

## Resumen IME

### Capítulo 1

Script	Pág
Importación de datos	6
Importación librerías	7
Data.frame	8
Modificación matrices	8-9
Filtros	9
Pivotar matriz	10

### Capítulo 2

#### Variabilidad y dispersión:

**OBS:** si existen demasiados valores atípicos en las gráficas, usar **mediana** como medida de tendencia central e **IQR** como dispersión.

Script	Pág
Media	2
Quantiles	4
Media-Var-Desv-IQR	5
Tabla contingencia	6
Histograma	10
Caja	12
Barras	12
Torta	13
Dispersión	14
Barras 2 variables	16
Mosaico 2 variables	17
Caja varios grupos	18

### Capítulo 3

**Variable aleatoria:** puede tomar diferentes números, ser continua o discreta.

**Valor esperado:** promedio variable aleatoria.

Script	Pág
Valor esperado	2

Distribución Normal	6
Prueba Normalidad	8

El resto del capítulo habla de distribuciones

Bernoulli, Geométrica, Binomial...

### Capítulo 4

**Alfa:** nivel de significancia

**Error Estándar (SE):**

$\text{desv muestra} / \sqrt{\text{tam muestra}}$

**Intervalo de confianza:**

$X = 1 - \alpha$

**Interpretación:** “se tiene x % de certeza de que el parámetro de la población se encuentra entre...”

**Pruebas de Hipótesis:**

$H_0: \mu_N = \mu_0$

$H_A: \mu_N \neq \mu_0$

**Interpretación:**

- Se falla al rechazar  $H_0 \rightarrow$  se acepta  $H_0$
- Se rechaza  $H_0$  en favor de  $H_A \rightarrow$  se acepta  $H_A$

**OBS:** al definir cualquier intervalo de confianza (ejem: [a,b]), si el valor de  $\mu_0$  se encuentra dentro de este, se falla al rechazar  $H_0$ .

**Valor p:**

**p-value:** Probabilidad de que la hipótesis alternativa sea verdadera cuando se considerando la hipótesis nula como verdadera.

### Interpretación:

- $p < \alpha$  : Se rechaza  $H_0$  , hay mucha evidencia a favor de  $H_A$
- $p > \alpha$  : Se falla al rechazar  $H_0$ , no hay evidencia suficiente a favor de  $H_A$

### Metodología prueba de hipótesis:

1. Plantear  $H_0$  y  $H_A$
2. Verificar si sigue una distribución normal
3. Obtener p-value.

Script	Pág
Valor p unilateral	9
Valor p bilateral	10
Prueba Normalidad	8

### Capítulo 5

#### Prueba Z:

Usada para inferir medias.

#### Condiciones:

- Muestra con al menos 30 obs, sino conocer la varianza poblacional.
- Obs deben ser independientes.
- Sigue una distribución normal.

**OBS:** Si las muestras fueron elegidas al azar, entonces son independientes, también lo son si estas representan un porcentaje de la población.

**OBS:** Para demostrar normalidad, usar **shapiro.test(x)** o **grafico Q-Q**

**OBS:** Resultado shapiro.test()

- $p < 0.05$ : No sigue una distribución normal.

- $p > 0.05$ : Sigue una distribución normal.

### Prueba T:

Cuando la muestra es pequeña y la varianza poblacional desconocida.

Condiciones:

- Obs son independientes
- Sigue una distribución normal.

### Prueba T dos muestras pareadas:

Se usa cuando un dato de un conjunto se encuentra relacionado con solo uno del otro conjunto.

Usar:  $\mu_1 - \mu_2 = 0$

### Prueba T dos muestras independientes:

Las observaciones no están relacionadas

Usar:  $\mu_1 - \mu_2 = d_0$

Script	Pág
Prueba Z una muestra	3
Prueba T una muestra	7
Muestras pareadas	9
Muestras Indep.	11

## Capítulo 6

**Error Tipo 1:** Rechazar  $H_0$  cuando es verdadera

- Alfa = probabilidad de cometer este error

**Error Tipo 2:** Aceptar  $H_0$  cuando  $H_A$  es verdadera

- Beta = probabilidad de cometer este error

### Poder estadístico:

Probabilidad de no cometer un error de tipo 2 (1-beta)

Probabilidad de rechazar  $H_0$  cuando esta es falsa

### Tamaño de efecto:

Que tan grande es la diferencia entre dos grupos.

**OBS:** mayor tamaño de efecto implica un mayor poder estadístico.

Script	Pág.
Poder teórico	8
Poder en R	11

## Capítulo 7

### Método de Wald o Wilson:

- Las observaciones de la muestra son independientes.
- Al menos 10 obs corresponden a éxito y al menos 10 a fracaso

**OBS:**  $p_0$  = proporción de estudio

- 70 % de la población, es decir,  $p_0 = 0.7$

**OBS:**

- $n \cdot p_0 = n$  de obs de éxito
- $n \cdot (1 - p_0) = n$  de obs de fracaso

### Método de Wald dos proporciones:

#### Condiciones:

- Cada proporción, por separado, sigue el modelo Normal.
- Las dos muestras son indep. Una de la otra.

Script	Pág.
Wald teórico	3
Wald 2 prop.	5
Wilson 1 prop.	8

## Capítulo 8

### Chi Cuadrado:

Utiliza tablas de contingencia entre dos variables categóricas, donde una presenta varios niveles

**Homogeneidad:** Se usa para ver si dos poblaciones tienen las mismas proporciones en los distintos niveles de la variable categórica.

**Bondad de ajuste:** Determina si una distribución se asemeja a una distribución esperada, usada para ver si una muestra es representativa de la población.

**Prueba de independencia:** Determina si dos variables categóricas de una misma población son independientes o no.

### Muestras Pequeñas:

**Fisher:** Ve si las variables son indep. o están relacionadas.

**McNemar:** usada para dos muestras pareadas

**Q de Cochran:** cuando son mas de dos muestras pareadas.

Script	Pág
Chi-Homogeneidad	4
Chi-Bondad	5
Chi-Indep	7
Fisher	9
McNemar	11
Q de Cochran	14

