# Distribuciones y Prueba de HIpotesis

Paola Neuta

2025-04-28

## Taller de distribuciones, normalidad y pruebas de hipotesis

En este Taller se realizon 8 puntos, utilizando conceptos como la media, la desviación estandar, los niveles de confianza, los intervalos de confianza, entre otros, para determinar conclusiones de ciertas poblaciones a partir de muestras estudiadas.

A continuación se presentarán los puntos desarrollados:

### 1) Estudio de la variable talla para Recien Nacidos

Para el primer punto se utiliza la función pnorm() la cual nos permite conocer la probabilidad de que un evento ocurra, teniendo en cuenta la media de una población, así como, su desviación estandar.

```
pnorm(45,50,2, lower.tail = TRUE)  # Probabilidad de que mida menos de 45 cm el recien nacido
```

```
## [1] 0.006209665
```

```
pnorm(52,50,2, lower.tail = FALSE) # Probabilidad de que mida a lo menos 52 cm
```

```
## [1] 0.1586553
```

```
ninos <- (pnorm(52,50,2) - pnorm(48,50,2)) * 100; ninos  # Porcentaje de recien nacidos que tienen
```

```
## [1] 68.26895
```

```
                                                         # talla entre 48 y 52 cm

floor(200 * ninos/100) # Cantidad de niños de talla entre 48 y 52 cm
```

```
## [1] 136
```

La probabilidad de que un recien nacido mida menos de 45 cm es de 0.62 %. La probabilidad de que un recien nacido mida a lo menos 52 cm es de 15.86 %. El porcentaje de recien nacidos que tienen una talla entre 48 y 52 cm es de 68.27 %. La cantidad de recien nacidos de talla entre 48 y 52 cm de 200 es 136. Podemos realizar estas conclusiones teniendo en cuenta que es una distribución normal.

### 2) ¿Cuántos días se espera tener una temperatura entre 39 y 42 °C?

Para el segundo punto se halla la probabilidad de que el evento ocurra, nuevamente usando la función pnorm(). Esta probabilidad se aplicada a la cantidad de dias de un mes como sigue:

```
temperture <- pnorm(42, 36, 3) - pnorm(39, 36, 3); temperture
```

```
## [1] 0.1359051
```

```
round(30 * temperture, 1) # Cantidad de días que tienen una temperatura entre 39 y 42 °C
```

```
## [1] 4.1
```

La probabilidad de tener una temperatura entre 39 y 42 °C es de 13.6 %. Es decir 4 días en el mes de abril.

### 3) Intervalo de confianza para la venta media por trabajador

Para determinnar el intervalo de confianza para la venta media por trabajador de la tienda de libros a un nivel de confianza de 90, es necesario importar la libreria BSDA y utilizar la función correspondiente a un test t-Student, debido a que el tamaño de la muestra escogida es menor a 30. Además, se halla el valor de la desviación estandar ya que se nos es dado el valor de la varianza.

De aquí se puede saber que el intervalo de confianza se encuentra entre 4.357 y 5.643.

### 4) Intervalo de confianza de la demanda diaria de un producto

Para el cuarto punto hay que hallar el valor de la media y la desviación estandar de la muestra, como no conocemos la desviación estandar de la población y además el tamaño de la muestra es menor a 30, nuevamente utilizamos el test t-Student para hallar el intervalo de confianza.

```
demandaProducto <- c(35,44,38,55,33,56,67,45,48,40,43)
demandaMedia <- mean(demandaProducto)
demandaDesviacion <- sd(demandaProducto)
tsum.test(mean.x = demandaMedia, s.x = demandaDesviacion, n.x = 11, conf.level = 0.95)
```

```
## Warning in tsum.test(mean.x = demandaMedia, s.x = demandaDesviacion, n.x = 11,
## : argument 'var.equal' ignored for one-sample test.
```

```
##
##  One-sample t-Test
##
## data:  Summarized x
## t = 14.976, df = 10, p-value = 3.551e-08
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  39.00127 52.63510
## sample estimates:
## mean of x
##  45.81818
```

Como podemos ver, el intervalo de confianza se encuentra entre 39.00 y 52.63.

**5) Hipotesis nula de la media de las deudas de una empresa**

Para este punto hay que decir si se rechaza o acepta la hipotesis nula. En este caso la hipotesis nula es la que se presenta en el enunciado, es decir, H0 = La media de las deudas por cobrar de una empresa es de 160.000 euros. Podemos utilizar el test t-Student y mirar cuál es el valor de p.

```
tsum.test(mean.x = 188000, s.x = 81000, n.x = 50, mu = 160000, alternative = "two.sided", conf.level = 0
```

```
## Warning in tsum.test(mean.x = 188000, s.x = 81000, n.x = 50, mu = 160000, :
## argument 'var.equal' ignored for one-sample test.
```

```
##
##  One-sample t-Test
##
## data:  Summarized x
## t = 2.4443, df = 49, p-value = 0.01816
## alternative hypothesis: true mean is not equal to 160000
## 95 percent confidence interval:
##  164980.1 211019.9
## sample estimates:
## mean of x
##    188000
```

Se obtiene que el valor de p es p = 0,018 < 0,05, esto quiere decir que la hipotesis nula se rechaza y la media es significativamente diferente a 160.000.

**6) Hipotesis de la vida media de un componente electrónico**

Para este sexto punto se presenta la afirmación de que la vida media de un componente electrónico es mayor a 1500 horas y se pide aceptar o rechazar la hipotesis. Como el tamalo de la muestra es muy grande, se puede utilizar la instruncción zsum.test() de una distribución normal.

```
zsum.test(mean.x = 1450, sigma.x = sqrt(650), n.x = 900, mu = 1500, alternative = "greater", conf.level
```

```
##
##  One-sample z-Test
##
## data:  Summarized x
## z = -58.835, p-value = 1
## alternative hypothesis: true mean is greater than 1500
## 95 percent confidence interval:
##  1448.602       NA
## sample estimates:
## mean of x
##      1450
```

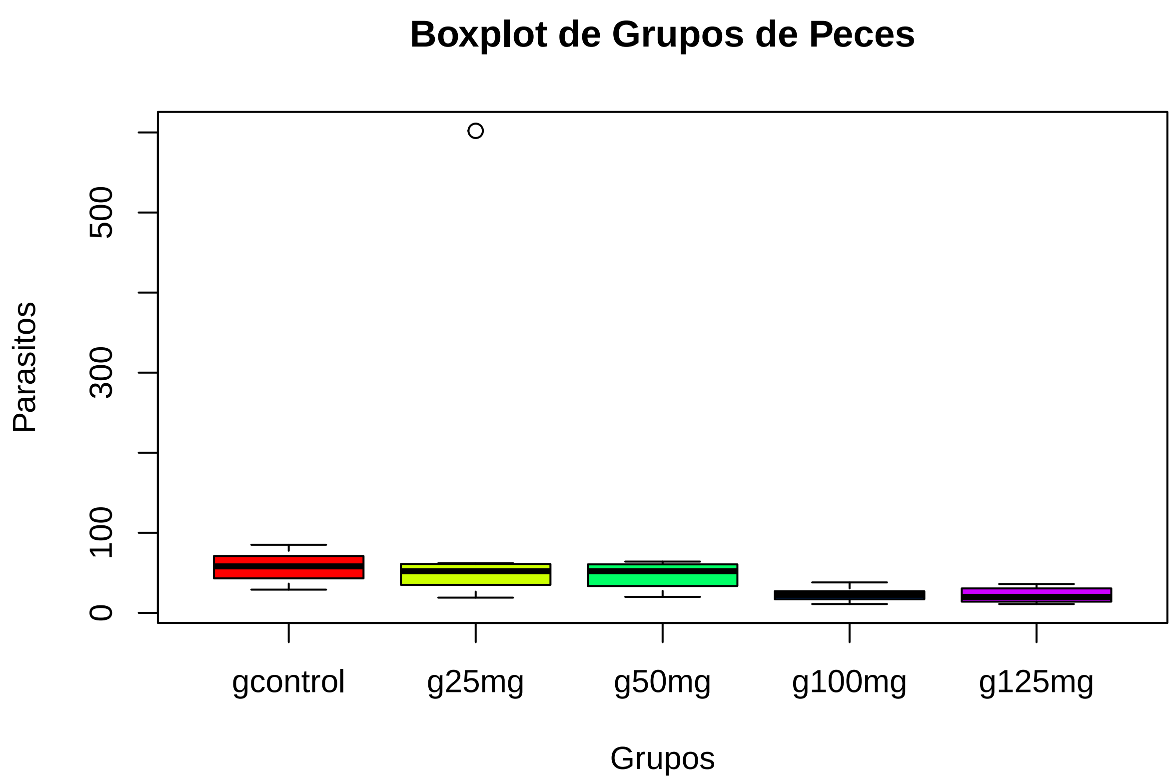
La hipotesis nula se toma como el caso contrario del enunciado, es decir, H0= La vida media de un componente electrónico es menor a 1500 horas. Queremos entonces rechazar esta hipotesis para saber que H1 es verdadero, sin embargo se obtiene que el valor de p > 0.05, entonces no se puede afirmar que la vida media de un componente electrónico es mayor a 1500 horas. Por lo tanto no se rechaza la hipotesis nula H0 de que la vida media de un comoponente sea menor a 1500.

**7) Analizar la normalidad de los datos de Pesos de perosnas por**

**8) Analisis de Normalidad y Homocedasticidad de Medicamento para Parasitos de peces criados en acuicultura**

```
gcontrol <- c(50, 65, 72, 46, 38, 29, 70, 85, 72, 40, 57, 59)
g25mg <- c(49, 47, 30, 602, 62, 60, 19, 28, 56, 62, 55, 40)
g50mg <- c(20, 59, 64, 61, 28, 47, 29, 41, 60, 57, 61, 38)
g100mg <- c(20, 23, 38, 31, 27, 16, 27, 18, 22, 12, 24, 11)
g125mg <- c(18, 30, 22, 26, 31, 11, 15, 12, 31, 36, 16, 13)

Medicamentos_df <- data.frame(gcontrol, g25mg, g50mg, g100mg, g125mg)
boxplot(as.matrix(Medicamentos_df), col = rainbow(5), xlab="Grupos",
        ylab="Parasitos", main = "Boxplot de Grupos de Peces")
```




```
shapiro.test(gcontrol)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  gcontrol
## W = 0.97409, p-value = 0.9486
```

```r
shapiro.test(g25mg)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  g25mg
## W = 0.41016, p-value = 4.136e-06
```

```r
shapiro.test(g50mg)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  g50mg
## W = 0.87882, p-value = 0.08465
```

```r
shapiro.test(g100mg)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  g100mg
## W = 0.97573, p-value = 0.9607
```

```r
shapiro.test(g125mg)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  g125mg
## W = 0.91021, p-value = 0.2147
```

```r
MedicamentosNormalidadP <- data.frame(Grupo = c("Control", "25mg", "50mg", "100mg", "125mg"),
                                      Pvalue = c(0.9486, 4.136e-06, 0.08465, 0.9607, 0.2147)); Medicamen
```

```
##     Grupo    Pvalue
## 1 Control 9.486e-01
## 2    25mg 4.136e-06
## 3    50mg 8.465e-02
## 4   100mg 9.607e-01
## 5   125mg 2.147e-01
```

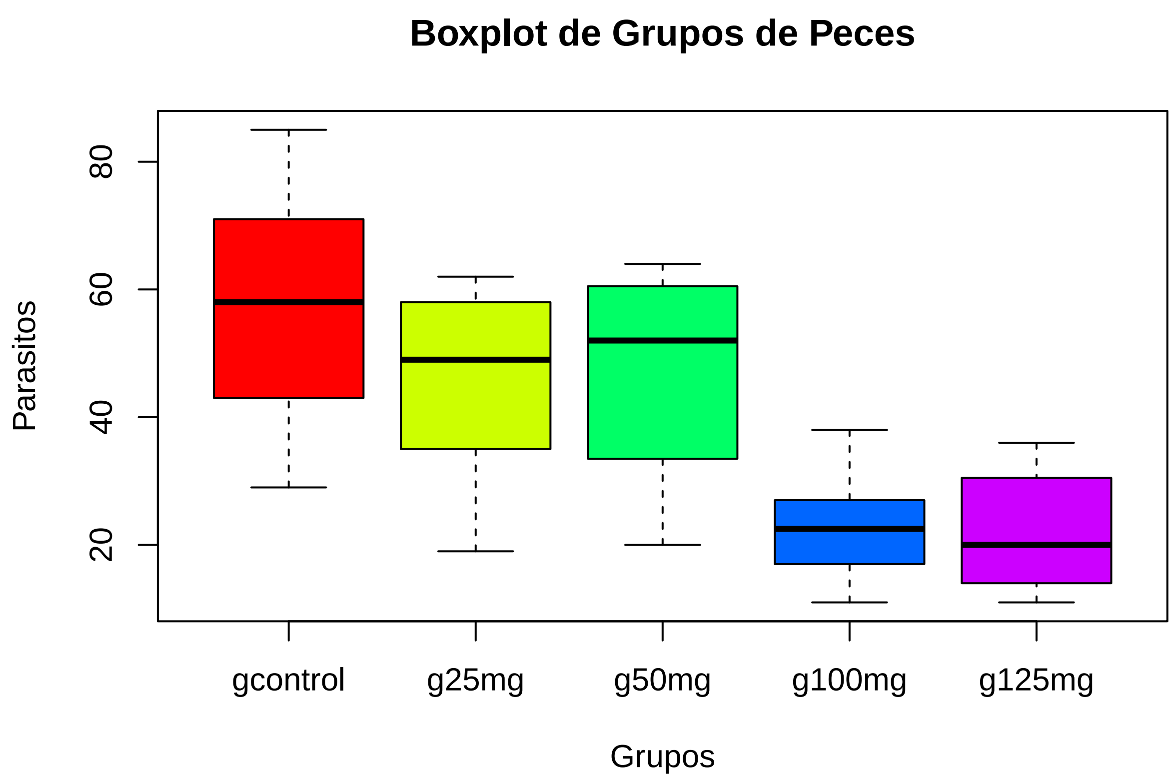
```r
# Elimino el valor atípico
g25mg <- c(49, 47, 30, 49, 62, 60, 19, 28, 56, 62, 55, 40)
Medicamentos_df <- data.frame(gcontrol, g25mg, g50mg, g100mg, g125mg)
shapiro.test(g25mg)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  g25mg
## W = 0.90623, p-value = 0.1908
```

```r
MedicamentosNormalidadP <- data.frame(Grupo = c("Control", "25mg", "50mg", "100mg", "125mg"),
                                      Pvalue = c(0.9486, 0.1908, 0.08465, 0.9607, 0.2147)); Medicamentos
```

```
##     Grupo  Pvalue
## 1 Control 0.94860
## 2    25mg 0.19080
## 3    50mg 0.08465
## 4   100mg 0.96070
## 5   125mg 0.21470
```

```r
boxplot(as.matrix(Medicamentos_df), col = rainbow(5), xlab="Grupos", ylab="Parasitos", main = "Boxplot d
```



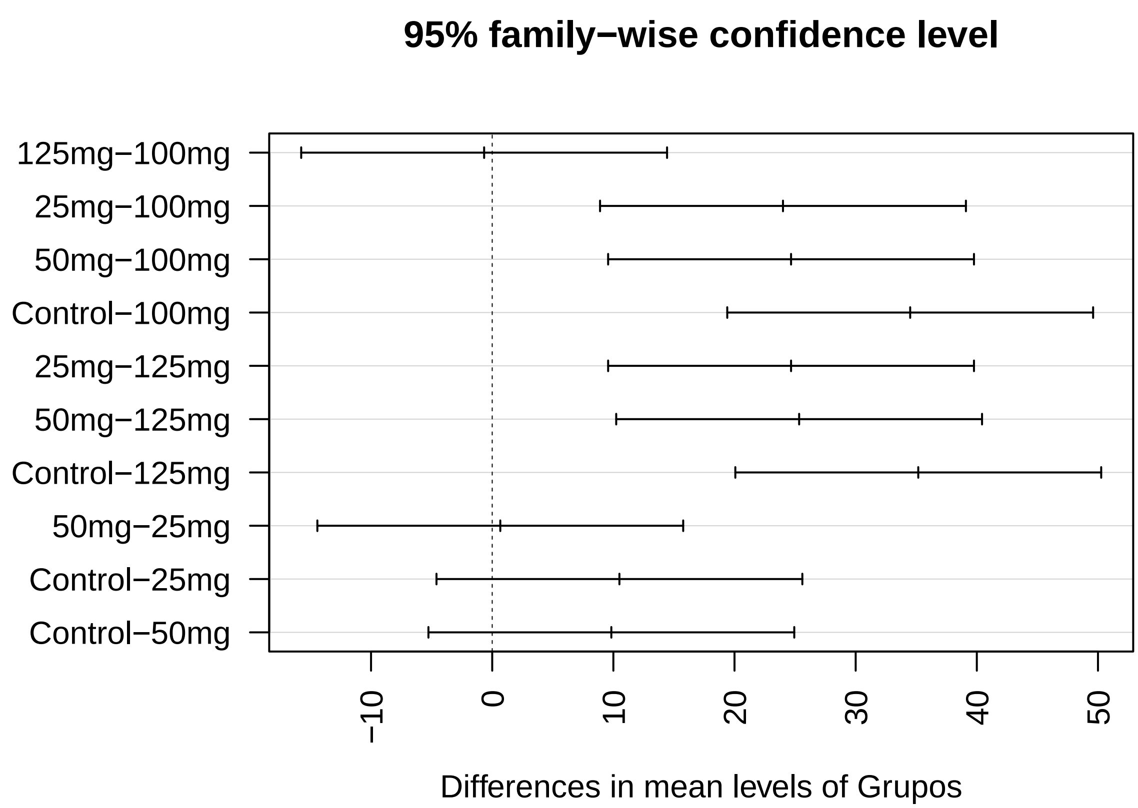

```r
#Homocedasticidad
Parasitos <- c(gcontrol, g25mg, g50mg, g100mg, g125mg)
Grupos <- factor(rep(c("Control", "25mg", "50mg", "100mg", "125mg"),
                     times = c(length(gcontrol), length(g25mg), length(g50mg), length(g100mg), length(g
bartlett.test(Parasitos ~ Grupos)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  Parasitos by Grupos
## Bartlett's K-squared = 9.1976, df = 4, p-value = 0.05635
```

```r
#Tabla Anova y gráfico de intervalos
AnovaTable <- aov(lm(Parasitos ~ Grupos)); summary(AnovaTable)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Grupos       4  12167    3042   17.69 2.19e-09 ***
## Residuals   55   9458     172
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
intervalos <- TukeyHSD(AnovaTable)
par(mar=c(5, 8, 4, 2))
plot(intervalos, las =2)
```




Fin.