CARRERA DE INGENIERÍA DE SISTEMAS

Diseño Completamente Aleatorio

ESTADÍSTICA APLICADA
PROF. MSC. VILMA S. ROMERO ROMERO
5TO CICLO | 23-11-2022

UNIVERSIDAD DE LIMA

Introducción

- Los modelos de diseño de experimentos son modelos estadísticos clásicos cuyo objetivo es evaluar si uno o más factores influyen en una variable de interés.
- Los diseños más utilizados son: Diseño completamente aleatorio, Diseño por bloques, Diseño Factorial, etc.
- Por ejemplo:
  - Estudiar la influencia de la marca de gasolina en la eficiencia del combustible del automóvil.
  - Evaluar el efecto de usar cinco fertilizantes diferentes en los rendimientos de los cultivos.
  - Estudiar la influencia en las ventas de un nuevo producto cuando se exhibe en tres lugares distintos dentro de varios supermercados.

Estudios observacionales vs experimentales

Estudio observacional

Observar lo que sucede en una situación particular, registrar datos de una o más variables de interés y realizar el análisis estadístico respectivo.

Por ejemplo:

- Analizar un grupo aleatorio de clientes que compran en Ripley.
  - Variables: Tiempo que demora en realizar la compra, sexo del cliente, monto gastado en la compra, cantidad de productos comprados, etc.
  - Análisis estadístico para determinar los factores que influyen en el gasto.
- Encuestas y sondeos de opinión pública. Ayudan a observar opiniones de los encuestados.

Estudios observacionales vs experimentales

Estudio experimental

- Se lleva a cabo un experimento bajo condiciones controladas para recopilar los datos.
- Se investiga los efectos de ciertas condiciones en individuos u objetos en la muestra.

Por ejemplo:

- Realizar un experimento para evaluar los efectos de un nuevo medicamento en la presión arterial.
- Realizar un experimento para evaluar si un nuevo aplicativo ayuda en el desarrollo de la comprensión lectora de niños de primaria.

Conceptos básicos

1) **Unidad experimental**
   - Sujeto u objeto sobre el que se experimenta el tratamiento.
   - Ejemplos: Una máquina, un paciente, un lote de material, etc.

2) **Factor**
   - Variable independiente de interés del experimentador de la cual se desea evaluar su efecto sobre la variable respuesta.
   - Ejemplos: Tipos de fertilizantes, marcas de gasolina, etc.

3) **Niveles de un factor o tratamientos**
   - Diferentes tipos o categorías específicas del factor.
   - Niveles de temperatura: $30°C$, $50°C$, $70°C$

4) **Tratamiento**
   - Nivel particular de un factor. Efecto que se desea estudiar.

UNIVERSIDAD DE LIMA

Diseño Completamente Aleatorio

Modelo de efectos fijos:

$$\boxed{Y_{ij} = \mu + \tau_j + e_{ij}}$$

Donde:

- $j = 1, 2, \ldots, k$ (cantidad total de tratamientos)
- $i = 1, 2, \ldots, n$ (número de repeticiones por tratamiento)
- $N = n \times k$ es el total de las observaciones.

- Y_{ij} es la i-ésima respuesta correspondiente al efecto del j-ésimo tratamiento.
- μ representa la media global y es común a todas las observaciones.
- τ_j es el efecto del j-ésimo tratamiento.
- e_{ij} es el componente aleatorio asociado al modelo.

UNIVERSIDAD DE LIMA

PH para la media de los tratamientos

Comparar más de dos tratamientos o niveles de un factor.

1) Hipótesis:
   $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$ (todos los tratamientos tienen la misma media común)
   H_1 : Al menos dos de las medias no son iguales

   Otra forma de expresar:

   $H_0 : \tau_1 = \tau_2 = \ldots = \tau_k = 0$ (los efectos de los tratamientos son 0)
   H_1 : Al menos un $\tau_j \neq 0$ (el factor tiene un efecto en la respuesta)

2) Especificar el nivel de significación α.

3) Calcular el valor del estadístico de prueba usando la tabla de análisis de varianza.

Tabla de Análisis de Varianza (ANOVA)

Fuente de variación	**Grados de libertad**	**Suma de Cuadrados**	**Cuadrado Medio**	$\mathbf{F_0}$	**P-value**
Debido a los tratamientos	$k-1$	SCTrat	$CMTrat = \frac{SCR}{k-1}$	$F_0 = \frac{CMTrat}{CME}$	$P(F > F_0)$
Debidoa al error	$k(n-1)$	SCE	$CME = \frac{SCE}{k(n-1)}$		
Total	$kn-1$	SCT			

4) Región crítica y regla de decisión:

$$RC = \langle F_{(k-1,k(n-1),1-\alpha)}; \infty \rangle \Rightarrow \textbf{Rechazar } H_0 \text{ si } F_0 > F_{(k-1,k(n-1),1-\alpha)}$$

Se puede calcular y utilizar el P-value:

$$\text{P-value} = P(F_{(k-1,k(n-1))} > F_0) \Rightarrow \textbf{Rechazar } H_0 \text{ si P-value} \leq \alpha$$

- Si el estadístico F no es significativo $\Rightarrow$ No hay evidencia estadística para concluir que la respuesta media difiere en cualquiera de los k niveles del factor.
- Si el estadístico F es significativo $\Rightarrow$ Se deben hacer más pruebas antes de sacar conclusiones.

Supuestos para el Análisis de Varianza (ANOVA

1) Para cada una de las k poblaciones, la variable respuesta se distribuye de manera normal.
2) Todas las k varianzas poblacionales son iguales. $(\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2)$
3) Las observaciones son independientes.

Evaluación de los Supuestos

- **Supuesto de normalidad**
  Evaluar que cada población se distribuya de manera normal usando la prueba de Anderson-Darling o la prueba de Shapiro-Wilk.

- **Supuesto de homogeneidad de varianzas**
  i) Hipótesis:
     H_0 : Las varianzas son homogéneas
     H_1 : Al menos una de las varianzas es distinta de las otras
  ii) Especificar el nivel de significancia α.
  iii) p-value utilizando el método de Bartlett.
  iv) Si p-value$> \alpha \Rightarrow$ No se rechaza la H_0 y se concluye que las varianzas de los tratamientos son homogéneas.

UNIVERSIDAD DE LIMA

Comparaciones Múltiples

El objetivo es realizar comparaciones estadísticas entre pares de medias de los tratamientos para determinar dónde ocurren las diferencias entre las medias.

$$\boxed{\begin{array}{l} H_0 : \mu_i = \mu_j \\ H_1 : \mu_i \neq \mu_j \end{array}}$$

Los procedimientos más comunes son:

- Diferencia mínima significativa de Fisher (Fisher's LSD)
- Prueba de Tukey (Tukey's test)
- Pairwise t-tests
- Método de Dunnett's

Intervalos de confianza de Tukey

Intervalos de confianza para la comparación por pares de múltiples medias.

- Si el intervalo de confianza incluye el valor cero $\Rightarrow$ No se puede rechazar la hipótesis de que las dos medias poblacionales son iguales.
- Si el intervalo de confianza no incluye el valor cero $\Rightarrow$ Se concluye que existe una diferencia significativa entre las medias poblacionales.

Recursos Adicionales

Devore, J. (2019). *Introducción a la probabilidad y estadística para ingeniería y ciencias*. Cengage, 1 edition. Tomado de http://webaloe.ulima.edu.pe/portalUL/bi/baseDatosEtech/index.jsp?BD=BI_RUTA_CENGAGE.

Devore, J., Farnum, N., and Doi, J. (2014). *Applied Statistics for Engineers and Scientists*. Cengage Learning, 3 edition.

Johnson, R. A. (2012). *Probabilidad y estadística para ingenieros*. Pearson Educación, 8 edition. Tomado de http://webaloe.ulima.edu.pe/portalUL/bi/baseDatosEtech/index.jsp?BD=BI_RUTA_PEARSON.

Millones, R., Barreno, E., Vásquez, F., and Castillo, C. (2016). *Estadística aplicada a la ingeniería y los negocios*. Lima: Fondo Editorial de la Universidad de Lima, 1 edition. Código Biblioteca U.Lima: 519.5 E7.