ST5202_Tut_1

Zhu Xu
User ID:E0337988
Matriculation ID:A0191344H

10/02/2019

Question_1:

```r
data_copier = read.table(
  "/Users/xuzhu/Desktop/Notes/Sem2/ST5202-Applied_Regression_Analysis/Tut/copier_maintenance.txt")
colnames(data_copier) = c("Y","X")
x = data_copier$X
y = data_copier$Y
reg = lm(Y~X, data =data_copier)
summary(reg)
```

```
##
## Call:
## lm(formula = Y ~ X, data = data_copier)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.7723  -3.7371   0.3334   6.3334  15.4039
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.5802     2.8039  -0.207    0.837
## X            15.0352     0.4831  31.123   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.914 on 43 degrees of freedom
## Multiple R-squared:  0.9575, Adjusted R-squared:  0.9565
## F-statistic: 968.7 on 1 and 43 DF,  p-value: < 2.2e-16
```

a)

The estimated regression function is:

$$b_0 = -0.5802; \quad b_1 = 15.0352$$

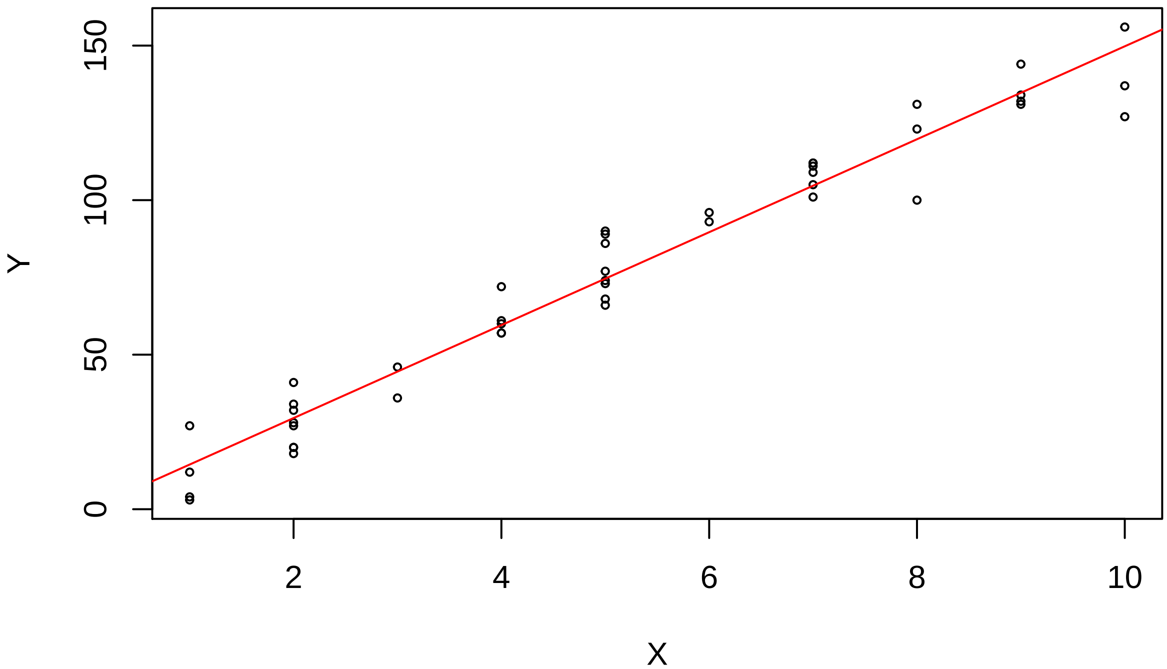
$$Y_i = b_0 + b_1 X_i + \epsilon_i \Rightarrow f(x) = -0.5802 + 15.0352x$$

b)

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

```
##                  2.5 %    97.5 %
## (Intercept) -6.234843  5.074529
## X           14.061010 16.009486
```

```r
b_0 = -0.5802
b_1 = 15.0352
x_new = seq(0,11, 0.1)
plot(data_copier$X,data_copier$Y, type="p", cex=0.5,  xlab="X", ylab="Y")
f = b_1*x_new+b_0
lines(x_new, f, lty=1, col="red", lwd=1)
```



Question_2:

a)

```
b_0 = -0.5802
b_1 = 15.0352
s2 = sum((y-(b_0+b_1*x))^2)/(45-2)
s_b_1 = sqrt(s2/sum((x-mean(x))^2))
s_b_1
```

```
## [1] 0.4830872
```

```
bound = qt(0.95, 43)   # 90%
c(b_1-bound*s_b_1, b_1+bound*s_b_1)
```

```
## [1] 14.2231 15.8473
```

b)

$H_0 : \beta_1 = 0; \quad H_a : \beta_1 \neq 0$

From a) we know $b_1 = 15.0352; \quad s\{b_1\} = 0.4831$

Thus $t^* = \frac{b_1 - \beta_1}{s\{b_1\}} = \frac{15.0352-0}{0.4831} = 31.122 > 1.6811 \quad \Rightarrow$ reject $H_0 : \beta_1 = 0$

c)

Yes

d)

$H_0 : \beta_1 \leq 14; \quad H_a : \beta_1 > 14$

$t^* = \frac{15.0352-14}{0.4831} = 2.1428 > 1.681$ conclude H_a, p-value=0.189

Question_3:

a)

```
X_h = 6
hat_Y_h = b_0+b_1*X_h
hat_Y_h # \hat{Y_h}=89.631
```

```
## [1] 89.631
```

```
n=45
MSE = sum((y-(b_0+b_1*x))^2)/(n-2)
s_hat_Y_h = sqrt(MSE*(1/n+(X_h-mean(x))^2/sum((x-mean(x))^2)))
s_hat_Y_h # s{\hat{Y_h}}=1.3964
```

```
## [1] 1.396411
```

```
c(hat_Y_h-bound*s_hat_Y_h, hat_Y_h+bound*s_hat_Y_h)
```

```
## [1] 87.28353 91.97847
```

```
# 87.2835\leq E{Y_h}\leq 91.9785
```

b)

```
s_pred = sqrt(MSE+s_hat_Y_h^2)
s_pred # s{pred}=9.022228
```

```
## [1] 9.022228
```

```
c(hat_Y_h-bound*s_pred, hat_Y_h+bound*s_pred)
```

```
## [1]  74.464 104.798
```

```
# 74.464\leq E[Y_{h(new)}]\leq 104.798
```

c)

$\frac{87.28353}{6} \leq c \leq \frac{91.97847}{6}, \quad \Rightarrow 14.5473 \leq c \leq 15.3298$

Question_4:

a)

```
aov.copier = aov(reg)
summary(aov.copier)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## X            1  76960   76960   968.7 <2e-16 ***
## Residuals   43   3416      79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

b)

```
hat_Y = b_0+b_1*x
SSR = sum((hat_Y-mean(y))^2) # SSR=76959.93
MSR = SSR/1 # MSR=76959.93
SSE = sum((y-hat_Y)^2) # SSE=3416.377
MSE = SSE/(n-2) # MSE=79.45063
# H_0:\beta_1=0; H_a:\beta_1=1
test_F = MSR/MSE # test_F=968.651
value = qf(0.9, 1, 43) # value=2.825999
if(test_F<=value){
  print("conclude H_0")
} else{
  print("conclude H_1")
}
```

```
## [1] "conclude H_1"
```

c)

95.75%, coefficient of determination.

d)

```
SSTO = SSE + SSR # SSTO=80376.31
R_square = SSR/SSTO
if(b_1>0){
  r = sqrt(R_square)
  paste("r=", r)
} else{
```

```
  r = -sqrt(R_square)
  paste("r=", r)
}
```

```
## [1] "r= 0.978516848778707"
```

Question_5:

a)

Prediction

b)

Mean response

c)

Prediction

Question_6:

a)

$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n}(Y_i - \hat{Y}_i) = \sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i) = \sum_{i=1}^{n} Y_i - nb_0 - b_1 \sum_{i=1}^{n} X_i = 0$

b)

$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n}(Y_i - \hat{Y}_i) = 0 \quad \Rightarrow \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$

c)

$\sum_{i=1}^{n} X_i e_i = \sum_{i=1}^{n} X_i(Y_i - b_0 - b_1 X_i) = \sum_{i=1}^{n} X_i Y_i - b_0 \sum_{i=1}^{n} X_i - b_1 \sum_{i=1}^{n} X_i^2 = 0$

d)

$\sum_{i=1}^{n} \hat{Y}_i e_i = \sum_{i=1}^{n}(b_0 + b_1 X_i)e_i = b_0 \sum_{i=1}^{n} e_i + b_1 \sum_{i=1}^{n} X_i e_i = 0$

e)

$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{\sum_{i=1}^{n} \hat{Y}_i}{n} = \frac{\sum_{i=1}^{n}(b_o + b_1 X_i)}{n} = b_0 + \frac{b_1 \sum_{i=1}^{n} X_i}{n} = b_0 + b_1 \bar{X}$
Thus the regreesion line always pass through the point$(\bar{X}, \bar{Y})$

Question_7:

a)

Let $Q = \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i) \quad \Rightarrow minimizing \quad \Rightarrow \frac{\partial Q}{\partial \beta_0} = \frac{\partial Q}{\partial \beta_1} = 0$

$\Rightarrow b_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}; \quad b_0 = \bar{Y} - b_1 \bar{X}$

Since let $k_i = \frac{X_i - \bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}, \quad \sum k_i = 0, \quad \Rightarrow b_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \sum k_i(Y_i - \bar{Y}) = \sum k_i Y_i$

The Y_i's are independently, normally distributed, then a linear combination of Y_i's is also normally distributed, thus b_1 is normally distributed, b_0 is also a normally distributed.

b)

$E[b_0] = E[\bar{Y} - b_1 \bar{X}] = \bar{Y} - \bar{X}E[b1] = \bar{Y} - \bar{X}\beta_1 = \beta_0$

c)

Since $Var[b_1] = Var[\sum k_i Y_i] = \sum k_i^2 Var[Y_i] = \sigma^2 \frac{1}{\sum(X_i - \bar{X})^2}$

Thus:

$Var[b_0] = Var[\bar{Y} - b_1 \bar{X}] = Var[\frac{\sum Y_i}{n}] + \bar{X}^2 Var[b_1] = \frac{\sigma^2}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \sigma^2(\frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2})$

d)

$b_0 \sim N(\beta_0, \sigma^2(\frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2}))$