

# Modelo de regresión lineal múltiple

Kevin Garcia - Alejandro Vargas

16 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 mediano de las viviendas | 32500  | 229600  | 294400  | 318400 | 419100  | 500000 | 120509.8        |
| Ingreso mediano                | 0.7054 | 2.87    | 3.697   | 3.905  | 4.554   | 15     | 1.735087        |
| Edad mediana                   | 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         |

Figura: Resumen estadístico

# Variable 'Valor mediano de las viviendas'

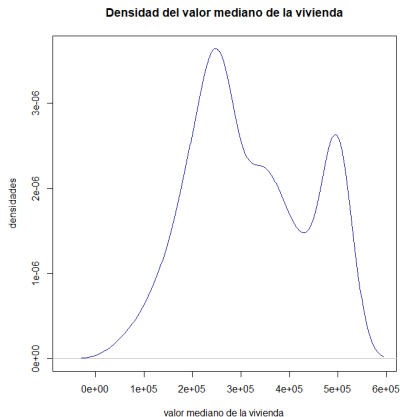
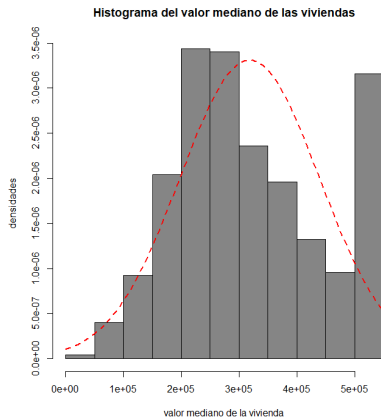


Figura: Histograma y densidad de la variable 'Valor mediano de las casas'

# Variable 'Valor mediano de las viviendas'

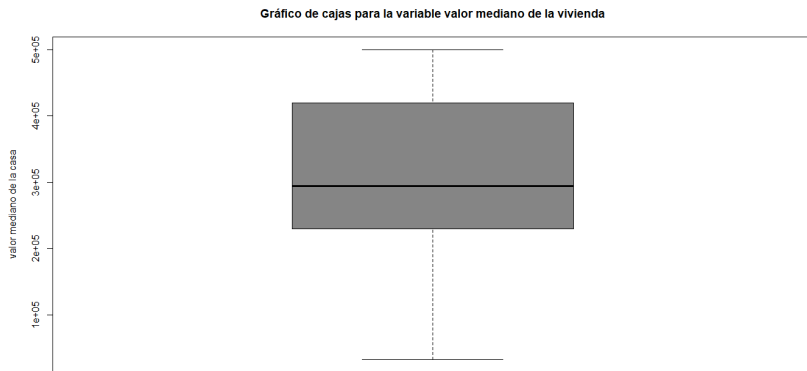
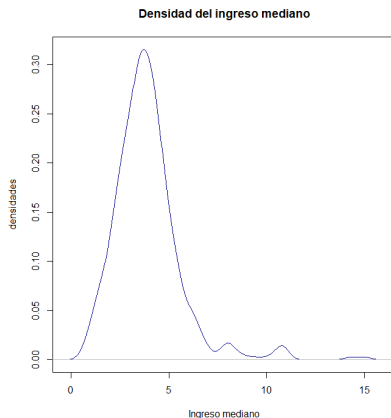
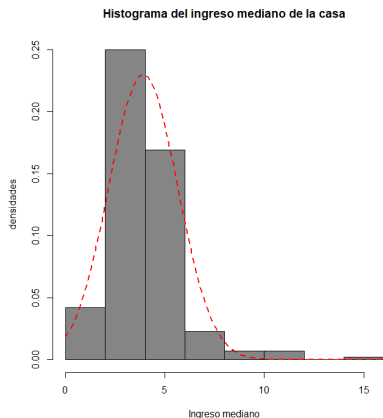


Figura: Gráfico de cajas para la variable 'Valor mediano de las casas'

# Variable 'Ingreso mediano'



**Figura:** Histograma y densidad de la variable 'Ingreso mediano'

# Variable 'Ingreso mediano'

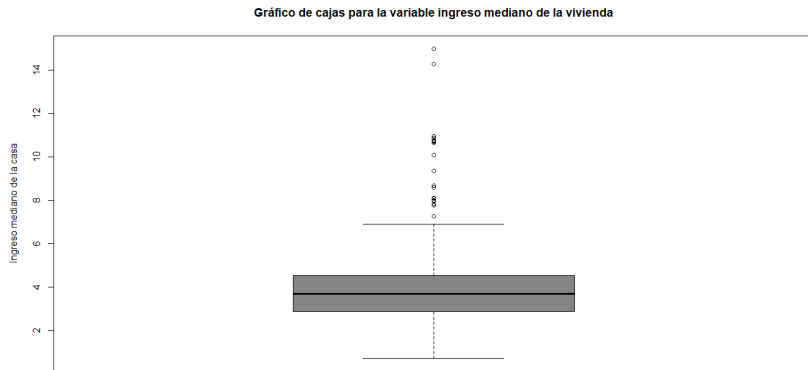
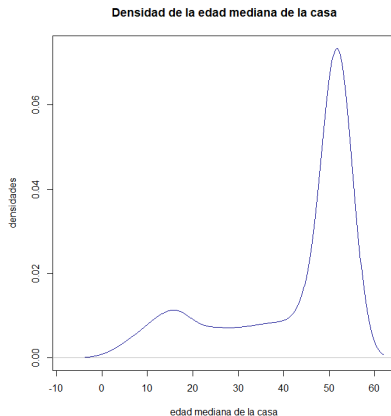
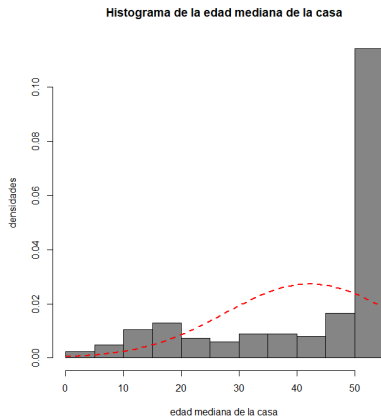


Figura: Gráfico de cajas para la variable 'Ingreso mediano de las casas'



# Variable 'Edad mediana de las viviendas'



**Figura:** Histograma y densidad de la variable 'Edad mediana de las viviendas'

# Variable 'Edad mediana de las viviendas'

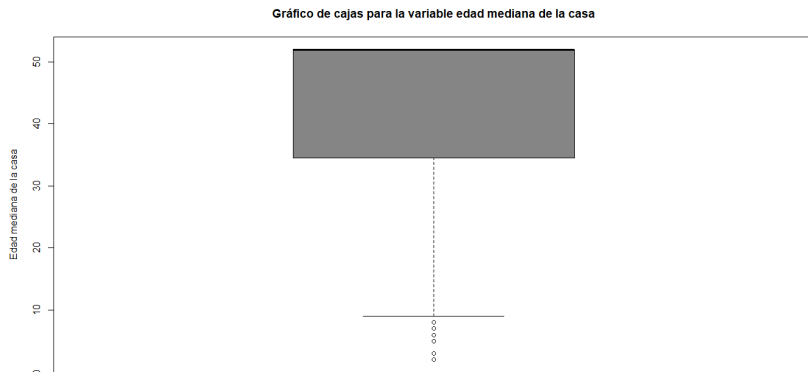
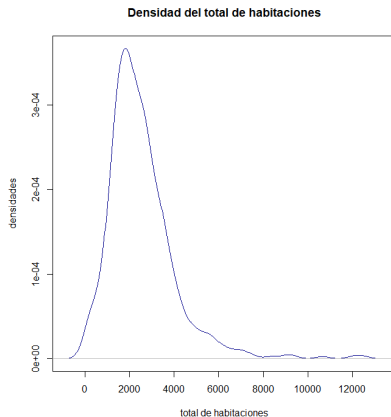
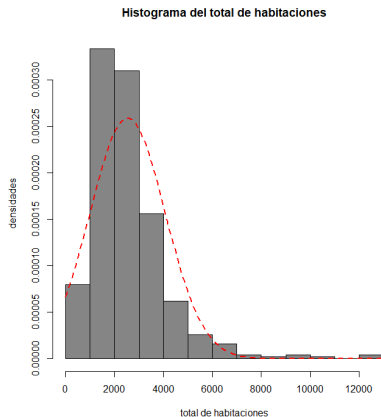


Figura: Gráfico de cajas para la variable 'Edad mediana de las viviendas'

# Variable 'Total de habitaciones'



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

# Variable 'Total de habitaciones'

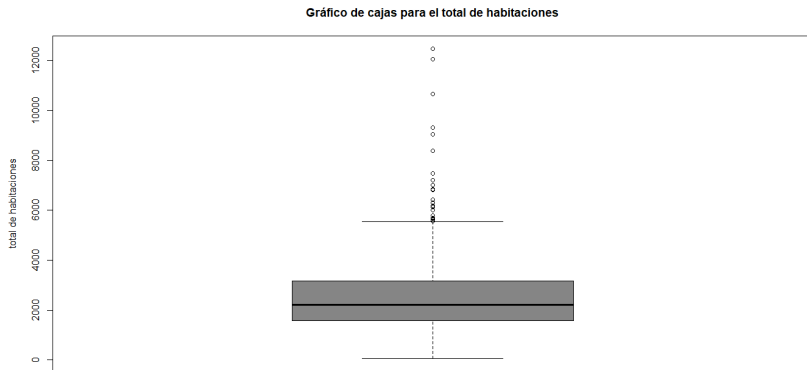
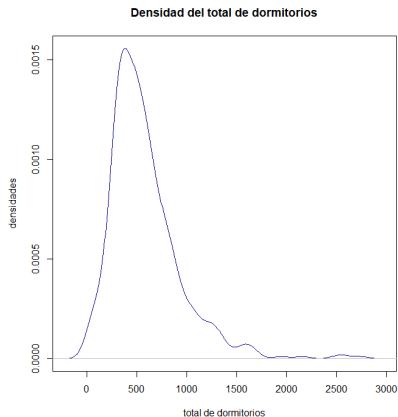
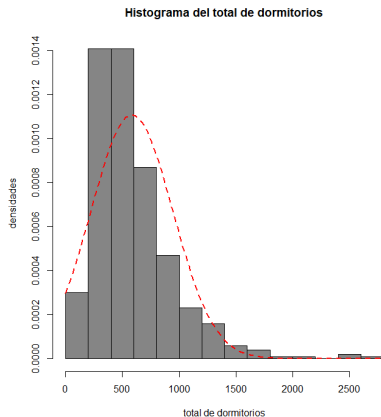


Figura: Gráfico de cajas para la variable 'Total de habitaciones'

# Variable 'Total de dormitorios'



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

# Variable 'Total de dormitorios'

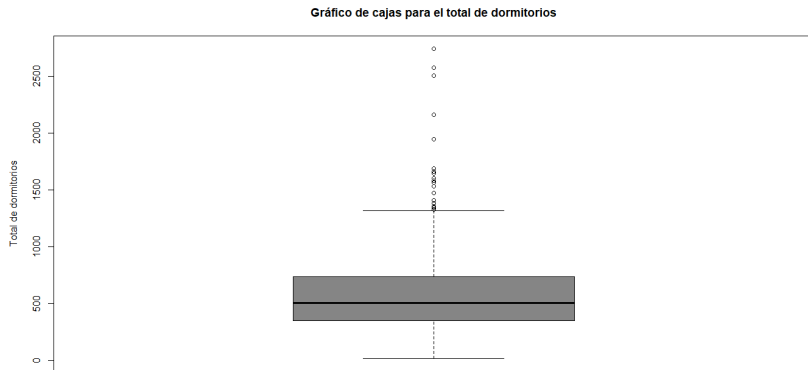


Figura: Gráfico de cajas para la variable 'Total de dormitorios'

# Variable 'Población'

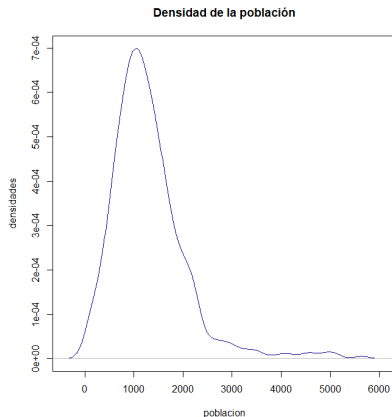
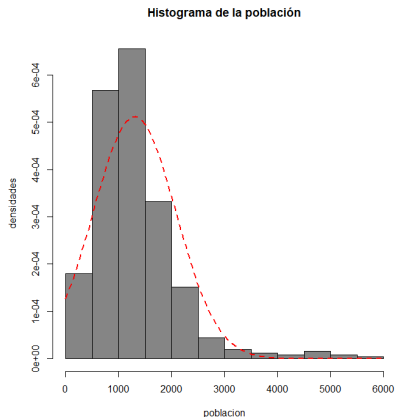


Figura: Histograma y densidad de la variable 'Población'

# Variable 'Población'

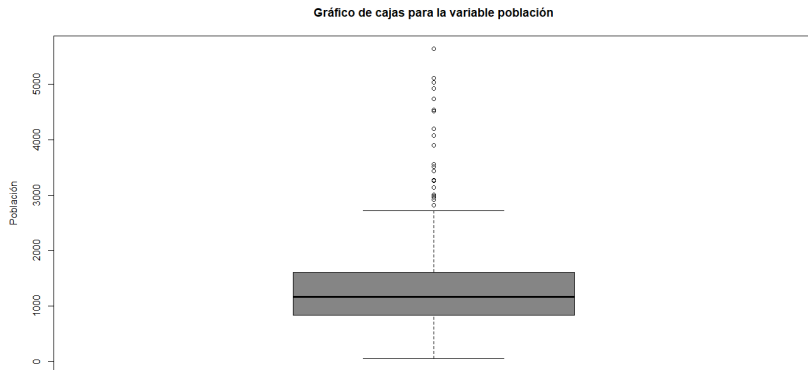
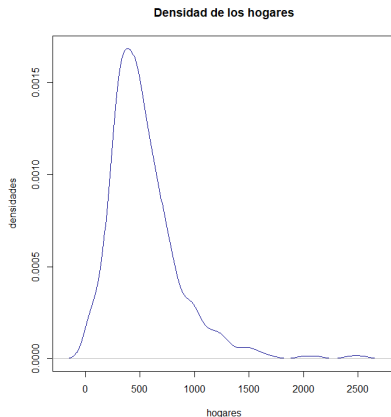
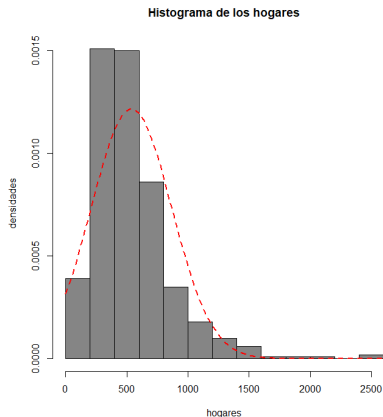


Figura: Gráfico de cajas para la variable 'Población'



# Variable 'Hogares'



**Figura:** Histograma y densidad de la variable 'Hogares'

# Variable 'Hogares'

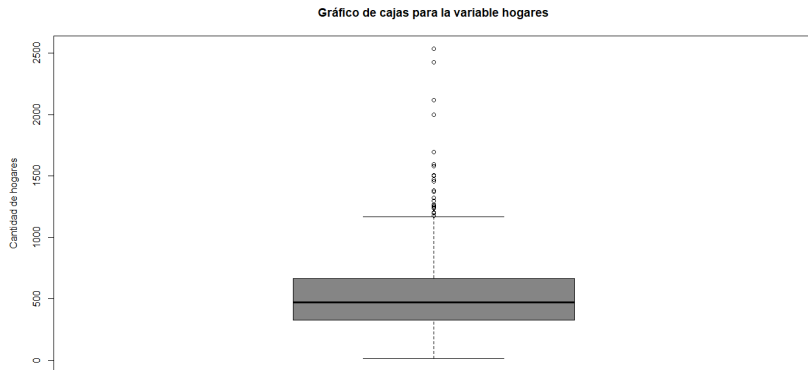


Figura: Gráfico de cajas para la variable 'Hogares'

# San Francisco

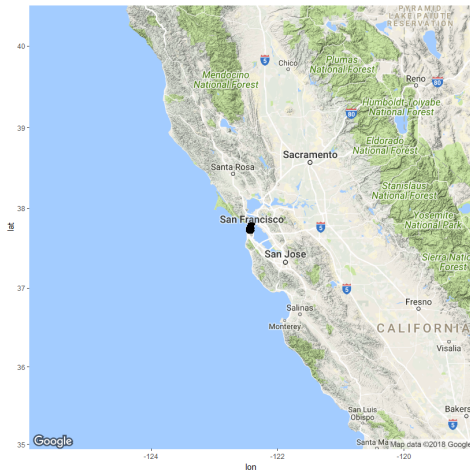


Figura: Mapa con los puntos dados de longitud y latitud

# San Diego

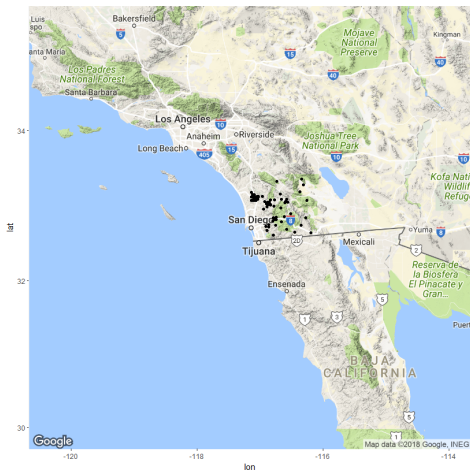


Figura: Mapa con los puntos dados de longitud y latitud

# Posibles relaciones entre las variables explicativas

- Variables 'Población' y 'Hogares':

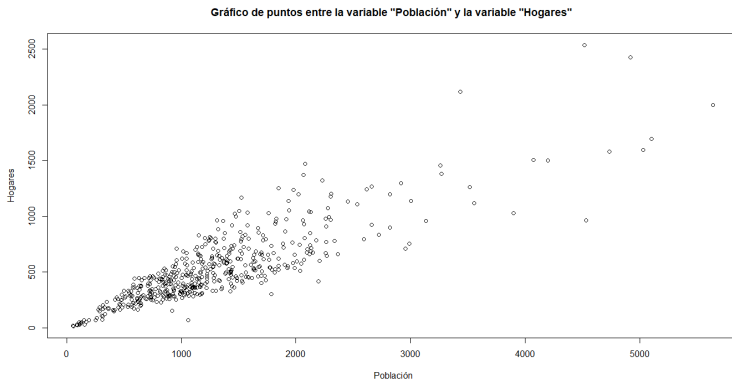


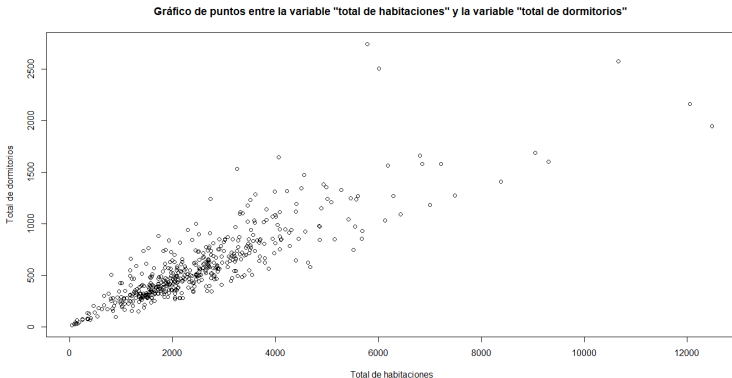
Figura: Gráfico de puntos entre las variables 'Población' y 'Hogares'

# 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 incluyendo todas las variables sin transformación y sin selección de variables 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 mediano de la casa,  $X_1$  =Ingreso mediano,  $X_2$  =Edad mediana 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

- $\beta_0$ : Representa el valor del intercepto con el eje y, esto quiere decir que -4113000 siempre va a afectar mi variable dependiente sin importar en cuantas unidades aumente mis otras variables.
- $\beta_1$ : por cada unidad que aumente de ingreso mediano se aumentan 22100 unidades del valor mediano de la casa.
- $\beta_2$ : por cada unidad que aumente de edad mediana de la vivienda se aumentan 1835 unidades del valor mediano de la casa.
- $\beta_3$ : por cada unidad que aumente de total de habitaciones se aumentan 30,35 unidades del valor mediano de la casa.
- $\beta_4$ : por cada unidad que aumente de total de dormitorios se disminuye 14,43 unidades del valor mediano de la casa.

- $\beta_5$ : por cada unidad que aumente de población se disminuye 124,5 unidades del valor mediano de la casa.
- $\beta_6$ : por cada unidad que aumente de hogares se disminuye 210,2 unidades del valor mediano de la casa.
- $\beta_7$ : por cada unidad que aumente de latitud se disminuye 33230 unidades del valor mediano de la casa.
- $\beta_8$ : por cada unidad que aumente de longitud se disminuye 45080 unidades del valor mediano de la casa.