ANÁLISIS DE DATOS EN PSICOLOGÍA II

Antonio Pardo
Rafael San Martín

PSICOLOGÍA PIRÁMIDE

Normalidad

El segundo de los supuestos afirma que las observaciones de cada tratamiento o grupo constituyen una muestra aleatoria extraída de una población normal. En muchos trabajos ha quedado probado que el estadístico F es muy robusto frente al incumplimiento de este supuesto (ver, por ejemplo, Box, 1953). En general, si las poblaciones de donde se muestrea, aun no siendo normales, son simétricas o tienen forma parecida (por ejemplo, todas positivamente asimétricas y leptocúrticas), el comportamiento del estadístico F es muy aceptable incluso con tamaños muestrales relativamente pequeños. Con todo, siempre es recomendable trabajar con tamaños muestrales moderadamente grandes; de esta forma tendremos garantizado un comportamiento aceptable del estadístico F incluso en aquellas situaciones en que las poblaciones originales muestren un marcado alejamiento de la normalidad.

Las desviaciones de la normalidad pueden detectarse, cuando son muy marcadas, utilizando sencillos métodos gráficos (ver Chambers, Cleveland, Kleiner y Tukey, 1983; o Iman y Conover, 1983). Pero si estamos interesados en contrastar la hipótesis de que una muestra procede de una población normalmente distribuida podemos utilizar la *prueba de normalidad* de Lillieffors (1967; ver Conover, 1980, págs. 357-361; o San Martín y Pardo, 1989, págs. 89-90).

Homocedasticidad (igualdad de varianzas)

El último de los supuestos referidos al modelo completamente aleatorizado afirma que las observaciones han sido extraídas de poblaciones con la misma varianza. Durante muchos años se ha venido aceptando, a partir de trabajos como el de Horsnell (1953) o Box (1954a), que el estadístico F es muy robusto frente al incumplimiento de este supuesto si los tamaños muestrales son iguales y no demasiado pequeños. Estudios más recientes, sin embargo, parecen confirmar que, cuando las varianzas poblacionales son distintas, el comportamiento del estadístico F puede resultar insatisfactorio incluso con tamaños muestrales iguales (Rogan y Keselman, 1977; Tomarken y Serlin, 1986; Wilcox, Charlin y Thompson, 1986; Harwell y otros, 1992). Y, desde luego, si los tamaños muestrales son diferentes, muchos trabajos (ver Glass, Peckham y Sanders, 1972, para una revisión) ponen de manifiesto que el estadístico F pierde su robustez frente a la heterogeneidad de varianzas: se convierte en muy conservador cuando las varianzas más grandes corresponden a los grupos de mayor tamaño (perdiendo, además, potencia) y en marcadamente liberal cuando las varianzas más grandes corresponden a los grupos de menor tamaño.

Estas consideraciones hacen más que recomendable detenerse a contrastar la hipótesis de igualdad de varianzas cuando se tiene intención de utilizar el estadístico F del ANOVA. Para ello, disponemos de varias pruebas de significación, pero no todas ellas son igualmente robustas frente al incumplimiento del supuesto de normalidad (más bien, la mayoría son poco robustas; ver O'Brien, 1981). Entre las más recomendables se encuentra la *prueba de Levene* (1960), que consiste en 1) transformar las puntuaciones originales Y_{ij} en desviaciones en valor absoluto respecto a las medias de sus respectivos grupos:

$$D_{ij} = |Y_{ij} - \bar{Y}_j|$$

y 2) aplicar el estadístico F del ANOVA a las puntuaciones transformadas. Si las varianzas son iguales, las desviaciones D_{ij} serán parecidas en todos los grupos y las medias de esas desviaciones nos servirán como referencia del grado de igualdad entre las varianzas. Una F significativa nos llevará al rechazo de la hipótesis nula H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_J^2$. Brown y Forsythe (1974a) han propuesto un procedimiento idéntico al de Levene, pero utilizando las *medianas* en lugar de las medias.

Posteriormente, O'Brien (1981) ha diseñado un procedimiento también muy aceptable desde el punto de vista de la robustez. Al igual que el de Levene, consiste en obtener el estadístico F del ANOVA utilizando, no las puntuaciones originales Y_{ij}, sino una transformación de las mismas:

$$r_{ij} = \frac{(n_j - 1{,}5)n_j(Y_{ij} - \bar{Y}_j)^2 - 0{,}5S_j^2(n_j - 1)}{(n_j - 1)(n_j - 2)}$$

siendo S_j^2 la varianza insesgada de cada grupo (en cada grupo, la media de las puntuaciones transformadas debe ser igual a la varianza insesgada).

Si alguno de estos procedimientos nos lleva al rechazo de la hipótesis de igualdad de varianzas, podemos utilizar, como primera aproximación al problema de la heterogeneidad de varianzas, la solución propuesta por Box (1954a). Con varianzas distintas, el estadístico F del ANOVA sigue distribuyéndose según el modelo de probabilidad F, pero con los grados de libertad alterados. Box ha demostrado que el verdadero punto crítico (pc) se encuentra entre:

$$_{1-\alpha}F_{1,n-1} \geqslant pc \geqslant {}_{1-\alpha}F_{J-1,N-J}$$

El punto crítico $_{1-\alpha}F_{J-1,N-J}$ es el que corresponde a la distribución del estadístico F cuando se cumplen los supuestos del ANOVA. En el caso de incumplimiento de los supuestos de normalidad y homocedasticidad, podemos utilizar el mayor de los dos puntos críticos (el obtenido con 1 y $n - 1$ grados de libertad). Si con esta estrategia obtenemos un resultado significativo, podremos concluir que las medias poblacionales son distintas sin preocuparnos de si las varianzas son iguales o no. Si obtenemos un resultado no significativo, entonces no podremos detener ahí nuestro análisis, pues la solución propuesta por Box es extremadamente conservadora. Podríamos estimar el número de grados de libertad que corresponden a la F concreta que estamos utilizando (el propio Box propone un método para obtener esas estimaciones). Pero también podemos utilizar alguno de los procedimientos diseñados como alternativas al estadístico F para afrontar situaciones de heterogeneidad de varianzas.

De entre esos procedimientos alternativos, el estadístico V_w propuesto por Welch (1951) ha sido considerado repetidamente como uno de los más apropiados[20] en términos de potencia y protección contra los errores de tipo I (ver, por ejemplo, Kohr y Games, 1974; o Tomarken y Serlin, 1986). La obtención del estadístico de Welch es tediosa, pero no difícil:

$$w_j = \frac{n_j}{S_j^2} \quad ; \quad \bar{Y}^* = \frac{\sum_j w_j \bar{Y}_j}{\sum_j w_j}$$

$$\Lambda = \frac{3\Sigma(1 - w_j/\Sigma w_j)^2/(n_j - 1)}{J^2 - 1}$$

[20] En el capítulo 9 consideramos algunos estadísticos *no paramétricos* que, cuando se incumplen los supuestos del estadístico F del ANOVA, constituyen una alternativa robusta y potente.