Modelo de regresión lineal múltiple

Kevin Garcia - Alejandro Vargas

10 de febrero de 2018

Análisis exploratorio de datos

Para trabajar con la base de datos denominada 'cadata', generamos un número aleatorio con la ayuda del software R, el cuál nos arrojó el número 15529, por tanto nuestra base de datos final, quedo con las 9 variables (columnas) y con las filas desde la 15529 hasta la 16028.

El objetivo del estudio es ajustar un modelo de regresión para la variable 'Valor medio de la casa', tomando como variables explicativas las variables 'Ingreso medio','Edad media de la vivienda', 'Total de habitaciones','Total de dormitorios','Población','Hogares','Latitud' y 'Longitud'.

Análisis exploratorio de datos

Previo al ajuste e interpretación del modelo, se llevo a cabo el respectivo análisis exploratorio de datos, para tener una idea de las descriptivas mas importantes de cada variable, su forma de distribución y su rango de valores.

Variable 'Valor medio de la casa'

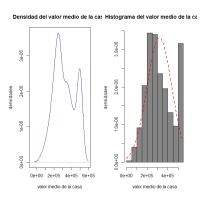


Figura: Histograma y densidad de la variable 'Valor medio de la casa'

Variable 'Ingreso medio'

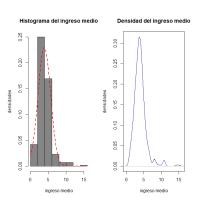


Figura: Histograma y densidad de la variable 'Ingreso medio'

Variable 'Edad media de la vivienda'

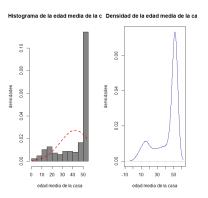


Figura: Histograma y densidad de la variable 'Edad media de la vivienda'

Variable 'Total de habitaciones'

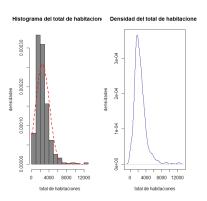


Figura: Histograma y densidad de la variable 'Total de habitaciones'

Variable 'Total de dormitorios'

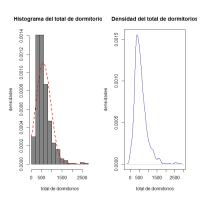


Figura: Histograma y densidad de la variable 'Total de dormitorios'

Variable 'Población'

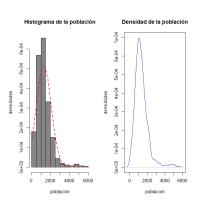


Figura: Histograma y densidad de la variable 'Población'

Variable 'Hogares'

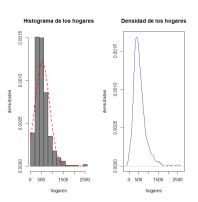


Figura: Histograma y densidad de la variable 'Hogares'

Variables 'Latitud'

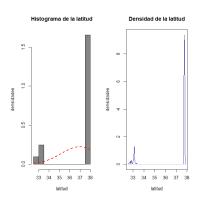


Figura: Histograma y densidad de la variable 'Latitud'

Variable 'Longitud'

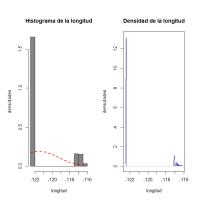


Figura: Histograma y densidad de la variable 'Longitud'

Modelo ajustado e interpretación