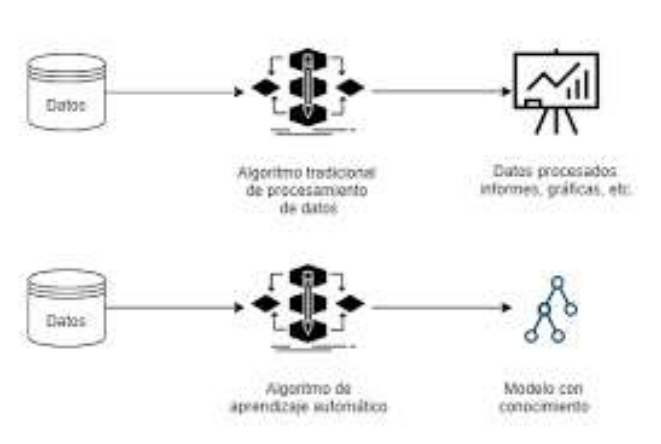


P303

Ing. Maximiliano Carsi Castrejón – Extracción y Conocimiento en Bases de Datos

DESCRIPCIÓN BREVE

Este documento trata sobre solucionar un problema en lenguaje de programación R

Luis Eduardo Bahena Castillo
9°C IDyGS

INTRODUCCIÓN

Práctica: Regresión Lineal con el Dataset Iris

Objetivo de la Práctica:

Aprender a calcular una regresión lineal simple y verificar los resultados utilizando R. Esta práctica incluye el cálculo de medias, varianza, covarianza y los coeficientes de regresión (β0 y β1), así como la predicción de nuevos valores. Los estudiantes verificarán sus resultados usando R y evaluarán el modelo con R^2.

Parte 1: Cálculo Digital en R

1. Selección de Variables

- **Variable Independiente (x)**: `Sepal.Length`
- **Variable Dependiente (y)**: `Petal.Length`

2. Pasos a Seguir en R

1. **Cargar el Dataset y Calcular Medias**:
   - Calcular la media de `Sepal.Length`
   - Calcular la media de `Petal.Length`
2. **Calcular la Varianza de x y la Covarianza de x y**:
   - Calcular la varianza de `Sepal.Length`
   - Calcular la covarianza entre `Sepal.Length` y `Petal.Length`
3. **Calcular los Coeficientes de Regresión (β0 y β1)**:
   - Calcular β1
   - Calcular β0
4. **Predicción de un Nuevo Valor**:
   - Utilizar los coeficientes calculados para predecir el `Petal.Length` para un nuevo valor de `Sepal.Length` (por ejemplo, xnuevo=5.5)
5. **Evaluación del Modelo**:
   - Calcular el R^2
   - Graficar los datos y la línea de regresión

Parte 2: Reporte

Estructura del Reporte:

1. **Introducción**:
   - Explicación breve del objetivo de la práctica y la importancia de la regresión lineal.
2. **Cálculos y Resultados**:
   - **Medias**:
     - Media de `Sepal.Length`
     - Media de `Petal.Length`
   - **Varianza y Covarianza**:

        - Varianza de `Sepal.Length`
        - Covarianza entre `Sepal.Length` y `Petal.Length`
    - **Coeficientes de Regresión**:
        - β1: β1
        - β0: β0
    - **Predicción**:
        - Predicción de `Petal.Length` para `Sepal.Length` = 5.5
    - **Evaluación del Modelo**:
        - R^2
3. **Gráficos**:
    - Gráfico de dispersión con la línea de regresión.
4. **Conclusiones**:
    - Resumen de los hallazgos.
    - Importancia de verificar los cálculos utilizando herramientas de software.

Entrega:

- **Reporte**: Subir un informe en formato PDF que incluya la introducción, cálculos y resultados, gráficos y conclusiones.
- **Código R**: Subir el código R utilizado para la verificación.

DESARROLLO

Introducción

El objetivo de esta práctica es aprender a calcular una regresión lineal simple y verificar los resultados utilizando R. En particular, usaremos la variable `Sepal.Length` como predictor (x) para predecir la variable `Petal.Length` (y) en el dataset `Iris`. Calcularemos las medias, varianza, covarianza y los coeficientes de regresión (β0\beta_0β0 y β1\beta_1β1), así como la predicción de nuevos valores. Evaluaremos el modelo con R2R^2R2 y visualizaremos los datos junto con la línea de regresión.

Cálculos y Resultados

Medias

- **Media de Sepal.Length:**

$$\bar{x} = mean(iris\$Sepal.Length)$$

- **Media de Petal.Length:**

$$\bar{y} = mean(iris\$Petal.Length)$$

Varianza y Covarianza

- **Varianza de Sepal.Length:**

$$Var(x) = var(iris\$Sepal.Length)$$

- **Covarianza entre Sepal.Length y Petal.Length:**

$$Cov(x, y) = cov(iris\$Sepal.Length,\ iris\$Petal.Length)$$

Coeficientes de Regresión

- **β1\beta_1β1:**

$$\beta 1 = \frac{Cov(x, y)}{Var(x)}$$

- **β0\beta_0β0:**

$$\beta 0 = \bar{y} - \beta 1 \cdot \bar{x}$$

Predicción

- **Predicción de Petal.Length para Sepal.Length = 5.5:**

$$\hat{y} = \beta 0 + \beta 1 \cdot 5.5$$

Evaluación del Modelo

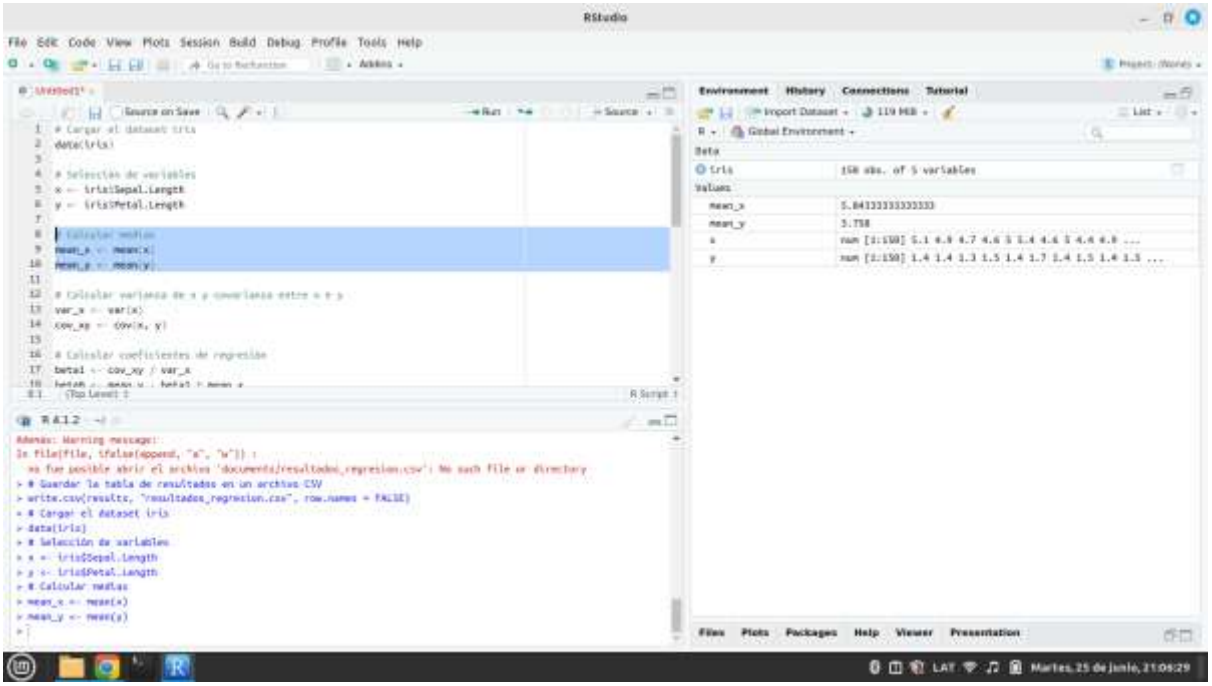
- **R^2:**

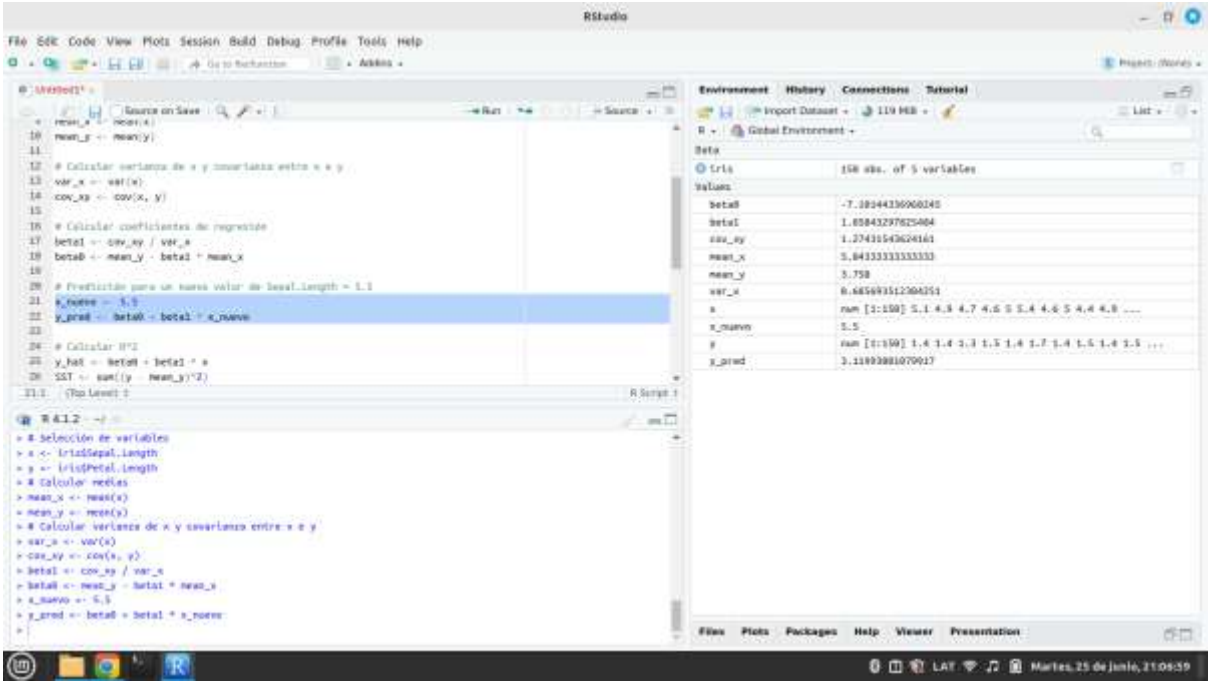
$$R2 = \frac{SSR}{SST} = 1 - \frac{\sum(yi - \hat{y}i)2}{\sum(yi - \bar{y})2}$$

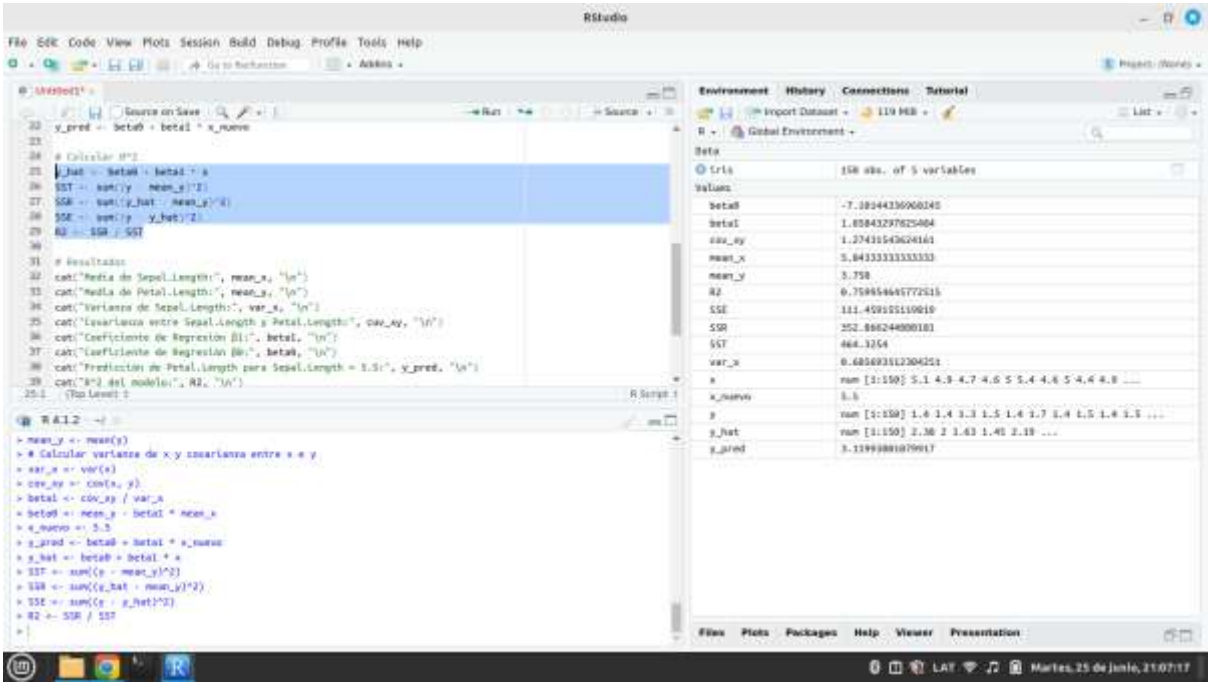
- **Gráfico de Dispersión con Línea de Regresión:**

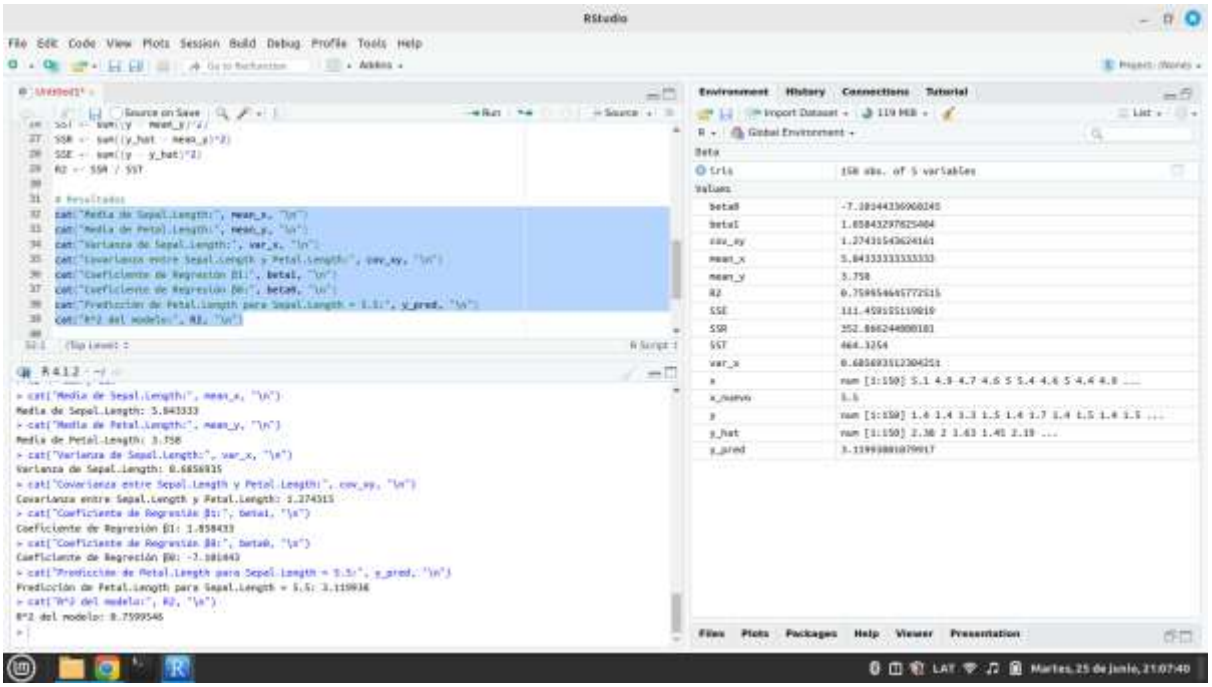
  Graficar los datos y la línea de regresión ajustada.

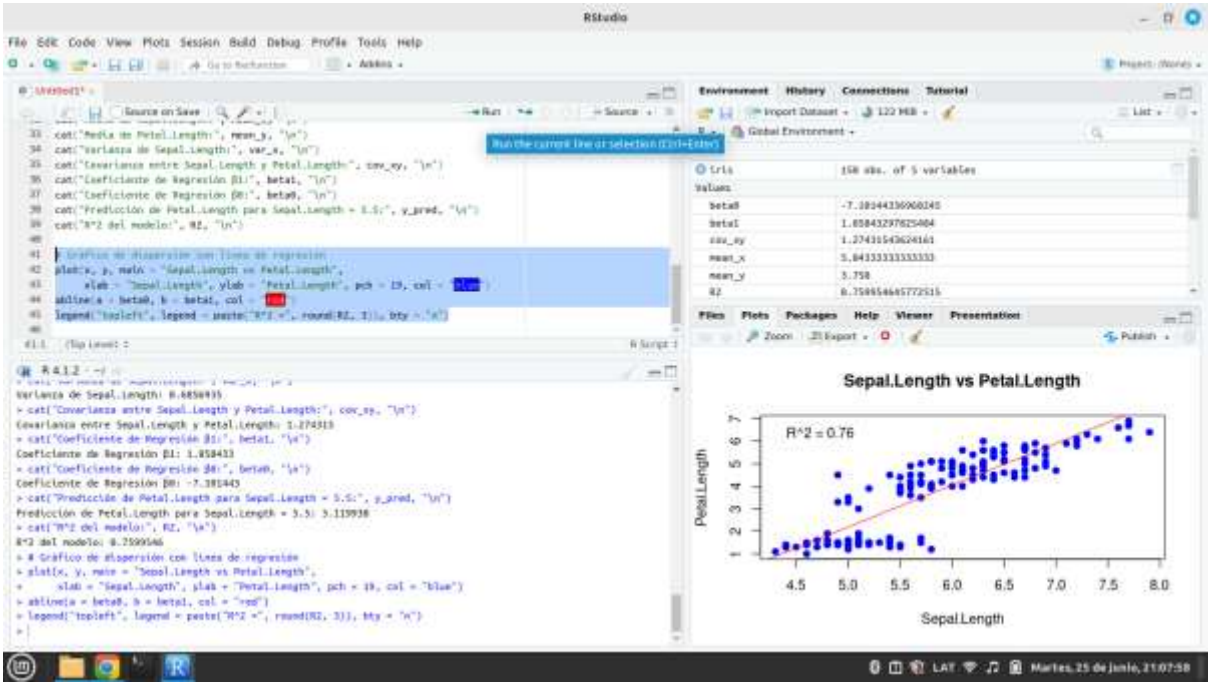
Capturas


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Environment History Connections Tutorial
Files Plots Packages Help Viewer Presentation


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Environment History Connections Tutorial
Files Plots Packages Help Viewer Presentation


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Environment History Connections Tutorial
Files Plots Packages Help Viewer Presentation
Martes, 25 de junio, 21:07:17


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Environment History Connections Tutorial
Files Plots Packages Help Viewer Presentation
Martes, 25 de junio, 21:07:40



Conclusiones

Este análisis muestra cómo se puede utilizar la regresión lineal para entender la relación entre dos variables. Los coeficientes de regresión obtenidos nos permiten predecir nuevos valores de Petal.Length basados en Sepal.Length. La evaluación del modelo con R^2 y la visualización gráfica ayuda a verificar la precisión del modelo. Es importante utilizar herramientas de software como R para realizar y verificar estos cálculos de manera eficiente y precisa.