

Linealización del radio efectivo y luminosidad media

Arturo Adrián Juárez Nava

Programación I

Profesor: Jose Manuel Nava Cervantes

Universidad de Guadalajara

Centro Universitario de Guadalajara (CUGDL)

Resumen

En este proyecto se estudia la relación que existe entre el radio efectivo y el brillo superficial de galaxias mediante un análisis estadístico y un modelo de regresión lineal, utilizando Python para procesar una base de datos astronómicos.

1. Introducción

El archivo `galaxias_data.xlsx` es una base de datos de galaxias que contiene 246,227 registros y 4 características o columnas. Las variables de interés en este proyecto son `raefcorkpg` (radio efectivo) y `muecorg` (brillo superficial). El objetivo es estudiar la relación entre estas variables.

2. Metodología

Se utilizaron las bibliotecas de Python `pandas`, `numpy`, `matplotlib` y `scipy` para el procesamiento y análisis de los datos. En primer lugar, se eliminaron los registros con valores nulos. Posteriormente, se calcularon estadísticas descriptivas para la variable `raefcorkpg`. A esta misma variable se le aplicó una transformación logarítmica en base 10 con el objetivo de facilitar el análisis y mejorar la linealidad de la relación con la variable `muecorg`. Finalmente, se ajustó un modelo de regresión lineal simple entre `muecorg` y $\log(\text{raefcorkpg})$ para evaluar la existencia de una relación lineal entre ambas variables.

2.1. Linealización

La transformación logarítmica que se realizó a la variable `raefcorkpg` se conoce como linealización. Esto funciona bajo la hipótesis de que las variables en cuestión tienen una relación no lineal, pero más específicamente, que la variable `raefcorkpg` se comporta de manera exponencial respecto de la variable `muecorg`. Esto último permite entonces transformar dicha relación no lineal en una relación aproximadamente lineal, lo que simplifica el análisis permitiendo usar herramientas como la regresión lineal.

3. Resultados

3.1. Datos estadísticos

Los datos estadísticos obtenidos se muestran en la siguiente tabla:

Estadístico	Valor
Media	5.3285
Mediana	4.9685
Moda	2.1750
Desviación estándar	2.8567
Varianza	8.1609
Mínimo	0.0020
Máximo	284.8071
Rango	284.8050
Covarianza	1.1190

Tabla 1: Estadísticas descriptivas de la variable `raefcorkpg`.

3.2. Regresión lineal

Se aplicó un modelo de regresión lineal donde se obtuvo la siguiente ecuación de la recta:

$$\text{muecorg} = 1.7693 \cdot \log(\text{raefcorkpg}) + 20.2922 \quad (1)$$

con un coeficiente de determinación $R^2 = 0.2264$

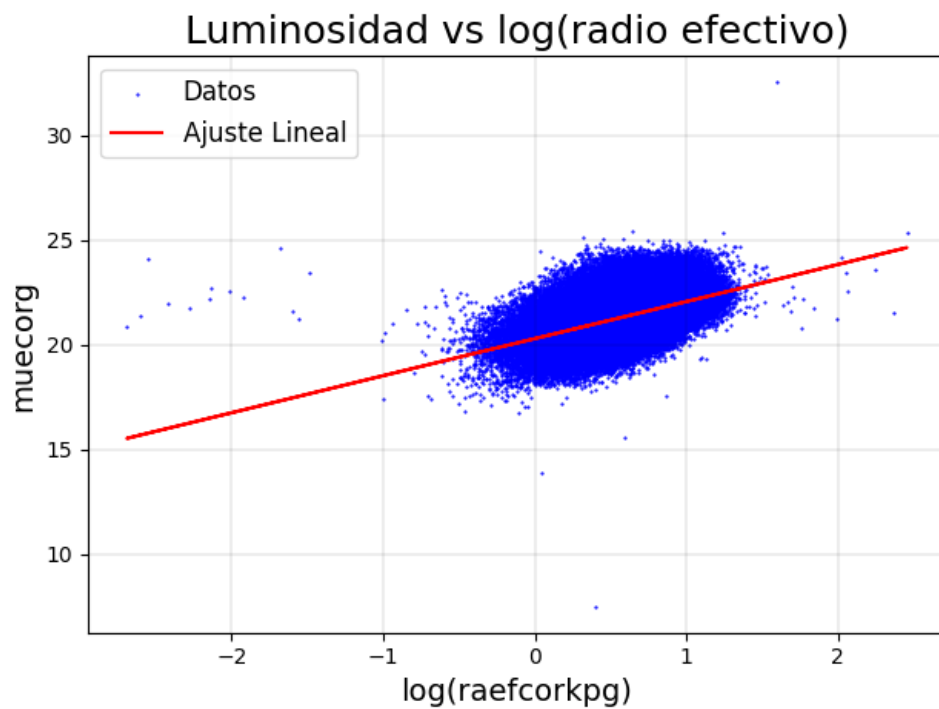


Figura 1: Ajuste lineal `muecorg` vs `log(raefcorkpg)`

4. Conclusiones

El modelo lineal entre la luminosidad superficial media (`muecorg`) y el logaritmo base 10 del radio efectivo (`log(raefcorkpg)`) mostró una correlación débil de $R^2 = 0.2264$, lo que sugiere que la relación entre estas variables no sigue una relación lineal simple o que la transformación logarítmica aplicada no fue suficiente para linealizarla. Este bajo coeficiente de determinación podría deberse a la presencia de ruido en los datos, surgiendo la necesidad de aplicar filtros más estrictos para limpieza de los datos, o a que se requiera un modelo no lineal.