Homework 3

Zach White

1. Agresti 5.17 (Death Penalty in North Carolina)

Part A Solution

The case most likely to receive the death penalty is one with a white defendent, white victim, and two different additional factors, which has the odds ratio of compared to the baseline. $e^{\hat{\alpha}+\hat{\beta}_2^D+\hat{\beta}_2^V+\hat{\beta}_3^F} = e^{-5.26+0.17+0.91+3.98}$. The resulting probability is the following: $\frac{e^{-5.26+0.17+0.913.98}}{1+e^{-5.26+0.17+0.91+3.98}} = .45$

Part B Solution

$$
\begin{array}{llll}
\hat{\alpha} = -0.2 & \hat{\beta}_1^D = -0.17 & \hat{\beta}_1^V = -0.91 & \hat{\beta}_1^F = -3.98 \\
 & \hat{\beta}_2^D = 0 & \hat{\beta}_2^V = 0 & \hat{\beta}_2^F = -2.02 \\
 & & & \hat{\beta}_3^F = 0
\end{array}
$$

Part C Solution

$$
\begin{array}{llll}
\hat{\alpha} = -2.72 & \hat{\beta}_1^D = -0.085 & \hat{\beta}_1^V = -0.455 & \hat{\beta}_1^F = -2 \\
 & \hat{\beta}_2^D = 0.085 & \hat{\beta}_2^V = 0.455 & \hat{\beta}_2^F = 0.02 \\
 & & & \hat{\beta}_3^F = 1.98
\end{array}
$$

2. Agresti 5.26 (Using OR to Approximate Relative Risk)

According to the model 5.1 that is desired, $\pi(x) = \frac{\exp\{\alpha+\beta x\}}{1+\exp\{\alpha+\beta x\}}$. If $\pi(x)$ is small, then that would mean that comparitatively, $\exp\{\alpha+\beta x\}$ is much smaller than $1+\exp\{\alpha+\beta x\}$, which also implies that $1 > \exp\{\alpha+\beta x\}$

$$
\begin{aligned}
\frac{\pi(x+1)}{\pi(x)} &= \frac{\exp\{\alpha+\beta(x+1)\}}{1+\exp\{\alpha+\beta(x+1)\}} \frac{1+\exp\{\alpha+\beta x\}}{\exp\{\alpha+\beta x\}} \\
&= \frac{\exp\{\alpha\}\exp\{\beta x\}\exp\{\beta\}}{\exp\{\alpha\}\exp\{\beta x\}} \frac{1+\exp\{\alpha+\beta x\}}{1+\exp\{\alpha+\beta(x+1)\}} \\
&= \exp\{\beta\} \frac{1+\exp\{\alpha+\beta x\}}{1+\exp\{\alpha+\beta(x+1)\}} \\
&\approx \exp\{\beta\} \text{ since } \frac{1+\exp\{\alpha+\beta x\}}{1+\exp\{\alpha+\beta(x+1)\}} \to 1
\end{aligned}
$$

3. Agresti 6.6 (Missing People)

Solution

I propose the following model to fit this data:

$$\log(\frac{\pi_i}{1-\pi_i}) = \beta_0 + \beta_1 x_i + \beta_2 \text{I}(\text{age} = 14\text{-}18) + \beta_3 \text{I}(\text{age} > 19) \tag{1}$$

$$x_{1i} = \begin{cases} 0 \text{ if male} \\ 1 \text{ if female} \end{cases} \tag{2}$$

where x_{1i}

```
still.miss = c(33,38,63,108,157,159)
total = c(3271,2486,7256,8877,5065,3520)
not.miss = total - still.miss
missing = c(rep(1,6),rep(0,6))
Freq = c(still.miss,not.miss)
female = c(0,1,0,1,0,1,0,1,0,1,0,1)
age1418 = c(0,0,1,1,0,0,0,0,1,1,0,0)
age19up = c(0,0,0,0,1,1,0,0,0,0,1,1)

missing.df = data.frame(missing,Freq,female,age1418,age19up)
missing.ind = expand.dft(missing.df)

missing.model = glm(missing ~ female + age1418 + age19up, data = missing.ind, family = binomial(link =
summary(missing.model)
```

```
##
## Call:
## glm(formula = missing ~ female + age1418 + age19up, family = binomial(link = "logit"),
##     data = missing.ind)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.3032  -0.2516  -0.1576  -0.1304   3.0891
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.56479    0.12783 -35.710  < 2e-16 ***
## female       0.38028    0.08689   4.377 1.21e-05 ***
## age1418     -0.19797    0.14241  -1.390    0.164
## age19up      1.12790    0.13252   8.511  < 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: 5570.1  on 30474  degrees of freedom
## Residual deviance: 5348.4  on 30471  degrees of freedom
## AIC: 5356.4
##
## Number of Fisher Scoring iterations: 7
```

The interpretation of of the coefficients are as follows: e^{β_0} represents the odds of a child still being missing after a year who is a male and less than 13.

e^{β_1} represents the odds-ratio of still being missing after a year between male and female.

e^{β_2} represents he odds-ratio of still being missing after a year between a child less than 13 years old and between 14-18.

e^{β_3} represents he odds-ratio of still being missing after a year between a child less than 13 years old and older than 19.

It seems that the effect of gender is significant and also there is a difference between between the age of less than 13 and older than 19. We can also include interactions because those might be of interest in our case.

```
missing.model.int = glm(missing ~ female + age1418 + age19up + female:age1418 + female:age19up, data = m
summary(missing.model.int)
```

```
##
## Call:
## glm(formula = missing ~ female + age1418 + age19up + female:age1418 +
##     female:age19up, family = binomial(link = "logit"), data = missing.ind)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.3040  -0.2510  -0.1565  -0.1321   3.0810
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -4.58620    0.17496 -26.213   <2e-16 ***
## female          0.42076    0.23945   1.757   0.0789 .
## age1418        -0.15153    0.21592  -0.702   0.4828
## age19up         1.14383    0.19283   5.932    3e-09 ***
## female:age1418 -0.07988    0.28761  -0.278   0.7812
## female:age19up -0.02948    0.26551  -0.111   0.9116
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5570.1  on 30474  degrees of freedom
## Residual deviance: 5348.3  on 30469  degrees of freedom
## AIC: 5360.3
##
## Number of Fisher Scoring iterations: 7
```

4. Agresti 6.32 (Residuals for Binary Data)

For ungrouped binary data, explain why when $\hat{\pi}_i$ is near 1, residuals are necessarily either small and positive or large and negative. What happens when $\hat{\pi}_i$ is near 0?

Solution

Under logistic regression and binary data, there are obviosly only two responses: success(1) or failure (0). Thus, when $\hat{\pi}_i$ is near 1 and the observed value is a success, the residual will of course be small and positive. On the other side of the spectrum, if $\hat{\pi}_i$ is near one and the observed value is zero, then the residual would be large and negative.

By similar reasoning as above, when $\hat{\pi}_i$ is near 0, there will be small and negative when a failure is observed or large and positive residuals when a success is observed.