ANALISIS DE VARIANZA

Como se menciono anteriormente, la forma general de esta prueba es:

$$H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_k$$

$$H_1\colon No\ todas\ las\ medias\ poblacionales\ son\ iguales$$

Donde μ_j es igual a la media de la j-ésima población. Para un diseño completamente aleatorizado, cada una de las *k* poblaciones o **tratamientos** se toma una muestra aleatoria simple de tamaño n_j. En cada muestra los datos son los siguientes:

El valor de la observación *i* del tratamiento *j* es igual a x_{ij}, donde el numero de observaciones es en el tratamiento *j* se denota por n_j. La media muestral del tratamiento *j* es $\bar{x}_j$, la varianza muestral del tratamiento *j* s_j^2y su desviación estándar s_j.

Las formulas para obtener la media y la varianza muestral del tratamiento *j* son:

$$\bar{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j}$$

$$s^2 = \frac{\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_j\ )^2}{n_j - 1}$$

La media muestral general es $\bar{\bar{x}}$, siendo la suma de todas las observaciones (de cada tratamiento), dividida en la cantidad total de observaciones $(n_T = n_1 + n_2 + \cdots + n_k)$:

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}}{n_T}$$

Cabe recordar que, si todas las muestras son de igual tamaño, la fórmula para calcular la media muestral general, se puede expresar como la sumatoria de los promedios de cada tratamiento dividido por el total de tratamientos *k*:

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}}{nk} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}/n}{k} = \frac{\sum_{j=1}^{k}\bar{x}_j}{k}$$

Cuando la hipótesis nula es verdadera la estimación de la media muestral general es la mejor estimación de la media poblacional μ.

CALCULO DE LA VARIANZA POBLACIONAL ENTRE TRATAMIENTOS

A la estimación de la σ^2 entre tratamientos se le denomina **cuadrado medio debido a los tratamientos** y se denota **CMTR**.

$$\mathbf{CMTR} = \frac{\sum_{\mathrm{i=1}}^{\mathrm{n}} n_j(\bar{x}_j - \bar{\bar{x}}\ )^2}{\mathrm{k} - 1}$$

El numerador es conocido como **suma de cuadrados debido a los tratamientos** denotado como **SCTR.** El denominador k-1 representa los grados de libertad de **SCTR.** Por ende $SCTR = \sum_{\mathrm{i}=1}^{\mathrm{n}} n_j(\bar{x}_j - \bar{\bar{x}}\ )^2$, entonces:

$$\mathbf{CMTR} = \frac{\boldsymbol{SCTR}}{\mathrm{k}-1}$$

Si nuestra hipótesis nula es verdadera **CMTR** proporciona una estimación insesgada de la varianza poblacional. Si no se cumple el estimador de la varianza poblacional es insesgado y sobreestima a la misma.

CALCULO DE LA VARIANZA POBLACIONAL DENTRO DE LOS TRATAMIENTOS

A la estimación de la σ^2 dentro de los tratamientos se le denomina **cuadrado medio debido al error** y se denota **CME**.

$$\boldsymbol{CME} = \frac{\sum_{j=1}^{k}(n_j-1)s_j^2}{n_T-k}$$

El numerador es conocido como **suma de cuadrados debido al error** denotado como **SCE.** El denominador representa los grados de libertad de **SCE.** Por ende $SCE = \sum_{j=1}^{k}(n_j-1)s_j^2$, entonces:

$$\mathbf{CME} = \frac{\boldsymbol{SCE}}{n_T-k}$$

Como el **CME** está basado en la variación dentro de cada tratamiento, no influye que la hipótesis nula sea verdadera o no. Por tanto siempre entrega un estimador insesgado de la varianza poblacional.

ESTADISTICO DE PRUEBA F: COMPARACIÓN DE LAS ESTIMACIONES E LA VARIANZA POBLACIONAL

CMTR y **CME** son estimaciones insesgadas e independientes de la varianza poblacional si nuestra hipótesis nula es verdadera, es decir que todas las medias poblacionales sean iguales. Cuando se tienen dos poblaciones normales, la distribución muestral del cociente de dos estimaciones independientes de la varianza poblacional sigue una distribución *F.* Sí la hipótesis nula es verdadera y se satisfacen los supuestos de ANOVA, la distribución de **CMTR/CME** sigue una distribución *F* con k-1 grados de libertad en el numerador y $n_T - k$ grados de libertad en el denominador. Pero si la hipótesis nula es falsa el valor del cociente de **CMTR/CME** será muy grande para haber sido tomado de una distribución *F* con k-1 grados de libertad en el numerador y $n_T - k$ grados de libertad en el denominador y se rechazará la hipótesis nula. Por lo tanto, nuestro estadístico de prueba es:

$$F = \frac{\mathbf{CMTR}}{\mathbf{CME}}$$

El estadístico sigue una distribución *F* con $k-1$ grados de libertad en el numerador y $n_T - k$ grados de libertad en el denominador.

Los métodos de valor-p y valor crítico se aplican de igual forma que con los test de hipótesis. Buscando el valor del estadístico F en la tabla, siempre en la cola superior.