Análisis de varianza
Diseño de Experimentos

Pontificia Universidad
JAVERIANA
Bogotá

VIGILADA MINEDUCACIÓN

Pontificia Universidad
JAVERIANA
Colombia

Introducción

Previamente se revisaron procedimientos para la validez y/o contraste de una variable respuesta bajo el efecto de un factor de dos niveles. Ahora se revisará el análisis de una variable respuesta bajo el efecto de un factor que tiene a niveles.

Para este tipo de problemas usaremos la metodología de **análisis de varianza (ANOVA, *ANalysis Of VAriance*).**

Peso porcentual del algodón	Número de corrida experimental				
15	1	2	3	4	5
20	6	7	8	9	10
25	11	12	13	14	15
30	16	17	18	19	20
35	21	22	23	24	25

* La ejecución de las corridas debería hacerse de forma aleatoria.

Análisis de varianza

- Modelo con efectos fijos:

Los a tratamientos/niveles son elegidos expresamente por el experimentador. En esta situación quieren probarse hipótesis acerca de las medias de los tratamientos, y las conclusiones se aplicarán únicamente a los niveles del factor considerados en el análisis.

- Modelo con efectos aleatorios:

Los a tratamientos son una muestra aleatoria de una población más grande de tratamientos. En esta situación sería deseable poder extender las conclusiones (las cuales se basan en la muestra de los tratamientos) a la totalidad de los tratamientos de la población, sea que se hayan considerado explícitamente en el análisis o no.

.

Análisis de varianza

Tabla 3-2 Datos típicos de un experimento de un solo factor

Tratamiento (nivel)		Observaciones			Totales	Promedios
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2.}$	$\bar{y}_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$ $\cdots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$y_{a.}$	$\bar{y}_{a.}$
					$y_{..}$	$\bar{y}_{..}$

Modelo de los efectos:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases}$$

Análisis de varianza

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_a$$
$$H_1: \mu_i \neq \mu_j \quad \text{para al menos un par } (i,\ j)$$

$$H_0: \tau_1 = \tau_2 = \ldots \tau_a = 0$$
$$H_1: \tau_i \neq 0 \quad \text{para al menos una } i$$

$$\mu_i = \mu + \tau_i$$

Suma de cuadrados

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2$$

$$\sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^{a} (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2$$

$$SS_T = SS_{\text{Tratamientos}} + SS_E$$

$$MS_{\text{Tratamientos}} = \frac{SS_{\text{Tratamientos}}}{a-1} \qquad MS_E = \frac{SS_E}{N-a}$$

Suma de cuadrados

Variabilidad explicada por el factor

Estadístico de prueba

Tabla 3-3 Tabla de análisis de varianza para el modelo con un solo factor y efectos fijos

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0
Entre los tratamientos	$SS_{\text{Tratamientos}} = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2$	$a-1$	$MS_{\text{Tratamientos}}$	$F_0 = \dfrac{MS_{\text{Tratamientos}}}{MS_E}$
Error (dentro de los tratamientos)	$SS_E = SS_T - SS_{\text{Tratamientos}}$	$N-a$	MS_E	
Total	$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$	$N-1$		

Variabilidad debida al error

*Si el factor tiene un efecto real, la variabilidad entre las medias de los grupos (MST) debería ser mayor que la variabilidad dentro de los grupos (MSE). Esto es porque el factor está causando diferencias reales entre los grupos.

Supuestos del modelo ANOVA – efectos fijos

1. **Normalidad**

- Los datos en cada grupo o tratamiento deben seguir una distribución normal.

ANOVA de efectos fijos asume que los errores (residuos) están distribuidos normalmente dentro de cada grupo. Aunque el ANOVA es relativamente robusto a las desviaciones de la normalidad cuando se tiene un tamaño de muestra grande, grandes desviaciones pueden afectar la validez de los resultados en muestras pequeñas.

Como comprobar:

Q-Q plots
Histogramas de los residuos para evaluar la normalidad
Prueba de Shapiro-Wilk o Kolmogorov-Smirnov para comprobar la normalidad de los residuos.

Supuestos del modelo ANOVA – efectos fijos

2. Homoscedasticidad (homogeneidad de varianzas)

- Las varianzas dentro de cada grupo deben ser aproximadamente iguales.

ANOVA asume que la variabilidad dentro de cada grupo es similar. Si las varianzas entre grupos son muy diferentes, esto puede llevar a una mala estimación de la variabilidad y a resultados erróneos.

Como comprobar:

Prueba de Bartlett
Gráficos de dispersión de los residuos para identificar posibles desigualdades en las varianzas.

Supuestos del modelo ANOVA – efectos fijos

3. Independencia

- Las observaciones dentro de cada grupo y entre grupos deben ser independientes entre sí.

ANOVA asume que cada observación en el conjunto de datos es independiente de las demás. La falta de independencia puede afectar la validez de los resultados, ya que puede llevar a una subestimación o sobreestimación de la variabilidad entre los grupos

Como comprobar:

Generalmente se asegura mediante el diseño del estudio. Asegúrate de que las observaciones se recojan de manera que cada una sea independiente.

Supuestos del modelo ANOVA – efectos fijos

Acciones para tomar ante no cumplimiento de supuestos:

- Transformaciones de datos: transformaciones como la logarítmica o la raíz cuadrada pueden ayudar a estabilizar las varianzas y aproximar los datos a una distribución normal.

- Métodos alternativos: si la homogeneidad de varianzas no se cumple, usar pruebas no paramétricas como la prueba de Kruskal-Wallis.

- Revisión del diseño del estudio: el diseño experimental debe ser adecuado y las observaciones independientes.

Pruebas Post-Hoc

ANOVA indica si hay al menos una diferencia significativa entre las medias de los grupos, pero no dice específicamente cuáles grupos difieren entre sí. Aquí es donde entran las pruebas post-hoc: su objetivo es identificar cuáles grupos específicos tienen diferencias significativas.

- Prueba de Tukey HSD
- Prueba de Bonferroni
- Prueba de Scheffé

Bibliografía

Gutierrez, H. Análisis y diseño de experimentos. México: Mc Graw Hill, 2004.

Montgomery. Diseño y análisis de Experimentos. 2 Edición. John Wiley and sons, 2002.