ACST 890- QUIZ 1 SOLUTIONS: 45535531-ROHAN MARCUS DEANS

GITHUB Username: 45535531

Repository Name: Quiz1 (https://github.com/45535531/Quiz1.git)

File Name: 45535531DeansRohanMarcusTHQ1.pdf

1.

```
#The function
bond_ting <- function(n,C,F) {
  #n is the number of coupons
  #c is the coupon payment
  #F is the face value payment
  #rates is a vector of interest rates which has a length of n

  #Times of payment
  times <- seq(from = 0.5,to = n*0.5, by = 0.5)

  #Index of th etimes
  index <- 1:length(times)

  #Rates for coupon
  rates  <- rnorm(n, 0.02,0.01)

  #Storing the cash flows
  money_ting <- rnorm(n)
```

```
for (k in index){

  #Pick interest for time k
  i <- rates[k]

  #Pick the actual time at time k
  t <- times[k]

  #Calculate discounting factor
  discounting_factor <- exp(-i*t)

  #Get the present value of the coupon payment
  cp <- C*discounting_factor

  #Store it in the vector
  money_ting[k] <- cp

}

#Present value of face value cf
fv <- F*exp(rates[n]*times[n])

#Add it to the rest
result <- fv + sum(money_ting)
```

```
  return(result)

}
```

for example:

```
> bond_ting(5,150,1000)
[1] 1777.334
```

2.

The chapters 3.1 and 3.2 cover the key ideas under Simple Linear Regression Models (SLRM) and Multiple Linear Regression Models (MLRM). Despite being an ancient statistical learning method, it serves the purpose of being a good jumping-off point to develop newer approaches in the form of generalizations or extensions of itself. Hence the need for a sound understanding of its usage and importance is vital. The relationship between response variable Y and predictor variable X can be mathematically written as $Y \approx \beta_0 + \beta_1 X + e$, where β_0 and β_1 are unknown constants representing the intercept and slope of the model respectively. Thus the dependency of Y on various X is assumed to be linear. Once known, we predict the response variable as $y_i\hat{}= \beta_0\hat{} + \beta_1\hat{}\ x_i$. Then $e_i = y_i - y_i\hat{}$ represents the residual and the residual sum of squares (RSS) is given by $RSS = e_1^2 + e_2^2 + \cdots + e_n^2$, which can be minimized by least squares approach using $\beta_1\hat{}$ and $\beta_0\hat{}$ as minimizing values.

It is clear that when assessing the accuracy of the coefficient estimates, standard errors are made use of in computing the confidence intervals and hypothesis tests as well. σ^2 is unknown but it can be estimated from the data using an estimate known as residual standard error(RSE), which is known to us by the formula $RSE = \sqrt{(RSS/(n-2))}$.

If we assume a 95% confidence interval, it means that the following interval will contain the true value of β_1: $[\beta_1\hat{} - 2 \cdot SE\ (\beta_1\hat{}),\ \beta_1\hat{} + 2 \cdot SE\ (\beta 1\hat{})]$.

For hypothesis tests, it usually involves testing the null hypothesis (H_0) versus the alternate hypothesis (H_1), where $H_0 = \beta_1 = 0$ (No relationship between X and Y) and $H_1 = \beta_1 \neq 0$ (Some form of relationship between X and Y). We usually test the null hypothesis by computing the t- test statistic with (n-2) degrees of freedom given by $t = (\beta 1\hat{} - 0)/SE\ (\beta 1\hat{})$. This allows us to measure the number of standard deviations that $\beta 1\hat{}$ is away from 0. The R^2 statistic provides an alternative measure of fit to the RSE's absolute measure of lack of fit since it is measured in units of y, taking the form of a proportion as follows: $R^2 = (TSS-RSS)/TSS = 1-(RSS/TSS)$. TSS is the total sum of squares and is given by $TSS = \sum (y_i - \bar{y})^2$. Correlation (r) is also a measure of the relationship between X and Y, which can also be used instead of R^2 for SLRM, such that $R^2 = r^2$.

In practice we usually have more than one predictor and hence the need for MLRM. The relationship for a MLRM is written similarly as $Y \approx \beta_0 + \beta_1 X_1 + \ldots .. + \beta_p X_p + e$ and we can make predictions using the formula $y = \beta_0\hat{} + \beta_1\hat{}\ x_1 + \ldots .. + \beta_p\hat{}\ x_p$. With regard to hypothesis testing, in the multiple regression setting with p predictors, since all the regression coefficients are zero, i.e. whether $\beta_1 = \beta_2 = \cdots = \beta_p = 0$, which is taken as our null hypothesis (H_0) versus alternate hypothesis (H_1): at least one β_j is non-zero. We compute the F-statistic: $F = [(TSS-RSS)/p]/[\ RSS/(n-p-1)]$. We use forward, backward or mixed selection when deciding on important variables. RSE for MLRM is given by: $RSE = \sqrt{(RSS/(n-p-1))}$.

3.

a).

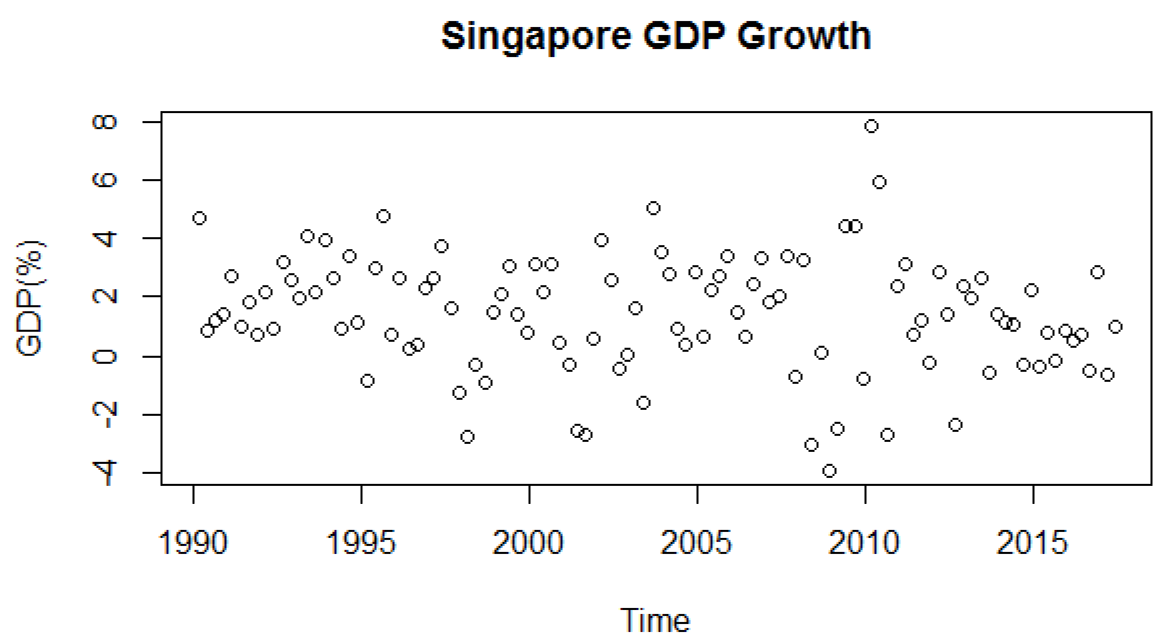
```
singapore_data <- read.csv(file="singapore.economy.csv",head=TRUE,sep=",")
```

b).

```
singapore_cleandata<-na.omit(singapore_data)
```

c).

```
plot(singapore_cleandata$time,singapore_cleandata$gdp,xlab="Time",ylab="GDP(%)",main="S
ingapore GDP Growth")
```



d).

```
period <- 1

index <- which(singapore_cleandata$period == period)

period1_data <- singapore_cleandata[index,]

mean1<-mean(period1_data$gdp)
```

```
[1] 1.702953

stdev1<-sqrt(var(period1_data$gdp))
[1] 1.685216

period <- 2
index <- which(singapore_cleandata$period == period)
period2_data <- singapore_cleandata[index,]
mean2<-mean(period2_data$gdp)

[1] 1.571337

stdev2<-sqrt(var(period2_data$gdp))
[1] 1.787153

period <- 3
index <- which(singapore_cleandata$period == period)
period3_data <- singapore_cleandata[index,]
mean3<-mean(period3_data$gdp)

[1] 1.04528

stdev3<-sqrt(var(period3_data$gdp))
[1] 2.40823
```

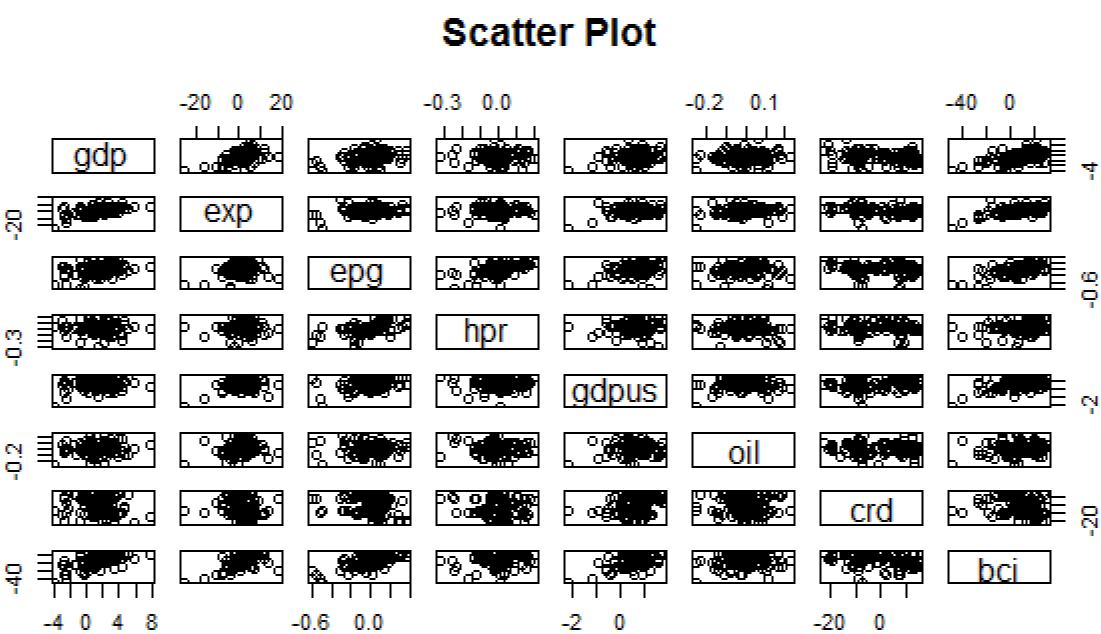
stat.table<-
data.table(period=c("1","2","3"),mean=c(mean1,mean2,mean3),stdev=c(stdev1,stdev2,stdev3))

stat.table

	period	mean	stdev
1	1	1.702953	1.685216
2	2	1.702953	1.685216
3	3	1.045280	2.408231

e).

pairs(~gdp+exp+epg+hpr+gdpus+oil+crd+bci,data=singapore_cleandata,main="Scatter Plot")



f).

mod1<-lm(singapore_cleandata$gdp~singapore_cleandata$exp,data=singapore_cleandata)

summary(mod1)

lm(formula = singapore_cleandata$gdp ~ singapore_cleandata$exp,

data = singapore_cleandata)

Residuals:

```
    Min      1Q  Median      3Q     Max 
-4.9294 -1.0044  0.2445  0.8869  5.6055 
```

Coefficients:

```
                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)               1.19832    0.16529   7.250 6.60e-11 ***
singapore_cleandata$exp  0.19076    0.02887   6.608 1.52e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Residual standard error: 1.693 on 108 degrees of freedom
Multiple R-squared:  0.2879,  Adjusted R-squared:  0.2813 
F-statistic: 43.66 on 1 and 108 DF,  p-value: 1.524e-09
```

Conclusion:

We get a model intercept of 1.19832. Our beta for exp is 0.19076. We can conclude that GDP for Singapor will increase by 0.19076 for every percentage increase in exp. Since p- value is very small the effect of export growth on GDP is significant.

g).

```
mod2<-lm(singapore_cleandata$gdp~singapore_cleandata$exp+singapore_cleandata$epg+singapore_cleandata$hpr+singapore_cleandata$oil+singapore_cleandata$gdpus+singapore_cleandata$crd,data=singapore_cleandata)

summary(mod2)

lm(formula = singapore_cleandata$gdp ~ singapore_cleandata$exp + 
    singapore_cleandata$epg + singapore_cleandata$hpr + singapore_cleandata$oil + 
```

```
    singapore_cleandata$gdpus + singapore_cleandata$crd, data = singapore_cleandata)

Residuals:
    Min      1Q  Median      3Q     Max
-5.1425 -0.8646  0.0995  0.9493  5.2931

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               1.287222   0.245125   5.251 8.17e-07 ***
singapore_cleandata$exp    0.156420   0.031059   5.036 2.03e-06 ***
singapore_cleandata$epg    3.964425   1.224407   3.238  0.00162 **
singapore_cleandata$hpr   -5.474249   2.094579  -2.614  0.01030 *
singapore_cleandata$oil    0.190507   1.794301   0.106  0.91565
singapore_cleandata$gdpus -0.046563   0.300998  -0.155  0.87736
singapore_cleandata$crd   -0.008538   0.016211  -0.527  0.59955
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.628 on 103 degrees of freedom
Multiple R-squared:  0.372,    Adjusted R-squared:  0.3354
F-statistic: 10.17 on 6 and 103 DF,  p-value: 8.144e-09
```

Conclusion:

This MLRM is not significant due to p-value. Increase in exp by 1% increases GDP by 0.15642, nit increase in hpr increases GDP by 3.9644, Real house price has the biggest effect on GDP.

h).

```
cric_val <- quantile(singapore_cleandata$gdp,0.05,type=1)

5%

-2.564473

state <- ifelse(singapore_cleandata$gdp < cric_val, "Crisis","Normal")

singapore_newdata<-data.frame(singapore_cleandata,state)    #or we can use
singapore_cleandata$state<-state

print(singapore_newdata)

training_data<-singapore_newdata[which(singapore_newdata$time<=2007,]

logmod<-glm(state~bci,data=training_data,family="binomial" )

summary(logmod)


Call:

glm(formula = state ~ bci, family = "binomial", data = training_data)


Deviance Residuals:

    Min        1Q    Median        3Q       Max

-2.52103   0.07219   0.09536   0.16072   0.82933


Coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)  3.64967    1.07407   3.398 0.000679 ***

bci          0.10395    0.05184   2.005 0.044936 *

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 18.046 on 67 degrees of freedom

Residual deviance: 12.370 on 66 degrees of freedom

AIC: 16.37

Number of Fisher Scoring iterations: 8

```
prediction<-predict(logmod)
actual<-training_data$gdp
CM<-confusionMatrix(prediction,actual,cutoff=0.5)
CM
```