

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.

Análisis exploratorio de datos

	Min.	1st Qu.	Mediana	Media	3rd Qu.	Max.
Valor medio de la casa	32500	229600	294400	318400	419100	500000
Ingreso medio	0.7054	2.87	3.697	3.905	4.554	15
Edad media	2	34.75	52	42.2	52	52
Total de habitaciones	52	1580	2218	2537	3166	12480
Total de dormitorios	13	346	506	583.9	735.2	2747
Población	55	836	1168	1303	1608	5640
Hogares	13	327.80	469.5	538.8	664.5	2538
Latitud	32.61	37.72	37.75	36.95	37.78	37.81
Longitud	-122.5	-122.4	-122.4	-121.5	-122.4	-116.2

Figura: Resumen estadístico

Variable 'Valor medio de la casa'

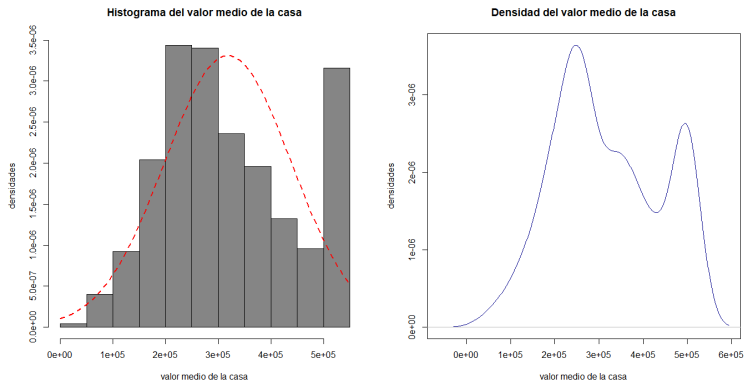


Figura: Histograma y densidad de la variable 'Valor medio de la casa'

Variable 'Ingreso medio'

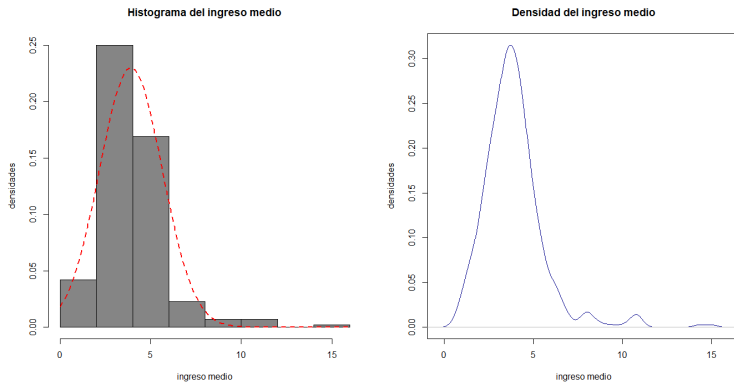


Figura: Histograma y densidad de la variable 'Ingreso medio'

Variable 'Edad media de la vivienda'

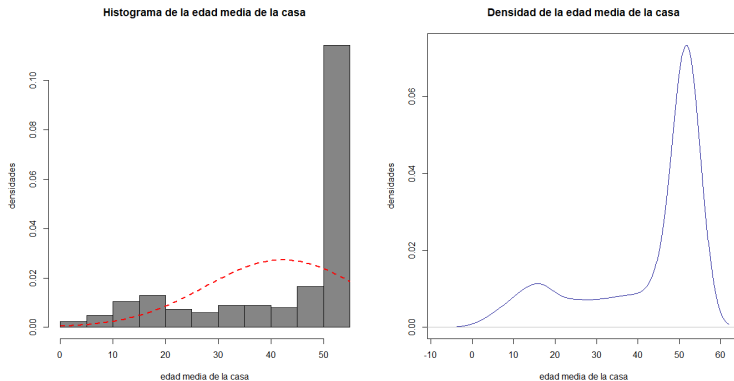


Figura: Histograma y densidad de la variable 'Edad media de la vivienda'

Variable 'Total de habitaciones'

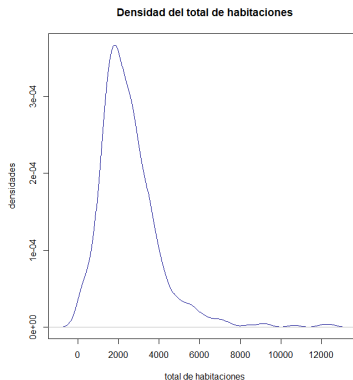
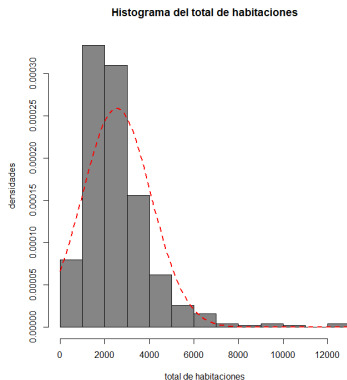


Figura: Histograma y densidad de la variable 'Total de habitaciones'

Variable 'Total de dormitorios'

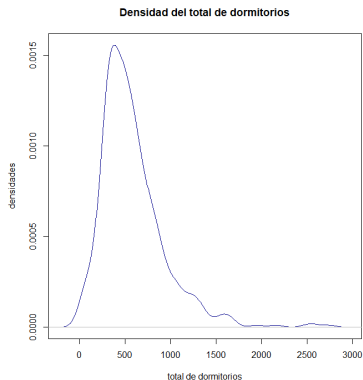
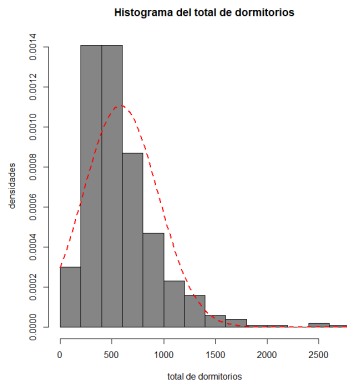


Figura: Histograma y densidad de la variable 'Total de dormitorios'

Variable 'Población'

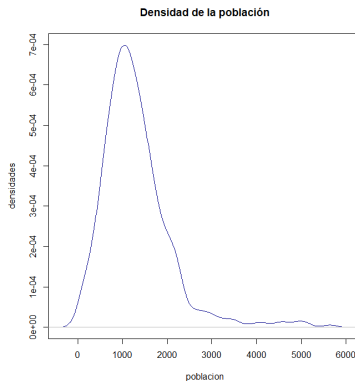
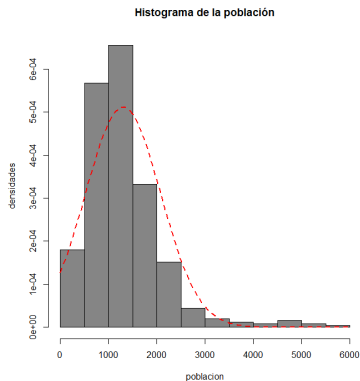


Figura: Histograma y densidad de la variable 'Población'

Variable 'Hogares'

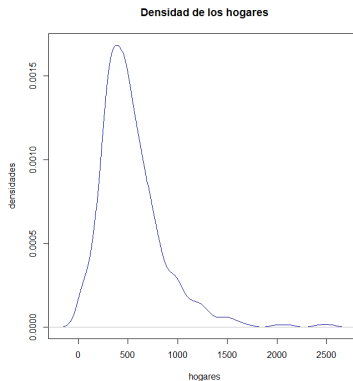
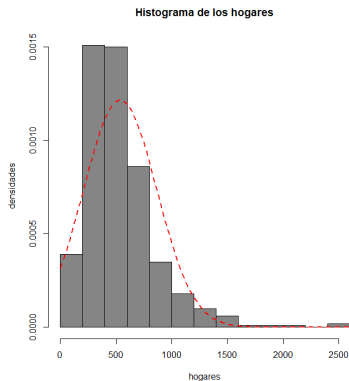


Figura: Histograma y densidad de la variable 'Hogares'

Variables 'Latitud'

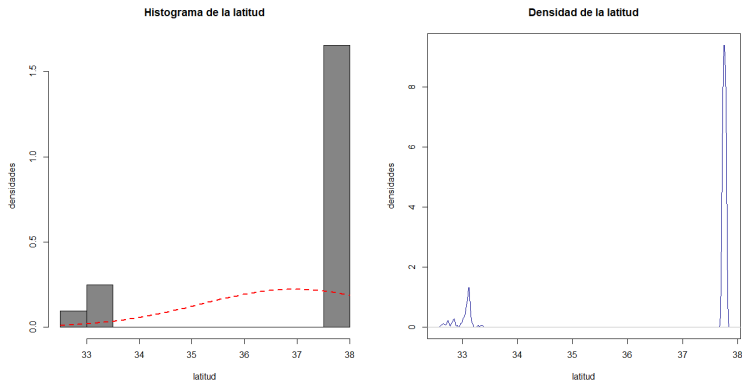


Figura: Histograma y densidad de la variable 'Latitud'

Variable 'Longitud'

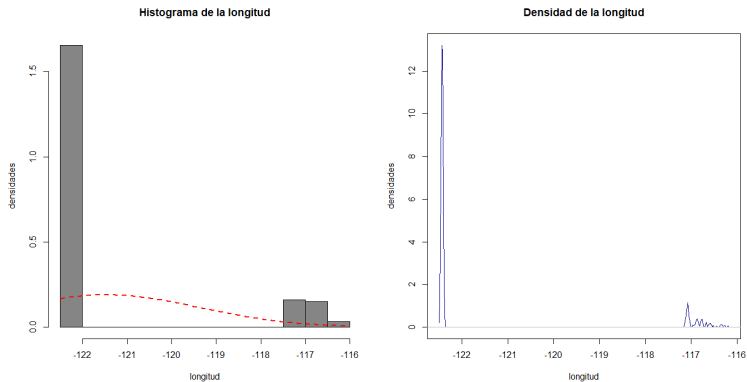


Figura: Histograma y densidad de la variable 'Longitud'

Modelo ajustado e interpretación

El modelo ajustado es:

$$\begin{aligned} \text{Valor medio de la casa} = & -4113000 + 22100(\text{Ingreso medio}) + 1835(\text{Edad media de la vivienda}) + 30,35(\text{Total de habitaciones}) \\ & - 14,43(\text{Total de dormitorios}) - 124,5(\text{Población}) + 210,2(\text{Hogares}) \\ & - 33230(\text{Latitud}) - 45080(\text{Longitud}) \end{aligned}$$