

# Prueba de hipótesis

Santos G

## Tabla de contenidos

|   |  |   |
|---|--|---|
| 1 | Contexto del proyecto                                | 1 |
| 2 | Carga de librerías y datos                           | 1 |
| 3 | Verificación de supuestos                            | 2 |
| 4 | Ejecución del Kruskal-Wallis global                  | 3 |
| 5 | Comparaciones post-hoc (Dunn, corrección Bonferroni) | 3 |
| 6 | Conclusiones Kruskal-Wallis                          | 5 |
| 7 | Anova con transformación logarítmica                 | 6 |
| 8 | Comparaciones post-hoc (Tukey HSD)                   | 7 |
| 9 | Conclusiones Anova (log-transformado)                | 8 |

## 1 Contexto del proyecto

En este análisis trabajaremos con el dataset `palmerpenguins`, que contiene información morfométrica de tres especies (*Adelie*, *Chinstrap* y *Gentoo*). Previamente, los datos fueron limpiados y verificados, de modo que ahora podemos avanzar hacia las pruebas estadísticas que comparan medias entre grupos.

**Pregunta de investigación:** ¿Existen diferencias significativas en el largo del pico (*bill\_length\_mm*) entre las tres especies de pingüinos?

## 2 Carga de librerías y datos

```
1 library(tidyverse) # dplyr, ggplot2, tidyr: manipulación de datos y gráficos
2 library(janitor)   # clean_names(): estandarizar nombres de variables
3 library(palmerpenguins) # dataset de ejemplo (penguins)
4 library(car)       # leveneTest(): prueba de homogeneidad de varianzas
5 library(rstatix)   # funciones rápidas para ANOVA, Kruskal-Wallis, post-hoc
6 library(ggpubr)    # visualización de resultados estadísticos
7
```

```

8 df_raw <- penguins %>% as_tibble() # guardo raw para auditoría
9 df <- df_raw %>% clean_names()

```

### 3 Verificación de supuestos

```

1 modelo <- aov(bill_length_mm ~ species, data = df)
2
3 # Normalidad de residuos
4 Normalidad <- shapiro.test(residuals(modelo))
5
6 # Homogeneidad de varianzas (Levene)
7 Homogeneidad <- leveneTest(bill_length_mm ~ species, data = df)
8
9 # Tablas
10 knitr::kable(
11   broom::tidy(Normalidad),
12   caption = "Test de normalidad de residuos (Shapiro-Wilk)"
13 )

```

Tabla 1: Test de normalidad de residuos (Shapiro-Wilk)

| statistic | p.value   | method                      |
|-----------|-----------|-----------------------------|
| 0.9890313 | 0.0113053 | Shapiro-Wilk normality test |

```

1 knitr::kable(
2   broom::tidy(Homogeneidad),
3   caption = "Test de homogeneidad de varianzas (Levene)"
4 )

```

Tabla 2: Test de homogeneidad de varianzas (Levene)

| statistic | p.value   | df | df.residual |
|-----------|-----------|----|-------------|
| 2.242456  | 0.1077705 | 2  | 339         |

En la **Tabla 1** se presentan los resultados del test Shapiro–Wilk ( $W = 0.989$ ,  $p = 0.0113$ ) el valor  $p$  es menor a 0.05; por lo tanto, se rechaza la normalidad de los residuos. En cuanto a la homogeneidad de varianzas, en la **Tabla 2** se reflejan los resultados del test de Levene ( $F = 2.242$ ,  $p = 0.108$ ) arrojó un valor  $p$  mayor a 0.05, lo que indica que no se rechaza este supuesto, es decir las varianzas entre grupos pueden considerarse similares.

En consecuencia, dado que el supuesto de normalidad no se cumple, aunque sí el de homogeneidad de varianzas, la opción más segura y robusta es realizar Kruskal–Wallis (prueba no paramétrica).

## 4 Ejecución del Kruskal-Wallis global

```
1 # Kruskal-Wallis (global)
2 kruskal_res <- df %>% kruskal_test(bill_length_mm ~ species)
3
4 # Formatear p-value como p < 0.001 si es muy pequeño
5 kruskal_res <- kruskal_res %>%
6   mutate(p = ifelse(p < 0.001, "< 0.001", round(p, 3)))
7
8 knitr::kable(kruskal_res, caption = "Prueba Kruskal-Wallis global")
```

Tabla 3: Prueba Kruskal-Wallis global

| .y.            | n   | statistic | df | p       | method         |
|----------------|-----|-----------|----|---------|----------------|
| bill_length_mm | 342 | 244.1367  | 2  | < 0.001 | Kruskal-Wallis |

En la **Tabla 3** se presentan los resultados de la prueba de Kruskal–Wallis:  $\chi^2 = 244.14$ ,  $gl = 2$ ,  $p = < 0.001$ . El valor  $p$  es mucho menor a 0.001, lo que indica diferencias altamente significativas en el largo del pico entre las tres especies de pingüinos.

```
1 # Tamaño de efecto (epsilon-squared)
2 Efe <- df %>% kruskal_effsize(bill_length_mm ~ species)
3 knitr::kable(Efe, caption = "Tamaño de efecto (epsilon-squared)")
```

Tabla 4: Tamaño de efecto (epsilon-squared)

| .y.            | n   | effsize   | method  | magnitude |
|----------------|-----|-----------|---------|-----------|
| bill_length_mm | 342 | 0.7142676 | eta2[H] | large     |

En la **Tabla 4** se muestra el tamaño de efecto ( $\epsilon^2$ ): 0.7142676. Este resultado confirma que la magnitud de estas diferencias es muy grande: la variable “especie” explica aproximadamente el 71% de la variación en el largo del pico. Esto significa que el largo del pico no varía al azar, sino que está fuertemente determinado por la especie.

## 5 Comparaciones post-hoc (Dunn, corrección Bonferroni)

```
1 # Dunn test con corrección Bonferroni
2 dunn_res <- df %>%
3   dunn_test(bill_length_mm ~ species, p.adjust.method = "bonferroni") %>%
4   mutate(p = ifelse(p < 0.001, "< 0.001", round(p, 3)),
5          p.adj = ifelse(p.adj < 0.001, "< 0.001", round(p.adj, 3)))
6
7 knitr::kable(dunn_res, caption = "Comparaciones post-hoc (Dunn, Bonferroni)")
```

Tabla 5: Comparaciones post-hoc (Dunn, Bonferroni)

| .y.            | group1    | group2    | n1  | n2  | statistic | p       | p.adj   | p.adj.signif |
|----------------|-----------|-----------|-----|-----|-----------|---------|---------|--------------|
| bill_length_mm | Adelie    | Chinstrap | 151 | 68  | 12.753511 | < 0.001 | < 0.001 | ****         |
| bill_length_mm | Adelie    | Gentoo    | 151 | 123 | 13.135630 | < 0.001 | < 0.001 | ****         |
| bill_length_mm | Chinstrap | Gentoo    | 68  | 123 | -         | 0.077   | 0.231   | ns           |
|                |           |           |     |     | 1.767498  |         |         |              |

En la **Tabla 5** se presentan las comparaciones por pares (Dunn, Bonferroni):

- *Adelie* vs *Chinstrap*:  $p_{\text{adj}} = < 0.001 \rightarrow$  significativo (\*\*\*\*).
- *Adelie* vs *Gentoo*:  $p_{\text{adj}} = < 0.001 \rightarrow$  significativo (\*\*\*\*).
- *Chinstrap* vs *Gentoo*:  $p_{\text{adj}} = 0.231 \rightarrow$  no significativo (ns).

Las comparaciones muestran que el largo del pico en *Adelie* difiere significativamente tanto de *Chinstrap* como de *Gentoo*. Mientras que entre *Chinstrap* y *Gentoo* no se detectan diferencias estadísticamente significativas en el largo del pico. Esto sugiere que la dieta de *Adelie* es distinta a la de las otras especies. Mientras que entre *Gentoo* y *Chinstrap*, podrían compartir parcialmente los mismos recursos tróficos. En conjunto, la prueba confirma una clara segregación de *Adelie* respecto a las otras especies, lo que reduce la competencia y favorece la coexistencia de las tres especies en el ecosistema antártico.

```

1 # Añadir posiciones y columnas útiles para gráficas
2 dunn_res <- dunn_res %>% add_xy_position(x = "species")
3
4 # Calcula un y-position razonable para las etiquetas
5 y_max <- max(df$bill_length_mm, na.rm = TRUE)
6 dunn_res <- dunn_res %>% mutate(y.position = y_max + seq(0.5, by = 0.5,
7   length.out = nrow(dunn_res)))
8
9 p <- ggplot(df, aes(x = species, y = bill_length_mm, fill = species)) +
10   geom_boxplot(alpha = 0.7) +
11   geom_jitter(width = 0.15, alpha = 0.5, size = 1) +
12   theme_minimal() +
13   theme(legend.position = "none") +
14   labs(x = "Especie", y = "Largo del pico (mm)",
15     title = "Comparación del largo del pico entre especies")
16
17 # Añadir anotaciones de p ajustadas
18 p + ggpubr::stat_pvalue_manual(dunn_res, label = "p.adj.signif",
19   tip.length = 0.01)

```

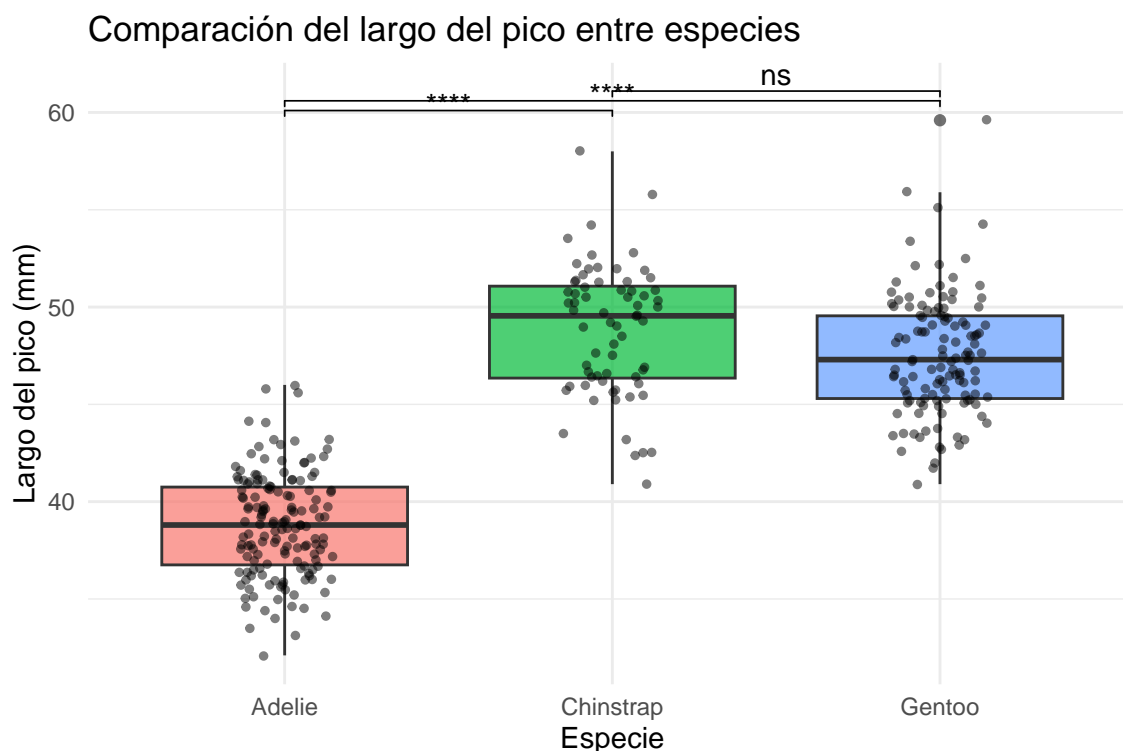


Figura 1: Boxplot de bill\_length\_mm por especie con comparaciones post-hoc (Dunn).

La **Figura 1** muestra la distribución del largo del pico en las tres especies de pingüinos (*Adelie*, *Chinstrap* y *Gentoo*). Se observa que los *Adelie* presentan valores menores y más concentrados (mediana cercana a 38–40 mm), mientras que los *Chinstrap* tienen los picos más largos (mediana en torno a 48–50 mm). Los *Gentoo* ocupan una posición intermedia, con valores cercanos a 46–48 mm y mayor variabilidad en comparación con *Adelie*.

Las comparaciones post-hoc de Dunn con corrección de Bonferroni evidencian diferencias altamente significativas en *Adelie* vs *Chinstrap* (\*\*\*\*) y *Adelie* vs *Gentoo* (\*\*\*\*). Por el contrario, no se encontraron diferencias significativas entre *Chinstrap* vs *Gentoo* (ns).

## 6 Conclusiones Kruskal-Wallis

- El análisis estadístico mediante Kruskal–Wallis confirmó diferencias significativas en la distribución del largo del pico entre especies de pingüinos ( $\chi^2 = 244.1$ ,  $p < 0.001$ ), con un tamaño de efecto muy grande ( $\epsilon^2 = 0.71$ ). Esto indica que la especie explica una proporción sustancial de la variabilidad observada en la morfología del pico.
- Las pruebas post-hoc mostraron que los *Adelie* difieren significativamente de *Chinstrap* y *Gentoo*, mientras que estas dos últimas no presentan diferencias estadísticamente significativas entre sí.
- Desde una perspectiva ecológica, los resultados sugieren que *Adelie* ocupa un nicho trófico diferenciado, asociado al consumo de krill y pequeños invertebrados, lo que reduce la competencia directa con las otras especies. En cambio, *Gentoo* y *Chinstrap*, al presentar longitudes de pico más similares, podrían solaparse en el aprovechamiento de presas de mayor tamaño (peces e invertebrados grandes).
- En conjunto, los hallazgos respaldan la hipótesis de una segregación morfológica y trófica que favorece la coexistencia de las tres especies en el ecosistema antártico.

## 7 Anova con transformación logarítmica

```
1 df <- df %>%
2   mutate(log_bill_length = log(bill_length_mm))
3 modelo_log <- aov(log_bill_length ~ species, data = df)
4
5 # Supuestos
6 shapiro_res <- shapiro.test(residuals(modelo_log))
7 levene_res <- leveneTest(log_bill_length ~ species, data = df)
8
9 # Tablas
10 knitr::kable(
11   broom::tidy(shapiro_res),
12   caption = "Test de normalidad de residuos (Shapiro-Wilk)"
13 )
```

Tabla 6: Test de normalidad de residuos (Shapiro-Wilk)

| statistic | p.value   | method                      |
|-----------|-----------|-----------------------------|
| 0.9945676 | 0.2675526 | Shapiro-Wilk normality test |

```
1 knitr::kable(
2   broom::tidy(levene_res),
3   caption = "Test de homogeneidad de varianzas (Levene)"
4 )
```

Tabla 7: Test de homogeneidad de varianzas (Levene)

| statistic | p.value  | df | df.residual |
|-----------|----------|----|-------------|
| 0.5610386 | 0.571145 | 2  | 339         |

En la **Tabla 6** se presentan los resultados del test Shapiro–Wilk ( $W = 0.995$ ,  $p = 0.268$ ) el valor  $p$  es mayor a 0.05; por lo tanto, no se rechaza la normalidad de los residuos. En cuanto a la homogeneidad de varianzas, en la **Tabla 7** se reflejan los resultados del test de Levene ( $F = 0.561$ ,  $p = 0.571$ ) arrojó un valor  $p$  mayor a 0.05, lo que indica que no se rechaza este supuesto, es decir las varianzas entre grupos pueden considerarse similares.

```
1 # Guardar en tidy para columnas limpias
2 anova_res <- broom::tidy(modelo_log) %>%
3   mutate(p.value = case_when(
4     p.value < 0.001 ~ "< 0.001",
5     TRUE ~ as.character(round(p.value, 3))
6   ))
7 knitr::kable(
8   anova_res,
9   caption = "Resultados de la prueba ANOVA (log-transformado)"
10 )
```

Tabla 8: Resultados de la prueba ANOVA (log-transformado)

| term      | df  | sumsq    | meansq    | statistic | p.value |
|-----------|-----|----------|-----------|-----------|---------|
| species   | 2   | 3.846076 | 1.9230382 | 427.5831  | < 0.001 |
| Residuals | 339 | 1.524639 | 0.0044975 | NA        | NA      |

En la **Tabla 8** se presentan los resultados de la prueba de ANOVA (log-transformado):  $F = 427.6$ ,  $gl = 2$ ,  $p < 0.001$ . El valor  $p$  es mucho menor a 0.001, lo que indica diferencias altamente significativas en el largo del pico entre las tres especies de pingüinos.

## 8 Comparaciones post-hoc (Tukey HSD)

```

1  tukey_res <- TukeyHSD(modelo_log)
2
3  tukey_table <- broom::tidy(tukey_res) %>%
4    mutate(
5      adj.p.value = case_when(
6        adj.p.value < 0.001 ~ "< 0.001",
7        TRUE ~ as.character(round(adj.p.value, 3))
8      ),
9      signif = case_when(
10       adj.p.value == "< 0.001" ~ "***",
11       as.numeric(adj.p.value) < 0.01 ~ "**",
12       as.numeric(adj.p.value) < 0.05 ~ "*",
13       TRUE ~ "ns"
14     )
15   )
16
17  knitr::kable(
18    tukey_table,
19    caption = "Comparaciones post-hoc (Tukey HSD)"
20  )

```

Tabla 9: Comparaciones post-hoc (Tukey HSD)

| term    | contrast             | null.value | estimate  | conf.low  | conf.high | adj.p.value | signif |
|---------|----------------------|------------|-----------|-----------|-----------|-------------|--------|
| species | Chinstrap-<br>Adelie | 0          | 0.2302359 | 0.2071801 | 0.2532916 | < 0.001     | ***    |
| species | Gentoo-Adelie        | 0          | 0.2029276 | 0.1837526 | 0.2221025 | < 0.001     | ***    |
| species | Gentoo-<br>Chinstrap | 0          | -         | -         | -         | 0.02        | *      |

En la **Tabla 9** se presentan los resultados de las comparaciones por pares (Tukey HSD):

- *Adelie vs Chinstrap*:  $p_{adj} = < 0.001 \rightarrow$  diferencia altamente significativa (\*\*\*).

- *Adelie* vs *Gentoo*:  $p_{\text{adj}} = < 0.001 \rightarrow$  diferencia altamente significativa (\*\*\*).
- *Chinstrap* vs *Gentoo*:  $p_{\text{adj}} = 0.02 \rightarrow$  diferencia significativa, aunque de magnitud menor (\*).

Los resultados muestran que el largo del pico en *Adelie* difiere de forma marcada respecto a *Chinstrap* y *Gentoo*. En contraste, *Chinstrap* y *Gentoo* presentan una diferencia estadísticamente significativa pero pequeña, lo que sugiere cierto solapamiento ecológico entre ambas especies.

En conjunto, la prueba confirma una clara segregación de *Adelie* en términos de morfología y, por extensión, de dieta (más especializada en krill e invertebrados pequeños). Por otro lado, *Chinstrap* y *Gentoo* parecen compartir parcialmente los mismos recursos tróficos, aunque mantienen diferencias detectables. Esto apoya la hipótesis de que la diversificación morfológica reduce la competencia y favorece la coexistencia de las tres especies en el ecosistema antártico.

## 9 Conclusiones Anova (log-transformado)

- El ANOVA con transformación logarítmica confirmó diferencias significativas en el largo del pico entre especies ( $F = 427.6$ ,  $p < 0.001$ ).
- El post-hoc Tukey corroboró que *Adelie* difiere de *Chinstrap* y *Gentoo*, y además detectó una diferencia débil pero significativa entre *Chinstrap* y *Gentoo* ( $p = 0.02$ ).
- Desde un punto de vista ecológico ambos enfoques concuerdan en que *Adelie* presenta un nicho trófico diferenciado (picos más cortos y especializados en krill e invertebrados pequeños). Por su parte, *Chinstrap* y *Gentoo* muestran solapamiento parcial, aunque el ANOVA sugiere cierta divergencia morfológica que podría traducirse en un aprovechamiento diferenciado de los recursos.

Esto complementa los hallazgos del Kruskal–Wallis: mientras este último no detectó diferencias entre *Chinstrap* y *Gentoo*, el ANOVA más sensible identificó una separación sutil.