

Modelo de regresión lineal múltiple

Kevin Garcia - Alejandro Vargas

13 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.	Desviación Est.
Valor medio de la casa	32500	229600	294400	318400	419100	500000	120509.8
Ingreso medio	0.7054	2.87	3.697	3.905	4.554	15	1.735087
Edad media	2	34.75	52	42.2	52	52	14.5956
Total de habitaciones	52	1580	2218	2537	3166	12480	1539.323
Total de dormitorios	13	346	506	583.9	735.2	2747	360.0191
Población	55	836	1168	1303	1608	5640	779.5774
Hogares	13	327.80	469.5	538.8	664.5	2538	328.173
Latitud	32.61	37.72	37.75	36.95	37.78	37.81	1.790967
Longitud	-122.5	-122.4	-122.4	-121.5	-122.4	-116.2	2.099794

Figura: Resumen estadístico

Variable 'Valor medio de la casa'

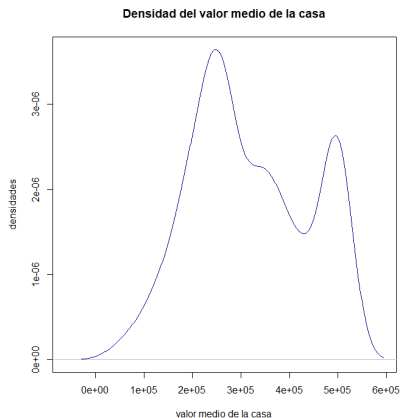
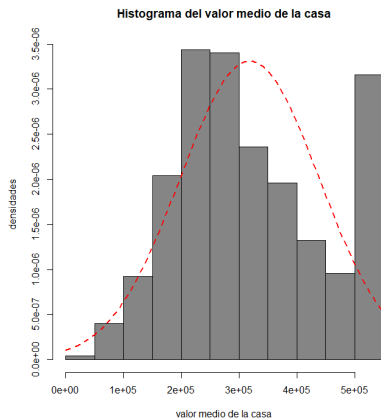


Figura: Histograma y densidad de la variable 'Valor medio de la casa'

Variable 'Ingreso medio'

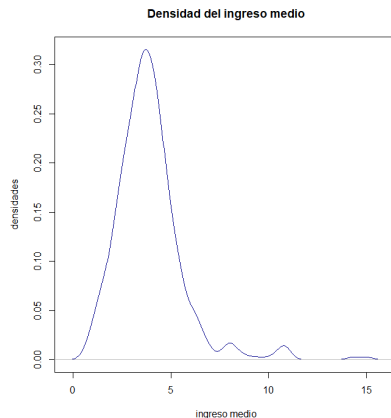
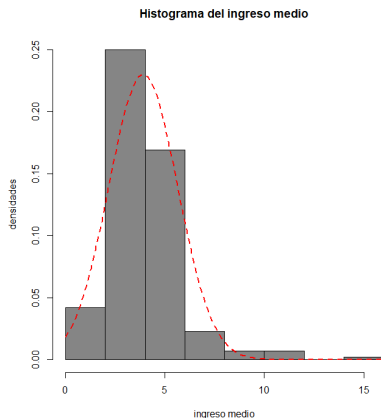


Figura: Histograma y densidad de la variable 'Ingreso medio'

Variable 'Edad media de la vivienda'

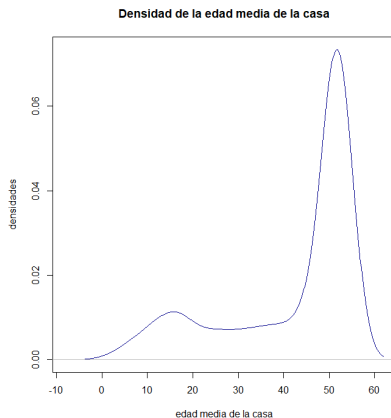
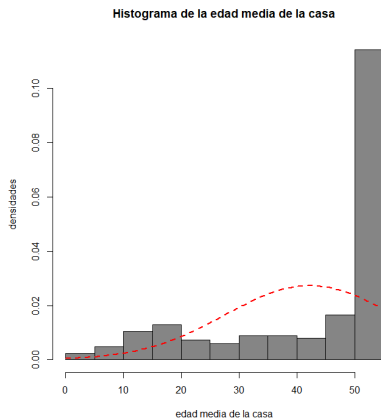


Figura: Histograma y densidad de la variable 'Edad media de la vivienda'

Variable 'Total de habitaciones'

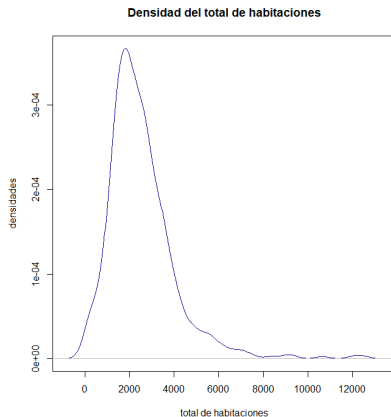
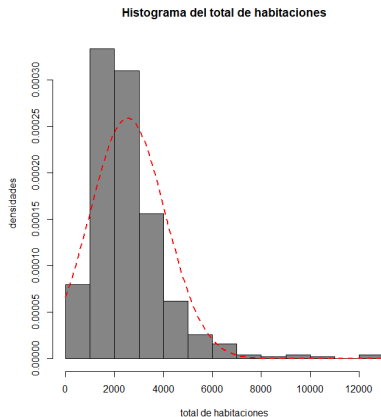


Figura: Histograma y densidad de la variable 'Total de habitaciones'

Variable 'Total de dormitorios'

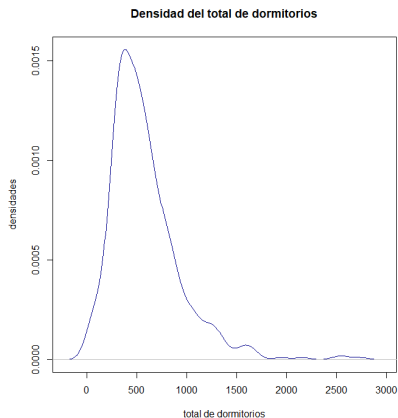
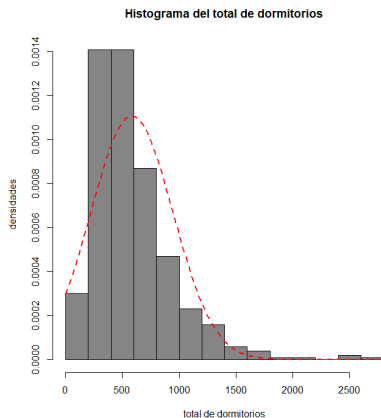


Figura: Histograma y densidad de la variable 'Total de dormitorios'

Variable 'Población'

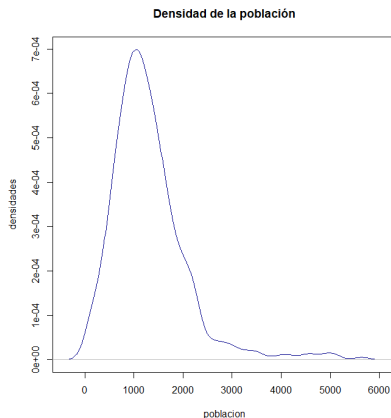
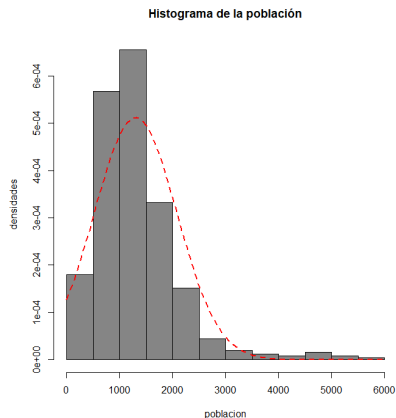


Figura: Histograma y densidad de la variable 'Población'

Variable 'Hogares'

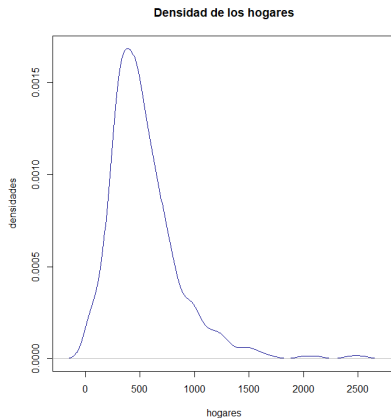
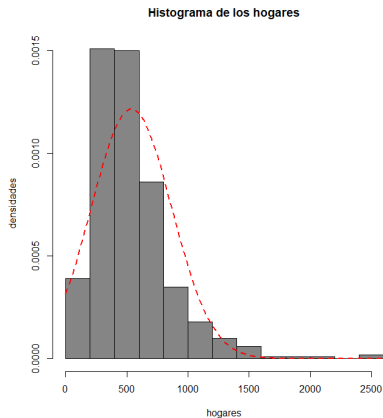


Figura: Histograma y densidad de la variable 'Hogares'

Variables 'Latitud'

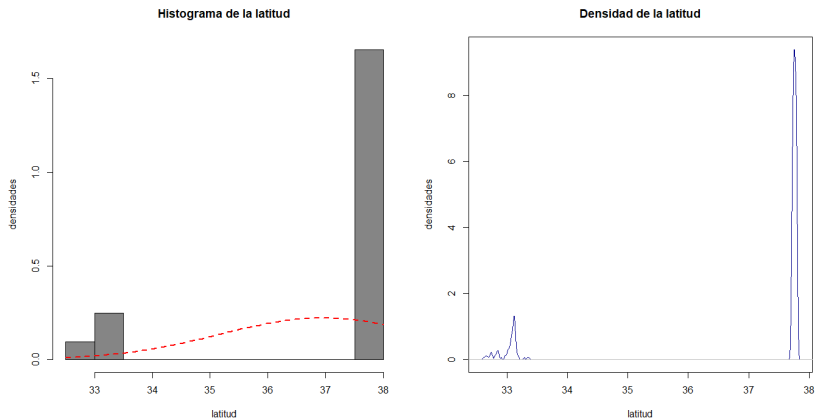


Figura: Histograma y densidad de la variable 'Latitud'

Variable 'Longitud'

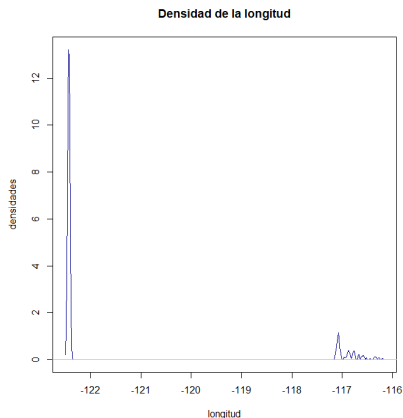
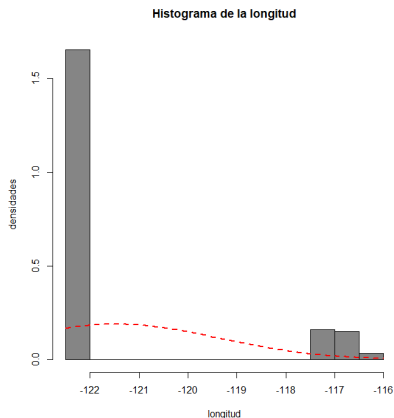


Figura: Histograma y densidad de la variable 'Longitud'

Posibles relaciones entre las variables explicativas

- Variables 'Población' y 'Hogares':

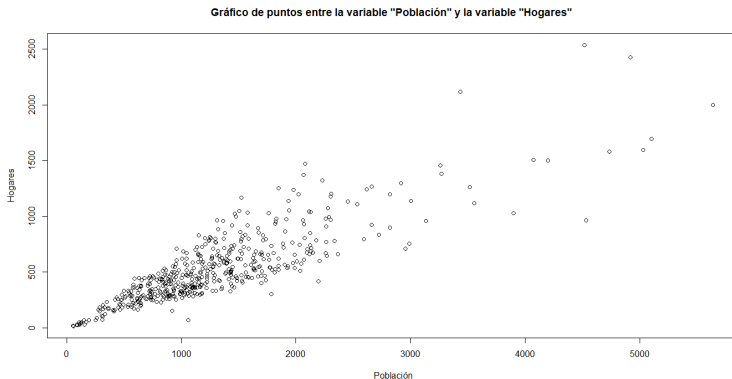


Figura: Gráfico de puntos entre las variables 'Población' y 'Hogares'

Posibles relaciones entre las variables explicativas

- correlación de Pearson: $r = 0,853918$
- correlación de Spearman: $\rho = 0,8395669$

Posibles relaciones entre las variables explicativas

- Variables 'total de habitaciones' y 'total de dormitorios':

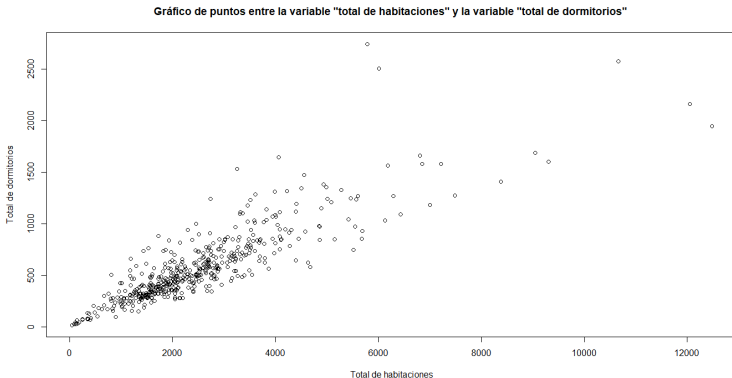


Figura: Gráfico de puntos entre las variables 'total de habitaciones' y 'total de dormitorios'

Posibles relaciones entre las variables explicativas

- correlación de Pearson: $r = 0,862072$
- correlación de Spearman: $\rho = 0,8716658$

Modelo ajustado e interpretación

El modelo ajustado es:

$$Y = -4113000 + 22100X_1 + 1835X_2 + 30,35X_3 - 14,43X_4 - 124,5X_5 + 210,2X_6 - 33230X_7 - 45080X_8$$

Donde: Y=Valor medio de la casa, X_1 =Ingreso medio, X_2 =Edad media de la vivienda, X_3 =Total de habitaciones, X_4 =Total de dormitorios, X_5 =Población, X_6 =Hogares, X_7 =Latitud, X_8 =Longitud.

Modelo ajustado e interpretación

- β_0 :
- β_1 :
- β_2 :
- β_3 :
- β_4 :
- β_5 :
- β_6 :
- β_7 :
- β_8 :
- β_9 :