# Anotaciones Examen Parcial

Jesus Mauricio Huayhua Flores

2024-05-12

## Librerias

```r
library(DescTools)
library(ggplot2)
library(MASS)
```

## Lectura de archivos

```r
# Para leer archivos xlsx
data <- read_xlsx(path="directory/file.xlsx")
head(data)
# Para leer archivos csv
data <- read.csv("directory/file.csv")
```

## Pruebas de bondad de ajuste

### Distribución multinominal

- Repetir n veces, de forma independiente
- $k(k \geqslant 2)$ resultados o categorias
- $C_1, C_2, \dots, C_k$ con $p_1, p_2, \dots, p_k$
- $\sum_{i=1}^{k} p_i = 1$
- Notación: $(X_1, X_2, \dots, X_k) \sim Mult(n, p_1, p_2, \dots, p_k)$

#### Función de probabilidad conjunta

Si $(X_1, X_2, \dots, X_k) \sim Mult(n, p_1, p_2, \dots, p_k)$

$$P(X_1 = x_1, \dots, X_k = x_k) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}$$

- $X_1, X_2, \dots, X_k$ **NO** son independientes - $X_1 \sim B(n, p_i), \forall i = 1, \dots, k$ - $E(X_i) = np_i$ - $V(X_i) = np_i(1 - p_i)$ - $Cov(X_i, X_j) = -np_ip_j, (si)\ i \neq j$

### Pruebas Chi-cuadrado

Utilizado para contrastar hipótesis acerca de los parámetros, de 1 o varias distribuciones multinominales.

$$W = \sum_{i=1}^{k} \frac{(X_i - E_i)^2}{E_i}$$

**Teorema 1**

Si **n** es grande, $\forall i = 1, \ldots, k$ se cumple que $E_i = np_i \geqslant 5$

$$W = \sum_{i=1}^{k} \frac{(X_i - E_i)^2}{E_i} = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i}$$

Tiene distribución Chi-cuadrado con (K - 1) grados de libertad $W \sim X^2_{(k-1)}$

## Prueba Chi-cuadrado para bondad de ajuste

Es la prueba para determinar si la variable que observamos se ajusta o no a una distribución teórica.

$$\begin{cases} H_0 : & F_Y = F_o \\ H_1 : & F_Y \neq F_o \end{cases}$$

Experimento multinominal de $k$ categorías excluyentes.

- $C_1, \ldots, C_k$ con probablidades $p_1, \ldots, p_k$
- Frecuencias observadas $x_1, \ldots, x_2$, donde $\sum_{i=1}^{k} x_i = n$
- $X_i$ números de veces en que ocurre la categoría $C_i$
- Contrastar si las frecuencias observadas de cada categoría difieren significativamente, de las frecuencias esperadas $e_i = np_i$

$$\begin{cases} H_0 : & \forall i : p_i = p_i^o \\ H_1 : & \exists i | p_i \neq p_i^o \end{cases}$$

Si $e_i = np_i^0 \geqslant 5$, la estadística de prueba:

$$W = \sum_{i=1}^{k} \frac{(X_i - np_i^0)^2}{np_i^0} = \sum_{i=1}^{k} \frac{(O_i - e_i^0)^2}{ei}$$

Punto crítico, se distribuye Chi-cuadrado con $k-1$ grados de libertad: $W \sim X^2_{k-1}$.

# Pasos para la prueba de Chi-cuadrado(x^2) de bondad de ajuste

1. Establecer las hipótesis sobre la función de distribución desconocida

$$\begin{cases} H_0 & : F_Y = F_0 \\ H_1 & : F_Y \neq F_0 \end{cases}, con\ F_0\ conocida$$

2. Obtener una tabla aleatoria de tamaño **n** de $Y$, con una tabla de distribución de frecuencias.

`| Intervalos | marca de clase | frecuencia observada | frecuencia esperada |`

3. Se calculan las frecuencias esperadas $E_i^0$
4. Se calcula Estadístico de prueba

$$U_0 = \sum_{i=1}^{k} \frac{(O_i - E_i^0)^2}{E_i^0} \sim X_(k - m - 1)^2$$

5. Se recha si $U_0 > X^2_{1-\alpha,k-m-1}$, donde $r$ es la cantidad de parametros desconocidos.

## Comandos utilices en R

**Opción1:**

```
normalityTest(horas ~ pastillas, test="shapiro.test", data=Dataset)
```

**Opción2:**

```
# Histograma
auxhist <- hist(X,xlab="Largo (mm)", main= "Histograma del Largo")   #   Al crear el histograma se guardan las características de la distribución.
# Prueba de bondad de ajuste
cortes_histo <- auxhist$breaks
probAcum_DistNormal <- pnorm(cortes_histo, mean = mean(X), sd = sd(X))
n <- length(probAcum_DistNormal)
intervalos <- data.frame(probAcum_DistNormal[-n],probAcum_DistNormal[-1])
prop.esperada <- intervalos[,2]-intervalos[,1]
freq.esperada <- (prop.esperada/sum(prop.esperada))*length(X)
freq.observada <- auxhist$counts
chisq.test(freq.observada, p=prop.esperada, rescale.p=TRUE, simulate.p.value = T)
```