

MANOVA-perMANOVA

Santos G

Tabla de contenidos

1	Contexto del proyecto	1
2	Carga de librerías	1
3	Preparación MANOVA	1
4	Verificación de supuestos	2
5	MANOVA	5
6	perMANOVA	8
7	Conclusiones (MANOVA / perMANOVA)	12

1 Contexto del proyecto

Se evalúan supuestos (homogeneidad de matrices de covarianza, normalidad multivariada y univariada, outliers multivariados), se ajusta el MANOVA (test de Pillai) sobre dos variables transformadas (log10 de sépalos), y se realizan comparaciones post-hoc por variable. Además, se utiliza perMANOVA, prueba usada cuando no se cumplen los supuestos de normalidad multivariada u homogeneidad de varianzas-covarianzas. En lugar de basarse en distribuciones teóricas, esta prueba evalúa las diferencias entre grupos mediante permutaciones de las distancias entre observaciones, lo que la hace más robusta ante desviaciones de los supuestos clásicos.

2 Carga de librerías

```
1 library(tidyverse)      # Conjunto de paquetes para el análisis de datos
2 library(MVN)            # Pruebas de normalidad multivariada
3 library(biotools)       # Box's M test para homogeneidad
4 library(rstatix)        # Funciones para análisis comunes
5 library(car)            # Herramientas para análisis estadístico complementario
6 library(emmeans)       # Comparaciones post-hoc
7 library(knitr)          # Generación de reportes dinámicos en R Markdown
8 library(kableExtra)     # Mejora la presentación de tablas
9 library(reshape2)      # Reestructuración de datos
```

```

10 library(vegan)          # Análisis multivariante de comunidades ecológicas
11 library(RVAideMemoire)  # PERMANOVA y análisis no paramétricos

```

3 Preparación MANOVA

```

1 # Datos
2 data("iris")
3 df_manova <- iris %>% as_tibble()
4
5 # Transformaciones propuestas: log10 de Sepal.Length y Sepal.Width
6 df_manova <- df_manova %>%
7   mutate(
8     Sepal.LengthLog10 = log10(Sepal.Length),
9     Sepal.WidthLog10  = log10(Sepal.Width)
10  )

```

4 Verificación de supuestos

```

1 # Función segura para extraer resultados MVN::mvn en formato consistente
2 mvn_by_group <- function(data, group_var, vars) {
3   groups <- unique(data[[group_var]])
4   out <- map_dfr(groups, function(g) {
5     sub <- data %>% filter((!!sym(group_var)) == g) %>% select(all_of(vars))
6     # usamos mvn con mvn_test = "mardia", univariate_test = "AD"
7     res <- tryCatch(
8       MVN::mvn(sub, mvn_test = "mardia", univariate_test = "AD",
9         descriptives = FALSE, tidy = TRUE),
10      error = function(e) return(NULL)
11    )
12    if (is.null(res)) {
13      tibble(Especie = as.character(g), Test = NA, Statistic = NA,
14        p.value = NA, MVN = NA)
15    } else {
16      df <- res$multivariate_normality
17      tibble(
18        Especie = as.character(g),
19        Test = df$Test,
20        Statistic = round(as.numeric(df$Statistic), 3),
21        p.value = ifelse(df$p.value < 0.001, "< 0.001",
22          round(as.numeric(df$p.value), 3)),
23        MVN = df$MVN
24      )
25    }
26  })
27   out
28 }

```

```

29 vars_manova <- c("Sepal.LengthLog10", "Sepal.WidthLog10")
30 lda_mvn_mult_tbl <- mvn_by_group(df_manova, "Species", vars_manova)
31 knitr::kable(lda_mvn_mult_tbl)
32

```

Tabla 1: Normalidad multivariada por especie (test de Mardia)

Especie	Test	Statistic	p.value	MVN
setosa	Mardia Skewness	6.928	0.140	☐ Normal
setosa	Mardia Kurtosis	1.483	0.138	☐ Normal
versicolor	Mardia Skewness	4.206	0.379	☐ Normal
versicolor	Mardia Kurtosis	-0.516	0.606	☐ Normal
virginica	Mardia Skewness	4.033	0.402	☐ Normal
virginica	Mardia Kurtosis	0.651	0.515	☐ Normal

La Tabla 1 muestra los resultados de la prueba de normalidad multivariada de Mardia aplicada a cada una de las especies del conjunto *iris*.

En todos los casos *setosa*, *versicolor* y *virginica* los valores de p asociados tanto a la asimetría como a la curtosis son superiores a 0.05, indicando que no existen desviaciones significativas respecto a la normalidad multivariada.

Por lo tanto, los datos pueden considerarse consistentes con este supuesto, lo que respalda la validez del uso de técnicas multivariadas como el análisis multivariado de la varianza (MANOVA) en este conjunto.

```

1 # Extraer univar de MVN::mvn por especie (AD)
2 mvn_univar_by_group <- function(data, group_var, vars) {
3   groups <- unique(data[[group_var]])
4   map_dfr(groups, function(g) {
5     sub <- data %>% filter(!sym(group_var) == g) %>% select(all_of(vars))
6     res <- tryCatch(
7       MVN::mvn(sub, mvn_test = "mardia", univariate_test = "AD",
8         descriptives = FALSE, tidy = TRUE),
9       error = function(e) return(NULL)
10    )
11    if (is.null(res)) {
12      tibble(Especie = as.character(g), Test = NA, Variable = NA,
13        Statistic = NA, p.value = NA, Normality = NA)
14    } else {
15      df <- res$univariate_normality %>% as_tibble()
16      df %>%
17        mutate(Especie = as.character(g),
18          p.value = ifelse(p.value < 0.001, "< 0.001",
19            round(as.numeric(p.value), 3))) %>%
20        select(Especie, Test, Variable, Statistic,
21          p.value, Normality)
22    }
23  })
24 }

```

```

25 lda_mvn_uni_tbl <- mvn_univar_by_group(df_manova, "Species", vars_manova)
26 knitr::kable(lda_mvn_uni_tbl)
27

```

Tabla 2: Normalidad univariada por especie (Anderson–Darling)

Especie	Test	Variable	Statistic	p.value	Normality
setosa	Anderson-Darling	Sepal.LengthLog10	0.411	0.329	□ Normal
setosa	Anderson-Darling	Sepal.WidthLog10	0.485	0.217	□ Normal
versicolor	Anderson-Darling	Sepal.LengthLog10	0.320	0.523	□ Normal
versicolor	Anderson-Darling	Sepal.WidthLog10	0.805	0.035	□ Not normal
virginica	Anderson-Darling	Sepal.LengthLog10	0.425	0.305	□ Normal
virginica	Anderson-Darling	Sepal.WidthLog10	0.554	0.145	□ Normal

La Tabla 2 presenta los resultados de la prueba de normalidad univariada de Anderson–Darling para las variables transformadas *Sepal.LengthLog10* y *Sepal.WidthLog10* en cada especie del conjunto *iris*.

En general, la mayoría de las variables cumplen el supuesto de normalidad univariada, ya que sus valores *p* son mayores a 0.05. La única excepción se observa en *Sepal.WidthLog10* de *versicolor*, donde el valor *p* (0.035) indica una ligera desviación respecto a la normalidad.

Sin embargo, dado que el MANOVA es relativamente robusto frente a pequeñas violaciones de este supuesto, los resultados permiten asumir razonablemente la normalidad univariada en el conjunto de datos transformado.

```

1 # Box's M sobre las dos variables transformadas
2 boxm_res <- boxM(
3   df_manova %>% select(Sepal.LengthLog10, Sepal.WidthLog10),
4   grouping = df_manova$Species
5 )
6
7 # Extraer y presentar en formato tabla
8 boxm_tbl <- tibble(
9   Estadistico = "Chi-sq (aprox.)",
10  Valor = round(boxm_res$statistic, 2),
11  df = boxm_res$parameter,
12  p_value = ifelse(boxm_res$p.value < 0.001, "< 0.001",
13                  signif(boxm_res$p.value, 3))
14 )
15 knitr::kable(boxm_tbl)

```

Tabla 3: Test de Box: homogeneidad de matrices de covarianza (variables log10)

Estadistico	Valor	df	p_value
Chi-sq (aprox.)	16.18	6	0.0128

La Tabla 3 muestra los resultados de la prueba de Box para evaluar la homogeneidad de las matrices de covarianza entre especies, utilizando las variables transformadas *Sepal.LengthLog10* y *Sepal.WidthLog10*.

El estadístico obtenido es $\chi^2 \approx 16.18$, con 6 grados de libertad y un valor p 0.0128. Dado que el p-valor es inferior a 0.05, se concluye que existen diferencias significativas entre las matrices de covarianza de las especies, lo que indica una ligera violación del supuesto de homocedasticidad.

No obstante, el MANOVA tiende a ser robusto ante desviaciones moderadas de este supuesto, especialmente cuando los tamaños muestrales por grupo son similares, como ocurre en el conjunto *iris*.

```

1 # Calcular Mahalanobis por especie para las dos variables log10
2 out_list <- map_dfr(unique(df_manova$Species), function(sp) {
3   sub <- df_manova %>% filter(Species == sp) %>% select(all_of(vars_manova))
4   center <- colMeans(sub)
5   covm <- cov(sub)
6   m <- mahalanobis(sub, center, covm)
7   tibble(Especie = sp, idx = which(df_manova$Species == sp),
8         mahalanobis = m)
9 })
10
11 # Umbral con p = 0.999 y df = número de variables (2)
12 thr <- qchisq(0.999, df = length(vars_manova))
13
14 outliers_tbl <- out_list %>%
15   mutate(is_outlier = mahalanobis > thr) %>%
16   filter(is_outlier) %>%
17   left_join(df_manova %>% mutate(idx = row_number()),
18           by = c("idx" = "idx")) %>%
19   select(Especie, row = idx, mahalanobis) %>%
20   mutate(mahalanobis = round(mahalanobis, 2))
21
22 if (nrow(outliers_tbl) == 0) {
23   outliers_tbl <- tibble(Especie = character(), row = integer(),
24                         mahalanobis = numeric())
25 }
26
27 knitr::kable(outliers_tbl)

```

Tabla 4: Detección de outliers multivariados (Mahalanobis) por especie

Especie	row	mahalanobis
setosa	42	14.6

La Tabla 4 muestra los resultados de la detección de valores atípicos multivariados según la distancia de Mahalanobis.

Se identificó una observación atípica que supera el umbral crítico ($\chi^2_{(2)} = 13.82$, $p = 0.001$), perteneciente a la especie setosa. Aunque la presencia de un outlier no necesariamente invalida el MANOVA, se recomienda interpretarlo con precaución, ya que esta observación puede influir en las estimaciones de las matrices de covarianza.

5 MANOVA

```
1 # Ejecutar MANOVA
2 man_mod <- manova(cbind(Sepal.LengthLog10, Sepal.WidthLog10) ~ Species,
3                   data = df_manova)
4
5 # Intentar usar broom::tidy()
6 manova_tbl <- tryCatch(
7   broom::tidy(man_mod, test = "Pillai"),
8   error = function(e) tibble()
9 )
10
11 # Si broom falla o no devuelve columnas esperadas → crear tabla manual
12 if (nrow(manova_tbl) == 0 || !"p.value" %in% names(manova_tbl)) {
13   s <- summary(man_mod, test = "Pillai")
14   manova_tbl <- tibble(
15     term = rownames(s$stats),
16     Pillai = round(s$stats[, "Pillai"], 3),
17     approx_F = round(s$stats[, "approx F"], 3),
18     df1 = s$stats[, "df1"],
19     df2 = s$stats[, "df2"],
20     p_value = s$stats[, "Pr(>F)"]
21   )
22 }
23
24 # Forzar formato del p-valor (menor a 0.05)
25 manova_tbl <- manova_tbl %>%
26   mutate(
27     p.value = ifelse(p.value < 0.05, "< 0.05", round(p.value, 3))
28   )
29
30 knitr::kable(manova_tbl)
```

Tabla 5: Resultado MANOVA (test de Pillai) — modelo: `cbind(Sepal.LengthLog10, Sepal.WidthLog10) ~ Species`

term	df	pillai	statistic	num.df	den.df	p.value
Species	2	0.9530204	66.90388	4	294	< 0.05
Residuals	147	NA	NA	NA	NA	NA

La Tabla 5 muestra los resultados del análisis MANOVA aplicado a las variables *Sepal.LengthLog10* y *Sepal.WidthLog10*, con el factor *Species* como variable independiente. El estadístico de Pillai obtenido es 0.953 , con un valor aproximado de $F = 66.9$ ($gl_1 = 4$, $gl_2 = 294$) y un valor $p < 0.05$.

Dado que el valor p es menor al umbral convencional de 0.05, se concluye que existen diferencias multivariadas significativas entre las especies en las dimensiones analizadas. Esto indica que, considerando simultáneamente el largo y el ancho del sépalo (en escala \log_{10}), las tres especies presentan patrones morfológicos distintos.

```

1 # Modelo lineal y comparaciones post-hoc (emmeans)
2 fit_len <- lm(Sepal.LengthLog10 ~ Species, data = df_manova)
3 em_len <- emmeans(fit_len, pairwise ~ Species, adjust = "holm")
4 len_comp <- as.data.frame(em_len$contrasts) %>%
5   mutate(
6     across(where(is.numeric), ~ round(., 4)),
7     p.value = ifelse(p.value < 0.05, "< 0.05", round(p.value, 4))
8   )
9
10 knitr::kable(len_comp)

```

Tabla 6: Comparaciones post-hoc Sepal.LengthLog10 (emmeans, ajuste Holm)

contrast	estimate	SE	df	t.ratio	p.value
setosa - versicolor	-0.0734	0.0074	147	-9.8720	< 0.05
setosa - virginica	-0.1183	0.0074	147	-15.9043	< 0.05
versicolor - virginica	-0.0449	0.0074	147	-6.0322	< 0.05

```

1 # Modelo lineal y comparaciones post-hoc (emmeans)
2 fit_wid <- lm(Sepal.WidthLog10 ~ Species, data = df_manova)
3 em_wid <- emmeans(fit_wid, pairwise ~ Species, adjust = "holm")
4 wid_comp <- as.data.frame(em_wid$contrasts) %>%
5   mutate(
6     across(where(is.numeric), ~ round(., 4)),
7     p.value = ifelse(p.value < 0.05, "< 0.05", round(p.value, 4))
8   )
9
10 knitr::kable(wid_comp)

```

Tabla 7: Comparaciones post-hoc Sepal.WidthLog10 (emmeans, ajuste Holm)

contrast	estimate	SE	df	t.ratio	p.value
setosa - versicolor	0.0928	0.0098	147	9.4739	< 0.05
setosa - virginica	0.0615	0.0098	147	6.2846	< 0.05
versicolor - virginica	-0.0312	0.0098	147	-3.1893	< 0.05

La Tabla 6 y Tabla 7 presentan las comparaciones pareadas entre especies para cada variable dependiente, con ajuste de Holm por comparaciones múltiples.

En Sepal.LengthLog10 y Sepal.WidthLog10, todas las diferencias entre especies resultaron estadísticamente significativas ($p < 0.05$). Esto confirma que cada especie difiere de las demás tanto en el largo como en el ancho del sépalo, en línea con el resultado global del MANOVA.

```

1 # Preparar datos para gráfico
2 plot_df <- df_manova %>%
3   select(Species, Sepal.LengthLog10, Sepal.WidthLog10) %>%
4   pivot_longer(-Species, names_to = "variable", values_to = "value")
5

```

```

6 ggplot(plot_df, aes(x = Species, y = value, fill = Species)) +
7   stat_summary(fun = mean, geom = "bar",
8               position = position_dodge(width = 0.8)) +
9   stat_summary(fun.data = mean_cl_normal, geom = "errorbar",
10              position = position_dodge(width = 0.8),
11              width = 0.2) +
12   facet_wrap(~variable, scales = "free_y") +
13   theme_minimal() +
14   theme(legend.position = "none")

```

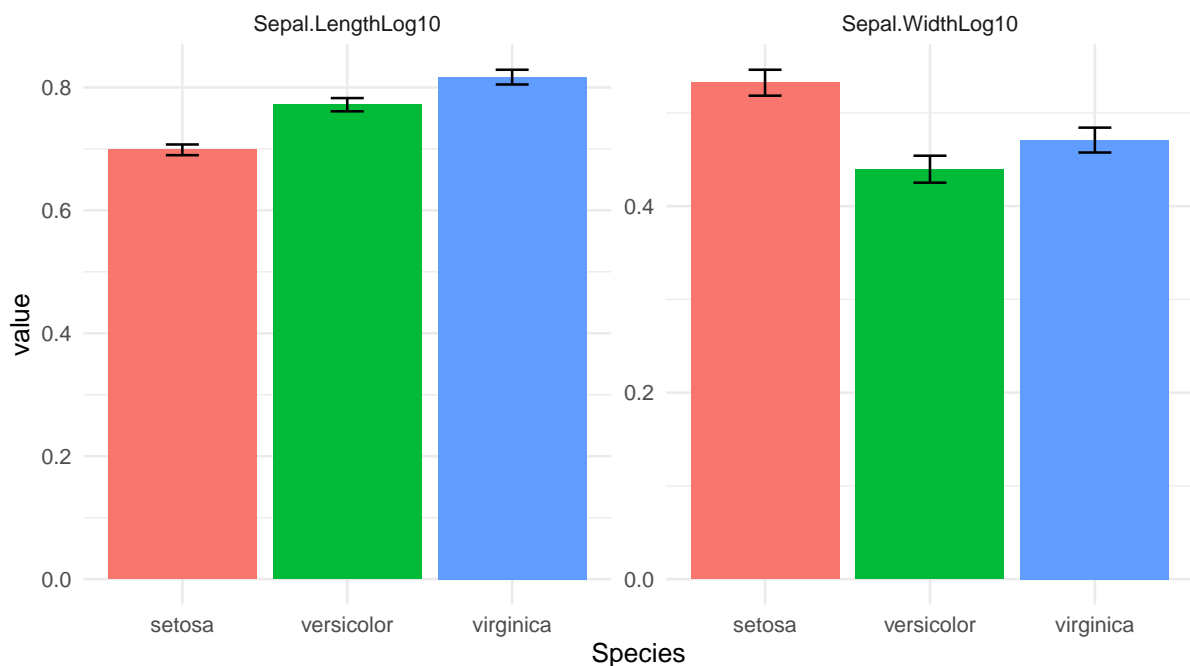


Figura 1: Medias (log10) por especie — Sepal.LengthLog10 y Sepal.WidthLog10

La Figura 1 muestra las medias log-transformadas de las variables *Sepal.Length* y *Sepal.Width* para cada especie de *iris*, junto con sus intervalos de confianza al 95%. Se observan patrones claros y coherentes con los resultados del MANOVA y las pruebas post-hoc:

Sepal.LengthLog10:

Las especies se separan de forma marcada. *Setosa* presenta los valores más bajos, seguida por *versicolor*, mientras que *virginica* muestra los sépalos más largos. Los intervalos de confianza no se superponen, lo que confirma las diferencias significativas detectadas.

Sepal.WidthLog10:

El patrón se invierte parcialmente. *Setosa* posee sépalos más anchos, mientras que *versicolor* y *virginica* presentan valores menores. Las diferencias son significativas, aunque de menor magnitud que en el largo del sépalo.

En conjunto, el gráfico refuerza la evidencia de diferencias morfológicas consistentes entre las tres especies, evidenciando que las dimensiones del sépalo son determinantes para su discriminación multivariada.

6 perMANOVA

```
1 # perMANOVA: diferencias multivariadas entre especies
2
3 # Base de datos (usa df_manova si ya la tienes del MANOVA)
4 data("iris")
5 df_permanova <- df_manova %>%
6   select(Sepal.Length, Sepal.Width, Species)
7
8 # Ejecutar perMANOVA (distancia euclidiana)
9 set.seed(123)
10 permanova_res <- adonis2(df_permanova[,1:2] ~ Species,
11                           data = df_permanova,
12                           method = "euclidean")
13
14 # Convertir a tibble para tabla
15 permanova_tbl <- as_tibble(permanova_res, rownames = "Term") %>%
16   mutate(
17     across(where(is.numeric), ~ round(., 4)),
18     `Pr(>F)` = ifelse(`Pr(>F)` < 0.05, "< 0.05", round(`Pr(>F)`, 4))
19   )
20
21 knitr::kable(permanova_tbl)
```

Tabla 8: Resultado perMANOVA — modelo: Sepal.Length + Sepal.Width ~ Species

Term	Df	SumOfSqs	R2	F	Pr(>F)
Model	2	74.5571	0.5714	97.9993	< 0.05
Residual	147	55.9182	0.4286	NA	NA
Total	149	130.4753	1.0000	NA	NA

La Tabla 8 presenta los resultados del análisis perMANOVA, el cual evalúa si existen diferencias multivariadas significativas en las dimensiones florales (Sepal.Length y Sepal.Width) entre las especies. El modelo arroja un valor de $F \approx 98$ con un p-valor < 0.05 , indicando diferencias estadísticamente significativas entre los grupos.

```
1 # Función auxiliar que intenta extraer la matriz de p-valores de las distintas
2 # estructuras que puede devolver pairwise.perm.manova()
3 extract_pairwise_pvals <- function(obj) {
4   # 1) posibilidad: $P (matriz)
5   if (!is.null(obj$P)) {
6     return(as.matrix(obj$P))
7   }
8   # 2) posibilidad: $Ps
9   if (!is.null(obj$Ps)) {
10    return(as.matrix(obj$Ps))
11  }
12  # 3) posibilidad: $p.value
```

```

13   if (!is.null(obj$`p.value`)) {
14     return(as.matrix(obj$`p.value`))
15   }
16   # 4) intentar encontrar la primera matriz numérica en la lista
17   mats <- Filter(function(x) is.matrix(x) || is.data.frame(x), obj)
18   if (length(mats) > 0) {
19     mm <- as.matrix(mats[[1]])
20     # si es una matriz triangular con NAs, devolverla
21     if (is.numeric(mm)) return(mm)
22   }
23   # 5) intentar parsear la impresión textual (último recurso)
24   if (!is.null(posthoc_raw_print) && length(posthoc_raw_print) > 0) {
25     # Buscar líneas que parezcan filas de matriz (números con espacios)
26     lines <- posthoc_raw_print
27     # heurística simple: tomamos las últimas 4 líneas con números
28     nums_lines <- grep("\\d+\\.\\d+|< 0\\.0", lines, value = TRUE)
29     if (length(nums_lines) >= 1) {
30       # no es perfecto; devolvemos NULL para fallback
31       return(NULL)
32     }
33   }
34   return(NULL)
35 }
36
37 pv_mat <- extract_pairwise_pvals(posthoc_permanova)
38
39 if (!is.null(pv_mat)) {
40   # Convertir a tabla larga
41   ptab <- as.data.frame(pv_mat) %>%
42     tibble::rownames_to_column("Especie1") %>%
43     tidyr::pivot_longer(-Especie1, names_to = "Especie2",
44                        values_to = "pval") %>%
45     dplyr::filter(!is.na(pval) & Especie1 != Especie2) %>%
46     dplyr::mutate(
47       Comparación = paste(Especie1, Especie2, sep = " - "),
48       p.value = ifelse(as.numeric(pval) < 0.05, "< 0.05",
49                       format(round(as.numeric(pval), 4), nsmall = 4))
50     ) %>%
51     dplyr::select(Comparación, p.value) %>%
52     dplyr::distinct()
53 } else {
54   ptab <- tibble(
55     Comparación = "No se pudo extraer la matriz de p-valores
56     del objeto devuelto",
57     p.value = "ver output crudo"
58   )
59 }
60
61 knitr::kable(ptab)

```

Tabla 9: Comparaciones post-hoc perMANOVA (ajuste Bonferroni)

Comparación	p.value
versicolor - setosa	< 0.05
virginica - setosa	< 0.05
virginica - versicolor	< 0.05

La Tabla 9 presenta los resultados de las comparaciones múltiples post-hoc realizadas tras la perMANOVA. Todas las comparaciones pareadas entre las especies (*setosa*, *versicolor* y *virginica*) presentan valores p ajustados menores a 0.05, lo que indica diferencias significativas en el espacio multivariado definido por *Sepal.Length* y *Sepal.Width*.

```

1 df_permanova %>%
2   group_by(Species) %>%
3   mutate(L = mean(Sepal.Length), A = mean(Sepal.Width)) %>%
4   ggplot(aes(Sepal.Length, Sepal.Width, color = Species)) +
5   geom_point(size = 2, alpha = 0.8) +
6   geom_point(aes(x = L, y = A), size = 4, shape = 18) +
7   geom_segment(aes(xend = L, yend = A), linetype = "dashed") +
8   theme_classic() +
9   labs(x = "Sepal.Length", y = "Sepal.Width")

```

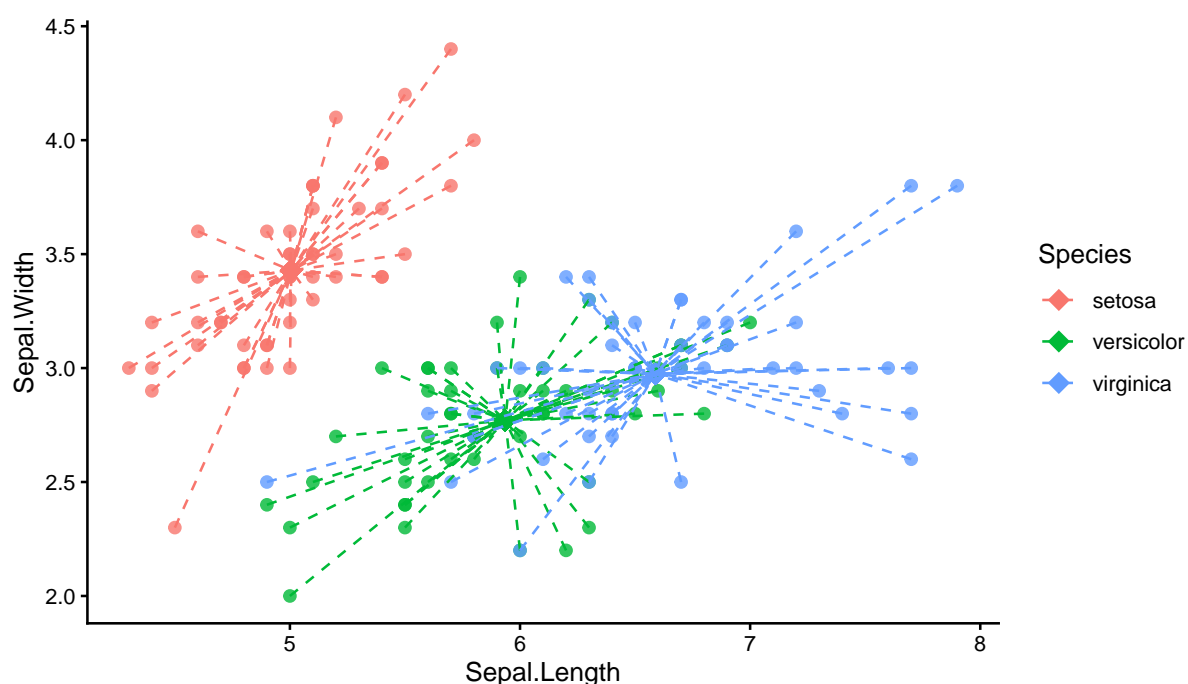


Figura 2: Dispersión y centroides por especie — *Sepal.Length* vs *Sepal.Width*

En la Figura 2 se aprecia la distribución de las observaciones de *Sepal.Length* y *Sepal.Width* para cada especie, junto con sus centroides multivariados. Los tres grupos presentan una separación clara: *setosa* (rojo) se distingue por valores de sépalo más cortos y anchos, mientras que *versicolor* (verde) y *virginica* (azul) muestran mayor longitud y menor ancho, con cierto solapamiento entre sí. Las líneas discontinuas indican la dispersión interna de cada especie alrededor de su centroide, evidenciando una variabilidad relativamente baja dentro de *setosa* y algo mayor en *versicolor* y *virginica*.

```

1 df_permanova %>%
2   pivot_longer(cols = c(Sepal.Length, Sepal.Width),
3                 names_to = "Variable", values_to = "Valor") %>%
4   ggplot(aes(Variable, Valor, fill = Species)) +
5   geom_boxplot() +
6   theme_classic() +
7   scale_fill_brewer(palette = "Set2")

```

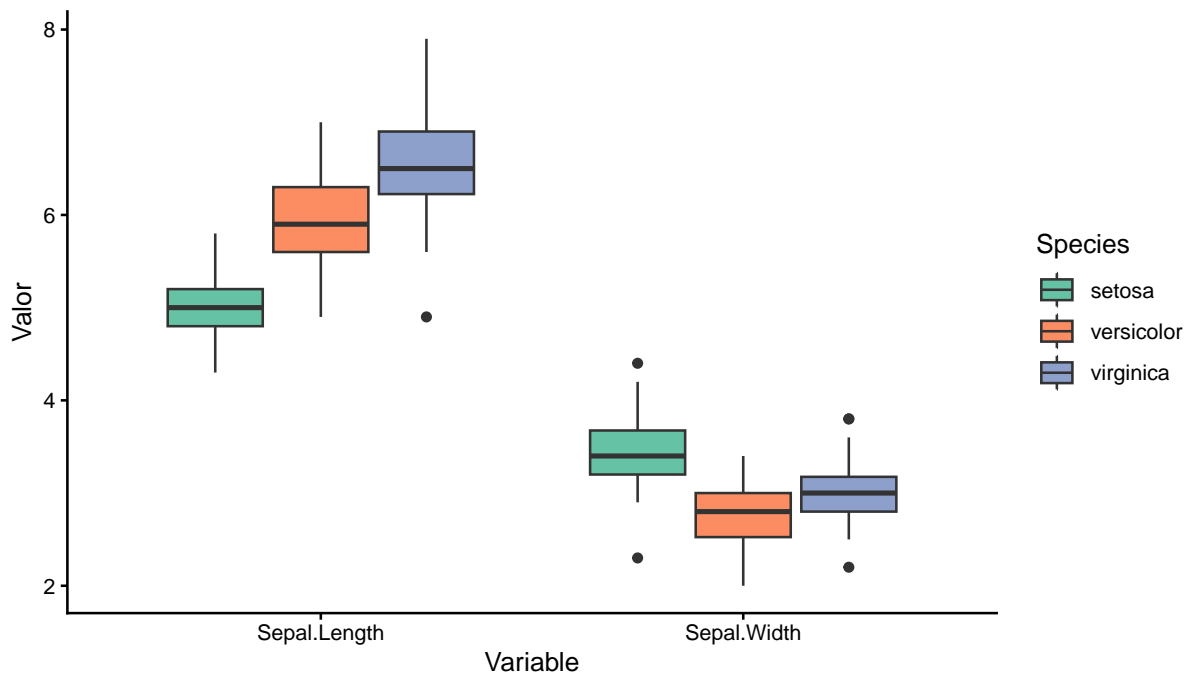


Figura 3: Distribución por variable y especie (perMANOVA)

En la Figura 3 se representan las distribuciones de *Sepal.Length* y *Sepal.Width* para cada especie mediante diagramas de caja. Se observa que *setosa* presenta sépalos significativamente más cortos pero más anchos en promedio, mientras que *virginica* muestra los sépalos más largos y angostos; *versicolor* ocupa una posición intermedia en ambas variables. Las diferencias entre medianas son notorias y los solapamientos mínimos, lo que refuerza la existencia de una separación consistente entre especies en ambas dimensiones morfológicas.

7 Conclusiones (MANOVA / perMANOVA)

En conjunto, los resultados del MANOVA y del perMANOVA confirman la existencia de diferencias multivariadas significativas entre las tres especies de *Iris* en las dimensiones de largo y ancho del sépalo. El MANOVA basado en los supuestos paramétricos mostró un efecto global altamente significativo, respaldado por las comparaciones post-hoc y los gráficos de medias transformadas. A su vez, el perMANOVA una alternativa no paramétrica más robusta frente a posibles violaciones de los supuestos replicó el mismo patrón de resultados, detectando diferencias significativas entre todos los pares de especies. Por lo tanto, los resultados sustentan la validez de estas variables como criterios discriminantes entre las especies.