EstimacionPrecioVivienda

AlbertoMarino

11 de diciembre de 2016

## Practica de estimación del precio de la vivienda

Tenemos doa objetivo básicos en la práctica: 1.- Analizar el efecto de la superficie de la vivienda en el precio de la vivienda 2.- Estimar el precio de venta de unos inmuebles de la cartera de la empresa.

1. En el primer caso nos interesa ver mas el comportamiento de las variables entre si, para versu relacion con el precio. Es una análisis mas estadístico (nos interesa una mayor explicacion de la relacion de las variables asi como de la variabilidad de los errores del modelo)
2. En el segundo caso nos interesa ser mas efectivos en la prediccion del precio, es un análisis mas enfocado al ML. Nos interesará mas el ajuste de los errores.

### Desarrollo

Inicializamos el dataset

data=read.csv("house\_train.csv",stringsAsFactors = FALSE)

Revisamos los datos y los estadisticos de cada feature del dataset

str(data)

## 'data.frame': 17384 obs. of 21 variables:  
## $ id : num 7.13e+09 6.41e+09 5.63e+09 2.49e+09 1.95e+09 ...  
## $ date : chr "20141013T000000" "20141209T000000" "20150225T000000" "20141209T000000" ...  
## $ price : num 221900 538000 180000 604000 510000 ...  
## $ bedrooms : int 3 3 2 4 3 4 3 3 3 3 ...  
## $ bathrooms : num 1 2.25 1 3 2 4.5 2.25 1.5 1 2.5 ...  
## $ sqft\_living : int 1180 2570 770 1960 1680 5420 1715 1060 1780 1890 ...  
## $ sqft\_lot : int 5650 7242 10000 5000 8080 101930 6819 9711 7470 6560 ...  
## $ floors : num 1 2 1 1 1 1 2 1 1 2 ...  
## $ waterfront : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ view : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ condition : int 3 3 3 5 3 3 3 3 3 3 ...  
## $ grade : int 7 7 6 7 8 11 7 7 7 7 ...  
## $ sqft\_above : int 1180 2170 770 1050 1680 3890 1715 1060 1050 1890 ...  
## $ sqft\_basement: int 0 400 0 910 0 1530 0 0 730 0 ...  
## $ yr\_built : int 1955 1951 1933 1965 1987 2001 1995 1963 1960 2003 ...  
## $ yr\_renovated : int 0 1991 0 0 0 0 0 0 0 0 ...  
## $ zipcode : int 98178 98125 98028 98136 98074 98053 98003 98198 98146 98038 ...  
## $ lat : num 47.5 47.7 47.7 47.5 47.6 ...  
## $ long : num -122 -122 -122 -122 -122 ...  
## $ sqft\_living15: int 1340 1690 2720 1360 1800 4760 2238 1650 1780 2390 ...  
## $ sqft\_lot15 : int 5650 7639 8062 5000 7503 101930 6819 9711 8113 7570 ...

head(data)

## id date price bedrooms bathrooms sqft\_living  
## 1 7129300520 20141013T000000 221900 3 1.00 1180  
## 2 6414100192 20141209T000000 538000 3 2.25 2570  
## 3 5631500400 20150225T000000 180000 2 1.00 770  
## 4 2487200875 20141209T000000 604000 4 3.00 1960  
## 5 1954400510 20150218T000000 510000 3 2.00 1680  
## 6 7237550310 20140512T000000 1225000 4 4.50 5420  
## sqft\_lot floors waterfront view condition grade sqft\_above sqft\_basement  
## 1 5650 1 0 0 3 7 1180 0  
## 2 7242 2 0 0 3 7 2170 400  
## 3 10000 1 0 0 3 6 770 0  
## 4 5000 1 0 0 5 7 1050 910  
## 5 8080 1 0 0 3 8 1680 0  
## 6 101930 1 0 0 3 11 3890 1530  
## yr\_built yr\_renovated zipcode lat long sqft\_living15 sqft\_lot15  
## 1 1955 0 98178 47.5112 -122.257 1340 5650  
## 2 1951 1991 98125 47.7210 -122.319 1690 7639  
## 3 1933 0 98028 47.7379 -122.233 2720 8062  
## 4 1965 0 98136 47.5208 -122.393 1360 5000  
## 5 1987 0 98074 47.6168 -122.045 1800 7503  
## 6 2001 0 98053 47.6561 -122.005 4760 101930

summary(data)

## id date price bedrooms   
## Min. :1.000e+06 Length:17384 Min. : 75000 Min. : 0.000   
## 1st Qu.:2.124e+09 Class :character 1st Qu.: 320000 1st Qu.: 3.000   
## Median :3.893e+09 Mode :character Median : 450000 Median : 3.000   
## Mean :4.574e+09 Mean : 539367 Mean : 3.369   
## 3rd Qu.:7.304e+09 3rd Qu.: 640000 3rd Qu.: 4.000   
## Max. :9.900e+09 Max. :7700000 Max. :10.000   
## bathrooms sqft\_living sqft\_lot floors   
## Min. :0.000 Min. : 290 Min. : 520 Min. :1.000   
## 1st Qu.:1.750 1st Qu.: 1420 1st Qu.: 5050 1st Qu.:1.000   
## Median :2.250 Median : 1910 Median : 7616 Median :1.500   
## Mean :2.115 Mean : 2080 Mean : 15092 Mean :1.494   
## 3rd Qu.:2.500 3rd Qu.: 2550 3rd Qu.: 10665 3rd Qu.:2.000   
## Max. :8.000 Max. :13540 Max. :1651359 Max. :3.500   
## waterfront view condition grade   
## Min. :0.000000 Min. :0.0000 Min. :1.000 Min. : 1.000   
## 1st Qu.:0.000000 1st Qu.:0.0000 1st Qu.:3.000 1st Qu.: 7.000   
## Median :0.000000 Median :0.0000 Median :3.000 Median : 7.000   
## Mean :0.007651 Mean :0.2361 Mean :3.411 Mean : 7.655   
## 3rd Qu.:0.000000 3rd Qu.:0.0000 3rd Qu.:4.000 3rd Qu.: 8.000   
## Max. :1.000000 Max. :4.0000 Max. :5.000 Max. :13.000   
## sqft\_above sqft\_basement yr\_built yr\_renovated   
## Min. : 290 Min. : 0.0 Min. :1900 Min. : 0.00   
## 1st Qu.:1200 1st Qu.: 0.0 1st Qu.:1952 1st Qu.: 0.00   
## Median :1560 Median : 0.0 Median :1975 Median : 0.00   
## Mean :1788 Mean : 292.2 Mean :1971 Mean : 83.11   
## 3rd Qu.:2210 3rd Qu.: 560.0 3rd Qu.:1997 3rd Qu.: 0.00   
## Max. :9410 Max. :4820.0 Max. :2015 Max. :2015.00   
## zipcode lat long sqft\_living15   
## Min. :98001 Min. :47.16 Min. :-122.5 Min. : 399   
## 1st Qu.:98033 1st Qu.:47.47 1st Qu.:-122.3 1st Qu.:1490   
## Median :98065 Median :47.57 Median :-122.2 Median :1840   
## Mean :98078 Mean :47.56 Mean :-122.2 Mean :1986   
## 3rd Qu.:98117 3rd Qu.:47.68 3rd Qu.:-122.1 3rd Qu.:2360   
## Max. :98199 Max. :47.78 Max. :-121.3 Max. :6210   
## sqft\_lot15   
## Min. : 651   
## 1st Qu.: 5100   
## Median : 7620   
## Mean : 12776   
## 3rd Qu.: 10065   
## Max. :871200

Revisamos y cambiamos el formato de las variables donde lo creemos necesario Cambiamos el formato de la fecha por si es necesario Nos quedamos con el año y lo trataremos como un factor, ya que puede aportar informacion a la valoracion Vemos que el año de renovacion el valo máximo es 2015 y la media 83.11. no tiene mucho sentido este dato zipcode es la localizacion. Podria ser tratado como un factor, pero hay demasiados. Puede inflir bastante. teniendo el zipcode, la latitud y la longitud al ser coordenadas, podríamos utilizarlas pero para hilar muy fino. de momento las dejamos fuera el id no deja de ser un campo de etiqueta, no aporta informacion

data$date= as.Date(substr(data$date, 0, 8), "%Y%m%d")  
data$date = as.factor(format(data$date,'%Y'))  
data$zipcode = as.factor(data$zipcode)

Modificamos el resto de variables que no son numericas, sino factores, donde el orden numerico no es el significado real de la variable

data$yr\_renovated = as.factor(data$yr\_renovated)  
data$yr\_built = as.factor(data$yr\_built)  
data$condition = as.factor(data$condition)  
data$floors = as.factor(data$floors)  
data$bathrooms = as.factor(data$bathrooms)  
data$bedrooms = as.factor(data$bedrooms)  
data$view = as.factor(data$view)  
data$waterfront = as.factor(data$waterfront)  
unique(data$grade)

## [1] 7 6 8 11 9 5 10 12 4 3 13 1

data$grade = as.factor(data$grade)

Seleccionamos pues las variables que consideramos optimas

houses = subset(data, select=c("date","price","bedrooms","sqft\_living","sqft\_lot","floors","waterfront","view","condition","grade","sqft\_above","sqft\_basement","yr\_built","yr\_renovated","zipcode","sqft\_living15","sqft\_lot15"))  
  
str(houses)

## 'data.frame': 17384 obs. of 17 variables:  
## $ date : Factor w/ 2 levels "2014","2015": 1 1 2 1 2 1 1 2 2 2 ...  
## $ price : num 221900 538000 180000 604000 510000 ...  
## $ bedrooms : Factor w/ 11 levels "0","1","2","3",..: 4 4 3 5 4 5 4 4 4 4 ...  
## $ sqft\_living : int 1180 2570 770 1960 1680 5420 1715 1060 1780 1890 ...  
## $ sqft\_lot : int 5650 7242 10000 5000 8080 101930 6819 9711 7470 6560 ...  
## $ floors : Factor w/ 6 levels "1","1.5","2",..: 1 3 1 1 1 1 3 1 1 3 ...  
## $ waterfront : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ view : Factor w/ 5 levels "0","1","2","3",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ condition : Factor w/ 5 levels "1","2","3","4",..: 3 3 3 5 3 3 3 3 3 3 ...  
## $ grade : Factor w/ 12 levels "1","3","4","5",..: 6 6 5 6 7 10 6 6 6 6 ...  
## $ sqft\_above : int 1180 2170 770 1050 1680 3890 1715 1060 1050 1890 ...  
## $ sqft\_basement: int 0 400 0 910 0 1530 0 0 730 0 ...  
## $ yr\_built : Factor w/ 116 levels "1900","1901",..: 56 52 34 66 88 102 96 64 61 104 ...  
## $ yr\_renovated : Factor w/ 70 levels "0","1934","1940",..: 1 46 1 1 1 1 1 1 1 1 ...  
## $ zipcode : Factor w/ 70 levels "98001","98002",..: 67 56 17 59 38 30 3 69 61 24 ...  
## $ sqft\_living15: int 1340 1690 2720 1360 1800 4760 2238 1650 1780 2390 ...  
## $ sqft\_lot15 : int 5650 7639 8062 5000 7503 101930 6819 9711 8113 7570 ...

head(houses)

## date price bedrooms sqft\_living sqft\_lot floors waterfront view  
## 1 2014 221900 3 1180 5650 1 0 0  
## 2 2014 538000 3 2570 7242 2 0 0  
## 3 2015 180000 2 770 10000 1 0 0  
## 4 2014 604000 4 1960 5000 1 0 0  
## 5 2015 510000 3 1680 8080 1 0 0  
## 6 2014 1225000 4 5420 101930 1 0 0  
## condition grade sqft\_above sqft\_basement yr\_built yr\_renovated zipcode  
## 1 3 7 1180 0 1955 0 98178  
## 2 3 7 2170 400 1951 1991 98125  
## 3 3 6 770 0 1933 0 98028  
## 4 5 7 1050 910 1965 0 98136  
## 5 3 8 1680 0 1987 0 98074  
## 6 3 11 3890 1530 2001 0 98053  
## sqft\_living15 sqft\_lot15  
## 1 1340 5650  
## 2 1690 7639  
## 3 2720 8062  
## 4 1360 5000  
## 5 1800 7503  
## 6 4760 101930

summary(houses)

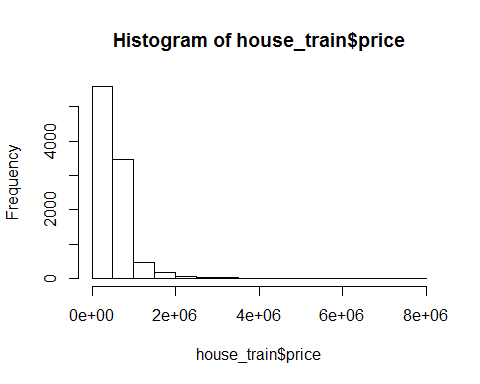
## date price bedrooms sqft\_living   
## 2014:11781 Min. : 75000 3 :7960 Min. : 290   
## 2015: 5603 1st Qu.: 320000 4 :5513 1st Qu.: 1420   
## Median : 450000 2 :2199 Median : 1910   
## Mean : 539367 5 :1272 Mean : 2080   
## 3rd Qu.: 640000 6 : 223 3rd Qu.: 2550   
## Max. :7700000 1 : 156 Max. :13540   
## (Other): 61   
## sqft\_lot floors waterfront view condition  
## Min. : 520 1 :8595 0:17251 0:15660 1: 24   
## 1st Qu.: 5050 1.5:1517 1: 133 1: 262 2: 134   
## Median : 7616 2 :6645 2: 805 3:11263   
## Mean : 15092 2.5: 135 3: 396 4: 4603   
## 3rd Qu.: 10665 3 : 488 4: 261 5: 1360   
## Max. :1651359 3.5: 4   
##   
## grade sqft\_above sqft\_basement yr\_built   
## 7 :7250 Min. : 290 Min. : 0.0 2014 : 462   
## 8 :4920 1st Qu.:1200 1st Qu.: 0.0 2005 : 381   
## 9 :2077 Median :1560 Median : 0.0 2006 : 371   
## 6 :1633 Mean :1788 Mean : 292.2 2004 : 347   
## 10 : 898 3rd Qu.:2210 3rd Qu.: 560.0 2007 : 339   
## 11 : 307 Max. :9410 Max. :4820.0 2003 : 336   
## (Other): 299 (Other):15148   
## yr\_renovated zipcode sqft\_living15 sqft\_lot15   
## 0 :16660 98038 : 490 Min. : 399 Min. : 651   
## 2014 : 71 98115 : 476 1st Qu.:1490 1st Qu.: 5100   
## 2003 : 31 98103 : 469 Median :1840 Median : 7620   
## 2005 : 31 98052 : 460 Mean :1986 Mean : 12776   
## 2013 : 31 98117 : 451 3rd Qu.:2360 3rd Qu.: 10065   
## 2000 : 27 98034 : 448 Max. :6210 Max. :871200   
## (Other): 533 (Other):14590

Dividimos entre train y test, ya que el segundo fichero de la práctica no da datos de precios

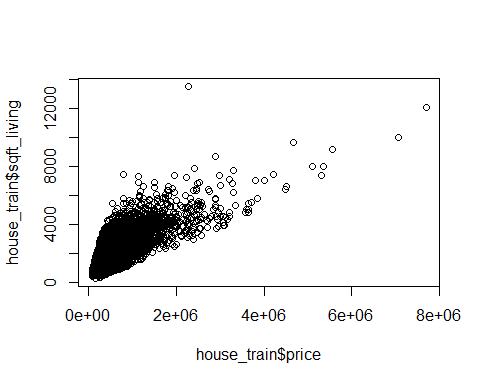
SAMPLE= sample.split(houses$price, SplitRatio = 0.5)  
house\_train = subset(houses, SAMPLE == TRUE)  
house\_test = subset(houses, SAMPLE == FALSE)

Analizamos el comportamiento de la variable dependiente a estudiar (precio) para ver si cumple las condiciones de normalidad (etc) para aplicar un modelo de regresion vemos el comporatamiento de las variables. No tiene pinta de ser normal, pero si una poisson

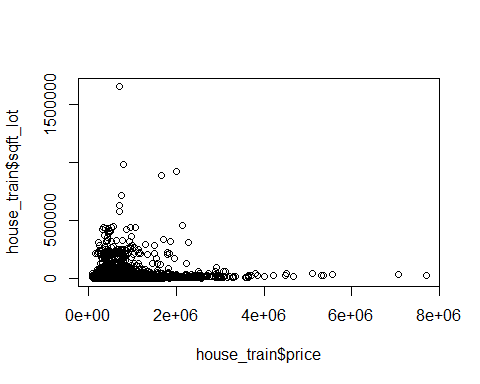
hist(house\_train$price)

 La plotear la variable dependiente a estimar, junto con su relacion con la superficie, parece que si hay una relacion Lineal a pesar de haber outliers

plot(house\_train$price,house\_train$sqft\_living)

 No asi con la superficie de la parcela

plot(house\_train$price,house\_train$sqft\_lot)



Aplicamos un modelo de regresion lineal para ver la relacion entre precio y metros cuadrados y estimar su comportamiento en base los parámetros del modelo y los residuos

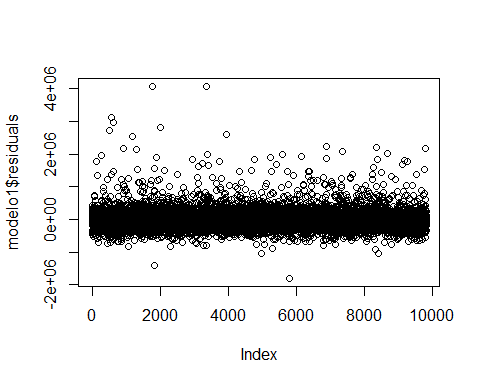
EstimaciÃ³n de ventas en funciÃ³n al precio

modelo1=lm(price~sqft\_living,data=house\_train)  
summary(modelo1)

##   
## Call:  
## lm(formula = price ~ sqft\_living, data = house\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1802294 -156009 -21830 111663 4077085   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -92199.299 6864.030 -13.43 <2e-16 \*\*\*  
## sqft\_living 308.308 2.956 104.30 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 283800 on 9821 degrees of freedom  
## Multiple R-squared: 0.5255, Adjusted R-squared: 0.5255   
## F-statistic: 1.088e+04 on 1 and 9821 DF, p-value: < 2.2e-16

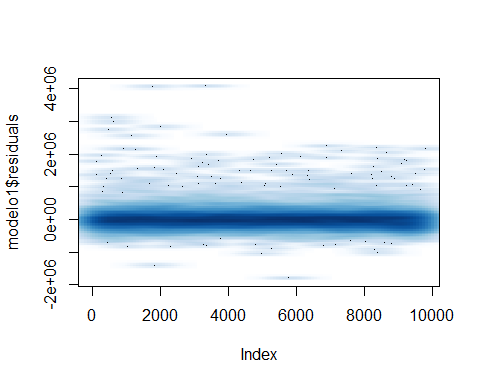
Segun este modelo, solo soy capaz de estimar el 5% de los resultados, por lo que tengo mucha variabilidad aunque la variable si aporte al modelo

plot(modelo1$residuals)



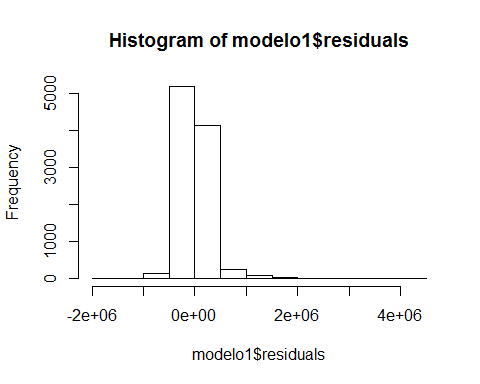
vemos que el comportamiento es bueno, aunque hay muchos valores dispersos por la parte alta

smoothScatter(modelo1$residuals)



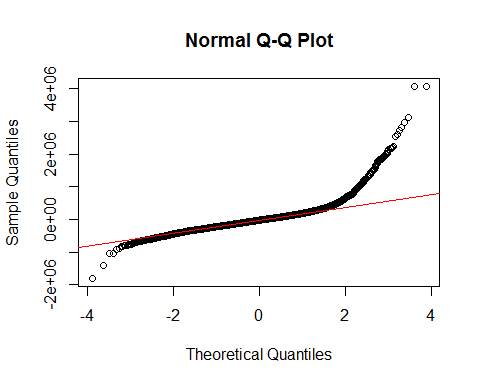
Se acercan a una normal muy picuda, es decir, que están muy concentrados.

hist(modelo1$residuals)



Vemos que por el qqnorm en los valores extremos, no podemos explicar el comportamiento.

qqnorm(modelo1$residuals); qqline(modelo1$residuals,col=2)



confint(modelo1,level=0.95)

## 2.5 % 97.5 %  
## (Intercept) -105654.2081 -78744.3896  
## sqft\_living 302.5137 314.1028

La relacion entre el precio y la superficie de metros cuadrádos útiles podría describirse como: precio = -55.283,995 + 285,540\*sqft\_living, con un nivel de confianza del 95% de que el factor que multimplica sqft\_living está entre 279,4704 - 291,6087 Lo cual quiere decir, que por cada metro cuadrado, umenta aproximadamente entre 280 y 290$ el precio eso en este modelo, que no es bueno y no está ajustado.

Se me ocurre en este punto que: +debemos aplicar una transformacion al precio, ya que no cumple la hipotesis de normalidad +tendremos que aplicar una rlm asi como suavizar los bordes, ya que tenemos problemas con los valores extremos

No hemos detectado cambios estructurales extremos

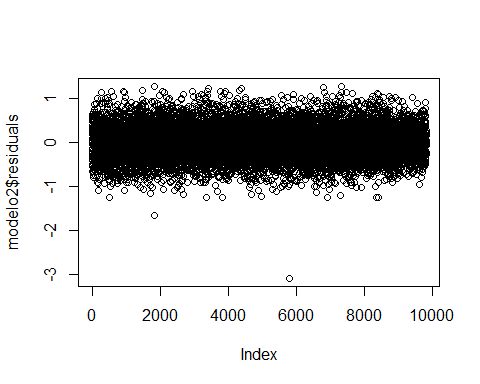
Puesto que la variable del precio no sigue valores normales, sino aprarentemente mas una distrbucion de poisson, aplicamos una transformacion logaritmica y comprobamos los resultados.

Aplicamos la transformacion y comprobamos

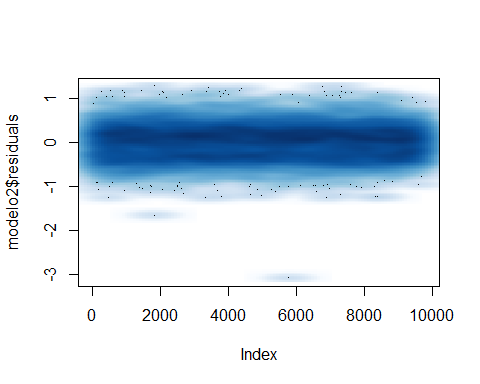
modelo2=lm(log(price)~sqft\_living,data=house\_train)  
summary(modelo2)

##   
## Call:  
## lm(formula = log(price) ~ sqft\_living, data = house\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.08139 -0.29239 0.01676 0.26650 1.26472   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.220e+01 9.354e-03 1304.6 <2e-16 \*\*\*  
## sqft\_living 4.075e-04 4.029e-06 101.1 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3867 on 9821 degrees of freedom  
## Multiple R-squared: 0.5102, Adjusted R-squared: 0.5102   
## F-statistic: 1.023e+04 on 1 and 9821 DF, p-value: < 2.2e-16

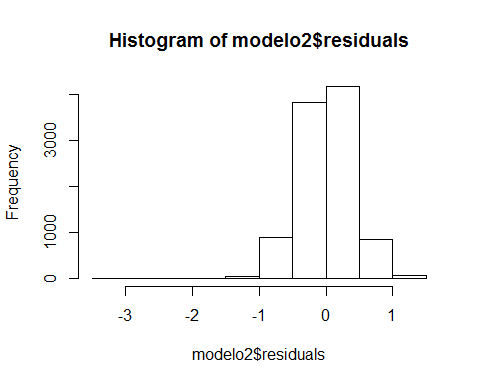
plot(modelo2$residuals)



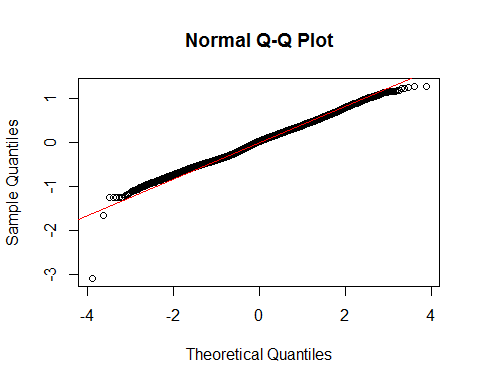
smoothScatter(modelo2$residuals)



hist(modelo2$residuals)



qqnorm(modelo2$residuals); qqline(modelo2$residuals,col=2)



confint(modelo2,level=0.95)

## 2.5 % 97.5 %  
## (Intercept) 1.218549e+01 1.222217e+01  
## sqft\_living 3.995802e-04 4.153741e-04

Vemos que el modelo transformado se ajusta mejora por las siguientes razones: + Los resuiduos siguen una distribucion normal entre 0 y uno.

* la gráfica de disperion de los errores está muy centrada en torno al cero
* la visualizacion del qqplot se ajusta muy bien excepto en los extremos que hay un poco de variacion.

Como primera conclusión, un modelo transformado ajusta mejor. La variabilidad de los errores (R2 ajustado) es practicamente la misma en ambos casos.Aplicamos semielasticidad al modelo (transformacion en una de las variables)

El siguente paso sería calcular las correlaciones parciales, para ver aquellas que mas aportan al modelo lo ideal seria utilizar la correlacion parcial, que elimina los efectos de una variable sobre las demas evitando asi la multicolinealidad, pero las librerías que he usado no me funcionan, asi que utilizaré la correlacion normal, intentado detectar el overfitting con el comportamiento de los residuos.

library(ggm)

## Loading required package: igraph

##   
## Attaching package: 'igraph'

## The following objects are masked from 'package:stats':  
##   
## decompose, spectrum

## The following object is masked from 'package:base':  
##   
## union

##   
## Attaching package: 'ggm'

## The following object is masked from 'package:igraph':  
##   
## pa

library(igraph)  
  
houses\_cor <- house\_train  
houses\_cor$date = as.numeric(houses\_cor$date)  
houses\_cor$bedrooms = as.numeric(houses\_cor$bedrooms)  
houses\_cor$floors = as.numeric(houses\_cor$floors)  
houses\_cor$waterfront = as.numeric(houses\_cor$waterfront)  
houses\_cor$view = as.numeric(houses\_cor$view)  
houses\_cor$condition = as.numeric(houses\_cor$condition)  
houses\_cor$grade = as.numeric(houses\_cor$grade)  
houses\_cor$yr\_built = as.numeric(houses\_cor$yr\_built)  
houses\_cor$yr\_renovated = as.numeric(houses\_cor$yr\_renovated)  
houses\_cor$zipcode = as.numeric(houses\_cor$zipcode)  
#library(parcor)  
#parcor(houses\_cor)  
  
  
cor(houses\_cor)

## date price bedrooms sqft\_living  
## date 1.0000000000 0.0005822182 -0.0006571629 -0.02783991  
## price 0.0005822182 1.0000000000 0.3329273588 0.72493069  
## bedrooms -0.0006571629 0.3329273588 1.0000000000 0.59873527  
## sqft\_living -0.0278399133 0.7249306864 0.5987352652 1.00000000  
## sqft\_lot 0.0072850339 0.0958115190 0.0392828125 0.17303840  
## floors -0.0176512033 0.2587374102 0.1881922071 0.35841782  
## waterfront -0.0147871751 0.2905470063 0.0003954258 0.12473511  
## view -0.0019401032 0.4111259111 0.1004373818 0.30740668  
## condition -0.0498565728 0.0341300142 0.0258481862 -0.06001510  
## grade -0.0306150368 0.6769137051 0.3888177920 0.77297129  
## sqft\_above -0.0279821413 0.6286029249 0.4993021578 0.88026700  
## sqft\_basement -0.0065236198 0.3562938222 0.3313087161 0.46698794  
## yr\_built -0.0022758870 0.0527573571 0.1608786183 0.30637463  
## yr\_renovated -0.0266468817 0.1421185863 0.0319172630 0.07595751  
## zipcode -0.0056125382 -0.0716119915 -0.1598402059 -0.19271193  
## sqft\_living15 -0.0256277410 0.5926419370 0.4179912855 0.75675948  
## sqft\_lot15 -0.0008009322 0.0801679966 0.0317808265 0.17969424  
## sqft\_lot floors waterfront view  
## date 0.007285034 -0.017651203 -0.0147871751 -0.001940103  
## price 0.095811519 0.258737410 0.2905470063 0.411125911  
## bedrooms 0.039282813 0.188192207 0.0003954258 0.100437382  
## sqft\_living 0.173038399 0.358417818 0.1247351142 0.307406679  
## sqft\_lot 1.000000000 -0.001706688 0.0143227875 0.080994196  
## floors -0.001706688 1.000000000 0.0271625751 0.035315093  
## waterfront 0.014322787 0.027162575 1.0000000000 0.420448481  
## view 0.080994196 0.035315093 0.4204484809 1.000000000  
## condition -0.011807395 -0.256761503 0.0063314479 0.043170500  
## grade 0.114730514 0.462022146 0.0898299539 0.267417185  
## sqft\_above 0.183335807 0.525263595 0.0885993377 0.185584745  
## sqft\_basement 0.023004124 -0.223543497 0.0977608572 0.301994076  
## yr\_built 0.059956574 0.486027879 -0.0285036105 -0.053957023  
## yr\_renovated 0.002194423 0.005642410 0.0853884685 0.099424590  
## zipcode -0.125026325 -0.028976871 0.0305339265 0.062825258  
## sqft\_living15 0.148712762 0.288095187 0.0933639124 0.303095406  
## sqft\_lot15 0.757265496 -0.011473661 0.0240513336 0.077326111  
## condition grade sqft\_above sqft\_basement  
## date -0.049856573 -0.03061504 -0.027982141 -0.006523620  
## price 0.034130014 0.67691371 0.628602925 0.356293822  
## bedrooms 0.025848186 0.38881779 0.499302158 0.331308716  
## sqft\_living -0.060015103 0.77297129 0.880266999 0.466987940  
## sqft\_lot -0.011807395 0.11473051 0.183335807 0.023004124  
## floors -0.256761503 0.46202215 0.525263595 -0.223543497  
## waterfront 0.006331448 0.08982995 0.088599338 0.097760857  
## view 0.043170500 0.26741719 0.185584745 0.301994076  
## condition 1.000000000 -0.13978721 -0.156047085 0.164335797  
## grade -0.139787209 1.00000000 0.767836562 0.198056189  
## sqft\_above -0.156047085 0.76783656 1.000000000 -0.008490186  
## sqft\_basement 0.164335797 0.19805619 -0.008490186 1.000000000  
## yr\_built -0.376763847 0.44033763 0.415046738 -0.127818536  
## yr\_renovated -0.068553216 0.02787546 0.037330501 0.090509373  
## zipcode -0.006570518 -0.18195427 -0.239388147 0.039996019  
## sqft\_living15 -0.099350815 0.71892927 0.731946055 0.230777254  
## sqft\_lot15 -0.004542521 0.11457309 0.189485170 0.025571025  
## yr\_built yr\_renovated zipcode sqft\_living15  
## date -0.002275887 -0.026646882 -0.005612538 -0.02562774  
## price 0.052757357 0.142118586 -0.071611991 0.59264194  
## bedrooms 0.160878618 0.031917263 -0.159840206 0.41799129  
## sqft\_living 0.306374630 0.075957511 -0.192711933 0.75675948  
## sqft\_lot 0.059956574 0.002194423 -0.125026325 0.14871276  
## floors 0.486027879 0.005642410 -0.028976871 0.28809519  
## waterfront -0.028503611 0.085388468 0.030533926 0.09336391  
## view -0.053957023 0.099424590 0.062825258 0.30309541  
## condition -0.376763847 -0.068553216 -0.006570518 -0.09935082  
## grade 0.440337627 0.027875456 -0.181954268 0.71892927  
## sqft\_above 0.415046738 0.037330501 -0.239388147 0.73194606  
## sqft\_basement -0.127818536 0.090509373 0.039996019 0.23077725  
## yr\_built 1.000000000 -0.200412984 -0.327766326 0.32473791  
## yr\_renovated -0.200412984 1.000000000 0.066227224 0.01322082  
## zipcode -0.327766326 0.066227224 1.000000000 -0.26996786  
## sqft\_living15 0.324737908 0.013220818 -0.269967856 1.00000000  
## sqft\_lot15 0.075481406 0.006206494 -0.141082254 0.18824579  
## sqft\_lot15  
## date -0.0008009322  
## price 0.0801679966  
## bedrooms 0.0317808265  
## sqft\_living 0.1796942368  
## sqft\_lot 0.7572654963  
## floors -0.0114736609  
## waterfront 0.0240513336  
## view 0.0773261109  
## condition -0.0045425207  
## grade 0.1145730866  
## sqft\_above 0.1894851699  
## sqft\_basement 0.0255710245  
## yr\_built 0.0754814063  
## yr\_renovated 0.0062064942  
## zipcode -0.1410822536  
## sqft\_living15 0.1882457948  
## sqft\_lot15 1.0000000000

Segun la matriz de correlaciones (sin tener en cuenta los efectos de la multicolinealidad),las variables que mas aportarían a la estimacion del precio serían: sqft\_living 0.702916354 grade 0.667468232 sqft\_above 0.605277522 sqft\_living15 0.583480821 view 0.391022681

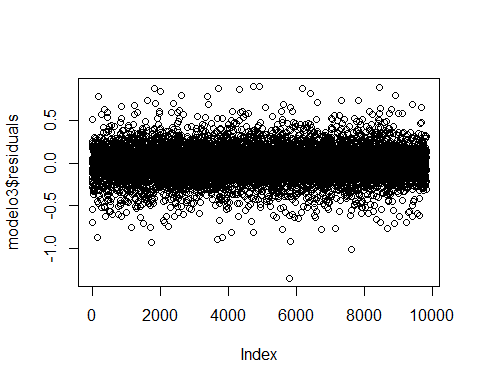
vamos a analizar la aportacion de las distintas variables al modelo, con varias simulaciones.

De esta manera, nos permitirá ir reduciendo aquellas que no aporten.

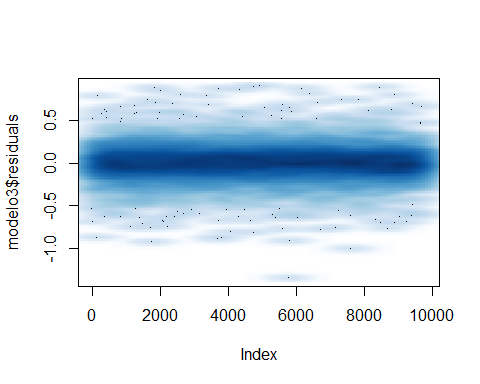
modelo3=lm(log(price)~.,data=house\_train)  
summary(modelo3)

##   
## Call:  
## lm(formula = log(price) ~ ., data = house\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.35311 -0.09400 0.00124 0.09928 0.90564   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.131e+01 1.867e-01 60.592 < 2e-16 \*\*\*  
## date2015 5.460e-02 3.948e-03 13.831 < 2e-16 \*\*\*  
## bedrooms1 -6.004e-02 8.596e-02 -0.699 0.484848   
## bedrooms2 -4.304e-02 8.446e-02 -0.510 0.610408   
## bedrooms3 4.537e-03 8.441e-02 0.054 0.957131   
## bedrooms4 1.712e-02 8.452e-02 0.203 0.839437   
## bedrooms5 -1.127e-02 8.480e-02 -0.133 0.894290   
## bedrooms6 -2.656e-02 8.614e-02 -0.308 0.757818   
## bedrooms7 -2.018e-01 9.515e-02 -2.121 0.033945 \*   
## bedrooms8 3.135e-02 1.143e-01 0.274 0.783947   
## bedrooms9 -1.679e-01 1.259e-01 -1.334 0.182281   
## bedrooms10 -4.369e-01 1.999e-01 -2.185 0.028907 \*   
## sqft\_living 1.431e-04 5.443e-06 26.286 < 2e-16 \*\*\*  
## sqft\_lot 9.054e-07 6.804e-08 13.307 < 2e-16 \*\*\*  
## floors1.5 -1.768e-02 8.125e-03 -2.175 0.029618 \*   
## floors2 -3.866e-02 6.817e-03 -5.672 1.45e-08 \*\*\*  
## floors2.5 -2.694e-02 2.208e-02 -1.220 0.222406   
## floors3 -1.509e-01 1.411e-02 -10.689 < 2e-16 \*\*\*  
## floors3.5 -2.751e-01 1.087e-01 -2.531 0.011375 \*   
## waterfront1 4.598e-01 2.562e-02 17.945 < 2e-16 \*\*\*  
## view1 1.006e-01 1.537e-02 6.542 6.37e-11 \*\*\*  
## view2 1.047e-01 9.255e-03 11.309 < 2e-16 \*\*\*  
## view3 1.620e-01 1.254e-02 12.923 < 2e-16 \*\*\*  
## view4 3.002e-01 1.896e-02 15.833 < 2e-16 \*\*\*  
## condition2 4.192e-02 5.489e-02 0.764 0.445073   
## condition3 1.692e-01 5.125e-02 3.301 0.000968 \*\*\*  
## condition4 2.261e-01 5.129e-02 4.409 1.05e-05 \*\*\*  
## condition5 3.051e-01 5.161e-02 5.912 3.50e-09 \*\*\*  
## grade3 1.246e-01 2.448e-01 0.509 0.610619   
## grade4 7.637e-03 2.092e-01 0.037 0.970883   
## grade5 1.668e-01 2.078e-01 0.802 0.422314   
## grade6 2.715e-01 2.076e-01 1.308 0.190994   
## grade7 4.184e-01 2.077e-01 2.015 0.043944 \*   
## grade8 5.176e-01 2.077e-01 2.492 0.012731 \*   
## grade9 6.207e-01 2.079e-01 2.985 0.002843 \*\*   
## grade10 6.819e-01 2.082e-01 3.275 0.001060 \*\*   
## grade11 7.068e-01 2.087e-01 3.387 0.000710 \*\*\*  
## grade12 7.301e-01 2.101e-01 3.474 0.000514 \*\*\*  
## grade13 6.902e-01 2.176e-01 3.172 0.001518 \*\*   
## sqft\_above 7.532e-05 6.271e-06 12.011 < 2e-16 \*\*\*  
## sqft\_basement NA NA NA NA   
## yr\_built1901 -7.430e-03 5.699e-02 -0.130 0.896269   
## yr\_built1902 -4.210e-02 5.889e-02 -0.715 0.474755   
## yr\_built1903 -1.345e-01 5.437e-02 -2.474 0.013387 \*   
## yr\_built1904 8.257e-02 5.696e-02 1.450 0.147190   
## yr\_built1905 3.585e-02 4.714e-02 0.760 0.446997   
## yr\_built1906 3.675e-03 4.672e-02 0.079 0.937310   
## yr\_built1907 3.647e-02 4.896e-02 0.745 0.456393   
## yr\_built1908 1.095e-01 4.883e-02 2.243 0.024913 \*   
## yr\_built1909 2.476e-02 4.513e-02 0.549 0.583297   
## yr\_built1910 1.065e-02 4.203e-02 0.253 0.799948   
## yr\_built1911 2.151e-02 4.425e-02 0.486 0.626853   
## yr\_built1912 -2.699e-02 4.758e-02 -0.567 0.570475   
## yr\_built1913 5.935e-02 4.964e-02 1.196 0.231860   
## yr\_built1914 3.384e-02 4.789e-02 0.707 0.479769   
## yr\_built1915 1.104e-02 4.926e-02 0.224 0.822710   
## yr\_built1916 1.670e-02 4.576e-02 0.365 0.715057   
## yr\_built1917 5.814e-02 5.025e-02 1.157 0.247249   
## yr\_built1918 -2.357e-02 4.363e-02 -0.540 0.589028   
## yr\_built1919 6.870e-02 4.626e-02 1.485 0.137556   
## yr\_built1920 -4.700e-03 4.410e-02 -0.107 0.915121   
## yr\_built1921 6.443e-02 4.669e-02 1.380 0.167658   
## yr\_built1922 7.743e-02 4.591e-02 1.686 0.091760 .   
## yr\_built1923 2.829e-02 4.854e-02 0.583 0.559978   
## yr\_built1924 6.199e-02 4.206e-02 1.474 0.140549   
## yr\_built1925 4.017e-02 4.105e-02 0.979 0.327816   
## yr\_built1926 4.668e-02 3.994e-02 1.169 0.242562   
## yr\_built1927 2.896e-02 4.411e-02 0.656 0.511561   
## yr\_built1928 3.572e-02 4.136e-02 0.863 0.387888   
## yr\_built1929 3.386e-02 4.343e-02 0.780 0.435681   
## yr\_built1930 1.689e-02 4.427e-02 0.382 0.702834   
## yr\_built1931 2.021e-02 4.745e-02 0.426 0.670263   
## yr\_built1932 1.010e-01 5.360e-02 1.884 0.059586 .   
## yr\_built1933 5.681e-02 5.583e-02 1.018 0.308874   
## yr\_built1934 -3.088e-02 8.843e-02 -0.349 0.726910   
## yr\_built1935 -5.100e-02 7.396e-02 -0.690 0.490474   
## yr\_built1936 5.236e-02 5.550e-02 0.943 0.345532   
## yr\_built1937 7.125e-02 5.095e-02 1.398 0.162050   
## yr\_built1938 9.712e-02 5.308e-02 1.830 0.067328 .   
## yr\_built1939 5.977e-02 4.294e-02 1.392 0.164014   
## yr\_built1940 4.093e-02 4.225e-02 0.969 0.332661   
## yr\_built1941 4.187e-02 4.111e-02 1.018 0.308491   
## yr\_built1942 -1.008e-02 3.925e-02 -0.257 0.797401   
## yr\_built1943 -3.543e-02 4.049e-02 -0.875 0.381548   
## yr\_built1944 -2.005e-03 4.128e-02 -0.049 0.961267   
## yr\_built1945 6.040e-02 4.382e-02 1.378 0.168153   
## yr\_built1946 7.936e-03 4.309e-02 0.184 0.853896   
## yr\_built1947 -3.502e-02 3.916e-02 -0.894 0.371261   
## yr\_built1948 9.215e-03 3.921e-02 0.235 0.814198   
## yr\_built1949 2.066e-03 4.008e-02 0.052 0.958887   
## yr\_built1950 6.784e-03 3.947e-02 0.172 0.863537   
## yr\_built1951 -2.241e-02 3.902e-02 -0.574 0.565726   
## yr\_built1952 -1.112e-02 3.981e-02 -0.279 0.779940   
## yr\_built1953 -5.732e-02 3.943e-02 -1.454 0.146091   
## yr\_built1954 -6.770e-02 3.855e-02 -1.756 0.079105 .   
## yr\_built1955 -6.143e-02 3.927e-02 -1.564 0.117832   
## yr\_built1956 -6.054e-02 4.006e-02 -1.511 0.130798   
## yr\_built1957 -2.968e-02 4.023e-02 -0.738 0.460744   
## yr\_built1958 -3.640e-02 3.940e-02 -0.924 0.355526   
## yr\_built1959 -7.900e-02 3.854e-02 -2.050 0.040428 \*   
## yr\_built1960 -5.171e-02 3.943e-02 -1.311 0.189823   
## yr\_built1961 -5.391e-02 4.042e-02 -1.334 0.182319   
## yr\_built1962 -6.200e-02 3.854e-02 -1.609 0.107650   
## yr\_built1963 -4.396e-02 3.865e-02 -1.137 0.255428   
## yr\_built1964 -6.892e-02 4.094e-02 -1.684 0.092293 .   
## yr\_built1965 -9.731e-02 4.027e-02 -2.417 0.015683 \*   
## yr\_built1966 -8.691e-02 3.916e-02 -2.220 0.026459 \*   
## yr\_built1967 -5.111e-02 3.846e-02 -1.329 0.183931   
## yr\_built1968 -5.637e-02 3.781e-02 -1.491 0.136035   
## yr\_built1969 -6.272e-02 3.884e-02 -1.615 0.106405   
## yr\_built1970 -8.177e-02 4.269e-02 -1.916 0.055455 .   
## yr\_built1971 -6.870e-02 4.472e-02 -1.536 0.124498   
## yr\_built1972 -3.246e-02 4.244e-02 -0.765 0.444367   
## yr\_built1973 -3.736e-02 4.210e-02 -0.887 0.374841   
## yr\_built1974 -6.405e-02 4.118e-02 -1.556 0.119857   
## yr\_built1975 -3.538e-02 4.054e-02 -0.873 0.382869   
## yr\_built1976 -2.172e-02 4.000e-02 -0.543 0.587083   
## yr\_built1977 -4.812e-02 3.789e-02 -1.270 0.204061   
## yr\_built1978 -6.731e-02 3.808e-02 -1.768 0.077153 .   
## yr\_built1979 -6.564e-02 3.855e-02 -1.703 0.088680 .   
## yr\_built1980 -6.847e-02 3.944e-02 -1.736 0.082627 .   
## yr\_built1981 -1.824e-02 4.061e-02 -0.449 0.653339   
## yr\_built1982 -1.252e-02 4.265e-02 -0.294 0.769023   
## yr\_built1983 2.495e-02 4.008e-02 0.622 0.533656   
## yr\_built1984 -4.289e-03 3.959e-02 -0.108 0.913731   
## yr\_built1985 6.022e-03 3.980e-02 0.151 0.879716   
## yr\_built1986 -3.672e-02 4.005e-02 -0.917 0.359254   
## yr\_built1987 -2.942e-02 3.942e-02 -0.746 0.455442   
## yr\_built1988 -1.693e-02 3.924e-02 -0.431 0.666198   
## yr\_built1989 -7.446e-03 3.897e-02 -0.191 0.848473   
## yr\_built1990 -3.381e-02 3.874e-02 -0.873 0.382828   
## yr\_built1991 -3.516e-02 4.046e-02 -0.869 0.384858   
## yr\_built1992 -4.064e-04 4.065e-02 -0.010 0.992024   
## yr\_built1993 8.894e-03 4.080e-02 0.218 0.827438   
## yr\_built1994 5.653e-03 3.962e-02 0.143 0.886540   
## yr\_built1995 5.322e-03 4.159e-02 0.128 0.898182   
## yr\_built1996 1.483e-02 4.011e-02 0.370 0.711649   
## yr\_built1997 -1.261e-02 4.114e-02 -0.306 0.759254   
## yr\_built1998 -1.535e-02 3.982e-02 -0.386 0.699852   
## yr\_built1999 9.090e-03 3.946e-02 0.230 0.817798   
## yr\_built2000 2.512e-02 3.972e-02 0.632 0.527145   
## yr\_built2001 -2.383e-03 3.912e-02 -0.061 0.951424   
## yr\_built2002 -2.143e-02 4.006e-02 -0.535 0.592742   
## yr\_built2003 -3.302e-03 3.806e-02 -0.087 0.930863   
## yr\_built2004 -2.299e-02 3.820e-02 -0.602 0.547236   
## yr\_built2005 -2.528e-02 3.799e-02 -0.665 0.505767   
## yr\_built2006 -1.136e-02 3.791e-02 -0.300 0.764476   
## yr\_built2007 -3.906e-02 3.802e-02 -1.027 0.304311   
## yr\_built2008 -2.016e-04 3.847e-02 -0.005 0.995818   
## yr\_built2009 2.268e-03 4.004e-02 0.057 0.954823   
## yr\_built2010 3.159e-02 4.260e-02 0.742 0.458335   
## yr\_built2011 5.019e-02 4.369e-02 1.149 0.250601   
## yr\_built2012 6.248e-02 4.153e-02 1.505 0.132462   
## yr\_built2013 6.727e-02 4.048e-02 1.662 0.096558 .   
## yr\_built2014 5.676e-02 3.722e-02 1.525 0.127330   
## yr\_built2015 -1.541e-02 5.389e-02 -0.286 0.774886   
## yr\_renovated1940 -1.215e-01 1.825e-01 -0.666 0.505560   
## yr\_renovated1948 7.331e-02 1.836e-01 0.399 0.689625   
## yr\_renovated1950 -2.711e-01 1.852e-01 -1.463 0.143390   
## yr\_renovated1951 1.792e-02 1.835e-01 0.098 0.922204   
## yr\_renovated1953 -2.200e-01 1.856e-01 -1.185 0.236001   
## yr\_renovated1955 8.100e-02 1.054e-01 0.768 0.442342   
## yr\_renovated1956 -1.750e-01 1.068e-01 -1.639 0.101270   
## yr\_renovated1957 4.617e-02 1.059e-01 0.436 0.662784   
## yr\_renovated1958 -1.174e-03 1.299e-01 -0.009 0.992789   
## yr\_renovated1959 -3.091e-01 1.845e-01 -1.676 0.093806 .   
## yr\_renovated1960 -3.389e-01 1.352e-01 -2.507 0.012176 \*   
## yr\_renovated1963 -3.317e-01 1.286e-01 -2.579 0.009936 \*\*   
## yr\_renovated1964 7.328e-04 1.062e-01 0.007 0.994494   
## yr\_renovated1965 6.343e-02 1.302e-01 0.487 0.626105   
## yr\_renovated1967 9.824e-04 1.289e-01 0.008 0.993919   
## yr\_renovated1968 -1.817e-01 9.166e-02 -1.982 0.047519 \*   
## yr\_renovated1969 -1.281e-01 1.291e-01 -0.992 0.321192   
## yr\_renovated1970 -1.529e-01 8.168e-02 -1.872 0.061179 .   
## yr\_renovated1971 9.730e-02 1.839e-01 0.529 0.596834   
## yr\_renovated1972 -2.070e-01 1.300e-01 -1.593 0.111222   
## yr\_renovated1973 -2.385e-01 1.287e-01 -1.853 0.063917 .   
## yr\_renovated1974 2.609e-01 1.916e-01 1.362 0.173285   
## yr\_renovated1975 1.642e-01 1.283e-01 1.279 0.200819   
## yr\_renovated1976 -2.717e-01 1.051e-01 -2.586 0.009715 \*\*   
## yr\_renovated1977 -6.189e-02 7.439e-02 -0.832 0.405427   
## yr\_renovated1978 -8.332e-02 1.080e-01 -0.772 0.440279   
## yr\_renovated1979 -5.655e-02 9.333e-02 -0.606 0.544575   
## yr\_renovated1980 7.870e-02 8.313e-02 0.947 0.343787   
## yr\_renovated1981 2.504e-02 1.050e-01 0.239 0.811433   
## yr\_renovated1982 8.654e-02 9.138e-02 0.947 0.343677   
## yr\_renovated1983 1.669e-01 9.135e-02 1.827 0.067779 .   
## yr\_renovated1984 -1.040e-03 6.469e-02 -0.016 0.987177   
## yr\_renovated1985 6.400e-02 9.322e-02 0.687 0.492360   
## yr\_renovated1986 5.473e-02 6.134e-02 0.892 0.372225   
## yr\_renovated1987 5.025e-02 6.975e-02 0.720 0.471259   
## yr\_renovated1988 1.891e-01 6.482e-02 2.917 0.003542 \*\*   
## yr\_renovated1989 1.949e-02 5.563e-02 0.350 0.726092   
## yr\_renovated1990 4.003e-02 6.118e-02 0.654 0.512912   
## yr\_renovated1991 9.126e-02 5.578e-02 1.636 0.101822   
## yr\_renovated1992 -1.001e-02 6.476e-02 -0.155 0.877172   
## yr\_renovated1993 8.342e-02 6.475e-02 1.288 0.197664   
## yr\_renovated1994 1.150e-02 6.482e-02 0.178 0.859119   
## yr\_renovated1995 1.215e-01 5.527e-02 2.199 0.027897 \*   
## yr\_renovated1996 -1.117e-01 7.459e-02 -1.497 0.134350   
## yr\_renovated1997 -4.320e-02 6.918e-02 -0.624 0.532386   
## yr\_renovated1998 1.193e-01 6.876e-02 1.736 0.082648 .   
## yr\_renovated1999 1.489e-01 7.448e-02 1.999 0.045633 \*   
## yr\_renovated2000 9.057e-02 4.321e-02 2.096 0.036121 \*   
## yr\_renovated2001 1.027e-02 5.544e-02 0.185 0.853091   
## yr\_renovated2002 3.181e-01 5.315e-02 5.985 2.24e-09 \*\*\*  
## yr\_renovated2003 6.775e-02 3.929e-02 1.724 0.084703 .   
## yr\_renovated2004 1.221e-01 5.782e-02 2.112 0.034691 \*   
## yr\_renovated2005 2.241e-02 4.748e-02 0.472 0.636994   
## yr\_renovated2006 1.046e-01 5.776e-02 1.810 0.070298 .   
## yr\_renovated2007 1.141e-01 4.716e-02 2.418 0.015606 \*   
## yr\_renovated2008 1.566e-01 5.789e-02 2.706 0.006823 \*\*   
## yr\_renovated2009 1.760e-01 6.071e-02 2.899 0.003754 \*\*   
## yr\_renovated2010 2.164e-01 6.429e-02 3.366 0.000765 \*\*\*  
## yr\_renovated2011 5.893e-02 8.155e-02 0.723 0.469941   
## yr\_renovated2012 3.256e-02 9.066e-02 0.359 0.719518   
## yr\_renovated2013 1.852e-01 4.726e-02 3.919 8.96e-05 \*\*\*  
## yr\_renovated2014 1.819e-01 2.871e-02 6.335 2.49e-10 \*\*\*  
## yr\_renovated2015 -2.514e-01 1.291e-01 -1.948 0.051466 .   
## zipcode98002 -7.760e-03 2.456e-02 -0.316 0.752051   
## zipcode98003 4.534e-02 2.157e-02 2.102 0.035604 \*   
## zipcode98004 1.123e+00 2.043e-02 54.948 < 2e-16 \*\*\*  
## zipcode98005 7.401e-01 2.534e-02 29.207 < 2e-16 \*\*\*  
## zipcode98006 6.581e-01 1.898e-02 34.668 < 2e-16 \*\*\*  
## zipcode98007 6.836e-01 2.664e-02 25.662 < 2e-16 \*\*\*  
## zipcode98008 6.729e-01 2.103e-02 31.994 < 2e-16 \*\*\*  
## zipcode98010 2.479e-01 2.933e-02 8.451 < 2e-16 \*\*\*  
## zipcode98011 4.501e-01 2.516e-02 17.892 < 2e-16 \*\*\*  
## zipcode98014 2.987e-01 3.035e-02 9.841 < 2e-16 \*\*\*  
## zipcode98019 3.200e-01 2.457e-02 13.022 < 2e-16 \*\*\*  
## zipcode98022 1.626e-02 2.371e-02 0.686 0.492814   
## zipcode98023 -2.287e-02 1.867e-02 -1.225 0.220611   
## zipcode98024 3.933e-01 3.277e-02 12.004 < 2e-16 \*\*\*  
## zipcode98027 5.188e-01 1.999e-02 25.956 < 2e-16 \*\*\*  
## zipcode98028 4.384e-01 2.170e-02 20.199 < 2e-16 \*\*\*  
## zipcode98029 5.739e-01 2.092e-02 27.430 < 2e-16 \*\*\*  
## zipcode98030 4.613e-02 2.179e-02 2.117 0.034288 \*   
## zipcode98031 8.004e-02 2.135e-02 3.749 0.000179 \*\*\*  
## zipcode98032 -1.489e-02 2.829e-02 -0.526 0.598703   
## zipcode98033 7.966e-01 1.909e-02 41.734 < 2e-16 \*\*\*  
## zipcode98034 5.734e-01 1.872e-02 30.631 < 2e-16 \*\*\*  
## zipcode98038 1.525e-01 1.854e-02 8.226 < 2e-16 \*\*\*  
## zipcode98039 1.249e+00 3.609e-02 34.592 < 2e-16 \*\*\*  
## zipcode98040 8.711e-01 2.173e-02 40.080 < 2e-16 \*\*\*  
## zipcode98042 5.793e-02 1.852e-02 3.128 0.001765 \*\*   
## zipcode98045 3.201e-01 2.385e-02 13.423 < 2e-16 \*\*\*  
## zipcode98052 6.446e-01 1.833e-02 35.171 < 2e-16 \*\*\*  
## zipcode98053 5.646e-01 1.988e-02 28.400 < 2e-16 \*\*\*  
## zipcode98055 1.228e-01 2.277e-02 5.396 6.97e-08 \*\*\*  
## zipcode98056 3.024e-01 1.944e-02 15.554 < 2e-16 \*\*\*  
## zipcode98058 1.719e-01 1.925e-02 8.933 < 2e-16 \*\*\*  
## zipcode98059 3.313e-01 1.920e-02 17.256 < 2e-16 \*\*\*  
## zipcode98065 3.786e-01 2.058e-02 18.400 < 2e-16 \*\*\*  
## zipcode98070 3.798e-01 3.044e-02 12.477 < 2e-16 \*\*\*  
## zipcode98072 4.951e-01 2.167e-02 22.846 < 2e-16 \*\*\*  
## zipcode98074 5.486e-01 1.948e-02 28.165 < 2e-16 \*\*\*  
## zipcode98075 5.441e-01 2.045e-02 26.603 < 2e-16 \*\*\*  
## zipcode98077 4.280e-01 2.370e-02 18.063 < 2e-16 \*\*\*  
## zipcode98092 -5.873e-03 1.977e-02 -0.297 0.766357   
## zipcode98102 9.470e-01 3.361e-02 28.178 < 2e-16 \*\*\*  
## zipcode98103 7.923e-01 1.934e-02 40.960 < 2e-16 \*\*\*  
## zipcode98105 9.203e-01 2.382e-02 38.627 < 2e-16 \*\*\*  
## zipcode98106 3.262e-01 2.151e-02 15.167 < 2e-16 \*\*\*  
## zipcode98107 8.391e-01 2.220e-02 37.790 < 2e-16 \*\*\*  
## zipcode98108 3.589e-01 2.621e-02 13.691 < 2e-16 \*\*\*  
## zipcode98109 9.724e-01 2.918e-02 33.322 < 2e-16 \*\*\*  
## zipcode98112 1.013e+00 2.213e-02 45.761 < 2e-16 \*\*\*  
## zipcode98115 7.972e-01 1.905e-02 41.844 < 2e-16 \*\*\*  
## zipcode98116 7.400e-01 2.123e-02 34.862 < 2e-16 \*\*\*  
## zipcode98117 7.905e-01 1.908e-02 41.438 < 2e-16 \*\*\*  
## zipcode98118 4.596e-01 1.930e-02 23.820 < 2e-16 \*\*\*  
## zipcode98119 9.821e-01 2.528e-02 38.845 < 2e-16 \*\*\*  
## zipcode98122 8.165e-01 2.328e-02 35.076 < 2e-16 \*\*\*  
## zipcode98125 5.954e-01 2.012e-02 29.588 < 2e-16 \*\*\*  
## zipcode98126 5.541e-01 2.141e-02 25.888 < 2e-16 \*\*\*  
## zipcode98133 4.724e-01 1.920e-02 24.609 < 2e-16 \*\*\*  
## zipcode98136 6.708e-01 2.225e-02 30.146 < 2e-16 \*\*\*  
## zipcode98144 6.546e-01 2.135e-02 30.658 < 2e-16 \*\*\*  
## zipcode98146 2.165e-01 2.185e-02 9.908 < 2e-16 \*\*\*  
## zipcode98148 2.049e-01 4.097e-02 5.002 5.78e-07 \*\*\*  
## zipcode98155 4.444e-01 1.938e-02 22.930 < 2e-16 \*\*\*  
## zipcode98166 3.057e-01 2.290e-02 13.349 < 2e-16 \*\*\*  
## zipcode98168 4.871e-02 2.245e-02 2.170 0.030062 \*   
## zipcode98177 5.859e-01 2.283e-02 25.663 < 2e-16 \*\*\*  
## zipcode98178 1.655e-01 2.209e-02 7.491 7.42e-14 \*\*\*  
## zipcode98188 9.549e-02 2.556e-02 3.736 0.000188 \*\*\*  
## zipcode98198 5.292e-02 2.163e-02 2.447 0.014433 \*   
## zipcode98199 8.330e-01 2.177e-02 38.256 < 2e-16 \*\*\*  
## sqft\_living15 7.613e-05 4.840e-06 15.729 < 2e-16 \*\*\*  
## sqft\_lot15 -1.557e-07 1.112e-07 -1.400 0.161412   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1795 on 9534 degrees of freedom  
## Multiple R-squared: 0.8975, Adjusted R-squared: 0.8944   
## F-statistic: 289.9 on 288 and 9534 DF, p-value: < 2.2e-16

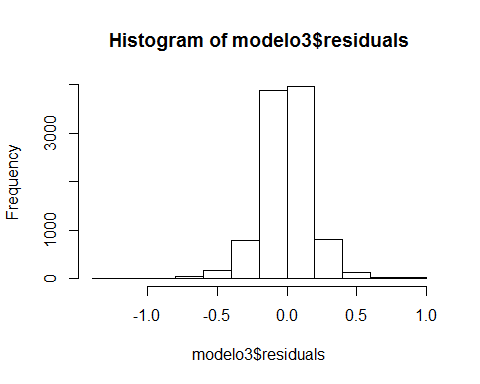
plot(modelo3$residuals)



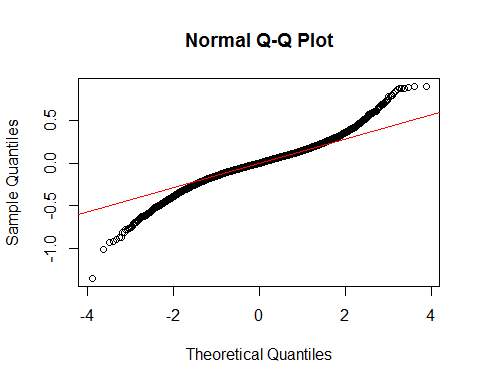
smoothScatter(modelo3$residuals)



hist(modelo3$residuals)



qqnorm(modelo3$residuals); qqline(modelo3$residuals,col=2)



confint(modelo3,level=0.95)

## 2.5 % 97.5 %  
## (Intercept) 1.094623e+01 1.167815e+01  
## date2015 4.686310e-02 6.233939e-02  
## bedrooms1 -2.285367e-01 1.084470e-01  
## bedrooms2 -2.086026e-01 1.225324e-01  
## bedrooms3 -1.609198e-01 1.699945e-01  
## bedrooms4 -1.485431e-01 1.827918e-01  
## bedrooms5 -1.775017e-01 1.549641e-01  
## bedrooms6 -1.954115e-01 1.422886e-01  
## bedrooms7 -3.883389e-01 -1.530190e-02  
## bedrooms8 -1.927771e-01 2.554772e-01  
## bedrooms9 -4.147352e-01 7.885963e-02  
## bedrooms10 -8.288439e-01 -4.496295e-02  
## sqft\_living 1.324164e-04 1.537567e-04  
## sqft\_lot 7.720199e-07 1.038756e-06  
## floors1.5 -3.360224e-02 -1.749101e-03  
## floors2 -5.202660e-02 -2.530162e-02  
## floors2.5 -7.022666e-02 1.633924e-02  
## floors3 -1.785359e-01 -1.232027e-01  
## floors3.5 -4.880633e-01 -6.207168e-02  
## waterfront1 4.096062e-01 5.100669e-01  
## view1 7.043436e-02 1.306970e-01  
## view2 8.652894e-02 1.228143e-01  
## view3 1.374391e-01 1.865906e-01  
## view4 2.630312e-01 3.373646e-01  
## condition2 -6.567591e-02 1.495117e-01  
## condition3 6.869443e-02 2.696108e-01  
## condition4 1.255823e-01 3.266464e-01  
## condition5 2.039217e-01 4.062472e-01  
## grade3 -3.551457e-01 6.044108e-01  
## grade4 -4.025093e-01 4.177842e-01  
## grade5 -2.405980e-01 5.741149e-01  
## grade6 -1.354678e-01 6.784972e-01  
## grade7 1.135344e-02 8.254532e-01  
## grade8 1.104017e-01 9.248058e-01  
## grade9 2.130979e-01 1.028290e+00  
## grade10 2.737486e-01 1.089957e+00  
## grade11 2.977279e-01 1.115910e+00  
## grade12 3.181563e-01 1.141962e+00  
## grade13 2.637004e-01 1.116711e+00  
## sqft\_above 6.303041e-05 8.761529e-05  
## sqft\_basement NA NA  
## yr\_built1901 -1.191375e-01 1.042776e-01  
## yr\_built1902 -1.575435e-01 7.334860e-02  
## yr\_built1903 -2.410792e-01 -2.792174e-02  
## yr\_built1904 -2.908188e-02 1.942250e-01  
## yr\_built1905 -5.655353e-02 1.282465e-01  
## yr\_built1906 -8.790258e-02 9.525158e-02  
## yr\_built1907 -5.951033e-02 1.324503e-01  
## yr\_built1908 1.381311e-02 2.052362e-01  
## yr\_built1909 -6.371265e-02 1.132336e-01  
## yr\_built1910 -7.174092e-02 9.304489e-02  
## yr\_built1911 -6.521902e-02 1.082410e-01  
## yr\_built1912 -1.202560e-01 6.626784e-02  
## yr\_built1913 -3.795399e-02 1.566632e-01  
## yr\_built1914 -6.002453e-02 1.277055e-01  
## yr\_built1915 -8.551466e-02 1.075877e-01  
## yr\_built1916 -7.298519e-02 1.063940e-01  
## yr\_built1917 -4.035319e-02 1.566395e-01  
## yr\_built1918 -1.091055e-01 6.195797e-02  
## yr\_built1919 -2.198044e-02 1.593847e-01  
## yr\_built1920 -9.114063e-02 8.174047e-02  
## yr\_built1921 -2.709804e-02 1.559601e-01  
## yr\_built1922 -1.257376e-02 1.674270e-01  
## yr\_built1923 -6.685540e-02 1.234431e-01  
## yr\_built1924 -2.045518e-02 1.444358e-01  
## yr\_built1925 -4.029629e-02 1.206370e-01  
## yr\_built1926 -3.161384e-02 1.249678e-01  
## yr\_built1927 -5.751431e-02 1.154295e-01  
## yr\_built1928 -4.536501e-02 1.168005e-01  
## yr\_built1929 -5.127931e-02 1.189936e-01  
## yr\_built1930 -6.989434e-02 1.036756e-01  
## yr\_built1931 -7.281113e-02 1.132214e-01  
## yr\_built1932 -4.081678e-03 2.060713e-01  
## yr\_built1933 -5.262326e-02 1.662521e-01  
## yr\_built1934 -2.042188e-01 1.424527e-01  
## yr\_built1935 -1.959794e-01 9.397611e-02  
## yr\_built1936 -5.644100e-02 1.611597e-01  
## yr\_built1937 -2.863009e-02 1.711268e-01  
## yr\_built1938 -6.928071e-03 2.011601e-01  
## yr\_built1939 -2.441004e-02 1.439509e-01  
## yr\_built1940 -4.188388e-02 1.237440e-01  
## yr\_built1941 -3.871983e-02 1.224651e-01  
## yr\_built1942 -8.701053e-02 6.685904e-02  
## yr\_built1943 -1.147979e-01 4.393538e-02  
## yr\_built1944 -8.292120e-02 7.891171e-02  
## yr\_built1945 -2.550238e-02 1.463007e-01  
## yr\_built1946 -7.653500e-02 9.240640e-02  
## yr\_built1947 -1.117778e-01 4.174705e-02  
## yr\_built1948 -6.764414e-02 8.607430e-02  
## yr\_built1949 -7.649157e-02 8.062351e-02  
## yr\_built1950 -7.058149e-02 8.414880e-02  
## yr\_built1951 -9.889430e-02 5.407235e-02  
## yr\_built1952 -8.915306e-02 6.690832e-02  
## yr\_built1953 -1.346119e-01 1.997664e-02  
## yr\_built1954 -1.432631e-01 7.868918e-03  
## yr\_built1955 -1.384159e-01 1.555807e-02  
## yr\_built1956 -1.390773e-01 1.799398e-02  
## yr\_built1957 -1.085421e-01 4.918673e-02  
## yr\_built1958 -1.136302e-01 4.082567e-02  
## yr\_built1959 -1.545577e-01 -3.446617e-03  
## yr\_built1960 -1.290071e-01 2.559382e-02  
## yr\_built1961 -1.331427e-01 2.532196e-02  
## yr\_built1962 -1.375434e-01 1.353437e-02  
## yr\_built1963 -1.197309e-01 3.180723e-02  
## yr\_built1964 -1.491602e-01 1.132360e-02  
## yr\_built1965 -1.762429e-01 -1.837948e-02  
## yr\_built1966 -1.636678e-01 -1.016213e-02  
## yr\_built1967 -1.265039e-01 2.428371e-02  
## yr\_built1968 -1.304830e-01 1.774649e-02  
## yr\_built1969 -1.388588e-01 1.341978e-02  
## yr\_built1970 -1.654377e-01 1.907375e-03  
## yr\_built1971 -1.563625e-01 1.895714e-02  
## yr\_built1972 -1.156446e-01 5.072676e-02  
## yr\_built1973 -1.198864e-01 4.516133e-02  
## yr\_built1974 -1.447616e-01 1.666368e-02  
## yr\_built1975 -1.148515e-01 4.409184e-02  
## yr\_built1976 -1.001241e-01 5.668059e-02  
## yr\_built1977 -1.223845e-01 2.614339e-02  
## yr\_built1978 -1.419609e-01 7.333355e-03  
## yr\_built1979 -1.412108e-01 9.932549e-03  
## yr\_built1980 -1.457894e-01 8.850863e-03  
## yr\_built1981 -9.783549e-02 6.135924e-02  
## yr\_built1982 -9.612962e-02 7.107978e-02  
## yr\_built1983 -5.361976e-02 1.035184e-01  
## yr\_built1984 -8.189409e-02 7.331601e-02  
## yr\_built1985 -7.198557e-02 8.403038e-02  
## yr\_built1986 -1.152220e-01 4.178606e-02  
## yr\_built1987 -1.066841e-01 4.784354e-02  
## yr\_built1988 -9.384504e-02 5.999042e-02  
## yr\_built1989 -8.383917e-02 6.894637e-02  
## yr\_built1990 -1.097418e-01 4.212622e-02  
## yr\_built1991 -1.144605e-01 4.414570e-02  
## yr\_built1992 -8.008732e-02 7.927458e-02  
## yr\_built1993 -7.108175e-02 8.886996e-02  
## yr\_built1994 -7.200770e-02 8.331385e-02  
## yr\_built1995 -7.620061e-02 8.684411e-02  
## yr\_built1996 -6.380087e-02 9.345610e-02  
## yr\_built1997 -9.325155e-02 6.803522e-02  
## yr\_built1998 -9.340838e-02 6.270438e-02  
## yr\_built1999 -6.825411e-02 8.643472e-02  
## yr\_built2000 -5.273826e-02 1.029730e-01  
## yr\_built2001 -7.907434e-02 7.430749e-02  
## yr\_built2002 -9.994499e-02 5.709373e-02  
## yr\_built2003 -7.791522e-02 7.131032e-02  
## yr\_built2004 -9.787792e-02 5.188863e-02  
## yr\_built2005 -9.975094e-02 4.918802e-02  
## yr\_built2006 -8.567687e-02 6.295816e-02  
## yr\_built2007 -1.135909e-01 3.547171e-02  
## yr\_built2008 -7.561227e-02 7.520900e-02  
## yr\_built2009 -7.621144e-02 8.074782e-02  
## yr\_built2010 -5.191064e-02 1.150951e-01  
## yr\_built2011 -3.544086e-02 1.358291e-01  
## yr\_built2012 -1.892007e-02 1.438791e-01  
## yr\_built2013 -1.207267e-02 1.466069e-01  
## yr\_built2014 -1.620508e-02 1.297210e-01  
## yr\_built2015 -1.210385e-01 9.021614e-02  
## yr\_renovated1940 -4.792367e-01 2.362271e-01  
## yr\_renovated1948 -2.864996e-01 4.331147e-01  
## yr\_renovated1950 -6.341485e-01 9.202290e-02  
## yr\_renovated1951 -3.417652e-01 3.776054e-01  
## yr\_renovated1953 -5.837556e-01 1.438534e-01  
## yr\_renovated1955 -1.256713e-01 2.876771e-01  
## yr\_renovated1956 -3.842705e-01 3.430772e-02  
## yr\_renovated1957 -1.613749e-01 2.537209e-01  
## yr\_renovated1958 -2.557918e-01 2.534437e-01  
## yr\_renovated1959 -6.706825e-01 5.245752e-02  
## yr\_renovated1960 -6.038626e-01 -7.397001e-02  
## yr\_renovated1963 -5.838097e-01 -7.953679e-02  
## yr\_renovated1964 -2.073907e-01 2.088562e-01  
## yr\_renovated1965 -1.917670e-01 3.186325e-01  
## yr\_renovated1967 -2.516974e-01 2.536623e-01  
## yr\_renovated1968 -3.613236e-01 -1.988093e-03  
## yr\_renovated1969 -3.811748e-01 1.250000e-01  
## yr\_renovated1970 -3.130626e-01 7.171166e-03  
## yr\_renovated1971 -2.632617e-01 4.578614e-01  
## yr\_renovated1972 -4.618125e-01 4.774501e-02  
## yr\_renovated1973 -4.907764e-01 1.380217e-02  
## yr\_renovated1974 -1.146591e-01 6.365494e-01  
## yr\_renovated1975 -8.738441e-02 4.157479e-01  
## yr\_renovated1976 -4.776687e-01 -6.578199e-02  
## yr\_renovated1977 -2.077161e-01 8.392869e-02  
## yr\_renovated1978 -2.949306e-01 1.283004e-01  
## yr\_renovated1979 -2.394889e-01 1.263903e-01  
## yr\_renovated1980 -8.424811e-02 2.416545e-01  
## yr\_renovated1981 -1.806865e-01 2.307657e-01  
## yr\_renovated1982 -9.259176e-02 2.656638e-01  
## yr\_renovated1983 -1.219800e-02 3.459221e-01  
## yr\_renovated1984 -1.278424e-01 1.257631e-01  
## yr\_renovated1985 -1.187230e-01 2.467255e-01  
## yr\_renovated1986 -6.549928e-02 1.749681e-01  
## yr\_renovated1987 -8.646709e-02 1.869648e-01  
## yr\_renovated1988 6.202024e-02 3.161286e-01  
## yr\_renovated1989 -8.955971e-02 1.285394e-01  
## yr\_renovated1990 -7.989836e-02 1.599674e-01  
## yr\_renovated1991 -1.806995e-02 2.005975e-01  
## yr\_renovated1992 -1.369597e-01 1.169401e-01  
## yr\_renovated1993 -4.350758e-02 2.103565e-01  
## yr\_renovated1994 -1.155478e-01 1.385574e-01  
## yr\_renovated1995 1.320031e-02 2.298634e-01  
## yr\_renovated1996 -2.579086e-01 3.453094e-02  
## yr\_renovated1997 -1.788128e-01 9.241731e-02  
## yr\_renovated1998 -1.543659e-02 2.541243e-01  
## yr\_renovated1999 2.891875e-03 2.949024e-01  
## yr\_renovated2000 5.861950e-03 1.752766e-01  
## yr\_renovated2001 -9.840199e-02 1.189336e-01  
## yr\_renovated2002 2.139125e-01 4.222678e-01  
## yr\_renovated2003 -9.273620e-03 1.447771e-01  
## yr\_renovated2004 8.790821e-03 2.354784e-01  
## yr\_renovated2005 -7.066760e-02 1.154835e-01  
## yr\_renovated2006 -8.665299e-03 2.177765e-01  
## yr\_renovated2007 2.161174e-02 2.065130e-01  
## yr\_renovated2008 4.317028e-02 2.701231e-01  
## yr\_renovated2009 5.698319e-02 2.949808e-01  
## yr\_renovated2010 9.040324e-02 3.424499e-01  
## yr\_renovated2011 -1.009274e-01 2.187847e-01  
## yr\_renovated2012 -1.451517e-01 2.102646e-01  
## yr\_renovated2013 9.256652e-02 2.778399e-01  
## yr\_renovated2014 1.255990e-01 2.381644e-01  
## yr\_renovated2015 -5.044552e-01 1.599891e-03  
## zipcode98002 -5.590190e-02 4.038282e-02  
## zipcode98003 3.052727e-03 8.762892e-02  
## zipcode98004 1.082687e+00 1.162792e+00  
## zipcode98005 6.903888e-01 7.897255e-01  
## zipcode98006 6.208885e-01 6.953107e-01  
## zipcode98007 6.313660e-01 7.357982e-01  
## zipcode98008 6.317037e-01 7.141612e-01  
## zipcode98010 1.903935e-01 3.053842e-01  
## zipcode98011 4.008260e-01 4.994620e-01  
## zipcode98014 2.391989e-01 3.581910e-01  
## zipcode98019 2.718037e-01 3.681377e-01  
## zipcode98022 -3.021411e-02 6.273722e-02  
## zipcode98023 -5.946435e-02 1.372591e-02  
## zipcode98024 3.291032e-01 4.575664e-01  
## zipcode98027 4.796643e-01 5.580317e-01  
## zipcode98028 3.958683e-01 4.809596e-01  
## zipcode98029 5.328957e-01 6.149224e-01  
## zipcode98030 3.415893e-03 8.884247e-02  
## zipcode98031 3.818812e-02 1.218960e-01  
## zipcode98032 -7.035529e-02 4.057296e-02  
## zipcode98033 7.591409e-01 8.339685e-01  
## zipcode98034 5.366999e-01 6.100879e-01  
## zipcode98038 1.161291e-01 1.887945e-01  
## zipcode98039 1.177830e+00 1.319336e+00  
## zipcode98040 8.284857e-01 9.136916e-01  
## zipcode98042 2.162590e-02 9.422857e-02  
## zipcode98045 2.733487e-01 3.668359e-01  
## zipcode98052 6.086786e-01 6.805312e-01  
## zipcode98053 5.256623e-01 6.036055e-01  
## zipcode98055 7.822055e-02 1.674707e-01  
## zipcode98056 2.642802e-01 3.404960e-01  
## zipcode98058 1.341875e-01 2.096386e-01  
## zipcode98059 2.936325e-01 3.688921e-01  
## zipcode98065 3.382880e-01 4.189611e-01  
## zipcode98070 3.201382e-01 4.394745e-01  
## zipcode98072 4.526350e-01 5.375996e-01  
## zipcode98074 5.104213e-01 5.867828e-01  
## zipcode98075 5.040525e-01 5.842411e-01  
## zipcode98077 3.815913e-01 4.744965e-01  
## zipcode98092 -4.461886e-02 3.287190e-02  
## zipcode98102 8.811015e-01 1.012858e+00  
## zipcode98103 7.543784e-01 8.302108e-01  
## zipcode98105 8.735660e-01 9.669669e-01  
## zipcode98106 2.840524e-01 3.683739e-01  
## zipcode98107 7.955882e-01 8.826392e-01  
## zipcode98108 3.074747e-01 4.102349e-01  
## zipcode98109 9.152442e-01 1.029655e+00  
## zipcode98112 9.691543e-01 1.055898e+00  
## zipcode98115 7.598602e-01 8.345523e-01  
## zipcode98116 6.983586e-01 7.815706e-01  
## zipcode98117 7.531461e-01 8.279392e-01  
## zipcode98118 4.218228e-01 4.974751e-01  
## zipcode98119 9.325440e-01 1.031662e+00  
## zipcode98122 7.709146e-01 8.621793e-01  
## zipcode98125 5.559365e-01 6.348253e-01  
## zipcode98126 5.121891e-01 5.961078e-01  
## zipcode98133 4.347762e-01 5.100349e-01  
## zipcode98136 6.272218e-01 7.144639e-01  
## zipcode98144 6.127443e-01 6.964509e-01  
## zipcode98146 1.736358e-01 2.592841e-01  
## zipcode98148 1.246092e-01 2.852140e-01  
## zipcode98155 4.063908e-01 4.823683e-01  
## zipcode98166 2.607853e-01 3.505572e-01  
## zipcode98168 4.701385e-03 9.272693e-02  
## zipcode98177 5.411617e-01 6.306691e-01  
## zipcode98178 1.221812e-01 2.087822e-01  
## zipcode98188 4.538943e-02 1.455931e-01  
## zipcode98198 1.052410e-02 9.532436e-02  
## zipcode98199 7.903033e-01 8.756653e-01  
## sqft\_living15 6.664635e-05 8.562272e-05  
## sqft\_lot15 -3.736164e-07 6.223120e-08

El modelo 3 nos indica cuales de las variables no son significativas para le modelo. Entre ellas: + aquellas que no tienen \*: bedrooms, sqft\_basement,yr\_built, yr\_renovated.

* Evidentemente, algunas de las variables anteriores, al ser factores, aportan cuando hay datos, pero no aportan cuando no los hay, con lo cual pueden deshecharse las variables

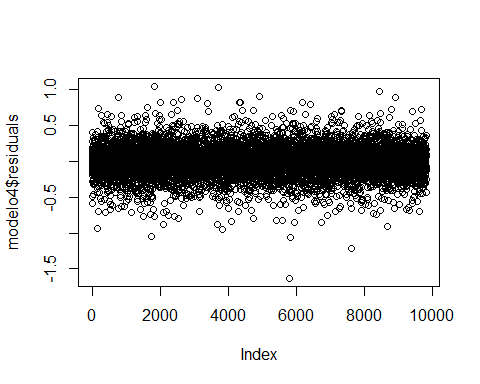
+En este caso el R2 es elevado R-squared: 0.8866, eso indica que puede explicarse el 88% de la variabilidad de los datos.

Eliminamos del modelo las variables que hemos visto que no aportan.

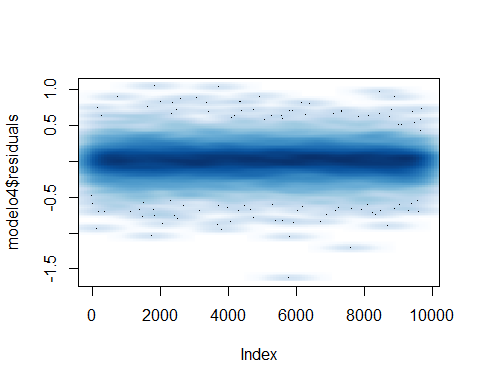
modelo4=lm(log(price)~sqft\_living+grade+sqft\_above+sqft\_living15+view+waterfront+zipcode,data=house\_train)  
summary(modelo4)

##   
## Call:  
## lm(formula = log(price) ~ sqft\_living + grade + sqft\_above +   
## sqft\_living15 + view + waterfront + zipcode, data = house\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.63146 -0.10271 0.00501 0.10594 1.04793   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.121e+01 1.969e-01 56.917 < 2e-16 \*\*\*  
## sqft\_living 1.754e-04 4.995e-06 35.112 < 2e-16 \*\*\*  
## grade3 2.169e-01 2.397e-01 0.905 0.365381   
## grade4 2.698e-01 2.041e-01 1.322 0.186241   
## grade5 4.592e-01 1.970e-01 2.331 0.019763 \*   
## grade6 5.583e-01 1.963e-01 2.844 0.004466 \*\*   
## grade7 6.839e-01 1.963e-01 3.484 0.000496 \*\*\*  
## grade8 7.597e-01 1.964e-01 3.869 0.000110 \*\*\*  
## grade9 8.608e-01 1.965e-01 4.380 1.20e-05 \*\*\*  
## grade10 9.147e-01 1.968e-01 4.648 3.39e-06 \*\*\*  
## grade11 9.272e-01 1.973e-01 4.700 2.64e-06 \*\*\*  
## grade12 9.372e-01 1.989e-01 4.712 2.49e-06 \*\*\*  
## grade13 8.842e-01 2.075e-01 4.260 2.06e-05 \*\*\*  
## sqft\_above 5.249e-05 5.653e-06 9.284 < 2e-16 \*\*\*  
## sqft\_living15 8.110e-05 5.038e-06 16.097 < 2e-16 \*\*\*  
## view1 1.085e-01 1.623e-02 6.684 2.45e-11 \*\*\*  
## view2 1.107e-01 9.754e-03 11.353 < 2e-16 \*\*\*  
## view3 1.760e-01 1.327e-02 13.262 < 2e-16 \*\*\*  
## view4 2.984e-01 2.007e-02 14.863 < 2e-16 \*\*\*  
## waterfront1 4.493e-01 2.684e-02 16.741 < 2e-16 \*\*\*  
## zipcode98002 1.501e-02 2.613e-02 0.574 0.565824   
## zipcode98003 3.351e-02 2.306e-02 1.453 0.146208   
## zipcode98004 1.113e+00 2.144e-02 51.888 < 2e-16 \*\*\*  
## zipcode98005 7.565e-01 2.678e-02 28.251 < 2e-16 \*\*\*  
## zipcode98006 6.606e-01 2.014e-02 32.795 < 2e-16 \*\*\*  
## zipcode98007 6.709e-01 2.823e-02 23.767 < 2e-16 \*\*\*  
## zipcode98008 6.644e-01 2.215e-02 29.993 < 2e-16 \*\*\*  
## zipcode98010 3.089e-01 3.074e-02 10.048 < 2e-16 \*\*\*  
## zipcode98011 4.304e-01 2.690e-02 15.999 < 2e-16 \*\*\*  
## zipcode98014 3.189e-01 3.223e-02 9.894 < 2e-16 \*\*\*  
## zipcode98019 3.298e-01 2.611e-02 12.632 < 2e-16 \*\*\*  
## zipcode98022 1.194e-01 2.471e-02 4.833 1.36e-06 \*\*\*  
## zipcode98023 -3.072e-02 1.990e-02 -1.544 0.122617   
## zipcode98024 4.599e-01 3.463e-02 13.280 < 2e-16 \*\*\*  
## zipcode98027 5.172e-01 2.128e-02 24.311 < 2e-16 \*\*\*  
## zipcode98028 4.254e-01 2.316e-02 18.365 < 2e-16 \*\*\*  
## zipcode98029 5.693e-01 2.222e-02 25.623 < 2e-16 \*\*\*  
## zipcode98030 4.768e-02 2.325e-02 2.050 0.040370 \*   
## zipcode98031 8.715e-02 2.271e-02 3.838 0.000125 \*\*\*  
## zipcode98032 -1.552e-02 3.008e-02 -0.516 0.606016   
## zipcode98033 8.003e-01 2.033e-02 39.366 < 2e-16 \*\*\*  
## zipcode98034 5.550e-01 1.989e-02 27.903 < 2e-16 \*\*\*  
## zipcode98038 1.685e-01 1.966e-02 8.571 < 2e-16 \*\*\*  
## zipcode98039 1.270e+00 3.800e-02 33.427 < 2e-16 \*\*\*  
## zipcode98040 8.773e-01 2.291e-02 38.300 < 2e-16 \*\*\*  
## zipcode98042 7.614e-02 1.968e-02 3.870 0.000110 \*\*\*  
## zipcode98045 3.382e-01 2.538e-02 13.325 < 2e-16 \*\*\*  
## zipcode98052 6.352e-01 1.954e-02 32.500 < 2e-16 \*\*\*  
## zipcode98053 5.675e-01 2.104e-02 26.971 < 2e-16 \*\*\*  
## zipcode98055 1.154e-01 2.426e-02 4.758 1.98e-06 \*\*\*  
## zipcode98056 3.176e-01 2.069e-02 15.349 < 2e-16 \*\*\*  
## zipcode98058 1.818e-01 2.050e-02 8.869 < 2e-16 \*\*\*  
## zipcode98059 3.365e-01 2.048e-02 16.432 < 2e-16 \*\*\*  
## zipcode98065 3.712e-01 2.181e-02 17.021 < 2e-16 \*\*\*  
## zipcode98070 4.447e-01 3.144e-02 14.146 < 2e-16 \*\*\*  
## zipcode98072 4.918e-01 2.294e-02 21.439 < 2e-16 \*\*\*  
## zipcode98074 5.443e-01 2.075e-02 26.236 < 2e-16 \*\*\*  
## zipcode98075 5.410e-01 2.184e-02 24.769 < 2e-16 \*\*\*  
## zipcode98077 4.549e-01 2.510e-02 18.121 < 2e-16 \*\*\*  
## zipcode98092 2.057e-02 2.103e-02 0.978 0.328015   
## zipcode98102 9.516e-01 3.480e-02 27.346 < 2e-16 \*\*\*  
## zipcode98103 7.825e-01 1.957e-02 39.973 < 2e-16 \*\*\*  
## zipcode98105 9.569e-01 2.445e-02 39.133 < 2e-16 \*\*\*  
## zipcode98106 3.204e-01 2.276e-02 14.081 < 2e-16 \*\*\*  
## zipcode98107 8.284e-01 2.287e-02 36.228 < 2e-16 \*\*\*  
## zipcode98108 3.519e-01 2.780e-02 12.660 < 2e-16 \*\*\*  
## zipcode98109 1.002e+00 3.036e-02 33.008 < 2e-16 \*\*\*  
## zipcode98112 1.039e+00 2.254e-02 46.100 < 2e-16 \*\*\*  
## zipcode98115 8.205e-01 1.951e-02 42.058 < 2e-16 \*\*\*  
## zipcode98116 7.616e-01 2.200e-02 34.620 < 2e-16 \*\*\*  
## zipcode98117 8.033e-01 1.968e-02 40.810 < 2e-16 \*\*\*  
## zipcode98118 4.674e-01 2.017e-02 23.172 < 2e-16 \*\*\*  
## zipcode98119 9.840e-01 2.613e-02 37.665 < 2e-16 \*\*\*  
## zipcode98122 8.062e-01 2.389e-02 33.744 < 2e-16 \*\*\*  
## zipcode98125 5.828e-01 2.101e-02 27.733 < 2e-16 \*\*\*  
## zipcode98126 5.626e-01 2.235e-02 25.172 < 2e-16 \*\*\*  
## zipcode98133 4.588e-01 2.013e-02 22.794 < 2e-16 \*\*\*  
## zipcode98136 6.706e-01 2.325e-02 28.843 < 2e-16 \*\*\*  
## zipcode98144 6.761e-01 2.211e-02 30.575 < 2e-16 \*\*\*  
## zipcode98146 2.128e-01 2.299e-02 9.253 < 2e-16 \*\*\*  
## zipcode98148 1.775e-01 4.319e-02 4.111 3.97e-05 \*\*\*  
## zipcode98155 4.447e-01 2.041e-02 21.783 < 2e-16 \*\*\*  
## zipcode98166 3.074e-01 2.408e-02 12.765 < 2e-16 \*\*\*  
## zipcode98168 3.262e-02 2.365e-02 1.379 0.167858   
## zipcode98177 5.763e-01 2.400e-02 24.016 < 2e-16 \*\*\*  
## zipcode98178 1.541e-01 2.316e-02 6.654 3.00e-11 \*\*\*  
## zipcode98188 7.499e-02 2.711e-02 2.766 0.005683 \*\*   
## zipcode98198 3.649e-02 2.302e-02 1.585 0.112929   
## zipcode98199 8.619e-01 2.254e-02 38.238 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1937 on 9734 degrees of freedom  
## Multiple R-squared: 0.8782, Adjusted R-squared: 0.8771   
## F-statistic: 797.7 on 88 and 9734 DF, p-value: < 2.2e-16

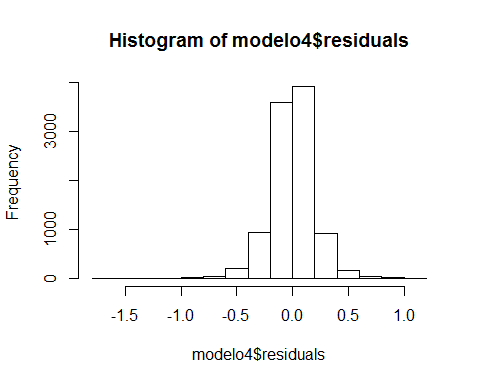
plot(modelo4$residuals)



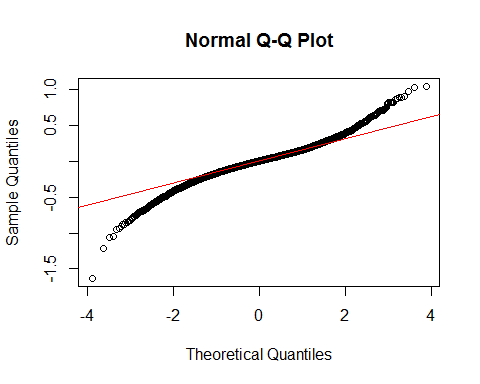
smoothScatter(modelo4$residuals)



hist(modelo4$residuals)



qqnorm(modelo4$residuals); qqline(modelo4$residuals,col=2)



confint(modelo4,level=0.95)

## 2.5 % 97.5 %  
## (Intercept) 1.082033e+01 1.159221e+01  
## sqft\_living 1.655844e-04 1.851660e-04  
## grade3 -2.528523e-01 6.867507e-01  
## grade4 -1.302909e-01 6.699098e-01  
## grade5 7.308170e-02 8.453630e-01  
## grade6 1.734747e-01 9.431155e-01  
## grade7 2.991496e-01 1.068657e+00  
## grade8 3.748283e-01 1.144647e+00  
## grade9 4.755665e-01 1.246039e+00  
## grade10 5.289450e-01 1.300415e+00  
## grade11 5.404779e-01 1.313964e+00  
## grade12 5.473138e-01 1.327064e+00  
## grade13 4.773725e-01 1.291017e+00  
## sqft\_above 4.140321e-05 6.356701e-05  
## sqft\_living15 7.122248e-05 9.097351e-05  
## view1 7.665635e-02 1.402708e-01  
## view2 9.161586e-02 1.298561e-01  
## view3 1.499986e-01 2.020286e-01  
## view4 2.590256e-01 3.377263e-01  
## waterfront1 3.966754e-01 5.018914e-01  
## zipcode98002 -3.622225e-02 6.623722e-02  
## zipcode98003 -1.169225e-02 7.871242e-02  
## zipcode98004 1.070475e+00 1.154531e+00  
## zipcode98005 7.039905e-01 8.089686e-01  
## zipcode98006 6.210873e-01 7.000527e-01  
## zipcode98007 6.155981e-01 7.262686e-01  
## zipcode98008 6.209516e-01 7.077919e-01  
## zipcode98010 2.486009e-01 3.691122e-01  
## zipcode98011 3.776646e-01 4.831281e-01  
## zipcode98014 2.557233e-01 3.820855e-01  
## zipcode98019 2.786396e-01 3.810005e-01  
## zipcode98022 7.098427e-02 1.678413e-01  
## zipcode98023 -6.972473e-02 8.281136e-03  
## zipcode98024 3.919742e-01 5.277286e-01  
## zipcode98027 4.755288e-01 5.589397e-01  
## zipcode98028 3.799913e-01 4.708011e-01  
## zipcode98029 5.257457e-01 6.128523e-01  
## zipcode98030 2.093133e-03 9.325953e-02  
## zipcode98031 4.264168e-02 1.316567e-01  
## zipcode98032 -7.448321e-02 4.345135e-02  
## zipcode98033 7.604633e-01 8.401666e-01  
## zipcode98034 5.160189e-01 5.939986e-01  
## zipcode98038 1.299813e-01 2.070672e-01  
## zipcode98039 1.195844e+00 1.344831e+00  
## zipcode98040 8.323745e-01 9.221730e-01  
## zipcode98042 3.757533e-02 1.147122e-01  
## zipcode98045 2.884366e-01 3.879403e-01  
## zipcode98052 5.968943e-01 6.735179e-01  
## zipcode98053 5.262711e-01 6.087647e-01  
## zipcode98055 6.786443e-02 1.629647e-01  
## zipcode98056 2.769964e-01 3.581063e-01  
## zipcode98058 1.416504e-01 2.220305e-01  
## zipcode98059 2.963989e-01 3.766930e-01  
## zipcode98065 3.284453e-01 4.139402e-01  
## zipcode98070 3.830813e-01 5.063215e-01  
## zipcode98072 4.468534e-01 5.367898e-01  
## zipcode98074 5.036507e-01 5.849884e-01  
## zipcode98075 4.981755e-01 5.838017e-01  
## zipcode98077 4.057019e-01 5.041210e-01  
## zipcode98092 -2.065374e-02 6.179843e-02  
## zipcode98102 8.833882e-01 1.019814e+00  
## zipcode98103 7.441020e-01 8.208436e-01  
## zipcode98105 9.089945e-01 1.004861e+00  
## zipcode98106 2.758054e-01 3.650168e-01  
## zipcode98107 7.835803e-01 8.732268e-01  
## zipcode98108 2.974029e-01 4.063736e-01  
## zipcode98109 9.425120e-01 1.061523e+00  
## zipcode98112 9.947341e-01 1.083084e+00  
## zipcode98115 7.822919e-01 8.587780e-01  
## zipcode98116 7.184669e-01 8.047093e-01  
## zipcode98117 7.646916e-01 8.418583e-01  
## zipcode98118 4.278316e-01 5.069042e-01  
## zipcode98119 9.328288e-01 1.035255e+00  
## zipcode98122 7.593383e-01 8.529999e-01  
## zipcode98125 5.415772e-01 6.239587e-01  
## zipcode98126 5.187490e-01 6.063652e-01  
## zipcode98133 4.193136e-01 4.982196e-01  
## zipcode98136 6.250196e-01 7.161700e-01  
## zipcode98144 6.327161e-01 7.194019e-01  
## zipcode98146 1.676828e-01 2.578278e-01  
## zipcode98148 9.288980e-02 2.622037e-01  
## zipcode98155 4.046624e-01 4.846927e-01  
## zipcode98166 2.601700e-01 3.545688e-01  
## zipcode98168 -1.374008e-02 7.897625e-02  
## zipcode98177 5.292963e-01 6.233803e-01  
## zipcode98178 1.087260e-01 1.995324e-01  
## zipcode98188 2.184858e-02 1.281294e-01  
## zipcode98198 -8.630277e-03 8.161810e-02  
## zipcode98199 8.176987e-01 9.060642e-01

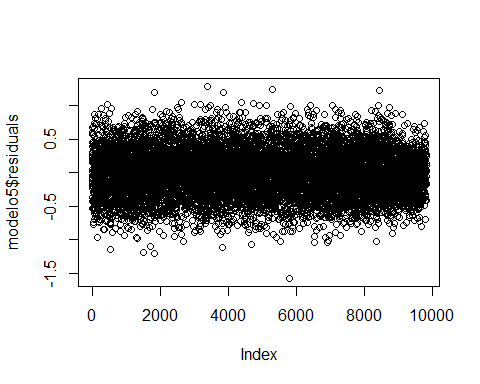
Vemos en este caso, que con muchas menos variables, somos capaces de explicar practicamente la misma variabilidad del modelo, sin embargo, el análisis de los residuos, vemos en el qqplot que se nos alejan de los extremos, con lo que el grado de predictividad del modelo no es del todo bueno

Eliminamos la variable del zipcode, el cual como podemos ver en los resultados, explicaba muy bien la variabilidad de los resultados, pero hacia que el modelo entrase en overfitting, de manera que ahoralos residuos se ajustan mucho mejor a la linea dentro del qqplot. Los residuos también están perfectamente ajustados a una normal con media 0.

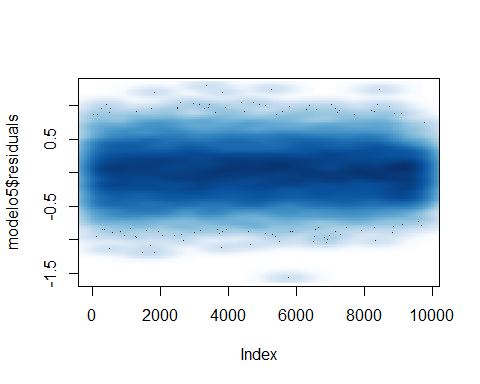
modelo5=lm(log(price)~sqft\_living+grade+sqft\_above+sqft\_living15+view+waterfront,data=house\_train)  
summary(modelo5)

##   
## Call:  
## lm(formula = log(price) ~ sqft\_living + grade + sqft\_above +   
## sqft\_living15 + view + waterfront, data = house\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.57690 -0.24186 0.00202 0.23175 1.29100   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.170e+01 3.438e-01 34.038 < 2e-16 \*\*\*  
## sqft\_living 2.474e-04 8.600e-06 28.766 < 2e-16 \*\*\*  
## grade3 1.256e-02 4.208e-01 0.030 0.976181   
## grade4 1.699e-01 3.576e-01 0.475 0.634682   
## grade5 3.676e-01 3.451e-01 1.065 0.286752   
## grade6 5.228e-01 3.438e-01 1.520 0.128440   
## grade7 7.319e-01 3.437e-01 2.129 0.033244 \*   
## grade8 9.197e-01 3.438e-01 2.675 0.007490 \*\*   
## grade9 1.132e+00 3.441e-01 3.289 0.001010 \*\*   
## grade10 1.305e+00 3.445e-01 3.788 0.000153 \*\*\*  
## grade11 1.431e+00 3.454e-01 4.142 3.47e-05 \*\*\*  
## grade12 1.533e+00 3.482e-01 4.403 1.08e-05 \*\*\*  
## grade13 1.655e+00 3.631e-01 4.557 5.26e-06 \*\*\*  
## sqft\_above -1.068e-04 9.385e-06 -11.375 < 2e-16 \*\*\*  
## sqft\_living15 7.480e-05 8.202e-06 9.120 < 2e-16 \*\*\*  
## view1 2.231e-01 2.836e-02 7.864 4.11e-15 \*\*\*  
## view2 1.406e-01 1.683e-02 8.352 < 2e-16 \*\*\*  
## view3 1.615e-01 2.296e-02 7.035 2.13e-12 \*\*\*  
## view4 2.744e-01 3.488e-02 7.866 4.05e-15 \*\*\*  
## waterfront1 4.024e-01 4.677e-02 8.605 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3436 on 9803 degrees of freedom  
## Multiple R-squared: 0.6141, Adjusted R-squared: 0.6134   
## F-statistic: 821.1 on 19 and 9803 DF, p-value: < 2.2e-16

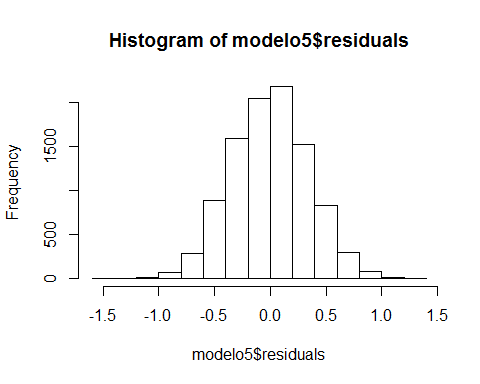
plot(modelo5$residuals)



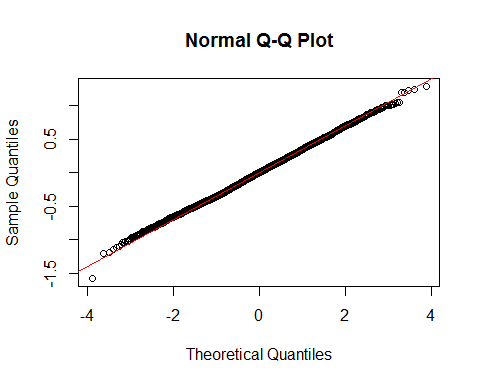
smoothScatter(modelo5$residuals)



hist(modelo5$residuals)



qqnorm(modelo5$residuals); qqline(modelo5$residuals,col=2)



confint(modelo5,level=0.95)

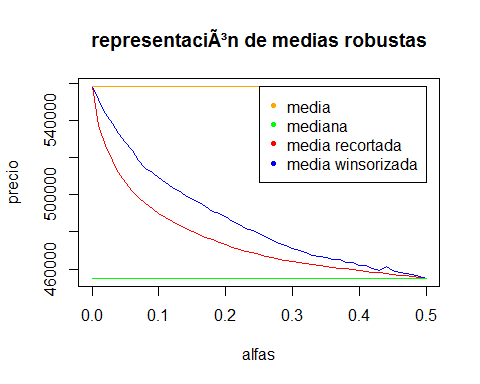
## 2.5 % 97.5 %  
## (Intercept) 1.102773e+01 1.237551e+01  
## sqft\_living 2.305211e-04 2.642351e-04  
## grade3 -8.123227e-01 8.374524e-01  
## grade4 -5.310522e-01 8.708879e-01  
## grade5 -3.087800e-01 1.043973e+00  
## grade6 -1.512144e-01 1.196741e+00  
## grade7 5.816708e-02 1.405726e+00  
## grade8 2.456949e-01 1.593641e+00  
## grade9 4.570935e-01 1.806091e+00  
## grade10 6.297366e-01 1.980283e+00  
## grade11 7.535835e-01 2.107639e+00  
## grade12 8.503923e-01 2.215409e+00  
## grade13 9.428711e-01 2.366511e+00  
## sqft\_above -1.251529e-04 -8.836023e-05  
## sqft\_living15 5.872503e-05 9.088028e-05  
## view1 1.674620e-01 2.786637e-01  
## view2 1.075940e-01 1.735825e-01  
## view3 1.164870e-01 2.064838e-01  
## view4 2.059833e-01 3.427242e-01  
## waterfront1 3.107617e-01 4.941229e-01

Vamos a intentar ver el comportamiento del modelo que mejor predicie los resultados (modelo 5) frente a un modelo robusto y comprarlos, para ver que efectivamente el comportamiento es similar, indicativo de que vamos por buen camino.

MediaWinsor<-function(x,probs=c(0.05,0.95)) {  
 xq<-quantile(x,probs=probs)  
 x[x < xq[1]]<-xq[1]  
 x[x > xq[2]]<-xq[2]  
 return(mean(x))  
}  
  
alphas=seq(from=0,to=0.5,by=0.01)  
medias\_recortadas=c()  
medias\_winsorizadas=c()  
for (alpha in alphas){  
 medias\_recortadas=c(medias\_recortadas,mean(house\_train$price, trim=alpha))  
 medias\_winsorizadas=c(medias\_winsorizadas,MediaWinsor(house\_train$price,probs=c(alpha,1-alpha)))  
}  
  
Estimadores=data.frame(alphas)  
Estimadores$media=mean(house\_train$price)  
Estimadores$mediana=median(house\_train$price)  
Estimadores$recortada=medias\_recortadas  
Estimadores$winsorizada=medias\_winsorizadas  
  
Estimadores

## alphas media mediana recortada winsorizada  
## 1 0.00 558446.4 455000 558446.4 558446.4  
## 2 0.01 558446.4 455000 537478.5 551154.6  
## 3 0.02 558446.4 455000 526418.1 543780.8  
## 4 0.03 558446.4 455000 518278.3 538635.7  
## 5 0.04 558446.4 455000 511824.7 532809.0  
## 6 0.05 558446.4 455000 506511.5 528353.4  
## 7 0.06 558446.4 455000 502062.9 523682.2  
## 8 0.07 558446.4 455000 498402.5 518480.7  
## 9 0.08 558446.4 455000 495336.7 514302.5  
## 10 0.09 558446.4 455000 492611.1 511840.7  
## 11 0.10 558446.4 455000 490147.0 509111.8  
## 12 0.11 558446.4 455000 487868.7 506808.3  
## 13 0.12 558446.4 455000 485784.7 504013.3  
## 14 0.13 558446.4 455000 483861.1 502185.0  
## 15 0.14 558446.4 455000 482015.2 499712.5  
## 16 0.15 558446.4 455000 480285.3 497571.9  
## 17 0.16 558446.4 455000 478665.9 495828.4  
## 18 0.17 558446.4 455000 477146.6 493640.8  
## 19 0.18 558446.4 455000 475717.7 491237.0  
## 20 0.19 558446.4 455000 474358.3 489799.1  
## 21 0.20 558446.4 455000 473081.3 488082.2  
## 22 0.21 558446.4 455000 471859.2 485772.7  
## 23 0.22 558446.4 455000 470738.6 484273.3  
## 24 0.23 558446.4 455000 469687.6 481789.7  
## 25 0.24 558446.4 455000 468702.4 480722.6  
## 26 0.25 558446.4 455000 467772.5 478948.7  
## 27 0.26 558446.4 455000 466908.8 477612.1  
## 28 0.27 558446.4 455000 466074.3 475555.0  
## 29 0.28 558446.4 455000 465336.1 473546.6  
## 30 0.29 558446.4 455000 464627.8 472457.8  
## 31 0.30 558446.4 455000 463983.9 471191.3  
## 32 0.31 558446.4 455000 463377.4 470273.1  
## 33 0.32 558446.4 455000 462821.1 468977.8  
## 34 0.33 558446.4 455000 462328.1 467598.5  
## 35 0.34 558446.4 455000 461855.7 466712.6  
## 36 0.35 558446.4 455000 461392.0 466070.0  
## 37 0.36 558446.4 455000 460977.6 465179.5  
## 38 0.37 558446.4 455000 460570.9 465331.0  
## 39 0.38 558446.4 455000 460111.1 463825.9  
## 40 0.39 558446.4 455000 459649.6 463782.0  
## 41 0.40 558446.4 455000 459183.9 461836.6  
## 42 0.41 558446.4 455000 458785.4 461810.1  
## 43 0.42 558446.4 455000 458392.8 460582.5  
## 44 0.43 558446.4 455000 458040.7 459295.4  
## 45 0.44 558446.4 455000 457677.8 461244.8  
## 46 0.45 558446.4 455000 457234.2 459273.3  
## 47 0.46 558446.4 455000 456793.7 458363.3  
## 48 0.47 558446.4 455000 456461.9 457437.5  
## 49 0.48 558446.4 455000 456055.6 456962.2  
## 50 0.49 558446.4 455000 455356.4 455987.1  
## 51 0.50 558446.4 455000 455000.0 455000.0

plot(Estimadores$recortada~Estimadores$alphas,ylim=c(min(Estimadores[,2:5]),max(Estimadores[,2:5])),col="red",type="l", main="representaciÃ³n de medias robustas", xlab="alfas",ylab="precio")  
lines(alphas,Estimadores$winsorizada,col="blue")  
lines(alphas,Estimadores$media,col="orange")  
lines(alphas,Estimadores$mediana,col="green")  
legend(x=mean(alphas),y=max(Estimadores[,2:5]),legend=c("media","mediana","media recortada","media winsorizada"),col=c("orange","green","red","blue"),pch=20)



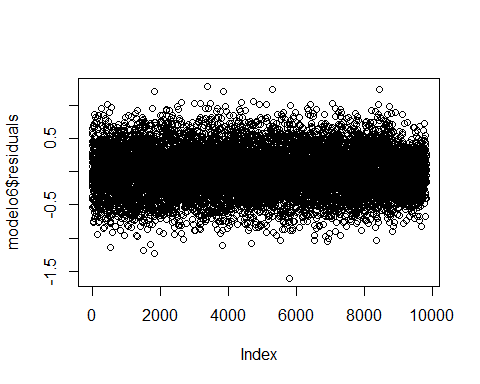
Generamos el modelo robusto para comprar

Vemos que el comporamiento de los residuos es idéntico, por lo cual, el modelo5 o el modelo6 son óptimos para hacer cualquier prediccion futura del precio respecto a las variables independientes encontradas.

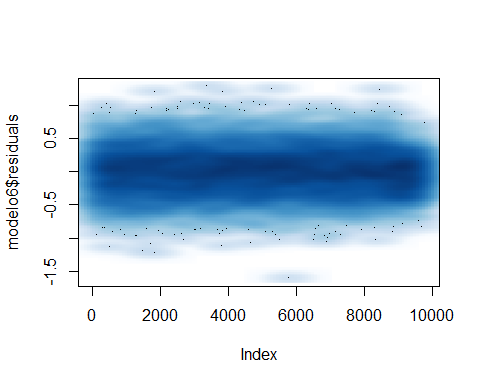
library(MASS)  
modelo6=rlm(log(price)~sqft\_living+grade+sqft\_above+sqft\_living15+view+waterfront,data=house\_train)  
summary(modelo6)

##   
## Call: rlm(formula = log(price) ~ sqft\_living + grade + sqft\_above +   
## sqft\_living15 + view + waterfront, data = house\_train)  
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.610285 -0.238336 0.004271 0.233071 1.292251   
##   
## Coefficients:  
## Value Std. Error t value   
## (Intercept) 11.7246 0.3545 33.0738  
## sqft\_living 0.0003 0.0000 29.5132  
## grade3 0.0068 0.4339 0.0156  
## grade4 0.2008 0.3687 0.5445  
## grade5 0.3831 0.3558 1.0767  
## grade6 0.5199 0.3545 1.4665  
## grade7 0.7306 0.3544 2.0612  
## grade8 0.9283 0.3545 2.6182  
## grade9 1.1464 0.3548 3.2309  
## grade10 1.3232 0.3552 3.7250  
## grade11 1.4595 0.3561 4.0980  
## grade12 1.5815 0.3590 4.4048  
## grade13 1.6946 0.3745 4.5255  
## sqft\_above -0.0001 0.0000 -12.9614  
## sqft\_living15 0.0001 0.0000 7.2583  
## view1 0.2206 0.0292 7.5426  
## view2 0.1446 0.0174 8.3330  
## view3 0.1592 0.0237 6.7256  
## view4 0.2829 0.0360 7.8661  
## waterfront1 0.3997 0.0482 8.2885  
##   
## Residual standard error: 0.3495 on 9803 degrees of freedom

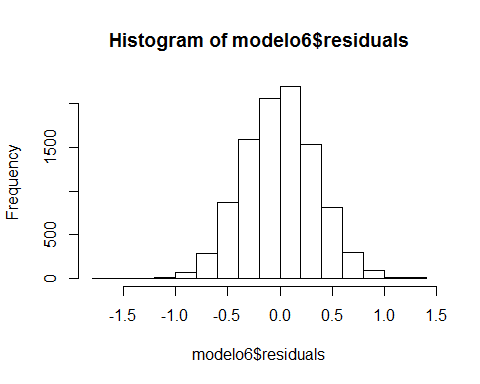
plot(modelo6$residuals)



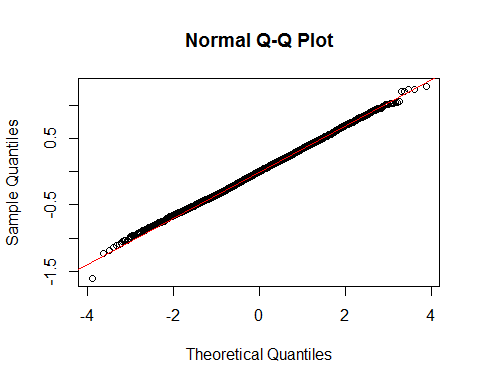
smoothScatter(modelo6$residuals)



hist(modelo6$residuals)



qqnorm(modelo6$residuals); qqline(modelo6$residuals,col=2)



confint(modelo6,level=0.95)

## 2.5 % 97.5 %  
## (Intercept) NA NA  
## sqft\_living NA NA  
## grade3 NA NA  
## grade4 NA NA  
## grade5 NA NA  
## grade6 NA NA  
## grade7 NA NA  
## grade8 NA NA  
## grade9 NA NA  
## grade10 NA NA  
## grade11 NA NA  
## grade12 NA NA  
## grade13 NA NA  
## sqft\_above NA NA  
## sqft\_living15 NA NA  
## view1 NA NA  
## view2 NA NA  
## view3 NA NA  
## view4 NA NA  
## waterfront1 NA NA

## Evaluacion del modelo

house\_train$prediccionLM=predict(modelo5,type="response")  
R2\_Train\_LM=1-sum((house\_train$price-house\_train$prediccionLM)^2)/sum((house\_train$price-mean(house\_train$price))^2)  
  
house\_test$prediccionLM=predict(modelo5,newdata=house\_test,type="response")  
R2\_Test\_LM=1-sum((house\_test$price-house\_test$prediccionLM)^2)/sum((house\_test$price-mean(house\_test$price))^2)  
  
R2\_Train\_LM

## [1] -1.838096

R2\_Test\_LM

## [1] -2.855937

Vemos que la diferencia del error en ambos conjuntos es bsatante pequeña, por lo que el comportamiento del modelo es aceptable > R2\_Train\_LM [1] -2.046132 > R2\_Test\_LM [1] -2.217989

Lo comparamos ahora con los resultados del modelo robusto

house\_train$prediccionRLM=predict(modelo6,type="response")  
R2\_Train\_RLM=1-sum((house\_train$price-house\_train$prediccionRLM)^2)/sum((house\_train$price-mean(house\_train$price))^2)  
  
house\_test$prediccionRLM=predict(modelo6,newdata=house\_test,type="response")  
R2\_Test\_RLM=1-sum((house\_test$price-house\_test$prediccionRLM)^2)/sum((house\_test$price-mean(house\_test$price))^2)  
  
R2\_Train\_RLM

## [1] -1.838096

R2\_Test\_RLM

## [1] -2.855937

R2\_Train\_RLM [1] -2.046132 R2\_Test\_RLM [1] -2.217989 Sale exactamente el mismo error que con el modelo de regresion lineal, por lo cual el comportamiento del modelo es estable.

SCE\_LM=sum((house\_test$price-house\_test$prediccionLM)^2) #MSE  
SCE\_RLM=sum((house\_test$price-house\_test$prediccionRLM)^2) #MSE  
  
SCE\_LM

## [1] 2.702975e+15

SCE\_RLM

## [1] 2.702975e+15

MAE\_LM=sum(abs(house\_test$price-house\_test$prediccionLM))  
MAE\_RLM=sum(abs(house\_test$price-house\_test$prediccionRLM))  
  
MAE\_LM

## [1] 3890632007

MAE\_RLM

## [1] 3890632021

## CONCLUSIONES A LOS REQUISITOS 1 y 2

* La relacion entre las variables puede decirse lineal, y representada como una ecuacion lineal (explicado anteriormente)
* Puesto que el precio no sigue una distribucion normal, lo mejor es transformarlo, por lo que la mejor relacion entre los factores que explican el precio, viene dada por la formula lineal con los factores del modleo4, que explican un 88% de la variabilidad. La interpretacion de aqui,puesto que se transforma el precio, es que el aumento de 1 unidad en cada variable afecta un % al precio
* El modelo que mejor predice un psible precio es el modelo5, donde los residuos se comportan mejor, y la comparativa tanto con el modelo robusto, como con el comportamiento entre los conjuntos de train y test, así como las distintas métricas, son muy similares, por tanto es un modelo estable.