Appendix D: R Programming Code

```
#display numeric value
options (scipen=999, digits=4)

#clear environment
rm(list = ls())

#Load often used Libraries
library(readr)
library(dplyr)
library(psych)
library(ggplot2)
library(caret)
library(lfe)
library(broom)
library(stargazer)
library(reshape2)
library(Matrix)
library(vtable)
library(lubridate)
library(zoo)
library(leaps)
library(fitdistrplus)
library(skimr)
library(forecast)
library(tidyr)

#1.Importing Comp Financial Data
comp_data <- read_csv("comp_financial.csv")
dim(comp_data)
str(comp_data)
summary(comp_data)

#2.Basic data cleaning
comp_data <- arrange(comp_data, gvkey, fyear)

#2.1. analyse an replace missing fyear NA
mean(is.na(comp_data$fyear)) #0.0009467 although very few
NA it is important to remove NAs in fyear

#fill in fyear if missing, based on Compustat's May cutoff
comp_data$fyear <- ifelse(
  is.na(comp_data$fyear),
  ifelse(
    as.numeric(format(comp_data$datadate, format = "%m"))
> 5,
    as.numeric(format(comp_data$datadate, format = "%Y")),
    as.numeric(format(comp_data$datadate, format = "%Y")) -
1), comp_data$fyear)

summary(comp_data$fyear)#no fyear NAs

#verification that each row have a gvkey and fyear
nrow(comp_data) - nrow(subset(comp_data, !is.na(gvkey) &
!is.na(fyear))) #0

#2.2 creating firm-year indices using gvkey and fyear
comp_data <- arrange(comp_data, gvkey, fyear)
comp_data$index <- paste(comp_data$gvkey,
comp_data$fyear, sep = "_")

#2.3 filtering only "INDL" data
comp_data <- filter(comp_data, indfmt == "INDL")

#2.4 identifying and removing duplicate indices
length(unique(comp_data$index)) #291933 different from
the base data set

comp_data_clean <- subset(
  comp_data, !(index %in% subset(comp_data,
duplicated(index) == 1)$index))
nrow(comp_data_clean) #261907 rows
length(unique(comp_data_clean$index)) #261907 firm-year
indices

#3. Importing executive compensation data
exec_data <- read.csv("new_exec_comp.csv")
dim(exec_data)
str(exec_data)
summary(exec_data)

#4. Basic data cleaning
exec_data <- arrange(exec_data, GVKEY, YEAR, CO_PER_ROL)

#4.1 Filtering to only including the data to CEO
exec_data <- exec_data[exec_data$CEOANN == "CEO", ]

#4.2 converting GVKEY to character
exec_data$GVKEY <- as.character(exec_data$GVKEY)
exec_data$GVKEY <- ifelse(nchar(exec_data$GVKEY) == 4,
paste0("00", exec_data$GVKEY), exec_data$GVKEY)
exec_data$GVKEY <- ifelse(nchar(exec_data$GVKEY) == 5,
paste0("0", exec_data$GVKEY), exec_data$GVKEY)

summary(exec_data$GVKEY)

#4.3 analyse missing Year values
mean(is.na(exec_data$YEAR)) #no NAs
nrow(exec_data) - nrow(subset(exec_data, !is.na(GVKEY) &
!is.na(YEAR))) #0

#4.4 creating firm-year indices using gvkey and year
exec_data <- arrange(exec_data, GVKEY, YEAR, CO_PER_ROL)
exec_data$index <- paste(exec_data$GVKEY,
exec_data$YEAR, sep = "_")

#4.5 identifying and removing duplicate indices
```

```
length(unique(exec_data$index)) #44940 different from the
base data set

exec_data_clean <- subset(
  exec_data,
  !(index %in% subset(exec_data, duplicated(index) ==
1)$index)
)
nrow(exec_data_clean) #44897 rows
length(unique(exec_data_clean$index)) #44897 firm-year
indices

#5.combining the two data set using inner join on index
data_comb <- inner_join(comp_data_clean, exec_data_clean,
by = "index")

#6. data cleaning on combined data set
data_comb <- arrange(data_comb, index)

#6.1 replace missing values with zero for ni, revt, oiadp, act,
lct, ch, lt, invt
data_comb_1 <- data_comb %>% mutate(ni = ifelse(is.na(ni),
0, ni), revt = ifelse(is.na(revt), 0, revt), oiadp =
ifelse(is.na(oiadp), 0, oiadp), act = ifelse(is.na(act), 0, act), lct=
ifelse(is.na(lct), 0, lct), ch = ifelse(is.na(ch), 0, ch), lt =
ifelse(is.na(lt), 0, lt), invt = ifelse(is.na(invt), 0, invt))

#6.2 create required lagged values
data_comb_2 <- arrange(data_comb_1, index)
data_comb_2 <- data_comb_1 %>% group_by(gvkey) %>%
mutate(at_lag = ifelse(fyear == lag(fyear) + 1, lag(at, n = 1),
NA), invt_lag = ifelse(fyear == lag(fyear) + 1, lag(invt, n = 1),
NA), seq_lag = ifelse(fyear == lag(fyear) +1, lag(seq, n = 1),
NA)) %>% ungroup()

#6.3 creating financial ratios: profit margin, operating profit,
ROE, current ratio, cash ratio, debt ratio, debt to equity ratio,
asset turnover, inventory turnover
data_comb_3 <- data_comb_2 %>%
mutate(net_profit_margin=ni/revt,operating_profit=oiadp/re
vt,ROE=ni/((seq+seq_lag)/2),
current_ratio=act/lct,cash_ratio=ch/lct,debt_ratio=lt/at,debt
_to_equity_ratio=lt/seq,
asset_turnover=revt/(at-lt),roa=ni/((at+at_lag)/2),
inventory_turnover=cogs/((invt+invt_lag)/2))

str(data_comb_3)

data_comb_4 <- data_comb_3 %>% group_by(CO_PER_ROL)
%>% mutate(ceo_years = YEAR - year(as.Date(BECAMECEO)),
count_ceo_years = n()) %>% ungroup() %>%
filter(!ceo_years<0)
summary(data_comb_4)

#7. Regression

#7.1 Select necessary variables for model estimation
data_reg_3 <- data_comb_4 %>% dplyr :: select(gvkey, fyear,
index, sic, ceo_years, at, net_profit_margin,
debt_to_equity_ratio, asset_turnover, inventory_turnover,
roa, current_ratio,
                                      SALARY, BONUS, OTHCOMP, RSTKGRNT,
OPTION_AWARDS_BLK_VALUE, LTIP)
summary(data_reg_3)
sum(is.infinite(data_reg_3$inventory_turnover))

#7.2 Remove NA and Infinite values in ratios
data_reg_3a <- data_reg_3 %>%
  filter(
    !is.na(at) &
     !is.infinite(inventory_turnover))

summary(data_reg_3a)

#7.3 remove outliers with truncation method
ggplot(data_reg_3a, aes(x = fyear, y = net_profit_margin)) +
geom_point() + geom_smooth(method = "lm", se = FALSE)
ggplot(data_reg_3a, aes(x = fyear, y = debt_to_equity_ratio))
+ geom_point() + geom_smooth(method = "lm", se = FALSE)
ggplot(data_reg_3a, aes(x = fyear, y = inventory_turnover)) +
geom_point() + geom_smooth(method = "lm", se = FALSE)
ggplot(data_reg_3a, aes(x = fyear, y = asset_turnover)) +
geom_point() + geom_smooth(method = "lm", se = FALSE)
ggplot(data_reg_3a, aes(x = fyear, y = current_ratio)) +
geom_point() + geom_smooth(method = "lm", se = FALSE)
ggplot(data_reg_3a, aes(x = fyear, y = roa)) + geom_point() +
geom_smooth(method = "lm", se = FALSE)
ggplot(data_reg_3a, aes(x = ceo_years)) + geom_density()

summary(data_reg_3a)

data_reg_3d_clean <- data_reg_3a %>%
filter(!net_profit_margin > quantile(net_profit_margin, 0.99,
na.rm = TRUE) & !net_profit_margin <
quantile(net_profit_margin, 0.01, na.rm = TRUE) &
!debt_to_equity_ratio > quantile(debt_to_equity_ratio, 0.99,
na.rm = TRUE) & !debt_to_equity_ratio <
quantile(debt_to_equity_ratio, 0.01, na.rm = TRUE) &
!inventory_turnover > quantile(inventory_turnover, 0.99,
na.rm = TRUE) & !inventory_turnover <
quantile(inventory_turnover, 0.01, na.rm = TRUE) &
!asset_turnover > quantile(asset_turnover, 0.99, na.rm =
TRUE) & !asset_turnover < quantile(asset_turnover, 0.01,
na.rm = TRUE) & !current_ratio > quantile(current_ratio,
0.99, na.rm = TRUE) & !current_ratio <
quantile(current_ratio, 0.01, na.rm = TRUE) & !roa >
quantile(roa, 0.99, na.rm = TRUE) & !roa < quantile(roa, 0.01,
na.rm = TRUE) &!ceo_years > 20) %>% mutate(othcomp_at =
```

```
OTHCOMP/at)  %>% mutate(salary_at = SALARY/at) %>%
mutate(bonus_at = BONUS/at) %>% mutate(LTIP_at =
LTIP/at)

summary(data_reg_3d_clean)

#8 multi-linear regression to analyse the effects of salary on
profit margin, roa, liquidity ratio, debt to equity ratio
#Dependent variable: net_profit_margin, roa, current_ratio,
debt_to_equity_ratio
#Independent variable: salary_at (salary/total assets)

#8.1. choosing training and test dataset
set.seed(1)
train1 <- sample_frac(data_reg_3d_clean, 0.75)
test1 <- anti_join(data_reg_3d_clean, train1)

#8.2. regressing salary_at with profit margin; control
varaibles: at, debt to equity ratio, ceo_years
lm1a_train <- lm(net_profit_margin ~ salary_at + at +
debt_to_equity_ratio + ceo_years, train1)

#8.2.1. perform stepwise regression
data_reg_stepwise_1a <- step(lm1a_train, direction = "both")
summary(data_reg_stepwise_1a)
data_reg_stepwise_pred_1a <-
predict(data_reg_stepwise_1a, test1)
accuracy(data_reg_stepwise_pred_1a,
test1$net_profit_margin)

#8.2.2. perform forward regression
data_reg_forward_1a <- step(lm1a_train, direction =
"forward")
summary(data_reg_forward_1a)
data_reg_forward_pred_1a <- predict(data_reg_forward_1a,
test1)
accuracy(data_reg_forward_pred_1a,
test1$net_profit_margin)

#8.2.3. perform backward regression
data_reg_backward_1a <- step(lm1a_train, direction =
"backward")
summary(data_reg_backward_1a)
data_reg_backward_pred_1a <-
predict(data_reg_backward_1a, test1)
accuracy(data_reg_backward_pred_1a,
test1$net_profit_margin)

#8.2.4. perform regression with fixed effects
data_reg_fixeff_1a <- felm(net_profit_margin ~ salary_at + at
+ debt_to_equity_ratio + ceo_years | gvkey + fyear |0 |
gvkey + fyear, train1)
summary(data_reg_fixeff_1a)
```

```
#8.2.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_1a <- car::vif(lm1a_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#8.2.6 resdiual analysis
resid_1a <- lm1a_train$residuals
plot(train1$salary_at, resid_1a) #relationship between
salary_at and residuals

fnorm_1a <- fitdist(resid_1a, "norm")
result_1a <- gofstat(fnorm_1a, discrete = FALSE)
result_1a
kscritvalue_1a <- 1.36/sqrt(length(train1$net_profit_margin))
kscritvalue_1a #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_1a)
plot(fnorm_1a) #residuals are not normally distributed
confint(lm1a_train, level = 0.95)

#8.2.7. Using stargazer for well-formatted regression output
stargazer(lm1a_train, data_reg_forward_1a,
data_reg_backward_1a, data_reg_stepwise_1a,
type="text",title="Regression Results",omit = c("Constant"),
digits=4,  no.space = TRUE, out="table1a.txt")

#8.3. regressing salary_at with return of asset; control
variables: total assets, debt to equity ratio, and CEO years
lm1b_train <- lm(roa ~ salary_at + at + debt_to_equity_ratio
+ ceo_years, train1)

#8.3.1. perform stepwise regression
data_reg_stepwise_1b <- step(lm1b_train, direction = "both")
summary(data_reg_stepwise_1b)
data_reg_stepwise_pred_1b <-
predict(data_reg_stepwise_1b, test1)
accuracy(data_reg_stepwise_pred_1b, test1$roa)

#8.3.2. perform forward regression
data_reg_forward_1b <- step(lm1b_train, direction =
"forward")
summary(data_reg_forward_1b)
data_reg_forward_pred_1b <- predict(data_reg_forward_1b,
test1)
accuracy(data_reg_forward_pred_1b, test1$roa)

#8.3.3. perform backward regression
data_reg_backward_1b <- step(lm1b_train, direction =
"backward")
summary(data_reg_backward_1b)
data_reg_backward_pred_1b <-
predict(data_reg_backward_1b, test1)
```

```
accuracy(data_reg_backward_pred_1b, test1$roa)

#8.3.4. perform regression with fixed effects
data_reg_fixeff_1b <- felm(roa ~ salary_at + at +
debt_to_equity_ratio + ceo_years | gvkey + fyear |0 | gvkey
+ fyear, train1)
summary(data_reg_fixeff_1b)

#8.3.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_1b <- car::vif(lm1b_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#8.3.6 resdiual analysis
resid_1b <- lm1b_train$residuals
plot(train1$salary_at, resid_1b) #relationship between
salary_at and residuals

fnorm_1b <- fitdist(resid_1b, "norm")
result_1b <- gofstat(fnorm_1b, discrete = FALSE)
result_1b
kscritvalue_1b <- 1.36/sqrt(length(train1$roa))
kscritvalue_1b #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_1b)
plot(fnorm_1b) #residuals are not normally distributed
confint(lm1b_train, level = 0.95)

#8.3.7. Using stargazer for well-formatted regression output
stargazer(lm1b_train, data_reg_forward_1b,
data_reg_backward_1b, data_reg_stepwise_1b,
type="text",title="Regression Results",omit = c("Constant"),
digits=4,  no.space = TRUE, out="table1b.txt")

#8.4. regressing salary_at with current ratio; control
variables: total assets, debt to equity ratio, and asset
turnover
lm1c_train <- lm(current_ratio ~ salary_at + at +
debt_to_equity_ratio + asset_turnover, train1)

#8.4.1. perform stepwise regression
data_reg_stepwise_1c <- step(lm1c_train, direction = "both")
summary(data_reg_stepwise_1c)
data_reg_stepwise_pred_1c <-
predict(data_reg_stepwise_1c, test1)
accuracy(data_reg_stepwise_pred_1c, test1$current_ratio)

#8.4.2. perform forward regression
data_reg_forward_1c <- step(lm1c_train, direction =
"forward")
summary(data_reg_forward_1c)
data_reg_forward_pred_1c <- predict(data_reg_forward_1c,
test1)
accuracy(data_reg_forward_pred_1c, test1$current_ratio)

#8.4.3. perform backward regression
data_reg_backward_1c <- step(lm1c_train, direction =
"backward")
summary(data_reg_backward_1c)
data_reg_backward_pred_1c <-
predict(data_reg_backward_1c, test1)
accuracy(data_reg_backward_pred_1c, test1$current_ratio)

#8.4.4. perform regression with fixed effects
data_reg_fixeff_1c <- felm(current_ratio ~ salary_at + at +
debt_to_equity_ratio + asset_turnover | gvkey + fyear |0 |
gvkey + fyear, train1)
summary(data_reg_fixeff_1c)

#8.4.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_1c <- car::vif(lm1c_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#8.4.6 resdiual analysis
resid_1c <- lm1c_train$residuals
plot(train1$salary_at, resid_1c) #relationship between
salary_at and residuals

fnorm_1c <- fitdist(resid_1c, "norm")
result_1c <- gofstat(fnorm_1c, discrete = FALSE)
result_1c
kscritvalue_1c <- 1.36/sqrt(length(train1$current_ratio))
kscritvalue_1c #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_1c)
plot(fnorm_1c) #residuals are not normally distributed
confint(lm1c_train, level = 0.95)

#8.4.7. Using stargazer for well-formatted regression output
stargazer(lm1c_train, data_reg_forward_1c,
data_reg_backward_1c, data_reg_stepwise_1c,
type="text",title="Regression Results",omit = c("Constant"),
digits=4,  no.space = TRUE, out="table1c.txt")

#8.5. regressing salary_at with debt to equity ratio; control
variables: total assets, net profit margin, and asset turnover
lm1d_train <- lm(debt_to_equity_ratio ~ salary_at + at +
net_profit_margin + asset_turnover, train1)

#8.5.1. perform stepwise regression
data_reg_stepwise_1d <- step(lm1d_train, direction = "both")
summary(data_reg_stepwise_1d)
```

```
data_reg_stepwise_pred_1d <-
predict(data_reg_stepwise_1d, test1)
accuracy(data_reg_stepwise_pred_1d,
test1$debt_to_equity_ratio)

#8.5.2. perform forward regression
data_reg_forward_1d <- step(lm1d_train, direction =
"forward")
summary(data_reg_forward_1d)
data_reg_forward_pred_1d <- predict(data_reg_forward_1d,
test1)
accuracy(data_reg_forward_pred_1d,
test1$debt_to_equity_ratio)

#8.5.3. perform backward regression
data_reg_backward_1d <- step(lm1d_train, direction =
"backward")
summary(data_reg_backward_1d)
data_reg_backward_pred_1d <-
predict(data_reg_backward_1d, test1)
accuracy(data_reg_backward_pred_1d,
test1$debt_to_equity_ratio)

#8.5.4. perform regression with fixed effects
data_reg_fixeff_1d <- felm(debt_to_equity_ratio ~ salary_at
+ at + net_profit_margin + asset_turnover | gvkey + fyear |0
| gvkey + fyear, train1)
summary(data_reg_fixeff_1d)

#8.5.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_1d <- car::vif(lm1d_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#8.5.6 resdiual analysis
resid_1d <- lm1d_train$residuals
plot(train1$salary_at, resid_1d) #relationship between
salary_at and residuals

fnorm_1d <- fitdist(resid_1d, "norm")
result_1d <- gofstat(fnorm_1d, discrete = FALSE)
result_1d
kscritvalue_1d <-
1.36/sqrt(length(train1$debt_to_equity_ratio))
kscritvalue_1d #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_1d)
plot(fnorm_1d) #residuals are not normally distributed
confint(lm1d_train, level = 0.95)

#8.5.7. Using stargazer for well-formatted regression output
stargazer(lm1d_train, data_reg_forward_1d,
data_reg_backward_1d, data_reg_stepwise_1d,
type="text",title="Regression Results",omit = c("Constant"),
digits=4,  no.space = TRUE, out="table1d.txt")

#9 multi-linear regression to analyse the effects of bonus on
profit margin, roa, liquidity ratio, debt to equity ratio
#Dependent variable: net_profit_margin, roa, current_ratio,
debt_to_equity_ratio
#Independent variable: bonus_at (bonus/total assets)

#9.1. choosing training and test dataset
set.seed(1)
train2 <- sample_frac(data_reg_3d_clean, 0.75)
test2 <- anti_join(data_reg_3d_clean, train2)

#9.2. regressing salary_at with profit margin; control
varaibles: at, debt to equity ratio, ceo_years
lm2a_train <- lm(net_profit_margin ~ bonus_at + at +
debt_to_equity_ratio + ceo_years, train2)

#9.2.1. perform stepwise regression
data_reg_stepwise_2a <- step(lm2a_train, direction = "both")
summary(data_reg_stepwise_2a)
data_reg_stepwise_pred_2a <-
predict(data_reg_stepwise_2a, test2)
accuracy(data_reg_stepwise_pred_2a,
test2$net_profit_margin)

#9.2.2. perform forward regression
data_reg_forward_2a <- step(lm2a_train, direction =
"forward")
summary(data_reg_forward_2a)
data_reg_forward_pred_2a <- predict(data_reg_forward_2a,
test2)
accuracy(data_reg_forward_pred_2a,
test2$net_profit_margin)

#9.2.3. perform backward regression
data_reg_backward_2a <- step(lm2a_train, direction =
"backward")
summary(data_reg_backward_2a)
data_reg_backward_pred_2a <-
predict(data_reg_backward_2a, test2)
accuracy(data_reg_backward_pred_2a,
test2$net_profit_margin)

#9.2.4. perform regression with fixed effects
data_reg_fixeff_2a <- felm(net_profit_margin ~ bonus_at + at
+ debt_to_equity_ratio + ceo_years | gvkey + fyear |0 |
gvkey + fyear, train2)
summary(data_reg_fixeff_2a)
```

#9.2.5. Checking for multicollinearity of controlled variables with VIF
vif_values_2a <- car::vif(lm2a_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#9.2.6 resdiual analysis
resid_2a <- lm2a_train$residuals
plot(train2$bonus_at, resid_2a) #relationship between salary_at and residuals

fnorm_2a <- fitdist(resid_2a, "norm")
result_2a <- gofstat(fnorm_2a, discrete = FALSE)
result_2a
kscritvalue_2a <- 1.36/sqrt(length(train2$net_profit_margin))
kscritvalue_2a #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_2a)
plot(fnorm_2a) #residuals are not normally distributed
confint(lm2a_train, level = 0.95)

#9.2.7. Using stargazer for well-formatted regression output
stargazer(lm2a_train, data_reg_forward_2a, data_reg_backward_2a, data_reg_stepwise_2a, type="text",title="Regression Results",omit = c("Constant"), digits=4, no.space = TRUE, out="table2a.txt")

#9.3. regressing bonus_at with return of asset; control variables: total assets, debt to equity ratio, and CEO years
lm2b_train <- lm(roa ~ bonus_at + at + debt_to_equity_ratio + ceo_years, train2)

#9.3.1. perform stepwise regression
data_reg_stepwise_2b <- step(lm2b_train, direction = "both")
summary(data_reg_stepwise_2b)
data_reg_stepwise_pred_2b <- predict(data_reg_stepwise_2b, test2)
accuracy(data_reg_stepwise_pred_2b, test2$roa)

#9.3.2. perform forward regression
data_reg_forward_2b <- step(lm2b_train, direction = "forward")
summary(data_reg_forward_2b)
data_reg_forward_pred_2b <- predict(data_reg_forward_2b, test2)
accuracy(data_reg_forward_pred_2b, test2$roa)

#9.3.3. perform backward regression
data_reg_backward_2b <- step(lm2b_train, direction = "backward")
summary(data_reg_backward_2b)
data_reg_backward_pred_2b <- predict(data_reg_backward_2b, test2)
accuracy(data_reg_backward_pred_2b, test2$roa)

#9.3.4. perform regression with fixed effects
data_reg_fixeff_2b <- felm(roa ~ bonus_at + inventory_turnover + asset_turnover + ceo_years | gvkey + fyear |0 | gvkey + fyear, train2)
summary(data_reg_fixeff_2b)

#9.3.5. Checking for multicollinearity of controlled variables with VIF
vif_values_2b <- car::vif(lm2b_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#9.3.6 resdiual analysis
resid_2b <- lm2b_train$residuals
plot(train2$bonus_at, resid_2b) #relationship between salary_at and residuals

fnorm_2b <- fitdist(resid_2b, "norm")
result_2b <- gofstat(fnorm_2b, discrete = FALSE)
result_2b
kscritvalue_2b <- 1.36/sqrt(length(train2$roa))
kscritvalue_2b #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_2b)
plot(fnorm_2b) #residuals are not normally distributed
confint(lm2b_train, level = 0.95)

#9.3.7. Using stargazer for well-formatted regression output
stargazer(lm2b_train, data_reg_forward_2b, data_reg_backward_2b, data_reg_stepwise_2b, type="text",title="Regression Results",omit = c("Constant"), digits=4, no.space = TRUE, out="table2b.txt")

#9.4. regressing bonus_at with current ratio; control variables: total assets, debt to equity ratio, and asset turnover
lm2c_train <- lm(current_ratio ~ bonus_at + at + debt_to_equity_ratio + asset_turnover, train2)

#9.4.1. perform stepwise regression
data_reg_stepwise_2c <- step(lm2c_train, direction = "both")
summary(data_reg_stepwise_2c)
data_reg_stepwise_pred_2c <- predict(data_reg_stepwise_2c, test2)
accuracy(data_reg_stepwise_pred_2c, test2$current_ratio)

#9.4.2. perform forward regression
data_reg_forward_2c <- step(lm2c_train, direction = "forward")
summary(data_reg_forward_2c)

```
data_reg_forward_pred_2c <- predict(data_reg_forward_2c,
test2)
accuracy(data_reg_forward_pred_2c, test2$current_ratio)

#9.4.3. perform backward regression
data_reg_backward_2c <- step(lm2c_train, direction =
"backward")
summary(data_reg_backward_2c)
data_reg_backward_pred_2c <-
predict(data_reg_backward_2c, test2)
accuracy(data_reg_backward_pred_2c, test2$current_ratio)

#9.4.4. perform regression with fixed effects
data_reg_fixeff_2c <- felm(current_ratio ~ bonus_at + at +
debt_to_equity_ratio + asset_turnover | gvkey + fyear |0 |
gvkey + fyear, train2)
summary(data_reg_fixeff_2c)

#9.4.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_2c <- car::vif(lm2c_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#9.4.6 resdiual analysis
resid_2c <- lm2c_train$residuals
plot(train2$bonus_at, resid_2c) #relationship between
salary_at and residuals

fnorm_2c <- fitdist(resid_2c, "norm")
result_2c <- gofstat(fnorm_2c, discrete = FALSE)
result_2c
kscritvalue_2c <- 1.36/sqrt(length(train2$current_ratio))
kscritvalue_2c #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_2c)
plot(fnorm_2c) #residuals are not normally distributed
confint(lm2c_train, level = 0.95)

#9.4.7. Using stargazer for well-formatted regression output
stargazer(lm2c_train, data_reg_forward_2c,
data_reg_backward_2c, data_reg_stepwise_2c,
type="text",title="Regression Results",omit = c("Constant"),
digits=4,  no.space = TRUE, out="table2c.txt")

#9.5. regressing salary_at with debt to equity ratio; control
variables: total assets, net profit margin, and asset turnover
lm2d_train <- lm(debt_to_equity_ratio ~ bonus_at + at +
net_profit_margin + asset_turnover, train2)

#9.5.1. perform stepwise regression
data_reg_stepwise_2d <- step(lm2d_train, direction = "both")
summary(data_reg_stepwise_2d)
data_reg_stepwise_pred_2d <-
predict(data_reg_stepwise_2d, test2)
accuracy(data_reg_stepwise_pred_2d,
test2$debt_to_equity_ratio)

#9.5.2. perform forward regression
data_reg_forward_2d <- step(lm2d_train, direction =
"forward")
summary(data_reg_forward_2d)
data_reg_forward_pred_2d <- predict(data_reg_forward_2d,
test2)
accuracy(data_reg_forward_pred_2d,
test2$debt_to_equity_ratio)

#9.5.3. perform backward regression
data_reg_backward_2d <- step(lm2d_train, direction =
"backward")
summary(data_reg_backward_2d)
data_reg_backward_pred_2d <-
predict(data_reg_backward_2d, test2)
accuracy(data_reg_backward_pred_2d,
test2$debt_to_equity_ratio)

#9.5.4. perform regression with fixed effects
data_reg_fixeff_2d <- felm(debt_to_equity_ratio ~ bonus_at
+ at + net_profit_margin + asset_turnover | gvkey + fyear |0
| gvkey + fyear, train2)
summary(data_reg_fixeff_2d)

#9.5.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_2d <- car::vif(lm2d_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#9.5.6 residual analysis
resid_2d <- lm2d_train$residuals
plot(train2$bonus_at, resid_2d) #relationship between
salary_at and residuals

fnorm_2d <- fitdist(resid_2d, "norm")
result_2d <- gofstat(fnorm_2d, discrete = FALSE)
result_2d
kscritvalue_2d <-
1.36/sqrt(length(train2$debt_to_equity_ratio))
kscritvalue_2d #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_2d)
plot(fnorm_2d) #residuals are not normally distributed
confint(lm2d_train, level = 0.95)

#9.5.7. Using stargazer for well-formatted regression output
```

```
stargazer(lm2d_train, data_reg_forward_2d,
data_reg_backward_2d, data_reg_stepwise_2d,
type="text",title="Regression Results",omit = c("Constant"),
digits=4,  no.space = TRUE, out="table2d.txt")

#10 multi-linear regression to analyse the effects of other
compensation on profit margin, roa, liquidity ratio, debt to
equity ratio
#Dependent variable: net_profit_margin, roa, current_ratio,
debt_to_equity_ratio
#Independent variable: othcomp_at (other
compensation/total assets)

#10.1. choosing training and test dataset
set.seed(1)
train3 <- sample_frac(data_reg_3d_clean, 0.75)
test3 <- anti_join(data_reg_3d_clean, train3)

#10.2. regressing othcomp_at with profit margin; control
varaibles: at, debt to equity ratio, ceo_years
lm3a_train <- lm(net_profit_margin ~ othcomp_at + at +
debt_to_equity_ratio + ceo_years, train3)

# 10.2.1. perform stepwise regression
data_reg_stepwise_3a <- step(lm3a_train, direction = "both")
summary(data_reg_stepwise_3a)
data_reg_stepwise_pred_3a <-
predict(data_reg_stepwise_3a, test3)
accuracy(data_reg_stepwise_pred_3a,
test3$net_profit_margin)

# 10.2.2. perform forward regression
data_reg_forward_3a <- step(lm3a_train, direction =
"forward")
summary(data_reg_forward_3a)
data_reg_forward_pred_3a <- predict(data_reg_forward_3a,
test3)
accuracy(data_reg_forward_pred_3a,
test3$net_profit_margin)

# 10.2.3. perform backward regression
data_reg_backward_3a <- step(lm3a_train, direction =
"backward")
summary(data_reg_backward_3a)
data_reg_backward_pred_3a <-
predict(data_reg_backward_3a, test3)
accuracy(data_reg_backward_pred_3a,
test3$net_profit_margin)

#10.2.4. perform regression with fixed effects
data_reg_fixeff_3a <- felm(net_profit_margin ~ othcomp_at
+ at + debt_to_equity_ratio + ceo_years | gvkey + fyear |0 |
gvkey + fyear, train3)
summary(data_reg_fixeff_3a)

#10.2.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_3a <- car::vif(lm3a_train) #no multicollinearity
issues as VIF values remain low across variables

#10.2.6 resdiual analysis
lm3a <- lm(net_profit_margin ~ othcomp_at, train3) #simple
linear regression with only other compensation
summary(lm3a)
plot(train3$othcomp_at, train3$net_profit_margin,
main="Relationship between
    other compensation and profit margin",
    xlab="othcomp_at", ylab="Profit Margin")
abline(lm3a, lwd=3, col="red")

resid_3a <- lm3a$residuals
plot(train3$othcomp_at, resid_3a) #relationship between
othcomp_at and residuals

fnorm_3a <- fitdist(resid_3a, "norm")
result_3a <- gofstat(fnorm_3a, discrete = FALSE)
result_3a
kscritvalue_3a <- 1.36/sqrt(length(train3$net_profit_margin))
kscritvalue_3a #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_3a)
plot(fnorm_3a) #residuals are not normally distributed
confint(lm3a_train, level = 0.95)

#10.2.7. Using stargazer for well-formatted regression output
stargazer(lm3a_train, data_reg_forward_3a,
data_reg_backward_3a, data_reg_stepwise_3a,
type="text",title="Regression Results", digits=4,  no.space =
TRUE, out="table3a.txt")

#10.3. regressing othcomp_at with roa; control varaibles: at,
debt to equity ratio, ceo_years
lm3b_train <- lm(roa ~ othcomp_at + at +
debt_to_equity_ratio + ceo_years, train3)

# 10.3.1. perform stepwise regression
data_reg_stepwise_3b <- step(lm3b_train, direction = "both")
summary(data_reg_stepwise_3b)
data_reg_stepwise_pred_3b <-
predict(data_reg_stepwise_3b, test3)
accuracy(data_reg_stepwise_pred_3b, test3$roa)

# 10.3.2. perform forward regression
data_reg_forward_3b <- step(lm3b_train, direction =
"forward")
summary(data_reg_forward_3b)
```

```
data_reg_forward_pred_3b <- predict(data_reg_forward_3b, test3)
accuracy(data_reg_forward_pred_3b, test3$roa)

# 10.3.3. perform backward regression
data_reg_backward_3b <- step(lm3b_train, direction = "backward")
summary(data_reg_backward_3b)
data_reg_backward_pred_3b <-
predict(data_reg_backward_3b, test3)
accuracy(data_reg_backward_pred_3b, test3$roa)

#10.3.4. perform regression with fixed effects
data_reg_fixeff_3b <- felm(roa ~ othcomp_at + at + debt_to_equity_ratio + ceo_years | gvkey + fyear |0 | gvkey + fyear, train3)
summary(data_reg_fixeff_3b)

#10.3.5. Checking for multicollinearity of controlled variables with VIF
vif_values_3b <- car::vif(lm3b_train) #no multicollinearity issues as VIF values remain low across variables

#10.3.6 resdiual analysis
lm3b <- lm(roa ~ othcomp_at, train3) #simple linear regression with only other compensation
summary(lm3b)
plot(train3$othcomp_at, train3$roa, main="Relationship between
    other compensation and ROA",
    xlab="othcomp_at", ylab="roa")
abline(lm3b, lwd=3, col="red")

resid_3b <- lm3b$residuals
plot(train3$othcomp_at, resid_3b) #relationship between othcomp_at and residuals

fnorm_3b <- fitdist(resid_3b, "norm")
result_3b <- gofstat(fnorm_3b, discrete = FALSE)
result_3b
kscritvalue_3b <- 1.36/sqrt(length(train3$roa))
kscritvalue_3b #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_3b)
plot(fnorm_3b) #residuals are not normally distributed
confint(lm3b, level = 0.95)

#10.3.7. Using stargazer for well-formatted regression output
stargazer(lm3b_train, data_reg_forward_3b, data_reg_backward_3b, data_reg_stepwise_3b, type="text",title="Regression Results", digits=4,  no.space = TRUE, out="table3b.txt")

#10.4. regressing othcomp_at with current ratio; control varaibles: at, debt to equity ratio, asset turnover
lm3c_train <- lm(current_ratio ~ othcomp_at + at + debt_to_equity_ratio + asset_turnover, train3)

# 10.4.1. perform stepwise regression
data_reg_stepwise_3c <- step(lm3c_train, direction = "both")
summary(data_reg_stepwise_3c)
data_reg_stepwise_pred_3c <-
predict(data_reg_stepwise_3c, test3)
accuracy(data_reg_stepwise_pred_3c, test3$current_ratio)

# 10.4.2. perform forward regression
data_reg_forward_3c <- step(lm3c_train, direction = "forward")
summary(data_reg_forward_3c)
data_reg_forward_pred_3c <- predict(data_reg_forward_3c, test3)
accuracy(data_reg_forward_pred_3c, test3$current_ratio)

# 10.4.3. perform backward regression
data_reg_backward_3c <- step(lm3c_train, direction = "backward")
summary(data_reg_backward_3c)
data_reg_backward_pred_3c <-
predict(data_reg_backward_3c, test3)
accuracy(data_reg_backward_pred_3c, test3$current_ratio)

#10.4.4. perform regression with fixed effects
data_reg_fixeff_3c <- felm(current_ratio ~ othcomp_at + at + debt_to_equity_ratio + asset_turnover | gvkey + fyear |0 | gvkey + fyear, train3)
summary(data_reg_fixeff_3c)

#10.4.5. Checking for multicollinearity of controlled variables with VIF
vif_values_3c <- car::vif(lm3c_train) #no multicollinearity issues as VIF values remain low across variables

#10.4.6 resdiual analysis
lm3c <- lm(current_ratio ~ othcomp_at, train3) #simple linear regression with only other compensation
summary(lm3c)
plot(train3$othcomp_at, train3$current_ratio, main="Relationship between
    other compensation and current ratio",
    xlab="othcomp_at", ylab="current ratio")
abline(lm3c, lwd=3, col="red")

resid_3c <- lm3c$residuals
plot(train3$othcomp_at, resid_3c) #relationship between othcomp_at and residuals

fnorm_3c <- fitdist(resid_3c, "norm")
```

```
result_3c <- gofstat(fnorm_3c, discrete = FALSE)
result_3c
kscritvalue_3c <- 1.36/sqrt(length(train3$current_ratio))
kscritvalue_3c #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_3c)
plot(fnorm_3c) #residuals are not normally distributed
confint(lm3c, level = 0.95)

#10.4.7. Using stargazer for well-formatted regression output
stargazer(lm3c_train, data_reg_forward_3c, data_reg_backward_3c, data_reg_stepwise_3c, type="text",title="Regression Results", digits=4,  no.space = TRUE, out="table3c.txt")


#10.5. regressing othcomp_at with debt to equity ratio; control varaibles: at, profit margin, asset turnover
lm3d_train <- lm(debt_to_equity_ratio ~ othcomp_at + at + net_profit_margin + asset_turnover, train3)
summary(lm3d_train)

# 10.5.1. perform stepwise regression
data_reg_stepwise_3d <- step(lm3d_train, direction = "both")
summary(data_reg_stepwise_3d)
data_reg_stepwise_pred_3d <- predict(data_reg_stepwise_3d, test3)
accuracy(data_reg_stepwise_pred_3d, test3$debt_to_equity_ratio)

# 10.5.2. perform forward regression
data_reg_forward_3d <- step(lm3d_train, direction = "forward")
summary(data_reg_forward_3d)
data_reg_forward_pred_3d <- predict(data_reg_forward_3d, test3)
accuracy(data_reg_forward_pred_3d, test3$debt_to_equity_ratio)

# 10.5.3. perform backward regression
data_reg_backward_3d <- step(lm3d_train, direction = "backward")
summary(data_reg_backward_3d)
data_reg_backward_pred_3d <- predict(data_reg_backward_3d, test3)
accuracy(data_reg_backward_pred_3d, test3$debt_to_equity_ratio)

#10.5.4. perform regression with fixed effects
data_reg_fixeff_3d <- felm(debt_to_equity_ratio ~ othcomp_at + at + net_profit_margin + asset_turnover | gvkey + fyear |0 | gvkey + fyear, train3)
summary(data_reg_fixeff_3d)

#10.5.5. Checking for multicollinearity of controlled variables with VIF
vif_values_3d <- car::vif(lm3d_train) #no multicollinearity issues as VIF values remain low across variables

#10.5.6 resdiual analysis
lm3d <- lm(debt_to_equity_ratio ~ othcomp_at, train3) #simple linear regression with only other compensation
summary(lm3d)
plot(train3$othcomp_at, train3$debt_to_equity_ratio, main="Relationship between
    other compensation and debt to equity ratio",
    xlab="othcomp_at", ylab="Debt to equity ratio")
abline(lm3d, lwd=3, col="red")

resid_3d <- lm3d$residuals
plot(train3$othcomp_at, resid_3d) #relationship between othcomp_at and residuals

fnorm_3d <- fitdist(resid_3d, "norm")
result_3d <- gofstat(fnorm_3d, discrete = FALSE)
result_3d
kscritvalue_3d <- 1.36/sqrt(length(train3$debt_to_equity_ratio))
kscritvalue_3d #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_3d)
plot(fnorm_3d) #residuals are not normally distributed
confint(lm3d, level = 0.95)

#10.5.7. Using stargazer for well-formatted regression output
stargazer(lm3d_train, data_reg_forward_3d, data_reg_backward_3d, data_reg_stepwise_3d, type="text",title="Regression Results", digits=4,  no.space = TRUE, out="table3d.txt")

#11 multi-linear regression to analyse the effects of restricted stock grant (RSTKGRNT) on profit margin, roa, liquidity ratio, debt to equity ratio

#Dependent variable: net_profit_margin, roa, current_ratio, debt_to_equity_ratio
#Independent variable: RSTKGRNT (restricted stock grant)

#11.1. Replacing NAs with 0s
data_reg_4d_clean <- data_reg_3d_clean %>% mutate(RSTKGRNT = ifelse(is.na(RSTKGRNT),0,RSTKGRNT))

#11.1.2. choosing training and test dataset
set.seed(1)
train4 <- sample_frac(data_reg_4d_clean, 0.75)
test4 <- anti_join(data_reg_4d_clean, train4)
```

```
#11.2. regressing RSTKGRNT with profit margin; control varaibles: at, debt to equity ratio, ceo_years
lm4a_train <- lm(net_profit_margin ~ RSTKGRNT + at + debt_to_equity_ratio + ceo_years, train4)

# 11.2.1. perform stepwise regression
data_reg_stepwise_4a <- step(lm4a_train, direction = "both")
summary(data_reg_stepwise_4a)
data_reg_stepwise_pred_4a <- predict(data_reg_stepwise_4a, test4)
accuracy(data_reg_stepwise_pred_4a, test4$net_profit_margin)

# 11.2.2. perform forward regression
data_reg_forward_4a <- step(lm4a_train, direction = "forward")
summary(data_reg_forward_4a)
data_reg_forward_pred_4a <- predict(data_reg_forward_4a, test4)
accuracy(data_reg_forward_pred_4a, test4$net_profit_margin)

# 11.2.3. perform backward regression
data_reg_backward_4a <- step(lm4a_train, direction = "backward")
summary(data_reg_backward_4a)
data_reg_backward_pred_4a <- predict(data_reg_backward_4a, test4)
accuracy(data_reg_backward_pred_4a, test4$net_profit_margin)

#11.2.4. perform regression with fixed effects
data_reg_fixeff_4a <- felm(net_profit_margin ~ RSTKGRNT + debt_to_equity_ratio + ceo_years | gvkey + fyear |0 | gvkey + fyear, train4)
summary(data_reg_fixeff_4a)

#11.2.5. Checking for multicollinearity of controlled variables with VIF
vif_values_4a <- car::vif(lm4a_train) #no multicollinearity issues as VIF values remain low across variables

#11.2.6 resdiual analysis
resid_4a<- lm4a_train$residuals
plot(train4$RSTKGRNT, resid_4a) #relationship between RSTKGRNT and residuals

fnorm_4a <- fitdist(resid_4a, "norm")
result_4a <- gofstat(fnorm_4a, discrete = FALSE)
result_4a
kscritvalue_4a <- 1.36/sqrt(length(train4$net_profit_margin))
kscritvalue_4a #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_4a)
plot(fnorm_4a) #residuals are not normally distributed
confint(lm4a_train, level = 0.95)

#11.2.7. Using stargazer for well-formatted regression output
stargazer(lm4a_train, data_reg_forward_4a, data_reg_backward_4a, data_reg_stepwise_4a, type="text",title="Regression Results",omit = c("Constant"), digits=4,  no.space = TRUE, out="table4a.txt")

#11.3. regressing RSTKGRNT with ROA ; control variables: ceo_years, inventory turnover, asset turnover; gvkey + fyear

lm4b_train <- lm(roa ~ RSTKGRNT + inventory_turnover + asset_turnover + ceo_years, train4)

# 11.3.1. perform stepwise regression
data_reg_stepwise_4b <- step(lm4b_train, direction = "both")
summary(data_reg_stepwise_4b)
data_reg_stepwise_pred_4b <- predict(data_reg_stepwise_4b, test4)
accuracy(data_reg_stepwise_pred_4b, test4$roa)

# 11.3.2. perform forward regression
data_reg_forward_4b <- step(lm4b_train, direction = "forward")
summary(data_reg_forward_4b)
data_reg_forward_pred_4b <- predict(data_reg_forward_4b, test4)
accuracy(data_reg_forward_pred_4b, test4$roa)

# 11.3.3. perform backward regression
data_reg_backward_4b <- step(lm4b_train, direction = "backward")
summary(data_reg_backward_4b)
data_reg_backward_pred_4b <- predict(data_reg_backward_4b, test4)
accuracy(data_reg_backward_pred_4b, test4$roa)

#11.3.4. perform regression with fixed effects
data_reg_fixeff_4b <- felm(roa ~ RSTKGRNT + inventory_turnover + asset_turnover + ceo_years | gvkey + fyear |0 | gvkey + fyear, train4)
summary(data_reg_fixeff_4b)

#11.3.5. Checking for multicollinearity of controlled variables with VIF
vif_values_4b <- car::vif(lm3a_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables
```

```
#11.3.6 resdiual analysis
resid_4b <- lm4b_train$residuals
plot(train4$RSTKGRNT, resid_4b) #relationship between
RSTKGRNT and residuals

fnorm_4b <- fitdist(resid_4b, "norm")
result_4b <- gofstat(fnorm_4b, discrete = FALSE)
result_4b
kscritvalue_4b<- 1.36/sqrt(length(train4$roa))
kscritvalue_4b #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_4b)
plot(fnorm_4b) #residuals are not normally distributed
confint(lm4b_train, level = 0.95)

#11.3.7. Using stargazer for well-formatted regression output
stargazer(lm4b_train, data_reg_forward_4b,
data_reg_backward_4b, data_reg_stepwise_4b,
type="text",title="Regression Results",omit = c("Constant"),
digits=4,  no.space = TRUE, out="table4b.txt")

 #11.4. regressing RSTKGRNT with Current ratio (liquidity) ;
control variables:  asset turnover, at, debt to equity ratio;
gvkey + fyear

  lm4c_train <- lm(current_ratio ~ RSTKGRNT + at +
debt_to_equity_ratio + asset_turnover, train4)

# 11.4.1. perform stepwise regression
data_reg_stepwise_4c <- step(lm4c_train, direction = "both")
summary(data_reg_stepwise_4c)
data_reg_stepwise_pred_4c <-
predict(data_reg_stepwise_4c, test4)
accuracy(data_reg_stepwise_pred_4c, test4$current_ratio)

# 11.4.2. perform forward regression
data_reg_forward_4c <- step(lm4c_train, direction =
"forward")
summary(data_reg_forward_4c)
data_reg_forward_pred_4c <- predict(data_reg_forward_4c,
test4)
accuracy(data_reg_forward_pred_4c, test4$current_ratio)

# 11.4.3. perform backward regression
data_reg_backward_4c <- step(lm4c_train, direction =
"backward")
summary(data_reg_backward_4c)
data_reg_backward_pred_4c <-
predict(data_reg_backward_4c, test4)
accuracy(data_reg_backward_pred_4c, test4$current_ratio)

#11.4.4. perform regression with fixed effects
data_reg_fixeff_4c <- felm(current_ratio ~ RSTKGRNT + at +
debt_to_equity_ratio + asset_turnover | gvkey + fyear |0 |
gvkey + fyear, train4)
summary(data_reg_fixeff_4c)

#11.4.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_4c <- car::vif(lm4b_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#11.4.6 resdiual analysis
resid_4c <- lm4c_train$residuals
plot(train4$RSTKGRNT, resid_4c) #relationship between
RSTKGRNT and residuals

fnorm_4c<- fitdist(resid_4c, "norm")
result_4c <- gofstat(fnorm_4c, discrete = FALSE)
result_4c
kscritvalue_4c <- 1.36/sqrt(length(train4$current_ratio))
kscritvalue_4c #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_4c)
plot(fnorm_4c) #residuals are not normally distributed
confint(lm4c_train, level = 0.95)

#11.4.7. Using stargazer for well-formatted regression output
stargazer(lm4c_train, data_reg_forward_4c,
data_reg_backward_4c, data_reg_stepwise_4c,
type="text",title="Regression Results",omit = c("Constant"),
digits=4,  no.space = TRUE, out="table4c.txt")



#11.5. regressing RSTKGRNT with Debt to equity ratio
(leverage) ; control variables:  at, net profit margin, asset
turnover; gvkey + fyear

lm4d_train <- lm(debt_to_equity_ratio ~ RSTKGRNT + at +
net_profit_margin + asset_turnover, train4)

# 11.5.1. perform stepwise regression
data_reg_stepwise_4d <- step(lm4d_train, direction = "both")
summary(data_reg_stepwise_4d)
data_reg_stepwise_pred_4d <-
predict(data_reg_stepwise_4d, test4)
accuracy(data_reg_stepwise_pred_4d,
test4$debt_to_equity_ratio)

# 11.5.2. perform forward regression
data_reg_forward_4d <- step(lm4d_train, direction =
"forward")
summary(data_reg_forward_4d)
```

```
data_reg_forward_pred_4d <- predict(data_reg_forward_4d, test4)
accuracy(data_reg_forward_pred_4d, test4$debt_to_equity_ratio)

# 11.5.3. perform backward regression
data_reg_backward_4d <- step(lm4d_train, direction = "backward")
summary(data_reg_backward_4d)
data_reg_backward_pred_4d <- predict(data_reg_backward_4d, test4)
accuracy(data_reg_backward_pred_4d, test4$debt_to_equity_ratio)

#11.5.4. perform regression with fixed effects
data_reg_fixeff_4d <- felm(debt_to_equity_ratio ~ RSTKGRNT + at + net_profit_margin + asset_turnover | gvkey + fyear |0 | gvkey + fyear, train4)
summary(data_reg_fixeff_4d)

#11.5.5. Checking for multicollinearity of controlled variables with VIF
vif_values_4d <- car::vif(lm4d_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#11.5.6 resdiual analysis
resid_4d <- lm4d_train$residuals
plot(train4$RSTKGRNT, resid_4d) #relationship between RSTKGRNT and residuals

fnorm_4d <- fitdist(resid_4d, "norm")
result_4d <- gofstat(fnorm_4d, discrete = FALSE)
result_4d
kscritvalue_4d <- 1.36/sqrt(length(train4$debt_to_equity_ratio))
kscritvalue_4d #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_4d)
plot(fnorm_4d) #residuals are not normally distributed
confint(lm4d_train, level = 0.95)

#11.5.7. Using stargazer for well-formatted regression output
stargazer(lm4d_train, data_reg_forward_4d, data_reg_backward_4d, data_reg_stepwise_4d, type="text",title="Regression Results",omit = c("Constant"), digits=4,  no.space = TRUE, out="table4d.txt")

#12 multi-linear regression to analyse the effects of other compensation on profit margin, roa, liquidity ratio, debt to equity ratio
#Dependent variable: net_profit_margin, roa, current_ratio, debt_to_equity_ratio
#Independent variable: OPTION_AWARDS_BLK_VALUE

#12.1.1 Replacing NAs with 0s
data_reg_5d_cleaner <- data_reg_3d_clean %>% mutate(OPTION_AWARDS_BLK_VALUE = ifelse(is.na(OPTION_AWARDS_BLK_VALUE),0,OPTION_AWARDS_BLK_VALUE))

#12.1.2 Choosing Training and Test Dataset
set.seed(1)
train5 <- sample_frac(data_reg_5d_cleaner, 0.75)
test5 <- anti_join(data_reg_5d_cleaner, train5)

#12.2 Net Profit Margin Regression Analysis

#12.2.1 regressing OPTION_AWARDS_BLK_VALUE with net profit margin; control variables: at, debt to equity ratio, ceo_years
lm5a_train <- lm(net_profit_margin ~ OPTION_AWARDS_BLK_VALUE + at + debt_to_equity_ratio + ceo_years, train5)
summary(lm5a_train)

#12.2.2 Stepwise Regression
data_reg_stepwise_5a <- step(lm5a_train, direction = "both")
summary(data_reg_stepwise_5a)
data_reg_stepwise_pred_5a <- predict(data_reg_stepwise_5a, test5)
accuracy(data_reg_stepwise_pred_5a, test5$net_profit_margin)
summary(data_reg_stepwise_pred_5a)

#12.2.3 Forward Regression
data_reg_forward_5a <- step(lm5a_train, direction = "forward")
summary(data_reg_forward_5a)
data_reg_forward_pred_5a <- predict(data_reg_forward_5a, test5)
accuracy(data_reg_forward_pred_5a, test5$net_profit_margin)

#12.2.4 Backward Regression
data_reg_backward_5a <- step(lm5a_train, direction = "backward")
summary(data_reg_backward_5a)
data_reg_backward_pred_5a <- predict(data_reg_backward_5a, test5)
accuracy(data_reg_backward_pred_5a, test5$net_profit_margin)

#12.2.5 perform regression with fixed effects
data_reg_fixeff_5a <- felm(net_profit_margin ~ OPTION_AWARDS_BLK_VALUE + at + debt_to_equity_ratio + ceo_years | gvkey + fyear | 0 | gvkey + fyear, train5)
```

```
summary(data_reg_fixeff_5a)

#12.2.6 Checking for multicollinearity of controlled variables with VIF
vif_values_5a <- car::vif(lm5a_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#12.2.7 residual analysis
resid_5a <- lm5a_train$residuals
plot(train5$OPTION_AWARDS_BLK_VALUE, resid_5a)
#relationship between othcomp_at and residuals

fnorm_5a <- fitdist(resid_5a, "norm")
result_5a <- gofstat(fnorm_5a, discrete = FALSE)
result_5a
kscritvalue_5a <- 1.36/sqrt(length(train5$net_profit_margin))
kscritvalue_5a #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_5a)
plot(fnorm_5a) #residuals are not normally distributed
confint(lm5a_train, level = 0.95)

#12.2.8 Using stargazer for well-formatted regression output
list5a <- list(lm5a_train, data_reg_forward_5a, data_reg_backward_5a, data_reg_stepwise_5a)
stargazer(list5a, type="text",title="Regression Results", digits=4, omit = c("Constant"), no.space = TRUE, out="table5a.txt")

#12.3 ROA Regression Analysis

#12.3.1 regressing OPTION_AWARDS_BLK_VALUE with roa; control variables: at, debt to equity ratio and ceo years
lm5b_train <- lm(roa ~ OPTION_AWARDS_BLK_VALUE + at + debt_to_equity_ratio + ceo_years, train5)
summary(lm5b_train)

#12.3.2 Stepwise Regression
data_reg_stepwise_5b <- step(lm5b_train, direction = "both")
summary(data_reg_stepwise_5b)
data_reg_stepwise_pred_5b <- predict(data_reg_stepwise_5b, test5)
accuracy(data_reg_stepwise_pred_5b, test5$roa)
summary(data_reg_stepwise_pred_5b)

#12.3.3 Forward Regression
data_reg_forward_5b <- step(lm5b_train, direction = "forward")
summary(data_reg_forward_5b)
data_reg_forward_pred_5b <- predict(data_reg_forward_5b, test5)
accuracy(data_reg_forward_pred_5b, test5$roa)

#12.3.4 Backward Regression
data_reg_backward_5b <- step(lm5b_train, direction = "backward")
summary(data_reg_backward_5b)
data_reg_backward_pred_5b <- predict(data_reg_backward_5b, test5)
accuracy(data_reg_backward_pred_5b, test5$roa)

#12.3.5 perform regression with fixed effects
data_reg_fixeff_5b <- felm(roa ~ OPTION_AWARDS_BLK_VALUE + at + debt_to_equity_ratio + ceo_years | gvkey + fyear | 0 | gvkey + fyear, train5)
summary(data_reg_fixeff_5b)

#12.3.6 Checking for multicollinearity of controlled variables with VIF
vif_values_5b <- car::vif(lm5b_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#12.3.7 residual analysis
resid_5b <- lm5b_train$residuals
plot(train5$OPTION_AWARDS_BLK_VALUE, resid_5b)
#relationship between OPTION AWARDS BLK VALUE and residuals

fnorm_5b <- fitdist(resid_5b, "norm")
result_5b <- gofstat(fnorm_5b, discrete = FALSE)
result_5b
kscritvalue_5b <- 1.36/sqrt(length(train5$roa))
kscritvalue_5b #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_5b)
plot(fnorm_5b) #residuals are not normally distributed
confint(lm5b_train, level = 0.95)

#12.3.8 Using stargazer for well-formatted regression output
list5b <- list(lm5b_train, data_reg_forward_5b, data_reg_backward_5b, data_reg_stepwise_5b)
stargazer(list5b, type="text",title="Regression Results", digits=4, omit = c("Constant"), no.space = TRUE, out="table5b.txt")

#12.4 Current Ratio Regression Analysis

#12.4.1 regressing OPTION_AWARDS_BLK_VALUE with current ratio; control variables: at, asset turnover and debt to equity ratio
lm5c_train <- lm(current_ratio ~ OPTION_AWARDS_BLK_VALUE + at + debt_to_equity_ratio + asset_turnover, train5)
summary(lm5c_train)
```

```
#12.4.2 Stepwise Regression
data_reg_stepwise_5c <- step(lm5c_train, direction = "both")
summary(data_reg_stepwise_5c)
data_reg_stepwise_pred_5c <-
predict(data_reg_stepwise_5c, test5)
accuracy(data_reg_stepwise_pred_5c, test5$current_ratio)
summary(data_reg_stepwise_pred_5c)

#12.4.3 Forward Regression
data_reg_forward_5c <- step(lm5c_train, direction =
"forward")
summary(data_reg_forward_5c)
data_reg_forward_pred_5c <- predict(data_reg_forward_5c,
test5)
accuracy(data_reg_forward_pred_5c, test5$current_ratio)

#12.4.4 Backward Regression
data_reg_backward_5c <- step(lm5c_train, direction =
"backward")
summary(data_reg_backward_5c)
data_reg_backward_pred_5c <-
predict(data_reg_backward_5c, test5)
accuracy(data_reg_backward_pred_5c, test5$current_ratio)

#12.4.5 perform regression with fixed effects
data_reg_fixeff_5c <- felm(current_ratio ~
OPTION_AWARDS_BLK_VALUE + at + debt_to_equity_ratio +
asset_turnover | gvkey + fyear | 0 | gvkey + fyear, train5)
summary(data_reg_fixeff_5c)

#12.4.6 Checking for multicollinearity of controlled variables
with VIF
vif_values_5c <- car::vif(lm5c_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#12.4.7 residual analysis
resid_5c <- lm5c_train$residuals
plot(train5$OPTION_AWARDS_BLK_VALUE, resid_5c)
#relationship between OPTION AWARDS BLK VALUE and
residuals

fnorm_5c <- fitdist(resid_5c, "norm")
result_5c <- gofstat(fnorm_5c, discrete = FALSE)
result_5c
kscritvalue_5c <- 1.36/sqrt(length(train5$current_ratio))
kscritvalue_5c #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_5c)
plot(fnorm_5c) #residuals are not normally distributed
confint(lm5c_train, level = 0.95)

#12.4.8 Using stargazer for well-formatted regression output
list5c <- list(lm5c_train, data_reg_forward_5c,
data_reg_backward_5c, data_reg_stepwise_5c)
stargazer(list5c, type="text",title="Regression Results",
digits=4, omit = c("Constant"), no.space = TRUE,
out="table5c.txt")

#12.5 Debt to Equity Ratio Regression Analysis

#12.5.1 regressing OPTION_AWARDS_BLK_VALUE with debt
to equity ratio; control variables: at and net profit margin
lm5d_train <- lm(debt_to_equity_ratio ~
OPTION_AWARDS_BLK_VALUE + at + net_profit_margin +
asset_turnover, train5)
summary(lm5d_train)

#12.5.2 Stepwise Regression
data_reg_stepwise_5d <- step(lm5d_train, direction = "both")
summary(data_reg_stepwise_5d)
data_reg_stepwise_pred_5d <-
predict(data_reg_stepwise_5d, test5)
accuracy(data_reg_stepwise_pred_5d,
test5$debt_to_equity_ratio)
summary(data_reg_stepwise_pred_5d)

#12.5.3 Forward Regression
data_reg_forward_5d <- step(lm5d_train, direction =
"forward")
summary(data_reg_forward_5d)
data_reg_forward_pred_5d <- predict(data_reg_forward_5d,
test5)
accuracy(data_reg_forward_pred_5d,
test5$debt_to_equity_ratio)

#12.5.4 Backward Regression
data_reg_backward_5d <- step(lm5d_train, direction =
"backward")
summary(data_reg_backward_5d)
data_reg_backward_pred_5d <-
predict(data_reg_backward_5d, test5)
accuracy(data_reg_backward_pred_5d,
test5$debt_to_equity_ratio)

#12.5.5 perform regression with fixed effects
data_reg_fixeff_5d <- felm(debt_to_equity_ratio ~
OPTION_AWARDS_BLK_VALUE + at + net_profit_margin +
asset_turnover | gvkey + fyear | 0 | gvkey + fyear, train5)
summary(data_reg_fixeff_5d)

#12.5.6 Checking for multicollinearity of controlled variables
with VIF
vif_values_5d <- car::vif(lm5d_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables
```

```
#12.5.7 residual analysis
resid_5d <- lm5d_train$residuals
plot(train5$OPTION_AWARDS_BLK_VALUE, resid_5d)
#relationship between OPTION AWARDS BLK VALUE and
residuals

fnorm_5d <- fitdist(resid_5d, "norm")
result_5d <- gofstat(fnorm_5d, discrete = FALSE)
result_5d
kscritvalue_5d <-
1.36/sqrt(length(train5$debt_to_equity_ratio))
kscritvalue_5d #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_5d)
plot(fnorm_5d) #residuals are not normally distributed
confint(lm5d_train, level = 0.95)

#12.5.8 Using stargazer for well-formatted regression output
list5d <- list(lm5d_train, data_reg_forward_5d,
data_reg_backward_5d, data_reg_stepwise_5d)
stargazer(list5d, type="text",title="Regression Results",
digits=4, omit = c("Constant"), no.space = TRUE,
out="table5d.txt")

#13 multi-linear regression to analyse the effects of long term
incentives pay on profit margin, roa, liquidity ratio, debt to
equity ratio
#Dependent variable: net_profit_margin, roa, current_ratio,
debt_to_equity_ratio
#Independent variable: LTIP_at (long term incentives
pay/total assets)

data_reg_6d_clean <- data_reg_3d_clean %>%
mutate(LTIP_at = ifelse(is.na(LTIP_at),0,LTIP_at))

#13.1. choosing training and test dataset
set.seed(1)
train6 <- sample_frac(data_reg_6d_clean, 0.75)
test6 <- anti_join(data_reg_6d_clean, train6)

#13.2. regressing LTIP_at with profit margin; control
variables: at, debt to equity ratio, ceo_years
lm6a_train <- lm(net_profit_margin ~ LTIP_at + at +
debt_to_equity_ratio + ceo_years, train6)
summary(lm6a_train)

# 13.2.1. perform stepwise regression
data_reg_stepwise_6a <- step(lm6a_train, direction = "both")
summary(data_reg_stepwise_6a)
data_reg_stepwise_pred_6a <-
predict(data_reg_stepwise_6a, test6)
accuracy(data_reg_stepwise_pred_6a,
test6$net_profit_margin)

# 13.2.2. perform forward regression
data_reg_forward_6a <- step(lm6a_train, direction =
"forward")
summary(data_reg_forward_6a)
data_reg_forward_pred_6a <- predict(data_reg_forward_6a,
test6)
accuracy(data_reg_forward_pred_6a,
test6$net_profit_margin)

# 13.2.3. perform backward regression
data_reg_backward_6a <- step(lm6a_train, direction =
"backward")
summary(data_reg_backward_6a)
data_reg_backward_pred_6a <-
predict(data_reg_backward_6a, test6)
accuracy(data_reg_backward_pred_6a,
test6$net_profit_margin)

#13.2.4. perform regression with fixed effects
data_reg_fixeff_6a <- felm(net_profit_margin ~ LTIP_at + at +
debt_to_equity_ratio + ceo_years | gvkey + fyear |0 | gvkey
+ fyear, train6)
summary(data_reg_fixeff_6a)

#13.2.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_6a <- car::vif(lm6a_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#13.2.6 residual analysis
resid_6a <- lm6a_train$residuals
plot(train6$LTIP_at, resid_6a) #relationship between LTIP_at
and residuals

fnorm_6a <- fitdist(resid_6a, "norm")
result_6a <- gofstat(fnorm_6a, discrete = FALSE)
result_6a
kscritvalue_6a <- 1.36/sqrt(length(train6$net_profit_margin))
kscritvalue_6a #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_6a)
plot(fnorm_6a) #residuals are not normally distributed
confint(lm6a_train, level = 0.95)

#13.2.7. Using stargazer for well-formatted regression output
stargazer(lm6a_train, data_reg_forward_6a,
data_reg_backward_6a, data_reg_stepwise_6a,
type="text",title="Regression Results",omit = c("Constant"),
digits=4,  no.space = TRUE, out="table6a.txt")
```

```
#13.3. regressing LTIP_at with return on assets; control
variables: at, debt to equity ratio, ceo_years
lm6b_train <- lm(roa ~ LTIP_at + at + debt_to_equity_ratio +
ceo_years, train6)
summary(lm6b_train)

# 13.3.1. perform stepwise regression
data_reg_stepwise_6b <- step(lm6b_train, direction = "both")
summary(data_reg_stepwise_6b)
data_reg_stepwise_pred_6b <-
predict(data_reg_stepwise_6b, test6)
accuracy(data_reg_stepwise_pred_6b, test6$roa)

# 13.3.2. perform forward regression
data_reg_forward_6b <- step(lm6b_train, direction =
"forward")
summary(data_reg_forward_6b)
data_reg_forward_pred_6b <- predict(data_reg_forward_6b,
test6)
accuracy(data_reg_forward_pred_6b, test6$roa)

# 13.3.3. perform backward regression
data_reg_backward_6b <- step(lm6b_train, direction =
"backward")
summary(data_reg_backward_6b)
data_reg_backward_pred_6b <-
predict(data_reg_backward_6b, test6)
accuracy(data_reg_backward_pred_6b, test6$roa)

#13.3.4. perform regression with fixed effects
data_reg_fixeff_6b <- felm(roa ~ LTIP_at + at +
debt_to_equity_ratio + ceo_years | gvkey + fyear |0 | gvkey
+ fyear, train6)
summary(data_reg_fixeff_6b)

#13.3.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_6b <- car::vif(lm6b_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#13.3.6 residual analysis
resid_6b <- lm6b_train$residuals
plot(train6$LTIP_at, resid_6b) #relationship between LTIP_at
and residuals

fnorm_6b <- fitdist(resid_6b, "norm")
result_6b <- gofstat(fnorm_6b, discrete = FALSE)
result_6b
kscritvalue_6b <- 1.36/sqrt(length(train6$roa))
kscritvalue_6b #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_6b)
plot(fnorm_6b) #residuals are not normally distributed
confint(lm6b_train, level = 0.95)

#13.3.7. Using stargazer for well-formatted regression output
stargazer(lm6b_train, data_reg_forward_6b,
data_reg_backward_6b, data_reg_stepwise_6b,
type="text",title="Regression Results",omit = c("Constant"),
digits=4,  no.space = TRUE, out="table6b.txt")

#13.4. regressing LTIP_at with current ratio; control variables:
at, debt to equity ratio, assets turnover
lm6c_train <- lm(current_ratio ~ LTIP_at + at +
debt_to_equity_ratio + asset_turnover, train6)
summary(lm6c_train)

# 13.4.1. perform stepwise regression
data_reg_stepwise_6c <- step(lm6c_train, direction = "both")
summary(data_reg_stepwise_6c)
data_reg_stepwise_pred_6c <-
predict(data_reg_stepwise_6c, test6)
accuracy(data_reg_stepwise_pred_6c, test6$current_ratio)

# 13.4.2. perform forward regression
data_reg_forward_6c <- step(lm6c_train, direction =
"forward")
summary(data_reg_forward_6c)
data_reg_forward_pred_6c <- predict(data_reg_forward_6c,
test6)
accuracy(data_reg_forward_pred_6c, test6$current_ratio)

# 13.4.3. perform backward regression
data_reg_backward_6c <- step(lm6c_train, direction =
"backward")
summary(data_reg_backward_6c)
data_reg_backward_pred_6c <-
predict(data_reg_backward_6c, test6)
accuracy(data_reg_backward_pred_6c, test6$current_ratio)

#13.4.4. perform regression with fixed effects
data_reg_fixeff_6c <- felm(current_ratio ~ LTIP_at + at +
debt_to_equity_ratio + asset_turnover | gvkey + fyear |0 |
gvkey + fyear, train6)
summary(data_reg_fixeff_6c)

#13.4.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_6c <- car::vif(lm6c_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#13.4.6 residual analysis
resid_6c <- lm6c_train$residuals
```

```
plot(train6$LTIP_at, resid_6c) #relationship between LTIP_at and residuals

fnorm_6c <- fitdist(resid_6c, "norm")
result_6c <- gofstat(fnorm_6c, discrete = FALSE)
result_6c
kscritvalue_6c <- 1.36/sqrt(length(train6$current_ratio))
kscritvalue_6c #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_6c)
plot(fnorm_6c) #residuals are not normally distributed
confint(lm6c_train, level = 0.95)

#13.4.7. Using stargazer for well-formatted regression output
stargazer(lm6c_train, data_reg_forward_6c, data_reg_backward_6c, data_reg_stepwise_6c, type="text",title="Regression Results",omit = c("Constant"), digits=4,  no.space = TRUE, out="table6c.txt")

#13.5. regressing LTIP_at with debt to equity ratio; control variables: at, net profit margin, assets turnover
lm6d_train <- lm(debt_to_equity_ratio ~ LTIP_at + at + net_profit_margin + asset_turnover, train6)
summary(lm6d_train)

# 13.5.1. perform stepwise regression
data_reg_stepwise_6d <- step(lm6d_train, direction = "both")
summary(data_reg_stepwise_6d)
data_reg_stepwise_pred_6d <- predict(data_reg_stepwise_6d, test6)
accuracy(data_reg_stepwise_pred_6d, test6$debt_to_equity_ratio)

# 13.5.2. perform forward regression
data_reg_forward_6d <- step(lm6d_train, direction = "forward")
summary(data_reg_forward_6d)
data_reg_forward_pred_6d <- predict(data_reg_forward_6d, test6)
accuracy(data_reg_forward_pred_6d, test6$debt_to_equity_ratio)

# 13.5.3. perform backward regression
data_reg_backward_6d <- step(lm6d_train, direction = "backward")
summary(data_reg_backward_6d)
data_reg_backward_pred_6d <- predict(data_reg_backward_6d, test6)
accuracy(data_reg_backward_pred_6d, test6$debt_to_equity_ratio)

#13.5.4. perform regression with fixed effects
data_reg_fixeff_6d <- felm(debt_to_equity_ratio ~ LTIP_at + at + net_profit_margin + asset_turnover | gvkey + fyear |0 | gvkey + fyear, train6)
summary(data_reg_fixeff_6d)

#13.5.5. Checking for multicollinearity of controlled variables with VIF
vif_values_6d <- car::vif(lm6d_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#13.5.6 residual analysis
resid_6d <- lm6d_train$residuals
plot(train6$LTIP_at, resid_6d) #relationship between LTIP_at and residuals

fnorm_6d <- fitdist(resid_6d, "norm")
result_6d <- gofstat(fnorm_6d, discrete = FALSE)
result_6d
kscritvalue_6d <- 1.36/sqrt(length(train6$debt_to_equity_ratio))
kscritvalue_6d #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_6d)
plot(fnorm_6d) #residuals are not normally distributed
confint(lm6d_train, level = 0.95)

#13.5.7. Using stargazer for well-formatted regression output
stargazer(lm6d_train, data_reg_forward_6d, data_reg_backward_6d, data_reg_stepwise_6d, type="text",title="Regression Results",omit = c("Constant"), digits=4,  no.space = TRUE, out="table6d.txt")
```