Modelos Generales Lineales: ANOVA

Fundamento del ANOVA

El ANOVA es una técnica estadística que descompone la varianza de una variable aleatoria X con distribución normal en sus componentes, a los que se conoce como fuentes de variación. Una de estas fuentes de variabilidad será la varianza intragrupo, la otra la varianza entre grupos.

El objetivo del investigador que aplica el ANOVA a su trabajo experimental es comparar las medias poblacionales, tal que:

H0: $\mu_1 = \mu_2 = \ldots = \mu_k$
Ha: al menos dos medias son diferentes

Ahora bien, el contraste de medias es realizado estudiando o analizando las dos fuentes de variación anteriormente mencionadas.

Uno de los supuestos básicos del ANOVA es que la varianza de la variable X en cada grupo experimental o **varianza intragrupo** es similar en todas las poblaciones (o grupos experimentales):

$$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2$$

siendo por tanto σ^2 la varianza común a las *k* poblaciones. Cuando un tratamiento es eficaz el resultado es que la variabilidad total de la variable X no sólo viene explicada por la varianza común o intragrupo, sino además por las diferencias existentes entre las poblaciones como consecuencia del efecto de los difrentes niveles de los tratamientos. Llamaremos a esta varianza como **varianza entre grupos**.

El ANOVA se basa en que tanto la varianza intragrupo como la varianza entre grupos son dos estimaciones independientes de la varianza total de la variable X. Por consiguiente, el test o prueba estadística del ANOVA consiste en la comparación de las dos estimaciones por medio de una prueba o test de la *F*:

$$F = \frac{\sigma^2_{ENTRE\ \ GRUPOS}}{\sigma^2_{INTRAGRUPO}}$$

y el siguiente contraste de hipótesis:

$$\text{H0: } \sigma^2_{ENTRE\ \ GRUPOS} = \sigma^2_{INTRAGRUPO}$$
$$\text{Ha: } \sigma^2_{ENTRE\ \ GRUPOS} \neq \sigma^2_{INTRAGRUPO}$$

Si en el contraste aceptamos la hipótesis nula H0 concluiremos que no hay diferencias entre las dos estimaciones de la varianza poblacional, y por tanto no existen diferencias significativas entre los efectos de los distintos niveles de un tratamiento. Es decir, aceptaremos:

$$\text{H0: } \mu_1 = \mu_2 = \ldots = \mu_k$$

Por el contrario, si en el ANOVA rechazamos la H0 en el *F*-test aceptando Ha esto significa que el valor de la varianza total estimado por $\sigma^2_{ENTRE\ \ GRUPOS}$ se dispara por encima del valor estimado por $\sigma^2_{INTRAGRUPO}$ como consecuencia de los efectos significativos de los distintos niveles de un tratamiento, y por tanto aceptaremos que:

Ha: al menos dos medias son diferentes

Modelo de ANOVA de 1 vía

A efectos prácticos el modelo de ANOVA de 1 vía estima la varianza entre grupos y la varianza intragrupo utilizando expresiones que reciben el nombre de *media de cuadrados*. Las medias de cuadrados tienen la forma de un cociente, el numerador recibe el nombre de suma de cuadrados y el denominador corresponde a lo que se conoce como grados de libertad.

En primer lugar introduciremos la siguiente nomenclatura: n es el número total de observaciones o datos, $\overline{x}$ la media aritmética de todas las observaciones, k es el número de grupos experimentales, n_i el número de observaciones de un grupo experimental i sometido al tratamiento *i*, $\overline{x_i}$ la media aritmética de las observaciones del grupo experimental *i*, *T* es la suma de todas las observaciones y T_i la suma de las observaciones del grupo experimental *i*.

En segundo lugar, y con la siguiente expresión, se calcula la suma de cuadrados total (SCTotal):

$$\text{SCTotal} = \sum_{i}^{k}\sum_{j}^{n} x_{ij}^{2} - \frac{\left(\sum_{i}^{k}\sum_{j}^{n} x_{ij}\right)^{2}}{n}$$

En tercer lugar se obtiene la suma de cuadrados de los tratamientos o SCT:

$$\text{SCT} = \frac{\sum_{i}^{k} T_i^2}{n_i} - \frac{\left(\sum_{i}^{k}\sum_{j}^{n} x_{ij}\right)^{2}}{n}$$

En cuarto y último lugar se calcula la suma de cuadrados del error (SCE). Su valor se obtiene a partir de las expresiones anteriores, ya que el modelo de ANOVA de 1 vía establece que:

SCTotal = SCT + SCE

Por consiguiente, tendremos que:

SCE = SCTotal – SCT

A partir de las sumas de cuadrados se obtienen las medias de cuadrados, es decir las expresiones que permiten estimar tanto la varianza entre grupos o $\sigma^2_{ENTRE\ \ GRUPOS}$ como la varianza intragrupo o $\sigma^2_{INTRAGRUPO}$. La expresión que permite estimar a la varianza entre grupos se le denomina como *media de cuadrados de los tratamientos* o MCT:

$$\text{MCT} = \frac{\text{SCT}}{k-1}$$

La expresión que estima a la varianza intragrupo se conoce como *media de cuadrados del error* o MCE:

$$\mathrm{MCE} = \frac{\mathrm{SCE}}{n - k}$$

El modelo de ANOVA concluye obteniéndose el siguiente cociente con distribución F de Fisher y $\nu_1 = k - 1$, $\nu_2 = n - k$ grados de libertad:

$$F = \frac{\mathrm{MCT}}{\mathrm{MCE}}$$

que permite realizar el siguiente contraste:

H0: $\mu_1 = \mu_2 = \ldots = \mu_k$
Ha: al menos dos medias son diferentes

Rafael Lahoz-Beltrá, Pilar López González-Nieto, Mariángeles Gómez Flechoso, María Eugenia Arribas, Mocoroa, Alfonso Muñoz Martín, María de la Luz García Lorenzo, Gloria Cabrera Gómez, Jose Antonio Alvarez Gómez, Andrea Caso Fraile, Jefferson Mark Orosco Dagan, Raul Merinero Palomar. Universidad Complutense de Madrid, 2017.

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivar 4.0 Internacional.