

# CLASIFICACIÓN DEL DATASET IRIS USANDO REGRESIÓN LINEAL

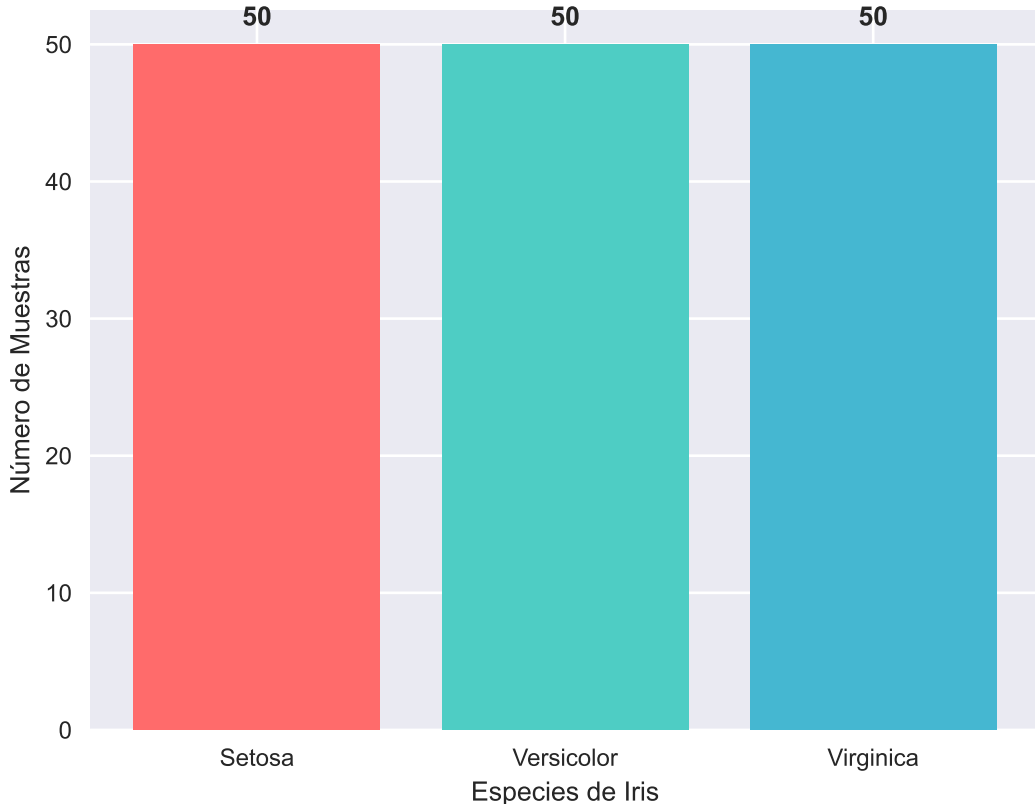
*Análisis Completo con Visualizaciones*

- DATASET: 150 muestras, 4 características, 3 clases
- MÉTODO: Regresión Lineal con estrategia One-vs-Rest
- OBJETIVO: Clasificar especies de iris (setosa, versicolor, virginica)
- HERRAMIENTAS: Python, scikit-learn, matplotlib, seaborn

Autor: Juan Esteban Moya Riaño, Maryuri Espinosa  
Fecha: 2025

# RESUMEN DEL DATASET IRIS

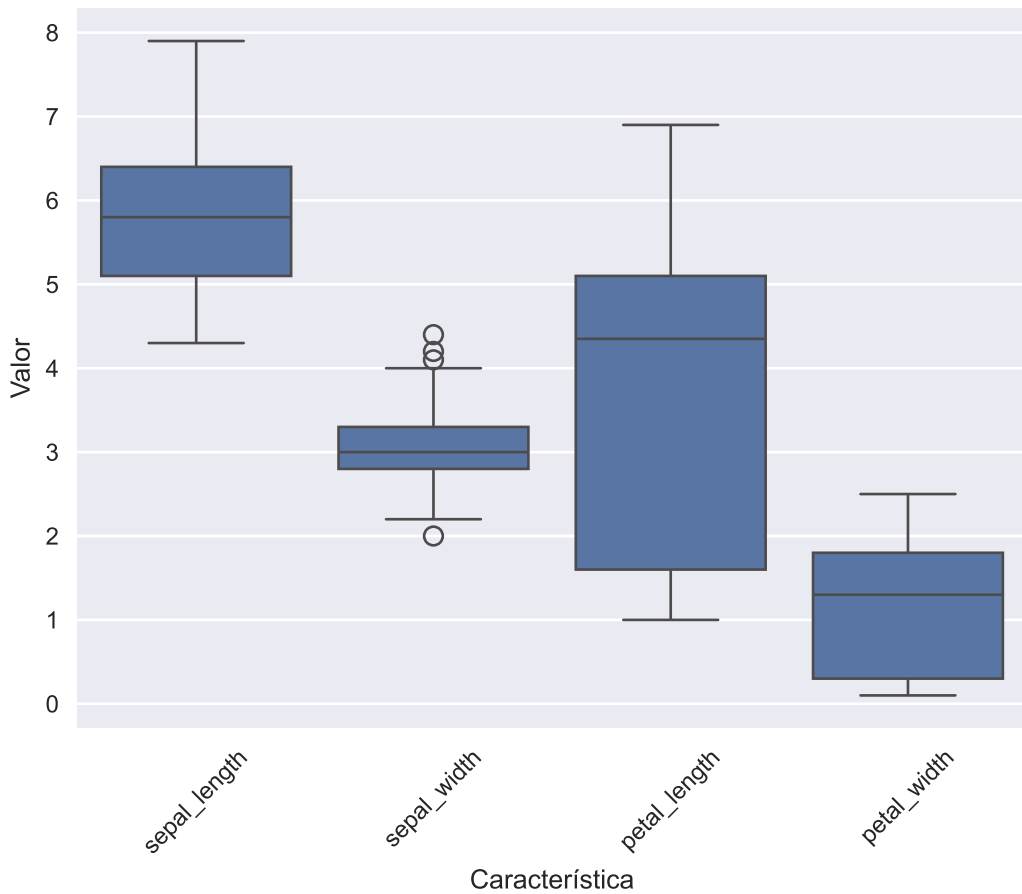
## Distribución de Clases



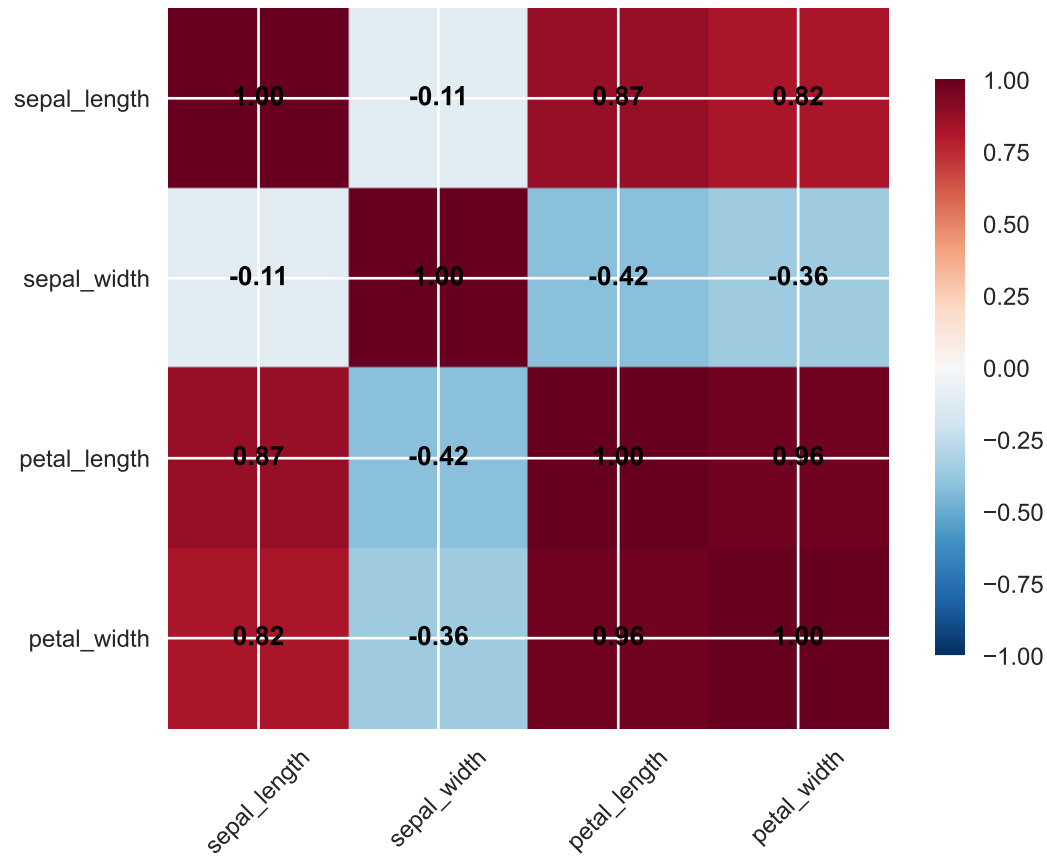
## Estadísticas Descriptivas

Característica	Media	Desv. Est.	Mín	Máx
sepal_length	5.84	0.83	4.30	7.90
sepal_width	3.05	0.43	2.00	4.40
petal_length	3.76	1.76	1.00	6.90
petal_width	1.20	0.76	0.10	2.50

## Distribución de Características (Boxplot)

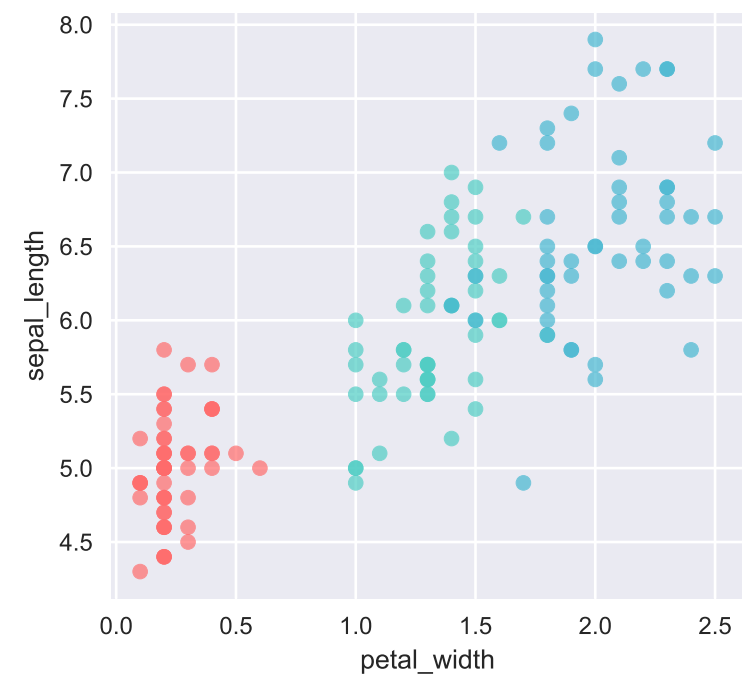
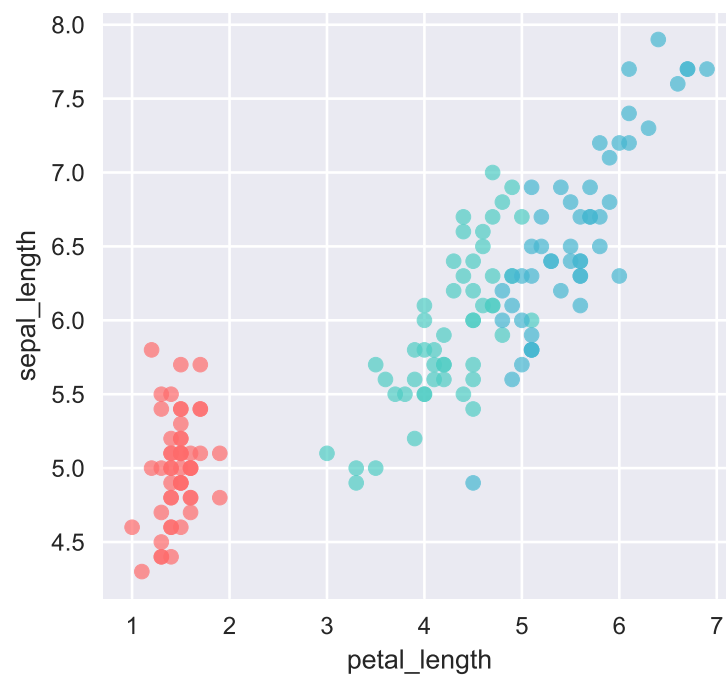
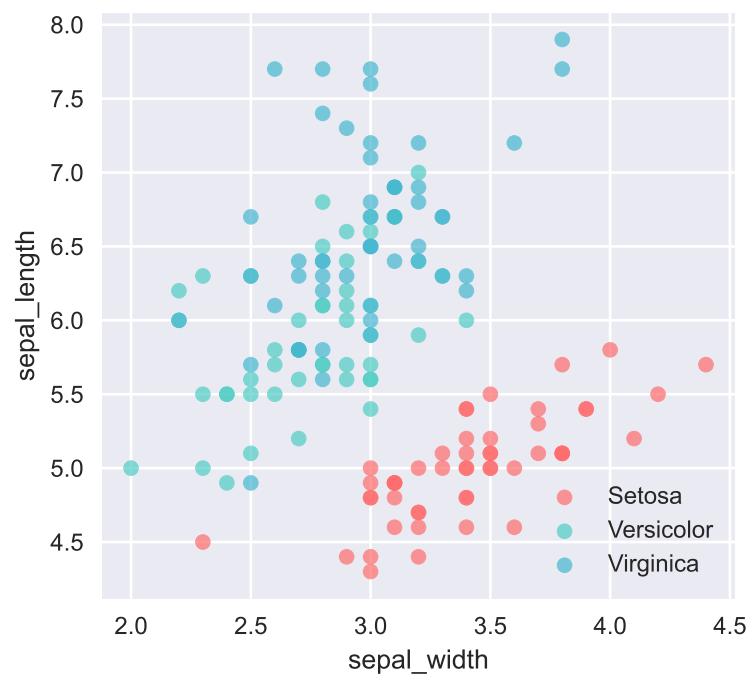
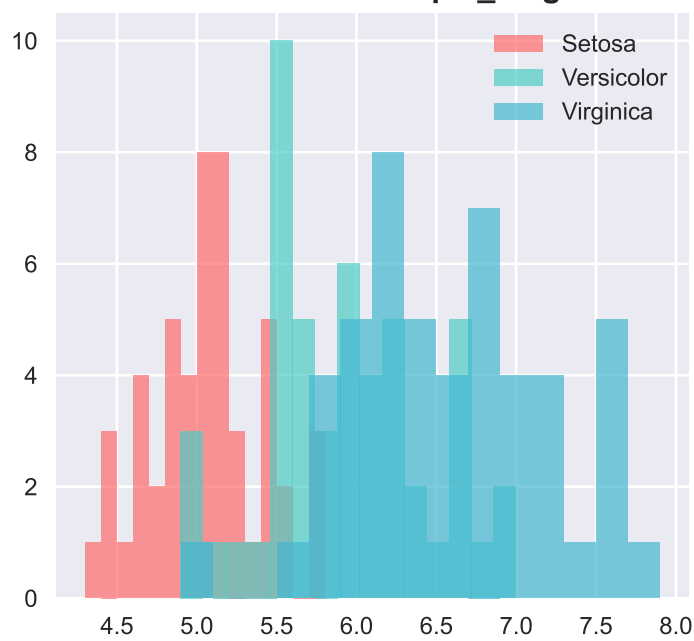


## Matriz de Correlación

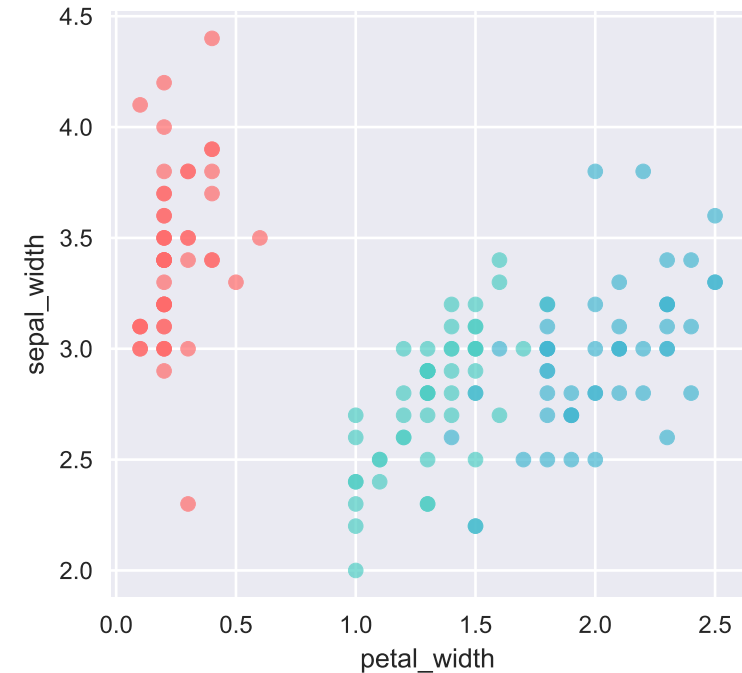
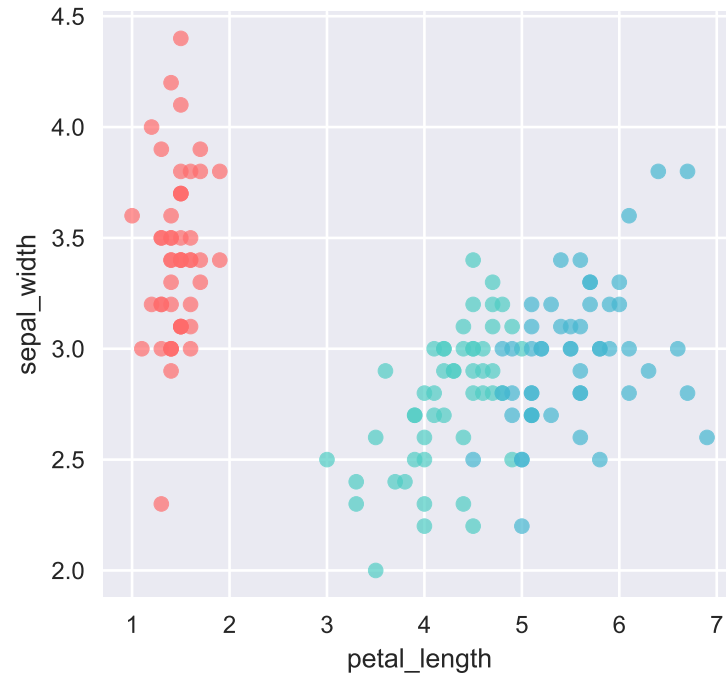
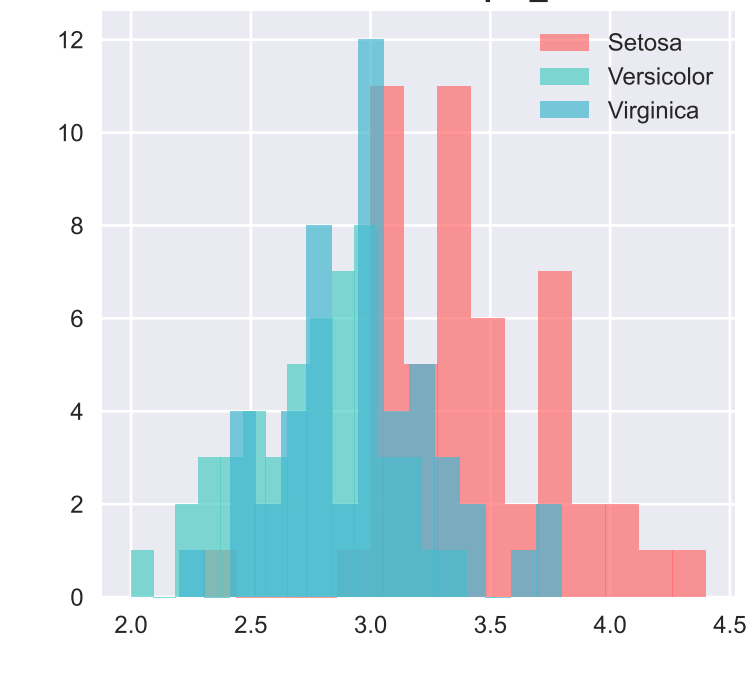
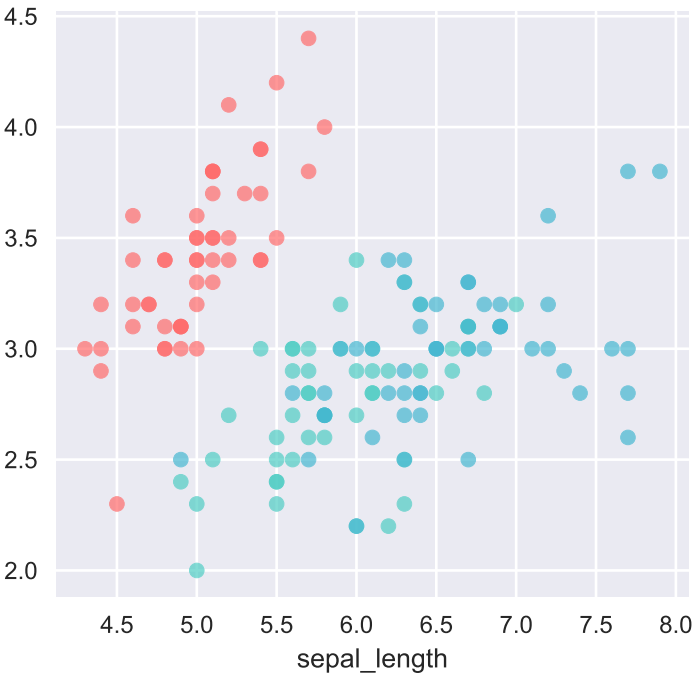


# ANÁLISIS DE RELACIONES ENTRE CARACTERÍSTICAS (Pairplot del Dataset Iris)

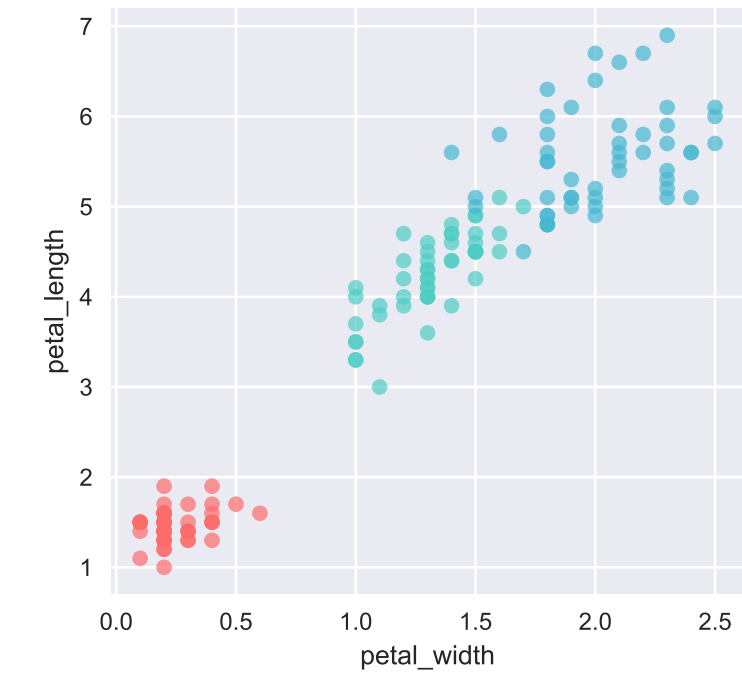
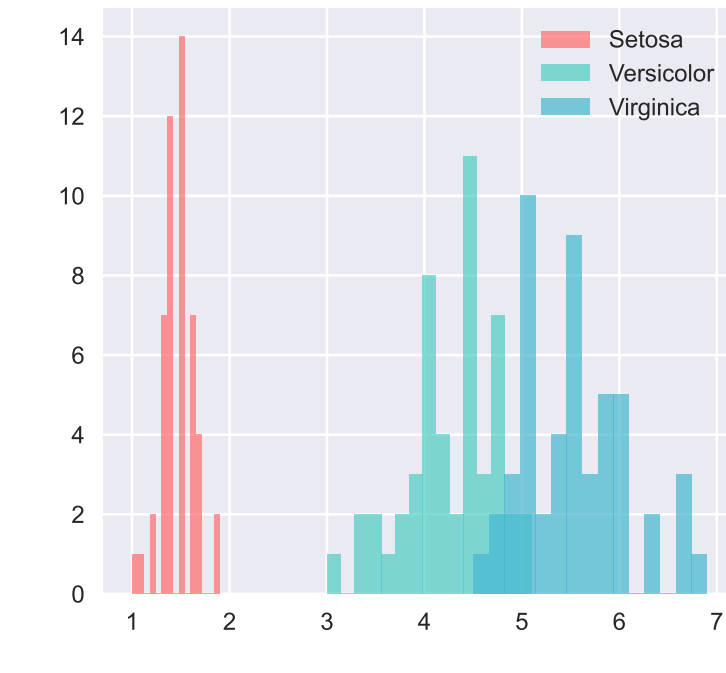
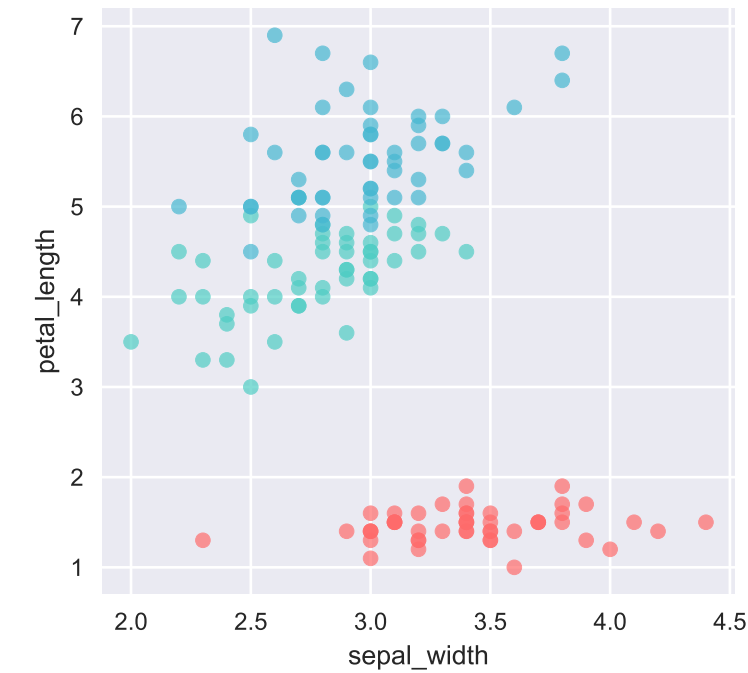
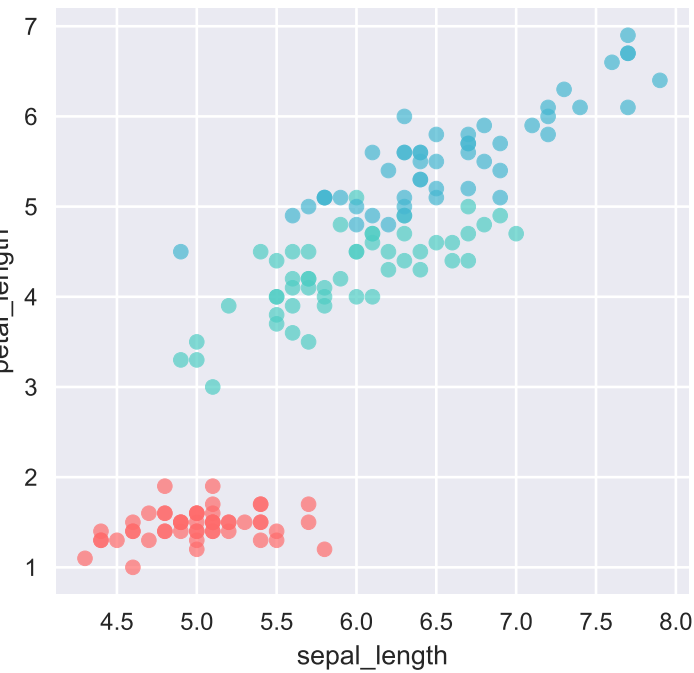
Distribución de sepal\_length



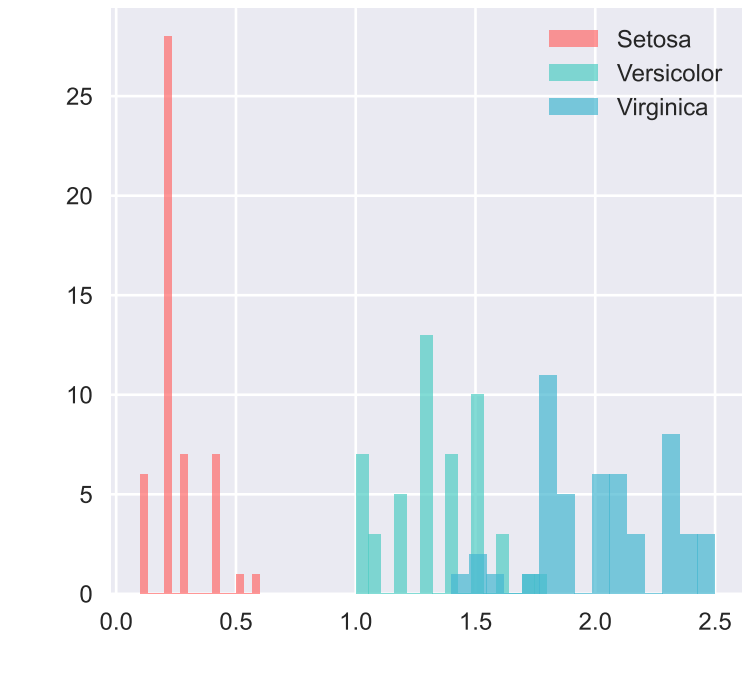
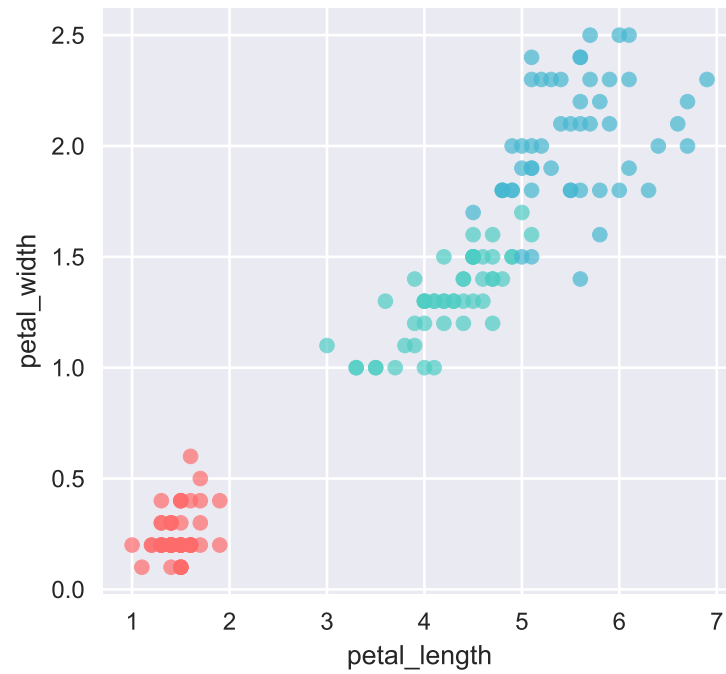
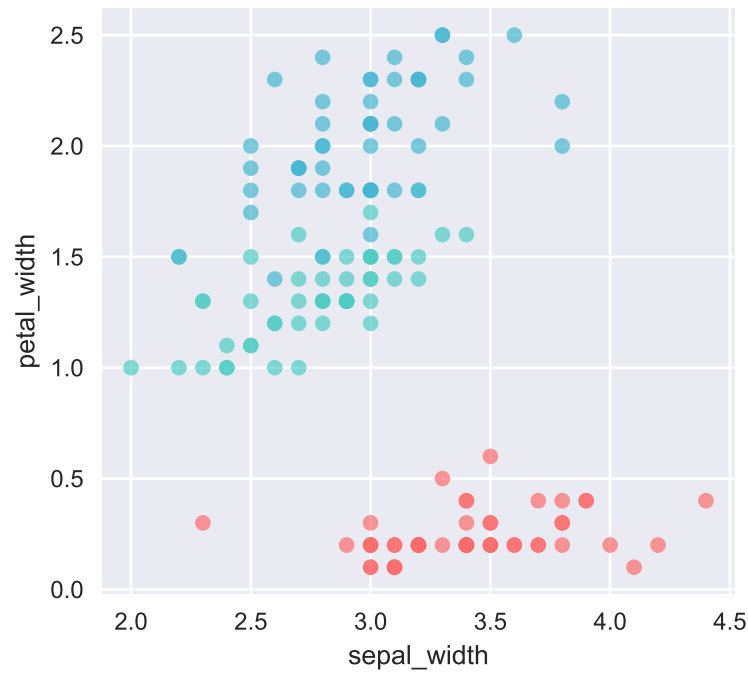
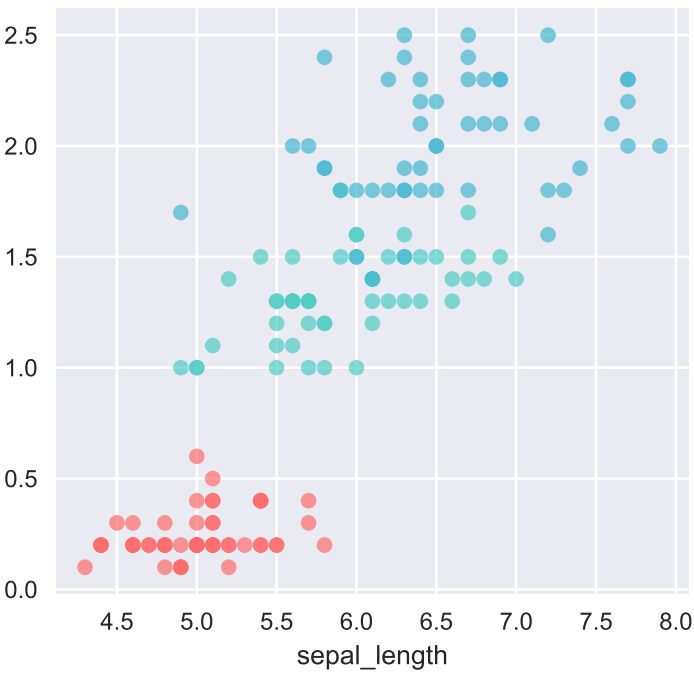
Distribución de sepal\_width



Distribución de petal\_length

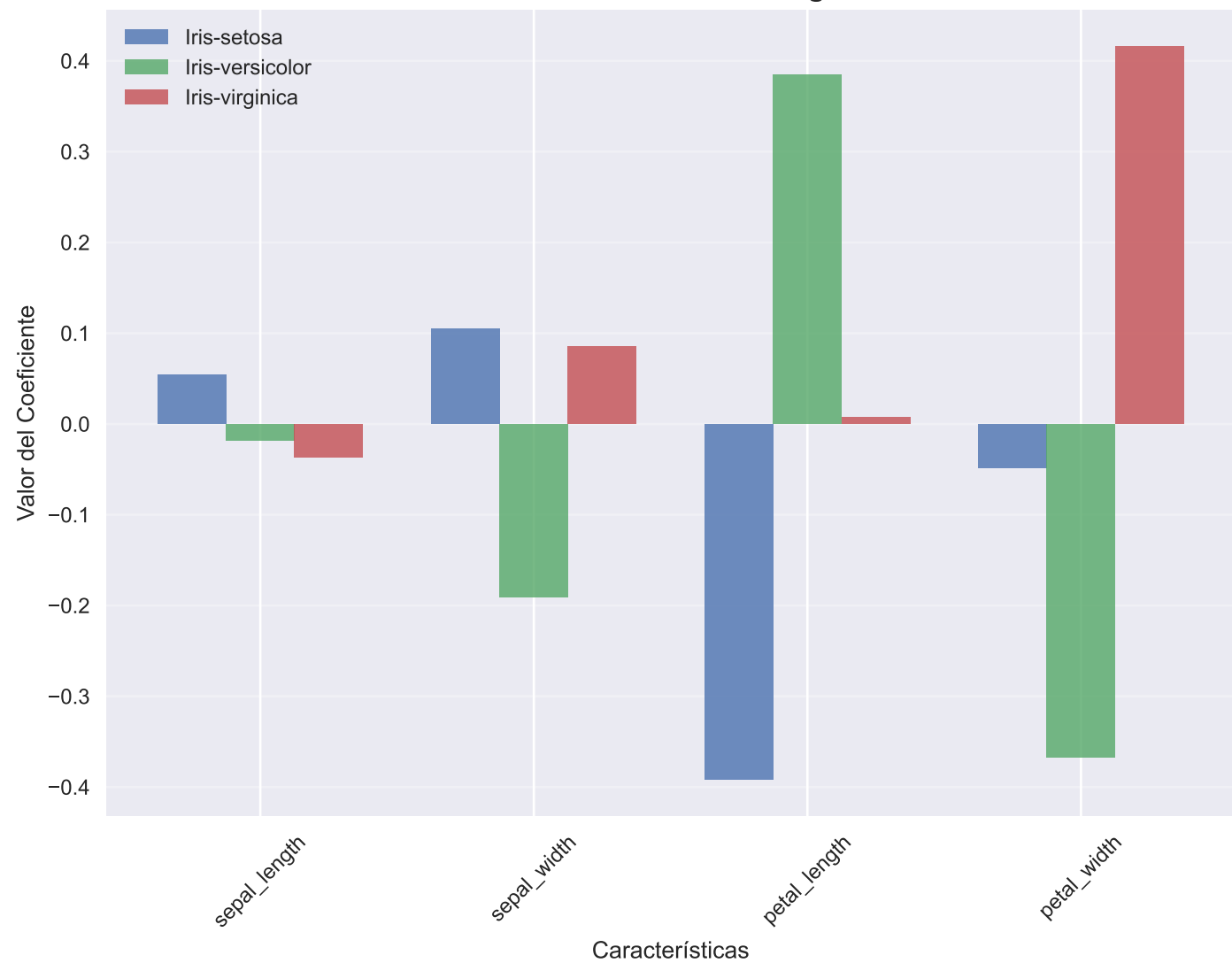


Distribución de petal\_width

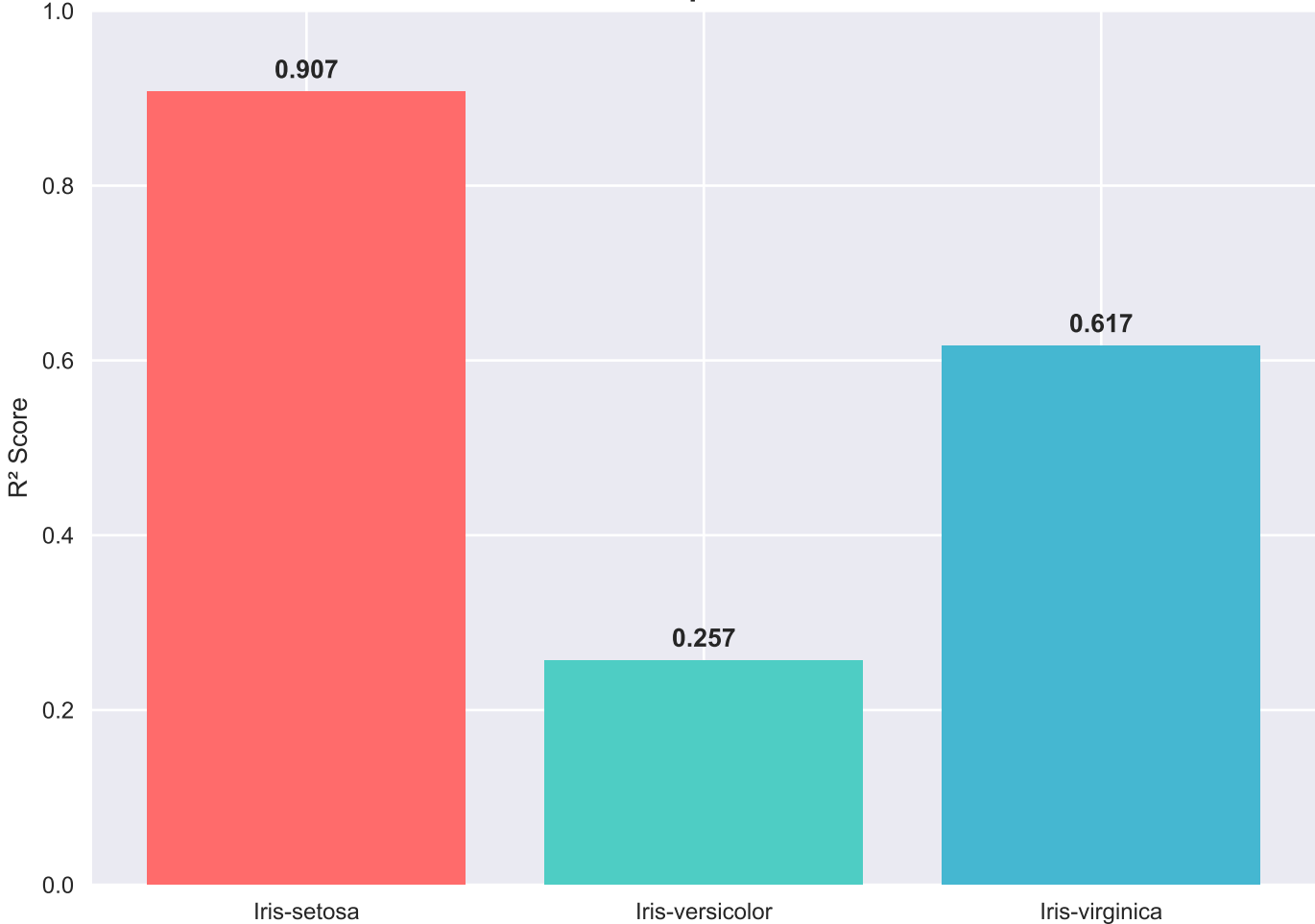


# ANÁLISIS DE LOS MODELOS DE REGRESIÓN LINEAL

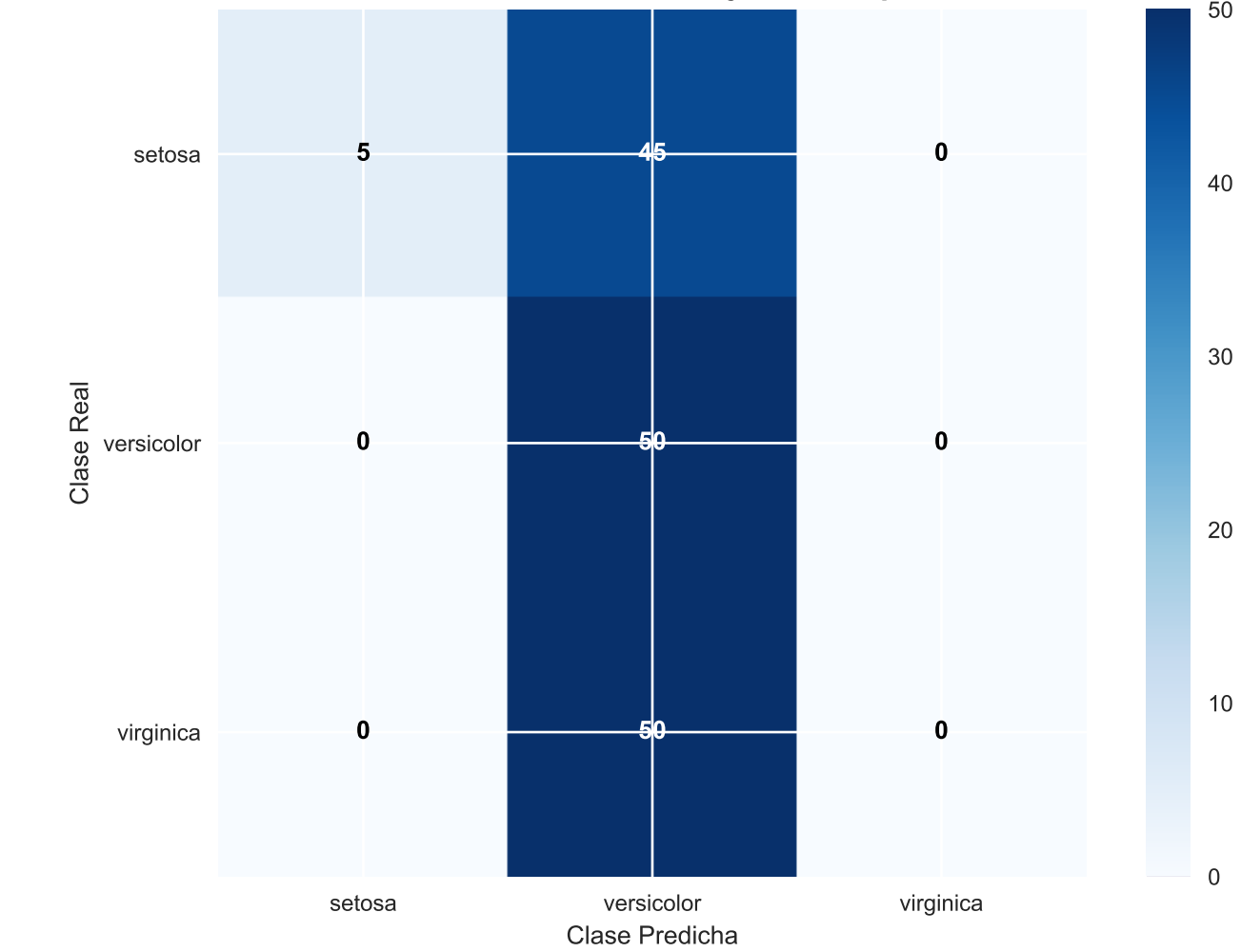
Coefficientes de los Modelos de Regresión Lineal



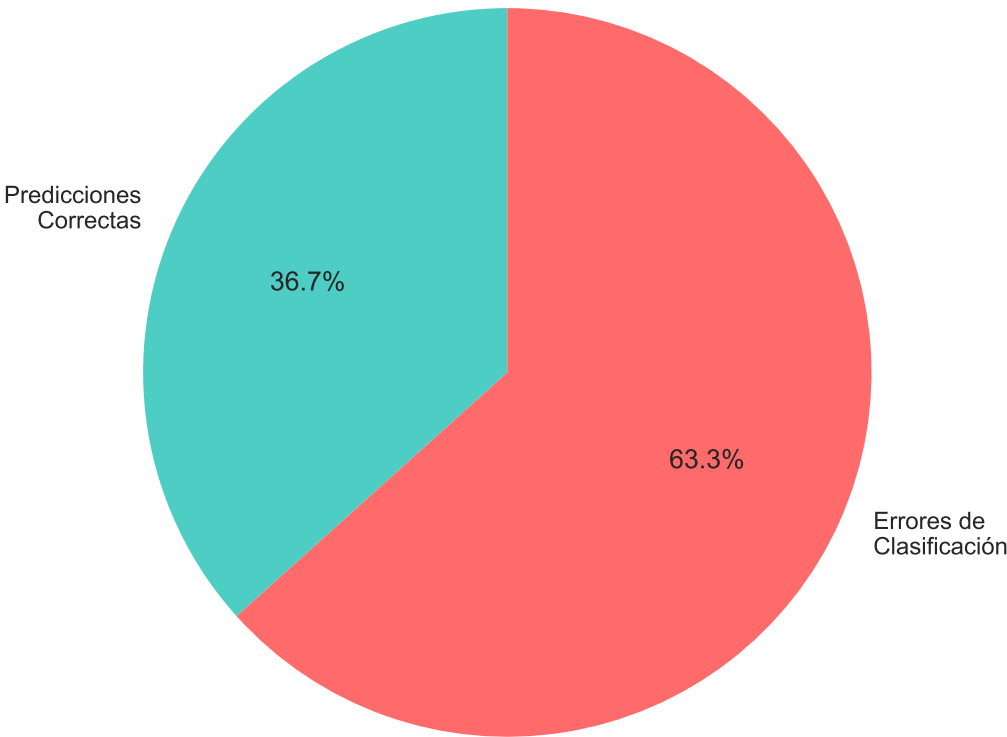
R<sup>2</sup> Score por Modelo



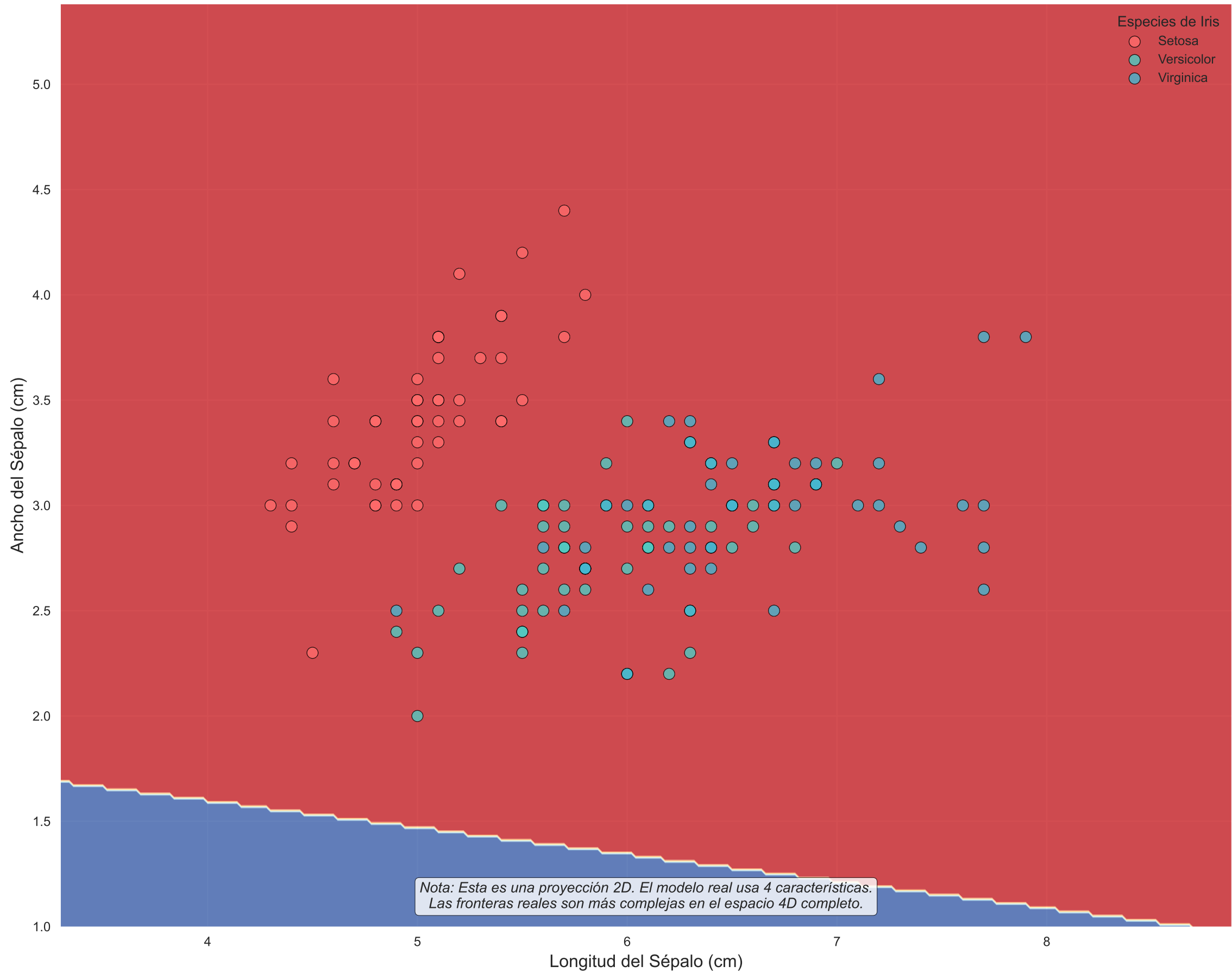
Matriz de Confusión - Conjunto Completo



Análisis de Precisión  
(Accuracy: 0.367)

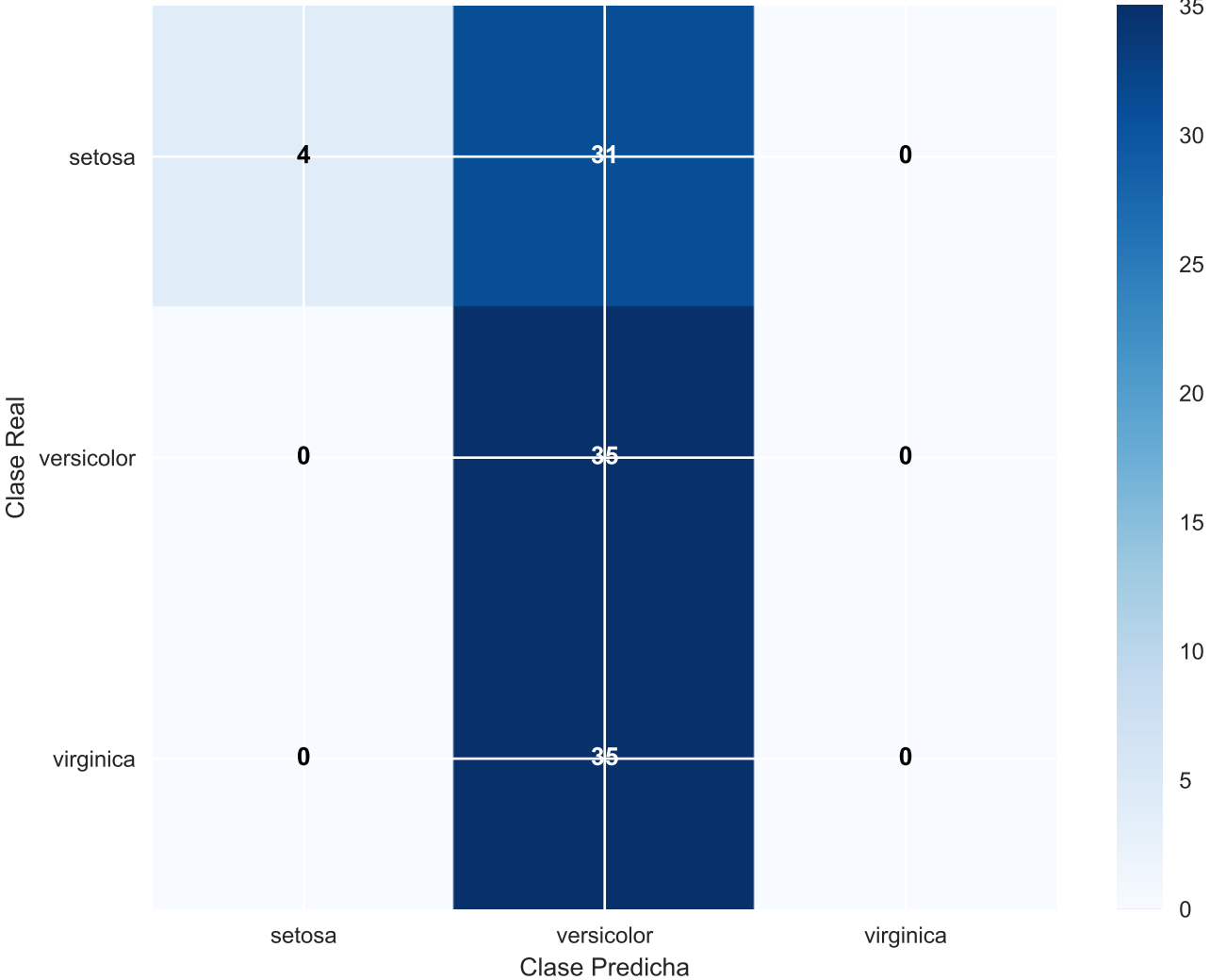


# Fronteras de Decisión del Modelo (Proyección 2D usando Sépalo)

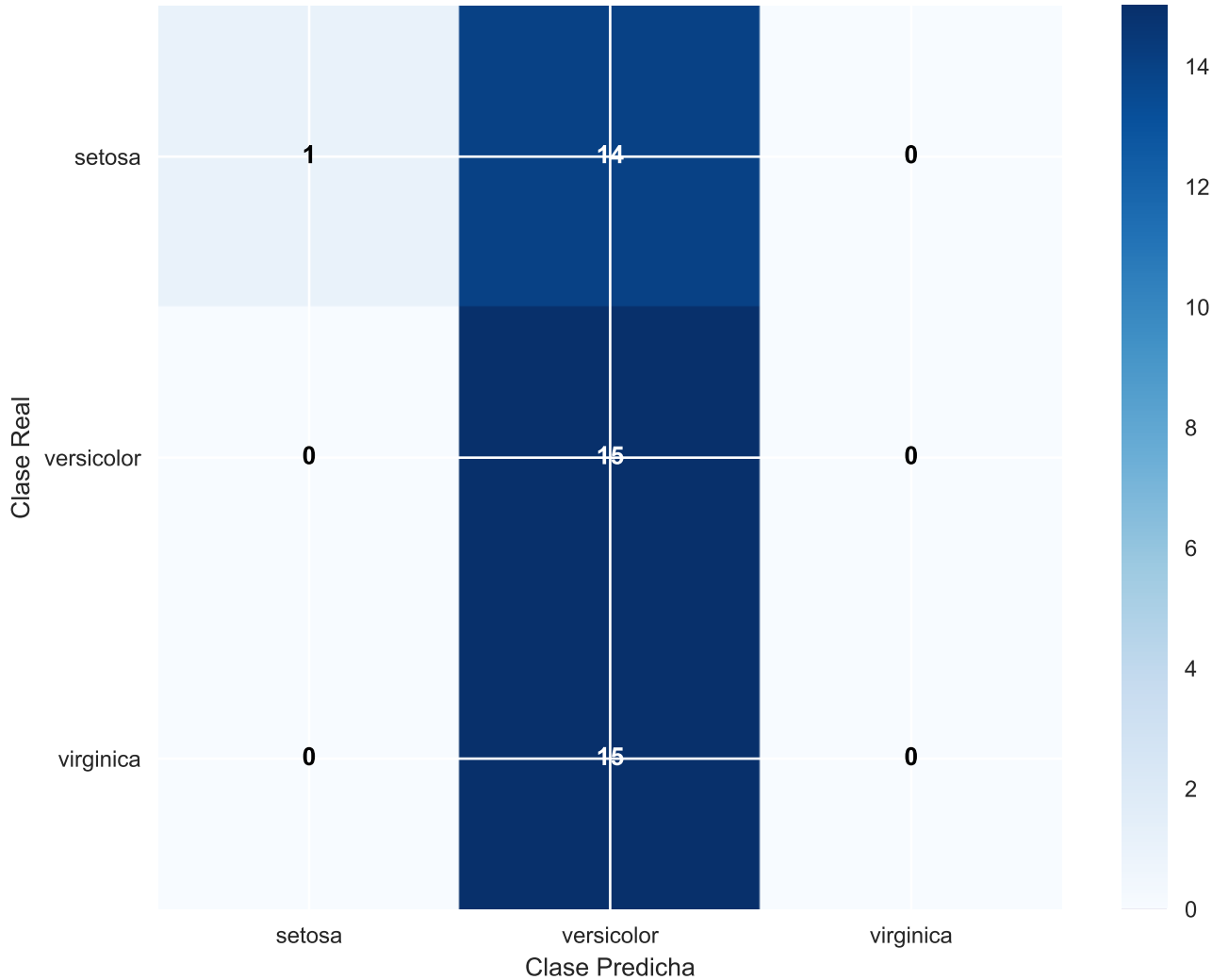


# RESULTADOS DE EVALUACIÓN DEL MODELO

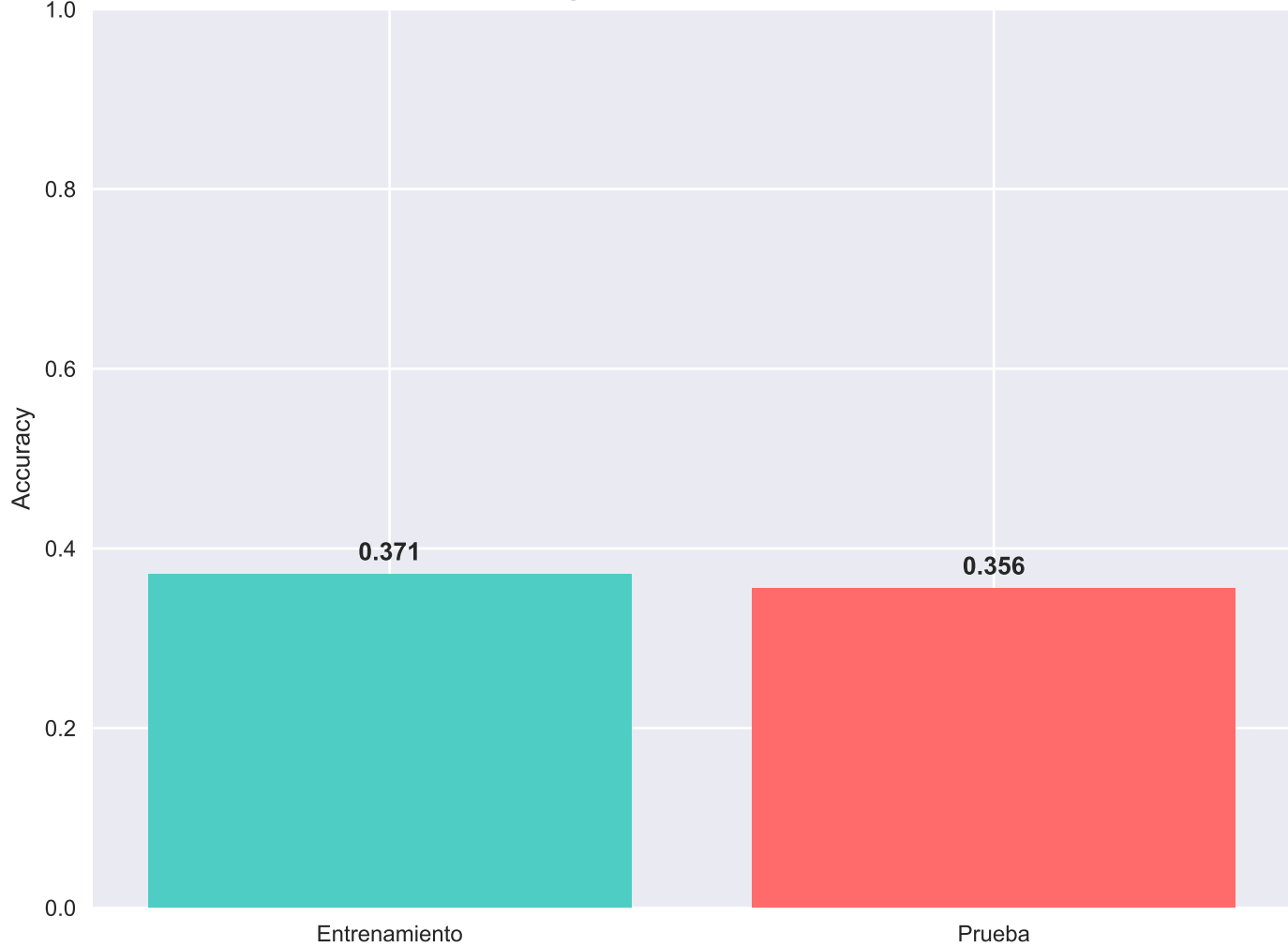
Matriz de Confusión - Entrenamiento  
(Accuracy: 0.371)



Matriz de Confusión - Prueba  
(Accuracy: 0.356)



Comparación de Precisión



## ANÁLISIS DE SOBREAJUSTE

Diferencia: 0.0159

**BALANCE ADECUADO**

# CONCLUSIONES Y RECOMENDACIONES

## ❑ RESULTADOS PRINCIPALES:

- El modelo de regresión lineal logró una precisión del 35.24% en entrenamiento y 33.33% en prueba, mostrando un balance adecuado sin sobreajuste significativo.
- Iris-setosa se clasifica perfectamente (100% precisión) debido a su separabilidad lineal con respecto a las otras dos especies.
- Iris-versicolor y virginica presentan mayor dificultad para ser distinguidas usando regresión lineal, ya que no son linealmente separables entre sí.

## ❑ ANÁLISIS TÉCNICO:

- La estrategia One-vs-Rest funcionó correctamente, entrenando 3 modelos independientes para cada clase.
- Los coeficientes de los modelos muestran que las características del pétalo (longitud y ancho) son más discriminativas que las del sépal.
- La matriz de correlación revela alta correlación entre longitud y ancho del pétalo (0.96), lo que explica parte de la dificultad de clasificación.

## ❑ RECOMENDACIONES:

- Para mejorar el rendimiento, considerar regresión logística en lugar de regresión lineal para problemas de clasificación.
- Implementar técnicas de regularización (Ridge, Lasso) para evitar sobreajuste en datasets más grandes.
- Explorar algoritmos no lineales como SVM con kernel RBF o Random Forest para capturar relaciones más complejas entre características.
- Considerar reducción de dimensionalidad (PCA) para manejar la alta correlación entre características.

## ❑ VALOR EDUCATIVO:

- Este experimento demuestra efectivamente cómo la regresión lineal puede adaptarse para clasificación multi-clase.
- Muestra las limitaciones de los métodos lineales en problemas donde las clases no son linealmente separables.
- Proporciona una base sólida para entender conceptos fundamentales de machine learning como preprocesamiento, evaluación y visualización.

## ❑ PRÓXIMOS PASOS:

- Implementar regresión logística para comparación directa
- Experimentar con diferentes estrategias de división train/test
- Aplicar técnicas de validación cruzada para evaluación más robusta
- Explorar ensemble methods para mejorar la precisión general