

Análisis de Varianza

Leonardo H. Añez Vladimirovna^{*}

Facultad de Ingeniería en Ciencias de la Computación y Telecomunicaciones
Universidad Autónoma Gabriel René Moreno

1 de diciembre de 2018

1. ANOVA

1.1. Analisis Multivariante

El Análisis Multivariante comprende un conjunto de técnicas o métodos estadísticos cuya finalidad es analizar simultáneamente información relativa a varias variables para cada individuo o elemento estudiado. Algunos de estos métodos son puramente descriptivos de los datos muestrales, mientras que otros utilizan dichos datos muestrales para realizar inferencias acerca de parámetros poblacionales.

Propósitos:

- Describir información de forma resumida.
- Agrupar observaciones o variables en subconjuntos homogéneos.
- Explorar la existencia de asociaciones entre variables.
- Explicar (o probar) comportamientos.

Existen diferentes clasificaciones de los métodos de Análisis Multivariante. Una de las más usuales distingue dos grandes grupos, según el objetivo del análisis: métodos de dependencia y métodos de interdependencia. Además, dentro de cada uno de estos grupos, la naturaleza de las variables juega un papel importante en la definición de los diversos métodos.

1.2. Analisis de Varianza

El Análisis de Varianza (*ANOVA*) es un método estadístico utilizado para probar diferencias entre dos o más medias. Puede parecer extraño que la técnica se llame *Análisis de varianza* en lugar de *Análisis de medias*. Como verá, el nombre es apropiado porque las inferencias sobre las medias se hacen analizando la varianza. Así, el Análisis de Varianza se puede describir de la siguiente manera:

El análisis de varianza (ANOVA) es una herramienta estadística utilizada para detectar diferencias entre las medias de los grupos experimentales.

En múltiples ocasiones el analista o investigador se enfrenta al problema de determinar si dos o más grupos son iguales, si dos o más cursos de acción arrojan resultados similares o si dos o más conjuntos de observaciones son parecidos. Es por esto que ANOVA se utiliza para probar diferencias generales más que específicas entre los medios.

^{*} correo: toborochi98@outlook.com

1.2.1. Comparación Simultanea

Necesitamos poder comparar simultaneamente todas la medias. El test que lo permite es el test ANOVA. Como su nombre indica, compara varianzas aunque lo que contrastamos sean medias. Para ello parte de 3 requisitos previos:

- **Independencia:** Las k muestras son independientes
- **Normalidad:** $X_i \sim N(\mu_i, \sigma_i^2), i = 1, \dots, k$
- **Homocedasticidad:** $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \sigma^2$

1.2.2. Fundamentos

•	•	•	•	•	•	•
•	•	•	•	•	•	•
•	•	•	•	•	•	•
•	•	•	•	•	•	•
•	•	•	•	•	•	•
Medias	\bar{x}_1	\bar{x}_2	\dots	\bar{x}_i	\dots	\bar{x}_k
Varianzas	•	•	•	•	•	•

1.3. ANOVA de un factor

ANOVA de un factor (también llamada ANOVA unifactorial o one-way ANOVA en inglés) es una técnica estadística que señala si dos variables (una independiente y otra dependiente) están relacionadas en base a si las medias de la variable dependiente son diferentes en las categorías o grupos de la variable independiente. Es decir, señala si las medias entre dos o más grupos son similares o diferentes.

1.3.1. Cuando usar ANOVA de un factor

Usamos ANOVA de un factor cuando queremos saber si las medias de una variable son diferentes entre los niveles o grupos de otra variable. Por ejemplo, si comparamos el número de hijos entre los grupos o niveles de clase social: los que son clase baja, clase trabajadora, clase media-baja, clase media-alta y clase alta. Es decir, vamos a comprobar mediante ANOVA si la variable “número de hijos” está relacionada con la variable “clase social”. Concretamente, se analizará si la media del número de hijos varía según el nivel de clase social a la que pertenece la persona.

Condiciones

- En ANOVA de un factor solo se relacionan dos variables: una variable dependiente (o a explicar) y una variable independiente (que en esta técnica se suele llamar factor)
- La variable dependiente es cuantitativa (escalar) y la variable independiente es categórica (nominal u ordinal).
- Se pide que las variables sigan la distribución normal, aunque como siempre esto es difícil de cumplir en investigaciones sociales.
- También que las varianzas (es decir, las desviaciones típicas al cuadrado) de cada grupo de la variable independiente sean similares (fenómeno que se conoce como homocedasticidad). Aunque esto es lo ideal, en la realidad cuesta de cumplir, e igualmente se puede aplicar ANOVA

1.3.2. Bases de ANOVA de un factor

ANOVA de un factor compara las medias de la variable dependiente entre los grupos o categorías de la variable independiente. Por ejemplo, comparamos las medias de la variable “Número de hijos” según los grupos o categorías de la variable “Clase social”.

Si las medias de la variable dependiente son iguales en cada grupo o categoría de la variable independiente, los grupos no difieren en la variable dependiente, y por tanto no hay relación entre las variables. En cambio, y siguiendo con el ejemplo, si las medias del número de hijos son diferentes entre los niveles de la clase social es que las variables están relacionadas.

1.3.3. Estadísticos calculados en ANOVA

Al aplicar ANOVA de un factor se calcula un estadístico o test denominado F y su significación. El estadístico F o F-test (se llama F en honor al estadístico Ronald Fisher) se obtiene al estimar la variación de las medias entre los grupos de la variable independiente y dividirla por la estimación de la variación de las medias dentro de los grupos. En **conclusión**, cuanto más difieren las medias de la variable dependiente entre los grupos de la variable independiente, más alto será el valor de F. Si hacemos varios análisis de ANOVA de un factor, aquel con F más alto indicará que hay más diferencias y por tanto una relación más fuerte entre las variables.

1.3.4. Interpretación

- **Significación:** si es menor de 0,05 es que las dos variables están relacionadas y por tanto que hay diferencias significativas entre los grupos
- **Valor de F:** cuanto más alto sea F, más están relacionadas las variables, lo que significa que las medias de la variable dependiente difieren o varían mucho entre los grupos de la variable independiente.

1.3.5. Ejemplo

En un experimento se compararon tres métodos de enseñar un idioma extranjero; para evaluar la instrucción, se administró una prueba de vocabulario de 50 preguntas a los 24 estudiantes del experimento repartidos de a ocho por grupo.

A) *¿Cuál es la variable respuesta y la explicativa en este estudio?*

La variable respuesta es el puntaje en la prueba de vocabulario. La variable explicativa son los métodos de enseñanza (auditivo, traducción y combinado). Es un factor con 3 niveles.

B) *Complete la tabla de ANOVA*

1.4. ANOVA de dos factores