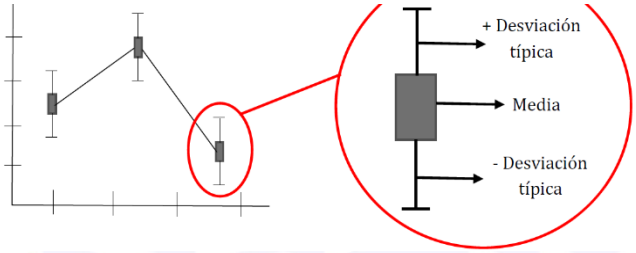
Ejemplo: calcula las diferentes medidas de dispersión de estos datos: 48-54-58-60-60.

PROPIEDADES DE MEDIA Y LA VARIANZA

- La media siempre estará entre el valor mínimo y el máximo.
- La varianza y la desviación típica siempre será mayor que 0.
- Si queremos hacer la media o la varianza total teniendo diferentes subgrupos:

$$\bar{x}_T = \frac{\sum \bar{x}_i \cdot N_i}{N} \qquad S_T^2 = \frac{\sum S_i^2 \cdot N_i}{N}$$

- Se puede representar la media y la desviación típica con el siguiente gráfico:



ASIMETRÍA

Se refiere a si los valores de la variable se concentran en una determinada zona de la distribución.

As>0 asimetría positiva

As=0 simetría

As<0 asimetría negativa

CURTOSIS

Es el grado de concentración de los valores alrededor del centro.

K>0 Leptocúrtica

K=0 Mesocúrtica

K<0 Platicúrtica

Ejemplo SPSS: Interpreta

Descriptivos

		Estadístico	Error estándar
Media		5,514	,3652
95% de intervalo de confianza para la media	Límite inferior	4,772	
	Límite superior	6,257	
Media recortada al 5%		5,516	
Mediana		5,000	
Varianza		4,669	
Desviación estándar		2,1608	
Mínimo		1,0	
Máximo		10,0	
Rango		9,0	
Rango intercuartil		3,0	
Asimetría		,137	,398
Curtosis		-,595	,778

4. TRANSFORMACIÓN DE PUNTUACIONES

MEDIA Y VARIANZA DE TRANSFORMACIONES LINEALES

	Media	Desviación típica	Varianza
Cambio de origen			
Cambio de escala			

Ejemplo: la nota media del examen de análisis ha sido de un 6 con desviación típica de 2. ¿Qué pasaría con la nota media si subiera un punto a todos los alumnos + un 10% adicional por buen comportamiento?

PUNTUACIONES TÍPICAS

$$z = \frac{x_i - \bar{x}}{S}$$

Ejemplo: un alumno del G1 ha sacado un 7 en Estadística, y otro del G2 ha sacado un 6. Sabiendo que la nota media del G1 ha sido de 6,5 con una desviación de 2 y la nota media del G2 ha sido 5 con una desviación de 3. ¿Quién tiene una mejor posición relativa?

ESCALAS DERIVADAS

$$T_i = a \cdot z_i + b$$

Escalas derivadas en Psicología:

$T = 10 \cdot z_i + 50$ $CI = 15 \cdot z_i + 100$ $S = 2 \cdot z_i + 5$

Ejemplo: obtén las puntuaciones derivadas del ejemplo anterior.

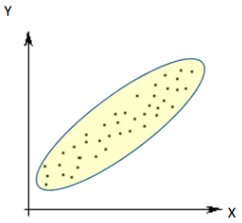
5. CORRELACIÓN LINEAL

LA RELACIÓN LINEAL

La **relación lineal** entre dos variables puede ser:

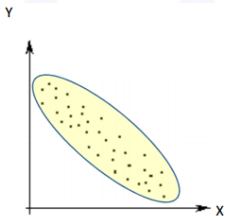
- **Positiva** (+ de una variable, + del otro o al revés), Ej : Horas de estudio (X) y nota en el examen (Y).

Horas de estudio	nota
3	2
8	4
12	7
16	9

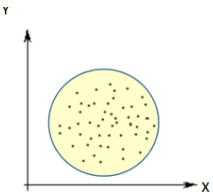


- **Negativa** (+ de una, - del otro) Ej.: ausencias a clase (X) y nota en el examen (Y).

Ausencias	nota
4	9
5	7
9	4
12	2



- **Nula.** Sin relación. Ex: altura (X) y nota (Y).



Además del análisis gráfico, podemos calcular una serie de indicadores que nos ayudarán a decidir si las variables están relacionadas (de forma lineal positiva o negativa) y con **qué intensidad.**

Covarianza: es una medida descriptiva de la relación lineal de dos variables (introducida por Pearson). En términos absolutos cuanto más grande más relación y el signo indica el sentido de la relación. Problema que presenta es la dependencia a las unidades de medida de las variables.

$$S_{xy} = \frac{\sum(x_i - \bar{x}) \cdot (y_j - \bar{y})}{N} = \frac{\sum x_i y_j}{N} - \bar{x}\bar{y}$$

Correlación (r): es una medida adimensional que resumen la relación lineal entre dos variables. Su valor oscila entre -1 y 1. El signo es el mismo que la Cov e indica el sentido de la relación. Si la relación lineal es exacta entre ambas variables los coeficiente de correlación será igual a 1 o -1. No le afectan las tranformaciones lineales a diferencia de la covarianza.

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$$

$$r_{xy} = \frac{\left(N \cdot \sum X_i \cdot Y_i\right) - \left(\sum X_i \cdot \sum Y_i\right)}{\left(\sqrt{N \cdot \sum X_i^2 - \left(\sum X_i\right)^2}\right)\left(\sqrt{N \cdot \sum Y_i^2 - \left(\sum Y_i\right)^2}\right)}$$

Propiedades de la correlación:

- Si la elevamos al cuadrado nos da la variabilidad común de X e Y
- Que sea 0 no implica que no haya relación entre X e Y, puede ser no lineal.
- Correlación no implica causalidad, ya que ambas variables pueden estar afectadas por otras variables.

Matriz de Varianzas-Covarianzas, *S*

$\boldsymbol{S}$	X	Y	$\cdots$	W
X	S_x^2	S_{xy}	$\cdots$	S_{xw}
Y		S_y^2	$\cdots$	S_{yw}
$\vdots$			$\ddots$	$\vdots$
W				S_w^2

Matriz de Correlaciones, *R*

$\boldsymbol{R}$	X	Y	$\cdots$	W
X	1	r_{xy}	$\cdots$	r_{xw}
Y		1	$\cdots$	r_{yw}
$\vdots$			$\ddots$	$\vdots$
W				1

Ejemplo: interpreta la relación entre la variable T y U y la relación entre X y W.

Matriz de Varianzas-Covarianzas, S

$\boldsymbol{S}$	X	Y	S	T	U	W
X	0,80	1,20	− 0,20	0	− 0,40	− 0,60
Y		2	− 0,60	0,20	− 0,80	− 0,60
S			3,20	− 2,40	1,60	0
T				2	− 1,20	0,60
U					0,96	− 0,08
W						2,64
Medias:	2	5	3	3	2,8	5,6

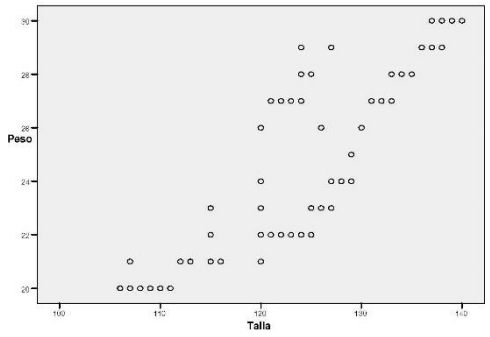
Matriz de Correlaciones, R

$\boldsymbol{R}$	X	Y	S	T	U	W
X	1,00	0,95	− 0,12	0,00	− 0,46	− 0,41
Y		1,00	− 0,24	0,10	− 0,58	− 0,26
S			1,00	− 0,95	0,91	0,00
T				1,00	− 0,87	0,26
U					1,00	− 0,05
W						1,00

Ejemplo SPSS:

Correlaciones

		Peso	Talla
Peso	Correlación de Pearson	1	,860
	Sig. (bilateral)		,000
	N	76	76
Talla	Correlación de Pearson	,860	1
	Sig. (bilateral)	,000	
	N	76	76



6. COMBINACIÓN LINEAL DE VARIABLES

DOS VARIABLES

$$Suma \qquad T_i = X_i + Y_i \qquad \overline{T} = \overline{X} + \overline{Y} \qquad S_T^2 = S_x^2 + S_y^2 + 2S_{xy}$$

$$Resta \qquad T_i = X_i - Y_i \qquad \overline{T} = \overline{X} - \overline{Y} \qquad S_T^2 = S_x^2 + S_y^2 - 2S_{xy}$$

J VARIABLES

$$Suma \quad T_i = X_{i1} + X_{i2} + X_{i3} \qquad \overline{T} = \overline{X_1} + \overline{X_2} + \overline{X_3} \qquad S_T^2 = S_1^2 + S_2^2 + +S_3^2 + 2 \cdot (S_{12} + S_{13} + S_{23})$$

$$Combinación\ lineal \qquad T_i = k_1 \cdot X_{i1} + k_2 \cdot X_{i2} \qquad \overline{T} = k_1 \cdot \overline{X_1} + k_2 \cdot \overline{X_2}$$

$$S_T^2 = k_1^2 \cdot S_1^2 + k_2^2 \cdot S_2^2 + 2 \cdot (k_1 \cdot k_2 \cdot S_{12})$$

Ejemplo: calcula las medias y las varianzas de:

a) U = X1 + X2

b) V = X2 – X3

c) W = X1 + X2 + X3

				Media:	Varianza:
X_1	1	3	5	3	2,67
X_2	4	5	3	4	0,67
X_3	2	6	7	5	4,67

$$S = \begin{matrix} X_1 \\ X_2 \\ X_3 \end{matrix} \begin{bmatrix} 2,67 & -0,67 & 3,33 \\ & 0,67 & -0,33 \\ & & 4,67 \end{bmatrix}$$

Este dosier está hecho para seguir la clase de prueba.

Si te apuntas al curso te enviaremos el dosier entero con todos los temas que faltan, ejercicios y exámenes de años anteriores

Más información en:

www.estadistix.com

Y si tienes cualquier consulta,
escríbenos un whatsapp al 644310902