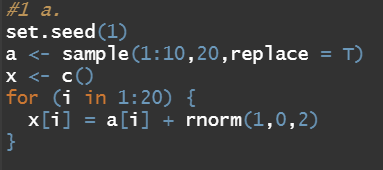
商業分析：SAS / R HW2

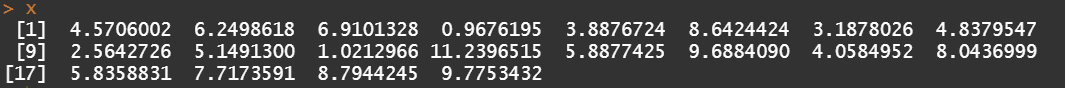
108208004 經濟三 白植允

1.[20pts] a. 生成一筆資料：

Xi=a+εXi=a+ε, i=1,...,20i=1,...,20

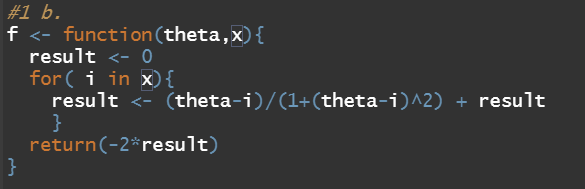
* aa 為0~10 任意數字。εε ∼N(0,2)∼N(0,2)
* 注意：XX必須在0~11內。





[20pts] b. Cauchy(θ, 1) 的密度函數，取log後一次微分如下，請寫出此function

f(θ)=−2∑i=1nθ−xi{1+(θ−xi)2}f(θ)=−2∑i=1nθ−xi{1+(θ−xi)2}



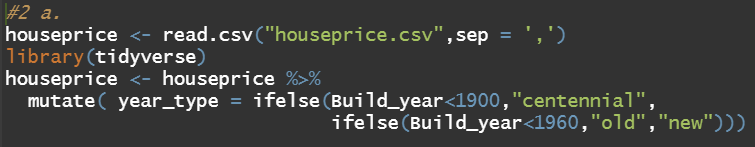
[10pts] c. 代入a生成的資料至b的function，並令θ=0.3θ=0.3。

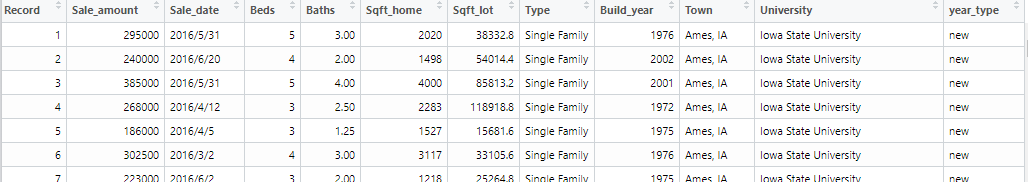




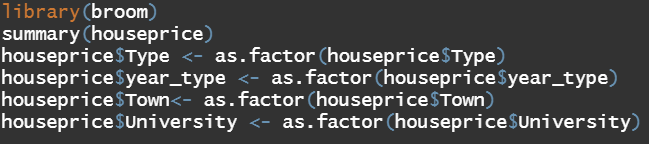
2.

[10pts] a. 根據Build\_year，建立一個新類別變數year\_type，1899年以前的房子為”centennial”，1900~1959年為”old”，1960年以上為”new”。



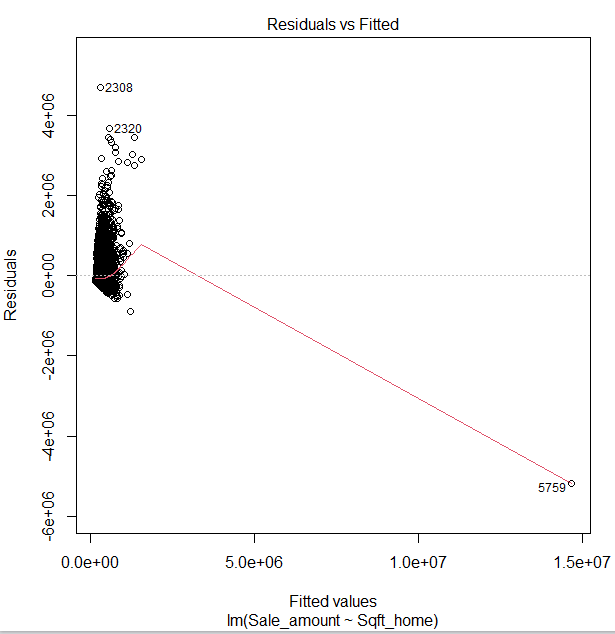


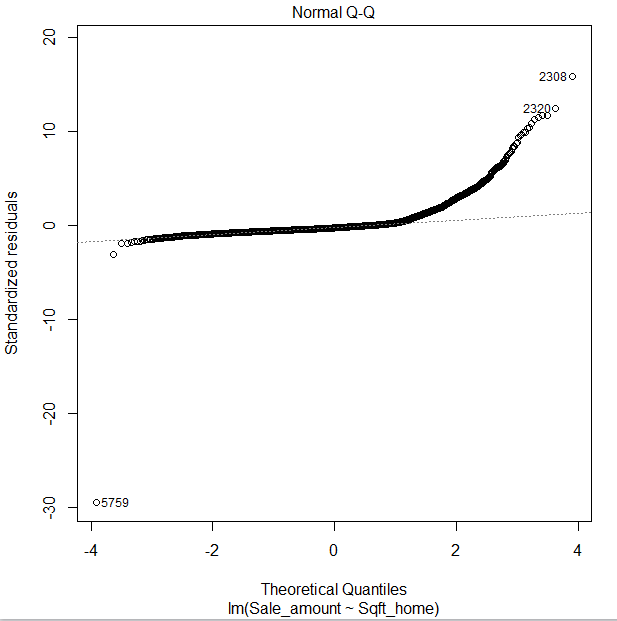
[40pts] b. 決定好你的最佳配適模型後，總結你的發現並根據解釋變數預測房屋價格。

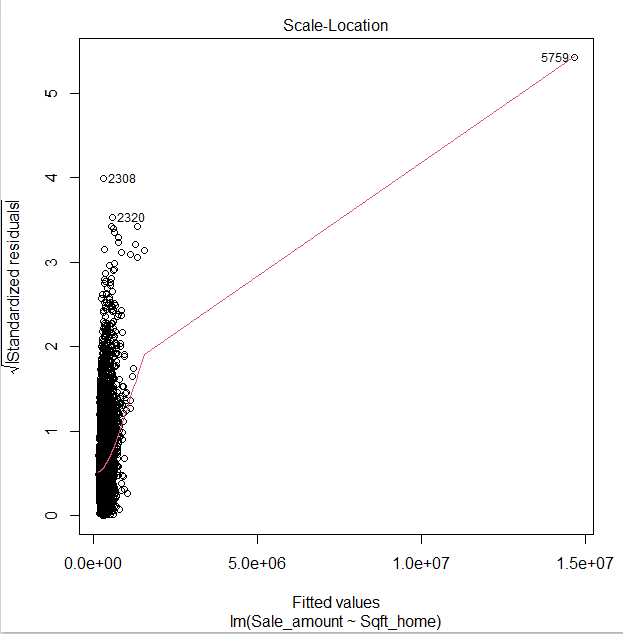


=>調整數據型態



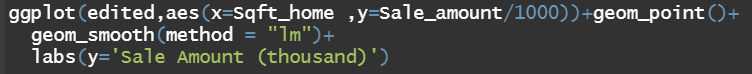


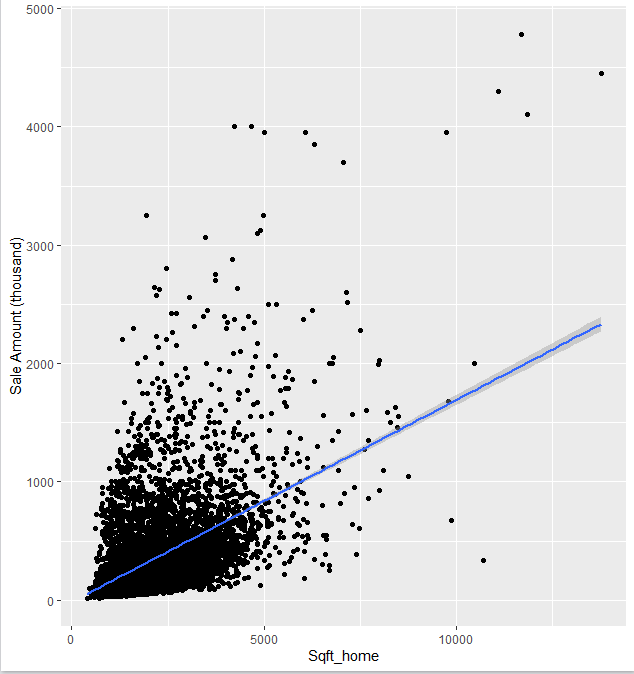




=>以Sqft\_home為參數做簡單線性迴歸，發現record :(2308,2320,5759)應該是outlier，且資料有heavy tailed的分布

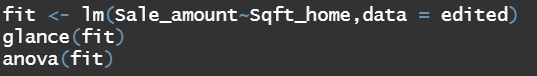


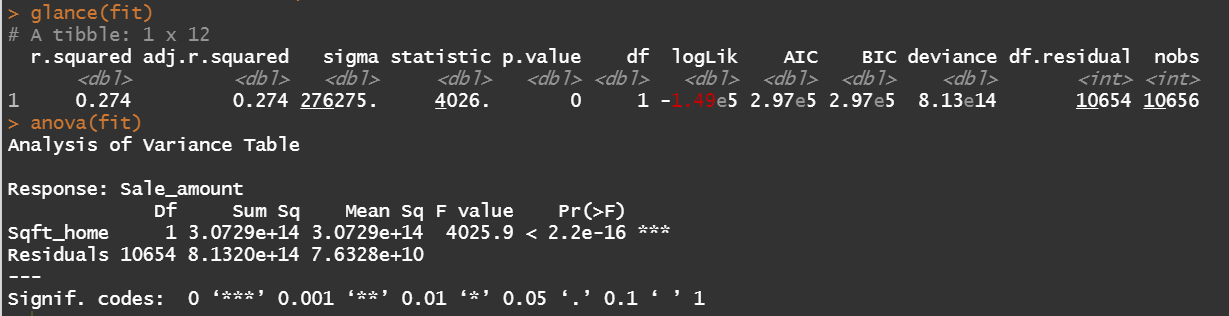
=>以edited代替原資料作分析



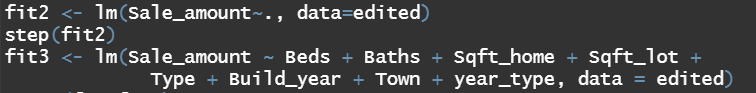
=>大致上有趨勢，看起來房子平方英尺數越大房屋價格也越大。

=>隨著Sqft\_home越來越大，變異數有越來越大的趨勢，模型有效性可能沒有到很好

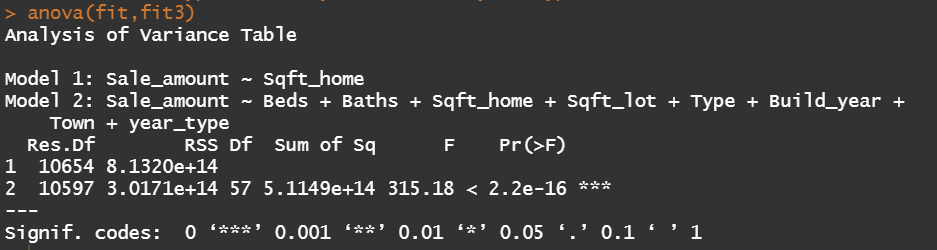




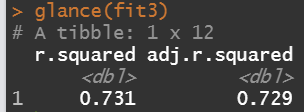
=>此模型能解釋約27.4%的房屋價格，而p-value小於0.001，有蠻高(>99.9%)的信心水準說明價格和房子多少平方英尺有線性關係



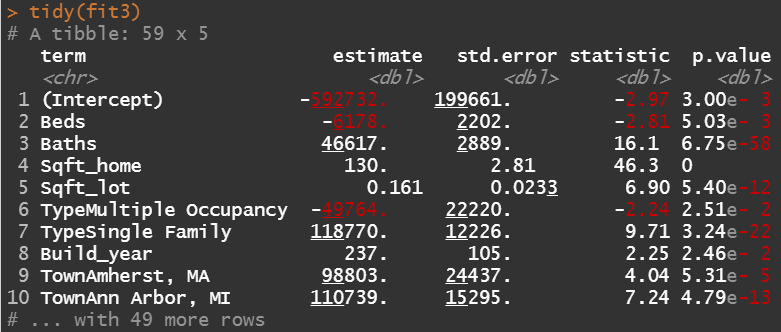
=>直接用電腦找出最適合的模型，其中有Beds , Baths , Sqft\_home , Sqft\_lot , Type , Build\_year , Town , year\_type做參數。



=>根據anova table，他的模型比較好(以beds,baths,sqft\_home,sqft\_lot,type,build\_year,town,year\_type做參數估計sale\_amount)，果然電腦比較厲害。



=>他的模型可以解釋約73%的房屋價格，他的模型真的比較好，看來這個模型比較適合做我的最佳配飾模型。



=>附上最適模型的部分參數估計表(Beds跟sale\_amount的關係好像不顯著，其中一個type也是)。

* 程式碼

#1 a.

set.seed(1)

a <- sample(1:10,20,replace = T)

x <- c()

for (i in 1:20) {

x[i] = a[i] + rnorm(1,0,2)

}

#1 b.

f <- function(theta,x){

result <- 0

for( i in x){

result <- (theta-i)/(1+(theta-i)^2) + result

}

return(-2\*result)

}

#1 c.

f(theta = 0.3,x)

#2 a.

houseprice <- read.csv("houseprice.csv",sep = ',')

library(tidyverse)

houseprice <- houseprice %>%

mutate( year\_type = ifelse(Build\_year<1900,"centennial",

ifelse(Build\_year<1960,"old","new")))

#2 b.

library(broom)

summary(houseprice)

houseprice$Type <- as.factor(houseprice$Type)

houseprice$year\_type <- as.factor(houseprice$year\_type)

houseprice$Town<- as.factor(houseprice$Town)

houseprice$University <- as.factor(houseprice$University)

plot(lm(Sale\_amount~Sqft\_home,data = houseprice))

edited <-houseprice[-c(2308,2320,5759),]

ggplot(edited,aes(x=Sqft\_home ,y=Sale\_amount/1000))+geom\_point()+

geom\_smooth(method = "lm")+

labs(y='Sale Amount (thousand)')

fit <- lm(Sale\_amount~Sqft\_home,data = edited)

glance(fit)

anova(fit)

fit2 <- lm(Sale\_amount~., data=edited)

step(fit2)

fit3 <- lm(Sale\_amount ~ Beds + Baths + Sqft\_home + Sqft\_lot +

Type + Build\_year + Town + year\_type, data = edited)

anova(fit,fit3)

glance(fit3)

tidy(fit3)