GitHub username: pjswagger
Name of repository: ACST_s1_2019
File name: 43889573PengJiajunTHQ1

```
#Question 1
Price_Bond <- function(C,F,n,i){
  j=seq(0.5,n,0.5)
  sum(C*exp(-i*j))+F*exp(-i[2*n]*n)
}
i <- rep(0.1,2)
```

#Q2 3.1Simple linear regression

Simple linear regression is an approach for predicting a quantitative response Y on a single predictor variable X and assume an approximately linear regression between Y and X. Mathematically, $Y \approx \beta_0 + \beta_1 X$. β_0 and β_1 are unknown and represent the intercept and slope term in the linear model. Those can be estimate by least square approach and used to predict Y, that is $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, which is also the least square line. The different between observed and predicted is calculated by the Residual sum of square (RSS) which is the sum of the residual $e_i = y_i - \hat{y}_i$. There is a population regression line $Y = \beta_0 + \beta_1 X + \varepsilon$, where ε is the mean-zero random error term that catch-all for what is missed from the simple model and independent to X. If we use the sample mean ($\hat{\mu}$) to estimate the true mean (μ), this estimate is unbiased. Also, the standard error of the $\hat{\mu}$, SE ($\hat{\mu}$) measure the accuracy of estimation of $\hat{\mu}$. We can also calculate how close the estimate $\hat{\beta}_0$ and $\hat{\beta}_1$ are to the true β_0 and β_1 by SE ($\hat{\beta}_0$) and SE($\hat{\beta}_1$). These standard errors can be used to calculate confidence interval and perform hypothesis tests using t-statistic and p-value. Residual Standard Error (RSE) and R^2 are usually used to quantify the extent to which the model fits the data after we rejected the null hypothesis. RSE measure the average amount that the response will deviate from the true regression line and also the lack of fit of the model to the data. RSE will should be small if the estimate response is close to the true response. R^2 on the other hand, provide a proportion of variability in Y that can be explained using X and independent of the scale of Y. $R^2 \in [0,1]$. R^2 statistic that is close to 1 indicate that large proportion of the variability in the response has been explained by the regression, vice versa. But the a good R^2 value depends on the application.

3.2 Multiple linear regression

Similar to single linear regression but with more 1 predictor in the model. Estimation of coefficients are done by minimise RSS and will be presented in matrix form. The interpretation of multiple linear regression is with certain predict fixed, every one unit change in another predictor (say X_2) will result β_2 change in the response. Sometime the coefficient of same predictor in multiple linear is a lot different form the simple linear since the correlation between predictors. F-statistic is used to make hypothesis test for whether the predictors are significant to the response. Forward, backward, and mixed selection are useful for variable selection. RSE and R^2 are used to measure model fit. Uncertainty in the model are measured by confidence and prediction interval. Prediction interval consider both reducible and irreducible errors whereas confidence interval consider only reducible error.

```
#Q3
a)
SingNA <- read.csv("singapore.economy.csv", header = T,)

b)
Sing <- na.omit(SingNA)

c)
plot(Sing$time, Sing$gdp,xlab = "Time",ylab = "GDP(%)", main = "Singapore GDP
growth")

d)
m <- aggregate(Sing$gdp, by=list(Sing$period),FUN=mean)
  sd <- aggregate(Sing$gdp, by=list(Sing$period),FUN=sd)
  stat.table <-cbind(m[,2],sd[,2])
  rownames(stat.table) <- c("Period 1","Period 2", "Period 3")
  colnames(stat.table) <- c("Mean","Sd")

e)
pairs(~Sing$gdp+Sing$exp+Sing$epg+Sing$hpr+Sing$gdpus+Sing$oil+Sing$crd+Sing$bci,
Sing)

f)
FittedlmGvE <- lm(Sing$gdp~Sing$exp)
  summary(FittedlmGvE)
  #Comment:The intercept is 1.19832, the coefficient of
  #the exp is 0.19076. The p-value of both are close to zero
  #so both of them are significant. R square is 28.8%, so
  #it might not suggest a strong relationship between response and predictor

g)
FittedlmGvM <-
lm(Sing$gdp~Sing$exp+Sing$epg+Sing$hpr+Sing$oil+Sing$gdpus+Sing$crd)
  summary(FittedlmGvM)
  #Comment: the coefficient of the predictors
  #show that exp and epg have a positive effect to the response
  #whereas hpr, gdpus and crd have negative effect.
  #For p-value, intercept, exp,  epg, hpr are significant
  #whereas gdpus and crd are not.
  #R-square is 37.2% indicate that the relationship between
  #response and predictor are not strong

h)
q <- quantile(Sing$gdp,probs = 0.05)
  state <- factor(ifelse(Sing$gdp<q,"crsis","normal"))
  train <- Sing[Sing$time<2008,]
  test <- Sing[Sing$time>=2008,]
  FittedState <- glm(Sing$state~Sing$bci,data=test, family = binomial)
  summary(FittedState)
  #confusion matrix
```

```
prob <- predict(FittedState, data = test, type = 'response')
contrasts(state)
pred <- ifelse(prob<0.05, "crisis","normal")
conf.mat <- table(pred, state);conf.mat
```