Tarea 3

Jairo Enrique Alba

28/3/2023

Con base en los datos de ofertas de vivienda descargadas del portal finca raiz para apartamento de estrato 4 con área construida menor a 200 (vivienda4.RDS) la inmobiliaria A&C requiere el apoyo en la construcción de un modelo que lo oriente sobre los precios de inmuebles, por lo tanto se realiza un análisis descriptivo correspondiente al precio y al área construida tambien se analiza el tipo de vivienda y la zona donde se encuentra ubicada.

También se establecerá un modelo lineal para predecir el precio del inmueble teniendo en cuenta el área construida, es decir el precio sera la variable respuesta y el área construida es la variable predictora, en otras palabras $y = \beta_0 + \beta_1 x + \epsilon$.

Análisis Exploratorio de Variables

- **Análisis Exploratorio de la Variable Precio de la Vivienda**

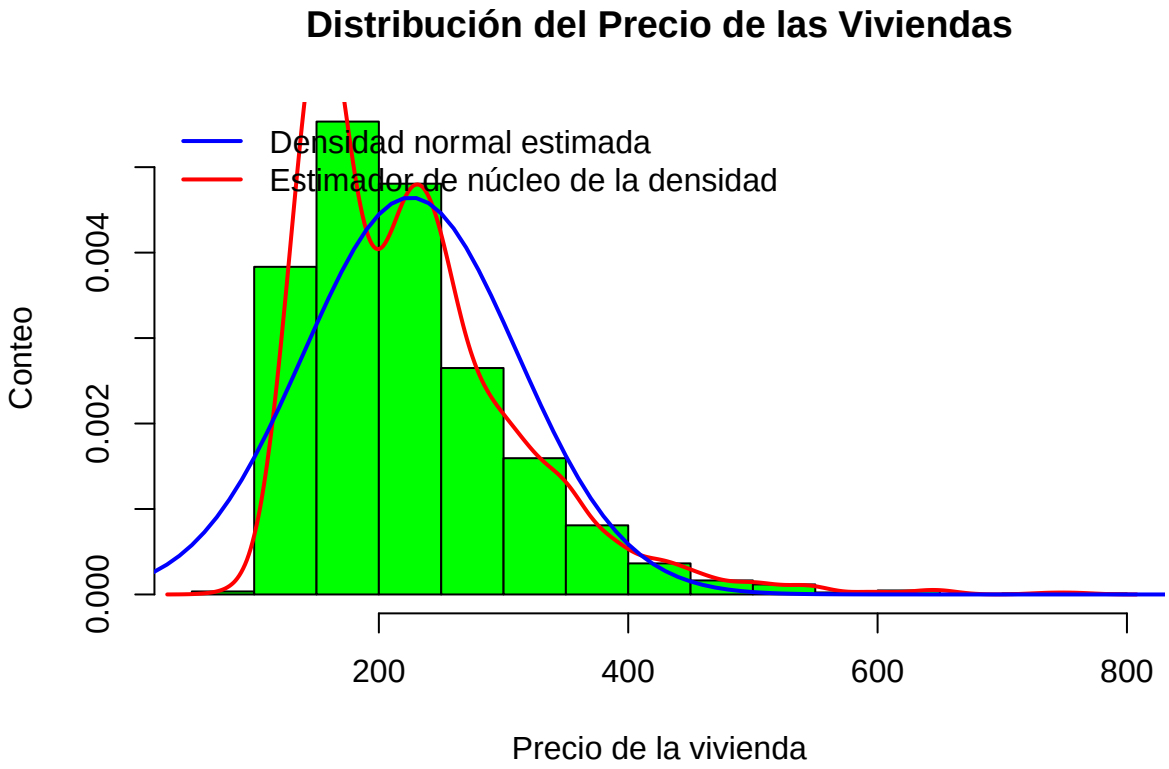
```
## [1] " El analisis exploratorio de datos nos arrojo los siguientes resultados:"
```

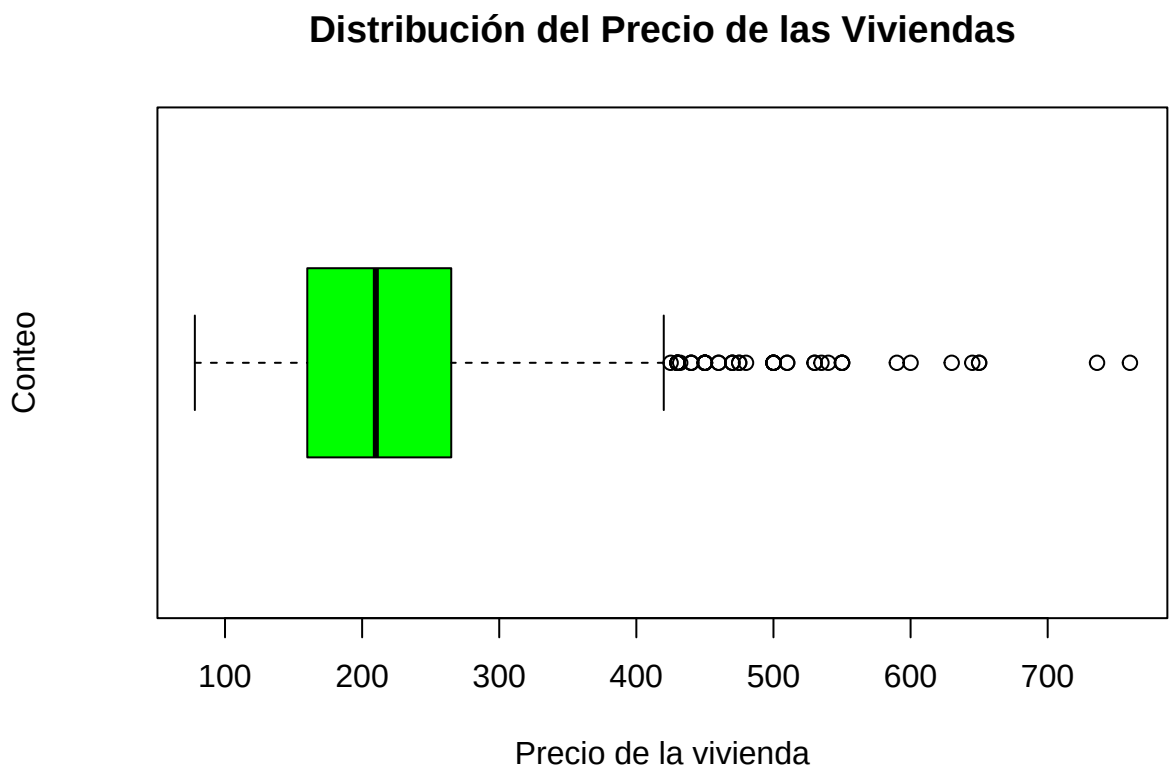
```
##     Minimo  Q1  Q2 Promedio Mediana  Q3 Maximo Coef_apertura varianza
## 25%     78 160 210 225.3746     210 265    760       9.74359 7376.274
##     Desv.Estandar Coef.Variacion Coef.Asimetria Coef.Curtosis
## 25%      85.88524          38.11       1.490777      3.573111
```

Estadístico	Valor
Mínimo	78
Cuartil 1	160
Cuartil 2 = Mediana	210
Promedio	255.3746
Cuartil 3	265
Máximo	760
Varianza	7376.274
Desviación estándar	85.8853
Coeficiente de Variación	38.11%
Coeficiente de asimetría	1.4908
Curtosis	3.5731

De las medidas encontradas evidenciamos un precio mínimo y maximo de las casas de 78 y 760 respectivamente con un promedio de precio de 255.38, además se observa que el 25%(427) de las casa presentaron un valor entre 78 y 160, mientras que el 50%(853) de las mismas tenian un precio entre 78 y 210, finalmente el 75%(1280) de las casas presentaron valores entre 78 y 265. Con lo descrito anteriormente se puede inferir que los precios de las casas se encuentran sesgados de manera positiva y lo podemos validar con el coeficiente de

asimetría 1.49, por otro lado la curtosis nos informa que la distribución es leptocúrtica 3,5731. Con respecto a la dispersión de los precios se tienen unos valores de 7376.274 y 85.89 para la varianza y desvición estándar.






```
## [1] " El límite inferior es:"

## 25%
## 2.5

## [1] " El bigote inferior es:"

## [1] 78

## [1] " El límite superior es:"

##   75%
## 422.5

## [1] " El bigote superior es:"

## [1] 422.5
```

Se observa que existen datos atípicos superiores, es decir aquellos precios superiores a 422.5, datos atípicos inferiores no se encontraron.

A continuación analizaremos la normalidad de la variable precio del inmueble:

```
library(normtest) ###REALIZA 5 PRUEBAS DE NORMALIDAD###
library(nortest) ###REALIZA 10 PRUEBAS DE NORMALIDAD###
library(moments) ###REALIZA 1 PRUEBA DE NORMALIDAD###
lillie.test(data$preciom)
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  data$preciom
## D = 0.098202, p-value < 2.2e-16
```

```
shapiro.test(data$preciom)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$preciom
## W = 0.8896, p-value < 2.2e-16
```

Las gráficas anteriores muestran una distribución sesgada positivamente, con la presencia de datos atípicos superiores los cuales los podemos evidenciar en el diagrama de cajas, dichos datos son aquellos precios de las casas superiores a 422.5, mediante la prueba de normalidad de Kolmogorov-Smirnov nos arrojo un p-valor de $2.2(10^{-16})$ confirmando que la distribución de los datos no es normal.

- **Análisis Exploratorio de la Variable área construida**

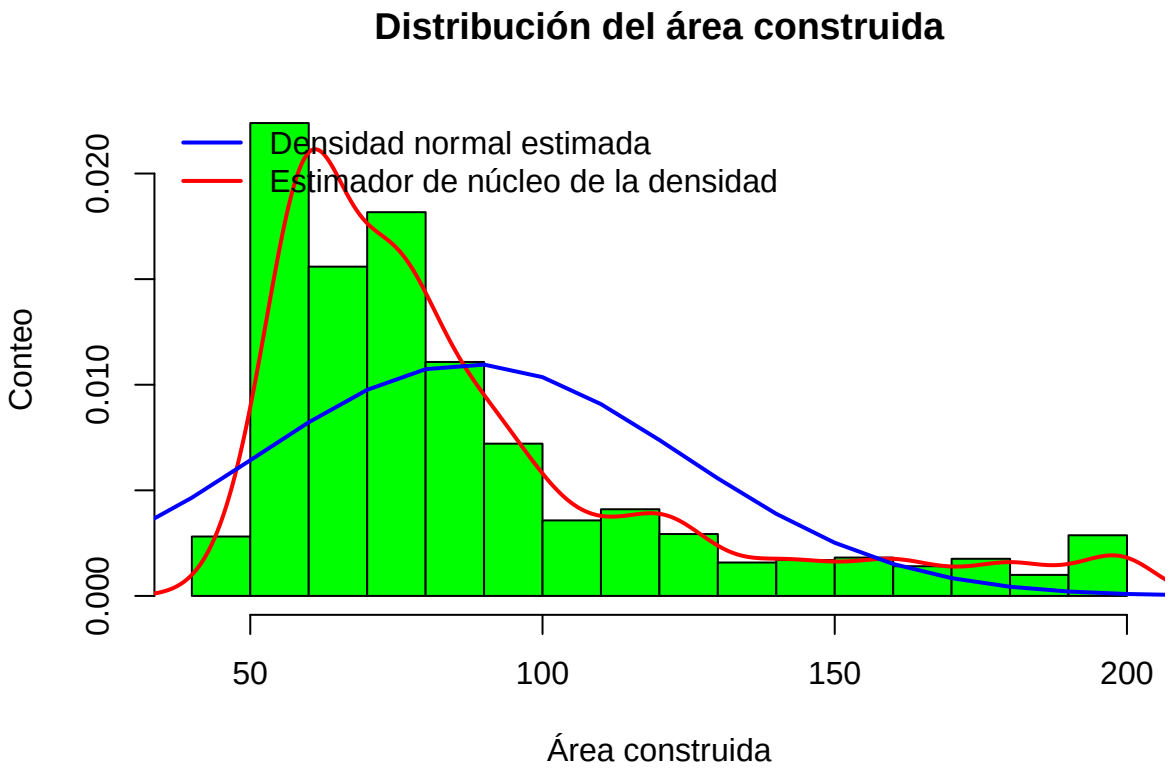
```
## [1] " El analisis exploratorio de datos nos arrojo los siguientes resultados:"
```

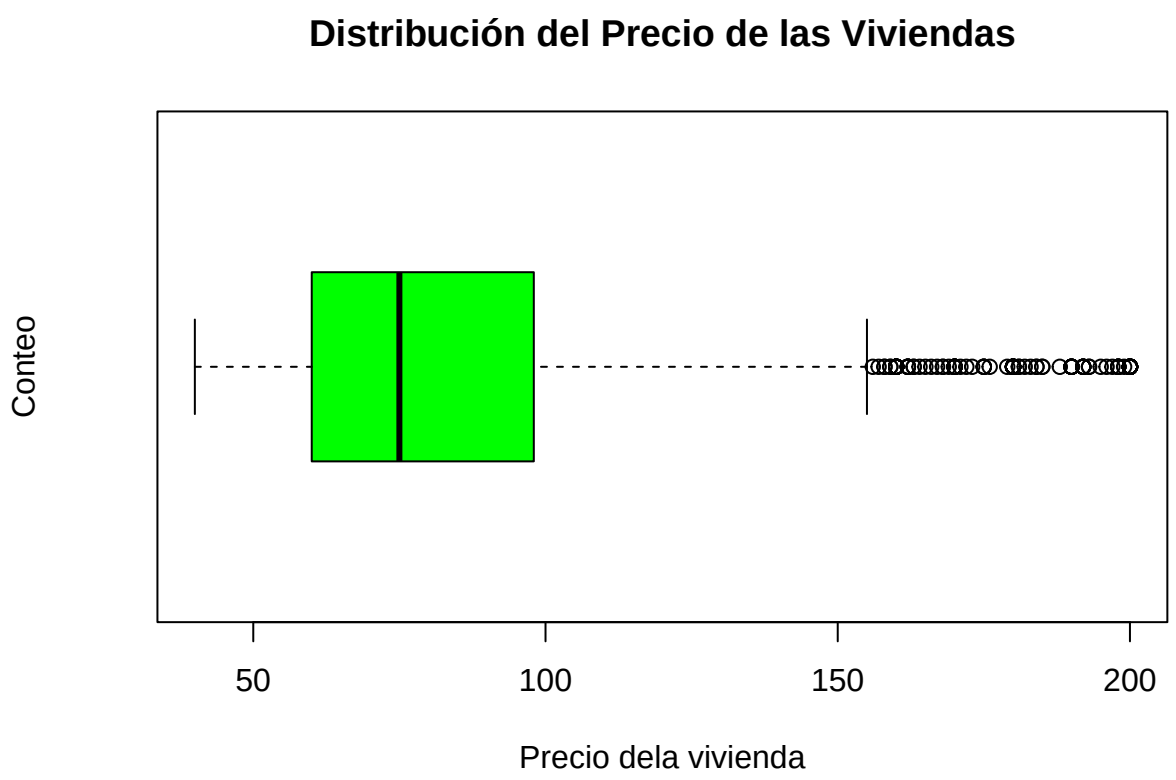
```
##     Minimoa Q1a Q2a Promedioa Medianaa Q3a Maximoa Coef_aperturaa varianzaa
## 25%      40  60  75  87.62954       75  98     200              5  1321.069
##     Desv.Estandara Coef.Variaciona Coef.Asimetriaa Coef.Curtosisa
## 25%       36.34651           41.48        1.532792       1.682735
```

Estadístico	Valor
Mínimo	40
Cuartil 1	60
Cuartil 2 = Mediana	75
Promedio	87.6295
Cuartil 3	98
Máximo	200
Varianza	1321.069
Desviación estándar	36.3465
Coeficiente de Variación	41.48%
Coeficiente de asimetría	1.533
Curtosis	1,683

De las medidas encontradas evidenciamos un área de construcción mínima y maxima de las casas de 40 y 60 respectivamente con un promedio de área de 87.6295, además se observa que el 25%(427) de las casa presentaron un área entre 40 y 60, mientras que el 50%(853) de las mismas tenian un área entre 40 y 75,

finalmente el 75%(1280) de las casas presentaron valores entre 40 y 98. Con lo descrito anteriormente se puede inferir que las áreas de las casas se encuentran sesgados de manera positiva y lo podemos validar con el coeficiente de asimetría 1.533, por otro lado la curtosis nos informa que la distribución es leptocúrtica 1,683. Con respecto a la dispersión de los precios se tienen unos valores de 1321.069 y 36.3465 para la varianza y desvición estándar.






```
## [1] " El límite inferior es:"

## 25%
##   3

## [1] " El bigote inferior es:"

## [1] 40

## [1] " El límite superior es:"

## 75%
## 155

## [1] " El bigote superior es:"

## [1] 155
```

Se observa que existen datos atípicos superiores, es decir aquellos superiore superiores a 155, datos atípicos inferiores no se encontraron.

A continuación analizaremos la normalidad de la variable área construida del inmueble:

```
library(normtest) ###REALIZA 5 PRUEBAS DE NORMALIDAD###
library(nortest) ###REALIZA 10 PRUEBAS DE NORMALIDAD###
library(moments) ###REALIZA 1 PRUEBA DE NORMALIDAD###
lillie.test(data$areaconst)
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  data$areaconst
## D = 0.17447, p-value < 2.2e-16
```

```
shapiro.test(data$areaconst)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$areaconst
## W = 0.8168, p-value < 2.2e-16
```

Las gráficas anteriores muestran una distribución sesgada positivamente, con la presencia de datos atípicos superiores los cuales los podemos evidenciar en el diagrama de cajas, dichos datos son aquellos precios de las casas superiores a 155, mediante la prueba de normalidad de Kolmogorov-Smirnov nos arrojo un p-valor de $2.2(10^{-16})$ confirmando que la distribución de los datos no es normal.

Análisis por Zona y tipo de vivienda

De los datos observados se evidencia que el 79% corresponde a la zona sur, 17% corresponde a la zona norte, el 4% zona oeste, el 0.5% zona centro y el 0.4%.

Zona	Frecuencia	Porcentaje
Sur	1344	79%
Norte	288	17%
Oeste	60	4%
Centro	8	0.05 %
Oriente	6	0.04

En cuanto al tipo de vivienda el 80% corresponde a apartamentos y 20% a casas.

Tipo	Frecuencia	Porcentaje
Apartamento	1363	80%
Casa	343	20%

Para un mejor modelo excluimos los datos atípicos

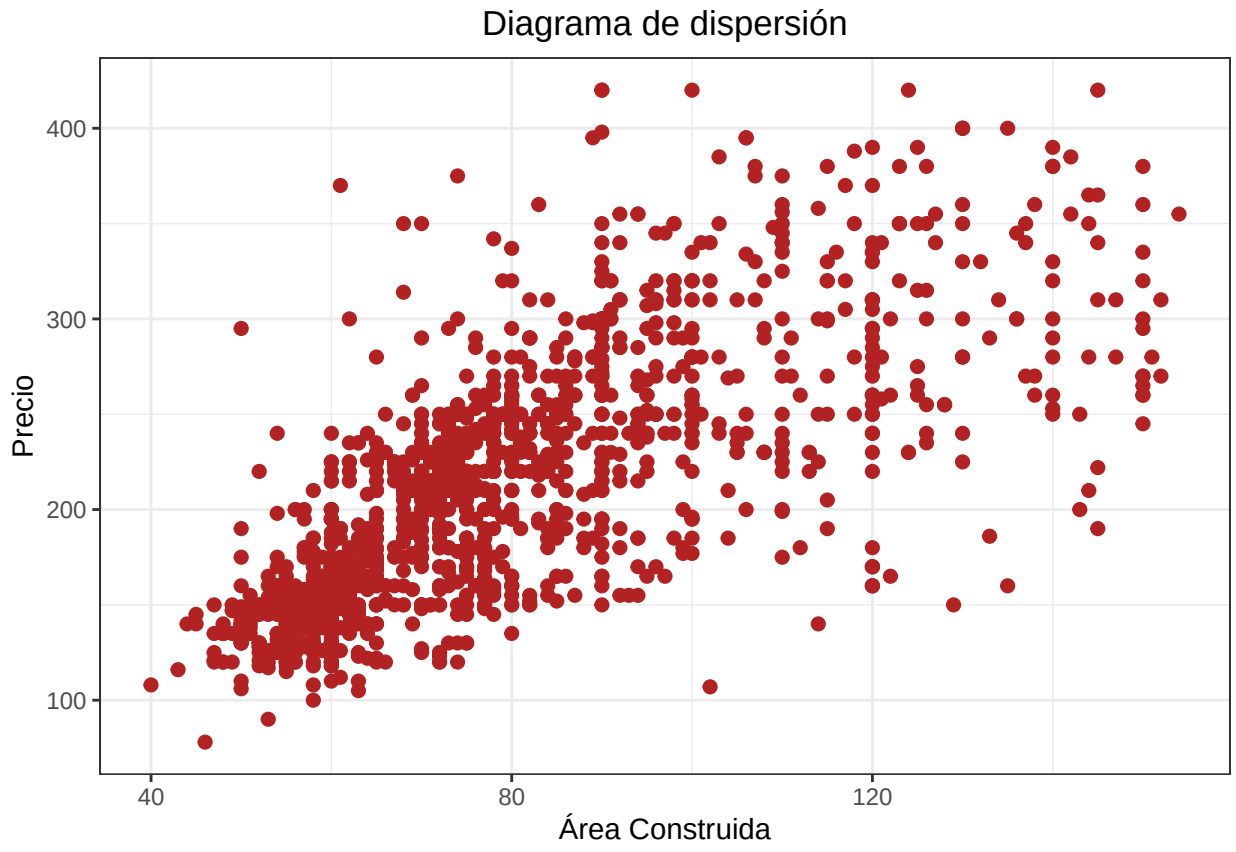
Para mejorar el modelo excluimos los datos atípicos

```r
# Datos sin datos atípicos

data1=subset(data, (data$preciom < 422.5  & data$areaconst < 155))

# pairs(data1) # para hacer gráficas cruzando las variables
```

Análisis Exploratorio Bivariado

Nube de Puntos o gráfica de dispersión

A continuación elaboramos la gráfica de puntos o de dispersión para análizar de manera gráfica si existe algún tipo de correlación entre la variable área construida y el precio de la casa.



De la gráfica anterior podemos evidenciar una correlación lineal positiva (directa) entre las variables **área construida** como variable predictora y la variable **precio de la vivienda** como variable respuesta, es decir a mayor área construida mayor será el precio.

```
##
##  Pearson's product-moment correlation
##
## data:  data1$areaconst and data1$preciom
## t = 41.902, df = 1543, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7053494 0.7520650
```

```
## sample estimates:
##       cor
## 0.7295571
```

De la salida anterior podemos concluir que:

- Visualizando los gráficos de dispersión podemos observar que la variable areaconst (área construida) está linealmente asociada con la variable respuesta preciom (precio de la vivienda), por lo que utilizaremos un modelo lineal.
- El coeficiente de correlación de Pearson es alta (r = 0.729557) y significativo (p-value = 2.2e-16). Ello indica una correlación entre ambas variables alta. Lo cual lo verificamos con el intervalo de confianza para dicho coeficiente (0.7053494, 0.7520650)
- Por lo tanto tiene sentido generar el modelo de regresión lineal ya que se cumplen los primeros requisitos.

Elaboración del modelo de regresión lineal simple

A continuación estableceremos un modelo lineal simple para predecir el precio de las casas teniendo en cuenta el área construida, de la forma

$$y = \beta_1 x + \beta_0 + \epsilon$$

donde x será el área de la casa e y el precio.

Veamos,

```
##
## Call:
## lm(formula = preciom ~ areaconst, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -165.85  -25.99   -5.80   26.99  197.13
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 46.76231    4.03417   11.59   <2e-16 ***
## areaconst    2.06729    0.04934   41.90   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 44.35 on 1543 degrees of freedom
## Multiple R-squared:  0.5323, Adjusted R-squared:  0.532
## F-statistic:  1756 on 1 and 1543 DF,  p-value: < 2.2e-16
```

Resumen del modelo

El resumen del modelo presenta los errores estándar, el valor del estadístico t y el correspondiente p-valor de los parámetros $\hat{\beta_0}$ y $\hat{\beta_1}$. El p-valor nos permite dterminar si los estimadores de los parámetros son significativamente distintos de cero, es decir que contribuyen al modelo. El parámetro $\hat{\beta_1}$ correspondiente a la pendiente suele ser el más util de estos modelos.

De los resultados obtenidos podemos concluir los siguiente:

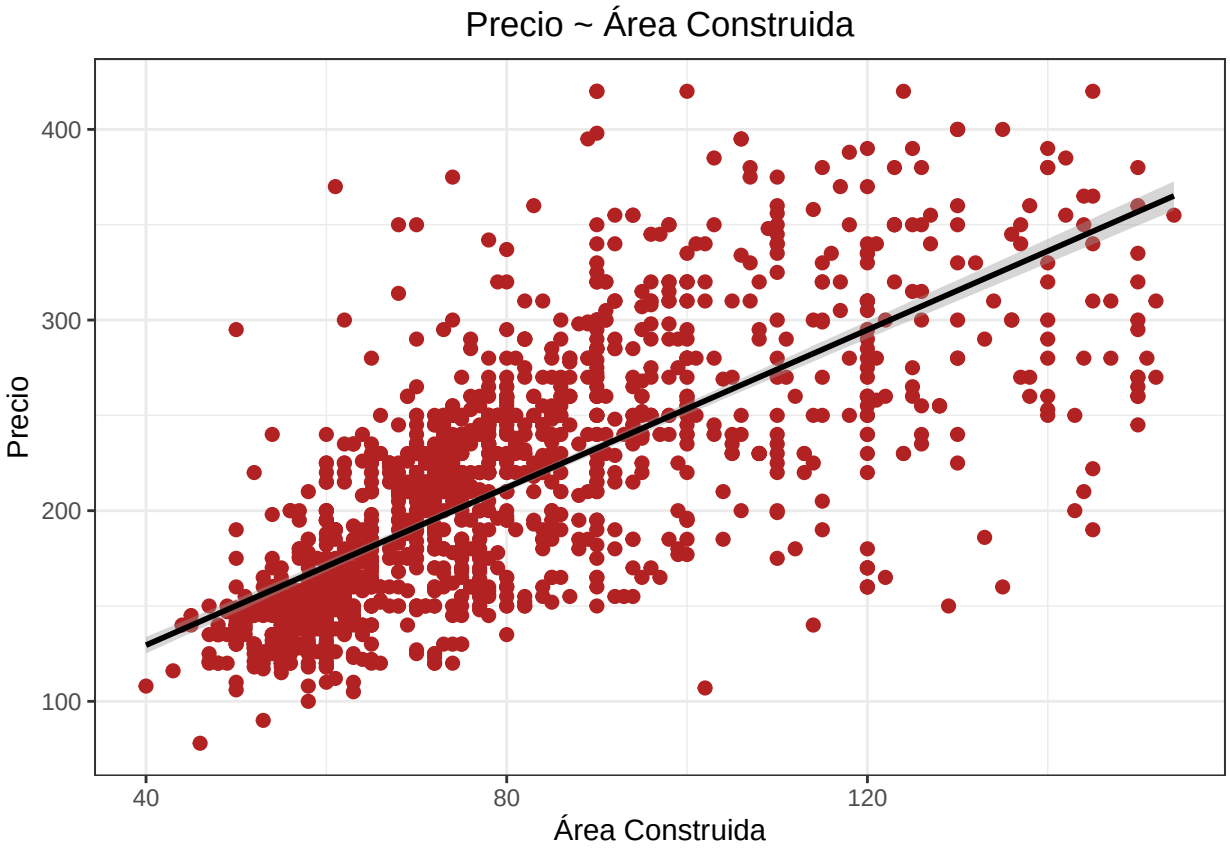
- Tanto el parámetro $\hat{\beta}_0$ correspondiente al corte con el eje y como $\hat{\beta}_1$ correspondiente a la pendiente son significativos, ya que los p-valores fueron $2.2(10^{-16})$
- El coeficiente de determinación R^2 indica que el modelo es capaz de explicar el 53% de la variabilidad presente en la variable respuesta (precio) mediante la variable predictora (área construida).
- El p-valor obtenido en el test F es $2.2(10^{-16})$ el cual determina que es significativamente superior la varianza explicada por el modelo en comparación con la varianza total, por lo tanto se puede aceptar el modelo como útil y válido.
- Ecuación del modelo estará determinada por: $Precio = 2.067(\acute{a}rea\ construida) + 46.76 + \epsilon$, es decir por cada unidad que se incrementa en el área, el precio aumenta 2.067 unidades

Intervalos de confianza para los parámetros del modelo

```
##                 2.5 %    97.5 %
## (Intercept) 38.849267 54.675346
## areaconst    1.970521  2.164067
```

Los intervalos de confianza para los parámetros $\hat{\beta}_0$ y $\hat{\beta}_1$son respectivamente (38.85, 54.68) y (1.97, 2.16), lo cual nos confirma la estimación que no incluyen el cero.

Representación gráfica del modelo



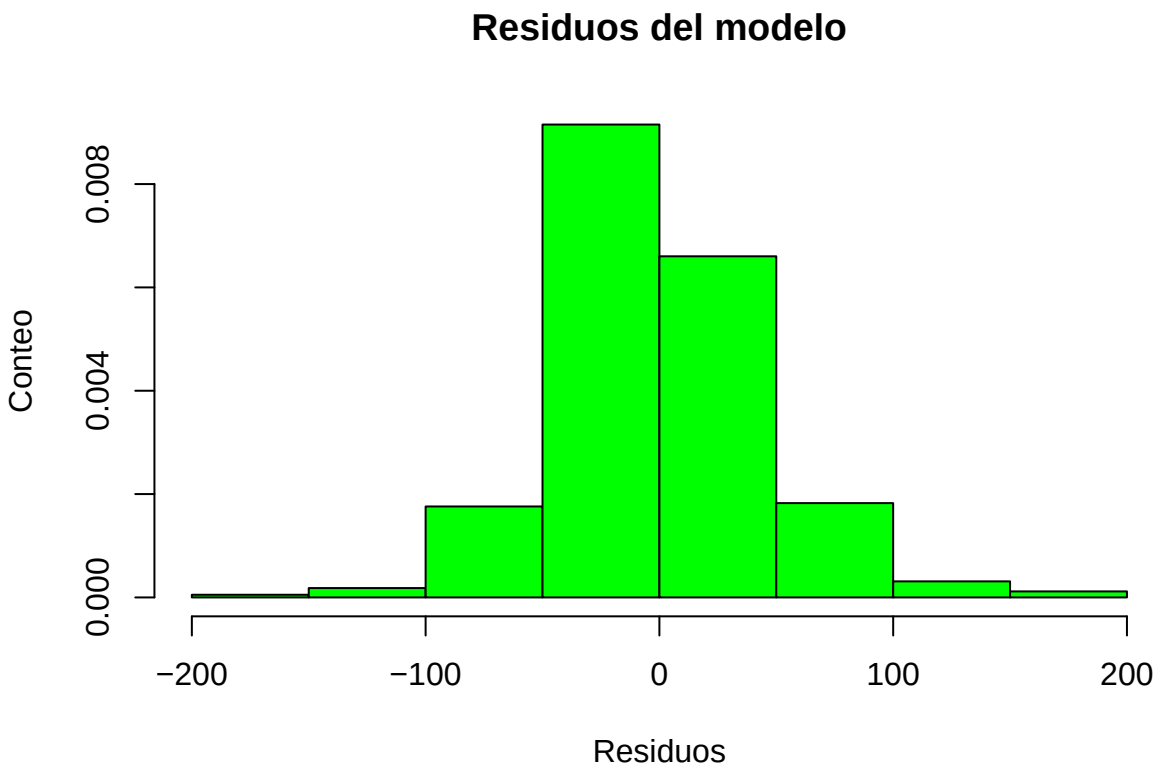
Verificación de las condiciones para aceptar el modelo

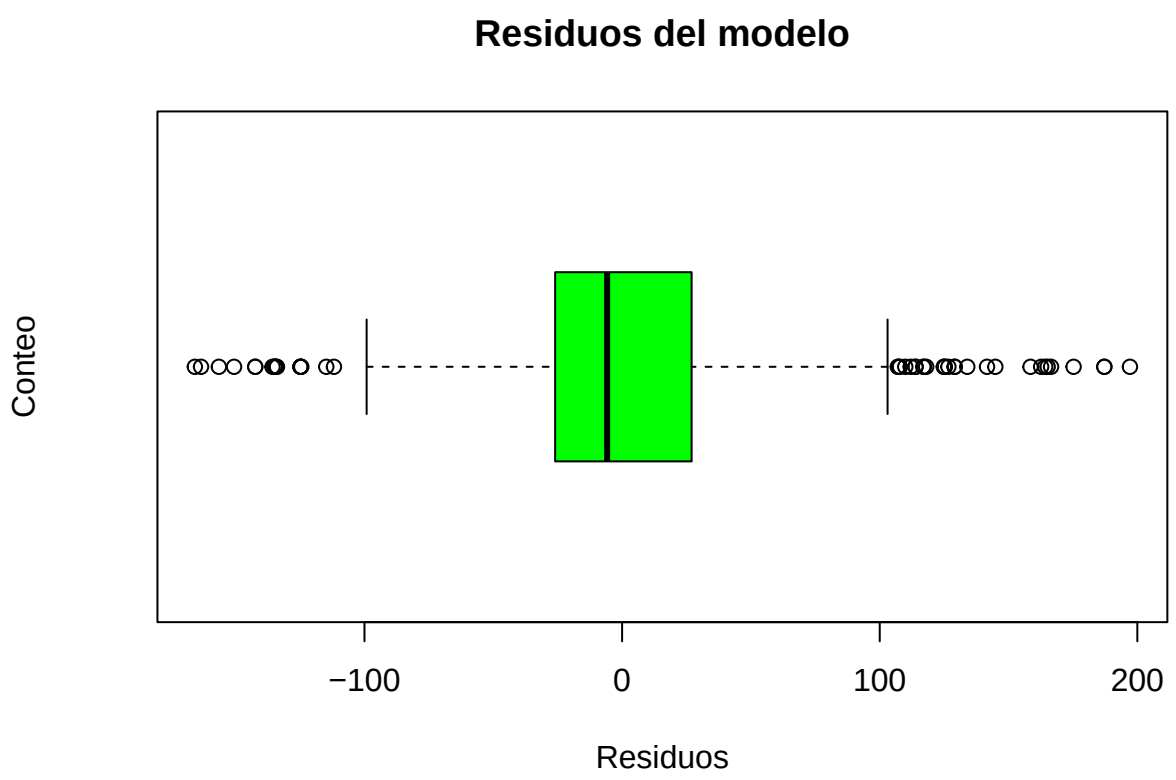
Para verificar los supuestos que debe cumplir los residuos para poder aplicar la teoría de tal forma que el modelo sea válido y confiable, se debe tener en cuenta los siguiente:

- Análisis de los residuos (distribución, variabilidad. . .)(plot(modelo))
- Test de hipótesis de Shapiro Wilk para el análisis de normalidad(shapiro.test(modelo$residuals))
- Test de contraste de homocedasticidad Breusch-Pagan (bptest(modelo))
- Detección de observaciones influyentes (influence.measures(modelo))
- Visualización de observaciones influyentes (influencePlot(modelo))
- Test de detección de outliers (outlierTest(modelo))
- Cálculo de residuos estudentizados (rstudent(modelo))

Normalidad de los residuos

```
##
##  Shapiro-Wilk normality test
##
## data:  modelo.lineal$residuals
## W = 0.97687, p-value = 4.558e-15
```

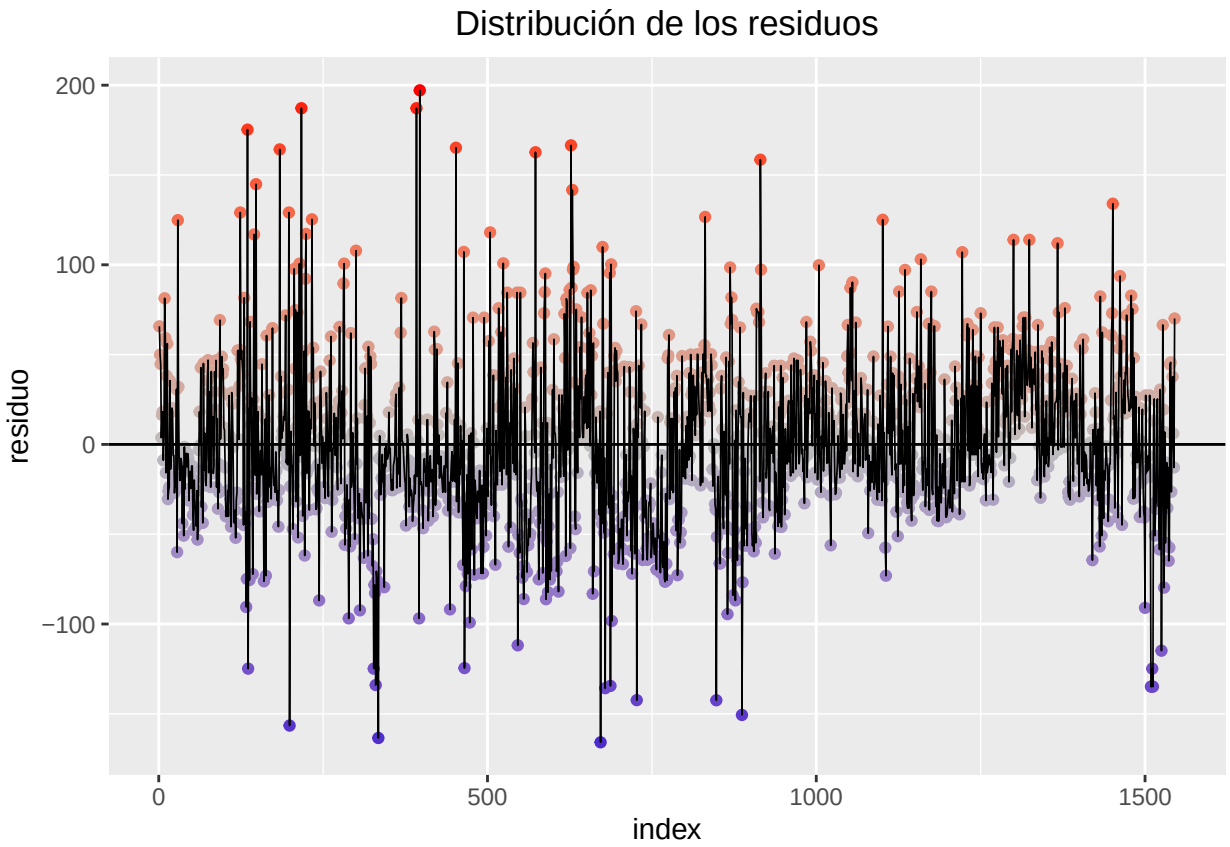




Observamos que los residuos para este modelo no se ditribuyen de manera normal ($p-valor = 4.558(10^{-15})$) también lo observamos a través del histograma y el diagrama de cajas.

Homocedasticidad de los residuos

```
##
##  studentized Breusch-Pagan test
##
## data:  modelo.lineal
## BP = 165.51, df = 1, p-value < 2.2e-16
```



La condición de homocedasticidad (supuesto de varianza constante) parece no cumplirse ya que p-valor es 2.2e-16.

En resumen, la normalidad de los residuos parece que no podemos aceptarla, y tampoco parecen seguir una clara tendencia según el orden de registro de las observaciones, tampoco la condición de homocedasticidad parece no cumplirse.

Sin embargo al observar algunos gráficos podríamos sospechar que algunas observacionesque pueden estar influyendo al modelo. Para analizar en qué medida pueda estar influyendo esta u otras observaciones, se **reajustará** el modelo excluyendo posibles observaciones sospechosas.

Dependiendo de la finalidad del modelo, la exclusión de posibles outliers debe analizarse con detalles, ya que estas observaciones podrían ser errores de medida, pero también podrían representar casos interesantes.

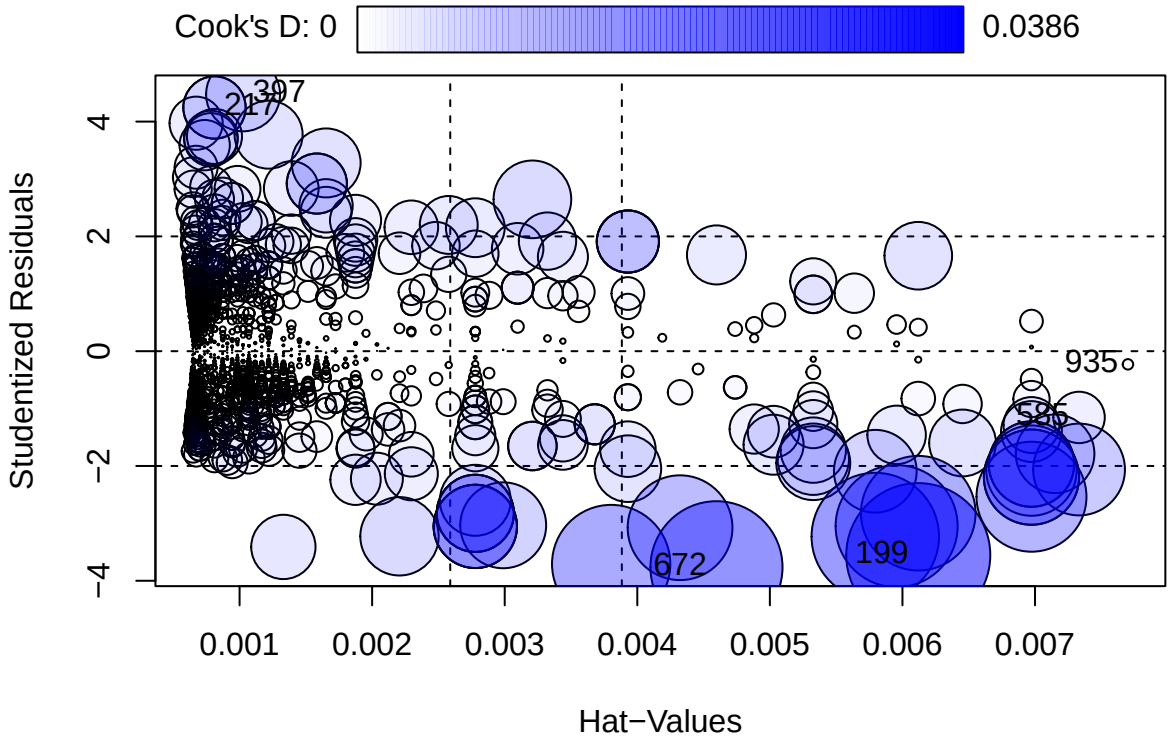
También podrian utilizarse otros tipos de modelos tales como: lineal-logarítmico, logarítmico-lineal, logarítmico-logarítmico.

Primero excluyamos los datos influyentes y ajustemos el modelo:

Datos que estan influenciando en el modelo

```
##  135  148  184  199  217  330  334  392  397  452  573  627  629  672  679  687
##  135  148  184  199  217  330  334  392  397  452  573  627  629  672  679  687
##  727  848  887  915 1451 1509 1510 1512
##  727  848  887  915 1451 1509 1510 1512
```

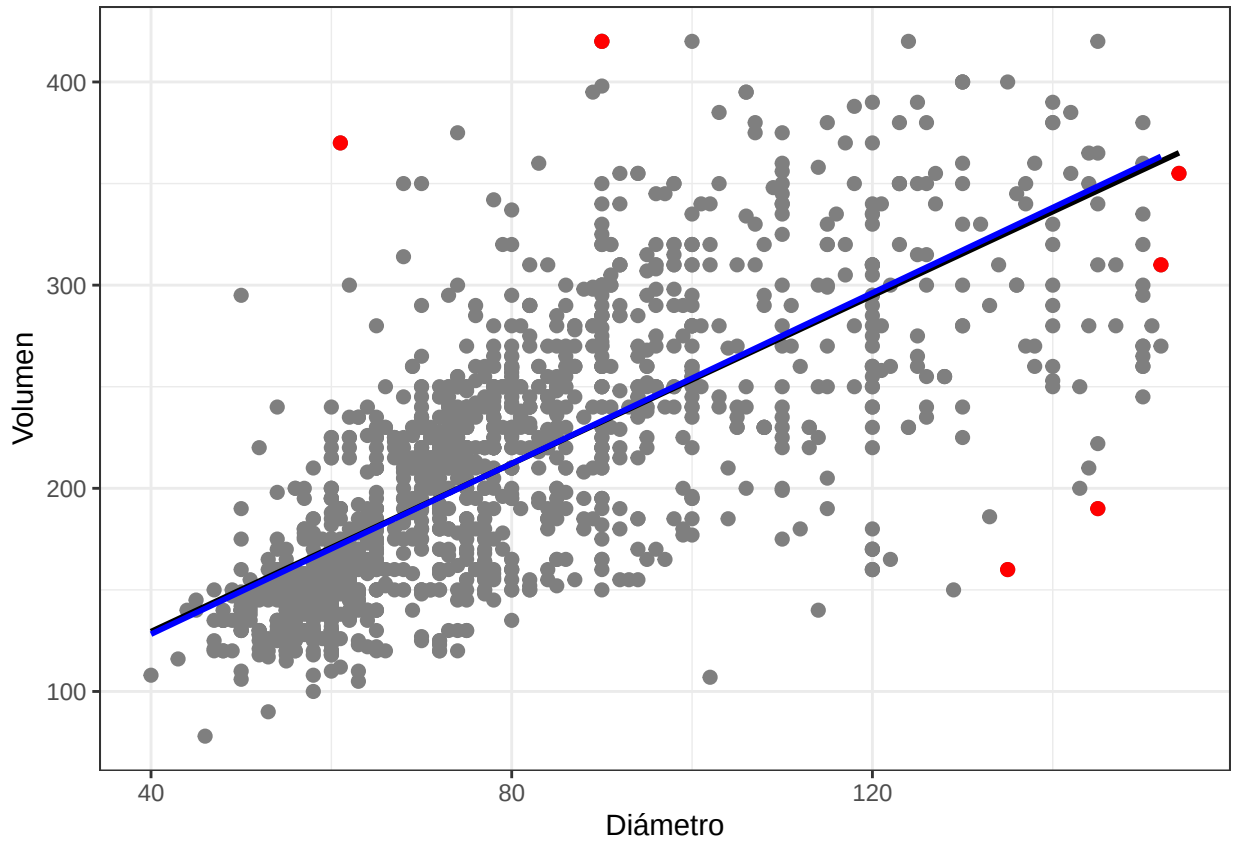
Valores con mayor influencia en el modelo son:




```
##        StudRes          Hat        CookD
## 199 -3.5535802 0.0061194259 0.0385849711
## 217  4.2457524 0.0008107531 0.0072335953
## 397  4.4748240 0.0010265453 0.0101630553
## 585 -1.1541749 0.0073322143 0.0049187078
## 672 -3.7643245 0.0045972710 0.0324454833
## 935 -0.2291391 0.0077010030 0.0002038638
```

En este análisis de los residuos estudentizados se logra detectar observaciones atípicas, la observaciones 199, 217,397, 585, 672, 935 parece estar influenciando en gran medida al modelo.

Procederemos a reajustar el modelo excluyendo dichas observaciones. Veamoslo en la siguiente gráfica:



Realizamos nuevamente el modelo lineal excluyendo aquellos valores que estaban influenciando en el modelo.

Modelo 2: excluyendo los valores más influyentes

```
##
## Call:
## lm(formula = preciom ~ areaconst, data = data1[c(-199, -217,
##     -397, -585, -672, -935), ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -165.086  -25.453   -5.205   26.410  186.803
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 44.22062    3.99148   11.08   <2e-16 ***
## areaconst    2.09973    0.04895   42.89   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43.48 on 1537 degrees of freedom
## Multiple R-squared:  0.5448, Adjusted R-squared:  0.5446
## F-statistic:  1840 on 1 and 1537 DF,  p-value: < 2.2e-16

##
```

```
##  Shapiro-Wilk normality test
##
## data:  modelo.lineal2$residuals
## W = 0.98156, p-value = 3.751e-13
```

Modelo 3: excluyendo todos los valores influyentes

```
##
## Call:
## lm(formula = preciom ~ areaconst, data = data1[c(-135, -148,
##     -184, -199, -217, -330, -334, -392, -397, -452, -573, -627,
##     -629, -672, -679, -687, -727, -848, -887, -915, -1451, -1509,
##     -1510, -1512), ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -131.591  -24.203   -3.937   26.047  127.924
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 38.14001    3.70381   10.30   <2e-16 ***
## areaconst    2.17552    0.04555   47.76   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 40.1 on 1519 degrees of freedom
## Multiple R-squared:  0.6003, Adjusted R-squared:    0.6
## F-statistic:  2281 on 1 and 1519 DF,  p-value: < 2.2e-16
```

```
##
##  Shapiro-Wilk normality test
##
## data:  modelo.lineal3$residuals
## W = 0.99326, p-value = 2.054e-06
```

Modelo 4: Modelo Lineal -logarítmico

```
##
## Call:
## lm(formula = preciom ~ log(areaconst), data = data1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -156.648  -22.897   -2.394   22.431  200.067
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -579.440     17.566  -32.99   <2e-16 ***
## log(areaconst)  182.290      4.053   44.97   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 42.66 on 1543 degrees of freedom
## Multiple R-squared:  0.5673, Adjusted R-squared:  0.567
## F-statistic:  2023 on 1 and 1543 DF,  p-value: < 2.2e-16
```

```
##
##  Shapiro-Wilk normality test
##
## data:  modelo.lineal4$residuals
## W = 0.97699, p-value = 5.066e-15
```

Modelo 5: Modelo Logaritmo -lineal

```
##
## Call:
## lm(formula = log(preciom) ~ areaconst, data = data1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.84704 -0.13303 -0.01285  0.15161  0.78257
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.5522588  0.0188297   241.8   <2e-16 ***
## areaconst   0.0094864  0.0002303    41.2   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.207 on 1543 degrees of freedom
## Multiple R-squared:  0.5238, Adjusted R-squared:  0.5235
## F-statistic:  1697 on 1 and 1543 DF,  p-value: < 2.2e-16
```

```
##
##  Shapiro-Wilk normality test
##
## data:  modelo.lineal5$residuals
## W = 0.9941, p-value = 8.142e-06
```

Modelo 6: Modelo Logaritmo - Logaritmo

```
##
## Call:
## lm(formula = log(preciom) ~ log(areaconst), data = data1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.87844 -0.12158  0.00196  0.13474  0.79874
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     1.62438    0.08051   20.18   <2e-16 ***
## log(areaconst)  0.84906    0.01858   45.70   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1955 on 1543 degrees of freedom
## Multiple R-squared:  0.5751, Adjusted R-squared:  0.5749
## F-statistic:  2089 on 1 and 1543 DF,  p-value: < 2.2e-16
```

```
##
##  Shapiro-Wilk normality test
##
## data:  modelo.lineal6$residuals
## W = 0.99213, p-value = 2.376e-07
```

Resumenes de los tistintos modelos construidos

Modelo	$\hat{\beta_0}$	$\hat{\beta_1}$	R^2	p-normalidad residuos	Modelo
Modelo 1	46.76	2.067	0.53	$4.558(10^{-15})$	$y = 46.76 + 2.067x$
Modelo 2	44.22	2.099	0.55	$3.175(10^{-13})$	$y = 44.22 + 2.099x$
Modelo 3	38.14	2.176	0.6	$2.054(10^{-6})$	$y = 38.14 + 2.176x$
Modelo 4	-579.44	182.29	0.57	$5.066(10^{-15})$	$y = -579.44 + 182.29log(x)$
Modelo 5	4.55	0.0095	0.52	$8.142(10^{-6})$	$log(y) = 4.55 + 0.0095x$
Modelo 6	1.6244	0.849	0.58	$2.376(10^{-7})$	$log(y) = 0.849 + 0.58log(x)$

Del resumen anterior se evidencia que el mejor modelo es el 3, $R^2 = 0.6$, sin embargo los residuos no se comportan de manera normal, parece que la mejor opción es buscar más datos que afectan el modelo y excluirlos, para buscar la satisfacción de los supuestos del modelo. Sin embargo utilizaremos el modelo 6 para para predecir algunas observaciones.

Uso del modelo 3 para predecir nuevas observaciones

```
# Precio PROMEDIO que esperaríamos del precio de una casa de 110 metros de área construida.
predict(modelo.lineal3, data.frame(areaconst = 110), interval = "confidence")
```

```
##        fit      lwr      upr
## 1 277.4476 273.9577 280.9375
```

```
# Precio esperado  del precio de una casa de 110 metros de área construida
predict(modelo.lineal3, data.frame(areaconst = 110), interval = "prediction")
```

```
##        fit      lwr      upr
## 1 277.4476 198.7176 356.1775
```

Por lo tanto se espera que en promedio el precio de un inmueble de área de 110 sea de 277,45. Podemos afirmar con un 95% de confianza que el verdadero valor promedio se encuentra entre (273.96- 280.94), mientras que el intervalo de predicción para una solo inmueble de área este entre (198.72 - 356.18) y su valor predicho es 277.45, por lo tanto un precio de 200 millones parece una atractiva oferta, sin embargo se debe tener en cuenta el tipo de vivienda(apartamento o casa) y la zona donde se encuentra ubicado(sur, norte, oeste,centro u oriente) y en la medida de lo posible establecer un modelo para la zona y el tipo de inmueble, además se debe pedir asesoría de expertos en el negocio.

Informe para los directivos.

La información encontrada de la matriz de datos correspondiente a los datos de ofertas de vivienda descargadas del portal Fincaraiz para inmuebles de estrato 4 con área construida menor a $200m^2$ se evidenció un precio mínimo y maximo de los inmuebles de 78 y 760 respectivamente con un promedio de precio de 255.38, además se observa que el 25%(427) de las casa presentaron un valor entre 78 y 160, mientras que el 50%(853) de las mismas tenian un precio entre 78 y 210, finalmente el 75%(1280) de las casas presentaron valores entre 78 y 265. Con lo descrito anteriormente se puede inferir que los precios de las casas se encuentran sesgados de manera positiva y lo podemos validar con el coeficiente de asimetría 1.49, por otro lado la curtosis nos informa que la distribución es leptocúrtica 3,5731. Con respecto a la dispersión de los precios se tienen unos valores de 7376.274 y 85.89 para la varianza y desvición estándar.

Con respecto al areá construida se obtuvo las siguientes medida: la construcción mínima y maxima de los inmuebles fueron de 40 y 60 respectivamente con un promedio de área de 87.6295, además se observa que el 25%(427) de las casa presentaron un área entre 40 y 60, mientras que el 50%(853) de las mismas tenian un área entre 40 y 75, finalmente el 75%(1280) de las casas presentaron valores entre 40 y 98. Con lo descrito anteriormente se puede inferir que las áreas de las casas se encuentran sesgados de manera positiva y lo podemos validar con el coeficiente de asimetría 1.533, por otro lado la curtosis nos informa que la distribución es leptocúrtica 1,683. Con respecto a la dispersión de los precios se tienen unos valores de 1321.069 y 36.3465 para la varianza y desvición estándar.

Además se evidencia que el 79% corresponde a la zona sur, 17% corresponde a la zona norte, el 4% zona oeste, el 0.5% zona centro y el 0.4%.

Zona	Frecuencia	Porcentaje
Sur	1344	79%
Norte	288	17%
Oeste	60	4%
Centro	8	0.05 %
Oriente	6	0.04

En cuanto al tipo de vivienda el 80% corresponde a apartamentos y 20% a casas.

Tipo	Frecuencia	Porcentaje
Apartamento	1363	80%
Casa	343	20%

También de la construcción de un modelo lineal para predecir el precio del inmuble a partir del área construida se obtuvo e evidenció que el mejor modelo es $y = 38.14 + 2.1755x + \epsilon$, es decir $precio = 38.14 + 2.1755(\acute{a}rea\ construida) + \epsilon$, dicho modelo es capaz de explicar el 60% ($R^2 = 0.6$) de la variabilidad presente en la variable respuesta (precio) mediante la variable predictora (área construida). Sin embargo los residuos no se comportan de manera normal, parece que la mejor opción es buscar más datos que afectan el modelo y excluirlos, para buscar la satisfacción de los supuestos del modelo.

Para la utilización del modelo se recomienda tener en cuenta el tipo de inmueble y la ubicación del mismo y en la medida de lo posible establecer un modelo particularizando las condiciones del inmueble (zona y tipo de inmueble), además se debe pedir asesoría de expertos en el negocio.