

ANALISIS DE DATOS CUANTITATIVOS - 2007

Dr. Juan Gagliardi

La orden MEDIAS

En las clases "El Módulo ANALYSIS" y "Análisis de datos con ANALYSIS" vimos como se utiliza esta orden. Brevemente la repasamos mencionando que la orden MEDIAS ofrece una tabla de datos continuos u ordinales y realiza las estadísticas adecuadas.

Esta orden también puede utilizarse con sólo una variable cuantitativa para describirla como por ejemplo:

EDA D	Frecuenci a	Porcentaj e	Porcentaje acumulado
31	1	0,1%	0,1%
34	2	0,2%	0,3%
36	4	0,4%	0,7%
38	7	0,7%	1,4%
39	3	0,3%	1,7%
40	7	0,7%	2,4%
41	10	1,0%	3,3%
42	16	1,6%	4,9%
.....
83	10	1,0%	97,2%
84	9	0,9%	98,0%
85	5	0,5%	98,5%
86	6	0,6%	99,1%
87	5	0,5%	99,6%
89	3	0,3%	99,9%
91	1	0,1%	100,0%
Total	1020	100,0%	100,0%

Observacion es	Total	Media	Varianz a	Desviación típica
1020	63626,000	62,378	133,324	11,5466
		0 4	8	
Mínim o	25%	Median a	75%	Máxim o Moda

31,000 54,000 63,000 71,000 91,000 63,000
 0 0 0 0 0 0

Además de ver cuál es la media y el desvío estándar de esta población, el rango con los valores mínimo y máximo y los percentilos 25 y 75, podemos ver que la media, la mediana y la moda son aproximadamente iguales ($62,4 \approx 63,0 \approx 63,0$) y por lo tanto aceptar que esta variable tiene una distribución gaussiana.

A diferencia de la versión anterior (Epi 6) el Epi 2000 no brinda el cálculo del error estándar de la media. El mismo puede calcularse a partir de la fórmula:

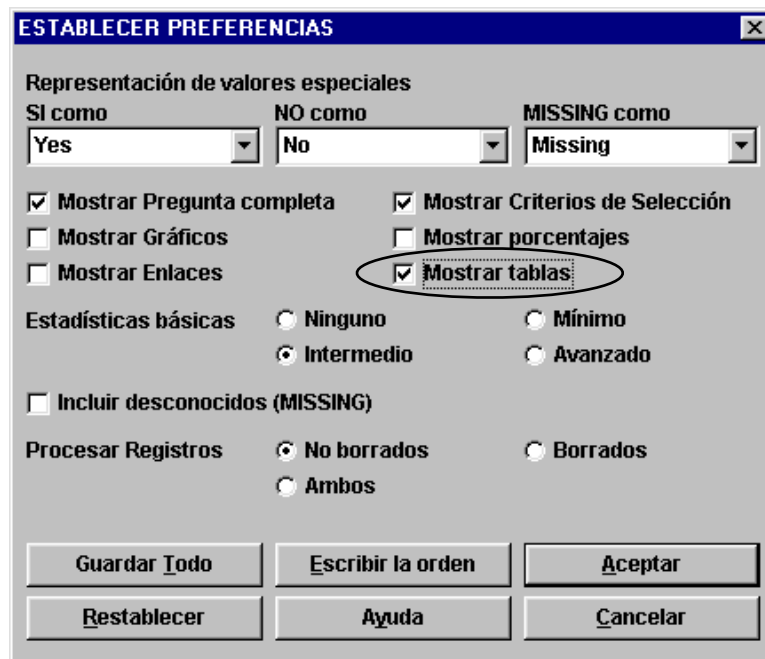
$$ES = \frac{DS}{\sqrt{N}}$$

En este caso el resultado sería:

$$ES = \frac{11.546}{\sqrt{1020}} = \frac{11.546}{31.937} = 0.362$$

Cuando la orden MEDIAS contiene una variable numérica y otra cualitativa que determina qué grupos va a comparar presenta en sus resultados una tabla comparando los datos, luego realiza un análisis descriptivo de cada grupo y finalmente la comparación de los datos cuantitativos. Así la orden MEANS EDAD HTA permite comparar la edad de los hipertensos (HTA="S") con la edad de los no hipertensos (HTA="N").

Recordemos también que si preferimos que no se muestre la tabla de datos debemos quitar el tilde de "Mostrar Tablas" de la ventana "Establecer Preferencias".



Veamos con un ejemplo.

La orden MEANS EDAD SEXO permite comparar la edad de los hombres con la edad de las mujeres. Nótese que no mostramos la tabla de datos ya que quitamos la tilde respectiva en la ventana de preferencias.

Los resultados se presentarán de la siguiente forma:

Estadísticas descriptivas de la variable de cruce

	Obs	Total	Media	Varianza	Desviación típica
F	321	21104,000	65,7445	120,9658	10,9984
M	699	42522,000	60,8326	131,5780	11,4707

	Mínimo	25%	Media	75%	Maximo	Mode
F	40,0000	57,0000	66,0000	74,0000	89,0000	68,0000
M	31,0000	53,0000	61,0000	69,0000	91,0000	60,0000

ANOVA, test paramétrico para comparación de medias

(Para datos normalmente distribuidos)

Variación	SC	gl	MS	Estadístico F
Entre grupos	5307,4562	1	5307,4562	41,3862
Intra grupo	130550,4693	1018	128,2421	
Total	135857,9255	1019		

Estadístico T = 6,4332

Valor p= 0,0000

Test de Bartlett para igualdad de Varianzas poblacionales

Chi cuadrado de Bartlett= 0,7667 df= 1 valor p=0,3813

Un valor de p pequeños (menor de 0,05) sugiere que las varianzas no son homogéneas y que el test ANOVA no es apropiado.

Test de dos muestras de Mann-Whitney/Wilconxon (Test de Kruskal-Wallis para dos grupos)

H de Kruskal-Wallis (equivalente a Chi cuadrado)= 37,3173

Grados de libertad= 1

Valor p= 0,0000

Dividiremos esta presentación de los resultados por partes para poder comprender mejor su significado.

La primera parte es:

Estadísticas descriptivas de la variable de cruce

	O bs	Total	Media	Varianz a	Desviación típica	
F	32 1	21104,000 0	65,744 5	120,965 8	10,9984	
M	69 9	42522,000 0	60,832 6	131,578 0	11,4707	
	Mínim o	25%	Media n	75%	Maximu m	Mode
F	40,000 0	57,000 0	66,000 0	74,000 0	89,0000	68,000 0
M	31,000 0	53,000 0	61,000 0	69,000 0	91,0000	60,000 0

El título nos indica que se trata de la estadística descriptiva de la variable que estamos evaluando. Luego podemos ver el análisis descriptivo de la variable en ambos grupos. Nos dice cuál es la edad de media de las mujeres ("F") y la de los hombres ("M"), sus respectivos desvíos estándar y las variancias de cada uno. Nos informa además el rango y los percentilos.

La segunda parte:

ANOVA, test paramétrico para comparación de medias

(Para datos normalmente distribuidos)

Variación	SC	gl	MS	Estadístico F
Entre grupos	5307,4562	1	5307,456 2	41,3862
Intra grupo	130550,469 3	101 8	128,2421	
Total	135857,925	101 5	9	

Estadístico T = 6,4332

Valor p= 0,0000

Nos muestra el test de ANOVA en el que el valor de p es equivalente al test de t de Student ya que sólo hay 2 muestras (edad de mujeres versus edad de hombres). Cuando los grupos son más de 2 el valor de t no se aplica y por lo tanto no se presenta en el resultado.

Variación indica la fuente de la variancia, *Intra* = dentro de los grupos e *Inter* = entre las medias. *SC* significa suma de cuadrados, *gl* grados de libertad y *MS* = cuadrados medios (CM). *Estadístico F*, *Estadístico T* y *valor-p* no necesitan aclaración.

El estadístico F surge de la división de los cuadrados medios. En este ejemplo:

$$F = CM_{em} / CM_{dg}$$

$$F = 5307.456 / 128.242$$

$$F = 41.386$$

Y como $F = t^2$, en este ejemplo:

$$t = \sqrt{F}$$

$$t = \sqrt{41.386}$$

$$t = 6,4332$$

La tercera parte de estos resultados está representada por el test homoscedasticidad, en este caso presentado como test de homogeneidad de las varianzas de Barlett.

Test de Bartlett para igualdad de Varianzas poblacionales

Chi cuadrado de	0,766	df=	valor
Bartlett=	7	1	p=0,3813

Un valor de p pequeños (menor de 0,05) sugiere que las varianzas no son homogéneas y que el test ANOVA no es apropiado.

Aquí nos informa que las variancias son iguales con un 95% de confianza por lo que podemos utilizar los resultados expresados por el test de ANOVA presentados más arriba. En el caso que los datos no tengan una distribución gaussiana (por ejemplo si hacemos la media del valor de CPK por cada sexo) nos presenta este mensaje:

Test de Bartlett para igualdad de Varianzas poblacionales

Chi cuadrado de	19,690	df=	valor
Bartlett=	5	1	p=0,0000

Un valor de p pequeños (menor de 0,05) sugiere que las varianzas no son homogéneas y que el test ANOVA no es apropiado.

Esto nos indica que el test de homoscedasticidad encuentra que las variancias son diferentes en forma significativa (valor p = 0,0000) y por lo tanto recomienda utilizar los resultados de los test no-paramétricos que brinda más abajo en lugar de ANOVA.

En la última parte se presentan los resultados de los test no-paramétricos:

Test de dos muestras de Mann-Whitney/Wilconxon (Test de Kruskal-Wallis para dos grupos)

H de Kruskal-Wallis (equivalente a Chi 37,317
cuadrado)= 3
Grados de libertad= 1
Valor p= 0,0000

El programa realiza también los cálculos para los test no paramétricos independientemente del tipo de distribución de las variables.

En el caso de la edad según el sexo muestra que la diferencia es también estadísticamente significativa, pero no debe tenerse en cuenta el resultado ya que la variable tiene una distribución gaussiana.