Contabilidad y Finanzas.

Asignatura: Estadística II **Año: 2do.**

Tema 1. Análisis de Varianza.

Sumario.

1.1. Introducción. Análisis de Varianza Supuestos del modelo.

1.2. Descomposición de la suma de cuadrados total. Varianza del error estándar (Se). Prueba F. Tabla de Análisis de Varianza.

1.3. EJERCITACIÓN.

Objetivos: Que el alumno se capaz de:

- Aplicar el análisis de varianza en la solución de problemas relacionados con el perfil de la profesión.
- Interpretar salidas de paquete de programas estadísticos relacionados con el análisis de varianza.

Método:

- Elaboración conjunta, trabajo independiente

Medios:

- Tiza y pizarra, Libro de texto.
- Forma de enseñanza: Conferencia

Bibliografía:

- Probabilidades y Estadística para ingenieros Ronald E. Walpole / Raymond Myers
- Materiales entregados por el profesor

1.1. Introducción. Supuestos del Modelo.

Es muy frecuente en la vida que un profesional tenga que decidir si entre varios procesos o tratamientos algunos de ellos producen resultados diferentes.

La pregunta de interés del profesional será ¿existen diferencias significativas entre las medias de los tratamientos? Si comprueba que existen diferencias significativas, entonces le interesará

saber ¿cuáles de las poblaciones o tratamientos comparados son diferentes?

Primero necesitamos saber cómo contestar la pregunta general. El procedimiento para determinar si existen diferencias significativas entre varias poblaciones o grupos se llama Análisis de Varianza, y nos vamos a referir a él usando las letras ANOVA por Analysis of Variance, en inglés. ANOVA es un nombre genérico y se usa para una variedad inmensa de modelos de comparación de medias, también conocido como diseño de experimentos. Por ahora sólo hablaremos del ANOVA simple, de una vía (one way ANOVA), que se refiere a que se evalúa una sola causa de efecto (los tratamientos) a la comparación de medias de más de dos tratamientos. Vamos a llamar factor a una variable cualitativa que usaremos para designar a los grupos o tratamientos a comparar. Los niveles del factor serán el número de tratamientos o grupos.

Para conocer cuáles es o son los tratamientos diferentes se necesitan otras pruebas como el test de Tukey , pero no es objeto de este curso.

Ejemplo 1. Un ingeniero desea comparar 4 máquinas que se dedican a fabricar juntas de gomas para ollas de presión. Para la comparación se analiza la resistencia a la flexión de las juntas, entonces son selecciona 4 juntas producidas de cada máquina y se mide la resistencia a la flexión de cada junta, obteniendo los siguientes resultados

Máquinas	Máquina No.1	Máquina No.2	Máquina No. 3	Máquina No.4
Resiste a la flexión Junta 1	4.3	4.0	2.0	3.5
Resiste a la flexión Junta 2	4.8	4.1	3.8	3.4
Resiste a la flexión Junta 3	3.9	3.7	3.6	4.0
Resiste a la flexión Junta 4	3.7	3.9	3.5	4.1

En el caso usaremos un análisis de varianza ANOVA simple de un factor (el valor de la característica que se mide resistencia a la flexión de las juntas) está influido por un solo efecto que en este caso serán la máquina que los produce que llamaremos (Tratamientos).

Si además de la resistencia a la flexión de las juntas se evaluarán otras variables como la elasticidad de la junta, entonces el Análisis sería Multivariado (Análisis de 2 factores).

El Modelo matemático se representa por la expresión siguiente:

$$y_{ij} = \mu + \sigma_1 + e_{ij}$$

Donde:

y_{ij}= Valor que toma la característica aleatoria "y"

μ= Valor que temaría la característica "y" si no estuviera presente el tratamiento y el error aleatorio

σ_1= Efecto del tratamiento.

e_{ij} = Error de la Medición de y_{ij}

El subíndice "i": se destina a los diferentes tratamientos en este caso serán 4 máquinas diferentes.

El subíndice "j": se designa a las veces que se repite el tratamiento en cada grupo, en este caso 4 juntas diferentes, se debe aclarar que en la clasificación simple no es necesario que todos los grupos tengan las mismas cantidades.

Supuestos del Modelo:

- Las "y_{ij}" siguen una Distribución Normal con Media ($\mu+\sigma_1$) y varianza σ^2 para todo "i; j".
- Las observaciones no estén correlacionadas, son independiente

Como se desea comprobar si alguno de los tratamientos (máquinas) produce efectos diferentes entonces la hipótesis a comprobar son:

- **<u>Prueba de Hipótesis</u>.**
- **Ho: $\mu_1 = \mu_2 = \mu_3 = \mu_4$**
- **Hi: al mejos una μ_i diferente**

1.2. DESCOMPOSICIÓN DE LA SUMA DE CUADRADOS TOTAL. VARIANZA DEL ERROR ESTÁNDAR. TABLA DE ANÁLISIS DE VARIANZA. PRUEBA F.

El análisis de varianza se define como una técnica en la que la variabilidad de un conjunto de datos se divide en varios componentes y cada uno de ellos se asocia a una fuente específica

de variación, de manera que durante el análisis es posible encontrar la magnitud con la que contribuye cada una de esas fuentes en la variación total.

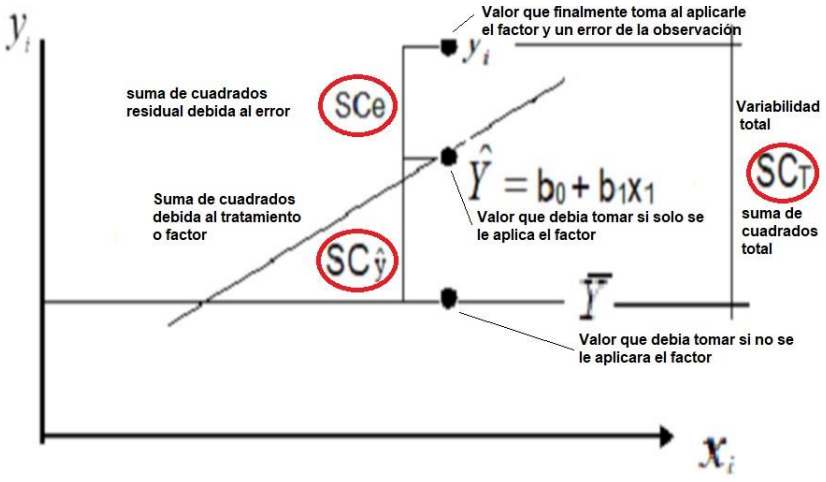
El nombre ANOVA es porque para comparar las medias de los grupos o tratamientos necesitamos identificar las distintas fuentes de variabilidad.

La variabilidad de la variable respuesta, sin referencia a ningún factor que la pudiera estar afectando, se conoce como variabilidad total.

La variabilidad de la variable respuesta que se atribuye a factores específicos se conoce como variabilidad explicada. Mide la variabilidad entre los diferentes grupos.

La variabilidad de la variable respuesta de las unidades (experimentales) dentro de cada nivel del factor se conoce como variabilidad no-explicada.

Se desprende que: Variabilidad total = variabilidad explicada + variabilidad no Explicada



Causas de variación	Suma de cuadrados (SC)	Grados de libertad (GL)	Cuadrado medio (CM)	F(calculada) Razón F	F(tabulada)	Pvalor
Debida al tratamiento	SC_{Trat}	p – 1	$CM_{trat.}=SC_{trat}/p-1$	CM_{trat}/CM_{res}	$F_{(1-\alpha;\ p-1;\ n-p)}$	
Residual o error	$SC_{residuo}$	n – p	$CM_{res}=SC_{res}/n-p$ Varianza del error estándar (Se)			
Total	SC_{total}	n - 1				

Para la Prueba de Hipótesis acerca de las medias usaremos el test estadístico F. Este test contrasta la variabilidad entre los grupos con la variabilidad que será natural dentro de los grupos.

$$F = \frac{\text{variabilidad ENTRE las medias muestrales}}{\text{variabilidad DENTRO de las muestras}}$$

Selección de Tablas Estádisticas: Pág. 23

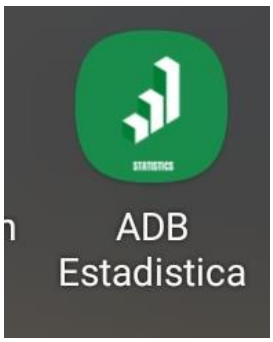
Para α=0.05 (p-1)= 4 Filas; (n-p)= 16 Columnas $F_{(1-\alpha;\, p-1;\, n-p)} = F_{(0.95;\, 3;\, 12)} = 3.49$

1.3. EJERCICIOS.

Ejemplo 1. Un ingeniero desea comparar 4 máquinas que se dedican a fabricar juntas de gomas para ollas de presión. Para la comparación se analiza la resistencia a la flexión de las juntas, entonces son selecciona 4 juntas producidas de cada máquina y se mide la resistencia a la flexión de cada junta, obteniendo los siguientes resultados

Máquinas	Máquina No.1	Máquina No.2	Máquina No. 3	Máquina No.4
Resiste a la flexión Junta 1	4.3	4.0	2.0	3.5
Resiste a la flexión Junta 2	4.8	4.1	3.8	3.4
Resiste a la flexión Junta 3	3.9	3.7	3.6	4.0
Resiste a la flexión Junta 4	3.7	3.9	3.5	4.1

Después de realizar el análisis de varianza con Excel o en otra aplicación se obtuvo el siguiente resultado



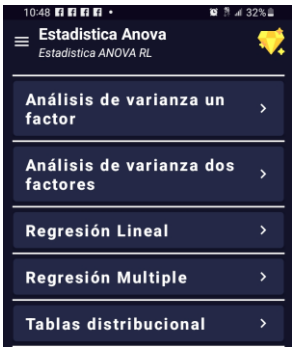






TABLA ANOVA

FUENTES DE VARIACION	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	VARIANZAS O MEDIA	ESTADISTICO TEST	p-value
ENTRE LOS GRUPOS	1.942	3	0.647	2.418	0.117
DENTRO DE LOS GRUPOS	3.212	12	0.268		
TOTAL	5.154	15			

Causas de variación	Suma de cuadrados (SC)	Grados de libertad (GL)	Cuadrado medio (CM)	F(calculada)	Pvalor
Debida al tratamiento	1.942	3	0.647	2.418	0.117
Residual o error	3.21	12	3767.78		
Total	5.154	15			

Existirá diferencias significativas entre los resultados de las 4 dietas con una confianza del 95%?

RESPUESTA:

Datos:

Tratamientos (Máquinas) (p)= 4

Muestras n= 16

(1-α)= 0.95 entonces α= 0.05

1er Paso: Planteamiento de la Hipótesis.

H_o: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ No hay diferencias entre las dietas

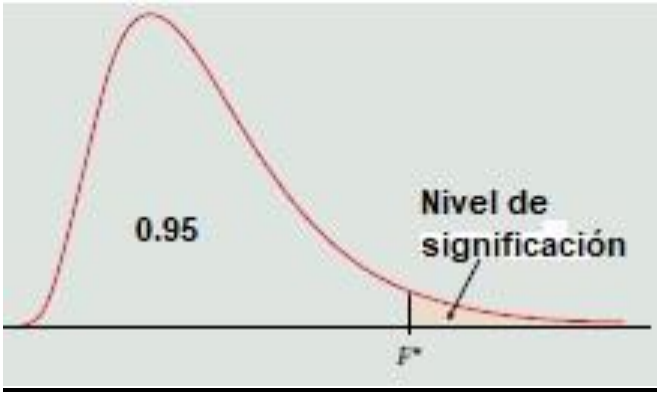
H_i: al menos un $\mu_i \neq$ alguna de las dietas es desigual

2Do Paso: Fijar el Nivel de significación α

α= 0.05

3er Paso: Determinar la Zona Crítica o Regla de Decisión

Si $F_{cal} > F_{(0.95;\ 3;\ 15)}(F_{crítico})$ entonces Rechazo H_o ó Si $P_{valor} < \alpha$ entonces Rechazo H_o

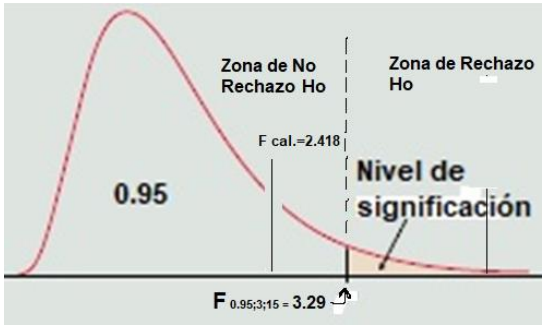


4to Paso: Calcular el Estadístico Z_{cal}

$F_{(0.95;\ 3;\ 15)}=3.29$

5to Paso: Toma de decisión.

$F_{cal} > F_{(0.95;\ 3;\ 15)}(F_{crítico})$



2.418 < 3.29 Entonces No Rechazo H_o

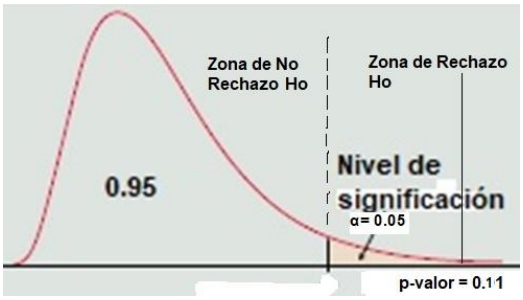
R/ No hay diferencias significativas en la resistencia a la flexión de las juntas producidas por las 4 máquinas

4to Paso: Toma de decisión.

Con P_{valor}

Si $P_{valor} < \alpha$ entonces Rechazo H_o

0.11 > 0.05 entonces No Entonces Rechazo H_o



R/ No hay diferencias significativas en la resistencia a la flexión de las juntas producidas por las 4 máquinas

CONCLUSIONES

- El análisis de varianza simple (oneway ó de un factor) analiza un solo efecto (los tratamientos) sobre la media de las muestras.
- El análisis de Varianza permite decidir si al menos uno de los tratamientos produce resultados diferentes a los demás, aunque no se pueda definir cual es el diferente, para ello sería necesario otras pruebas como la Prueba de Tukey o la Prueba de Duncan. que no se estudia en este curso.

ESTUDIO INDEPENDIENTE

Definir si hay diferencias significativas en la resistencia a flexión de las juntas que se producen en cada una de las 6 máquinas

Máquina

1	2	3	4	5	6
17.5	16.4	20.3	14.6	17.5	18.3
16.9	19.2	15.7	16.7	19.2	16.2
15.8	17.7	17.8	20.8	16.5	17.5
18.6	15.4	18.9	18.9	20.5	20.1