Stat632 HW6

Brandon Keck

```r
library(ggplot2)
library(readr)
county_votes16 <- read_csv("county_votes16.csv")
```

Exercise 1.

(a)

```r
glm1 <- glm(trump_win ~ obama_pctvotes, data = county_votes16, family = binomial)
summary(glm1)
```

```
Call:
glm(formula = trump_win ~ obama_pctvotes, family = binomial,
    data = county_votes16)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.3777   0.0025   0.0206   0.1159   2.4832

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)    20.12971    1.04450   19.27   <2e-16 ***
obama_pctvotes -0.37149    0.01971  -18.85   <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: 2703.37  on 3111  degrees of freedom
Residual deviance:  736.42  on 3110  degrees of freedom
AIC: 740.42

Number of Fisher Scoring iterations: 8
```
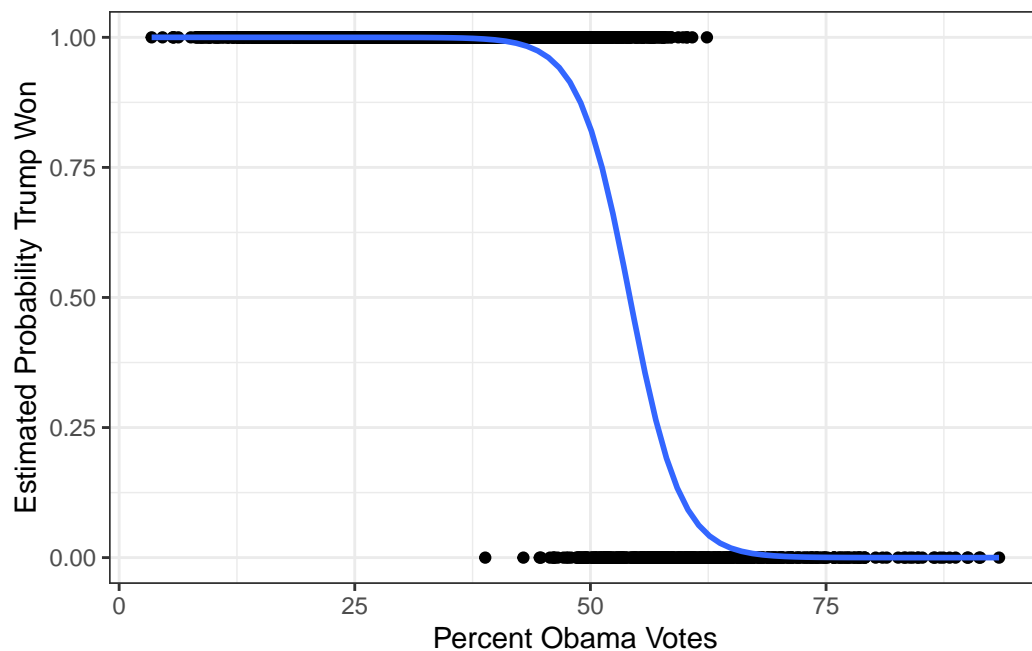
Estimated Logistic Regression Equation:

$$\widehat{p}(x) = \frac{e^{20.12971 - 0.37149x}}{1 + e^{20.12971 - 0.37149x}}$$

(b)

```
ggplot(county_votes16, aes(obama_pctvotes, trump_win)) +
  geom_point() +
  geom_smooth(method = "glm", method.args = list(family = "binomial"), se = F) +
  xlab("Percent Obama Votes") +
  ylab("Estimated Probability Trump Won") +
  theme_bw()
```

```
`geom_smooth()` using formula = 'y ~ x'
```

(c)

```r
new_x <- data.frame(obama_pctvotes = c(40, 50, 60))

predict(glm1, newdata = new_x, type = "response")
```

```
        1         2         3
0.9948835 0.8256735 0.1034357
```

40%: When Obama has 40% of the votes in a specific county, the estimated probability that Trump will win that county is 99.50%

50%: When Obama has 50% of the votes in a specific county, the estimated probability that Trump will win is 82.57%.

60%: When Obama has 60% of the votes in a specific county, the estimated probability that Trump will win that county is 10.35%.

(d)

```r
exp(-0.37149)
```

```
[1] 0.6897059
```

For each 1 percentage point increase in Obama's 2012 vote share in a county, the log-odds of Trump winning the county in 2016 decreases by approximately 0.6897059.

Exercise 2

(a)

```r
glm2 <- glm(trump_win ~ pct_pop65 + pct_black + pct_white + pct_hispanic + pct_asian + hig
summary(glm2)
```

```
Call:
glm(formula = trump_win ~ pct_pop65 + pct_black + pct_white +
    pct_hispanic + pct_asian + highschool + bachelors + income,
    family = binomial, data = county_votes16)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.2155   0.0648   0.1350   0.3170   2.9283

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.760459   1.721412   1.604 0.108802
pct_pop65    -0.020445   0.017910  -1.142 0.253632
pct_black    -0.035455   0.007739  -4.581 4.63e-06 ***
pct_white     0.084759   0.007873  10.765  < 2e-16 ***
pct_hispanic -0.083716   0.007005 -11.952  < 2e-16 ***
pct_asian    -0.160999   0.046158  -3.488 0.000487 ***
highschool   -0.042242   0.020994  -2.012 0.044204 *
bachelors    -0.193758   0.014444 -13.415  < 2e-16 ***
income        0.048985   0.008503   5.761 8.39e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2703.4  on 3111  degrees of freedom
Residual deviance: 1269.4  on 3103  degrees of freedom
AIC: 1287.4

Number of Fisher Scoring iterations: 7
```

(b)

```
glm3 <- glm(trump_win ~ pct_black + pct_white + pct_hispanic + pct_asian + highschool + ba
summary(glm3)
```

```
Call:
glm(formula = trump_win ~ pct_black + pct_white + pct_hispanic +
    pct_asian + highschool + bachelors + income, family = binomial,
```

```
    data = county_votes16)

Deviance Residuals:
    Min      1Q   Median       3Q      Max
-3.2004   0.0653   0.1351   0.3205   2.9272

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.644244   1.716668   1.540 0.123479
pct_black    -0.036567   0.007694  -4.753 2.01e-06 ***
pct_white     0.081996   0.007479  10.963  < 2e-16 ***
pct_hispanic -0.082609   0.006918 -11.942  < 2e-16 ***
pct_asian    -0.152133   0.044877  -3.390 0.000699 ***
highschool   -0.043707   0.020911  -2.090 0.036606 *
bachelors    -0.192417   0.014389 -13.373  < 2e-16 ***
income        0.050576   0.008376   6.039 1.56e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2703.4  on 3111  degrees of freedom
Residual deviance: 1270.7  on 3104  degrees of freedom
AIC: 1286.7

Number of Fisher Scoring iterations: 7
```

(c)

- pct_black: (-0.0366) Counties with a higher percentage of Black residents are less likely to vote for Trump.

- pct_white: (0.0819) Counties with a higher percentage of White residents are more likely to vote for Trump

- pct_hispanic: (-0.0826) Higher Hispanic population is associated with lower likelihood of Trump winning

- pct_asian: (-0.1521) Higher Asian population also predicts lower odds of Trump winning

- highschool: (-0.0437) A higher percentage of residents with at least a high school diploma is slightly associated with decreased odds of Trump winning.

- income: (0.0506) Higher per capita income is associated with a higher chance of Trump winning