UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO

Instituto de Ciencias Económico Administrativas

– Área Académica: Economía

– Tema: Análisis de la varianza (ANOVA)

– Profesor(a):
  - Dr. Juan Roberto Vargas Sánchez

– Periodo: Enero Junio 2019

Resumen

- El análisis de la varianza permite determinar si diferentes tratamientos muestran diferencias significativas o por el contrario puede suponerse que sus medias poblacionales no difieren. El análisis de la varianza permite superar las limitaciones de hacer contrastes bilaterales por parejas que son un mal método para determinar si un conjunto de variables con $n > 2$ difieren entre sí.

- **Palabras Clave**: Poblaciones Independientes, Poblaciones con desviación estándar iguales, Distribución normal

Tema: Prueba de los signos

Abstract

The analysis of the variance allows to determine if different treatments show significant differences or on the contrary it can be assumed that their population means do not differ. The analysis of variance allows us to overcome the limitations of making bilateral contrasts by pairs that are a bad method to determine if a set of variables with n> 2 differ from each other.

Keywords: Independent Populations, Populations with equal standard deviation, Normal distribution

Objetivo General

- Que los estudiantes desarrollen un ANOVA de una vía.

Objetivos Específicos

- Que los estudiantes organicen los datos en una tabla ANOVA y realicen pruebas de hipótesis entre tres o más medias de tratamiento.

- # ANOVA... ¿Para qué sirve?

"Sirve para determinar si varias medias muestrales provienen de una sola población o de poblaciones con medias diferentes." Lind et al (2008).

Conceptos

Tratamiento (k): es una propiedad o característica que nos permite distinguir entre sí a las distintas poblaciones.

Variación de tratamiento (SST): Suma de las diferencias elevadas al cuadrado entre la media de cada tratamiento y la media total o global.

Variación aleatoria (SSE): Suma de las diferencias elevadas al cuadrado entre cada observación y su media de tratamiento.

Variación total (SS total): Suma de las diferencias elevadas al cuadrado entre cada observación y la media global.

Estadístico de prueba

Es la razón de los 2 estimados de la varianza poblacional:

$$F = \frac{SST/_{k-1}}{SSE/_{n-k}}$$

Donde:

SST es la variación aleatoria

SSE es la variación del tratamiento

k es el número de tratamientos

n es el total del datos

El valor crítico de F se calcula

suponiendo una prueba de cola derecha, ya

que los valores grandes de F corresponden

a diferencias significativas entre medias.

Con k muestras, cada una con n

valores, el número de grados de libertad

se obtiene de la siguiente manera.

Grados de libertad:

(k = número de muestras y n = tamaño de la muestra)

$$\text{numeradorde grados de libertad} = k - 1$$

$$\text{denominadorde grados de libertad} = k(n - 1)$$

Autoevaluación 12.3

Los siguientes datos son las colegiaturas por semestre (en miles de dólares) de una muestra de universidades privadas en varias regiones de Estados Unidos. Con un nivel de significancia de 0.05, ¿se puede concluir que hay una diferencia en las colegiaturas medias de las diversas regiones?

Noreste (en miles de dólares)	Sureste (en miles de dólares)	Oeste (en miles de dólares)
10	8	7
11	9	8
12	10	6
10	8	7
12		6

a) Formule las hipótesis nula y alternativa.
b) ¿Cuál es la regla de decisión?
c) Elabore una tabla ANOVA. ¿Cuál es el valor del estadístico de prueba?
d) ¿Cuál es su decisión respecto de la hipótesis nula?
e) ¿Puede existir una diferencia significativa entre la colegiatura media en el noreste en comparación con la del oeste? Si la hay, desarrolle un intervalo de confianza de 95% para esa diferencia.

a)

$H_0 =$
$las\ colegiaturas\ medias\ en\ las\ 3\ regiones\ son\ iguales$

$H_i = las\ colegiaturas\ \ medias\ difieren$

b) Si el estadístico F es mayor al valor crítico, H_0 se rechaza, y se concluye que las colegiaturas difieren.

c) Se calcula la media global:

$$\bar{X}_G = \frac{124}{14} = 8.85$$

Seguido de las medias de cada tratamiento:

$\bar{X}_1 = 11$ $\bar{X}_2 = 8.75$ $\bar{X}_3 = 6.8$

Posteriormente las variaciones poblacionales:

SS total $\Sigma(\bar{X} - \bar{X}_G)^2 = \mathbf{53.715}$

SSE $\Sigma(\bar{X} - \bar{X}_c)^2 = \mathbf{9.55}$

SST $53.715 - 9.55 = \mathbf{44.165}$

$$F = \frac{\frac{44.165}{2}}{\frac{9.55}{11}} = \frac{22.0825}{0.86} = 25.43$$

Fuente	Suma de cuadrados	*gl*	Media cuadrática	*F*
Tratamiento	44.16	2	22.08	25.43
Error	9.55	11	0.8682	
Total	53.71	13		

$$F_{\propto} = \frac{2}{11} = 3.98$$

***El numerador indica los grados de libertad** k-1*

***El denominador indica los grados de libertad** n-k*

Posteriormente el valor crítico se obtiene de la tabla de distribución de probabilidad F al 0.05 nivel de significancia.

Como $F = 25.43 > F_{\propto} = 3.98$ *la hipótesis nula se rechaza y se concluye que las medias en las 3 regiones son diferentes.*

Referencias Bibliográficas

- Lind, D. A., Marchal, W. G., & Wathen, S. A. (2008). Estadística aplicada a los negocios y la economía (13ª ed.). Madrid, España: McGraw-Hill.

- Triola, M. A. R. I. O. F. (2009). Estadística elemental (10ª ed.). México, México: Pearson.