Assignment6 chapter3&4

Sun, Hao-Che

2018年10月25日

# 3.14

## (a)

set.seed(1)  
x1=runif(100)  
x2=0.5\*x1+rnorm(100)/10  
y=2+2\*x1+0.3\*x2+rnorm(100)

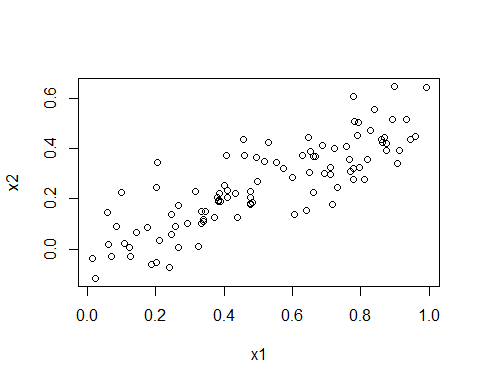
The

cor(x1,x2)

## [1] 0.8351212

## (b)

plot(x1,x2)



(c)

regg=lm(y~x1+x2)  
summary(regg)

##   
## Call:  
## lm(formula = y ~ x1 + x2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.8311 -0.7273 -0.0537 0.6338 2.3359   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.1305 0.2319 9.188 7.61e-15 \*\*\*  
## x1 1.4396 0.7212 1.996 0.0487 \*   
## x2 1.0097 1.1337 0.891 0.3754   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.056 on 97 degrees of freedom  
## Multiple R-squared: 0.2088, Adjusted R-squared: 0.1925   
## F-statistic: 12.8 on 2 and 97 DF, p-value: 1.164e-05

Through coding, the coefficients are close to the real parameters, and we can also discover that except for x2, we can reject the null hypothesis because of its p-value.

(d)(e)

regg2=lm(y~x1)  
summary(regg2)

##   
## Call:  
## lm(formula = y ~ x1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.89495 -0.66874 -0.07785 0.59221 2.45560   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.1124 0.2307 9.155 8.27e-15 \*\*\*  
## x1 1.9759 0.3963 4.986 2.66e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.055 on 98 degrees of freedom  
## Multiple R-squared: 0.2024, Adjusted R-squared: 0.1942   
## F-statistic: 24.86 on 1 and 98 DF, p-value: 2.661e-06

regg3=lm(y~x2)  
summary(regg3)

##   
## Call:  
## lm(formula = y ~ x2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.62687 -0.75156 -0.03598 0.72383 2.44890   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.3899 0.1949 12.26 < 2e-16 \*\*\*  
## x2 2.8996 0.6330 4.58 1.37e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.072 on 98 degrees of freedom  
## Multiple R-squared: 0.1763, Adjusted R-squared: 0.1679   
## F-statistic: 20.98 on 1 and 98 DF, p-value: 1.366e-05

In two cases, we can reject the null hypothesis because their p-values are extremely small.

(f)

No, because the x1 and x2 have collinearity, so in the condition of (c), the x2 might be replaced by x1, but if we predict by only one predictor, they are all significant.

(g)

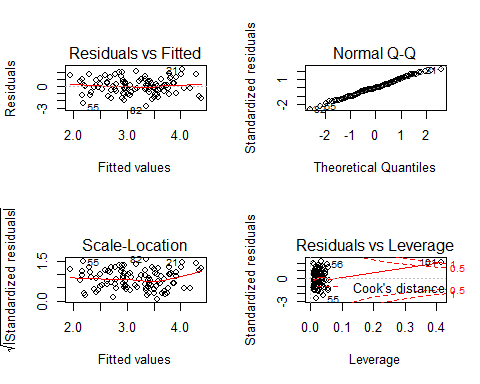
x1 = c(x1, 0.1)  
x2 = c(x2, 0.8)  
y = c(y, 6)  
lm(y~x1+x2)

##   
## Call:  
## lm(formula = y ~ x1 + x2)  
##   
## Coefficients:  
## (Intercept) x1 x2   
## 2.2267 0.5394 2.5146

summary(lm(y~x1+x2))

##   
## Call:  
## lm(formula = y ~ x1 + x2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.73348 -0.69318 -0.05263 0.66385 2.30619   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.2267 0.2314 9.624 7.91e-16 \*\*\*  
## x1 0.5394 0.5922 0.911 0.36458   
## x2 2.5146 0.8977 2.801 0.00614 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.075 on 98 degrees of freedom  
## Multiple R-squared: 0.2188, Adjusted R-squared: 0.2029   
## F-statistic: 13.72 on 2 and 98 DF, p-value: 5.564e-06

par(mfrow=c(2,2))  
plot(lm(y~x1+x2))



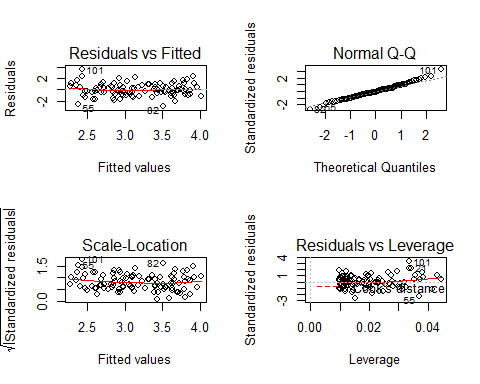
lm(y~x1)

##   
## Call:  
## lm(formula = y ~ x1)  
##   
## Coefficients:  
## (Intercept) x1   
## 2.257 1.766

summary(lm(y~x1))

##   
## Call:  
## lm(formula = y ~ x1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.8897 -0.6556 -0.0909 0.5682 3.5665   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.2569 0.2390 9.445 1.78e-15 \*\*\*  
## x1 1.7657 0.4124 4.282 4.29e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.111 on 99 degrees of freedom  
## Multiple R-squared: 0.1562, Adjusted R-squared: 0.1477   
## F-statistic: 18.33 on 1 and 99 DF, p-value: 4.295e-05

par(mfrow=c(2,2))  
plot(lm(y~x1))



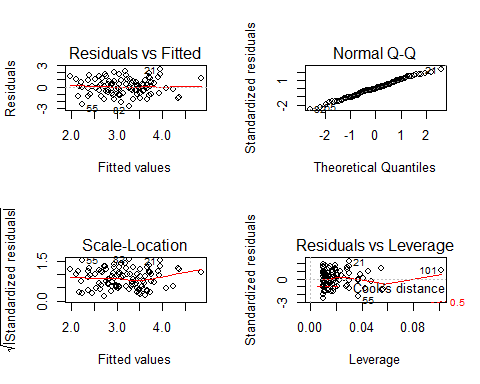
lm(y~x2)

##   
## Call:  
## lm(formula = y ~ x2)  
##   
## Coefficients:  
## (Intercept) x2   
## 2.345 3.119

summary(lm(y~x2))

##   
## Call:  
## lm(formula = y ~ x2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.64729 -0.71021 -0.06899 0.72699 2.38074   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.3451 0.1912 12.264 < 2e-16 \*\*\*  
## x2 3.1190 0.6040 5.164 1.25e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.074 on 99 degrees of freedom  
## Multiple R-squared: 0.2122, Adjusted R-squared: 0.2042   
## F-statistic: 26.66 on 1 and 99 DF, p-value: 1.253e-06

par(mfrow=c(2,2))  
plot(lm(y~x2))

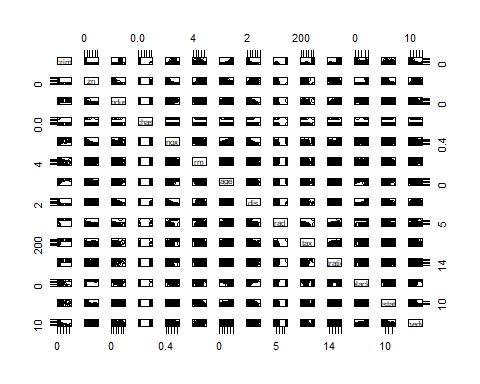


Above of these graphics and models, we can discover that the phenomenon happened in the last questions also happened, and except for y~x1, the other two models don’t have apparent outliers, and the first and third model has high-leverage.

3.15

(a)

library(MASS)  
attach(Boston)  
plot(Boston)

 ##(b)

multi=lm(crim~.,data=Boston)  
summary(multi)

##   
## Call:  
## lm(formula = crim ~ ., data = Boston)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -9.924 -2.120 -0.353 1.019 75.051   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 17.033228 7.234903 2.354 0.018949 \*   
## zn 0.044855 0.018734 2.394 0.017025 \*   
## indus -0.063855 0.083407 -0.766 0.444294   
## chas -0.749134 1.180147 -0.635 0.525867   
## nox -10.313535 5.275536 -1.955 0.051152 .   
## rm 0.430131 0.612830 0.702 0.483089   
## age 0.001452 0.017925 0.081 0.935488   
## dis -0.987176 0.281817 -3.503 0.000502 \*\*\*  
## rad 0.588209 0.088049 6.680 6.46e-11 \*\*\*  
## tax -0.003780 0.005156 -0.733 0.463793   
## ptratio -0.271081 0.186450 -1.454 0.146611   
## black -0.007538 0.003673 -2.052 0.040702 \*   
## lstat 0.126211 0.075725 1.667 0.096208 .   
## medv -0.198887 0.060516 -3.287 0.001087 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.439 on 492 degrees of freedom  
## Multiple R-squared: 0.454, Adjusted R-squared: 0.4396   
## F-statistic: 31.47 on 13 and 492 DF, p-value: < 2.2e-16

We can discover that zn, dis, rad, black, medv have significant relationship, so the null hypothesis of these predictors can be rejected.