

Curso: Estadística en la investigación científica
Presenta: César Enrique Martínez Sánchez
Docente: Dr. Marco Aurelio González Tagle

ANOVA de estroncio en cinco cuerpos de agua

2025-09-22

Descripción del problema

Un investigador midió la concentración de **estroncio** (mg/ml) en cinco cuerpos de agua, con **6 réplicas por sitio (n = 6)**. Se desea evaluar si existen diferencias significativas entre las medias de los sitios mediante un **ANOVA de una vía**, seguido de comparaciones múltiples con **LSD** y **Tukey HSD**.

Datos

```
# Tabla de datos (n = 6 por sitio)
df <- data.frame(
  sitio = factor(rep(c("Grayson's Pond", "Beaver Lake", "Angler's Cove", "Appletree Lake", "Rock R
estroncio = c(
  28.2, 33.2, 36.4, 34.6, 29.1, 31.0,
  39.6, 40.8, 37.9, 37.1, 43.6, 42.4,
  46.3, 42.1, 43.5, 48.8, 43.7, 40.1,
  41.0, 44.1, 46.4, 40.2, 38.6, 36.3,
  56.3, 54.1, 59.4, 62.7, 60.0, 57.3
)
)

# Paquetes que usaremos
req <- c("dplyr", "ggplot2", "emmeans", "multcomp", "multcompView", "broom", "knitr", "kableExtra")
to_install <- req[!sapply(req, requireNamespace, quietly = TRUE)]
if (length(to_install)) install.packages(to_install, repos = "https://cloud.r-project.org")
invisible(lapply(req, library, character.only = TRUE))

# Resumen de medias y desviaciones por sitio
tabla_resumen <- df %>%
  group_by(sitio) %>%
  summarise(n = n(),
            media = mean(estroncio),
            sd = sd(estroncio)) %>%
```

Curso: Estadística en la investigación científica
Presenta: César Enrique Martínez Sánchez
Docente: Dr. Marco Aurelio González Tagle

```
arrange(desc(media))
```

```
knitr::kable(tabla_resumen, digits = 2, caption = "Resumen por sitio (n, media, sd)") %>%  
  kableExtra::kable_styling(full_width = FALSE)
```

Table 1: Resumen por sitio (n, media, sd)

| sitio | n | media | sd |
|----------------|---|-------|------|
| Rock River | 6 | 58.30 | 3.04 |
| Angler's Cove | 6 | 44.08 | 3.08 |
| Appletree Lake | 6 | 41.10 | 3.67 |
| Beaver Lake | 6 | 40.23 | 2.53 |
| Grayson's Pond | 6 | 32.08 | 3.21 |

Hipótesis

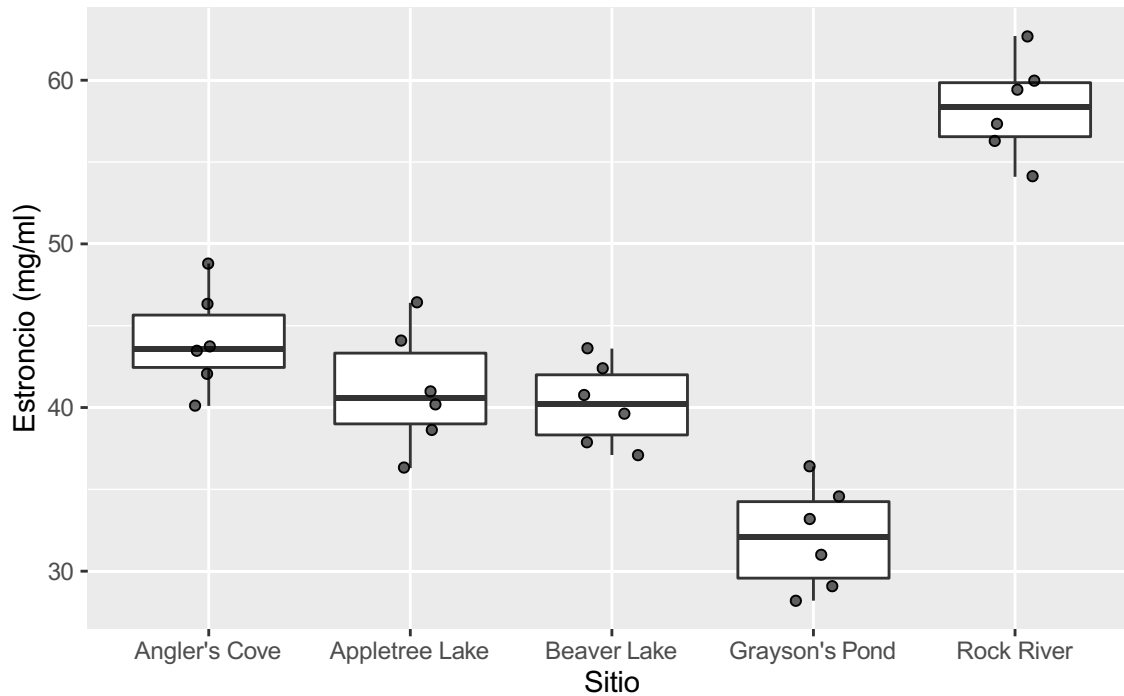
- **Ho (nula):** Las medias de concentración de estroncio son iguales en los cinco sitios:
 $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$.
- **H1 (alternativa):** Al menos una de las medias difiere.

Visualización rápida

```
ggplot(df, aes(x = sitio, y = estroncio)) +  
  geom_boxplot() +  
  geom_jitter(width = 0.15, alpha = 0.6) +  
  labs(x = "Sitio", y = "Estroncio (mg/ml)", title = "Distribución por sitio")
```

Curso: Estadística en la investigación científica
Presenta: César Enrique Martínez Sánchez
Docente: Dr. Marco Aurelio González Tagle

Distribución por sitio



ANOVA de una vía

```
modelo <- aov(estroncio ~ sitio, data = df)
anova_tabla <- broom::tidy(modelo) # tabla en formato ordenado
knitr::kable(anova_tabla, digits = 4, caption = "Tabla ANOVA (aov)") %>%
  kableExtra::kable_styling(full_width = FALSE)
```

Table 2: Tabla ANOVA (aov)

| term | df | sumsq | meansq | statistic | p.value |
|-----------|----|----------|----------|-----------|---------|
| sitio | 4 | 2193.442 | 548.3605 | 56.1546 | 0 |
| Residuals | 25 | 244.130 | 9.7652 | NA | NA |

Decisión: Si el **p-valor** del efecto de *sitio* es menor que 0.05, **se rechaza H_0** y hay evidencia de diferencias entre medias.

Prueba LSD (= 0.05)

La prueba LSD es equivalente a comparar medias con **t** sin corrección por multiplicidad.

Curso: Estadística en la investigación científica
Presenta: César Enrique Martínez Sánchez
Docente: Dr. Marco Aurelio González Tagle

```
emm <- emmeans(modelo, ~ sitio)

# 1) Comparaciones pareadas tipo LSD (sin ajuste)
pares_lsd <- pairs(emm, adjust = "none")
pares_lsd_tabla <- broom::tidy(pares_lsd)
knitr::kable(pares_lsd_tabla, digits = 4, caption = "Comparaciones pareadas (LSD, sin ajuste)",
  kableExtra::kable_styling(full_width = FALSE)
```

Table 3: Comparaciones pareadas (LSD, sin ajuste)

| term | contrast | null.value | estimate | std.error | df | statistic | p.value |
|-------|---------------------------------|------------|----------|-----------|----|-----------|---------|
| sitio | Angler's Cove - Appletree Lake | 0 | 2.9833 | 1.8042 | 25 | 1.6536 | 0.1107 |
| sitio | Angler's Cove - Beaver Lake | 0 | 3.8500 | 1.8042 | 25 | 2.1339 | 0.0428 |
| sitio | Angler's Cove - Grayson's Pond | 0 | 12.0000 | 1.8042 | 25 | 6.6512 | 0.0000 |
| sitio | Angler's Cove - Rock River | 0 | -14.2167 | 1.8042 | 25 | -7.8798 | 0.0000 |
| sitio | Appletree Lake - Beaver Lake | 0 | 0.8667 | 1.8042 | 25 | 0.4804 | 0.6351 |
| sitio | Appletree Lake - Grayson's Pond | 0 | 9.0167 | 1.8042 | 25 | 4.9977 | 0.0000 |
| sitio | Appletree Lake - Rock River | 0 | -17.2000 | 1.8042 | 25 | -9.5334 | 0.0000 |
| sitio | Beaver Lake - Grayson's Pond | 0 | 8.1500 | 1.8042 | 25 | 4.5173 | 0.0001 |
| sitio | Beaver Lake - Rock River | 0 | -18.0667 | 1.8042 | 25 | -10.0138 | 0.0000 |
| sitio | Grayson's Pond - Rock River | 0 | -26.2167 | 1.8042 | 25 | -14.5311 | 0.0000 |

```
# 2) Letras de agrupamiento (LSD ~ sin ajuste)
cld_lsd <- multcomp::cld(emm, adjust = "none", Letters = letters)
knitr::kable(cld_lsd[, c("sitio", ".group")], col.names = c("Sitio", "Grupo (LSD)"),
  caption = "Agrupamiento por letras (LSD, sin ajuste)" %>%
  kableExtra::kable_styling(full_width = FALSE)
```

Table 4: Agrupamiento por letras (LSD, sin ajuste)

| | Sitio | Grupo (LSD) |
|---|----------------|-------------|
| 4 | Grayson's Pond | a |
| 3 | Beaver Lake | b |
| 2 | Appletree Lake | bc |
| 1 | Angler's Cove | c |
| 5 | Rock River | d |

```
# 3) Cálculo manual del valor LSD (balanceado: n = 6 por grupo)
an <- summary(modelo)[[1]]
MSE <- an["Residuals", "Mean Sq"]
gl_error <- an["Residuals", "Df"]
n <- 6
tcrit <- qt(0.975, gl_error) # alfa/2 = 0.025 (prueba bilateral)
```

Curso: Estadística en la investigación científica
Presenta: César Enrique Martínez Sánchez
Docente: Dr. Marco Aurelio González Tagle

```
LSD_valor <- tcrit * sqrt(2 * MSE / n)

data.frame(MSE = MSE, gl_error = gl_error, t_0.975 = tcrit, LSD = LSD_valor) %>%
  knitr::kable(digits = 4, caption = "LSD calculado manualmente") %>%
  kableExtra::kable_styling(full_width = FALSE)
```

Table 5: LSD calculado manualmente

| MSE | gl_error | t_0.975 | LSD |
|--------|----------|---------|--------|
| 9.7652 | 25 | 2.0595 | 3.7158 |

Prueba de Tukey HSD ($\alpha = 0.05$)

Tukey controla el error familiar (FWER) y es más conservador que LSD.

```
# Base R
tukey_base <- TukeyHSD(modelo)
tukey_df <- as.data.frame(tukey_base$sitio)
tukey_df$contraste <- rownames(tukey_df)
names(tukey_df) <- sub("^p adj$", "p.adj", names(tukey_df))
tukey_df <- tukey_df[, c("contraste", "diff", "lwr", "upr", "p.adj")]
knitr::kable(tukey_df, digits = 4, caption = "Tukey HSD (base R)") %>%
  kableExtra::kable_styling(full_width = FALSE)
```

Table 6: Tukey HSD (base R)

| | contraste | diff | lwr | upr | p.adj |
|-------------------------------|-------------------------------|----------|----------|---------|--------|
| Appletree Lake-Angler's Cove | Appletree Lake-Angler's Cove | -2.9833 | -8.2820 | 2.3153 | 0.4791 |
| Beaver Lake-Angler's Cove | Beaver Lake-Angler's Cove | -3.8500 | -9.1486 | 1.4486 | 0.2376 |
| Grayson's Pond-Angler's Cove | Grayson's Pond-Angler's Cove | -12.0000 | -17.2986 | -6.7014 | 0.0000 |
| Rock River-Angler's Cove | Rock River-Angler's Cove | 14.2167 | 8.9180 | 19.5153 | 0.0000 |
| Beaver Lake-Appletree Lake | Beaver Lake-Appletree Lake | -0.8667 | -6.1653 | 4.4320 | 0.9885 |
| Grayson's Pond-Appletree Lake | Grayson's Pond-Appletree Lake | -9.0167 | -14.3153 | -3.7180 | 0.0003 |
| Rock River-Appletree Lake | Rock River-Appletree Lake | 17.2000 | 11.9014 | 22.4986 | 0.0000 |
| Grayson's Pond-Beaver Lake | Grayson's Pond-Beaver Lake | -8.1500 | -13.4486 | -2.8514 | 0.0011 |
| Rock River-Beaver Lake | Rock River-Beaver Lake | 18.0667 | 12.7680 | 23.3653 | 0.0000 |
| Rock River-Grayson's Pond | Rock River-Grayson's Pond | 26.2167 | 20.9180 | 31.5153 | 0.0000 |

```
# emmeans con ajuste Tukey
pares_tukey <- pairs(emm, adjust = "tukey")
pares_tukey_tabla <- broom::tidy(pares_tukey)
knitr::kable(pares_tukey_tabla, digits = 4, caption = "Comparaciones pareadas (ajuste Tukey)")
```

Curso: Estadística en la investigación científica

Presenta: César Enrique Martínez Sánchez

Docente: Dr. Marco Aurelio González Tagle

kableExtra::**kable_styling**(full_width = FALSE)

Curso: Estadística en la investigación científica
Presenta: César Enrique Martínez Sánchez
Docente: Dr. Marco Aurelio González Tagle

Table 7: Comparaciones pareadas (ajuste Tukey)

| term | contrast | null.value | estimate | std.error | df | statistic | adj.p.value |
|-------|---------------------------------|------------|----------|-----------|----|-----------|-------------|
| sitio | Angler's Cove - Appletree Lake | 0 | 2.9833 | 1.8042 | 25 | 1.6536 | 0.4791 |
| sitio | Angler's Cove - Beaver Lake | 0 | 3.8500 | 1.8042 | 25 | 2.1339 | 0.2376 |
| sitio | Angler's Cove - Grayson's Pond | 0 | 12.0000 | 1.8042 | 25 | 6.6512 | 0.0000 |
| sitio | Angler's Cove - Rock River | 0 | -14.2167 | 1.8042 | 25 | -7.8798 | 0.0000 |
| sitio | Appletree Lake - Beaver Lake | 0 | 0.8667 | 1.8042 | 25 | 0.4804 | 0.9885 |
| sitio | Appletree Lake - Grayson's Pond | 0 | 9.0167 | 1.8042 | 25 | 4.9977 | 0.0003 |
| sitio | Appletree Lake - Rock River | 0 | -17.2000 | 1.8042 | 25 | -9.5334 | 0.0000 |
| sitio | Beaver Lake - Grayson's Pond | 0 | 8.1500 | 1.8042 | 25 | 4.5173 | 0.0011 |
| sitio | Beaver Lake - Rock River | 0 | -18.0667 | 1.8042 | 25 | -10.0138 | 0.0000 |
| sitio | Grayson's Pond - Rock River | 0 | -26.2167 | 1.8042 | 25 | -14.5311 | 0.0000 |

```
# Letras de agrupamiento con ajuste Tukey
cld_tukey <- multcomp::cld(emm, adjust = "tukey", Letters = letters)
knitr::kable(cld_tukey[, c("sitio", ".group")], col.names = c("Sitio", "Grupo (Tukey)"),
  caption = "Agrupamiento por letras (Tukey)" %>%
  kableExtra::kable_styling(full_width = FALSE)
```

Table 8: Agrupamiento por letras (Tukey)

| | Sitio | Grupo (Tukey) |
|---|----------------|---------------|
| 4 | Grayson's Pond | a |
| 3 | Beaver Lake | b |
| 2 | Appletree Lake | b |
| 1 | Angler's Cove | b |
| 5 | Rock River | c |

```
# Cálculo manual de HSD (Tukey) para balanceado
k <- nlevels(df$sitio)
qcrit <- qtukey(0.95, k, gl_error)
HSD_valor <- qcrit * sqrt(MSE / n)

data.frame(MSE = MSE, k = k, gl_error = gl_error, q_0.95 = qcrit, HSD = HSD_valor) %>%
  knitr::kable(digits = 4, caption = "HSD (Tukey) calculado manualmente") %>%
  kableExtra::kable_styling(full_width = FALSE)
```

Table 9: HSD (Tukey) calculado manualmente

| MSE | k | gl_error | q_0.95 | HSD |
|-----|---|----------|--------|-----|
|-----|---|----------|--------|-----|

Curso: Estadística en la investigación científica
Presenta: César Enrique Martínez Sánchez
Docente: Dr. Marco Aurelio González Tagle

| | | | | |
|--------|---|----|--------|--------|
| 9.7652 | 5 | 25 | 4.1534 | 5.2986 |
|--------|---|----|--------|--------|

Curso: Estadística en la investigación científica
Presenta: César Enrique Martínez Sánchez
Docente: Dr. Marco Aurelio González Tagle

Interpretación

A partir de las tablas:

1. **¿Se rechaza H_0 ?** Revisar el p-valor del ANOVA. Si $p < 0.05$, **sí**.
2. **¿Qué pares difieren?** Comparar los resultados de *LSD* y de *Tukey*. Tukey suele marcar **menos** pares como significativos que *LSD*.
3. **¿Qué cuerpo presenta las concentraciones más altas y más bajas?**

```
# 1) Ganador y mínimo
```

```
ganador <- tabla_resumen$sitio[which.max(tabla_resumen$media)]  
minimo <- tabla_resumen$sitio[which.min(tabla_resumen$media)]
```

```
# 2) Pairs significativos segun Tukey (p.adj < 0.05)
```

```
sig_tukey <- subset(tukey_df, p.adj < 0.05)$contraste
```

```
# 3) Pairs significativos segun LSD (p < 0.05 sin ajuste)
```

```
sig_lsd <- subset(pares_lsd_tabla, p.value < 0.05)$contrast
```

```
cat(sprintf("- Sitio con **mayor** media: %s\n", ganador))
```

```
## - Sitio con **mayor** media: Rock River
```

```
cat(sprintf("- Sitio con **menor** media: %s\n\n", minimo))
```

```
## - Sitio con **menor** media: Grayson's Pond
```

```
cat("**Pares significativos según Tukey (p < 0.05):**\n")
```

```
## **Pares significativos según Tukey (p < 0.05):**
```

```
if (length(sig_tukey) == 0) cat("Ninguno\n") else print(sig_tukey)
```

```
## [1] "Grayson's Pond-Angler's Cove" "Rock River-Angler's Cove"
```

```
## [3] "Grayson's Pond-Appletree Lake" "Rock River-Appletree Lake"
```

```
## [5] "Grayson's Pond-Beaver Lake" "Rock River-Beaver Lake"
```

```
## [7] "Rock River-Grayson's Pond"
```

```
cat("\n**Pares significativos según LSD (p < 0.05, sin ajuste):**\n")
```

```
##
```

Curso: Estadística en la investigación científica

Presenta: César Enrique Martínez Sánchez

Docente: Dr. Marco Aurelio González Tagle

Pares significativos según LSD ($p < 0.05$, sin ajuste):

Curso: Estadística en la investigación científica
Presenta: César Enrique Martínez Sánchez
Docente: Dr. Marco Aurelio González Tagle

```
if (length(sig_lsd) == 0) cat("Ninguno\n") else print(sig_lsd)
```

```
## [1] "Angler's Cove - Beaver Lake"      "Angler's Cove - Grayson's Pond"  
## [3] "Angler's Cove - Rock River"       "Appletree Lake - Grayson's Pond"  
## [5] "Appletree Lake - Rock River"      "Beaver Lake - Grayson's Pond"  
## [7] "Beaver Lake - Rock River"         "Grayson's Pond - Rock River"
```

Comentario ambiental (breve)

- Diferencias elevadas y consistentes en un sitio pueden sugerir **influencia geológica local, descargas o procesos hidrológicos** distintos.
- Si un sitio como *Rock River* exhibe concentraciones significativamente mayores, podría requerir **monitoreo adicional** y evaluación de **posibles fuentes**.
- Sitios con medias similares (misma **letra** en Tukey) pueden considerarse **estadísticamente equivalentes** en concentración media.

Apéndice: reproducibilidad

Para compilar este documento a **PDF**:

```
# Instala TinyTeX (solo la primera vez)  
if (!tinytex::is_tinytex()) tinytex::install_tinytex()  
# Renderiza a PDF (desde R o RStudio)  
rmarkdown::render("tarea_anova.Rmd", output_format = "pdf_document")
```