

## COVARIANZA EN LOS DISEÑOS DE GRUPOS NATURALES

El control de las variables se debe llevar a cabo tanto al inicio como durante el proceso de la investigación, para así controlar las variables extrañas. Para ello uno de los métodos más utilizados es el del azar, aunque en ocasiones en educación esto no es posible, ya que nos situamos ante grupos naturales, que no aseguran el control de las variables extrañas.

Para poder controlar a las variables extrañas en este caso se plantean estrategias como la que veremos ahora. Así podemos disponer lo que se llama covariables, o sea variables que varían a la vez con la variable dependiente, ya que están relacionadas entre sí como se observa al realizar una correlación significativa. Si podemos analizar estas covarianzas podemos contrastar que las diferencias finales son significativas o no, vamos si se cumple la hipótesis nula o no, como ocurre con la prueba F o ANAVA, ya que permite ajustar las diferencias a las diferencias iniciales no controladas.

Así en esta situación deberemos emplear la prueba F de análisis de covarianzas. Como siempre intentaremos comprenderlo a partir de casos prácticos.

Se puede comprobar que la hipótesis de que el método de aprendizaje de la lectura el A, el B o el C, influyen en los resultados de final de curso. Se desea contrastar la hipótesis para un  $\alpha < 0,01$ . El investigador selecciona tres aulas, les asigna aleatoriamente los métodos y a los efectos de probar su hipótesis, escoge al azar 6 niños. Y como el colegio existen datos de exploraciones psicopedagógicas previas sobre aptitudes, decide utilizar tales datos como si fuese la covariable. Así obtendremos toda una serie de datos donde  $x$  será la covariable, la variable que analizaremos a parte de la variable dependiente original y la variable dependiente será la variable  $Y$ .

Así obtendremos los siguientes datos:

	Grupo a		Grupo B		Grupo c		total	
	x	Y	x	Y	x	Y	x	Y
$\Sigma X$	43	32	47	41	37	25	127	98
$\Sigma X^2$	323	178	375	283	239	111	937	572
	237		324		162		723	

Ahora pasaremos a calcular la prueba F, donde tenemos dos variables una variable dependiente o variante  $Y$  y una covariable  $x$ , covariable que correlacione con la variable dependiente, como se ve con el coeficiente de correlación.

Para ello nos proporcionan una tabla de sumatorios.

Debemos calcular como en los otros casos las variabilidades, las medias cuadráticas y la  $F$ , tanto para el total como entre grupos y dentro de los grupos, y ello no solo para la variable  $Y$  sino para la  $X$  y para su producto  $XY$ . Si se tratara de un ANAVA simple solo tendríamos que calcularlas para  $Y$ .

Veamos primero las tres  $SCT$ , con las formulas

$$S.C_T = \sum_{i=1}^N X_i^2 - \frac{\left(\sum_{i=1}^N X_i\right)^2}{N}$$

Sacaremos primero la SCT de x = 937-(127<sup>2</sup>:18)= 937-896,05= 40,94

SCT de y =572- (98<sup>2</sup>:18)= 572 -533.55 = 38.44

SCT de xy= 723 - ( (127 x 98):18) = 723 - 691.44 = 31.56

Despues realizaremos la varianza cuadrática de entre los grupos

$$S.C.E = \sum_{i=1}^G \frac{\left(\sum_{j=1}^n X_{ij}\right)^2}{n} - \frac{\left(\sum_{i=1}^N X_i\right)^2}{N}$$

SCE de x = ((43<sup>2</sup>+ 47<sup>2</sup>+ 37<sup>2</sup>):6 - 896.05= 904.5 - 896.05 = 8.44

SCE de y = ((32<sup>2</sup>+41<sup>2</sup>+25<sup>2</sup>):6) - 533.55= 555 - 533.55= 21.44

SCE de xy = ( (43x32) + (47x41) + (37x25) : 6 ) - ( 127x98):18= 704.67 - 691.44 =13.23

Ahora las SCd:

SCT-SCE= SCD

SCD de x = 40.94 -8.44=32.5

SCD de y =38.44-21.44=17

SCD de xy =31.56-13.23 =18.33

Con estos datos por separado podríamos hacer las anavas simples , la anava de la x la anava de la y. Pero si hiciésemos eso siempre nos quedaría la duda de que el efecto se deba a no haber el control inicial.

$$SC_{Y.X_T} = SC_{T(Y)} - \frac{SC_{T(XY)}^2}{SC_{T(X)}}$$

$$SC_{Y.X_D} = SC_{D(Y)} - \frac{SC_{D(XY)}^2}{SC_{D(X)}}$$

$$SC_{Y.X_E} = SC_{Y.X_T} - SC_{Y.X_D}$$

MC<sub>Y.X</sub>, esto es, M.C. de la variable Y ajustada: MC<sub>Y.XT</sub>; MC<sub>Y.XE</sub>; MC<sub>Y.XE</sub>

Estas serán las formulas que tendremos que aplicar ahora para saber la anava de la variable dependiente Y y quedarnos seguros de que no se ve afectada por la otra variable sería el calculo de las medias cuadráticas y varianzas cuadráticas ajustadas:

$$\text{La SC de y ajustada a x total} = 31.56^2 - (31.56^2 : 40.94^2) = 38.45 - 24.33 = 14.11$$

$$\text{La SC de y ajustada a x de la diferencia} = 17 - (18.33^2 : 32.5) = 17 - 10.34 = 6.66$$

$$\text{La SCde y ajustada a x de entregrupos} = 14.11 - 6.66 = 7.45$$

Fuente de variación	SCyx ajustada	GL	MCyx	F
Entre grupos	7.45	2	3.73	7.77
Dentro de grupos	6.66	4	0.48	
total	14.11	16		

El valor de F empirico se comparará con el valor teórico de la tabla para 2 y 14 gl. Y dará 6,51 como el empírico es de 7,7 > 6,51 y rechazamos la Ho para un alfa 0.01

Ahora tenemos que saber si estas diferencias existentes entre x e y se ven concretamente en los grupos que trabajamos el A y el B. Y así lo primero que debemos buscar es el coeficiente de regresión dentro de los grupos

$$\text{Media}_{Y.X} = \bar{Y} - b_p (\bar{X} - \bar{X}_G) \text{ donde:}$$

$\bar{X}$  = media de la variable X para cada uno de sus grupos.

$\bar{X}_G$  = media general de la variable X.

$\bar{Y}$  = media de la variable Y para cada uno de sus grupos.

$b_p$  = coeficiente de regresión dentro de los grupos, sabiendo que:

$$b_p = \frac{SC_{D(Y)}}{SC_{D(X)}}$$

Contraste entre medias de Y ajustadas, por ejemplo para los grupos A y B:

$$F = \frac{(\text{Media}_{Y.X(A)} - \text{Media}_{Y.X(B)})^2}{MC_{Y.X_D} \left[ \frac{2}{n} + \frac{(\bar{X}_A - \bar{X}_B)^2}{SC_{D(X)}} \right]}$$

Donde Media Y.X(A)= media ajustada de Y en el grupo A, siendo F con 1 y con [g(n-1)-1] g.l.

Primero la b es el coeficiente de regresión dentro del grupo

$$b_p = \frac{SC_{D(Y)}}{SC_{D(X)}}$$

$$18.33 : 32.5 = 0.564$$

Ahora veremos las medias ajustadas

$$Media_{Y.X} = \bar{Y} - b_p(\bar{X} - \bar{X}_G)$$

$$Media \text{ de } y \text{ ajustada a } x \text{ en A} = (32:6) - 0.564 ((43:6) - (127:181)) = 5.33 - 0.564 (7.17 - 7.05) = 5.33 - 0.07 = 5.26$$

$$\text{Ahora la otra media } y \text{ ajustada en } x \text{ en el grupo B} = (41:6) - 0.564 ((47:6) - (127:18)) = 6.38 - 0.564 (7.83 - 7.05) = 6.38 - 0.44 = 6.39$$

Por ultimo el contrasta de f para ajustadas

$$F = \frac{(Media_{Y.X(A)} - Media_{Y.X(B)})^2}{MC_{Y.X_D} \left[ \frac{2}{n} + \frac{(\bar{X}_A - \bar{X}_B)^2}{SC_{D(X)}} \right]}$$

Donde Media Y.X(A)= media ajustada de Y en el grupo A, siendo F con 1 y con [g(n-1)-1] g.l.

$$F = (6.39 - 5.26)^2 / 0.48(2/6 + (7.17 - 7.83)^2 / 32.5) = 1.28 : 0.48 (0.333 + 0.013) = 1.28 : 0.48 \times 0.346 = 1.28 : 0.166 = 7.71$$

Asi se rechaza la hipótesis nula porque este valor empírico es mas grande que el valor teórico que es 3.74 y por tanto si son significativas . Esto mismo se debería hacer para los grupos B y C y para el grupo A y C.