Tema 2. Análisis de la varianza de una vía

1. Introducción

Los modelos que se considerarán en el análisis de la varianza para la comparación de diferentes tratamientos o niveles asociados a un sólo factor son:

- Modelo de efectos fijos
- Modelo de efectos aleatorios.

Estos modelos son útiles para la clasificación en un sólo sentido. La realización de los experimentos se considera en un orden aleatorio, es decir, se considera el diseño experimental completamente aleatorizado.

Se representerá la observación $j-$ésima del tratamiento i mediante el siguiente modelo estadístico lineal:

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i = 1, \dots, k; \quad j = 1, \dots, n, \tag{1}$$

siendo n el tamaño muestral o número de observaciones de la variable respuesta bajo el tratamiento i. Se supone que todos los tratamientos poseen un parámetro común μ que representa el promedio global. El efecto del tratamiento i se representa mediante el parámetro τ_i. Finalmente, la componente aleatoria de error se representa mediante la variables aleatorias $\epsilon_{ij},\ i = 1, \dots, k;\ j = 1, \dots, n$, que se suponen independientes y distribuidas según una normal con media cero y varianza σ^2. Se intentará estimar los efectos de los tratamientos, así como plantear hipótesis apropiadas sobre los mismos. Más concretamente, a partir del modelo (1), se pueden abordar dos situaciones respecto al efecto de los tratamientos:

- Si los tratamientos han sido seleccionados específicamente por el experimentador, se plantearán hipótesis sobre las medias de los tratamientos y las conclusiones se aplicarán a dichos tratamientos. No siendo extensivas a tratamientos similares que no hayan sido probados. Se estimarán asimismo los parámetros del modelo (μ, τ_i, σ^2). Este modelo se conoce con el nombre de *modelo de efectos fijos.*

- Por otra parte, se puede considerar una selección aleatoria de los tratamientos, partiéndose entonces de una muestra aleatoria de una población mayor de tratamientos. En este ámbito, se pueden transladar las conclusiones a todos los tratamientos de la población, independientemente de que hayan sido o no explícitamente considerados. En este caso, los parámetros τ_i son variables aleatorias. Por tanto, el objetivo primordial será investigar las características esenciales de su distribución. En particular, su variabilidad. De ahí que se planteen hipótesis sobre la varianza de los efectos. Este modelo se conoce con el nombre de modelo de *efectos aleatorios.*

2. Modelo de efectos fijos

Los efectos τ_i, $i = 1, \ldots, k$, de los k tratamientos considerados se interpretan como desviaciones respecto a la media global. Por tanto, se tiene la restricción

$$\sum_{i=1}^{k} \tau_i = 0.$$

El planteamiento de hipótesis de comparación entre los tratamientos se traduce entonces en hipótesis de comparación o igualdad entre los parámetros τ_i, $i = 1, \ldots, k$. Formalmente, el valor esperado del $i-$ésimo tratamiento se descomopone en la suma

$$E[y_{ij}] = \mu_i = \mu + \tau_i, \quad i = 1, \ldots, k,$$

de la media general y el efecto de dicho tratamiento, para cualquier $j = 1, \ldots, n$, es decir, son equivalentes los siguientes contrastes

$$\begin{aligned} H_0 : \mu_1 &= \cdots = \mu_k \\ H_1 : \mu_i &\neq \mu_j, \quad \text{para algún par } (i, j) \end{aligned}$$

$$
\begin{aligned}
H_0 : \tau_1 &= \cdots = \tau_k = 0 \\
H_1 : \tau_i &\neq 0, \quad \text{para algún par } (i, j).
\end{aligned}
$$

Mediante la técnica de análisis de la varianza se descompone la variabilidad total de los datos en términos de la variabilidad entre tratamientos y dentro de tratamientos. Formalmente, la suma total de cuadrados o variabilidad total viene dada por

$$
SS_T = \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \overline{y}_{..})^2, \tag{2}
$$

siendo $\overline{y}_{..}$ la media muestral global definida por

$$
\overline{y}_{..} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij},
$$

y denotando por $N = kn$. Nótese, que

$$
\frac{SS_T}{N-1} = \frac{SS_T}{kn-1}
$$

representa la varianza muestral de y. Tras operar en (2) se obtiene la siguiente descomposición:

$$
\begin{aligned}
\sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \overline{y}_{..})^2 &= n \sum_{i=1}^{k} (\overline{y}_{i.} - \overline{y}_{..})^2 + \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \overline{y}_{i.})^2 \\
&= SS_{\text{tratamientos}} + SS_{\text{error}}. \qquad (3)
\end{aligned}
$$

El primer término de la descomposición anterior proporciona una medida empírica de las diferencias entre las medias de los tratamientos. El segundo término de la descomposición representa una medida empírica de la variabilidad ocasionada por el error aleatorio asociado a la medición de cada tratamiento. El número total de observaciones involucradas en la suma total es $N = kn$, contabilizando, por tanto, $N - 1$ grados de libertad para SS_T y $k - 1$ grados de libertad para $SS_{\text{tratamientos}}$. Finalmente, en la estimación del error experimental para cada tratamiento, se tienen $n - 1$ grados de libertad, dado que se han considerado n réplicas y $k(n-1) = N-k$ grados de libertad en la estimación de SS_{error}. En realidad, $SS_{\text{error}}/(N-k)$ proporciona una estimación de la varianza poblacional σ^2 común de los k tratamientos. Similarmente, si las medias de los tratamientos coinciden el estadístico

$$\frac{SS_{\text{Tratamientos}}}{k-1}$$

proporciona una estimación de la varianza poblacional σ^2. Se tienen así dos estimaciones de la varianza poblacional respectivamente basadas en la variabilidad interna de los tratamientos y en la variabilidad entre tratamientos. Si el nivel medio de los tratamientos coincide, estas dos estimaciones deben ser similares. Efectivamente, se puede comprobar que

$$E\left(\frac{SS_{\text{Tratamientos}}}{k-1}\right) = \sigma^2 + \frac{n\sum_{i=1}^{k}\tau_i^2}{k-1} \tag{4}$$

$$\tag{5}$$

$$E\left(SS_{\text{error}}/(N-k)\right) = \sigma^2.$$

La ecuación (4) nos indica que el valor esperado de la media de cuadrados entre tratamientos coincide con la varianza poblacional σ^2 cuando los niveles medios de los tratamientos coinciden, es decir, cuando $\tau_i = 0$, para $i = 1, \ldots, k$.

Estudiaremos primeramente la distribución de los estadísticos involucrados en el contraste de comparación de niveles medios de los tratamientos. Puesto que los errores ϵ_{ij} son variables aleatorias independientes, normalmente distribuidas con media cero y varianza σ^2, las observaciones también son variables aleatorias normales, independientes con media $\mu + \tau_i$ para $i = 1, \ldots, k$, y varianza σ^2. Se demuestra entonces que SS_T/σ^2 sigue una distribución χ^2 con $N-1$ grados de libertad y SS_E/σ^2 sigue una distribución χ^2 con $N-k$ grados de libertad. Finalmente, si la hipótesis nula $H_0 : \tau_i = 0,\ i = 1, \ldots, k$, se cumple, entonces $SS_{\text{Tratamientos}}/\sigma^2$ sigue una distribución χ^2 con $k-1$ grados de libertad. Puesto que las tres sumas de cuadrados anteriores no son independientes, se aplica un resultado estadístico (Teorema de Cochran) que permite comprobar que SS_E/σ^2 y

$SS_{\text{Tratamientos}}/\sigma^2$ son variables aleatorias independientes. Por tanto, el estadístico

$$F_0 = \frac{SS_{\text{Tratamientos}}/k-1}{SS_E/N-k}$$

sigue una distribución F-Snedecor con $k-1$ y $N-k$ grados de libertad. Puesto que cuando H_0 es falsa el valor esperado del numerador del estadístico F_0 será superior a σ^2, que define el valor esperado del denominador, valores grandes de dicho estadístico proporcionan evidencia empírica para rechazar la hipótesis nula. Es decir, se realizará un diseño unilateral superior de la región crítica que rechazará H_0 si

$$F_0 > F_{\alpha,k-1,N-k},$$

para un tamaño α del test (α representa la probabilidad de rechazar H_0 siendo cierta), donde $F_{\alpha,k-1,N-k}$ es el percentil $1-\alpha$ de la distribución F-Snedecor con $k-1$ y $N-k$ grados de libertad.

Normalmente en el cálculo del estadístico F_0 se utilizarán las siguientes expresiones simplificadas de SS_T y $SS_{\text{Tratamientos}}$

$$\begin{aligned} SS_T &= \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} \\ SS_{\text{Tratamientos}} &= \sum_{i=1}^{k} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N}. \end{aligned} \tag{6}$$

Entonces se tiene que

$$SS_E = SS_T - SS_{\text{Tratamientos}}.$$

2.1. Estimación de los parámetros

Se aborda en esta sección la aproximación del modelo

$$y_{ij} = \mu + \tau_i + \epsilon_{ij},$$

que se supone subyace a los datos, mediante la estimación de sus parámetros. Dicha estimación se realizará por mínimos-cuadrados, es decir, se determinan los valores $\hat{\mu}$ y $\hat{\tau}_i$, para $i = 1, \ldots, k$, que minimizan la función de pérdida

$$L = \sum_{i=1}^{k} \sum_{j=1}^{n} \epsilon_{ij}^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \mu - \tau_i)^2.$$

Para el cálculo de dichos valores se deriva L respecto a μ y a τ_i, $i = 1, \ldots, k$. Igualando a cero las derivadas se obtiene

$$\begin{aligned} -2 \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \widehat{\mu} - \widehat{\tau}_i) &= 0 \\ -2 \sum_{j=1}^{n} (y_{ij} - \widehat{\mu} - \widehat{\tau}_i) &= 0, \quad i = 1, \ldots, k. \end{aligned} \tag{7}$$

Operando y simplificando, considerando a restricción

$$\sum_{i=1}^{k} \widehat{\tau}_i = 0, \tag{8}$$

se obtienen las siguientes soluciones

$$\begin{aligned} \widehat{\mu} &= \overline{y}_{..} \\ \widehat{\tau}_i &= \overline{y}_{i\cdot} - \overline{y}_{..}, \quad i = 1, \ldots, k. \end{aligned} \tag{9}$$

Nótese que las ecuaciones dadas en (7) no son independientes, y por tanto, no existe una solución única. Se ha considerado pues la restricción (8) que tiene sentido a partir de la interpretación de los efectos de los tratamientos como desviaciones de la media general. Independientemente de que la solución obtenida no es única y depende de la restricción considerada, hay funciones de los parámetros que sí se estiman de forma única y que son fundamentales para la comparación de los tratamientos, que era nuestro objetivo inicial, tal es el caso de la diferencia entre la media de los tratamientos. Este hecho se utilizará para la formulación de intervalos de confianza que vendrán dados en términos de funciones estimables de los parámetros. Específicamente, en la estimación por intervalos de confianza del nivel medio del $i-$ésimo tratamiento, puesto que

$$\widehat{\mu_i} = \hat{\mu} + \widehat{\tau}_i = \overline{y}_{i\cdot},$$

bajo la hipótesis de normalidad de los errores aleatorios, se tiene que $\overline{y}_{i\cdot}$ se distribuye según una normal con media μ_i y varianza σ^2/n. Considerando como estimación de σ^2 la media de cuadrados dentro de cada tratamiento, se tiene que el estadístico $\overline{y}_{i\cdot}/\sqrt{SS_E/n(N-k)}$ se distribuye según una $t-$Student con $N-k$ grados de libertad. Por tanto, un intervalo de confianza a nivel de confianza $1-\alpha$ vendría dado por la siguiente expresión:

$$\left[\overline{y}_{i\cdot} \pm t_{\alpha/2,N-k}\sqrt{SS_E/n(N-k)}\right],$$

siendo $t_{\alpha/2,N-k}$ el percentil $1-\alpha/2$ de la distribución $t-$Student con $N-k$ grados de libertad. Dada la independencia entre los diferentes tratamientos, de forma similar se construye un intervalo de confianza para la diferencia de niveles medios de los tratamientos que vendrá dado por

$$\left[\overline{y}_{i\cdot} - \overline{y}_{j\cdot} \pm t_{\alpha/2,N-k}\sqrt{2SS_E/n(N-k)}\right].$$

En todo el desarrollo estadístico anterior debe tenerse en consideración que la hipótesis de normalidad e independencia de los errores aleatorios $\epsilon_{ij},\ i=1,\ldots,k; j=1,\ldots,n$, que se distribuyen de forma idéntica con media cero y varianza σ^2, debe cumplirse. En caso contrario, el análisis estadístico anterior no es válido. Por tanto, en la Sección 4 de este tema se describirá brevemente el análisis de los residuos para investigar la validez del modelo

$$y_{ij} = \mu + \tau_i + \epsilon_{ij},$$

bajo las hipótesis anteriores. Así, la fórmula del residuo

$$\epsilon_{ij} = y_{ij} - \widehat{y}_{ij} = y_{ij} - \hat{\mu} - \widehat{\tau}_i = \overline{y}_{\cdot\cdot} + (\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot}) = \overline{y}_{i\cdot}$$

constituirá una herramienta clave en dicho análisis.

2.2. Tamaños muestrales desiguales

En algunas ocasiones el número de observaciones disponibles para cada tratamiento es diferente. Las fórmulas de análisis de la varianza deben ser adaptadas a esta situación de la siguiente forma: Denotando por n_i el número de observaciones recolectadas del tratamiento i, para $i = 1, \ldots, k$, y por $N = \sum_{i=1}^{k} n_i$, se tiene que

$$\begin{aligned} SS_T &= \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} \\ SS_{\text{Tratamientos}} &= \sum_{i=1}^{k} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N}. \end{aligned} \tag{10}$$

En general, es conveniente disponer de un diseño equilibrado donde los tamaños muestrales de todos los tratamientos coincidan para que el análisis estadístico sea más robusto frente a pequeas desviaciones de la hipótesis de igualdad de varianzas de los diferentes tratamientos, así como para obtener una mayor potencia de la prueba de igualdad.

2.3. Métodos de comparación múltiple

Cuando la hipótesis de igualdad entre los niveles medios de los tratamientos se rechaza, es fundamental detectar dónde se ubican las diferencias. Es decir, cuáles son los tratamientos que producen efectos diferentes en la variable respuesta. Para ello se diseñan diferentes contrastes de hipótesis que permiten realizar las comparaciones pertinentes. Generalmente, el planteamiento de contrastes de comparación múltiple de medias de tratamientos se basan en estadísticos definidos a partir de combinaciones lineales de totales de tratamientos, es decir, se construten a partir de estadísticos de la forma

$$C = \sum_{i=1}^{k} c_i y_{i\cdot},$$

siendo $\sum_{i=1}^{k} c_i = 0$. Por ejemplo, si se quiere contrastar si los niveles medios de los tratamientos i y j coinciden, se considerará la combinación lineal $y_{i\cdot} - y_{j\cdot}$, contrastando su igualdad a cero. En otras ocasiones, la comparación se formulará entre grupos de tratamientos, por ejemplo, los tratamientos i, j y k producen el mismo efecto que los tratamientos l, m y n, en este caso, se considerará la combinación lineal

$$y_{i\cdot} + y_{j\cdot} + y_{k\cdot} - y_{l\cdot} - y_{m\cdot} - y_{n\cdot},$$

contrastándose la igualdad a cero de dicha combinación. Usualmente, puesto que cada combinación define la formulación de un posible test de comparación, dichas combinaciones se denominan contrastes. La suma de cuadrados asociada a un contraste se define entonces como:

$$SS_C = \frac{\left(\displaystyle\sum_{i=1}^{k} c_i y_{i\cdot}\right)^2}{n\displaystyle\sum_{i=1}^{k} c_i^2},$$

con un grado de libertad. En el caso de tamaños muestrales desiguales se considera la restricción

$$\sum_{i=1}^{k} n_i c_i = 0,$$

y la suma de cuadradrados viene dada por

$$SS_C = \frac{\left(\displaystyle\sum_{i=1}^{k} c_i y_{i\cdot}\right)^2}{\displaystyle\sum_{i=1}^{k} n_i c_i^2}.$$

El test de comparación se formula en términos del cociente entre la suma de cuadrados del contraste y la media de cuadrados del error. Se tiene así un estadístico con distribución $F-$Snedecor con 1 y $N-k$ grados de libertad.

En el diseño de contrastes de comparación, un criterio puede ser la ortogonalidad entre los contrastes, es decir, dados dos contrastes definidos mediante los pesos $c_i,\ i = 1, \ldots, k$ y $d_i,\ i = 1, \ldots, k$, se considera la restricción

$$\sum_{i=1}^{k} c_i d_i = 0,$$

o bien

$$\sum_{i=1}^{k} n_i c_i d_i = 0,$$

para tamaños muestrales desiguales. Dichos contrastes serán pues referidos como contrastes ortogonales. Con este procedimiento se pueden considerar $k-1$ contrastes ortogonales que descompongan en $k-1$ componentes independientes, con un sólo grado de libertad, la suma de cuadrados asociada a la variabilidad entre tratamientos. Obteniéndose así pruebas independientes asociadas a cada unos de los contrastes. La selección de coeficientes para el diseño de contrastes ortogonales debe ser previa a la observación de los datos, puesto que el planteamiento de dichos contrastes siguiendo únicamente las diferencias empíricas de mayor magnitud entre totales (o promedios) de los mismos produce un incremento del error tipo 1 del test. Para mantener un tamaño fijo del test u error tipo I en todas las posibles comparaciones, el método de Sceffé proporciona el siguiente diseño: Para m comparaciones de interés sobre las medias de los k tratamientos, dadas por

$$\Gamma_T = \sum_{i=1}^{k} c_{iT}\mu_i,$$

con $T = 1, \ldots, m$, se tienen los contrastes correspondientes, en términos de los promedios empíricos para cada tratamiento,

$$C_T = \sum_{i=1}^{k} c_{iT}\bar{y}_{i\cdot}, \quad T = 1, \ldots, m,$$

siendo el error estándar de cada contraste

$$S_{C_T} = \sqrt{(SS_E/(N-k))\sum_{i=1}^{k} c_{iT}^2/n_i},$$

donde, como antes, n_i denota el número de observaciones del i-ésimo tratamiento para $i = 1, \ldots, k$. El diseño de la región crítica viene entonces dado en términos del valor

$$S_{\alpha,T} = S_{C_T}\sqrt{(k-1)F_{\alpha,k-1,N-k}}.$$

Equivalentemente, con este procedimiento se rechaza la igualdad a cero de la combinación de niveles medios definida en el contraste Γ_T si se verifica

$$|C_T| > S_{\alpha,T},$$

para cada $T = 1, \ldots, m$. Los intervalos de confianza simultáneos definidos mediante el método de Sceffé vienen dados entonces por

$$C_T - S_{\alpha,T} \leq \Gamma_T \leq C_T + S_{\alpha,T}, \quad T = 1, \ldots, m,$$

donde la probabilidad de que las desigualdades anteriores sean verdaderas simultáneamente es al menos $1 - \alpha$.

Finalmente, introduciremos el planteamiento de cuatro procedimientos para contrastar la iguadad de medias de tratamientos por parejas. Es decir,

$$H_0 : \mu_i = \mu_j, \; i \neq j.$$

- *Mínima diferencia significativa*

  Estadístico
  $$t_0 = \frac{\bar{y}_{i\cdot} - \bar{y}_{j\cdot}}{\sqrt{(SS_E/(N-k))\,(1/n_i + 1/n_j)}}.$$

  Valor crítico (mínima diferencia significativa)

  $$MDS = t_{\alpha/2,N-k}\sqrt{(SS_E/(N-k))\,(1/n_i + 1/n_j)}.$$

  Procedimiento: Se rechaza H_0 si

  $$|y_{i\cdot} - y_{j\cdot}| > MDS.$$

- *Prueba de intervalos múltiples de Duncan*

  Estadístico
  $$D_0 = \frac{\bar{y}_{(i)\cdot} - \bar{y}_{(j)\cdot}}{\sqrt{SS_E/n_h(N-k)}}, \quad i > j,$$

  donde $n_h = \dfrac{k}{\displaystyle\sum_{i=1}^{k}(1/n_i)}$, y denotando por $\bar{y}_{(i)\cdot}$, $i = 1, \ldots, k$, las medias muestrales de los tratamientos en orden ascendente, es decir, ordenadas de menor a mayor (por orden de magnitud), $\bar{y}_{(i)\cdot} > \bar{y}_{(j)\cdot}$ si $i > j$.

  Valor crítico

  $$R_q = r_\alpha(q, f)\sqrt{SS_E/n_h(N-k)}, \quad q = 2, \ldots, k,$$

donde $r_\alpha(q,f)$ se calcula a partir de la tabla de intervalos significativos de Duncan para un nivel de significación α y un número f de grados de libertad del error.

Procedimiento: Se compara primeramente la diferencia entre el valor medio más alto $\bar{y}_k$ y el más pequeño $\bar{y}_1$ frente al intervalo mínimo significativo R_k. Posteriormente, la diferencia $\bar{y}_k - \bar{y}_2$ frente a R_{k-1} y así se continua comparando sucesivamente la media mayor $\bar{y}_k$ con las restantes de forma ordenada. Se continua con las diferencias entre $\bar{y}_{k-1}$ y las restantes medias, de forma ordenada, comenzando con $\bar{y}_1$ y comparando con R_{k-1} y así sucesivamente se van realizando las $k(k-1)/2$ comparaciones. En este procedimiento, si una pareja de medias se ubica entre dos parejas de medias que no difieren significativamente, dicha pareja de medias se considera que no difiere significativamente.

- *Prueba de Newman-Keuls*

  Estadístico
$$NK_0 = \frac{\bar{y}_{(i)\cdot} - \bar{y}_{(j)\cdot}}{\sqrt{SS_E/n_h(N-k)}}, \quad i > j,$$
  donde $n_h = \dfrac{k}{\sum_{i=1}^{k}(1/n_i)}$, y denotando por $\bar{y}_{(i)\cdot}$, $i = 1,\dots,k$, las medias muestrales de los tratamientos en orden ascendente, es decir, ordenadas de menor a mayor (por orden de magnitud), $\bar{y}_{(i)\cdot} > \bar{y}_{(j)\cdot}$ si $i > j$.

  Valor crítico
$$K_q = k_\alpha(q,f)\sqrt{SS_E/n_h(N-k)}, \quad q = 2,\dots,k,$$
  donde $k_\alpha(q,f)$ representa el percentil $1-\alpha$ de la distribución de
$$k(q,f) = \frac{\bar{y}\text{max} - \bar{y}_{\text{min}}}{\sqrt{SS_E/n_h(N-k)}},$$
  denotando por $\bar{y}_{\text{max}}$ e $\bar{y}_{\text{min}}$ los valores máximo y mínimo del grupo de q medias muestrales, siendo f los grados de libertad del error.

  Procedimiento: Es similar al anterior exceptuando el cálculo de las diferencias críticas, según se ha visto en el punto anterior.

- *Prueba de Tukey*

  Estadístico

  $$TU_0 = \frac{\bar{y}_{(i)\cdot} - \bar{y}_{(j)\cdot}}{\sqrt{SS_E/n_h(N-k)}}, \quad i > j,$$

  Valor crítico

  $$T_\alpha = q_\alpha(k,f)\sqrt{SS_E/n_h(N-k)},$$

  donde $q_\alpha(k,f)$ se define como en la prueba de Newman-Keuls.

  Procedimiento: Se tiene la misma metodología que en la prueba anterior. La única diferencia es que la distancia crítica es constante, dado que el valor crítico $q_\alpha(k,f)$ no depende del número de medias que estén en un grupo.

En algunas ocasiones interesa comparar cada uno de los tratamientos con un tratamiento control. En este caso, se realizan $k-1$ comparaciones

$$H_0 : \mu_i = \mu_k; \quad H_1 : \mu_i \neq \mu_k, \quad i = 1, \ldots, k-1.$$

Las comparaciones se realizan en términos de una modificación del test $t-$Student, considerando las diferencias

$$|\bar{y}_{i\cdot} - \bar{y}_{k\cdot}|, \quad i = 1, \ldots, k-1,$$

donde si se verifica que

$$|\bar{y}_{i\cdot} - \bar{y}_{k\cdot}| > d_\alpha(k-1,f)\sqrt{SS_E/(N-k)\,(1/n_i + 1/n_k)},$$

la hipótesis nula sobre la igualdad de los niveles medios de los tratamientos i y k (de control) es rechazada para un tamaño α del test.

3. Modelo de efectos aleatorios

El planteamiento del modelo de efectos aleatorios se produce cuando de una población amplia de tratamientos se selecciona una muestra aleatoria de

k tratamientos para inferir resultados sobre la población global de tratamientos asociados a un factor. El modelo estadístico lineal se formula entonces como sigue

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i = 1, \ldots, k; \quad j = 1, \ldots, n,$$

siendo ahora τ_i y ϵ_{ij} variables aleatorias. Se supone que dichas variables son independientes con varianzas σ_τ^2 y σ^2 respectivamente. La varianza de una observación

$$V(y_{ij}) = \sigma_\tau^2 + \sigma^2$$

se descompone entonces en la suma de las componentes de la varianza, σ_τ^2 y σ^2.

En el modelo de efectos aleatorios se sigue verificando la descomposición de la variabilidad total (SS_T) de las observaciones en términos de la variabilidad entre los tratamientos ($SS_{\text{tratamientos}}$) y variabilidad dentro de tratamientos (SS_E).

El contraste de igualdad de los tratamientos se realizará en términos de las varianzas de los efectos, puesto que $\sigma_\tau^2 = 0$ indica que todos los tratamientos son idénticos. En cambio, $\sigma_\tau^2 > 0$ indica la existencia de variabilidad entre los tratamientos. Se contrastará pues

$$H_0 : \sigma_\tau^2 = 0; \quad H_1 : \sigma_\tau^2 > 0.$$

Nuevamente, bajo la hipótesis nula, el estadístico

$$F_0 = \frac{SS_{\text{tratamientos}}/k-1}{SS_E/N-k}$$

sigue una distribución $F-$Snedecor con $k-1$ y $N-k$ grados de libertad. La región crítica del test, así como la validez del estadístico F_0 para contrastar la hipótesis nula referida se deriva de los siguientes valores esperados:

$$\begin{aligned} E\left[SS_{\text{tratamientos}}/k-1\right] &= \sigma^2 + n\sigma_\tau^2 \\ E\left[SS_E/N-k\right] &= \sigma^2. \end{aligned}$$

Efectivamente, de las ecuaciones anteriores se tiene que bajo la hiótesis alternativa H_1, el valor esperado del numerador es mayor que el valor esperado del denominador. Por tanto, se rechazará H_0 para valores grandes de F_0, considerándose una región crítica unilateral superior.

De lo anterior se deducen las siguientes expresiones de los estimadores de las componentes de la varianza

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{SS_E}{N-k} \\
\hat{\sigma}_\tau^2 &= \frac{\left(SS_{\text{tratamientos}}/k-1\right) - (SS_E/N-k)}{n}.
\end{aligned} \tag{11}$$

Cuando el número de réplicas de cada experimento es diferente, en la ecuación (11) se reemplaza n por

$$\widetilde{n} = \frac{1}{k-1}\left[\sum_{i=1}^{k} n_i - \frac{\sum_{i=1}^{k} n_i^2}{\sum_{i=1}^{k} n_i}\right].$$

Las estimaciones anteriores de las componentes de la varianza son válidas incluso cuando la hipótesis de normalidad no se cumple (aunque sí la de independencia). De hecho, estos estimadores son insesgados y óptimos dentro del grupo de funciones cuadráticas de las observaciones. Un valor negativo de $\hat{\sigma}_\tau^2$ puede interpretarse como un valor real nulo o bien, que el modelo considerado no es adecuado.

En relación con la definición de intervalos de confianza para las componentes de la varianza, consideraremos las siguientes distribuciones:

$$\begin{aligned}
\frac{SS_E}{\sigma^2} &\sim \chi^2(N-k) \\
\frac{\left(SS_{\text{tratamientos}}/k-1\right)/(n\sigma_\tau^2+\sigma^2)}{SS_E/(N-k)\sigma^2} &\sim F(k-1, N-k).
\end{aligned} \tag{12}$$

A partir de dichas distribuciones se formulan los intervalos de confianza

$$\begin{aligned}
\frac{SS_E}{\chi^2_{\alpha/2,N-k}} &\leq \sigma^2 \leq \frac{SS_E}{\chi^2_{1-\alpha/2,N-k}} \\
L &\leq \frac{\sigma_\tau^2}{\sigma^2} \leq U,
\end{aligned} \tag{13}$$

para un nivel de confianza $1-\alpha$, siendo

$$\begin{aligned} L &= \frac{1}{n}\left(\frac{SS_{\text{tratamientos}}/k-1}{SS_E/(N-k)}\frac{1}{F_{\alpha/2,k-1,N-k}}-1\right) \\ U &= \frac{1}{n}\left(\frac{SS_{\text{tratamientos}}/k-1}{SS_E/(N-k)}\frac{1}{F_{1-\alpha/2,k-1,N-k}}-1\right). \end{aligned} \tag{14}$$

De la ecuación anterior se deduce un intervalo de confianza para la proporción $\sigma_\tau^2/(\sigma_\tau^2+\sigma^2)$

$$\frac{L}{1+L} \leq \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2} \leq \frac{U}{1+U}.$$

4. Validez del modelo

En esta sección se proporciona un breve introducción sobre los métodos estadísticos usuales para la comprobación de las hipótesis que definen los modelos de efectos fijos y efectos aleatorios, es decir, para verificar que el modelo

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i=,\dots,k, \quad j=1,\dots,n,$$

(en los dos casos referidos) describe de forma apropiada los datos. Asimismo, se consideran algunas alternativas cuando dichas hipótesis no se cumplen.

En relación con la verificación de la hipótesis de normalidad de los residuos $\epsilon_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i\cdot}$, como métodos gráficos, se pueden considerar la construcción de un histograma de los residuos, la representación de la distribución acumulada de los residuos sobre papel de probabilidad normal. Las desviaciones moderadas de la normalidad en el análisis de la varianza para el modelo de efectos fijos no son excesivamente importantes puesto que la prueba F de comparación se ve poco afectada por dichas desviaciones, así como los test de comparaciones múltiples. En este sentido se puede decir que el análisis de la varianza en el modelo de efectos fijos es un método robusto frente a, por ejemplo, desviaciones moderadas o distribuciones sesgadas. (No es así cuando se presentan colas empíricas más pesadas para los errores). Sin embargo, el modelo de efectos aleatorios se ve bastante más afectado por la violación de dicha hipótesis, especialmente en la estimación por intervalos de confianza de las componentes de la varianza. La presencia de *outliers* en los residuos debe ser analizada e interpretada cuidadosamente y nunca deben ser eleminados del análisis sin una investigación o justificación real previa. Su detección se puede realizar mediante el análisis de los residuos estandarizados $d_{ij} = \epsilon_{ij}/(\sqrt{SS_E/N-k})$, puesto que, bajo la hipótesis de normalidad, deben encontrarse, prácticamente todos, dentro de los límites de $\pm 3\sigma$.

La representación gráfica secuencial de los residuos, de acuerdo al orden temporal en el que se recogieron los datos, es útil para la detección de correlaciones entre ellos. Por ejemplo, la presencia de rachas positivas o negativas alerta sobre la posible violación de la hipótesis de independencia. También se puede detectar, mediante dicha graficación, el incremento de dispersión en el tiempo, debido a un comportamiento más errático de los residuos y, por tanto, a un cambio de la varianza en el tiempo. Adicionalmente, para detectar varianzas cambiantes, un recurso es representar los residuos frente a los valores ajustados $\hat{y}_{ij} = \bar{y}_{i\cdot}$. Esta gráfica no debe mostrar ningún patrón. Por ejemplo, si se aprecia un aumento de la varianza de las observaciones cuando aumenta su magnitud, se produce una violación de la hipótesis de varianza constante que puede ser ocasionada por un sesgo en la distribución y una violación de la hipótesis de normalidad. La violación de la hipótesis de varianza constante afecta en menor grado a la prueba F en el caso de modelo de efectos fijos con tamaño muestral constante. Sin embargo, para tamaño muestral variable, las varianzas cambiantes afectan más seriamente. El modelo de efectos aleatorios se ve seriamente afectado por la presencia de varianzas cambiantes que dificultan el desarrollo de la inferencia sobre las componentes de la varianza. Para resolver este problema se suele realizar una transformación de los datos que iguale las varianzas. Las conlusiones e interpretaciones del análisis de la varianza se aplicarán a las poblaciones transformadas.

4.1. Control sobre la hipótesis de varianza constante y transformaciones muestrales

Para el estudio de la hipótesis de varianza constante a partir de los datos se puede plantear un test donde se contraste la igualdad de las varianzas de todos los tratamientos. Es decir,

$$H_0 : \sigma_1^2 = \cdots = \sigma_k^2.$$

Se puede considerar el siguiente estadístico con distribución χ^2 :

$$\chi_0^2 = 2{,}3026\frac{q}{c},$$

siendo

$$\begin{aligned}
q &= (N-k)\log_{10} S_p^2 - \sum_{i=1}^{k}(n_i-1)\log_{10} S_i^2 \\
c &= 1+\frac{1}{3(k-1)}\left(\sum_{i=1}^{k}(n_i-1)^{-1}-(N-k)^{-1}\right) \\
S_p^2 &= \frac{\sum_{i=1}^{k}(n_i-1)S_i^2}{N-k}, \qquad (15)
\end{aligned}$$

donde S_i^2 es la varianza muestral de la $i-$ésima población. Dado que cuando existen grandes diferencias entre las varianzas muestrales, el valor de q es grande y es nulo cuando todas las varianzas muestrales son iguales, se rechazará H_0 para valores grandes de χ_0^2, es decir, si

$$\chi_0^2 > \chi_{\alpha,k-1}^2.$$

Como anteriormente, $\chi_{\alpha,k-1}^2$ denota el percentil $1-\alpha$ de la distribución χ^2 con $k-1$ grados de libertad. Este contraste es muy sensible a la violación de la hipótesis de normalidad. Por tanto, habrá que verificar previamente que dicha hipótesis se cumple.

La determinación de transformaciones de los datos que permitan estabilizar la varianza se puede inducir a partir de la relación funcional entre la media y la varianza. Por ejemplo, si la desviación estándar es proporcional a alguna potencia de la media poblacional

$$\sigma \propto \mu^{\alpha},$$

considerando una transformación f potencial de los datos, es decir,

$$f(y) = y^{\lambda},$$

se tiene entonces que

$$\sigma_{f(y)} = \mu^{\lambda+\alpha-1}.$$

Por tanto, para $\lambda = 1 - \alpha$ se tiene una transformación f de los datos que proporciona varianza constante. La estimación de α a partir de los datos se puede realizar a partir de la ecuación

$$\log \sigma_{y_i} = \log C + \alpha \log \mu_i.$$

La estimación de por máxima verosimilitud del exponente λ que define la tranformación de los datos puede realizarse, junto con la estimación de la media general y los efectos de los tratamientos, a partir del análisis de la varianza para diferentes valores de λ sobre el modelo

$$\begin{aligned} y^{(\lambda)} &= \frac{y^{\lambda} - 1}{\lambda \bar{y}_G^{\lambda-1}}, \quad \lambda \neq 0, \\ y^{(\lambda)} &= \bar{y}_G \ln y, \quad \lambda = 0, \end{aligned} \tag{16}$$

donde $\bar{y}_G$ denota la media geométrica. La estimación por máxima verosimilitud de λ se obtiene a partir del valor que minimiza la suma de cuadrados del error $SS_E(\lambda)$.

4.2. Otros aspectos importantes en el diseño

En el ajuste de un modelo apropiado para los datos se deben tener en cuenta algunos otros aspectos adicionales a los señalados en las secciones anteriores tales como:

- Detección de variables adicionales que puedan afectar a la respuesta.
- Detección de los niveles o tratamientos que afectan a la variabilidad de la respuesta (efectos potenciales de dispersión).
- Selección del tamaño muestral a partir de
  - Curvas características de operación que representan la probabilidad de error tipo II frente a una cantidad relacionada con el parámetro que define la distribución del estadístico del contraste F bajo la hipótesis alternativa, para un tamaño muestral fijo. Se consideran tamaños que proporcionen curvas operativas más discriminativas.
  - Precisión deseada en la estimación por intervalos de confianza.

Ajuste de superficies de respuesta. Enfoque de regresión

Este aspecto será estudiado en el Tema 3. Básicamente, el ajuste de ecuaciones a los datos para poder realizar una interpolación entre niveles se desarrolla a partir del enfoque empleado en el análisis estadístico mediante regresión. Bajo este enfoque, se puede realizar un análisis de la varianza. La ecuación que define la regresión puede ser lineal o no lineal, unidimensional o multidimensional, etc. Todos estos aspectos serán estudiados con detalle en el tema señalado.