Homework #5

Neshma

11/2/2020

Homework #5 - Neshma Simon

Study Group: Fareha & Hertz

```
{r}
dat_use <- subset(acs2017_ny,use_varb)
use_varb <- (AGE >= 25) & (AGE <= 55) & (LABFORCE == 2) & (WKSWORK2 > 4) & (UHRSWORK >= 35) & (CITIZEN

# We were looking for women, who have at least one college degree and are citizens.
```

```
{r}
dat_use <- subset(acs2017_ny,use_varb)
model_1 <- lm(INCWAGE ~ AGE + I(AGE^2) + I(AGE^3) + I(AGE^4) + I(AGE^5) + I(AGE^6) )
summary(model_1)

- Call:
lm(formula = INCWAGE ~ AGE + I(AGE^2) + I(AGE^3) + I(AGE^4) +
    I(AGE^5) + I(AGE^6))

Residuals:
   Min     1Q Median     3Q    Max
-58984 -27574  -8046   5983 637058

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.320e+05  2.655e+04  -4.970 6.69e-07 ***
AGE          1.531e+04  3.883e+03   3.943 8.04e-05 ***
I(AGE^2)    -7.676e+02  2.204e+02  -3.483 0.000497 ***
I(AGE^3)     2.586e+01  6.259e+00   4.131 3.61e-05 ***
I(AGE^4)    -4.806e-01  9.447e-02  -5.087 3.64e-07 ***
I(AGE^5)     4.280e-03  7.234e-04   5.917 3.29e-09 ***
I(AGE^6)    -1.434e-05  2.208e-06  -6.491 8.53e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 63110 on 163151 degrees of freedom
  (33427 observations deleted due to missingness)
Multiple R-squared:  0.09037,   Adjusted R-squared:  0.09034
F-statistic:  2701 on 6 and 163151 DF,  p-value: < 2.2e-16
```

```
require(stargazer)
stargazer(model_1, type = "text")
```

```
# Through this I'm trying to look at higher polynomials for age up to 6 in order to see if the higher p
#The data shows that though the polynomials for age increase, The plot shows that there is no correlati
```

```
{r}
dat_use <- subset(acs2017_ny, use_varb)
model_2 <- lm(INCWAGE ~ I(AGE^2) + female + CITIZEN)
summary(model_2)
- Call:
lm(formula = INCWAGE ~ I(AGE^2) + female + CITIZEN)

Residuals:
   Min     1Q Median     3Q    Max
-49842 -31795 -17784  11396 624765

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.067e+04  3.294e+02 153.853   <2e-16 ***
I(AGE^2)    -3.243e+00  8.198e-02 -39.562   <2e-16 ***
female      -1.513e+04  3.245e+02 -46.615   <2e-16 ***
CITIZEN     -2.507e+02  1.581e+02  -1.586    0.113
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 65380 on 163154 degrees of freedom
  (33427 observations deleted due to missingness)
Multiple R-squared:  0.02362,	Adjusted R-squared:  0.02361
F-statistic:  1316 on 3 and 163154 DF,  p-value: < 2.2e-16
```

```
{r}
dat_use <- subset(acs2017_ny, use_varb)
model_3 <- lm(INCWAGE ~ I(AGE^3) + female + CITIZEN)
summary(model_3)
-Call:
lm(formula = INCWAGE ~ I(AGE^3) + female + CITIZEN)

Residuals:
   Min     1Q Median     3Q    Max
-50056 -31928 -16699  10481 630971

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.026e+04  2.908e+02 172.841   <2e-16 ***
I(AGE^3)    -5.016e-02  9.380e-04 -53.473   <2e-16 ***
female      -1.484e+04  3.233e+02 -45.904   <2e-16 ***
CITIZEN     -3.681e+02  1.575e+02  -2.336   0.0195 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 65130 on 163154 degrees of freedom
  (33427 observations deleted due to missingness)
Multiple R-squared:  0.03124,	Adjusted R-squared:  0.03122
F-statistic:  1754 on 3 and 163154 DF,  p-value: < 2.2e-16
```

```
{r}
dat_use <- subset(acs2017_ny, use_varb)
```

```
model_4 <- lm(INCWAGE ~ I(AGE^4) + female + CITIZEN)
summary(model_4)
-Call:
lm(formula = INCWAGE ~ I(AGE^4) + female + CITIZEN)

Residuals:
   Min     1Q Median     3Q    Max
-49228 -32103 -16394  11083 643069

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.927e+04  2.737e+02 180.001  < 2e-16 ***
I(AGE^4)    -6.672e-04  1.097e-05 -60.840  < 2e-16 ***
female      -1.468e+04  3.225e+02 -45.505  < 2e-16 ***
CITIZEN     -4.474e+02  1.572e+02  -2.847  0.00442 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 64960 on 163154 degrees of freedom
  (33427 observations deleted due to missingness)
Multiple R-squared:  0.03613,   Adjusted R-squared:  0.03611
F-statistic:  2038 on 3 and 163154 DF,  p-value: < 2.2e-16
```

```
{r}
dat_use <- subset(acs2017_ny, use_varb)
model_5 <- lm(INCWAGE ~ I(AGE^5) + female + CITIZEN)
summary(model_5)
-Call:
lm(formula = INCWAGE ~ I(AGE^5) + female + CITIZEN)

Residuals:
   Min     1Q Median     3Q    Max
-48129 -31866 -16770  11852 652397

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.814e+04  2.646e+02 181.919  < 2e-16 ***
I(AGE^5)    -8.082e-06  1.274e-07 -63.412  < 2e-16 ***
female      -1.461e+04  3.222e+02 -45.350  < 2e-16 ***
CITIZEN     -4.795e+02  1.570e+02  -3.054  0.00226 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 64900 on 163154 degrees of freedom
  (33427 observations deleted due to missingness)
Multiple R-squared:  0.03797,   Adjusted R-squared:  0.03795
F-statistic:  2146 on 3 and 163154 DF,  p-value: < 2.2e-16
```

```
{r}
dat_use <- subset(acs2017_ny, use_varb)
model_6 <- lm(INCWAGE ~ I(AGE^6) + female + CITIZEN)
summary(model_6)
-Call:
lm(formula = INCWAGE ~ I(AGE^6) + female + CITIZEN)
```

```
Residuals:
   Min     1Q Median     3Q    Max
-47075 -31453 -17114  12389 658652

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.708e+04  2.593e+02 181.537   <2e-16 ***
I(AGE^6)    -9.213e-08  1.464e-09 -62.937   <2e-16 ***
female      -1.462e+04  3.223e+02 -45.360   <2e-16 ***
CITIZEN     -4.788e+02  1.570e+02  -3.049   0.0023 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 64910 on 163154 degrees of freedom
  (33427 observations deleted due to missingness)
Multiple R-squared:  0.03762,	Adjusted R-squared:  0.0376
F-statistic:  2126 on 3 and 163154 DF,  p-value: < 2.2e-16
```

```
stargazer(model_1, model_2, model_3, model_4, model_5, model_6, type= "text")
=======================================================================================================
                                                                                                    Depe
                    -----------------------------------------------------------------------------------

                                 (1)                            (2)                            (3)
-------------------------------------------------------------------------------------------------------
AGE                         15,311.430***
                             (3,882.985)

I(AGE2)                      -767.580***                     -3.243***
                              (220.401)                       (0.082)

I(AGE3)                       25.856***                                                     -0.050***
                               (6.259)                                                       (0.001)

I(AGE4)                       -0.481***
                               (0.094)

I(AGE5)                        0.004***
                               (0.001)

I(AGE6)                      -0.00001***
                              (0.00000)

female                                                    -15,127.730***                 -14,841.340***
                                                            (324.525)                      (323.309)

CITIZEN                                                      -250.704                      -368.051**
                                                            (158.118)                      (157.528)

Constant                   -131,968.200***                50,671.860***                  50,261.850***
                            (26,551.430)                    (329.352)                      (290.798)

-------------------------------------------------------------------------------------------------------
Observations                   163,158                        163,158                        163,158
```

```
R2                              0.090                       0.024                       0.031
Adjusted R2                     0.090                       0.024                       0.031
Residual Std. Error   63,110.280 (df = 163151)      65,384.140 (df = 163154)      65,128.780 (df = 16315
F Statistic         2,701.493*** (df = 6; 163151) 1,315.896*** (df = 3; 163154) 1,753.535*** (df = 3; 1
=======================================================================================================
```

```
dat_use <- subset(acs2017_ny, use_varb)
model_2 <- lm(INCWAGE ~ I(AGE^2) + female + CITIZEN)
summary(model_2)
- Call:
lm(formula = INCWAGE ~ I(AGE^2) + female + CITIZEN)

Residuals:
   Min     1Q Median     3Q    Max
-49842 -31795 -17784  11396 624765

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.067e+04  3.294e+02 153.853   <2e-16 ***
I(AGE^2)    -3.243e+00  8.198e-02 -39.562   <2e-16 ***
female      -1.513e+04  3.245e+02 -46.615   <2e-16 ***
CITIZEN     -2.507e+02  1.581e+02  -1.586    0.113
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 65380 on 163154 degrees of freedom
  (33427 observations deleted due to missingness)
Multiple R-squared:  0.02362,   Adjusted R-squared:  0.02361
F-statistic:  1316 on 3 and 163154 DF,  p-value: < 2.2e-16
```

```
NNobs <- length(INCWAGE)
set.seed(12345)
graph_obs <- (runif(NNobs) < 0.1)
dat_graph <-subset(dat_use,graph_obs)

plot(INCWAGE ~ jitter(AGE, factor = 2), pch = 16, col = rgb(0.5, 0.5, 0.5, alpha = 0.2), ylim = c(0,150
to_be_predicted1 <- data.frame(AGE = 20:65, female = 1, educ_college = 1, educ_advdeg = 1)
to_be_predicted1$yhat <- predict(model_1, newdata = to_be_predicted1)
lines(yhat ~ AGE, data = to_be_predicted1)
-See Plot #1
```

```
dat_use <- subset(acs2017_ny, use_varb)
model_2 <- lm(INCWAGE ~ I(AGE^3) + female + CITIZEN)
summary(model_2)

NNobs <- length(INCWAGE)
set.seed(12345)
graph_obs <- (runif(NNobs) < 0.1)
dat_graph <-subset(dat_use,graph_obs)

plot(INCWAGE ~ jitter(AGE, factor = 3), pch = 16, col = rgb(0.5, 0.5, 0.5, alpha = 0.2), ylim = c(0,150
to_be_predicted2 <- data.frame(AGE = 20:65, female = 1, educ_college = 1, educ_advdeg = 1, CITIZEN = 1)
to_be_predicted2$yhat <- predict(model_2, newdata = to_be_predicted2)
lines(yhat ~ AGE, data = to_be_predicted2)
-See Plot #3
```

```
# This data shows that there's a negative correlation between age and icnome, however, in this we're loo

dat_use <- subset(acs2017_ny, use_varb)
model_3 <- lm(INCWAGE ~ I(AGE^3) + female + CITIZEN)
summary(model_3)

NNobs <- length(INCWAGE)
set.seed(12345)
graph_obs <- (runif(NNobs) < 0.1)
dat_graph <-subset(dat_use,graph_obs)

plot(INCWAGE ~ jitter(AGE, factor = 3), pch = 16, col = rgb(0.5, 0.5, 0.5, alpha = 0.2), ylim = c(0,1500
to_be_predicted2 <- data.frame(AGE = 20:65, female = 1, educ_college = 1, educ_advdeg = 1, CITIZEN = 1)
to_be_predicted2$yhat <- predict(model_3, newdata = to_be_predicted2)
lines(yhat ~ AGE, data = to_be_predicted2)
- See Plot #3
# I repeated the code again but with the different polynomial for age to see if the polynomial makes a
```