Subject: Probability and Statistics

Class X: Variance Analysis

UdeA

Bioengineering

Francisco José Campuzano Cardona

Bioengineerer, MSc in Engineering

UdeA

El análisis de varianza es una prueba estadística que permite conocer si los factores de un experimento dado, tienen efecto sobre la variable respuesta. Estas pruebas se basan en modelos lineales de la siguiente forma:

$$y_{ijk\ldots} = \mu + \tau_i + \beta_j + \cdots + \epsilon_{ijk\ldots}$$

Observación ijk…-esima estimada

Media poblacional

Efecto del factor τ, en el nivel i-ésimo.

Efecto del factor β, en el nivel j-ésimo.

Error aleatorio en de la observación ijk…-esima

UdeA

El Análisis de Varianza por ANOVA es una prueba paramétrica que tiene varios supuestos:

Previo a la aplicación de la Prueba

1. **Las Observaciones son normales (variable respuesta)**
2. **La variable respuesta es homoscedástica entre niveles de los factores.**
3. **Hay independencia entre puntos experimentales.**

Posterior a la aplicación de la Prueba

1. **Los errores o residuos son normales con media 0**

2. **Los errores o residuos son independientes y aleatorios**

Prueba de Runs. (WALD-WOLFOWITZ RunsTest)

Determina si una secuencia de datos están distribuidos de forma aleatoria, o si hay rachas o patrones.

Ho: *La secuencia es aleatoria*
H1: *La secuencia no es aleatoria.*

Es de utilidad para corroborar que los residuos de un ANOVA son aleatorios.

Pasos en el Análisis de Varianza:

1. Estructurar los datos.

Previamente se debió establecer un diseño experimental, donde, por ejemplo, si se trataba de un diseño factorial, entonces se tendría una tabla con las diferentes corridas experimentales, que son las combinaciones posibles de los diferentes niveles de los diferentes factores, con sus réplicas. Entonces a cada corrida experimental, se le asigna su respectiva observación de la variable respuesta.

UdeA

Pasos en el Análisis de Varianza:

1. Estructurar los datos.

Run	Factor 1	Factor 2	Variable respuesta
1	1	1	3.5
2	1	0	2.4
3	0	0	5.8
4	1	1	3.6
5	0	1	2.3
6	0	1	2.4
7	0	0	6.1
8	1	0	2.5

UdeA

Pasos en el Análisis de Varianza:

2. Corroborar los supuesto de Normalidad y Homocedasticidad de la variable respuesta entre niveles de los factores.

3. Determinar si hay interacciones (gráfico)

4. Realizar el ANOVA.

5. Corroborar los supuestos de Normalidad e independencia de los residuos o errores. Normalidad se puede analizar con un gráfico de probabilidad normal. ***

*** si el modelo es de efectos fijos, entonces no es tan importante la normalidad
*** si el modelo es de efectos aleatorios, sí es muy importante la normalidad

Un modelo ANOVA es de efectos fijos, cuando se eligen los niveles de los factores de forma arbitraria. En este caso las conclusiones del análisis solo son válidas para los valores analizados.

Un modelo ANOVA es de efectos aleatorios cuando los niveles de los factores se eligen de forma aleatoria de entre un rango de posibles valores. En este caso los resultados del análisis son extrapolables a todo el rango de valores de los factores.

Para cada efecto, el ANOVA tiene la siguientes hipótesis

Ho: *El efecto es igual a cero, es decir no es significativo*

H1: *El efecto es diferente de cero, es decir es significativo.*

UdeA

Ejemplo:

En el proceso de limpieza de unas piezas de metal, se realiza un estudio para determinar si 3 factores influyen en la conductividad eléctrica y sus efectos. Los factores que se cree afectan la conductividad son:

1. Concentración de ácido sulfúrico: se usan 4 niveles 0%, 6%, 12% y 18%
2. Concentración de sal: se usan 3 niveles: 0%, 10% y 20%
3. Temperatura del baño: se usan 2 niveles 80°F y 100°F.

Además de realizan dos réplicas

UdeA

Codificación de efectos en el modelo ols (Python)

Si se tienen dos variables llamadas: “Factor1” y “Factor2” (nombres de las columnas del Df)

Entonces ***‘C(Factor1)’*** significa: efecto del factor 1

Ahora si quiero considerar las interacciones multiplicativas de los dos efectos uso ***‘C(Factor1)*C(Factor2)’***

Ahora **‘C(Factor1)*C(Factor2)’** es equivalente a
‘C(Factor1) + C(Factor2)+ C(Factor1):C(Factor2)’

Efecto del factor 1

Efecto del factor 2

Efecto de la interacción de los dos factores

UdeA

Comparaciones Múltiples

Cuando se hace un ANOVA y en el experimento se tienen 3 o más niveles en algún factor que resulte tener efectos significativos, resulta útil realizar una prueba de comparaciones múltiples, para determinar si hay diferencias entre los diferentes niveles de ese factor.

Prueba HSD de Tukey ***
Prueba LSD de Fisher (t_test)

$$\boldsymbol{H_0}: \mu_i = \mu_j$$

$$\boldsymbol{H_1}: \mu_i \neq \mu_j$$

UdeA

Prueba Kruskal Wallis

Cuando no se logran cumplir la totalidad de los supuestos de un ANOVA se debe recurrir a una prueba no paramétrica como esta. Tiene el supuesto de independencia.

Esta prueba no permite analizar las interacciones, de modo que se deben analizar los factores por separado. Y determinar si hay diferencias para cada nivel.

Ho: *La media de las n muestras es igual. (el efecto no es significativo)*
H1: *La media de al menos 1 muestra es diferente de las demás. (el efecto es significativo)*

Subject: Probability and Statistics

UdeA

Bioengineering

¡Thanks!

Francisco José Campuzano Cardona

Bioengineering. MSc in Engineering