```r
#============================================== SLR and MLR ==========================================
# This is solution to in class problem for SLR and MLR

#Read the excel file and sheet
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
#install.packages("readxl")                                                    # install R package for function used to read the excel file
library(readxl)                                                               # load package
data_Q1<-read_excel("L16_InClass_Problems_solution.xlsx",sheet="Q1")          # read excel file and sheet

#install.packages("corrplot")
library(corrplot)
#check correlation among the variables
corr1<-cor(data_Q1)
corrplot(corr1,method="circle", type="upper")

#install.packages("jtools")                                                   # install and load R package for model output
library(jtools)

# SLR and MLR
model1_1<-lm(Y~X2+X3, data=data_Q1)                                           # between Y and X2,X3
summary (model1_1)                                                            # model output using built in function
summ(model1_1)                                                                # mode output using the package jtools

model1_2<-lm(Y~X2+X3+X4, data=data_Q1)                                        # between Y and X2,X3,X4
summ(model1_2)

model1_3<-lm(Y~log(X1), data=data_Q1)                                         # between Y and logX1
summ(model1_3)

model1_4<-lm(Y~X1, data=data_Q1)                                              # between Y and X1
summ(model1_4)

# Calibrate 10y=log(X)
data_Q3<-read_excel("L16_InClass_Problems_solution.xlsx",sheet="Q3")          # read excel file and sheet
model3<-lm(Y~log10(X),data=data_Q3)                                           # SLR
summ(model3)

# Calibrate Y=(a+bx)**-1
data_Q4<-read_excel("L16_InClass_Problems_solution.xlsx",sheet="Q4")          # read excel file and sheet
model4<-lm(1/Y~X,data=data_Q4)                                                # SLR
summ(model4)
# ============================================== END ==========================================
# This is solution to in class problem for binary logit
bi_logit<-read_excel("L16_InClass_Problems_solution.xlsx",sheet="Binary logit")    # read excel sheet


#dummy code the variables to make them dichotomous
#install.packages("dummy_cols")
library(fastDummies)
bi_logit<-dummy_cols(bi_logit,select_columns=c("drvrage","income"),remove_selected_columns=TRUE)
str(bi_logit)
```

```r

#check point-biserial correlation among the variables
corr2<-cor(bi_logit[,-c(1,2)])                                                  #
plot correlation coefficients
corrplot(corr2,method="circle", type="upper")

# specify categorical variables in the data
str(bi_logit)
library(dplyr)
cat_vars<-dplyr::select(bi_logit,choice,drvrage_1,drvrage_2,drvrage_3,drvrage_4,drvrage_5
,gender,maritsta,income_1,income_2,income_3,income_4,income_5)
cat_vars<-names(cat_vars)
bi_logit[,cat_vars]<-lapply(bi_logit[,cat_vars],factor)
str(bi_logit)

logitm1<-glm(choice~nopasngr+drvrage_2+drvrage_3+drvrage_4+drvrage_5+gender+maritsta+noch
ild+income_2+income_3+income_4+income_5+yearcar+fueleff+flowrate.Arterial+flowrate.Twolan
e+nosignal.Arterial+nosignal.Twolane,family=binomial,data=bi_logit)
summ(logitm1)                                                                   #
drvrage_1 and income_1 as base categories

#install.packages("MASS")
library(MASS)                                                                   # we
use stepwise elimination of variables based on AIC values using stepAIC from MASS package
logitm2<-stepAIC(logitm1)                                                       #
p-values alone are not adequate for deciding the inclusion of variable in the model
summary(logitm2)
summ(logitm2)

exp(coef(logitm2))                                                              #
odds ratio for the predictors


```