

Estadística II - Taller 03 Semestre: 2024-01

Profesores: Johnatan Cardona Jimenez, Freddy Hernández Barajas, Raul Alberto Perez

Monitor: Ronald Palencia

El conjunto de datos (kc_house_data.csv) utilizado en este análisis contiene información sobre los precios de las casas y sus características asociadas. Aquí hay un breve resumen del conjunto de datos:

- id: Un identificador único para cada propiedad listada.
- date: La fecha en la que la casa fue vendida.
- price: El precio de venta de la casa.
- bedrooms: El número de dormitorios en la casa.
- bathrooms: El número de baños en la casa, a menudo en formato decimal para representar baños parciales (por ejemplo, 1.5 para un baño completo y un aseo).
- sqft_living: El área habitable en pies cuadrados de la casa.
- sqft_lot: El tamaño total del terreno en pies cuadrados.
- floors: El número de pisos en la casa.

Para el desarrollo de este taller trabajaremos con las variables Price y sqft_lot.

A continuación se muestran las primeras 6 observaciones o registros del conjunto de datos

id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors
7129300520	2014-10-13	221900	3	1.00	1180	5650	1
6414100192	2014-12-09	538000	3	2.25	2570	7242	2
5631500400	2015-02-25	180000	2	1.00	770	10000	1
2487200875	2014-12-09	604000	4	3.00	1960	5000	1
1954400510	2015-02-18	510000	3	2.00	1680	8080	1
7237550310	2014-05-12	1225000	4	4.50	5420	101930	1

Parte teorica

Realizar un repaso de los conceptos mas importantes.

Parte practica

Sean los siguientes modelos:

- a) Modelo1: Price vs sqft_living
 - b) Modelo2: $\ln(\text{Price})$ vs $\ln(\text{sqft_living})$.
- 1) Escriba el modelo de regresión lineal asociado a cada modelo, explique los supuestos, y a partir de los resultados de R escriba la ecuación de regresión ajustada. Pruebe la significancia de la regresión, calcule e interprete el coeficiente de determinación, además realice la prueba de falta y ajuste para cada modelo.
 - 2) Con base en los resultados de R analice y concluya acerca de la validez de estos modelos (Realizar la validación de los supuestos de los modelos) y elija el mejor modelo.
 - 3) Estime los precios promedios cuando los metros cuadrados de la propiedad son 100 y 291.
 - 4) Para las estimaciones anteriores calcule intervalos de predicción y del 95%.