STAT3032

Mingming Xu

2019/2/20

Problem 1

```r
library(alr4)
```

```
## Loading required package: car
## Loading required package: carData
## Loading required package: effects
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```r
data=oldfaith
```

(a) The explanatory and response variables are Duration and Interval.

```r
names(oldfaith)
```

```
## [1] "Duration" "Interval"
```

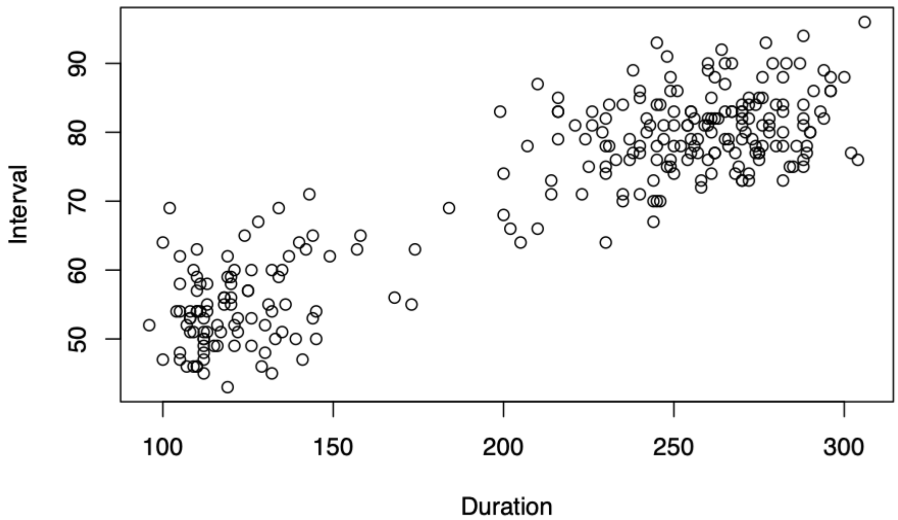
(b)

```r
model1=lm(Interval~Duration, data = oldfaith)
summary(model1)
```

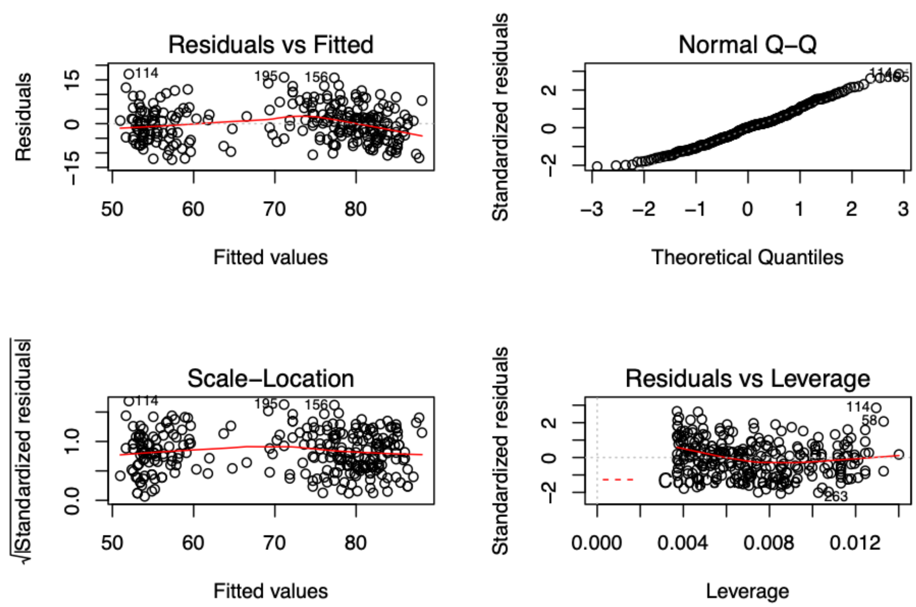
```
##
## Call:
## lm(formula = Interval ~ Duration, data = oldfaith)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.3337  -4.5250   0.0612   3.7683  16.9722
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.987808   1.181217   28.77   <2e-16 ***
## Duration     0.176863   0.005352   33.05   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.004 on 268 degrees of freedom
## Multiple R-squared:  0.8029, Adjusted R-squared:  0.8022
## F-statistic:  1092 on 1 and 268 DF,  p-value: < 2.2e-16
```

(c)Four assumptions to check :

```r
plot(Interval~Duration, data = oldfaith)
```




```r
par(mfrow=c(2,2))
plot(model1)
```



Linearity:Yes.We could see it from scatterplot.

Independence:Data information" We can assume this data was collected in an independent manner as multiple volunteers recorded the data for this month. "

Normality:Yes. From the normal QQ plot, the observations follow the line perfectly.

Equal variance: Yes.From the residuals vs Fitted plot, no discernible pattern.

(d)$\hat{Interval} = 33.987808 + 0.176863 * Duration$

(e)

(i)$S = 6.004$

(ii)Because $SE(\hat{\beta}_1) = S/sqrt(SXX)$, $SXX = 1258488$

```
SXX=(6.004/(0.005352))^2
(6.004/(0.005352))*(6.004/(0.005352))
```

```
## [1] 1258488
```

(iii) Because $SXY = SXX * \hat{\beta}_1$, $SXY = 7115611$

```
SXY=SXX/0.176863
SXX/0.176863
```

```
## [1] 7115611
```

(f)

(i)Coefficient hypothesis test:

$$H_0 : \beta_1 = 1 \quad vs. \quad H_A : \beta_1 \neq 1$$

(ii)Test statistic : $t = (\hat{\beta}_1 - 1)/SE(\hat{\beta}_1)$ ~t_{n-2}

```
(0.171863-1)/0.005352
```

```
## [1] -154.7341
```

(iii) P-value

```
(1-pt(154.7341,268))*2
```

```
## [1] 0
```

(iv) Conclusion:

Because the p-value is less than alpha 0.05, we reject the null hypothesis.

(g) confidence

(h) prediction

(i) The prediction interval is wider.Because the standard error of point estimate of prediction interval is larger than the other.

Problem 2

(d) Because $R^2 = SSreg/SST = 1 - RSS/SST$,with R^2larger, SSreg larger and RSS smaller, it means that the fitted modle better ,having stronger explaination ability to the relationship between response and explonary variables.

Problem 3

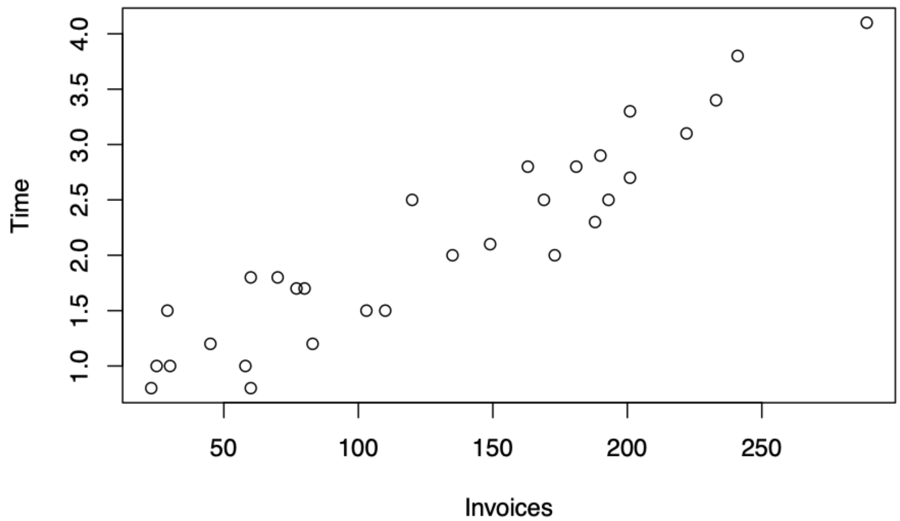
```
dat=read.table('http://gattonweb.uky.edu/sheather/book/docs/datasets/invoices.txt', sep='\t',header=TRU

names(dat)
```

```
## [1] "Day"      "Invoices" "Time"
```

(a)

```r
plot(Time~Invoices, data=dat)
```



(b)

```r
model2=lm(Time~Invoices, data=dat)
summary(model2)
```

```
##
## Call:
## lm(formula = Time ~ Invoices, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59516 -0.27851  0.03485  0.19346  0.53083
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6417099  0.1222707   5.248 1.41e-05 ***
## Invoices    0.0112916  0.0008184  13.797 5.17e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3298 on 28 degrees of freedom
## Multiple R-squared:  0.8718, Adjusted R-squared:  0.8672
## F-statistic: 190.4 on 1 and 28 DF,  p-value: 5.175e-14
```

(c) $RSS = (n-2) * \hat{\sigma^2}$

```r
28*0.3298*0.3298
```

```
## [1] 3.045505
```

(d)

```r
confint(model2, level = 0.95)
```

```
##                   2.5 %     97.5 %
## (Intercept) 0.391249620 0.89217014
## Invoices    0.009615224 0.01296806
```

(e)

```r
new.dat=data.frame(Invoices=30)
predict(model2,newdata = new.dat,interval = 'prediction',level=0.95)
```

```
##         fit       lwr      upr
## 1 0.9804592 0.2736022 1.687316
```

(f) It is possible. Because 95% of the prediction interval contains response values instead of all response values in it.We have 5% ppossibility that response values could be outside the prediction interval.

Problem 4

```r
library(alr4)
data("Rateprof")
```

(a) The value of quality is 1.674419, the value of clarity is 3.360465.

```r
Rateprof[76,]
```

```
##    gender numYears numRaters numCourses pepper discipline    dept  quality
## 76 female       11        86          8     no        Hum English 1.674419
##    helpfulness  clarity easiness raterInterest sdQuality sdHelpfulness
## 76    1.686047 3.360465  1.55814      2.814815  1.075851      1.180843
##    sdClarity sdEasiness sdRaterInterest
## 76  1.058417  0.8059288        1.225878
```

(b) $h_{ii} = 1/n + (x_i - \bar{x})^2/SXX$

```r
mod=lm(quality~clarity,data = Rateprof)
summary(mod)
```

```
##
## Call:
## lm(formula = quality ~ clarity, data = Rateprof)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.74449 -0.09193  0.00517  0.10351  0.45234
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.22097    0.04040   5.469 8.43e-08 ***
## clarity      0.95163    0.01114  85.435  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1828 on 364 degrees of freedom
## Multiple R-squared:  0.9525, Adjusted R-squared:  0.9524
## F-statistic:  7299 on 1 and 364 DF,  p-value: < 2.2e-16
```

```r
mean(Rateprof$clarity)
```

```
## [1] 3.524552
```

```r
xbar=mean(Rateprof$clarity)
n=nrow(Rateprof)
SXX=(0.1828/0.01114)^2
hii=1/n+(3.360465-xbar)^2/SXX
1/366+(3.360465-xbar)^2/SXX
```

```
## [1] 0.002832232
```

SO $h_{ii} = 0.01170816$

(c)

```r
hatvalues(mod)[76]
```

```
##          76
## 0.002832233
```

My result is consistent with my (b).

(d) $D_i = (r_i^2/2) * (h_{ii}/(1 - h_{ii}))$

$r_i = \hat{e}_i/(s * sqrt(1 - h_{ii}))$

$\hat{e}_i = y_i - \hat{y}_i$

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 * x_i$

```r
yihat=0.22097+0.95163*3.360465
eihat=1.674419-yihat
ri=eihat/(0.1828*sqrt(1-hii))
Di=ri^2/2*(hii/(1-hii))
ri^2/2*(hii/(1-hii))
```

```
## [1] 0.1296992
```

(e)

```r
cooks.distance(mod)[76]
```

```
##       76
## 0.129732
```

My result is consistent with my (d).