

Actividad2.7

Ricardo Kaleb Flores Alfonso, A01198716

2024-10-12

```
library(ggplot2)
library(stats)
```

1) 0) Se importan los datos

```
df <- read.csv("resistencia.csv")
```

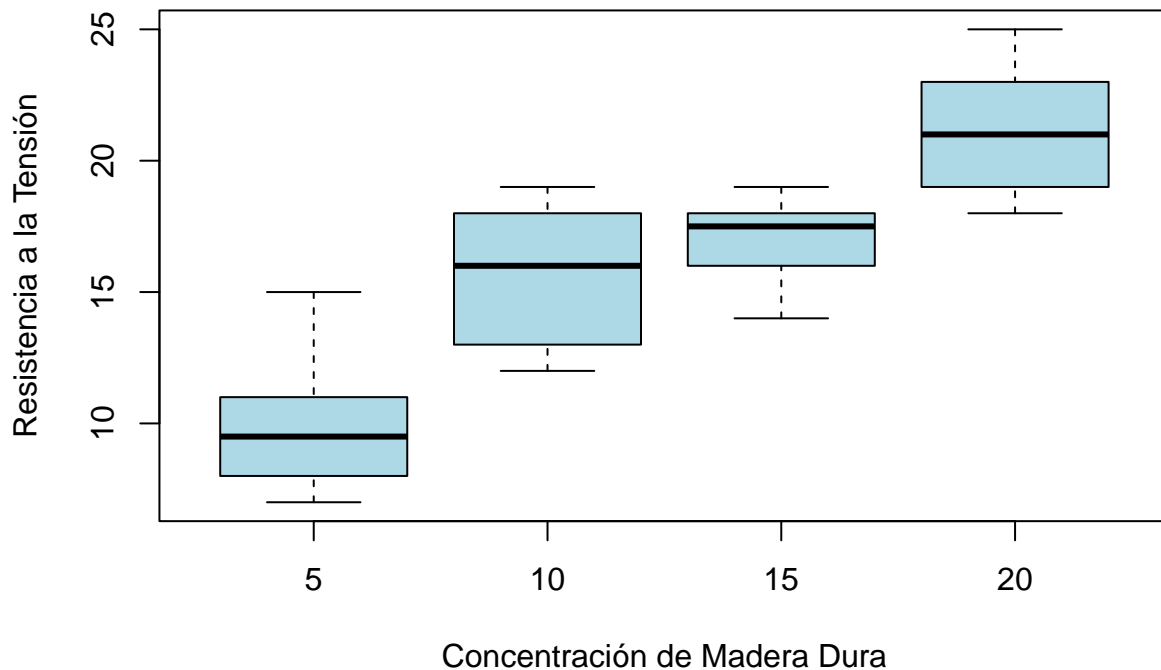
1) Análisis exploratorio

```
# Calcular la media de la resistencia por tratamiento (Concentración de Madera Dura)
mean_resistance <- aggregate(Resistencia ~ Concentracion, data = df, FUN = mean)
print(mean_resistance)
```

```
##   Concentracion Resistencia
## 1             5    10.00000
## 2            10    15.66667
## 3            15    17.00000
## 4            20    21.16667
```

```
# Crear el boxplot de Resistencia a la Tensión por Concentración de Madera Dura
boxplot(Resistencia ~ Concentracion, data = df,
        main = "Boxplot de Resistencia a la Tensión por Concentración de Madera Dura",
        xlab = "Concentración de Madera Dura",
        ylab = "Resistencia a la Tensión",
        col = "lightblue", border = "black")
```

Boxplot de Resistencia a la Tensión por Concentración de Madera Dura



Se observa que mientras aumenta la concentración de madera, el promedio de resistencia también aumenta, lo que sugiere una relación lineal positiva entre las variables.

2) Hipótesis estadística

Hipótesis nula H_0 : No hay diferencias significativas en la resistencia a la tensión entre los diferentes niveles de concentración de madera dura.

Hipótesis alternativa H_1 : Existen diferencias significativas en la resistencia a la tensión entre al menos dos niveles de concentración de madera dura.

3) Prueba de Anova

```
#Calcular la media global
media_global <- mean(df$Resistencia)

#Calcular la suma de cuadrados entre grupos
ss_entre <- sum(aggregate(Resistencia ~ Concentracion, df, function(x) length(x) * (mean(x) - media_global)^2))

#Calcular la suma de cuadrados intra grupos
ss_intra <- sum((df$Resistencia - ave(df$Resistencia, df$Concentracion))^2)

#Calcular las medias de cuadrados
df_entre <- length(unique(df$Concentracion)) - 1 # Grados de libertad entre grupos
df_intra <- nrow(df) - length(unique(df$Concentracion)) # Grados de libertad intra grupos
```

```

ms_entre <- ss_entre / df_entre
ms_intra <- ss_intra / df_intra

# Calcular el estadístico F
F_value <- ms_entre / ms_intra

# Calcular el p-valor basado en la distribución F
p_value <- pf(F_value, df_entre, df_intra, lower.tail = FALSE)

# Mostrar resultados
cat("Suma de cuadrados entre grupos:", ss_entre, "\n")

## Suma de cuadrados entre grupos: 382.7917
cat("Suma de cuadrados intra grupos:", ss_intra, "\n")

## Suma de cuadrados intra grupos: 130.1667
cat("Estadístico F:", F_value, "\n")

## Estadístico F: 19.60521
cat("p-valor:", p_value, "\n")

## p-valor: 3.592578e-06

```

4) Prueba de Anova en R

```

modelo <- aov(Resistencia ~ as.factor(Concentracion), data = df)
summary(modelo)

##               Df Sum Sq Mean Sq F value    Pr(>F)
## as.factor(Concentracion)  3  382.8   127.60    19.61 3.59e-06 ***
## Residuals                20   130.2     6.51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Mediante las pruebas de Anova se obtuvo un valor de $p < 0.05$, por lo que se rechaza la hipótesis nula, la hipótesis alternativa muestra que existe diferencia significativa entre al menos dos niveles de concentración de madera

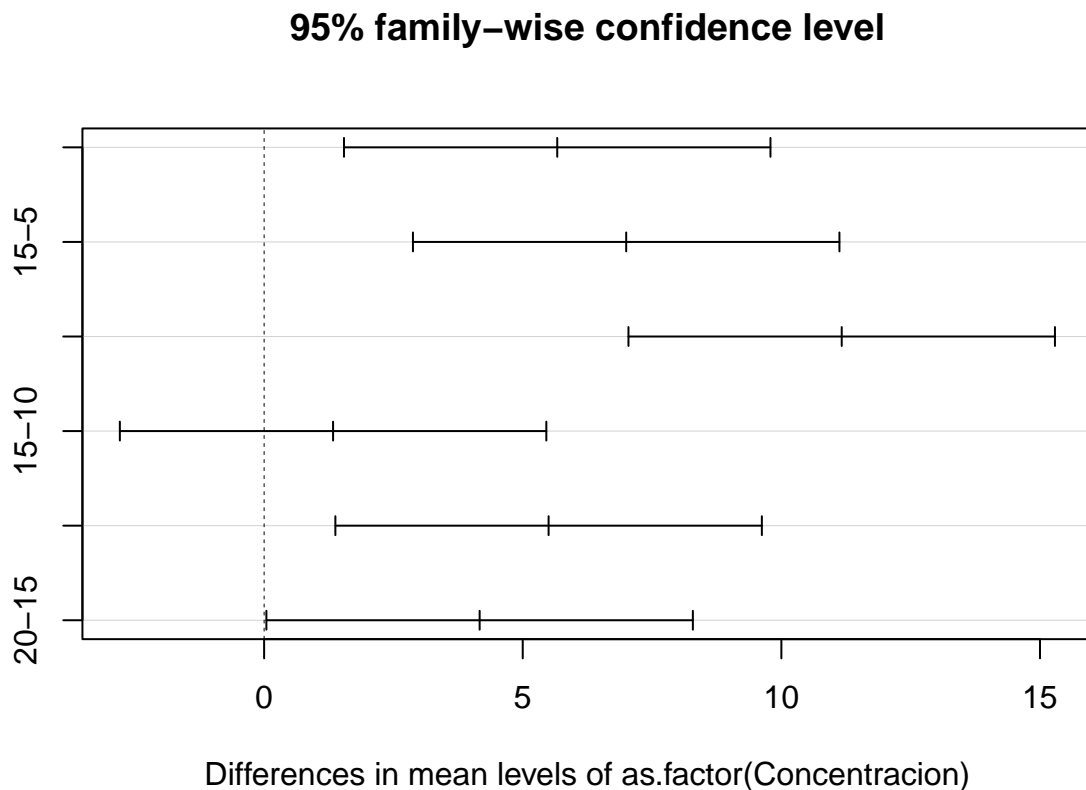
5) Diferencias por pares

```

tukey_result <- TukeyHSD(modelo)

plot(tukey_result)

```



El gráfico Tukey muestra las diferencias entre los niveles. Si los intervalos de confianza no incluyen 0, hay diferencias significativas entre esos pares de concentraciones, es por esto que se observa que existe una diferencia significativa entre los niveles 15 y 5, así como los niveles 20 y 15.

6) Validación de supuestos

6.1) Normalidad

Hipótesis nula H_0 : No existe normalidad en los residuos del modelo

Hipótesis alternativa H_1 : La normalidad en los residuos son normales

```
# Test de Shapiro-Wilk para normalidad
shapiro_test <- shapiro.test(residuals(modelo))
print(shapiro_test)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(modelo)
## W = 0.96624, p-value = 0.5757
```

El test de Shapiro-Wilk obtuvo un p-valor de 0.5757, por lo que no hay evidencia para rechazar la normalidad de los residuos. Así que el supuesto se cumple

6.2) Homocedasticidad

Hipótesis nula H_0 : Existe homocedasticidad en los residuos

Hipótesis alternativa H_1 : La varianza varía entre los grupos, heterocedasticidad

```
library(lmtest)

## Cargando paquete requerido: zoo
##
## Adjuntando el paquete: 'zoo'
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric
# Test de Bartlett para homocedasticidad
bartlett_test <- bartlett.test(Resistencia ~ Concentracion, data = df)
print(bartlett_test)

##
## Bartlett test of homogeneity of variances
##
## data: Resistencia by Concentracion
## Bartlett's K-squared = 1.1352, df = 3, p-value = 0.7686
```

El test de Bartlett obtuvo un valor de $p = 0.7686 > 0.05$, no hay evidencia para rechazar la hipótesis nula, por lo que el supuesto de homocedasticidad se cumple.

6.3) Independencia

Hipótesis nula H_0 : Las observaciones se obtuvieron de manera independiente

Hipótesis alternativa H_1 : Las observaciones tienen correlación entre ellas.

```
tabla <- table(df$Resistencia, df$Concentracion)

chisq.test(tabla)

## Warning in chisq.test(tabla): Chi-squared approximation may be incorrect
##
## Pearson's Chi-squared test
##
## data: tabla
## X-squared = 46, df = 48, p-value = 0.5551
```

Se obtuvo un p valor de 0.5551, lo que sugiere que no hay evidencia para rechazar la independencia de las observaciones.

7) Intervalos de confianza

```
intervalos_confianza <- tapply(df$Resistencia, df$Concentracion, function(x) t.test(x)$conf.int)
print(intervalos_confianza)

## $`5`
## [1] 7.031748 12.968252
## attr(,"conf.level")
## [1] 0.95
##
## $`10`
```

```
## [1] 12.72325 18.61008
## attr(,"conf.level")
## [1] 0.95
##
## $`15`
## [1] 15.12271 18.87729
## attr(,"conf.level")
## [1] 0.95
##
## $`20`
## [1] 18.39674 23.93659
## attr(,"conf.level")
## [1] 0.95
```

Estos intervalos muestran que las medias de los diferentes niveles de concentración están separadas, lo que da evidencias para demostrar los resultados en el ANOVA y la prueba de Tukey. # 8) Conclusión

El análisis sugiere que aumentar la concentración de madera dura aumenta significativamente la resistencia a la tensión. Todos los supuestos del modelo ANOVA se cumplieron, lo que valida la robustez del análisis.

En resumen, los diferentes niveles de concentración tienen efectos significativamente diferentes sobre la resistencia, y este efecto se ha validado estadísticamente y mediante los intervalos de confianza.