

```

#Data upload
setwd("~/Desktop/Škola/Bakalarka/Data_Insolcentrum")
library(readxl)
mydata <- read_xlsx("DATA.xlsx")

#checking the data at the first glance
summary (mydata)

#checking high value at totaldebt
summary(mydata$total_debt)
max(mydata$total_debt)

#extreme observation identified - delete observation
data <- mydata[-c(29), ]
max(data$total_debt)

##defining the dummy variables

#status dummy
success_d <- as.numeric((data$status=="USP" ))
success_d
data$success_d <- success_d

#female dummy
female_d <- as.numeric((data$gender=="F" ))
female_d
data$female_d <- female_d

#dependants dummy
depen_d <- as.numeric((data$dependants=="A" ))
depen_d
data$depen_d <- depen_d

#debt relief of spouses dummy
spouses_d <- as.numeric((data$debt_spouses=="A" ))
spouses_d
data$spouses_d <- spouses_d

#year dummies
data$y2008_12 <- as.numeric(data$proposal_year== 2008,2009,2010,2011,2012)
data$y2013_17 <- as.numeric(data$proposal_year== 2013,2014,2015,16,2017)

#descriptive statistics table for LaTeX
library(stargazer)
data2 <- data.frame(data)

```

```

stargazer(data2, omit.summary.stat = c("N", "p25", "p75"), median = TRUE )

summary(data)

#histogram - number of creditors
hist <- hist(data2$creditors, breaks=20,main=" ",xlab="Number of Creditors",xlim=c(0,50),
ylim=c(0,50),xaxt="n",cex.lab=0.7)
axis(1, at=seq(0, 50, by=3))
hist
text(hist$mid, hist$count, labels=hist$count,adj=c(0.5, -0.5), cex=0.65)

#preparation for corr matrix
data_corr <- data2
data_corr$status <- NULL
data_corr$age_cat <- NULL
data_corr$dependants <- NULL
data_corr$debt_spouses <- NULL
data_corr$gender <- NULL
data_corr$total_sec <- NULL
data_corr$total_debt <- NULL
data_corr$total_unsec <- NULL
data_corr$number <- NULL
data_corr$u_bank <- NULL
data_corr$u_nonbank <- NULL
data_corr$u_assignment <- NULL
data_corr$u_others <- NULL
data_corr$u_exe <- NULL
data_corr$s_bank <- NULL
data_corr$s_nonbank <- NULL
data_corr$s_assignment <- NULL
data_corr$s_others <- NULL
data_corr$s_exe <- NULL
data_corr$proposal_year <- NULL
data_corr$region_ID <- NULL
data_corr$born <- NULL
data_corr$success_d <- NULL
data_corr$region_MS<- as.numeric(data$region_ID== 1,3,6,7,13,14)
data_corr$region_C <- as.numeric(data$region_ID== 2,4,5,8,9,12,11)
data_corr$region_P <- as.numeric(data$region_ID== 10)

data_corr$bankp <- NULL
data_corr$nonbankp <- NULL
data_corr$assignmentp <- NULL
data_corr$othersp <- NULL
data_corr$exep <- NULL
data_corr$bank <- NULL

```

```

data_corr$nonbank <- NULL
data_corr$assignment <- NULL
data_corr$others <- NULL
data_corr$exe <- NULL
data_corr$u_exe <- NULL
data_corr$u_nonbank <- NULL
data_corr$u_bank<- NULL
data_corr$u_assignment <- NULL
data_corr$s_assignment <- NULL
data_corr$s_bank <- NULL
data_corr$s_nonbank <- NULL
data_corr$s_exe <- NULL
data_corr$u_others <- NULL
data_corr$s_others <- NULL

#corr matrix
library(ggcorrplot)
corr <- round(cor(data_corr), 1)
ggcorrplot(title = "", hc.order = TRUE, cor(data_corr), tl.cex=8, lab_size=6)

ggcorrplot(corr,hc.order=TRUE , type = "upper",
           lab = TRUE, tl.cex=9,lab_size=2)

#sum of unsecured and secured debts
data2$bank <- data2$u_bank + data2$s_bank
data2$nonbank <- data2$u_nonbank + data2$s_nonbank
data2$assignment <- data2$u_assignment + data2$s_assignment
data2$others <- data2$u_others + data2$s_others
data2$exe <- data2$u_exe + data2$s_exe

data2$bankp <- data2$u_bankp + data2$s_bankp
data2$nonbankp <- data2$u_nonbankp + data2$s_nonbankp
data2$assignmentp <- data2$u_assignmentp + data2$s_assignmentp
data2$othersp <- data2$u_othersp + data2$s_othersp
data2$exep <- data2$u_exep + data2$s_exep

#region ID dummies
data2$region_MS<- as.numeric(data$region_ID== 1,3,6,7,13,14)
data2$region_C <- as.numeric(data$region_ID== 2,4,5,8,9,12,11)
data2$region_P <- as.numeric(data$region_ID== 10)

##### LOGIT #####
model_logit2 <- (glm(success_d ~ bankp + nonbankp + assignmentp + exep , family =
binomial
(link = "logit"), data = data2))

```

```

summary(model_logit2)

model_logit22 <- (glm(success_d ~ kids+avr_inc + creditors+female_d + spouses_d+
  bankp + nonbankp + assignmentp + exep
  , family = binomial
  (link = "logit"), data = data2))
summary(model_logit22)

model_logit23 <- (glm(success_d ~ kids+avr_inc+ creditors+female_d + spouses_d+
  bankp + nonbankp + assignmentp + exep
  +region_MS+region_C
  +y2013_17, family = binomial
  (link = "logit"), data = data2))
summary(model_logit23)

model_logit24 <- (glm(success_d ~ kids+avr_inc+ + creditors+female_d + spouses_d+
  bankp + nonbankp + assignmentp + exep+
  +y2013_17, family = binomial
  (link = "logit"), data = data2))
summary(model_logit24)

model_logit25 <- (glm(success_d ~ kids +avr_inc +bankp+ nonbankp + assignmentp + +
  exep
  +y2013_17, family = binomial
  (link = "logit"), data = data2))
summary(model_logit25)

```

```

##pseudo R2 manually
pseudoR2_logit25 <- 1 - (model_logit25$deviance) / (model_logit25$null.deviance)
pseudoR2_logit25
pseudoR2_logit23 <- 1 - (model_logit23$deviance) / (model_logit23$null.deviance)
pseudoR2_logit23
pseudoR2_logit24 <- 1 - (model_logit24$deviance) / (model_logit24$null.deviance)
pseudoR2_logit24

##pseudo R2 function
rsquared <- function(created_model) {
  dev <- created_model$deviance
  null_dev <- created_model$null.deviance
  model_n <- length(created_model$fitted.values)
  R_I <- 1 - dev / null_dev
  R_cs <- 1 - exp(-(null_dev - dev) / model_n)
  R_n <- R_cs / (1 - exp(-(null_dev / model_n)))
  cat("Pseudo R-squared for logistic regression model\n\n")
  cat("Hosmer and Lemeshow R-squared\t", round(R_I, 3), "\n")
}

```

```

cat("Cox and Snell R-squared\t\t\t", round(R_cs, 3), "\n")
cat("Nagelkerke R-squared\t\t\t", round(R_n, 3), "\n")
}

rsquared(model_logit25)
rsquared(model_logit24)

### APEs and PEAs

##LOGIT24
#PEA
m1_PEA <- logitmfx(model_logit24,data=data2, atmean = TRUE)
m1_PEA
#APE
m1_APE <-logitmfx(model_logit24,data=data2,atmean = FALSE)
m1_APE

##LOGIT25
#PEA
m3_PEA <- logitmfx(model_logit25,data=data2, atmean = TRUE)
m3_PEA
#APE
m3_APE <-logitmfx(model_logit25,data=data2,atmean = FALSE)
m3_APE

##LOGIT14

#PEA
m2_PEA <- logitmfx(model_logit14,data=data2, atmean = TRUE)
m2_PEA
#APE
m2_APE <- logitmfx(model_logit14,data=data2,atmean = FALSE)
m2_APE

##### LPM MODEL #####
##only debt structure
model_LPM_structure <- (lm(success_d ~ kids +avr_inc +bankp+ nonbankp + assignmentp +
+ exep
+ y2013_17, data = data2))
summary(model_LPM_structure)

#Dummy variable trap - the debt structure

```

```

model_LPM3 <- (lm(success_d ~ kids+avr_inc+creditors+female_d + spouses_d+
                     bankp+ nonbankp + assignmentp + + exep
                     +y2013_17, data = data2))
summary(model_LPM3)

# Outlier function
outlierKD <- function(dt, var) {
  var_name <- eval(substitute(var),eval(dt))
  na1 <- sum(is.na(var_name))
  m1 <- mean(var_name, na.rm = T)
  par(mfrow=c(2, 2), oma=c(0,0,3,0))
  boxplot(var_name, main="With outliers")
  hist(var_name, main="With outliers", xlab=NA, ylab=NA)
  outlier <- boxplot.stats(var_name)$out
  mo <- mean(outlier)
  var_name <- ifelse(var_name %in% outlier, NA, var_name)
  boxplot(var_name, main="Without outliers")
  hist(var_name, main="Without outliers", xlab=NA, ylab=NA)
  title("Outlier Check", outer=TRUE)
  na2 <- sum(is.na(var_name))
  cat("Outliers identified:", na2 - na1, "n")
  cat("Propotion (%) of outliers:", round((na2 - na1) / sum(!is.na(var_name))*100, 1), "n")
  cat("Mean of the outliers:", round(mo, 2), "n")
  m2 <- mean(var_name, na.rm = T)
  cat("Mean without removing outliers:", round(m1, 2), "n")
  cat("Mean if we remove outliers:", round(m2, 2), "n")
  response <- readline(prompt="Do you want to remove outliers and to replace with NA?
[yes/no]: ")
  if(response == "y" | response == "yes"){
    dt[as.character(substitute(var))] <- invisible(var_name)
    assign(as.character(as.list(match.call())$dt), dt, envir = .GlobalEnv)
    cat("Outliers successfully removed", "n")
    return(invisible(dt))
  } else{
    cat("Nothing changed", "n")
    return(invisible(var_name))
  }
}

# outliers identification
data4 <- data2
outlierKD(data4,bankp) #no outliers
outlierKD(data4,nonbankp) #no outliers
outlierKD(data4,assignmentp) # 44 outliers
outlierKD(data4,othersp) #18 outliers
outlierKD(data4,creditors) #16 outliers

```

```

outlierKD(data4,exep) #31 outliers
outlierKD(data4,avr_inc) #3 outliers

data4 <- na.omit(data4)

##### LPM MODEL - SUBSAMPLE #####
#LPM model without outliers - debt structure + 2 control variables - (model_LPM_structure)
model_LPM_structure_O <- (lm(success_d ~ kids+avr_inc+bankp+nonbankp+
assignmentp+ + exep
+y2013_17, data = data4))
summary(model_LPM_structure_O)

#LPM model without region dummies - (model_LPM3)
model_LPM3o <- (lm(success_d ~ kids+avr_inc+creditors+female_d + spouses_d+
bankp+nonbankp+assignmentp+ + exep
+y2013_17, data = data4))
summary(model_LPM3o)

##### OLS ASSUMPTION CHECK #####
### Testing for serial autocorrelation
require(lmtest)
bgtest(model_LPM_structure, order=1)
bgtest(model_LPM_structure_O, order=1)
bgtest(model_LPM3o, order=1)
bgtest(model_LPM3, order=1)

### Testing for homoskedasticity
test_homoskedasticity <- bptest(model_LPM_structure, data=data2)
test_homoskedasticity

test_homoskedasticity <- bptest(model_LPM_structure_O, data=data4)
test_homoskedasticity

test_homoskedasticity <- bptest(model_LPM3, data=data2)
test_homoskedasticity

test_homoskedasticity <- bptest(model_LPM3o, data=data4)
test_homoskedasticity

### Correcting autocorrelation and heteroskedasticity
library(sandwich)

LPM_structure <- coeftest(model_LPM_structure, vcov =
vcovHAC(model_LPM_structure,type="HC3"))

```

```

LPM_structure_O <- coeftest(model_LPM_structure_O, vcov =
vcovHAC(model_LPM_structure_O,type="HC3"))

LPM3 <- coeftest(model_LPM3, vcov = vcovHAC(model_LPM3,type="HC3"))

LPM3o <- coeftest(model_LPM3o, vcov = vcovHAC(model_LPM3o,type="HC3"))

##LaTeX - structure model
stargazer(model_LPM_structure, model_LPM_structure_O)

##### LOGIT MODELS - SUBSAMPLE #####
model_logit2o <- (glm(success_d ~ bankp + nonbankp + assignmentp + exep , family =
binomial
(link = "logit"), data = data4))
summary(model_logit2o)

model_logit22o <- (glm(success_d ~ kids+avr_inc + creditors+female_d + spouses_d+
bankp + nonbankp + assignmentp + exep
, family = binomial
(link = "logit"), data = data4))
summary(model_logit22o)

model_logit23o <- (glm(success_d ~ kids+avr_inc+ creditors+female_d + spouses_d+
bankp + nonbankp + assignmentp + exep
+region_MS+region_C
+y2013_17, family = binomial
(link = "logit"), data = data4))
summary(model_logit23o)

model_logit24o <- (glm(success_d ~ kids+avr_inc+ creditors+female_d + spouses_d+
bankp + nonbankp + assignmentp + exep+
+y2013_17, family = binomial
(link = "logit"), data = data4))
summary(model_logit24o)

model_logit25o <- (glm(success_d ~ kids +avr_inc +bankp+ nonbankp + assignmentp + +
exep
+y2013_17, family = binomial
(link = "logit"), data = data4))
summary(model_logit25o)

lrtest(model_logit2o,model_logit22o)
lrtest(model_logit22o,model_logit23o)

```

```

lrttest(model_logit23o,model_logit24o)
lrttest(model_logit22o,model_logit24o)
lrttest(model_logit24o,model_logit25o)

##pseudo R2 manually
pseudoR2_logit25o <- 1 - (model_logit25o$deviance) / (model_logit25o$null.deviance)
pseudoR2_logit25o
pseudoR2_logit24o <- 1 - (model_logit24o$deviance) / (model_logit24o$null.deviance)
pseudoR2_logit24o

### APEs and PEAs

##LOGIT25 - subsample
#PEA
m3_PEAo <- logitmfx(model_logit25o,data=data4, atmean = TRUE)
m3_PEAo
#APE
m3_APEo <- logitmfx(model_logit25o,data=data4,atmean = FALSE)
m3_APEo

##LOGIT23 - subsample
#PEA
m2_PEAo <- logitmfx(model_logit23o,data=data4, atmean = TRUE)
m2_PEAo
#APE
m2_APEo <- logitmfx(model_logit23o,data=data4,atmean = FALSE)
m2_APEo

##LOGIT24 - subsample
#PEA
m1_PEAo <- logitmfx(model_logit24o,data=data4, atmean = TRUE)
m1_PEAo
#APE
m1_APEo <- logitmfx(model_logit24o,data=data4,atmean = FALSE)
m1_APEo

##LOGIT for LaTeX
stargazer(model_logit25,model_logit24, title="Results of logit models", align=TRUE)

##LOGIT + LPM for LaTex
stargazer(model_logit25,LPM_structure, model_logit24,LPM3, title="Results", align=TRUE)

##Logit + LPM for LaTeX - subsample

```

```
stargazer(model_logit25o,LPM_structure_O, model_logit24o, LPM3o, title="Results of  
models on a subsample", align=TRUE)  
  
library(texreg)  
  
##PEA, APE  
l5<-list(m3_PEA, m3_APE,m1_PEA,m1_APE)  
texreg(l5,digits=5)  
## PEA, APE - subsample  
l6<-list(m3_PEAo, m3_APEo,m1_PEAo,m1_APEo)  
texreg(l6,digits=5)
```