```
# Multivariate Statistics: Exercise 3

source("world.R") ## Dataset must be in the same directory of the R-Script.

summary(world)

lm.full <- lm(PopGrowth ~ GiveBirth + Women + LifeexpF + LifeexpM + InfMort +
              InhabDoc + Calorie + BabyUnderw, data=world)

summary(lm.full)
plot(lm.full)

# The model requirements are fullfilled:
# Sstandardized residuals are normally distributed (see Q-Q plot) -> possible outliers
(uae,ind,dko)
# Cook's distance to data point 'uae' is large. It is a leverage point and might be an
outlier.


world.scaled <- world
world.scaled[,3:11] <- scale(world[,3:11])

lm.full.scaled=lm(PopGrowth ~ GiveBirth + Women + LifeexpF + LifeexpM + InfMort +
              InhabDoc + Calorie + BabyUnderw, data=world.scaled)

summary(lm.full.scaled)

# Scaling of varibles is not needed. Selection of most significant regressors is the same.

# Stepwise elinination of regressors with highest p-value leads to:

lm.v1 <- lm(PopGrowth ~ GiveBirth + Women + InfMort +
              InhabDoc + Calorie, data=world)
summary(lm.v1)
## Comparision of the full model and reduced model
anova(lm.full, lm.v1)
## Stepweise regression
lm.step <- step(lm.full)
summary(lm.step)
# gives the same model as lm.v1

library(leaps)
lm.regsubset <- regsubsets(PopGrowth ~ GiveBirth + Women + LifeexpF + LifeexpM + InfMort +
                             InhabDoc + Calorie + BabyUnderw, data=world, nbest = 2,
nvmax = 8)
summary(lm.regsubset)
plot(lm.regsubset)
```