

Algunas ideas sobre el ANOVA con 2 factores

1 Introducción

El modelo más simple para analizar es el aditivo:

$$y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk} \quad (i = 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, K) \quad (1)$$

con

$$\sum_i \alpha_i = \sum_j \beta_j = 0.$$

Desde ahora suponemos que las columnas representen tratamientos y las filas unidades o bloques. Generalmente interesa comparar los tratamientos.

Si uno consigue ajustar un modelo así, la inferencia es más fácil, y especialmente la descripción es más simple. Sea

$$\xi_{ij} = E y_{ijk}$$

. Notar que este modelo tiene la consecuencia de que

$$\xi_{ij} > \xi_{kj} \text{ para algún } j \Rightarrow \xi_{ij} > \xi_{kj} \quad \forall j \quad (2)$$

o sea que las consecuencias son fuertes.

El estimador de α_i es $\hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{...}$ donde “ $\bar{}$ ” indica promedio sobre el índice. Los valores ajustados son

$$\hat{\xi}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j. \quad (3)$$

Para evaluar visualmente la aditividad, se grafican para cada i las \bar{y}_{ij} en función de j (y lo mismo al revés).

La visibilidad mejora si se ordenan las j en orden creciente de $\hat{\beta}_j$, o mejor aún si se grafica \bar{y}_{ij} vs. $\hat{\beta}_j$. Si hay aditividad, las curvas son aproximadamente paralelas.

1.1 ¿Se puede convivir con la interacción?

Supongamos que tenemos una tabla que no tiene cruzamientos notorios, y podemos testear la interacción, y resulta significativa. ¿Qué hacer?. Si – digamos – las diferencias entre tratamientos son del orden de 10, y las interacciones del orden de 0.5, está claro que las interacciones no van a afectar las comparaciones entre tratamientos, por lo que podemos trabajar como si todo fuera aditivo.

Más precisamente: el modelo es ahora

$$\xi_{ij} = E y_{ij} = \mu + \alpha_i + \beta_j + \eta_{ij} \quad (4)$$

donde η_{ij} son las interacciones. Para comparar los tratamientos j y k necesito

$$\xi_{ij} - \xi_{ik} = (\beta_j - \beta_k) + (\eta_{ij} - \eta_{ik}),$$

y si $|\eta_{ij} - \eta_{ik}| \leq 0.5|\beta_j - \beta_k|$, entonces $\text{sign}(\xi_{ij} - \xi_{ik}) = \text{sign}(\beta_j - \beta_k)$.

Si la tabla es chica, puedo examinar las η_{ij} . Si no, un criterio es comparar las “mean squares”. Sean M_{int} y M_{trat} los cuadrados medios (suma de cuadrados dividida por sus grados de libertad) para interacciones y efectos, respectivamente. Entonces si $M_{\text{int}}/M_{\text{trat}}$ es “chico” (por ejemplo ≤ 0.1) puedo despreocuparme de las interacciones.

2 “No aditividad transformable”

Si (1) no vale, pero sí vale para una transformación monótona $g(y)$, entonces las distancias verticales entre curvas no son constantes, pero mantienen el signo, y suelen tener forma de embudo. Aquí sigue valiendo (2). Si dos curvas están muy próximas, inevitablemente se cruzarán, simplemente por aleatoriedad.

Entonces se puede tener aditividad aplicando el modelo (1) a $g(y)$. Lo más simple sería intentar con la familia $g(y) = y^\gamma$. Se pueden probar distintos valores de γ y quedarse con el que da mayor paralelismo. Una forma directa es usar Box-Cox; pero este método intenta conseguir muchas cosas al mismo tiempo, no necesariamente el mejor γ para Box-Cox será el que dé mayor aditividad.

Nada garantiza que la transformación que produce aditividad no dé residuos heteroscedásticos o asimétricos. Estamos en una situación de “la chancha y los 20”. Mi criterio es que lo importante es asegurarse la *forma adecuada* del modelo. En una situación general de regresión $Y = h(X) + e$, pienso que lo principal es modelar bien la h (la parte determinística del modelo) aunque e (la parte estocástica) dé luego trabajo.

Si los errores no parecen cumplir las condiciones ortodoxas, eso se puede remediar usando métodos de inferencia más seguros como los tests no paramétricos o el bootstrap.

3 No aditividad irremediable

El problema es cuando hay cruzamientos escandalosos entre curvas distantes. Aquí ya no vale (2).

Aquí se plantean dos problemas:

- a) testear si la no aditividad observada es real o producto del azar
- b) si es real, describirla

Si $K > 1$, se puede testear (a) contra alternativas generales. Si no, hay que usar alternativas más restringidas (ver más abajo). Cualquiera sea el método usado para testear, hay que cuidarse (como siempre) de no dejarse arrastrar por un test. Si las interacciones son “significativamente distintas de cero”, pero tan pequeñas que (2) sigue valiendo (o sea, son menores que las diferencias entre medias), no vale la pena renunciar al modelo aditivo. Y si son lo bastante

grandes como para tenerlas en cuenta, no basta con decir “hay interacciones”, sino que hay que tratar de describirlas. Frecuentemente pasa que (1) no vale para toda la tabla, pero sí separadamente para subtablas.

Si no hay aditividad, y tampoco se la puede conseguir transformando, no es el fin del mundo. Significa simplemente que (2) no es cierto, por lo que el tratamiento i puede ser mejor que el k en algunos casos, y peor en otros. Esto *forma parte de las conclusiones* a extraer de los datos.

4 El caso $K = 1$

Un análisis útil es ordenar las filas y las columnas en orden creciente de sus medias (o sus medianas), y mirar los residuos $r_{ij} = y_{ij} - \hat{\xi}_{ij}$ en su lugar en la tabla (cuando $K > 1$ se puede hacer con $\bar{y}_{ij} - \hat{\xi}_{ij}$).

Si el ajuste es bueno, no debiera haber estructura visible (salvo la que imponen las restricciones lineales que cumplen los residuos); en particular, los signos debieran aparecer desordenados. Pero pueden aparecer estructuras que indican la necesidad de introducir interacciones. En particular, si hay bloques de signos en los rincones de la tabla, es decir

$$\begin{array}{cccccc} + & + & + & - & - & - \\ + & + & + & - & - & - \\ - & - & - & + & + & + \\ - & - & - & + & + & + \end{array} \quad (5)$$

En estos casos, el gráfico de residuos vs. ajustados muestra una estructura marcada.

4.1 Modelos multiplicativos para la interacción

La forma más simple de modelar la interacción en (4) es suponer

$$\gamma_{ij} = a_i b_j \quad (6)$$

donde los a_i y b_j son parámetros a estimar. Este modelo es un caso particular de un método más general llamado “biplot” que se ve en Análisis Multivariado. Se pueden deducir los estimadores de mínimos cuadrados de a_i y b_j , pero no consideraremos este tema.

Una versión más simple de este esquema es “one degree of freedom for nonadditivity” de Tukey (1949) donde

$$\gamma_{ij} = \lambda \alpha_i \beta_j. \quad (7)$$

Es fácil verificar que bajo este modelo, los signos de los residuos tienen la forma (5).

Se puede probar que los estimadores de μ, α_i, β_j son los habituales. Entonces queda λ como la pendiente de una recta por el origen entre r_{ij} y $\hat{\alpha}_i \hat{\beta}_j$, y se lo

puede estimar con

$$\hat{\lambda} = \frac{\sum_i \sum_j r_{ij} \hat{\alpha}_i \hat{\beta}_j}{\sum_i \sum_j (\hat{\alpha}_i \hat{\beta}_j)^2}.$$

Se puede probar que la inferencia sobre λ es como para cualquier recta por el origen. Así se puede testear si $\lambda = 0$.

El modelo (7) se puede examinar visualmente graficando los IJ residuos r_{ij} contra los productos $\hat{\alpha}_i \hat{\beta}_j$. Si vale (7) el gráfico sería una recta por el origen, con pendiente λ . Otro examen es calcular los nuevos residuos

$$y_{ij} - (\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \lambda \hat{\alpha}_i \hat{\beta}_j)$$

y ver si la estructura ha desaparecido.

Como siempre, no hay que dejarse llevar por un valor-p. Lo que importa no es sólo si $\lambda \neq 0$, sino que sea lo bastante grande para ser útil. Si es así, los residuos de (7) deben ser mucho menores que los de (1). Esto se puede verificar comparando los respectivos cuadrados medios.

Supongamos que las interacciones resultan lo bastante grandes como para adoptar el modelo de Tukey. Sea $\xi_{ij} = Ey_{ij}$. Para comparar el tratamiento j con el k hacemos

$$\hat{\xi}_{ij} - \hat{\xi}_{kj} = (\hat{\beta}_j - \hat{\beta}_k) (1 + \lambda \hat{\alpha}_i). \quad (8)$$

Si valores altos indican “mejor”, entonces el tratamiento j será mejor que el k para las unidades i con $1 + \lambda \hat{\alpha}_i > 0$, y viceversa.

Es posible testear $\xi_{ij} - \xi_{kj} > 0$, pero no he hecho la cuenta detallada de cómo hacerlo ni la he visto en la literatura. Se puede hacer inferencia mediante el *bootstrap*, pero eso está lejos de este curso.

4.2 El modelo de partición

Es posible que el análisis exploratorio sugiera que existen dos regímenes, dados por dos subtablas aditivas, o sea que para algún m

$$Ey_{ij} = \begin{cases} \alpha_i + \beta_j^{(1)} & \text{si } i \leq m \\ \alpha_i + \beta_j^{(2)} & \text{si } i > m \end{cases},$$

donde $(\beta_1^{(k)}, \dots, \beta_J^{(k)})$, $k = 1, 2$ son dos conjuntos distintos de parámetros, con $\sum_j \beta_j^{(k)} = 0$.

Se puede probar que este modelo es un caso particular de (6).

Con los medios disponibles, si el modelo de Tukey da un buen ajuste, una forma barata de obtener la partición es separar las unidades i según el signo de $1 + \lambda \hat{\alpha}_i$ en (8). Generalmente esas subtablas suelen ser bastante aditivas, dado que por (8) no debería haber cruzamientos en cada subtabla. Este criterio, por ser sistemático, parece mejor que particionar a ojo.

Una vez elegidas las subtablas, se aplica el modelo aditivo en cada una y se hace la correspondiente inferencia.

Naturalmente, esta inferencia no es del todo correcta, porque las subtablas fueron elegidas en función de los datos, y en consecuencia las distribuciones de los estadísticos no son las mismas que si hubieran sido elegidas a priori. Pero me parece más útil dar conclusiones de validez aproximada para un modelo creible, que dar conclusiones supuestamente exactas para un modelo que no describe la realidad (recordar el concepto del “error de tipo III” de Tukey: dar la respuesta correcta para el problema errado).