

```
setwd("~/shared-folder/HPC/APPSTAT/Exams/2023-07-07/E3")
rm(list = ls())

expenses <- read.table("expenditure.txt", header = TRUE)
dim(expenses)
head(expenses)

attach(expenses)

lm <- lm(avg_exp ~ income + age + perc_taxes + owns_house)
summary(lm)

coefficients(lm)
betas <- rbind(beta0 = coefficients(lm)[1], beta1 = coefficients(lm)[2], beta2 =
coefficients(lm)[3], beta3 = coefficients(lm)[4], beta4 = coefficients(lm)[5])
betas

sum(residuals(lm)^2) / lm$df

svg("residuals.overall.svg", width = 10, height = 10)
par(mfrow = c(2, 2))
plot(lm)
dev.off()

library(MVN)
shapiro.test(residuals(lm))

svg("residuals.vs.variables.svg", width = 10, height = 5)
par(mfrow = c(1, 3))
plot(income, residuals(lm), main = "Income vs Residuals")
plot(age, residuals(lm), main = "Age vs Residuals")
plot(perc_taxes, residuals(lm), main = "% Taxes vs Residuals")
dev.off()

age2 <- age^2

lm.2 <- lm(avg_exp ~ income + age + age2 + perc_taxes + owns_house)
summary(lm.2)

svg("residuals.overall.2.svg", width = 10, height = 10)
par(mfrow = c(2, 2))
plot(lm.2)
dev.off()

lm.3 <- lm(avg_exp ~ income + age + age2 + owns_house)
summary(lm.3)

par(mfrow = c(2, 2))
plot(lm.3)

library(glmnet)
set.seed(20230707)

lambda <- 10^seq(10, -2, length.out = 100)

x <- model.matrix(avg_exp ~ income + age + perc_taxes + owns_house, data = expenses)[, -1]
y <- avg_exp

fit.lasso <- glmnet(x, y, lambda = lambda)
summary(fit.lasso)
```