

Análisis de Varianza

ANOVA

MTRA. MELODY TREVIÑO RODRIGUEZ

Conceptos introductorios

Estadística



Descriptiva

Objetivo Principal: se centra en describir y resumir las características principales de un conjunto de datos y proporcionar una visión clara y concisa sobre la estructura de dichos datos, como la tendencia central, la dispersión y la forma de la distribución.

Métodos Utilizados: Incluye técnicas como medidas de tendencia central (media, mediana, moda), medidas de dispersión (varianza, desviación estándar, rango), y representaciones gráficas (histogramas, diagramas de dispersión, diagramas de caja).

Inferencial

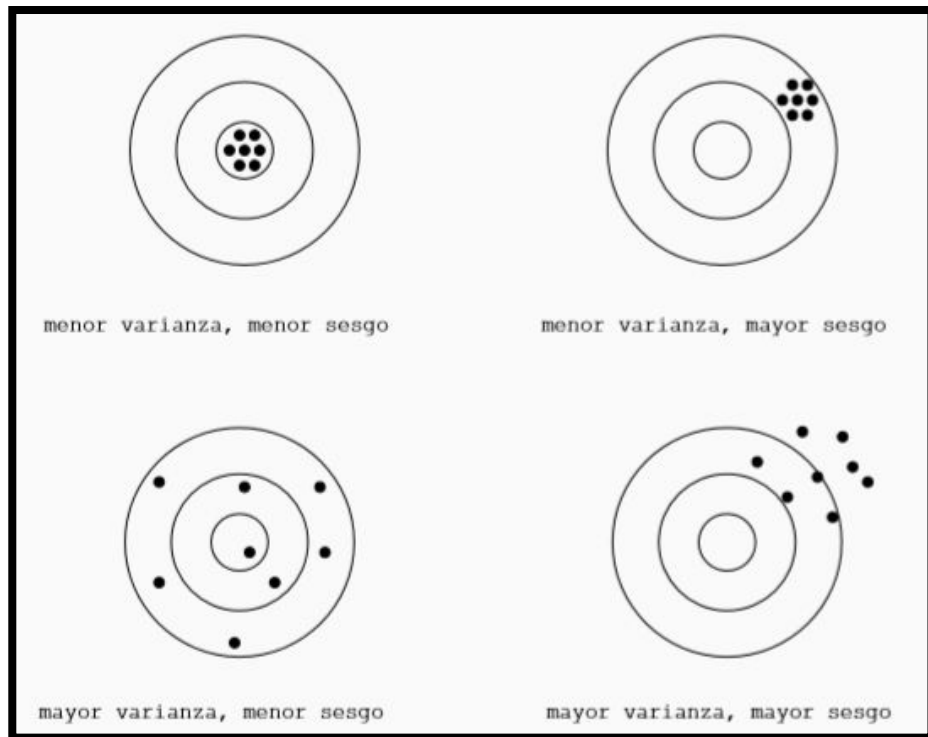
Objetivo Principal: se ocupa de hacer inferencias y generalizaciones sobre una población basándose en una muestra representativa de esa población. Busca sacar conclusiones más allá de los datos observados haciendo predicciones o estimaciones de la población.

Métodos Utilizados: Involucra técnicas como pruebas de hipótesis, intervalos de confianza, regresión, análisis de varianza, entre otros. Estas técnicas se utilizan para tomar decisiones informadas sobre las características de la población a partir de la información limitada proporcionada por la muestra.

Conceptos introductorios

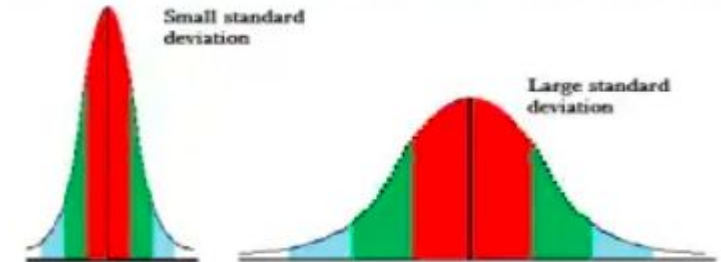
¿Qué es Varianza?

Es una medida de dispersión que indica cuánto se alejan los valores de una muestra o población respecto a su media



Importancia de estudiar la Variabilidad:

- Permite entender la extensión o dispersión de los datos, para comprender la "forma" de la distribución.



- Es esencial para la toma de decisiones y la realización de predicciones (si los datos son muy variables, las predicciones pueden ser menos confiables).
- Al comparar diferentes grupos o poblaciones, la variabilidad proporciona información sobre cuán distintos son estos grupos en términos de la dispersión de sus datos

Análisis de Varianza (ANOVA)

¿Qué es?

es una técnica estadística utilizada para comparar las medias de tres o más grupos diferentes y determinar si existen diferencias significativas entre ellos.

Su objetivo principal es analizar si la **variabilidad observada entre los grupos es mayor que la variabilidad esperada debido al azar.**

- En el ANOVA se utilizan las **pruebas de hipótesis.**
- Generalmente la hipótesis nula asume que no hay diferencias significativas entre las medias de los grupos, mientras que la hipótesis alternativa sugiere que al menos un grupo difiere significativamente de los demás.

1. Plantear hipótesis.

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D$$

$$H_A : \mu_i \neq \mu_j$$

Análisis de Varianza (ANOVA)

Tipos de ANOVA

ANOVA de un factor (One-Way ANOVA)

se compara la media de una variable entre tres o más grupos independientes.

ANOVA de dos factores (Two-Way ANOVA)

cuando se tienen dos variables independientes que afectan la variable dependiente. Puede analizar la influencia de cada variable por separado y su interacción.

ANOVA de medidas repetidas (Repeated Measures ANOVA)

Aplicado cuando las mismas medidas se toman en más de dos momentos diferentes o bajo más de una condición.

Entre otros.

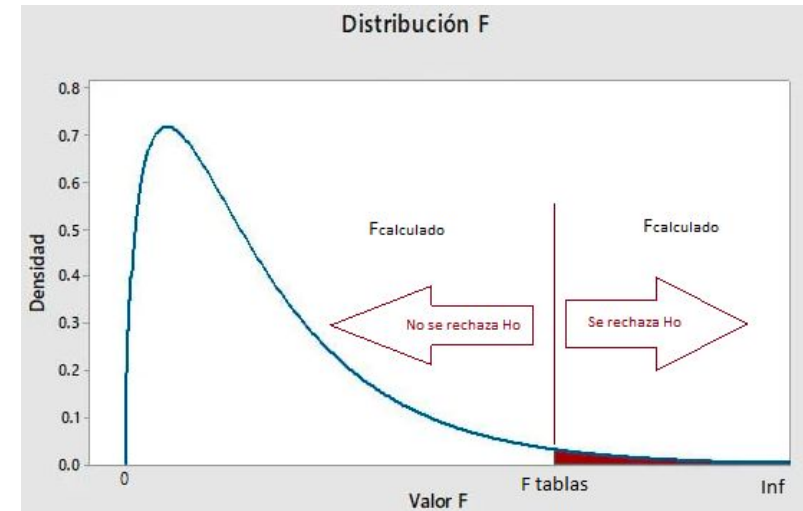
Estadística de Prueba F

Nombrado en honor al científico inglés Ronald Fisher; es un cociente de dos varianzas.

$$F_{prueba} = \frac{S_x^2}{S_w^2}$$

Valor de decisión.

$F_{calculado} > F_{tablas} \rightarrow$ se rechaza H_0



Análisis de Varianza (ANOVA)

Procedimiento general

1. Planteamiento del problema
2. Establecimiento de hipótesis
3. Elaboración de ANOVA
4. Evaluación de hipótesis y conclusión
5. Revisión de supuestos

Tabla ANOVA

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F
Factor	$SS_F = \sum_{i=1}^k n_i \cdot (\bar{y}_i - \bar{y})^2$	$k - 1$	$MSE_F = \frac{SS_F}{k - 1}$	$F = \frac{MSE_F}{MSE_E}$
Error	$SS_E = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	$N - k$	$MSE_E = \frac{SS_E}{N - k}$	
Total	$SS_T = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$	$N - 1$		

Donde:

- n_i es el tamaño de la muestra i .
- N es el número total de observaciones.
- k es el número de grupos diferentes del análisis de la varianza.
- y_{ij} es el valor j del grupo i .
- \bar{y}_i es la media del grupo i .
- \bar{y} es la media de todos los datos analizados.

Análisis de Varianza (ANOVA)

Ejercicio - ANOVA de un factor

1) Se lleva a cabo un estudio estadístico para comparar las puntuaciones que sacan cuatro alumnos en tres asignaturas diferentes (A, B y C). En la siguiente tabla se detallan las puntuaciones obtenidas por cada alumno en un test cuya puntuación máxima es 20. Realiza un análisis de la varianza para comparar las notas obtenidas por cada alumno en cada asignatura.

Alumno	Asignatura		
	A	B	C
1	14	13	19
2	12	14	17
3	14	10	16
4	10	14	19

2) La hipótesis nula es: las medias de las puntuaciones de las tres asignaturas son iguales vs

La hipótesis alternativa es: alguna de las medias es diferente

$$\begin{cases} H_0 : \mu_A = \mu_B = \mu_C = \mu \\ H_1 : \exists \mu_i \neq \mu \quad i = A, B, C \end{cases}$$

Análisis de Varianza (ANOVA)

Ejercicio - ANOVA de un factor

3) Elaboración de ANOVA

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	Estadístico F
Tratamiento o factor	70.167	2	35.083	11.079
Error	28.5	9	3.167	
Total	98.667	11		



Hoja de cálculo
de Microsoft Excel

4) Evaluación de hipótesis y conclusión

F tablas:

$$F_{(1-\alpha)_{gld}}^{gln} = F_{(1-0.5)_9}^2 = 4.256$$

Valor de decisión.

F calculado = 11.079 > F tablas = 4.256

Entonces se rechaza Ho;

Por lo tanto, alguna de las medias de los grupos estudiados es diferente a las otras.

Tabla D.9: VALORES CRÍTICOS DE LA DISTRIBUCIÓN F (0,05)

área a la derecha del valor crítico = 0,05

g.d.l	Grados de libertad del Numerador								
	1	2	3	4	5	6	7	8	9
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677
8	5,318	4,459	4,066	3,838	3,687	3,581	3,500	3,438	3,388
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020

Análisis de Varianza (ANOVA)

Ejercicio – ANOVA de un factor

5) Revisión de supuestos

Para poder hacer un análisis de la varianza (ANOVA) se deben cumplir los siguientes supuestos:

- **Independencia:** los valores observados son independientes entre sí. Una manera de garantizar la independencia de las observaciones es añadiendo aleatoriedad en el proceso de muestreo.
- **Homocedasticidad:** debe haber homogeneidad en las varianzas, es decir, la variabilidad de los residuos es constante.
- **Normalidad:** los residuos deben estar distribuidos normalmente.
- **Continuidad:** la variable dependiente debe ser continua.

Análisis de Varianza (ANOVA)

Tarea

1) Realizar el siguiente ejercicio:

Los miembros de un equipo ciclista se dividen al azar en tres grupos que entrenan con métodos diferentes. El primer grupo realiza largos recorridos a ritmo pausado, el segundo grupo realiza series cortas de alta intensidad y el tercero trabaja en el gimnasio con pesas y se ejercita en el pedaleo de alta frecuencia. Después de un mes de entrenamiento se realiza un test de rendimiento consistente en un recorrido cronometrado de 9 Km. Los tiempos empleados fueron los siguientes:

Método		
I	II	II
15	14	13
16	13	12
14	15	11
15	16	14
17	14	11

A un nivel de confianza del 95% ¿Puede considerarse que los tres métodos producen resultados equivalentes? O por el contrario ¿Hay algún método superior a los demás?

2) Investigar cómo se prueban los supuestos del ANOVA (*Independencia, Homocedasticidad, Normalidad y Continuidad*) y probarlos para el ejercicio visto en clase y para el ejercicio de tarea: