Homework6

Fareha Sameen

11/9/2020

Group: Neshma and Hertz

```
attach(acs2017_ny)
model_v1 <- lm(INCWAGE ~ AGE)


model_v2 <- lm(acs2017_ny$INCWAGE ~ acs2017_ny$AGE)

model_v3 <- lm(INCWAGE ~ AGE, data = acs2017_ny)
model_logit1 <- glm(LABFORCE ~ AGE,family = binomial, data = acs2017_ny)
```

```
# In this logit model, we check to see the status of women with ages between 25 and 55 who are in the la
```

```
acs2017_ny$LABFORCE <- as.factor(acs2017_ny$LABFORCE)
levels(acs2017_ny$LABFORCE) <- c("NA","Not in LF","in LF")
```

```
acs2017_ny$age_bands <- cut(acs2017_ny$AGE,breaks=c(0,25,35,45,55,65,100))
table(acs2017_ny$age_bands,acs2017_ny$LABFORCE)

pick_use1 <- (acs2017_ny$AGE >25) & (acs2017_ny$AGE <= 55)
dat_use1 <- subset(acs2017_ny, pick_use1)
```

```
             NA Not in LF in LF
  (0,25]   31680     11717 13256
  (25,35]      0      4271 20523
  (35,45]      0      4064 18924
  (45,55]      0      5406 21747
  (55,65]      0     10563 18106
  (65,100]     0     28701  5880
```

```
dat_use1$LABFORCE <- droplevels(dat_use1$LABFORCE)
# The product of this shows how may women in different age groups are in the labor force and how many a
```

```
model_logit1 <- glm(LABFORCE ~ AGE + I(AGE^2) + female + AfAm + Asian + race_oth + Hispanic
            + educ_hs + educ_somecoll + educ_college + educ_advdeg
            + MARST,
            family = binomial, data = dat_use1)
summary(model_logit1)
```

```
Call:
glm(formula = LABFORCE ~ AGE + I(AGE^2) + female + AfAm + Asian +
    race_oth + Hispanic + educ_hs + educ_somecoll + educ_college +
    educ_advdeg + MARST, family = binomial, data = dat_use1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
```

```
-2.5832   0.3544   0.4898   0.6531   1.4508

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.4685228  0.2448988   1.913 0.055732 .
AGE            0.0234645  0.0120812   1.942 0.052109 .
I(AGE^2)      -0.0003617  0.0001469  -2.463 0.013782 *
female        -0.6718196  0.0204038 -32.926  < 2e-16 ***
AfAm          -0.2242340  0.0279895  -8.011 1.13e-15 ***
Asian         -0.1303415  0.0373695  -3.488 0.000487 ***
race_oth      -0.0836070  0.0331696  -2.521 0.011716 *
Hispanic       0.1499195  0.0312545   4.797 1.61e-06 ***
educ_hs        0.9072897  0.0309561  29.309  < 2e-16 ***
educ_somecoll  1.4703761  0.0349971  42.014  < 2e-16 ***
educ_college   1.9526149  0.0370063  52.764  < 2e-16 ***
educ_advdeg    2.3771878  0.0436527  54.457  < 2e-16 ***
MARST         -0.0653027  0.0046961 -13.906  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 71408  on 74934  degrees of freedom
Residual deviance: 64991  on 74922  degrees of freedom
AIC: 65017

Number of Fisher Scoring iterations: 5

# In this we're using education and marital status as variables to see how education and marital status

# We used this to see the change in results without variables such as education, race, and marital statu
model_logit2 <- glm(LABFORCE ~ AGE, family = binomial, data = dat_use1)
summary(model_logit2)
Call:
glm(formula = LABFORCE ~ AGE, family = binomial, data = dat_use1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9021   0.6032   0.6259   0.6520   0.6735

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.864808   0.044962  41.475   <2e-16 ***
AGE         -0.009033   0.001064  -8.489   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 71408  on 74934  degrees of freedom
Residual deviance: 71336  on 74933  degrees of freedom
AIC: 71340
```

```
Number of Fisher Scoring iterations: 4
```

```
# The follwing show model_logit3,4,5 through which we tried seeing the impact of different variables on
model_logit3 <- glm(LABFORCE ~ AGE + I(AGE^2) + female + AfAm + Asian + race_oth + Hispanic + educ_hs +
summary(model_logit3)
```

```
Call:
glm(formula = LABFORCE ~ AGE + I(AGE^2) + female + AfAm + Asian +
    race_oth + Hispanic + educ_hs + educ_somecoll + educ_college +
    educ_advdeg + FAMSIZE, family = binomial, data = dat_use1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6280   0.3457   0.4955   0.6553   1.4369

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.2451623  0.2377430  -1.031 0.302444
AGE            0.0325788  0.0120427   2.705 0.006825 **
I(AGE^2)      -0.0004113  0.0001470  -2.798 0.005146 **
female        -0.6745928  0.0204295 -33.021  < 2e-16 ***
AfAm          -0.2863821  0.0275035 -10.413  < 2e-16 ***
Asian         -0.1305968  0.0374776  -3.485 0.000493 ***
race_oth      -0.0957544  0.0332063  -2.884 0.003931 **
Hispanic       0.1218593  0.0312831   3.895  9.8e-05 ***
educ_hs        0.9266151  0.0309441  29.945  < 2e-16 ***
educ_somecoll  1.5032480  0.0349743  42.982  < 2e-16 ***
educ_college   2.0028058  0.0370521  54.054  < 2e-16 ***
educ_advdeg    2.4468515  0.0436390  56.070  < 2e-16 ***
FAMSIZE        0.0664596  0.0058801  11.302  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 71408  on 74934  degrees of freedom
Residual deviance: 65053  on 74922  degrees of freedom
AIC: 65079

Number of Fisher Scoring iterations: 5
```

```
model_logit4 <- glm(LABFORCE ~ AGE + I(AGE^2) + female + AfAm + Asian + race_oth + Hispanic + educ_hs +
summary(model_logit4)
```

```
Call:
glm(formula = LABFORCE ~ AGE + I(AGE^2) + female + AfAm + Asian +
    race_oth + Hispanic + educ_hs + educ_somecoll + educ_college +
    educ_advdeg + MARST + FAMSIZE, family = binomial, data = dat_use1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6024   0.3462   0.4904   0.6533   1.4693

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.3905680  0.2451999   1.593  0.11119
AGE            0.0173019  0.0121191   1.428  0.15339
```

```
I(AGE^2)      -0.0002714  0.0001475  -1.840  0.06580 .
female        -0.6803756  0.0204615 -33.252  < 2e-16 ***
AfAm          -0.2308655  0.0280294  -8.237  < 2e-16 ***
Asian         -0.1493175  0.0374716  -3.985 6.75e-05 ***
race_oth      -0.0881713  0.0331875  -2.657  0.00789 **
Hispanic       0.1362644  0.0313217   4.350 1.36e-05 ***
educ_hs        0.9118417  0.0309869  29.427  < 2e-16 ***
educ_somecoll  1.4796793  0.0350513  42.215  < 2e-16 ***
educ_college   1.9729277  0.0371666  53.083  < 2e-16 ***
educ_advdeg    2.4022062  0.0438382  54.797  < 2e-16 ***
MARST         -0.0527822  0.0050385 -10.476  < 2e-16 ***
FAMSIZE        0.0425891  0.0062339   6.832 8.38e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 71408  on 74934  degrees of freedom
Residual deviance: 64943  on 74921  degrees of freedom
AIC: 64971

Number of Fisher Scoring iterations: 5
```

```
model_logit5 <- glm(LABFORCE ~ AGE + I(AGE^2) + female + AfAm + Asian + race_oth + Hispanic + educ_hs +
summary(model_logit5)
detach(acs2017_ny)
```

```
Call:
glm(formula = LABFORCE ~ AGE + I(AGE^2) + female + AfAm + Asian +
    race_oth + Hispanic + educ_hs + educ_somecoll + educ_college +
    educ_advdeg + MARST + FAMSIZE + RELATE + RELATED, family = binomial,
    data = dat_use1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6565   0.3228   0.4711   0.6318   1.7573

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.0322603  0.2498763   4.131 3.61e-05 ***
AGE            0.0130771  0.0123295   1.061  0.28886
I(AGE^2)      -0.0002720  0.0001500  -1.814  0.06974 .
female        -0.7592838  0.0210160 -36.129  < 2e-16 ***
AfAm          -0.1795140  0.0286135  -6.274 3.52e-10 ***
Asian         -0.1025814  0.0379683  -2.702  0.00690 **
race_oth      -0.0921546  0.0335918  -2.743  0.00608 **
Hispanic       0.1667801  0.0317853   5.247 1.55e-07 ***
educ_hs        0.8232360  0.0317256  25.949  < 2e-16 ***
educ_somecoll  1.3072221  0.0359129  36.400  < 2e-16 ***
educ_college   1.7886759  0.0380032  47.066  < 2e-16 ***
educ_advdeg    2.1959502  0.0446189  49.216  < 2e-16 ***
MARST          0.0098808  0.0054456   1.814  0.06961 .
FAMSIZE        0.0067252  0.0064006   1.051  0.29339
RELATE        -0.0846649  0.0882969  -0.959  0.33763
RELATED       -0.0002771  0.0008608  -0.322  0.74754
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 71408  on 74934  degrees of freedom
Residual deviance: 63347  on 74919  degrees of freedom
AIC: 63379

Number of Fisher Scoring iterations: 5
```

#Comparing the outputs, we can tell that the coefficient estimates are different for each variable tested, the standard error, the p and z vlaues also change. It is evident through the results that women with higher education degrees are more likely to be in the labor force. Furthermore, each variable has an effect on the labor force.