

```

library(haven)
library(foreign)
library(lmtest)
library(Hmisc)
library(plm)
library(dplyr)
library(stargazer)
library(lfe)
library(tidyr)
library(readr)

install.packages("coefplot")
library(coefplot)

#set the path

setwd("C:\\Users\\smjofess\\Documents\\Uni Dokumente\\Labour Economics PS4")

# Clear desk
rm(list=ls())

##### Question 4.1 #####
# write some verbal answers in your solution pdf...

##### Question 4.2 #####

#read in the data
twins_long <- read_dta("twins_long.dta")

#descriptive analysis
attach(twins_long)
summary(twins_long)

# plot for age
table(age)

fit2 <- lm(lwage ~ poly(age, degree = 2, raw = TRUE), data = data.frame(age, lwage))
fit4 <- lm(lwage ~ poly(age, degree = 4, raw = TRUE), data = data.frame(age, lwage))
age_seq <- seq(min(age), max(age), length.out = 100)
lwage_pred2 <- predict(fit2, newdata = data.frame(age = age_seq))
lwage_pred4 <- predict(fit4, newdata = data.frame(age = age_seq))

plot(0, type = "n", xlim = range(age_seq), ylim = range(c(lwage_pred2, lwage_pred4)),
     xlab = "Age", ylab = "log wage", main = "Life-Cycle Profiles with Polynomial
     Fits")
lines(age_seq, lwage_pred2, col = "red", lwd = 2, lty = 1, label = "2nd Order")
lines(age_seq, lwage_pred4, col = "blue", lwd = 2, lty = 2, label = "4th Order")
legend("bottomright", legend = c("2nd Order", "4th Order"), col = c("red", "blue"), lty
      = 2:3, cex = 0.8)

# plot for educ
table(educ)

fit2 <- lm(lwage ~ poly(educ, degree = 2, raw = TRUE), data = data.frame(educ, lwage))
fit4 <- lm(lwage ~ poly(educ, degree = 4, raw = TRUE), data = data.frame(educ, lwage))
educ_seq <- seq(min(educ), max(educ), length.out = 100)
lwage_pred2 <- predict(fit2, newdata = data.frame(educ = educ_seq))
lwage_pred4 <- predict(fit4, newdata = data.frame(educ = educ_seq))

plot(0, type = "n", xlim = range(educ_seq), ylim = range(c(lwage_pred2, lwage_pred4)),
     xlab = "Years of Education", ylab = "log wage", main = "Education--Wage Profiles
     with Polynomial Fits")
lines(educ_seq, lwage_pred2, col = "red", lwd = 2, lty = 1, label = "2nd Order")
lines(educ_seq, lwage_pred4, col = "blue", lwd = 2, lty = 2, label = "4th Order")
legend("bottomright", legend = c("2nd Order", "4th Order"), col = c("red", "blue"), lty
      = 2:3, cex = 0.8)

```

```

# plot for potential experience (usually people force within 0-40 years)
potexp <- pmin(pmax(age - educ - 6,0),40)
table(potexp)

fit2 <- lm(lwage ~ poly(potexp, degree = 2, raw = TRUE), data = data.frame(potexp,
lwage))
fit4 <- lm(lwage ~ poly(potexp, degree = 4, raw = TRUE), data = data.frame(potexp,
lwage))
potexp_seq <- seq(min(potexp), max(potexp), length.out = 100)
lwage_pred2 <- predict(fit2, newdata = data.frame(potexp = potexp_seq))
lwage_pred4 <- predict(fit4, newdata = data.frame(potexp = potexp_seq))

plot(0, type = "n", xlim = range(potexp_seq), ylim = range(c(lwage_pred2,
lwage_pred4)),
      xlab = "Potential Experience", ylab = "log wage", main = "Life-Cycle Profiles with
Polynomial Fits")
lines(potexp_seq, lwage_pred2, col = "red", lwd = 2, lty = 1, label = "2nd Order")
lines(potexp_seq, lwage_pred4, col = "blue", lwd = 2, lty = 2, label = "4th Order")
legend("bottomright", legend = c("2nd Order", "4th Order"), col = c("red", "blue"), lty
= 2:3, cex = 0.8)

```

Question 4.3

```

#creating variable: age squared
twins_long$age_sq <- (twins_long$age*twins_long$age)

#creating a linear model
regr <- lm(lwage ~ educ + age + age_sq + male + white , data = twins_long)

summary(regr)

```

Question 4.4

```

coeftest(regr, vcov = vcovHC, type = "HC1")

coeftest(regr, vcov = vcovHC, cluster = ~ famid)

```

Question 4.5

```

table(educ)

twins_long$educ_factor <- factor(twins_long$educ, levels = c(12, 8, 10, 11, 13, 14, 15,
16, 17, 18, 19, 20)) # 12=High-School will be the omitted category in the regression
table(twins_long$educ,twins_long$educ_factor)

regr2 <- lm(lwage ~ educ_factor + age + age_sq + male + white, data=subset(twins_long))
summary(regr2)

```

```

twins_long$coefficients <- coef(regr2)[twins_long$educ_factor]
twins_long$coefficients[twins_long$educ_factor == 12] <- 0
plot(twins_long$educ, twins_long$coefficients, xlim = range(educ),
xlab = "Years of Education", ylab = "log wage relative to HS=12", main = "Conditional
Education--Wage Profiles
(when Regression-Adjusting)", cex.lab=0.75, cex.main=0.85)
abline(lm(twins_long$coefficients ~ educ),col='red', lwd = 2, lty = 2)

```

Question 4.6

```

regr3 <- lm(lwage ~ educ + factor(famid), data = twins_long)

regr3 <- lm(lwage ~ educ + factor(famid) + factor(twin), data = twins_long)

summary(regr3)

```

Question 4.7

```
#read in the wide data
twins <- read_dta("twins.dta")

#differences in education
twins$dif_edu <- (twins$educ1-twins$educ2)

#differences in wages
twins$dif_lwage <- (twins$lwage1-twins$lwage2)

#linear model for first differences
first_dif_regr <- lm(dif_lwage ~ dif_edu, data = twins)

summary(first_dif_regr)
```