Ejercicio ANOVA

Denisse Aurora Alvarado Reyna

2025-09-22

```r
# Introduccion
# Este analisis evalua las diferencias en concentraciones de estroncio (mg/ml)
# entre cinco cuerpos de agua mediante ANOVA de una via y pruebas post-hoc

library(tidyverse)
library(agricolae)
library(car)
library(ggplot2)

grayson <- c(28.2, 33.2, 36.4, 34.6, 29.1, 31.0)
beaver <- c(39.6, 40.8, 37.9, 37.1, 43.6, 42.4)
angler <- c(46.3, 42.1, 43.5, 48.8, 43.7, 40.1)
appletree <- c(41.0, 44.1, 46.4, 40.2, 38.6, 36.3)
rock <- c(56.3, 54.1, 59.4, 62.7, 60.0, 57.3)

datos <- data.frame(
  concentracion = c(grayson, beaver, angler, appletree, rock),
  sitio = factor(rep(c("Grayson", "Beaver", "Angler", "Appletree", "Rock"), each = 6))
)

str(datos)
```

```
## 'data.frame':    30 obs. of  2 variables:
##  $ concentracion: num  28.2 33.2 36.4 34.6 29.1 31 39.6 40.8 37.9 37.1 ...
##  $ sitio        : Factor w/ 5 levels "Angler","Appletree",..: 4 4 4 4 4 4 3 3 3 3 ...
```

```r
head(datos, 10)
```

```
##    concentracion    sitio
## 1           28.2  Grayson
## 2           33.2  Grayson
## 3           36.4  Grayson
## 4           34.6  Grayson
## 5           29.1  Grayson
## 6           31.0  Grayson
## 7           39.6   Beaver
## 8           40.8   Beaver
## 9           37.9   Beaver
## 10          37.1   Beaver
```

```
# Estadisticas descriptivas
estadisticas <- datos %>%
  group_by(sitio) %>%
  summarise(
    n = n(),
    media = mean(concentracion),
    desv_est = sd(concentracion),
    varianza = var(concentracion),
    minimo = min(concentracion),
    maximo = max(concentracion),
    .groups = 'drop'
  )

print(estadisticas)
```

```
## # A tibble: 5 x 7
##   sitio         n media desv_est varianza minimo maximo
##   <fct>     <int> <dbl>    <dbl>    <dbl>  <dbl>  <dbl>
## 1 Angler        6  44.1     3.08     9.49   40.1   48.8
## 2 Appletree     6  41.1     3.67    13.4    36.3   46.4
## 3 Beaver        6  40.2     2.53     6.40   37.1   43.6
## 4 Grayson       6  32.1     3.21    10.3    28.2   36.4
## 5 Rock          6  58.3     3.04     9.22   54.1   62.7
```

```
cat("Media general:", mean(datos$concentracion), "mg/ml\n")
```

```
## Media general: 43.16 mg/ml
```

```
cat("Numero total de observaciones:", nrow(datos), "\n")
```
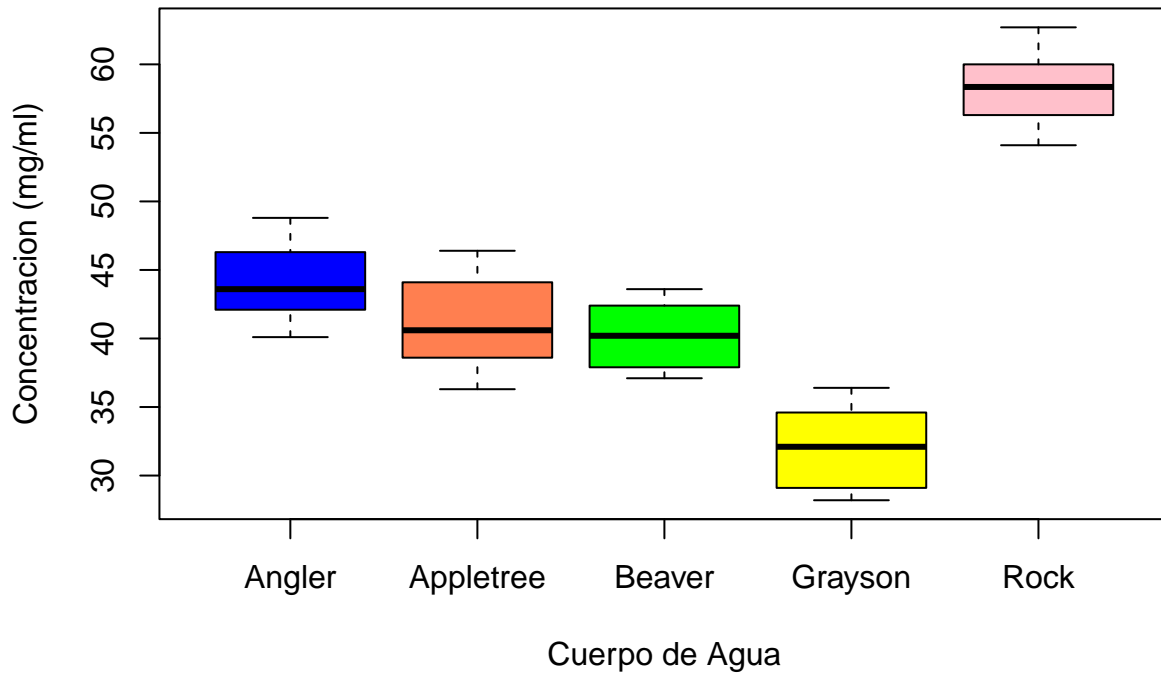
```
## Numero total de observaciones: 30
```

```
cat("Numero de grupos:", length(unique(datos$sitio)), "\n")
```

```
## Numero de grupos: 5
```

```
# Visualizacion
boxplot(concentracion ~ sitio, data = datos,
        main = "Concentraciones de Estroncio por Sitio",
        xlab = "Cuerpo de Agua",
        ylab = "Concentracion (mg/ml)",
        col = c("blue", "coral", "green", "yellow", "pink"))
```

Concentraciones de Estroncio por Sitio


```r
# ANOVA
modelo_anova <- aov(concentracion ~ sitio, data = datos)
summary(modelo_anova)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## sitio        4 2193.4   548.4   56.16 3.95e-12 ***
## Residuals   25  244.1     9.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Verificacion de supuestos
residuos <- residuals(modelo_anova)
shapiro.test(residuos)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuos
## W = 0.96415, p-value = 0.3935
```

```r
leveneTest(concentracion ~ sitio, data = datos)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
```

```
## group   4   0.1846 0.9442
##         25
```

```
# Pruebas post-hoc
prueba_lsd <- LSD.test(modelo_anova, "sitio", alpha = 0.05)
print(prueba_lsd)
```

```
## $statistics
##    MSerror Df  Mean        CV  t.value      LSD
##     9.7652 25 43.16 7.240343 2.059539 3.715779
##
## $parameters
##         test p.ajusted name.t ntr alpha
##    Fisher-LSD      none  sitio   5  0.05
##
## $means
##            concentracion      std r       se      LCL      UCL  Min  Max     Q25
## Angler          44.08333 3.080530 6 1.275748 41.45588 46.71079 40.1 48.8 42.450
## Appletree       41.10000 3.666061 6 1.275748 38.47255 43.72745 36.3 46.4 39.000
## Beaver          40.23333 2.530349 6 1.275748 37.60588 42.86079 37.1 43.6 38.325
## Grayson         32.08333 3.205256 6 1.275748 29.45588 34.71079 28.2 36.4 29.575
## Rock            58.30000 3.036445 6 1.275748 55.67255 60.92745 54.1 62.7 56.550
##              Q50    Q75
## Angler     43.60 45.650
## Appletree  40.60 43.325
## Beaver     40.20 42.000
## Grayson    32.10 34.250
## Rock       58.35 59.850
##
## $comparison
## NULL
##
## $groups
##            concentracion groups
## Rock            58.30000      a
## Angler          44.08333      b
## Appletree       41.10000     bc
## Beaver          40.23333      c
## Grayson         32.08333      d
##
## attr(,"class")
## [1] "group"
```

```
n_por_grupo <- 6
gl_error <- 25
alpha <- 0.05
t_critico <- qt(1 - alpha/2, gl_error)
cme <- summary(modelo_anova)[[1]][2, 3]
lsd_valor <- t_critico * sqrt(cme * (2/n_por_grupo))
cat("Valor LSD =", round(lsd_valor, 4), "mg/ml\n")
```

```
## Valor LSD = 3.7158 mg/ml
```

4

```
# Tukey
prueba_tukey <- TukeyHSD(modelo_anova, conf.level = 0.95)
print(prueba_tukey)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = concentracion ~ sitio, data = datos)
##
## $sitio
##                          diff        lwr        upr     p adj
## Appletree-Angler   -2.9833333  -8.281979   2.315312 0.4791100
## Beaver-Angler      -3.8500000  -9.148645   1.448645 0.2376217
## Grayson-Angler    -12.0000000 -17.298645  -6.701355 0.0000053
## Rock-Angler        14.2166667   8.918021  19.515312 0.0000003
## Beaver-Appletree   -0.8666667  -6.165312   4.431979 0.9884803
## Grayson-Appletree  -9.0166667 -14.315312  -3.718021 0.0003339
## Rock-Appletree     17.2000000  11.901355  22.498645 0.0000000
## Grayson-Beaver     -8.1500000 -13.448645  -2.851355 0.0011293
## Rock-Beaver        18.0666667  12.768021  23.365312 0.0000000
## Rock-Grayson       26.2166667  20.918021  31.515312 0.0000000
```

```
tukey_test <- HSD.test(modelo_anova, "sitio", alpha = 0.05)
print("Grupos homogeneos (Tukey):")
```

```
## [1] "Grupos homogeneos (Tukey):"
```

```
print(tukey_test$groups)
```

```
##          concentracion groups
## Rock          58.30000      a
## Angler        44.08333      b
## Appletree     41.10000      b
## Beaver        40.23333      b
## Grayson       32.08333      c
```

```
k <- 5
q_critico <- qtukey(0.95, k, gl_error)
tukey_hsd <- q_critico * sqrt(cme/n_por_grupo)
cat("Valor Tukey HSD =", round(tukey_hsd, 4), "mg/ml\n")
```

```
## Valor Tukey HSD = 5.2986 mg/ml
```

```
# Grafico de Tukey
tukey_df <- data.frame(prueba_tukey$sitio)
tukey_df$Comparacion <- rownames(tukey_df)

ggplot(tukey_df, aes(x = diff, y = Comparacion)) +
  geom_point() +
  geom_errorbarh(aes(xmin = lwr, xmax = upr), height = 0.2) +
```

```r
geom_vline(xintercept = 0, linetype = "dashed", color = "red") +
labs(title = "IC95% Prueba de Tukey",
     x = "Diferencias entre medias de sitio",
     y = "Comparaciones") +
theme_minimal()
```

IC95% Prueba de Tukey
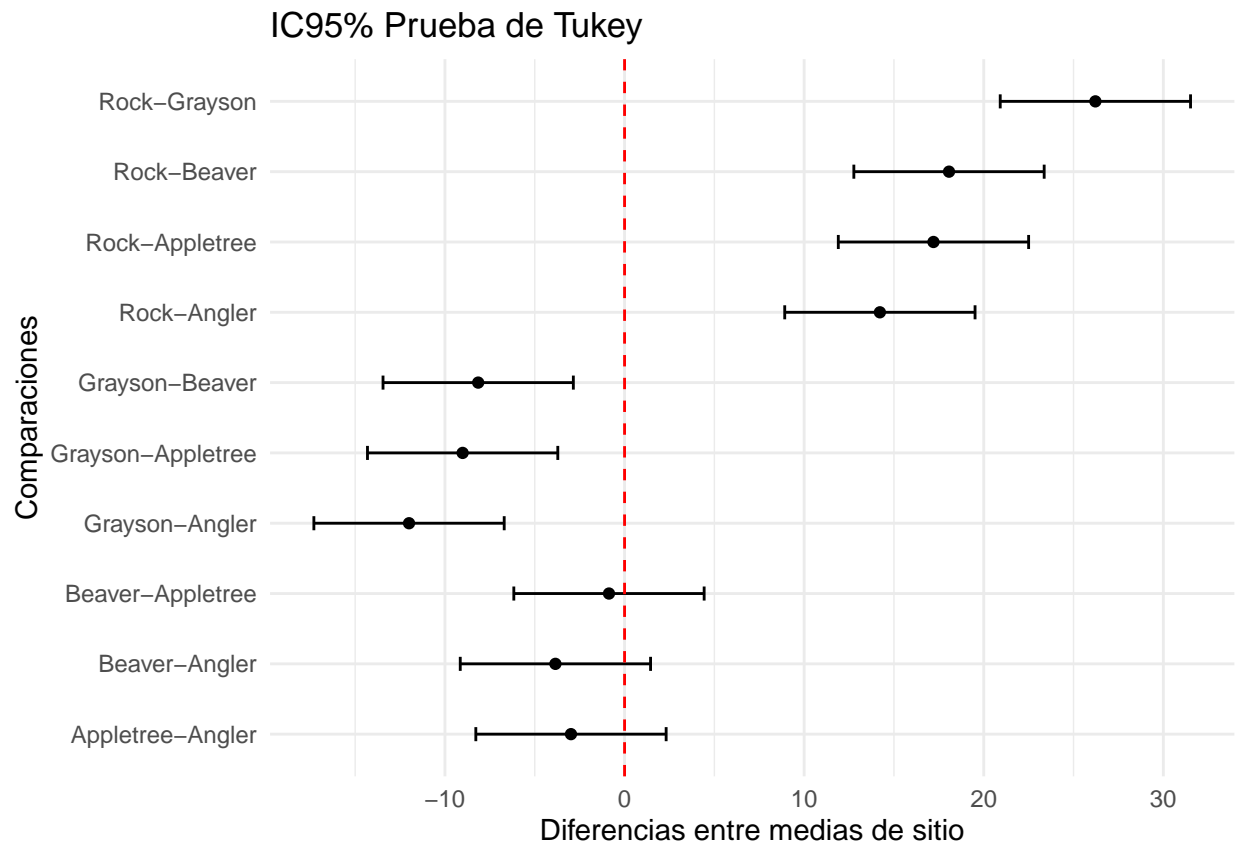

```r
# Comparacion de metodos
cat("COMPARACION DE METODOS:\n")
```

```
## COMPARACION DE METODOS:
```

```r
cat("LSD (alpha=0.05):", round(lsd_valor, 3), "mg/ml\n")
```

```
## LSD (alpha=0.05): 3.716 mg/ml
```

```r
cat("Tukey HSD:", round(tukey_hsd, 3), "mg/ml\n")
```

```
## Tukey HSD: 5.299 mg/ml
```

```r
cat("Tukey es mas conservador que LSD\n")
```

```
## Tukey es mas conservador que LSD
```

```
# Conclusiones
# ANOVA: F(4,25)=56.16, p<0.001. Se rechaza H0
# Rock River: concentraciones mas altas (58.30 mg/ml)
# Graysons Pond: mejores condiciones (32.08 mg/ml)
```