

INTRODUCCIÓN A LA ESTADÍSTICA. ENERO 2009 Primera semana. Código de carrera 43. Código de asignatura 203.

PRIMERA PARTE: PREGUNTAS TEÓRICAS

1.- Describa los principales tipos de muestreo y sus características.

Respuesta.-

Los muestreos probabilísticos se caracterizan porque las unidades muestrales se seleccionan al azar, mediante un procedimiento aleatorio. Entre los principales tenemos:

Muestreo aleatorio simple con o sin reposición. La muestra se obtiene elemento a elemento y todos los elementos de la población tienen la misma probabilidad de ser seleccionados.

Muestreo estratificado. La población se agrupa en estratos con características similares. La probabilidad de que un individuo sea seleccionado dependerá del estrato al que pertenezca.

Muestreo por conglomerados. Dividimos la población en grupos heterogéneos de individuos (conglomerados), normalmente áreas geográficas. Únicamente se estudia una muestra de conglomerados. Dentro del conglomerado se analizan todas las unidades del mismo.

Muestreo sistemático. En una población con N individuos y un tamaño muestral de n, todas las unidades tienen la misma probabilidad de ser seleccionadas. Se ordenan aleatoriamente y se obtiene aleatoriamente un número del 1 al n (arranque aleatorio); a partir del arranque aleatorio se va eligiendo un individuo cada $\frac{N}{n}$ lugares.

Muestreo polietápico. Le llamaremos *bietápico* cuando en una primera etapa seleccionamos los conglomerados y en una segunda etapa elegimos aleatoriamente una muestra dentro de los conglomerados. En general polietápico cuando el proceso anterior se generaliza a n etapas.

Muestreos no probabilísticos en los que la muestra no se extrae por procedimientos aleatorios. Un tipo de muestreo no probabilístico es por ejemplo el **muestreo por cuotas**, en el que se le dan a los entrevistadores unas cuotas, de edad o sexo de los individuos o de tamaño y grupo de actividad de las empresas, dejándole mayor o menor grado de libertad para que seleccionen las unidades muestrales en campo.

2.- Ventajas e inconvenientes de la mediana. En distribuciones simétricas, ¿qué relación tiene con la media aritmética?

Respuesta.-

Ventajas:

No se ve afectada por la dispersión de los valores extremos

Es calculable aun cuando se desconozcan los valores extremos (por ejemplo en distribuciones de edades en que los intervalos extremos son del tipo “<15” ó “>60”)

Inconvenientes:

Salvo los valores adyacentes, los demás no intervienen en su determinación. Por ejemplo, la mediana de las distribuciones A = {100, 101, 102, 103} y B = {1, 101, 102, 1000} es la misma

$$\frac{101+102}{2} = 101,5.$$

En una distribución simétrica, la mediana y la media aritmética coinciden.

3.- Momentos respecto al origen en una distribución bidimensional

Respuesta.-

Consideremos la variable bidimensional (X,Y) que toma los valores (x_i, y_j) con frecuencias n_{ij} , $1 \leq i \leq r$; $1 \leq j \leq s$ y sea $N = \sum_{(i,j)=(1,1)}^{(r,s)} n_{ij}$. Se define el momento de orden p, q, respecto del origen

como:
$$a_{pq} = \frac{1}{N} \sum_{(i,j)=(1,1)}^{(r,s)} x_i^p y_j^q n_{ij}.$$

Algunos momentos notables son $a_{10} = \frac{1}{N} \sum_{(i,j)=(1,1)}^{(r,s)} x_i n_{ij}$ y $a_{01} = \frac{1}{N} \sum_{(i,j)=(1,1)}^{(r,s)} y_j n_{ij}$ que son las medias

respectivas de las distribuciones marginales de la X y de la Y. También $a_{20} = \frac{1}{N} \sum_{(i,j)=(1,1)}^{(r,s)} x_i^2 n_{ij}$ y

$a_{02} = \frac{1}{N} \sum_{(i,j)=(1,1)}^{(r,s)} y_j^2 n_{ij}$ que son las medias de los cuadrados de las variables marginales, respectivamente.

Finalmente $a_{11} = \frac{1}{N} \sum_{(i,j)=(1,1)}^{(r,s)} x_i y_j n_{ij}$, media de los productos de las variables.

4.- Métodos para la desestacionalización de series temporales

Los métodos para desestacionalizar una serie temporal se basan en la construcción de un índice de variación estacional. A continuación el proceso de desestacionalización consiste en dividir cada valor de la serie por el índice correspondiente (y multiplicar por 100 si el índice está en %). La serie resultante es la serie desestacionalizada.

Para la construcción del índice de variación estacional podemos utilizar el método de la razón a la media móvil, en el caso de que la serie tenga carácter multiplicativo, o el método de la tendencia por ajuste por mínimo cuadrático si la serie es de carácter aditivo.

SEGUNDA PARTE: PROBLEMAS

1. Obtener a través de los momentos, las rectas de regresión mínimo cuadráticas asociadas a la siguiente tabla de correlación:

<div style="border: 1px solid black; padding: 2px;">X \ Y</div>	0	3	6	TOTAL
1	1	5	2	8
2	4	4	1	9
TOTAL	5	9	3	17

Y obtener la varianza explicada por la regresión y la varianza residual, comentando sus resultados

Solución.-

Efectuamos los cálculos adecuados con las distribuciones marginales para obtener los momentos respecto del origen:

X \ Y	0	3	6	$n_{i\cdot}$	$x_i n_{i\cdot}$	$x_i^2 n_{i\cdot}$
1	1	5	2	8	8	8
2	4	4	1	9	18	36
$n_{\cdot j}$	5	9	3	17	26	44
$y_j n_{\cdot j}$	0	27	18	45		
$y_j^2 n_{\cdot j}$	0	81	108	189		

$x_i y_j n_{ij}$	0	15	12
	0	24	12
			63

obteniéndose:

$$a_{10} = \frac{26}{17} = 1,53$$

$$m_{20} = a_{20} - a_{10}^2 = 0,25$$

$$a_{20} = \frac{44}{17} = 2,59$$

$$m_{02} = a_{02} - a_{01}^2 = 4,11$$

$$a_{01} = \frac{45}{17} = 2,65$$

$$m_{11} = a_{11} - a_{10} \cdot a_{01} = -0,34$$

$$a_{02} = \frac{189}{17} = 11,12$$

$$R = \frac{m_{11}}{\sqrt{m_{20} \cdot m_{02}}} = -0,3385$$

$$a_{11} = \frac{63}{17} = 3,71$$

$$R^2 = 0,1146$$

Recta de regresión de Y/X:

$$y - a_{01} = \frac{m_{11}}{m_{20}} (x - a_{10}) \Leftrightarrow y - 2,65 = \frac{-0,34}{0,25} (x - 1,53) \Leftrightarrow y = -1,375x + 4,75$$

Recta de regresión de X/Y:

$$x - a_{10} = \frac{m_{11}}{m_{02}} (y - a_{01}) \Leftrightarrow x - 1,53 = \frac{-0,34}{4,11} (y - 2,65) \Leftrightarrow x = -0,083y + 1,75 \Leftrightarrow y = -12x + 21$$

$$\text{Varianza explicada por la regresión} = S_{yt}^2 = m_{02} \cdot R^2 = 4,11 \cdot 0,1146 = 0,471$$

$$\text{Varianza residual} = S_{ry}^2 = m_{02} - S_{yt}^2 = 4,11 - 0,471 = 3,64$$

La mayor parte de la varianza de la variable dependiente resulta ser residual (concretamente un $\frac{3,64}{4,11} \cdot 100 = 88,5\%$) y consecuentemente una pequeña parte (un $R^2 \cdot 100 = 11,5\%$) estará explicada por la regresión. Significa esto que el ajuste no es bueno y las rectas de regresión no pueden usarse de forma fiable para hacer estimaciones.

2. En la siguiente tabla se muestra la serie del Valor añadido bruto, en millones de euros, generado por el sector industrial español en el periodo 2000 a 2006, donde el Índice de volumen establece la evolución real del sector, es decir, una vez eliminada la influencia de los precios.

Período (t)	VAB Industria	Índice de volumen
2000	103.415	100,00
2001	108.985	103,22
2002	111.846	103,07
2003	115.154	104,34
2004	119.293	105,00
2005	124.568	106,05
2006	132.419	109,08

- Elabore un índice (año base 2002) de la evolución del VAB a precios corrientes y constantes.
- Elabore un índice (año base 2002) de la evolución de los precios en el periodo. ¿En qué año los precios experimentaron el mayor crecimiento?
- Calcule la tasa de variación media acumulativa del VAB del sector entre 2000 y 2006 (en volumen).

Solución.-

a) El índice (año base 2002) de la evolución del VAB a precios corrientes lo obtenemos dividiendo cada valor del VAB por 111846 y multiplicando por 100.

Años	Indice VAB a precios corrientes
2000	92,46
2001	97,44
2002	100,00
2003	102,96
2004	106,66
2005	111,37
2006	118,39

Dividiendo cada VAB por su índice de volumen, obtenemos una relación de precios corrientes, de la que a su vez, podemos obtener una relación de números índices de precios en base año 2002, con la que “deflactaremos” los índices VAB a precios corrientes para obtener la relación de índices VAB a precios constantes:

Años	$\frac{\text{VAB}}{\text{Indice de volumen}} = \text{precios corrientes}$	Indice de precios (base 2002)	Indice VAB a precios constantes
2000	1034,15	95,30	97,02
2001	1055,85	97,30	100,15
2002	1085,15	100,00	100,00
2003	1103,64	101,70	101,23
2004	1136,12	104,70	101,87
2005	1174,62	108,24	102,89
2006	1213,96	111,87	105,83

b) El índice (año base 2002) de la evolución de precios es el calculado en el apartado anterior. De esa relación calcularemos las tasas de crecimiento:

Años	Indice de precios (base 2002)	Tasas de crecimiento (en %)
2000	95,30	
2001	97,30	$\frac{97,3 - 95,3}{95,3} \cdot 100 = 2,10$
2002	100,00	$\frac{100 - 97,3}{97,3} \cdot 100 = 2,77$
2003	101,70	$\frac{101,7 - 100}{100} \cdot 100 = 1,70$
2004	104,70	$\frac{104,7 - 101,7}{101,7} \cdot 100 = 2,94$
2005	108,24	$\frac{108,24 - 104,7}{104,7} \cdot 100 = 3,39$
2006	111,87	$\frac{111,87 - 108,24}{104,7} \cdot 100 = 3,35$

Se observa que el mayor crecimiento de precios se experimentó en el año 2005

c) La tasa de variación del VAB (en volumen) de todo el periodo 2000-2006 sería, en tanto por uno: $\frac{109,08-100}{100} = 0,0908$. Esto significa que un determinado volumen V_0 del año 2000, se va a convertir en $V_0 + V_0 \cdot 0,0908 = V_0(1 + 0,0908) = V_0 \cdot 1,0908$ en el año 2006. Por tanto, si x es la tasa de variación media acumulativa, tiene que cumplirse que $V_0(1+x)^6 = V_0 \cdot 1,0908$, de donde $1+x = \sqrt[6]{1,0908} \Leftrightarrow x = \sqrt[6]{1,0908} - 1 = 1,014590648 - 1 = 0,014590648$ que equivale aproximadamente a un 1,46%.