Estadística Descriptiva

Grado en ADE / Economía UAM

Apuntes

1. CONCEPTO Y CONTENIDO DE LA ESTADÍSTICA

OBJETO DE LA ESTADÍSTICA

La **estadística** se preocupa de los métodos de recogida y descripción de datos, así como de generar técnicas para el análisis de la información.

TIPOS DE VARIABLES

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

Variables cualitativas/categóricas/atributos: son aquellas que recogen una característica que no se puede expresar mediante una cantidad, aunque sí con una categoría. Ej.: Sexo (hombre/mujer), color de ojos (azules/marrones/verdes...), etc. En función de la escala de medida las v. cualitativas pueden ser:

- **Nominales:** tienen un conjunto de categorías sin ningún tipo de jerarquía.
- **Ordinales:** tienen un conjunto de categorías con una jerarquía u orden.

Variables cuantitativas/numéricas: son aquellas variables que recogen como información una cantidad numérica de lo que se está observando. Ej.: Edad, peso, tensión arterial, número de hijos, etc.
En función de la naturaleza de los valores, las v. cuantitativas pueden ser:

- **Discretas:** tienen un conjunto finito de valores por ejemplo si únicamente toman números enteros.
- **Continuas:** el conjunto de posibles valores entre dos números fijos es infinito.

En función de la escala de medida, las v. cuantitativas pueden ser:

- **Intervalo:** los valores no parten de cero.
- **Razón:** los valores parten de cero.

Los datos que recogemos de las variables pueden ser:

- **Transversales o de sección cruzada:** recogidas en un mismo momento temporal. Ejemplo: Análisis de las características de los alumnos de una clase de Estadística el primer día de clase.
- **Longitudinales o de serie temporal:** Recogidas en diversos momentos temporales. Ejemplo: evolución del PIB catalán a lo largo del tiempo.
- **Datos de panel:** cuando hay una combinación de ambos tipos de datos

POBLACIÓN Y MUESTRA

Individuo: elemento bien definido que presenta la característica a estudiar. Pueden ser personas, animales, objetos o grupos. *Por ejemplo: un alumno de la clase*
Muestra: subconjunto representativo de individuos de la población extraído por algún método válido de muestreo. Intentaremos extrapolar los resultados a la población*. Por ejemplo: 20 alumnos cogidos al azar.*
Población: el conjunto de individuos sobre los que queremos hacer el estudio. *Por ejemplo: la clase entera.*

Parámetro: son los valores que caracterizan la población. Se calculan en los estudios censal.
Estadísticos: son los valores que se refieren a la muestra. Se calculan en los estudios muestrales.

Estadística descriptiva: se encarga de describir o resumir unos datos mediante tablas, gráficos o indicadores para su comprensión.

Estadística inferencial: se encarga de generalizar unos resultados muestrales a toda la población.

ETAPAS DEL ANÁLISIS ESTADÍSTICO

1.- Delimitación del objetivo del estudio: hay que definir claramente el objetivo que se persigue.

2.- Especificación de las características: deben especificarse cuáles serán las variables de interés.

3.- Identificación del colectivo: esta fase del proceso consiste en identificar, los elementos del colectivo objeto de la investigación.

4.- Recogida de datos: el resultado de medir u observar la variable en los diferentes elementos. Hay dos métodos de recogida: método indirecto (ya han sido recogidos anteriormente en algún estudio previo) y método directo (recogidos por el investigador).

5.- Ordenación de los datos: los datos deben ser revisados para detectar errores o incoherencias. Si la información es poco manejable se recurre a tablas o cuadros (tabulación).

6.- Representación gráfica: como complemento a la tabulación, se utilizan las imágenes o gráficos que hacen más rápida y accesible la información.

7.- Análisis e interpretación de resultados: aplicar los procedimientos estadísticos adecuados para interpretar de diferentes maneras los resultados.

2. DISTRIBUCIONES DE FRECUENCIAS Y REPRESENTACIONES GRÁFICAS

ESTADÍSTICA UNIDIMENSIONAL

DISTRIBUCIONES DE FRECUENCIAS

Frecuencia absoluta (n_i): número de veces que se repite un determinado valor. Puede ser individual o acumulada (N_i) (en les variables nominales no tiene sentido acumular).

Frecuencia relativa (f_i): proporción que representa las apariciones de ese valor respecto al total $f_r = \frac{n_i}{n}$

También puede ser individual o acumulada (F_i). Si se multiplica por 100 nos dará un porcentaje.

DATOS CUANTITATIVOS CONTÍNUOS AGRUPADOS EN INTERVALOS

Recorrido o rango: $Re = X_{max} - X_{min}$

Amplitud del intervalo: $c_i = L_i - L_{i-1}$

Si la amplitud es constante $Re = nº_{intervalos} \cdot c_i$ *.Se suelen hacer entre 5-15 intervalos*

Marca de clase: $x_i = \frac{L_i + L_{i-1}}{2}$

Ejemplo numérico discreto, número de hijos por familia: 2, 1, 1, 0, 1, 2, 2, 0, 1, 1.

x_i	n_i	N_i	f_i	F_i
0				
1				
2				

Ejemplo: numérico continuo, notas de un examen: 3,4; 5,3; 6,8; 9,1; 7,2; 8,3; 1,9; 5,1; 4,4; 3,2.

	x_i	n_i	N_i	f_i	F_i	c_i
[0-2)						
[2-4)						
[4-6)						
[6-8)						
[8-10)						

Ejemplo: completa la siguiente tabla de frecuencias.

x_i	n_i	f_i	N_i	F_i
0	2			
1			7	
2				0,65
3	3	0,15		
4		0,2		
Total				

REPRESENTACIONES GRÁFICAS

VARIABLES CUANTITATIVAS SIN AGRUPAR EN INTERVALOS

Diagrama de barras

Diagramas acumulativos

VARIABLES CUANTITATIVAS AGRUPADAS EN INTERVALOS

Histograma (área n_i) $d_i=n_i/c_i$

Polígonos acumulativos

VARIABLES CUALITATIVAS/ATRIBUTOS

Diagrama de barras

Diagrama de sectores

Diagrama de rectángulos

Pictograma

TEST TEMA 1 y 2

1. De la Encuesta de Presupuestos Familiares se extraen los siguientes datos sobre las personas que integran las familias españolas: (N) Número de personas que integran el núcleo familiar; (E) Estudios; (P) Profesión; (C) Estado Civil; (I) Ingresos Anuales. ¿Cuáles de estos caracteres se consideran atributos?

A. E, P e I.

B. N e I.

C. E, P y C.

2. Señale la afirmación correcta en relación con las distribuciones de frecuencias:

A. A partir de la distribución de frecuencias de un atributo es posible calcular su media aritmética.

B. Las frecuencias relativas acumuladas nos indican el número de observaciones menores o iguales que un valor determinado.

C. Las frecuencias relativas nos dicen la proporción de observaciones correspondientes a cada valor o modalidad.

3. Dada una distribución de frecuencias agrupada en intervalos…

A. …la marca de clase es el valor medio del intervalo.

B. …el valor representativo de cada intervalo se denomina moda.

C. …el intervalo modal siempre es único.

4. A partir de la siguiente distribución se puede afirmar que…

x_i	f_i	N_i
0	0,5	
1	0,2	7
2	0,3	

A. …existe un total de 10 observaciones.

B. …la n=35.

C. …el valor 1 se repite 7 veces.

5. En la siguiente distribución de frecuencias, la frecuencia absoluta del 2 es:

x_i	n_i	f_i	N_i	F_i
0	2			
1			7	
2				0,65
3	3	0,15		
4		0,2		
Total				

A. 4

B. 5

C. 6

6. Se pretende realizar una representación gráfica para describir la situación de los distintos tipos de contratos temporales en la Comunidad de Madrid (en prácticas, eventual por circunstancias de la producción, por obra,…). ¿Cuál de los siguientes gráficos no sería adecuado?

A. Histograma.

B. Diagrama de sectores.

C. Diagrama de rectángulos.

7. Señale la afirmación falsa en relación con los histogramas:

A. Las alturas (de los rectángulos) de un histograma en el que los intervalos son de la misma amplitud se deben representar con frecuencias absolutas o con densidades de frecuencia.

B. El rectángulo más alto corresponde siempre al intervalo de mayor densidad de frecuencia y, por tanto, al de mayor frecuencia absoluta.

C. Las alturas (de los rectángulos) de un histograma en el que los intervalos son de distinta amplitud deberían representar densidades de frecuencia.

1. CONCEPTO Y CONTENIDO DE LA ESTADÍSTICA

OBJETO DE LA ESTADÍSTICA

La **estadística** se preocupa de los métodos de recogida y descripción de datos, así como de generar técnicas para el análisis de la información.

TIPOS DE VARIABLES

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

Variables cualitativas/categóricas/atributos: son aquellas que recogen una característica que no se puede expresar mediante una cantidad, aunque sí con una categoría. Ej.: Sexo (hombre/mujer), color de ojos (azules/marrones/verdes...), etc. En función de la escala de medida las v. cualitativas pueden ser:

- **Nominales:** tienen un conjunto de categorías sin ningún tipo de jerarquía.
- **Ordinales:** tienen un conjunto de categorías con una jerarquía u orden.

Variables cuantitativas/numéricas: son aquellas variables que recogen como información una cantidad numérica de lo que se está observando. Ej.: Edad, peso, tensión arterial, número de hijos, etc.
En función de la naturaleza de los valores, las v. cuantitativas pueden ser:

- **Discretas:** tienen un conjunto finito de valores por ejemplo si únicamente toman números enteros.
- **Continuas:** el conjunto de posibles valores entre dos números fijos es infinito.

En función de la escala de medida, las v. cuantitativas pueden ser:

- **Intervalo:** los valores no parten de cero.
- **Razón:** los valores parten de cero.

Los datos que recogemos de las variables pueden ser:

- **Transversales o de sección cruzada:** recogidas en un mismo momento temporal. Ejemplo: Análisis de las características de los alumnos de una clase de Estadística el primer día de clase.
- **Longitudinales o de serie temporal:** Recogidas en diversos momentos temporales. Ejemplo: evolución del PIB catalán a lo largo del tiempo.
- **Datos de panel:** cuando hay una combinación de ambos tipos de datos

POBLACIÓN Y MUESTRA

Individuo: elemento bien definido que presenta la característica a estudiar. Pueden ser personas, animales, objetos o grupos. *Por ejemplo: un alumno de la clase*
Muestra: subconjunto representativo de individuos de la población extraído por algún método válido de muestreo. Intentaremos extrapolar los resultados a la población*. Por ejemplo: 20 alumnos cogidos al azar.*
Población: el conjunto de individuos sobre los que queremos hacer el estudio. *Por ejemplo: la clase entera.*

Parámetro: son los valores que caracterizan la población. Se calculan en los estudios censal.
Estadísticos: son los valores que se refieren a la muestra. Se calculan en los estudios muestrales.

Estadística descriptiva: se encarga de describir o resumir unos datos mediante tablas, gráficos o indicadores para su comprensión.

Estadística inferencial: se encarga de generalizar unos resultados muestrales a toda la población.

ETAPAS DEL ANÁLISIS ESTADÍSTICO

1.- Delimitación del objetivo del estudio: hay que definir claramente el objetivo que se persigue.

2.- Especificación de las características: deben especificarse cuáles serán las variables de interés.

3.- Identificación del colectivo: esta fase del proceso consiste en identificar, los elementos del colectivo objeto de la investigación.

4.- Recogida de datos: el resultado de medir u observar la variable en los diferentes elementos. Hay dos métodos de recogida: método indirecto (ya han sido recogidos anteriormente en algún estudio previo) y método directo (recogidos por el investigador).

5.- Ordenación de los datos: los datos deben ser revisados para detectar errores o incoherencias. Si la información es poco manejable se recurre a tablas o cuadros (tabulación).

6.- Representación gráfica: como complemento a la tabulación, se utilizan las imágenes o gráficos que hacen más rápida y accesible la información.

7.- Análisis e interpretación de resultados: aplicar los procedimientos estadísticos adecuados para interpretar de diferentes maneras los resultados.

2. DISTRIBUCIONES DE FRECUENCIAS Y REPRESENTACIONES GRÁFICAS

ESTADÍSTICA UNIDIMENSIONAL

DISTRIBUCIONES DE FRECUENCIAS

Frecuencia absoluta (n_i): número de veces que se repite un determinado valor. Puede ser individual o acumulada (N_i) (en les variables nominales no tiene sentido acumular).

Frecuencia relativa (f_i): proporción que representa las apariciones de ese valor respecto al total $f_r = \frac{n_i}{n}$

También puede ser individual o acumulada (F_i). Si se multiplica por 100 nos dará un porcentaje.

DATOS CUANTITATIVOS CONTÍNUOS AGRUPADOS EN INTERVALOS

Recorrido o rango: $Re = X_{max} - X_{min}$

Amplitud del intervalo: $c_i = L_i - L_{i-1}$

Si la amplitud es constante $Re = nº_{intervalos} \cdot c_i$ *.Se suelen hacer entre 5-15 intervalos*

Marca de clase: $x_i = \frac{L_i + L_{i-1}}{2}$

Ejemplo numérico discreto, número de hijos por familia: 2, 1, 1, 0, 1, 2, 2, 0, 1, 1.

x_i	n_i	N_i	f_i	F_i
0				
1				
2				

Ejemplo: numérico continuo, notas de un examen: 3,4; 5,3; 6,8; 9,1; 7,2; 8,3; 1,9; 5,1; 4,4; 3,2.

	x_i	n_i	N_i	f_i	F_i	c_i
[0-2)						
[2-4)						
[4-6)						
[6-8)						
[8-10)						

Ejemplo: completa la siguiente tabla de frecuencias.

x_i	n_i	f_i	N_i	F_i
0	2			
1			7	
2				0,65
3	3	0,15		
4		0,2		
Total				

REPRESENTACIONES GRÁFICAS

VARIABLES CUANTITATIVAS SIN AGRUPAR EN INTERVALOS

Diagrama de barras

Diagramas acumulativos

VARIABLES CUANTITATIVAS AGRUPADAS EN INTERVALOS

Histograma (área n_i) $d_i=n_i/c_i$

Polígonos acumulativos

VARIABLES CUALITATIVAS/ATRIBUTOS

Diagrama de barras

Diagrama de sectores

Diagrama de rectángulos

Pictograma

TEST TEMA 1 y 2

1. De la Encuesta de Presupuestos Familiares se extraen los siguientes datos sobre las personas que integran las familias españolas: (N) Número de personas que integran el núcleo familiar; (E) Estudios; (P) Profesión; (C) Estado Civil; (I) Ingresos Anuales. ¿Cuáles de estos caracteres se consideran atributos?

A. E, P e I.

B. N e I.

C. E, P y C.

2. Señale la afirmación correcta en relación con las distribuciones de frecuencias:

A. A partir de la distribución de frecuencias de un atributo es posible calcular su media aritmética.

B. Las frecuencias relativas acumuladas nos indican el número de observaciones menores o iguales que un valor determinado.

C. Las frecuencias relativas nos dicen la proporción de observaciones correspondientes a cada valor o modalidad.

3. Dada una distribución de frecuencias agrupada en intervalos…

A. …la marca de clase es el valor medio del intervalo.

B. …el valor representativo de cada intervalo se denomina moda.

C. …el intervalo modal siempre es único.

4. A partir de la siguiente distribución se puede afirmar que…

A. …existe un total de 10 observaciones.

B. …la n=35.

C. …el valor 1 se repite 7 veces.

x_i	f_i	N_i
0	0,5	
1	0,2	7
2	0,3	

5. En la siguiente distribución de frecuencias, la frecuencia absoluta del 2 es:

A. 4

B. 5

C. 6

x_i	n_i	f_i	N_i	F_i
0	2			
1			7	
2				0,65
3	3	0,15		
4		0,2		
Total				

6. Se pretende realizar una representación gráfica para describir la situación de los distintos tipos de contratos temporales en la Comunidad de Madrid (en prácticas, eventual por circunstancias de la producción, por obra,…). ¿Cuál de los siguientes gráficos no sería adecuado?

A. Histograma.

B. Diagrama de sectores.

C. Diagrama de rectángulos.

7. Señale la afirmación falsa en relación con los histogramas:

A. Las alturas (de los rectángulos) de un histograma en el que los intervalos son de la misma amplitud se deben representar con frecuencias absolutas o con densidades de frecuencia.

B. El rectángulo más alto corresponde siempre al intervalo de mayor densidad de frecuencia y, por tanto, al de mayor frecuencia absoluta.

C. Las alturas (de los rectángulos) de un histograma en el que los intervalos son de distinta amplitud deberían representar densidades de frecuencia.

3. MEDIDAS DE POSICIÓN, DISPERSIÓN, FORMA Y DESIGUALDAD

MEDIDAS DE POSICIÓN. MOMENTOS UNIDIMENSIONALES

MEDIDAS DE POSICIÓN CENTRAL

Media aritmética: es el centro de gravedad de los datos.

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

Problemas: le afectan mucho los valores extremos

Media aritmética ponderada: sirve para hacer una media cuando los elementos no tienen el mismo peso.

$$\bar{x}_p = \frac{\sum x_i \cdot w_i}{\sum w_i}$$

Media geométrica: se utiliza para promediar porcentajes, tasas, números índices, etc, casos en que la variable presenta variaciones acumulativas, que evolucionan exponencialmente.

$$G = \sqrt{\prod x_i^{n_i}}$$

Por ejemplo, un dinero invertido durante n años a un interés anual i_1, i_2, …, i_n. Si deseamos saber el interés medio del periodo, no debemos calcular la media aritmética de los n tantos de interés anuales, pues nos daría un valor erróneo. Haremos:

$$i = \sqrt[n]{(1+i_1)\cdot(1+i_2)\ldots(1+i_n)} = G - 1$$

Problemas: Es de significado estadístico menos intuitivo que la media aritmética, su cálculo es más difícil y en ocasiones no queda determinada (si algún valor x es 0)

Mediana: es el valor de la distribución, tal que una vez ordenados todos los valores observados de menor a mayor (en sentido creciente), ocupa el lugar central.

$$Pos_{Md} = \frac{N+1}{2} \qquad Me = L_{i-1} + c_i \cdot \frac{N/2 - N_{i-1}}{n_i}$$

Moda: es el valor que más se repite.

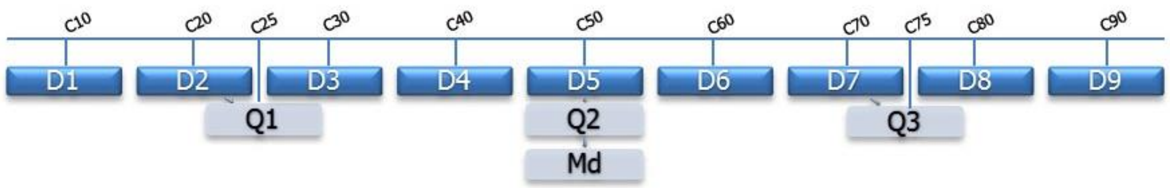
$$Mo_{(si\ c_i =)} = L_{i-1} + c_i \cdot \frac{n_{i+1}}{n_{i-1} + n_{i+1}}$$

MEDIDAS DE POSICIÓN NO CENTRAL

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

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

Centiles o percentiles (P): los 99 valores que dividen la distribución en 100 partes iguales.



$$Q_{r/q} = L_{i-1} + c_i \frac{(r \cdot n)/k - N_{i-1}}{n_i}$$

MOMENTOS DE UNA DISTRIBUCIÓN DE FRECUENCIAS

Los momentos son medidas que caracterizan a una distribución de frecuencias. Dos distribuciones son iguales si tienen todos sus momentos iguales.

Momentos respecto al origen:

$$a_r = \frac{\sum x_i^r \cdot n_i}{N} \qquad \boldsymbol{a_1} = \boldsymbol{\bar{x}} = \frac{\sum x_i^1 \cdot n_i}{N} \qquad a_2 = \frac{\sum x_i^2 \cdot n_i}{N} \;\ldots$$

Momentos respecto a la media:

$$m_r = \frac{\sum (x_i - \bar{x})^r \cdot n_i}{N} \qquad m_1 = \frac{\sum (x_i - \bar{x})^1 \cdot n_i}{N} \qquad \boldsymbol{m_2} = \boldsymbol{S^2} = \frac{\sum (x_i - \bar{x})^2 \cdot n_i}{N} = m_2 = a_2 - {a_1}^2 \quad \ldots$$

Ejemplo: calcula todas las medidas de posición y los momentos con la siguiente tabla de frecuencias.

x_i	X_i	n_i	N_i	f_i	F_i	a_i
[0-2)	1	1	1	0,1	0,1	2
[2-4)	3	2	3	0,2	0,3	2
[4-6)	5	3	6	0,3	0,6	2
[6-8)	*7*	*2*	*8*	*0,2*	*0,8*	2
[8-10)	*9*	*2*	*10*	*0,2*	*1*	2

MEDIDAS DE DISPERSIÓN

MEDIDAS DE DISPERSIÓN ABSOLUTAS

Recorrido o Amplitud (Re): es la diferencia entre el máximo y el mínimo.

$$Re = X_{max} - X_{min} = X_n - X_1$$

Rango intercuartílico (Rq): es la diferencia entre el tercer cuartil y el primero.

$$Re_I = C_3 - C_1$$

Desviación media respecto a la media ($D_{\bar{x}}$): es la media aritmética de los valores absolutos de las diferencias entre los valores de la variable y la media aritmética.

$$D_{\bar{x}} = \frac{\sum |x_i - \bar{x}| \cdot n_i}{N}$$

Desviación media respecto a la mediana (D_{Me}): es la media aritmética de las desviaciones en valor absoluto de los valores de la variable a la mediana.

$$D_{Me} = \frac{\sum |x_i - Me| \cdot n_i}{N}$$

Varianza (S^2): es la media aritmética de los cuadrados de las diferencias de los valores de la variable con respecto a su media aritmética y coincide con el momento de segundo orden respecto a la media.

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

La desviación estándar/típica (S): es la raíz de la varianza.

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

Siempre se verifica que $D_{Me} \leq D_{\bar{x}} \leq S$

En distribuciones simétricas: $\bar{x} \pm S \rightarrow 68\%$; $\bar{x} \pm 2S \rightarrow 95\%$; $\bar{x} \pm 3S \rightarrow 99{,}7\%$

MEDIDAS DE DISPERSIÓN RELATIVAS

Coeficiente de apertura (C_{Ap}): es el cociente entre los 2 valores extremos de una distribución.

$$C_{Ap} = \frac{X_n}{X_1}$$

Recorrido relativo (R_r): es el cociente entre el recorrido de la variable y su media aritmética.

$$R_r = \frac{Re}{\bar{x}}$$

Recorrido semi-intercuartílico (Re_{SI}): es el cociente entre el recorrido intercuartílico y la suma del primer y tercer cuartil.

$$Re_{SI} = \frac{C_3 - C_1}{C_3 + C_1}$$

Coeficiente de variación de Pearson (CV): Es el más empleado de los índices de dispersión relativa y se define como el cociente entre la desviación típica y la media aritmética.

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

A mayor coeficiente de variación, menor representatividad de la media.

Ejemplo: calcula todas las medidas de dispersión con la siguiente tabla de frecuencias.

x_i	X_i	n_i	N_i	f_i	F_i	a_i
[0-2)	1	1	1	0,1	0,1	2
[2-4)	3	2	3	0,2	0,3	2
[4-6)	5	3	6	0,3	0,6	2
[6-8)	7	2	8	0,2	0,8	2
[8-10)	9	2	10	0,2	1	2

TRANFORMACIONES

	Media	Moda	Desviación típica	Varianza
Cambio de origen				
Cambio de escala				

Ejemplo: en el presente ejercicio se decide aumentar el precio de un producto en 5€. Después, debito a un incremento impositivo, se aumenta un 3% adicional sobre el nuevo precio. Si durante el anterior el precio medio era de 60 euros, con una desviación de 4, ¿cuáles serán el nuevo precio medio, varianza y desviación típica?

ESTANDARIZACIÓN

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

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?

MEDIDAS DE FORMA

ASIMETRÍA

As>0 asimetría positiva

As=0 simetría

As<0 asimetría negativa

Coeficientes de Asimetría:

$$As_{Fisher} = g_1 = \frac{m_3}{S^3} \qquad As_{Pearson} = \frac{\bar{x} - Mo}{S}$$

CURTOSIS

Las medidas de curtosis, apuntamiento o concentración central tratan de estudiar la distribución de frecuencias en la zona central de la distribución. Se aplican únicamente a distribuciones **campaniformes**, es decir unimodales aproximadamente simétricas.

C>0 leptocúrtica

C=0 mesocúrtica

C<0 platicúrtica

Coeficientes de curtosis:

$$C_c = g_2 = \frac{m_4}{S^4} - 3$$

MEDIDAS DE DESIGUALDAD/CONCENTRACIÓN

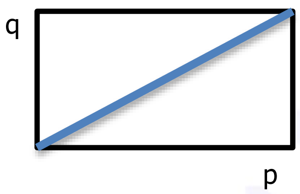
Tratan de poner de relieve el grado de igualdad en el reparto del total de los valores de la variable.

ÍNDICE DE GINI

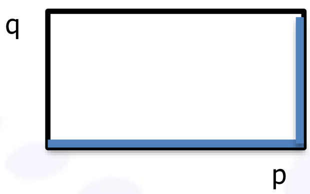
$$I_{Gini} = \frac{\sum(p_i - q_i)}{\sum p_i} \qquad \begin{cases} p\text{: porcentaje de los datos} \\ q\text{: cantidad de la variable del porcentaje } p \\ 0 = \text{equidistribución y } 1 = \text{máxima desigualdad / concentración} \end{cases}$$

CURVA DE LORENZ

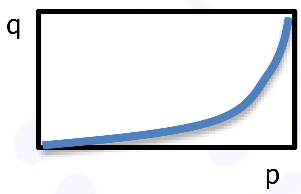
Equidistribución



Máxima concentración



Ejemplo:



Ejemplo: calcula el índice de Gini y la curva de Lorenz para la variable "nota" teniendo en cuenta que:

- El 10% con peores notas de la clase se han llevado el 6% de los puntos en el examen final.

- El 30% con peores notas se han llevado el 21% de los puntos.

- El 60% con peores notas se han llevado el 46% de los puntos.

- El 85% con peores notas se han llevado el 80% de los puntos.

Ejemplo: calcula el índice de Gini para la variable renta.

P (habitantes)	Q (renta)
0,1	0,01
0,3	0,07
0,5	0,21
0,7	0,43
0,9	0,75

Transformaciones:

TEST TEMA 3

1. Dada la siguiente distribución de frecuencias, señale la afirmación correcta:

x_i	n_i	f_i	N_i
2	10	0,25	
3			
4	15	0,375	30
5		0,25	

A. La media aritmética es igual a 3,625.

B. La moda es igual a 3.

C. La mediana es igual a 3.

2. Tenemos información correspondiente al número de horas diarias trabajadas por una persona contratada durante doce días en una determinada empresa: 8, 8, 7, 8, 7, 9, 9, 5, 6, 7, 8, 8.

A. El valor de la mediana es 9.

B. El valor de la mediana es 8 y el del tercer cuartil también.

C. El valor de la mediana es 9 y el del tercer cuartil 7.

3. Se ha calculado el percentil 85 sobre las estadísticas de siniestralidad laboral en el sector de la construcción durante el último año y se ha obtenido el valor 2,5. El significado de este dato es…

A. … el 2,5 por ciento de los trabajadores en el sector de la construcción sufren menos de 85 accidentes al año.

B. … el 15 por ciento de los trabajadores en el sector de la construcción sufren menos de 2,5 accidentes al año.

C. … el 85 por ciento de los trabajadores en el sector de la construcción sufren menos de 2,5 accidentes al año.

4. La suma de las desviaciones de cada valor de la variable respecto a la media aritmética ponderada por su frecuencia relativa…

A. …vale cero.

B. …es la desviación absoluta media.

C. …toma una valor positivo si la distribución es asimétrica a la derecha.

5. Señale la afirmación falsa en relación con las medidas de dispersión:

A. Las medidas de dispersión absolutas no se ven afectadas por cambios en las unidades de medida de la variable.

B. Las medidas de dispersión cuantifican la variabilidad de los valores que ha tomado una variable.

C. A diferencia de la varianza, la desviación típica se expresa en las mismas unidades de medida que la variable.

6. Si dos distribuciones con valores positivos tienen igual desviación típica, con medias distintas…

A. …la media menor es menos representativa de su distribución.

B. …la media mayor es menos representativa de su distribución.

C. …ambas tienen igual dispersión

7. Si dos distribuciones de frecuencias tienen el mismo coeficiente de variación, esto significa que…

A. …tienen la misma desviación típica.

B. …tienen la misma media.

C. …ambas medias son igualmente representativas de sus respectivas distribuciones.

8. Señale qué afirmación es falsa:

A. El coeficiente de variación de Pearson es una medida de dispersión relativa, y sirve para comparar la representatividad de las medias aritméticas de dos distribuciones.

B. La varianza es una medida de dispersión absoluta, y sirve para comparar las medias aritméticas en dos distribuciones.

C. La desviación absoluta media respecto a la media es una medida de dispersión absoluta, y se define como la media aritmética de las desviaciones absolutas entre los valores de la variable y su media.

9. Una variable estadística X tiene media 15, varianza 4 y moda 12. Definida la variable Y = 3X – 5, señale la afirmación correcta respecto a dicha variable:

A. Tiene media 40, desviación típica 1 y moda 36.

B. Tiene media 45, varianza 12 y moda 36.

C. Tiene media 40, desviación típica 6 y moda 31.

10. Si X media = 2 y $Sx^2 = 9$, y definimos Y = 2X + 1, entonces:

A. SY = 7.

B. Cv (X) = Cv (Y)

C. Cv (Y) = 1,2.

11. Se desea comparar el nivel académico de dos estudiantes de la misma titulación de diferentes universidades para la concesión de una beca de postgrado. Se le concederá…

A. … al que tenga una nota media más alta.

B. … al que tenga menor dispersión en las notas de su expediente.

C. … al que tenga la nota media tipificada más alta.

12. El coeficiente de variación de Pearson…

A. …es una medida de dispersión que varía entre 0 y 1.

B. …es una medida de concentración relativa.

C. …es invariante ante la transformación Y = aX con a > 0.

13. Ante un cambio de origen y escala, el coeficiente de asimetría de Fisher…

A. …sólo se ve afectado por el cambio de escala.

B. …se ve afectado por ambos.

C. …permanece invariante.

14. Ante un cambio de origen y escala, el coeficiente de apuntamiento…

A. …sólo se ve afectado por el cambio de escala.

B. …se ve afectado por ambos.

C. …permanece invariante.

15. Sean g_1 y g_2 el coeficiente de asimetría de Fisher y el coeficiente de apuntamiento, respectivamente. Si calculados ambos en una distribución de frecuencias encontramos que $g_1 = 0,5$ y $g_2 = -2$, entonces…

A. …la distribución es asimétrica positiva y platicúrtica.

B. …la distribución es asimétrica positiva y leptocúrtica.

C. …la distribución es asimétrica negativa y leptocúrtica.

16. Si para una distribución de frecuencias se verifica que $g_2 < -3$, entonces…

A. …la distribución es platicúrtica.

B. …la distribución es leptocúrtica.

C. …el coeficiente g2 no puede tomar valores menores que -3.

17. De una distribución de frecuencias se conoce que la varianza es 10 y el momento central de orden 4 vale 200. Por tanto,…

A. …la distribución es platicúrtica.

B. …la distribución es mesocúrtica.

C. …la distribución es leptocúrtica.

18. Dada una distribución de salarios…

A. …se dice que existe concentración mínima o equidistribución cuando todos los asalariados perciben la misma cantidad y, por tanto, el índice de Gini es igual a uno y la curva de Lorenz coincide con la bisectriz del primer cuadrante.

B. …se dice que existe concentración máxima cuando todos los salariados perciben la misma cantidad y, por tanto, el índice de Gini es igual a cero y la curva de Lorenz coincide con la bisectriz del primer cuadrante.

C. …se dice que existe concentración mínima o equidistribución cuando todos los asalariados perciben la misma cantidad y, por tanto, el índice de Gini es igual a cero y la curva de Lorenz coincide con la bisectriz del primer cuadrante.

19. Una empresa ha decidido subir todos los salarios de sus empleados de manera que el año próximo cobrarán 50 € más cada mes. El índice de Gini:

A. Aumenta

B. Disminuye

C. No varía

20. Una empresa ha decidido subir todos los salarios de sus empleados proporcionalmente a su salario actual y el año próximo pasarán a cobrar un 3 % más cada mes. El índice de Gini.

A. Aumenta

B. Disminuye

C. No varía

21. Una empresa ha decidido subir todos los salarios de sus empleados proporcionalmente a su salario actual y el año próximo pasarán a cobrar un 3 % más cada mes. La curva de Lorenz:

A. Se acerca más a la diagonal principal

B. Se aleja de la diagonal principal

C. No varía.

4. DISTRIBUCIÓN DE FRECUENCIAS BIDIMENSIONAL

DISTRIBUCIÓN BIDIMENSIONAL DE FRECUENCIAS

Tabla de correlación: es para variables cuantitativas

Tabla de contingencia: es para variables cualitativas

	Suj. 1	Suj. 2	Suj. 3	Suj. 4	Suj. 5	Suj. 6	Suj. 7	Suj.8	Suj. 9	Suj.10
Hijos (x)	1	2	0	3	1	3	2	0	1	2
Coches (y)	0	1	1	1	2	0	2	1	1	1

DISTRIBUCIONES CONJUNTAS

Frecuencias absolutas conjuntas:

x\y	0	1	2	
0	0	2	0	2
1	1	1	1	3
2	0	2	1	3
3	1	1	0	2
	2	6	2	10

Frecuencias relativas conjuntas:

x\y	0	1	2	
0	0	0,2	0	0,2
1	0,1	0,1	0,1	0,3
2	0	0,2	0,1	0,3
3	0,1	0,1	0	0,2
	0,2	0,6	0,2	1

DISTRIBUCIONES MARGINALES

x	ni
0	2
1	3
2	3
3	2
	10

y	ni
0	
1	
2	

DISTRIBUCIONES CONDICIONADAS

$X_i/(Y=2)$	ni	fi
0	0	0
1	1	0,5
2	1	0,5
3	0	0
	2	1

$Y_i/(X=0)$	ni	fi
0		
1		
2		

...

INDEPENDENCIA ESTADÍSTICA

Decimos que dos variables Xe Y son independientes estadísticamente si la frecuencia relativa conjunta es el producto de las frecuencias relativas marginales.

$$\frac{n_{ij}}{N} = \frac{n_i}{N} \cdot \frac{n_j}{N}$$

También decimos que dos variables son independientes si las frecuencias relativas condicionadas son iguales a las frecuencias relativas marginales:

$$f_{i/j} = \frac{n_{ij}}{n_{i.}} = \frac{n_{.j}}{N}$$

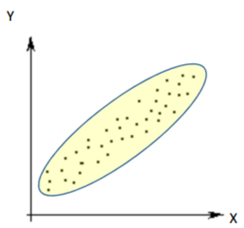
REPRESENTACIONES GRÁFICAS

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).

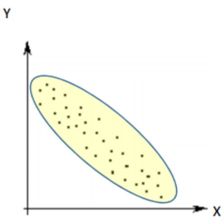
Horas de estudio	nota
3	2
8	4
12	7
16	9

-

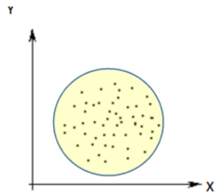


- **Negativa** (+ de una, - del otro) Ej.: ausencias a clase (X) y nota en el examen (Y).

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.**

MOMENTOS BIDIMENSIONALES. COVARIANZA

Momentos respecto al origen:

$$a_{rs} = \frac{\sum\sum x_i^r \cdot y_j^s \cdot n_{ij}}{N} \qquad a_{10} = \frac{\sum\sum x_i^1 \cdot y_j^0 \cdot n_{ij}}{N} = \boldsymbol{\bar{x}} \qquad a_{01} = \frac{\sum\sum x_i^0 \cdot y_j^1 \cdot n_{ij}}{N} = \boldsymbol{\bar{y}} \ \ldots$$

Momentos respecto a la media:

$$m_{rs} = \frac{\sum\sum (x_i - \bar{x})^r \cdot (y_i - \bar{y})^s \cdot n_{ij}}{N}$$

$$Orden\ 1: \quad m_{10} = \frac{\sum\sum (x_i - \bar{x})^1 \cdot (y_i - \bar{y})^0 \cdot n_{ij}}{N} = 0 \qquad m_{01} = \frac{\sum\sum (x_i - \bar{x})^0 \cdot (y_i - \bar{y})^1 \cdot n_{ij}}{N} = 0$$

$$Orden\ 2: \quad m_{20} = \frac{\sum\sum (x_i - \bar{x})^2 \cdot (y_i - \bar{y})^0 \cdot n_{ij}}{N} = \boldsymbol{S_x^2} \qquad m_{02} = \frac{\sum\sum (x_i - \bar{x})^0 \cdot (y_i - \bar{y})^2 \cdot n_{ij}}{N} = \boldsymbol{S_y^2}$$

$$m_{11} = \frac{\sum\sum (x_i - \bar{x})^1 \cdot (y_i - \bar{y})^1 \cdot n_{ij}}{N} = \boldsymbol{S_{xy}}$$

La **covarianza** o momento respecto a la media de segundo orden r=1 s=1 (S_{xy})

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

Ejemplo: calcula la covarianza con los siguientes datos:

	Suj. 1	Suj. 2	Suj. 3	Suj. 4	Suj. 5	Suj. 6	Suj. 7	Suj.8	Suj. 9	Suj.10
x	1	2	0	3	1	3	2	0	1	2
y	0	1	1	1	2	0	2	1	1	1

x\y	0	1	2	
0	0	2	0	2
1	1	1	1	3
2	0	2	1	3
3	1	1	0	2
	2	6	2	10

TEST TEMA 4

1.- El fin de considerar simultáneamente dos caracteres de un mismo colectivo es:

a) ahorrarse tiempo de cálculo

b) estudiar las posibles relaciones entre ellos

c) no tiene ningún fin específico

2.- Cuando no existe relación entre dos variables, se dice que las variables son:

a) de baja correlación

b) altamente relacionadas

c) independientes

3.- Cuando la relación entre dos variables es perfecta, se dice que:

a) las variables están relacionadas funcionalmente, su relación puede ser expresada bajo la forma y=f(x)

b) las variables tienen algún tipo de relación, que puede ser expresada bajo la forma y=f(x)+e, siendo e un residual

c) la relación entre ellas es inexistente

4.- En la siguiente tabla de correlación, n_{23} nos dice:

a) el número de veces que se presenta conjuntamente un 2 y un 3

b) el número de veces que se presenta x_3 con y_2

c) el número de veces que se presenta x_2 con y_3

X \ Y	y_1	y_2	y_3	$n_{i.}$
x_1	n_{11}	n_{12}	n_{13}	$n_{1.}$
x_2	n_{21}	n_{22}	n_{23}	$n_{2.}$
x_3	n_{31}	n_{32}	n_{33}	$n_{3.}$
n.j	$n_{.1}$	$n_{.2}$	$n_{.3}$	N

5.- En la misma tabla anterior, la frecuencia marginal de y_3 es:

a) $n_{.3}$

b) $n_{3.}$

c) $n_{.3}/N$

6.- Dos variables, X e Y, son independientes estadísticamente sí:

a) la frecuencia relativa conjunta es igual al cociente de las frecuencias relativas marginales

b) todas las frecuencias relativas condicionadas son iguales a sus correspondientes frecuencias relativas marginales

c) ninguna de las dos anteriores

7.- Son siempre nulos:

a) los momentos bidimensionales respecto a la media de orden 1

b) los momentos bidimensionales respecto al origen de orden 1

c) los momentos bidimensionales respecto al origen de orden 2

8.- Escoja la opción correcta:

a) si X e Y son independientes, el valor de su covarianza es distinto de cero

b) si la covarianza entre X e Y es cero, las variables son independientes siempre

c) si X e Y son independientes, su covarianza es cero, pero el recíproco no siempre es cierto

9.- Calcule la covarianza de las variables X e Y de la siguiente tabla de correlación:

X \ Y	5	8
10	4	2
20	6	3

a) la covarianza es cero pero las variables no son independientes

b) la covarianza es cero y las variables son estadísticamente independientes

c) la covarianza es 55, por lo que las variables son dependientes

10.- Calcule la covarianza de las variables X e Y de la siguiente tabla de correlación:

X \ Y	5	10
100	8	4
200	2	6

a) la covarianza es cero pero las variables no son independientes

b) la covarianza es 50, por lo que las variables son dependientes

c) la covarianza es -25, por lo que las variables son dependientes

Este dosier está hecho para seguir la clase de prueba.

Si te apuntas al curso te enviaremos por correo 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