Assignment2 STAT 353

Zimeng Ming V00844078

2019/10/7

Question 1 and 3 is on the paper attached.

Question2:
a)

```
gpa<-read.table("2.7.txt",header = T)

fit1<-lm(gpa$GPA~gpa$GMAT)
summary(fit1)
```

```
##
## Call:
## lm(formula = gpa$GPA ~ gpa$GMAT)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.98608 -0.25048 -0.04539  0.47659  0.64531
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.157611   2.014430   1.071    0.309
## gpa$GMAT    0.001931   0.003510   0.550    0.594
##
## Residual standard error: 0.5326 on 10 degrees of freedom
## Multiple R-squared:  0.02937,    Adjusted R-squared:  -0.06769
## F-statistic: 0.3026 on 1 and 10 DF,  p-value: 0.5943
```

So the R_square is 0.02937 the model is GPA=2.157611+0.001931(GMAT)+e

b):
the second person fitted value is: 2.157611+0.001931(540)=3.20034
so the estimate GPA for seconde person is: 3.2

c):

```
anova(fit1)
```

```
## Analysis of Variance Table
##
## Response: gpa$GPA
##           Df  Sum Sq  Mean Sq F value Pr(>F)
## gpa$GMAT   1 0.08585 0.085847  0.3026 0.5943
## Residuals 10 2.83698 0.283698
```

The p-value is significantly lager then 0.05, which is 0.5943, so the sigificant eveidence shows that do not reject H0:the beta1 is 0, so we can conclude that the GMAT is not an important predictor variable.

Question 4 a)
Assumptions:
** x is fixed
** E(Ei)=0
** V(Ei)= Theata^2 which is a constant
** cov(ei,ej)=0 for i not equal to j(Ei are independent)

```
data<-read.table("2.12.txt",header = T)


fit2<-lm(data$y~data$x)
summary(fit2)
```

```
##
## Call:
## lm(formula = data$y ~ data$x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16217 -0.10178 -0.07266  0.03979  0.49064
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.228090   0.137840  -1.655    0.137
## data$x       0.994757   0.005219 190.585 6.43e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2067 on 8 degrees of freedom
## Multiple R-squared:  0.9998, Adjusted R-squared:  0.9998
## F-statistic: 3.632e+04 on 1 and 8 DF,  p-value: 6.429e-16
```

So the model is y=-0.228090+0.994757x+e

b):

```
confint(fit2, level = 0.95)
```

```
##                  2.5 %     97.5 %
## (Intercept) -0.5459503 0.08977054
## data$x       0.9827204 1.00679271
```

The 95% confidence interval for the intercept(beta0) is : (-0.5459503,0.08977054)

c): The 95% confidence interval for the slope of the model(beta1) is : (0.9827204 ,1.00679271)

d): i): If the x=0, y=0.228090, which is not a no calcium present. also the CI for the intercept(beta0) contains zero, also shows there should be no corresponding result. ii): As 1 is included in the confidence interval for slope(beta1), so the slope could be 1.

e):

```
#we need beta0 is 0 so we use the zero intercept.
fit3<-lm(data$y~0+data$x)
summary(fit3)
```

```
##
## Call:
## lm(formula = data$y ~ 0 + data$x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.24861 -0.19054 -0.09167  0.00104  0.49827
##
## Coefficients:
##        Estimate Std. Error t value Pr(>|t|)
## data$x 0.987153   0.002704   365.1   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2258 on 9 degrees of freedom
## Multiple R-squared:  0.9999, Adjusted R-squared:  0.9999
## F-statistic: 1.333e+05 on 1 and 9 DF,  p-value: < 2.2e-16
```

```
confint(fit3, level = 0.95)
```

```
##            2.5 %    97.5 %
## data$x 0.9810362 0.9932693
```

Comment: The new model line is y=0.987153x+e
here the new 95% confidence interval for the slope of the model(beta1) is : (0.9810362, 0.9932693)

f): As we change the model, we remove the intercept term so the slope will change either. The new CI for the beta1 is not include 1 since the model slope is change.