Tema 13

DISEÑOS CON VARIABLES COVARIADAS

SUPUESTO

Un investigador está interesado en estudiar la eficacia de un tratamiento de relajación muscular. Decide utilizar un *diseño pre-test / post-test con grupo control*. Selecciona a 10 alumnos de Psicología y mide su tono muscular antes de la intervención, asignándolos aleatoriamente y de forma equilibrada a la situación de *Grupo de Terapia* y *Grupo de Control*. Una vez finalizada la intervención con el grupo experimental, vuelve a medir el tono muscular de ambos grupos. El objetivo es analizar si el tono muscular de los sujetos que completaron la terapia está ahora más relajado que el de los sujetos que no recibieron el entrenamiento en relajación. El investigador se plantea si el tono muscular previo de cada sujeto está correlacionado con el tono muscular expresado en la medición del *post-test*. Por ello, decide incluir en su diseño el nivel de tono muscular que cada sujeto manifestaba en la medición de *pre-test* para poder llegar así a conclusiones más válidas, controlando esa posible fuente de sesgo. Los resultados del *pretest* se recogen en la variable X y los del *postest* en la variable Y, tanto para el tratamiento a_1 (*Grupo de Control*) como para el tratamiento a_2 (*Grupo de Terapia*). ¿Las diferencias entre los dos grupos son estadísticamente significativas?

Tabla *Matriz de resultados*

Grupo de Control a_1		*Grupo de Terapia a_2*	
Pre-test(X)	*Post-test(Y)*	*Pre-test(X)*	*Post-test(Y)*
1	5	5	14
3	8	7	17
3	7	7	16
1	2	5	11
2	3	6	12
$M_{a_1} = 2$	$M_{a_1} = 5$	$M_{a_2} = 6$	$M_{a_2} = 14$

$M_X = 4$ $\quad$ $M_Y = 9.5$

Error de estimación bajo la *hipótesis nula*:

$$\hat{Y} = M \rightarrow E_{H_0} = Y - M = \begin{pmatrix} 1 & 5 \\ 3 & 8 \\ 3 & 7 \\ 1 & 2 \\ 2 & 3 \\ 5 & 14 \\ 7 & 17 \\ 7 & 16 \\ 5 & 11 \\ 6 & 12 \end{pmatrix} - \begin{pmatrix} 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \end{pmatrix} = \begin{pmatrix} -3 & -4.5 \\ -1 & -1.5 \\ -1 & -2.5 \\ -3 & -7.5 \\ -2 & -6.5 \\ 1 & 4.5 \\ 3 & 7.5 \\ 3 & 6.5 \\ 1 & 1.5 \\ 2 & 2.5 \end{pmatrix}$$

La ***suma de cuadrados total*** del modelo será:

$$\mathbf{SC}_{\text{total}} = \mathbf{E}_{H0}^{2} = \begin{pmatrix} -3 & -1 & -1 & -3 & -2 & 1 & 3 & 3 & 1 & 2 \\ -4.5 & -1.5 & -2.5 & -7.5 & -6.5 & 4.5 & 7.5 & 6.5 & 1.5 & 2.5 \end{pmatrix} \begin{pmatrix} -3 & -4.5 \\ -1 & -1.5 \\ -1 & -2.5 \\ -3 & -7.5 \\ -2 & -6.5 \\ 1 & 4.5 \\ 3 & 7.5 \\ 3 & 6.5 \\ 1 & 1.5 \\ 2 & 2.5 \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{48} & \mathbf{106} \\ \mathbf{106} & \mathbf{254.5} \end{pmatrix}$$

⇒A partir de la **SC**$_{\text{total}}$ se obtiene fácilmente la correlación entre la variable dependiente Y y su covariada X.

$$\boxed{r_{XY} = \frac{SP_{XY}}{\sqrt{SC_Y\,SC_X}} = \frac{106}{\sqrt{48 \cdot 254.5}} = 0.9590502}$$

⇒A partir de la correlación puede calcularse la proporción de varianza de la variable dependiente explicada por la covariada.

$$\boxed{\eta^2_{XY} = r^2_{XY} = 0.\ 9590502^2 = 0.9197773}$$

Por tanto, el 91.98% de la variabilidad de la variable dependiente se conoce a partir de las puntuaciones en la covariada.

⇒La proporción de **varianza no explicada**:

$$\Lambda_{XY} = 1 - \eta^2_{XY} = 1 - 0.\ 9197773 = 0.0802227$$

⇒La *suma de cuadrados total* correspondiente a la variable dependiente es **254.5** ($SC_{totalY} = 254.5$), esta *suma de cuadrados* hay que ajustarla descontando la varianza atribuida a la variable covariada:

$$SC_{totalY.X} = SC_{totalY} \cdot \eta^2_{XY}$$

La suma de cuadrados de la variable dependiente que está determinada por su relación con la covariada es:

$$SC_{totalY.X} = SC_{totalY} \cdot \eta^2_{XY} =$$
$$= 254.5 \cdot 0.9197773 = 234.0833$$

Del valor total de la suma de cuadrados (254.5) observada en la variable dependiente, 234.0833 viene explicado por las puntuaciones en la variable covariada.

⇒La ***suma de cuadrados ajustada*** ($SC^{*}_{\text{total}Y}$):

$$\boxed{SC^{*}_{\text{total}Y} = SC_{\text{total}Y}\,\Lambda_{XY}}$$

Aplicando este cálculo a nuestros datos.

$$SC^{*}_{\text{total}Y} = SC_{\text{total}Y}\,\Lambda_{XY} = 254.5 \cdot 0.0802227 = \boxed{20.4167}$$

⇒Efecto del tratamiento:

$$\mathbf{A}_{(1)} = \mathbf{M} - \mathbf{M}_a = \begin{pmatrix} 2 & 5 \\ 2 & 5 \\ 2 & 5 \\ 2 & 5 \\ 2 & 5 \\ 6 & 14 \\ 6 & 14 \\ 6 & 14 \\ 6 & 14 \\ 6 & 14 \end{pmatrix} - \begin{pmatrix} 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \end{pmatrix} = \begin{pmatrix} -2 & -4.5 \\ -2 & -4.5 \\ -2 & -4.5 \\ -2 & -4.5 \\ -2 & -4.5 \\ 2 & 4.5 \\ 2 & 4.5 \\ 2 & 4.5 \\ 2 & 4.5 \\ 2 & 4.5 \end{pmatrix}$$

⇒La suma de cuadrados correspondiente a la aplicación del tratamiento será entonces:

$$\mathbf{SC}_{entre} = \mathbf{A}' \cdot \mathbf{A} = \begin{pmatrix} -2 & -2 & -2 & -2 & -2 & 2 & 2 & 2 & 2 & 2 \\ -4.5 & -4.5 & -4.5 & -4.5 & -4.5 & 4.5 & 4.5 & 4.5 & 4.5 & 4.5 \end{pmatrix} \begin{pmatrix} -2 & -4.5 \\ -2 & -4.5 \\ -2 & -4.5 \\ -2 & -4.5 \\ -2 & -4.5 \\ 2 & 4.5 \\ 2 & 4.5 \\ 2 & 4.5 \\ 2 & 4.5 \\ 2 & 4.5 \end{pmatrix}$$

$$= \begin{pmatrix} 40 & 90 \\ 90 & 202.5 \end{pmatrix}$$

⇒La ***varianza residual*** después de aplicar el modelo de la *hipótesis alternativa* se obtiene sustrayendo de los valores de las variables dependientes la media general (**M**) y la variación explicada por el efecto del tratamiento (**A**).

$$\mathbf{E}_{H1} = \mathbf{Y} - \mathbf{M} - \mathbf{A} = \begin{pmatrix} 1 & 5 \\ 3 & 8 \\ 3 & 7 \\ 1 & 2 \\ 2 & 3 \\ 5 & 14 \\ 7 & 17 \\ 7 & 16 \\ 5 & 11 \\ 6 & 12 \end{pmatrix} - \begin{pmatrix} 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \\ 4 & 9.5 \end{pmatrix} - \begin{pmatrix} -2 & -4.5 \\ -2 & -4.5 \\ -2 & -4.5 \\ -2 & -4.5 \\ -2 & -4.5 \\ 2 & 4.5 \\ 2 & 4.5 \\ 2 & 4.5 \\ 2 & 4.5 \\ 2 & 4.5 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 1 & 3 \\ 1 & 2 \\ -1 & -3 \\ 0 & -2 \\ -1 & 0 \\ 1 & 3 \\ 1 & 2 \\ -1 & -3 \\ 0 & -2 \end{pmatrix}$$

⇒y la $\mathbf{SC}_{error}$ se obtiene elevando al cuadrado el error de estimación bajo la *hipótesis alternativa*.

$$\mathbf{SC}_{error} = \mathbf{E}_{H_1}{}^2 = \begin{pmatrix} -1 & 1 & 1 & -1 & 0 & -1 & 1 & 1 & -1 & 0 \\ 0 & 3 & 2 & -3 & -2 & 0 & 3 & 2 & -3 & -2 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 1 & 3 \\ 1 & 2 \\ -1 & -3 \\ 0 & -2 \\ -1 & 0 \\ 1 & 3 \\ 1 & 2 \\ -1 & -3 \\ 0 & -2 \end{pmatrix} = \begin{pmatrix} 8 & 16 \\ 16 & 52 \end{pmatrix}$$

⇒Bajo la **<u>hipótesis alternativa</u>** obtenemos el coeficiente de correlación entre las dos variables aplicando el mismo procedimiento que bajo el *modelo restringido*:

$$\boxed{r_{12} = \frac{SP_{XY}}{\sqrt{SC_Y\, SC_X}} = \frac{16}{\sqrt{8 \cdot 52}} = 0.7844645}$$

constatamos que los errores de las dos variables están también fuertemente relacionados, aunque en este caso <u>se reduce</u> el coeficiente de correlación a 0.7845.

⇒Por tanto la proporción de ***varianza de error*** de la variable dependiente explicado por su covariada es:

$$\eta^2_{XY} = r^2_{XY} = 0.7559289^2 = 0.6153861$$

⇒y la ***varianza no explicada***

$$\Lambda_{XY} = 1 - \eta^2_{XY} = 1 - 0.6153861 = 0.3846154$$

⇒Podemos estimar la **suma de cuadrados del error** que viene determinada por la covariada,

$$SC_{errorY.X} = SC_{errorY} \cdot \eta^2_{XY} = 52 \cdot 0.6153861 = 32.0$$

⇒el resto, corresponde a la suma de cuadrados del error, pero ya ajustada por la presencia de la variable covariada,

$$SC^*_{errorY} = SC_{errorY} \cdot \Lambda_{XY} = 52 \cdot 0.3846154 = 20.0$$

Ajuste de la varianza explicada:

La suma de cuadrados correspondiente al efecto del tratamiento, después de ajustar el efecto de la covariada, se obtiene descontando del componente total ajustado el error ajustado:

$$\boxed{SC^*_{entre Y} = SC^*_{total Y} - SC^*_{error Y}}$$

$$= 20.4 - 20.0 = 0.4$$

⇒Analicemos la descomposición que hemos realizado de la variabilidad de la variable dependiente en función de las dos hipótesis y considerando la varianza que ya está explicada por la variable covariada:

Descomposición de la suma de cuadrados de la variable dependiente en función de los dos modelos y ajustando el efecto de la covariada

Suma de Cuadrados	r_{XY}	η^2	SC_Y	$SC_{Y.X}$	SC^*_Y
$SC_{total} = \begin{pmatrix} 48 & 106 \\ 106 & 254.5 \end{pmatrix}$	0.9590502	0.9197773	254.5	234.1	20.4
$SC_{error} = \begin{pmatrix} 8 & 16 \\ 16 & 52 \end{pmatrix}$	0.7844645	0.6153861	52.0	32.0	20.0
$SC_{entre} = \begin{pmatrix} 40 & 90 \\ 90 & 202.5 \end{pmatrix}$	–	–	202.5	*182.5*	*0.4*

Prueba de la hipótesis

Tabla *Análisis de la covarianza aj ustando el ef ecto del nive tono muscular en el pre-test*

Fuente	*SC*	*gl*	*MC*	*Razón F*	*p*
X	32.000	1	32.000	11.200	< 0.05
A^*	0.417	1	0.417	0.146	> 0.05
Error*	20.000	7	2.857		
Total*	20.417	8			

ANOVA

Análisis de la varianza entre la situación pre-test / post-test

Fuente	*SC*	*gl*	*MC*	*Razón F*	*p*	$\eta 2$
A	202.50	1	202.50	31.154	< 0.05	0.796
Error	52.00	8	6.50			
Total	254.5	9				

ANCOVA

Análisis de la covarianza ajustando el efecto del PRE-TES

Fuente	*SC*	*gl*	*MC*	*Razón F*	*p*
X	32.000	1	32.000	11.200	< 0.05
A^*	0.417	1	0.417	0.146	> 0.05
Error*	20.000	7	2.857		
Total*	20.417	8			

Supuestos básicos del análisis de la covarianza

Especialmente destacan tres consideraciones.

- En primer lugar tiene que existir una *relación entre la variable dependiente y la variable covariada* ya que, en caso contrario no tendría sentido realizar el ajuste.
- Con el segundo supuesto se comprueba que *la variable covariada no mantiene un efecto estadísticamente significativo en la variable tratamiento*.
- Por último, hay que garantizar que *no existe un efecto de interacción entre los tratamientos y la variable covariada en el valor de la variable dependiente*.