

5.3 ANÁLISIS DE VARIANZA

CONTENIDOS:

5.3.1. Prueba de análisis de varianza.

5.3.2 Comparaciones múltiples.

OBJETIVOS:

- Determinar los pasos a seguir al realizar una prueba de análisis de varianza
- Plantear hipótesis para la prueba de comparaciones múltiples.
- Interpretar el nivel de significación de la prueba de hipótesis.
- Redactar una conclusión con los resultados obtenidos de la prueba de hipótesis.
- Realizar pruebas de hipótesis en problemas prácticos.

5.3.1 PRUEBA DE ANÁLISIS DE VARIANZA.

CONCEPTOS CLAVES: Parámetro. Estimador. Hipótesis. Estadístico de prueba. Nivel de significación. Región de rechazo. Conclusión.

RESUMEN DE CONCEPTOS Y PROPIEDADES:

Sean $X_i \sim N(\mu_i, \sigma^2)$ $i=1,2,\dots,k$, k poblaciones normales independientes. Se selecciona una muestra de tamaño n_i de cada población obteniendo los siguientes datos.

Suma de cuadrados total
$$SCT = \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}^2 - \frac{x_{..}^2}{n}$$

Suma de cuadrados entre
$$SCE = \sum_{i=1}^k \frac{x_{i.}^2}{n_i} - \frac{x_{..}^2}{n}$$

Suma de cuadrados dentro
$$SCD = SCT - SCE$$

Pasos a seguir al realizar la prueba de hipótesis:

P1: Plantear hipótesis. Hipótesis nula $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$

v/s Hipótesis alternativa $H_A : \text{Alguna diferencia existe}$

P2: Estadístico de prueba:
$$F_0 = \frac{\frac{SCE}{k-1}}{\frac{SCD}{n-k}} \sim F(k-1, n-k)$$

P3: Establecer un nivel de significación: $\alpha = P(\text{Rechazar } H_0 / H_0 \text{ es verdadero})$

P4: Región de rechazo de H_0
$$R = \left\{ x / x > F_{(1-\alpha; k-1, n-k)} \right\}$$

P5: Decisión: Si $F_0 \in R \Rightarrow$ se rechaza H_0 al nivel de significación α

P6: Conclusión: Se debe interpretar la decisión tomada en Paso 5.

5.3.2 PRUEBA DE COMPARACIONES MÚLTIPLES.

CONCEPTOS CLAVES: Parámetro. Estimador. Hipótesis. Estadístico de prueba. Nivel de significación. Región de rechazo. Conclusión.

RESUMEN DE CONCEPTOS Y PROPIEDADES:

Si se rechaza la hipótesis nula $H_0: \mu_1 = \mu_2 = \dots = \mu_k$, es necesario encontrar cuáles medias poblacionales difieren, para lo cual se deben probar las siguientes hipótesis para cada par de medias

P1: Plantear hipótesis. Hipótesis nula $H_0: \mu_i = \mu_j$
v/s Hipótesis alternativa $H_0: \mu_i \neq \mu_j, \quad i < j$

P2: Estadístico de prueba:
$$t_0 = \frac{|\bar{X}_i - \bar{X}_j|}{\sqrt{\frac{SCD}{n-k} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \sim t(n-k)$$

P3: Establecer un nivel de significación: $\alpha = P(\text{Rechazar } H_0 / H_0 \text{ es verdadero})$

P4: Región de rechazo de H_0
$$R = \left\{ x / x > t_{(1-\frac{\alpha}{2}; n-k)} \right\}$$

P5: Decisión: Si $t_0 \in R \Rightarrow$ se rechaza H_0 al nivel de significación α

P6: Conclusión: Se debe interpretar la decisión tomada en Paso 5.

También se puede considerar como criterio de decisión la comparación de

$|\bar{X}_i - \bar{X}_j|$ con el valor $c_{ij} = \left[t_{1-\frac{\alpha}{2}}(n-k) \right] \sqrt{\frac{SCD}{n-k} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$

Luego si $|\bar{X}_i - \bar{X}_j| > c_{ij}$ entonces se rechaza H_0 al nivel de significación α

EJERCICIO RESUELTO, PASO A PASO:

Ejercicio 1: (Aplicación en Ciencias de la Ingeniería)

En una forja se utilizan tres hornos para derretir metales. Se supone que todos los hornos deberían operar a la misma temperatura. Se anotan las temperaturas en sucesivos calentamiento de los hornos, obteniendo los siguientes resultados:

Horno 1: 491.5 498.3 498.1 493.5 493.6

Horno 2: 488.5 484.65 479.9 477.35

Horno 3: 490.1 484.8 488.25 473.0 471.85 478.65

¿Proporcionan los datos evidencia suficiente para indicar una diferencia entre las temperaturas medias de los hornos?. De ser así ¿Qué pares de hornos producen esta diferencia? Use un nivel de significación de 0.05

Esquema de solución

Paso 1: Leer cuidadosamente el enunciado del problema.

Paso 2: Identificar la variable en estudio y los parámetros involucrados.

Sea X_i = Temperatura del horno i . En este caso se debe suponer que $X_i \sim N(\mu_i, \sigma^2)$ $i = 1, 2, 3$ donde μ_i es la temperatura promedio en el horno i y σ es la desviación estándar común de la temperatura de los hornos.

Paso 3: Cálculos previos.

Con los datos de las muestras se obtiene

$$\sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}^2 = 3545941,223 \quad x_{..} = \sum_{i=1}^3 \sum_{j=1}^{n_i} x_{ij} = 7292,05$$

$$x_{1.} = 2475 \quad x_{2.} = 1930,4 \quad x_{3.} = 2886,65 \quad \text{luego}$$

$$SCT = \sum_{i=1}^3 \sum_{j=1}^{n_i} x_{ij}^2 - \frac{x_{..}^2}{n} = 3545941,223 - \frac{(7292,05)^2}{15} = 1008.34$$

$$SCE = \sum_{i=1}^k \frac{x_{i.}^2}{n_i} - \frac{x_{..}^2}{n} = \left(\frac{(2475)^2}{5} + \frac{(1930,4)^2}{4} + \frac{(2886,65)^2}{6} \right) - \frac{(7292,05)^2}{15} = 594.53$$

$$SCD = SCT - SCE = 1008.34 - 594.53 = 413.812$$

Paso 4: Leer la pregunta y revisar cual de los conceptos se debe usar para obtener lo pedido.

Para responder la pregunta se debe realizar una prueba de análisis de varianza

Paso 5: Realizar la prueba siguiendo los seis pasos.

P1: Plantear hipótesis.

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad \text{v/s} \quad H_A : \text{Alguna diferencia existe}$$

$$\text{P2: Estadístico de prueba; } F_0 = \frac{\frac{SCE}{k-1}}{\frac{SCD}{n-k}} = \frac{\frac{594.53}{2}}{\frac{413.812}{12}} = 8.62$$

P3: Nivel significación; $\alpha = 0.05$

P4: Región de rechazo de H_0 v/s H_A

$$R = \{x / x > F_{(1-\alpha, k-1, n-k)}\} = \{x / x > F_{(0.95, 2, 12)}\} = \{x / x > 3.89\}$$

P5: Decisión. Como $F_0 = 8.62 > 3.89 \Rightarrow F_0 \in R \Rightarrow$ Se rechaza H_0 al nivel de significación 0.05

P6: Conclusión. Con 95% de confianza alguna diferencia existe en las temperaturas medias entre los hornos.

Prueba de comparaciones múltiples

Como se rechazó la hipótesis nula en la prueba de análisis de varianza, se debe comparar las medias para cada par de hornos

Esquema de solución

Paso 1: Leer la pregunta y revisar cual de los conceptos se debe usar para obtener lo pedido. Para responder la segunda pregunta se debe realizar una prueba de comparaciones múltiples

Debemos calcular las medias muestrales de las temperaturas para cada horno,

$$\bar{X}_1 = 4495 \quad \bar{X}_2 = 482.6 \quad \bar{X}_3 = 481.108$$

Paso 2: Realizar la prueba.

$$c_{1,2} = [t_{0.975}(12)] \sqrt{\frac{413.812}{12} \left(\frac{1}{5} + \frac{1}{4} \right)} = 8.583$$

$$c_{1,2} = [t_{0.975}(12)] \sqrt{\frac{413.812}{12} \left(\frac{1}{5} + \frac{1}{6} \right)} = 7.7476$$

$$c_{1,2} = [t_{0.975}(12)] \sqrt{\frac{413.812}{12} \left(\frac{1}{4} + \frac{1}{6} \right)} = 8.259$$

Comparación entre Horno 1 y Horno 2

Para $H_0 : \mu_1 = \mu_2$ v/s $H_A : \mu_1 \neq \mu_2$ se tiene que

$$|\bar{X}_1 - \bar{X}_2| = 12.4 > 8.583 \quad \text{luego se rechaza } H_0 \text{ al nivel } 0.05$$

Comparación entre Horno 1 y Horno 3

Para $H_0 : \mu_1 = \mu_3$ v/s $H_A : \mu_1 \neq \mu_3$ se tiene que

$$|\bar{X}_1 - \bar{X}_3| = 13.8917 > 7.7476 \quad \text{luego se rechaza } H_0 \text{ al nivel } 0.05$$

Comparación entre Horno 2 y Horno 3

Para $H_0 : \mu_2 = \mu_3$ v/s $H_A : \mu_2 \neq \mu_3$ se tiene que

$$|\bar{X}_2 - \bar{X}_3| = 1.4917 < 8.259 \quad \text{luego no se rechaza } H_0 \text{ al nivel } 0.05$$

Conclusión. Con 95% de confianza hay diferencia en las temperaturas medias entre el horno 1 y horno 2, y entre el horno 1 y horno 3, y no hay diferencia en las temperaturas medias entre el horno 2 y horno 3.

EJERCICIOS PROPUESTOS:

1. (Aplicación en Ciencias de la Ingeniería)

En un experimento para determinar el efecto de las burbujas de aire sobre la resistencia del asfalto, las burbujas de aire se controlan en tres niveles: Bajo (2 a 4%), Medio (4 a 6%) y Alto (6 a 8%). Los datos obtenidos se presentan en la tabla siguiente:

Nivel de Burbujas de aire	Resistencia al asfalto					
Bajo	106	90	103	90	89	88
Medio	80	69	94	81	70	83
Alto	78	80	62	69	76	85

- ¿Estos diferentes niveles de burbujas de aire afectan de la misma manera a la resistencia promedio del asfalto? Utilice $\alpha = 0.01$.
- Si estos tres niveles burbujas de aire afectan de manera significativamente diferente a la resistencia promedio del asfalto ¿Qué niveles producen esta diferencia significativa? Use $\alpha = 0.01$.

2. (Aplicación en Ciencias del Mar)

Para realizar un estudio de contaminación por plomo en especímenes que habitan en cuatro sectores costeros diferentes, se selecciona muestras aleatorias de especímenes en los cuatro sectores y se miden los porcentajes de plomo observados en ellos. Los resultados se muestran en la tabla siguiente:

Sector A	0.027	0.025	0.029	0.026	
Sector B	0.025	0.028	0.030	0.027	0.024
Sector C	0.034	0.029	0.032	0.031	0.036
Sector D	0.030	0.033	0.031		

- Pruebe la hipótesis de no diferencia en los porcentajes medios de plomo entre los sectores Use $\alpha = 0.01$.
 - Compare las medias por pares, al 5% de significación.
- ### 3. (Aplicación en Ciencias del Mar)
- Se desea comparar el crecimiento (en cms.) de cholgas mediante 5 tipos de dieta. Se sospecha que al elevar el contenido de nutrientes en la dieta se incrementará el crecimiento de las cholgas. También se sabe que el contenido de nutrientes debe variar aproximadamente entre 10 y 40% para que la dieta no tenga resultados adversos. Se prueban cinco niveles de porcentaje de nutrientes: 15, 20, 25, 30 y 35% en cinco muestras en cada nivel de nutrientes, obteniendo los siguientes resultados:

% nutriente	Observaciones					Total
15	0.7	0.7	1.5	1.1	0.9	4.9
20	1.2	1.7	1.2	1.8	1.8	7.7
25	1.4	1.8	1.8	1.9	1.9	8.8
30	1.9	2.5	2.2	1.9	2.3	10.8
35	0.7	1.0	1.1	1.5	1.1	5.4

- ¿Es afectado el crecimiento promedio de las cholgas el variar el contenido en los porcentajes de nutrientes? Use $\alpha = 0.05$.
- ¿Qué porcentaje de nutrientes tienen diferencias en las medias?