**ANOVA**

Un análisis de varianza (**ANOVA**) prueba la hipótesis de que las medias de dos o más poblaciones son iguales. Los **ANOVA** evalúan la importancia de uno o más factores al comparar las medias de la variable de respuesta en los diferentes niveles de los factores.

El análisis de la varianza (o **Anova**: Analysis of variance) es un método **para** comparar dos o más medias, **que** es necesario porque cuando se quiere comparar más de dos medias es incorrecto utilizar repetidamente el contraste basado en la t de Student.

**ANÁLISIS DE LA VARIANZA CON UN FACTOR (ANOVA)**

El análisis de la varianza permite contrastar la hipótesis nula de que las medias de K poblaciones (K >2) son iguales, frente a la hipótesis alternativa de que por lo menos una de las poblaciones difiere de las demás en cuanto a su valor esperado. Este contraste es fundamental en el análisis de resultados experimentales, en los que interesa comparar los resultados de K 'tratamientos' o 'factores' con respecto a la variable dependiente o de interés.



El Anova requiere el cumplimiento los siguientes supuestos:

* Las poblaciones (distribuciones de probabilidad de la variable dependiente correspondiente a cada factor) son normales.
* Las K muestras sobre las que se aplican los tratamientos son independientes.
* Las poblaciones tienen todas igual varianza (homoscedasticidad).

El ANOVA se basa en la descomposición de la variación total de los datos con respecto a la media global (SCT), que bajo el supuesto de que H0 es cierta es una estimación de  obtenida a partir de toda la información muestral, en dos partes:

* Variación dentro de las muestras (SCD) o Intra-grupos, cuantifica la dispersión de los valores de cada muestra con respecto a sus correspondientes medias.
* Variación entre muestras (SCE) o Inter-grupos, cuantifica la dispersión de las medias de las muestras con respecto a la media global.

Las expresiones para el cálculo de los elementos que intervienen en el Anova son las siguientes:

MediaGlobal:   
Variación Total: 

Variación Intra-grupos: 

Variación Inter-grupos: 

Siendo xij el i-ésimo valor de la muestra j-ésima; nj el tamaño de dicha muestra y su media.

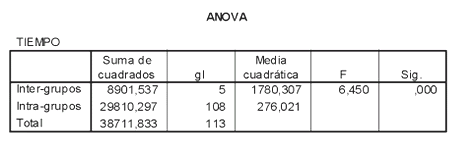
Cuando la hipótesis nula es cierta SCE/K-1 y SCD/n-K son dos estimadores insesgados de la varianza poblacional y el cociente entre ambos se distribuye según una F de Snedecor con K-1 grados de libertad en el numerador y N-K grados de libertad en el denominador. Por lo tanto, si H0 es cierta es de esperar que el cociente entre ambas estimaciones será aproximadamente igual a 1, de forma que se rechazará H0 si dicho cociente difiere significativamente de 1.

La secuencia para realizar un ANOVA es:

Analizar

Comparar medias

ANOVA de un factor



En el cuadro de resultados del ANOVA, el valor del estadístico de prueba, F=6,450, es significativamente distinto de 1 para cualquier nivel de significación y, por lo tanto, se rechaza la hipótesis nula de igualdad de medias .

## Grados de libertad

Los grados de libertad pueden descomponerse al igual que la suma de cuadrados. Así, GLtotal = GLentre + GLdentro. Los GLentre se calculan como: a - 1, donde a es el número de tratamientos o niveles del factor. Los GLdentro se calculan como N - a, donde N es el número total de observaciones o valores de la variable medida (la variable respuesta).

Pruebas de significación

El análisis de varianza lleva a la realización de pruebas de significación estadística, usando la denominada [distribución F](https://es.wikipedia.org/wiki/Distribuci%C3%B3n_F) de Snedecor.

Tabla ANOVA

Una vez que se han calculado las sumas de cuadrados, las medias cuadráticas, los grados de libertad y la F, se procede a elaborar una tabla que reúna la información, denominada "Tabla de Análisis de varianza o ANOVA", que adopta la siguiente forma:

{\displaystyle =+}

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Fuente de variación** | **Suma de cuadrados** | **Grados de libertad** | **Cuadrado medio** | **F** |
| Intergrupo | {\displaystyle SS\_{\hbox{Factores}}} | N° de grupos -1 | {\displaystyle T={\frac {SS\_{\hbox{Factores}}}{t-1}}}Suma de cuadrados interg / g de l | Cuadrados medios inter / cuadrados medios intrag{\displaystyle F={\frac {T}{E}}} |
| Intragrupo o Error | {\displaystyle SS\_{\hbox{Error}}} | diferencia | {\displaystyle E={\frac {SS\_{\hbox{Error}}}{N-t}}}Suma de cuadrados intra / g de l | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Total | {\displaystyle SS\_{\hbox{Total}}} | N - 1 | \_\_\_\_\_\_\_\_\_\_\_\_\_ |  |