Stat3032HW1

Mingming Xu

2019/2/13

Problem 1

```r
library(alr4)
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

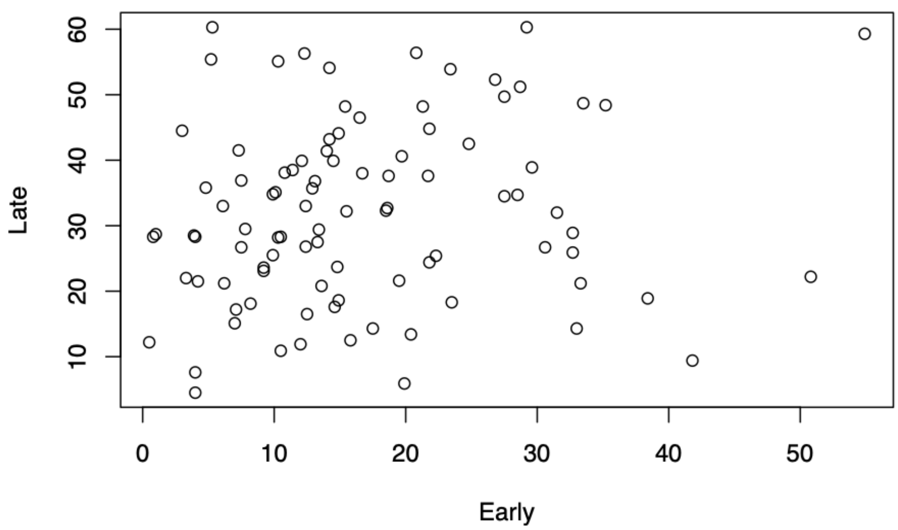
```
## Loading required package: effects
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```r
data(ftcollinssnow)
```

Part a

```r
plot(Late~Early,data=ftcollinssnow )
```



Part b

```r
model=lm(Late~Early,data = ftcollinssnow)
summary(model)
```

```
##
## Call:
## lm(formula = Late ~ Early, data = ftcollinssnow)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.7416  -9.3898  -0.1393   8.8177  30.5857
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  28.6358     2.6149  10.951   <2e-16 ***
## Early         0.2035     0.1310   1.553    0.124
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.72 on 91 degrees of freedom
## Multiple R-squared:  0.02581,    Adjusted R-squared:  0.01511
## F-statistic: 2.411 on 1 and 91 DF,  p-value: 0.1239
```

Part c

Test statistic: t=10.951

P-value:<2e-16

Conclusion:p value<0.05, reject the null hypothesis.

Part d

Test statistic:t=0.2035

P-value=0.124

Conclusion:p value>0.05, refuse to reject the null hypothesis.

Part e

I believe that we cannot use it.Because in the part c, we refuse to reject the hypothesis beta1= 0 .Hence, we cannot use the coefficent beta1 value to predict late season snowfall.

Problem 2

```
xbar=mean(ftcollinssnow$Early)
ybar=mean(ftcollinssnow$Late)
SXY=sum((ftcollinssnow$Early-xbar)*(ftcollinssnow$Late-ybar))
SXX=sum((ftcollinssnow$Early-xbar)^2)
RSS=sum((ftcollinssnow$Late-28.6358-0.2035*ftcollinssnow$Early)^2)
Betahat1=SXY/SXX
Betahat0=ybar-Betahat1*xbar
sigmahatsq=RSS/91
```

xbar=16.7440860215054

ybar=32.0430107526882

SXY=2229.01365591398

SXX=10954.0692473118

RSS=17118.83205735

Betahat1=0.203487270856991

Betahat0=28.6358023861774

sigmahatsq=188.119033597253

Problem 3

SXX=$\sum_{i=1}^{n}(x_i-\bar{x})^2$=$\sum_{i=1}^{n}(x_i^2-2*x_i*\bar{x}+\bar{x}^2)$=$\sum_{i=1}^{n}x_i^2-2\sum_{i=1}^{n}x_i*\bar{x}+\sum_{i=1}^{n}\bar{x}^2$=$\sum_{i=1}^{n}x_i^2-2*n*\bar{x}*\bar{x}+n*\bar{x}^2$=$\sum_{i=1}^{n}x_i^2-n*\bar{x}^2$=$\sum_{i=1}^{n}x_i^2-n*\bar{x}*\bar{x}$=$\sum_{i=1}^{n}x_i^2$-$\sum_{i=1}^{n}x_i*\bar{x}$=$\sum_{i=1}^{n}x_i*(x_i-\bar{x})$

Problem 4

(a)

```
tcrit=qt(0.975,28)
0.6417099+c(-1,1)*tcrit*0.1222707
```

```
## [1] 0.3912497 0.8921701
```

(b)

Test statistic:

```
t=(0.0112916-0.01)/0.0008184
```

P-value:

```
2*(1-pt(t,28))
```

```
## [1] 0.1257517
```

Conclusion:

P-value is larger than 0.05,the value of alpha, so refuse to reject the null hypothesis.There is no sufficient evidence that processing time has a relationship with the number of invoices.

(c)

For 130 invoices,$\hat{Time} = 0.6417099 + 0.0112916 * Invoices$, the point estimate is 0.6417099+0.112916*130=15.32079

```
xbar=130
tcrit=qt(0.975,28)
sxx=(0.3928/0.0008184)^2
sepred=0.3298*sqrt(1+1/30+((130-xbar)^2)/sxx)
(0.6417099+0.0112916*130)+c(-1,1)*tcrit* sepred
```

```
## [1] 1.422886 2.796350
```