```
library(MASS)
library(tidyverse)
library(GGally)
library(labelled)
library(dplyr)
library(rstatix)
library(ivreg)
library(simstudy)
library(ivmodel)
library(ggplot2)
library(car)
library(lmtest)
library(tseries)
library(ggfortify)
library(plotly)

setwd("C:\\Users\\fesjo\\OneDrive\\Dokumente\\Uni\\Labour Economics WS25-26\\Problem Sets")

rm(list = ls())



####### Task 1.a.) #######
# Generate log wages (ln wage) and interpret a table with descriptive statistics for education,
motivation,
# hours and ln wage. Also calculate correlations and plot the density of wages as well as a
histogram
# for the years of education.

load("ps1_clean_data.Rda")

#a)
# df1 stands for dataframe 1. generate the log of wage.
df1$ln_wage = log(df1$wage)

#Get summary statistics:
df1 %>%
  get_summary_stats(
    education, motivation, hours, ln_wage,  # columns to calculate for
    type = "common")

#Get correlations
cor(df1[,c(1,2,5,6)], use="pairwise", method="pearson")

#Some Graphs
density <- density(df1$ln_wage)
fig <- plot_ly(x = ~density$x, y = ~density$y, type = 'scatter', mode = 'lines', fill =
'tozeroy')
fig <- fig %>% layout(xaxis = list(title = 'Log Wage'),
                      yaxis = list(title = 'Density'))

fig2 <- plot_ly(x = ~df1$education, type = "histogram",alpha = 0.6)%>% layout(bargap=0.1,xaxis
=list(dtick = 1,tickmode = "linear"))
fig2 <- fig2 %>% layout(xaxis = list(title = 'Years of education'),
                        yaxis = list(title = 'Frequency'))
fig2

#If you want to export your graphs as .pdf:
#use the export button in your Viewer, then go to save as image and save it.
#Than you have to use an pdf writer such as adobe pdf writer and create an pdf with another
file, here we want this file to be our image we safed earlier.
#than you can use this .pdf for example to paste it into your LaTex document.


#b)
#Compare descriptive statistics for individuals who have a motivation above or equal to zero
versus below.
#Again comment on what you find.
```

```
#Create dummy
df1$M <- ifelse(df1$motivation >=0, 1, 0)

df1 %>%
  group_by(M) %>%
  get_summary_stats(
    education, motivation, hours, ln_wage,  # columns to calculate for
    type = "common")

df1$E <- ifelse(df1$education >=14, 1, 0)

df1 %>%
  group_by(E) %>%
  get_summary_stats(
    education, motivation, hours, ln_wage,
    type = "common")

#c)
#Plot a scatter of the hours_worked against ln_wage. What do you notice?

plot_ly(data=df1, x = ~ln_wage, y = ~hours, name = 'Observations', type = 'scatter',mode =
'markers')

####### Regressions #######

#d)
#Bivariate Regression
simple_OLS <- lm(hours ~ ln_wage,data = df1)
summary(simple_OLS)

#Confidence intervals
confint(simple_OLS)
Std_error_wage <- sqrt(diag(vcov(simple_OLS)))[2]
c("lower (2.5%)" = simple_OLS$coef[2] - qt(0.975, df = simple_OLS$df) * Std_error_wage,
  "upper (97.5%)" = simple_OLS$coef[2] + qt(0.975, df = simple_OLS$df) * Std_error_wage)


df1$hours_hat = fitted(simple_OLS)
df1$res = simple_OLS$residuals

#Plot of x, y and the regression line
plot_ly(data=df1, x = ~ln_wage, y = ~hours, name = 'Observations', type = 'scatter',mode =
'markers')%>%
  add_trace(data=df1,y = ~hours_hat, name = 'OLS', mode = 'lines')

#e)
#Adding education to the regression
summary(lm(hours ~ ln_wage + education,data = df1))

#f)
# Reg with all variables
summary(lm(hours ~ ln_wage + education + motivation ,data = df1))
```