## Math-661: Assignment 4 – Sample Solution

1. **Exercise 1 – Agresti # 6.20**

   The following R output shows output from fitting a cumulative logit model to data from the US 2008 General Social Survey. For subject $i$, let

   - $y_i$ = belief in existence of heaven (1 = yes, 2 = unsure, 3 = no),
   - $x_{i1}$ = gender (1 = female, 0 = male) and
   - $x_{i2}$ = race (1 = black, 0 = white).

```
> cbind(race, gender, y1, y2, y3)
     race gender  y1  y2 y3
[1,]    1      1  88  16  2
[2,]    1      0  54   7  5
[3,]    0      1 397 141 24
[4,]    0      0 235 189 39

> summary(vglm(cbind(y1,y2,y3) ~ gender+race, family=cumulative(parallel=T)))
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1  0.07631    0.08963   0.851    0.395
(Intercept):2  2.32238    0.13522  17.175  < 2e-16 ***
gender         0.76956    0.12253   6.281 3.37e-10 ***
race           1.01645    0.21059   4.827 1.39e-06 ***
---
Residual deviance: 9.2542 on 4 degrees of freedom
Log-likelihood: -23.3814 on 4 degrees of freedom
```

   (a) **State the model fitted here and interpret the race and gender effects.**

   A proportional odds (cumulative logit) model is fit and the estimated equations for the two cumulative logits are:

$$\log \frac{\hat{P}(y_i \le 1)}{1-\hat{P}(y_i \le 1)} = 0.076 + 0.770\ Female + 1.016\ Black$$

$$\log \frac{\hat{P}(y_i \le 2)}{1-\hat{P}(y_i \le 2)} = 2.322 + 0.770\ Female + 1.016\ Black$$

   Controlling for race, the estimated log-odds of a response in the "yes" direction rather than the "no" direction for a female is 0.770 higher than a male (i.e., the estimated odds ratio for a female versus a male is $\exp(0.770) = 2.159$).
   Controlling for gender, the estimated odds for a black person to respond in the "yes" direction rather than the "no" direction is $\exp(1.016) = 2.763$ times higher than for a white person.

   (b) **Test goodness-of-fit and construct confidence intervals for the effects.**

The goodness-of-fit test

$$H_0: \text{ model fits the data well} \qquad H_1: \text{ model does not fit the data well}$$

compares the current model to the saturated model, which has $df = 8$ (4 parameters for each of the baseline logit equation). The current model has $df = 4$, thus the goodness-of-fit $df = 8 - 4 = 4$. The residual deviance is 9.2542, which leads to a $p$-value $= 0.055$. The model does not appear to fit the data well.

```
> 1-pchisq(9.2542, 4)
[1] 0.05505042
```

Although the model does not fit the data well, let's still go ahead and construct 95% Wald confidence intervals for the effects:

- Given race, the odds of responding in the "yes" direction for a female are between 1.70 and 2.75 times the odds for a male.

$$\exp\left[0.770 \pm 1.96(0.123)\right] = (1.698, 2.745)$$

- Controlling for gender, the odds of responding in the "yes" direction for a black person are between 1.83 and 4.18 times the odds for a white person.

$$\exp\left[1.016 \pm 1.96(0.211)\right] = (1.829, 4.175)$$

2. **Exercise 2 – Agresti # 6.21**

Refer to the previous exercise. Consider the model

$$\log \frac{\pi_{ij}}{\pi_{i3}} = \alpha_j + \beta_j^G x_{i1} + \beta_j^R x_{i2}, \qquad j = 1, 2.$$

(a) **Fit the model and report prediction equations for**

$$\log \frac{\pi_{i1}}{\pi_{i3}}, \ \log \frac{\pi_{i2}}{\pi_{i3}}, \ \log \frac{\pi_{i1}}{\pi_{i2}}.$$

Let us first create the data:

```
> race=c(rep(1,2), rep(0,2))
> gender=c(rep(1:0,2))
> y1 = c(88, 54, 397, 235)
> y2 = c(16, 7, 141, 189)
> y3 = c(2, 5, 24, 39)
> dat = data.frame(cbind(race, gender, y1, y2, y3))
> dat
  race gender  y1  y2 y3
1    1      1  88  16  2
2    1      0  54   7  5
3    0      1 397 141 24
4    0      0 235 189 39
```

Let's fit the baseline logit (multinomial logit) model:

```
> summary(vglm(cbind(y1,y2,y3) ~ gender+race, family=multinomial))

               Estimate Std. Error z value Pr(>|z|)
(Intercept):1    1.7943     0.1675  10.712  < 2e-16 ***
(Intercept):2    1.5309     0.1717   8.918  < 2e-16 ***
gender:1         1.0339     0.2587   3.997 6.41e-05 ***
gender:2         0.3087     0.2697   1.145    0.252
race:1           0.6727     0.4114   1.635    0.102
race:2          -0.4757     0.4533  -1.049    0.294
---
Residual deviance: 6.0748 on 2 degrees of freedom
Log-likelihood: -21.7917 on 2 degrees of freedom
```

The prediction equations are:

$$\begin{aligned}
\log \frac{\hat{\pi}_{i1}}{\hat{\pi}_{i3}} &= 1.794 + 1.034\ Female + 0.673\ Black \\
\log \frac{\hat{\pi}_{i2}}{\hat{\pi}_{i3}} &= 1.531 + 0.309\ Female - 0.476\ Black \\
\log \frac{\pi_{i1}}{\pi_{i2}} &= \log \frac{\hat{\pi}_{i1}}{\hat{\pi}_{i3}} - \log \frac{\hat{\pi}_{i2}}{\hat{\pi}_{i3}} = 0.263 + 0.725\ Female + 1.149\ Black
\end{aligned}$$

(b) **Using the "yes" and "no" response categories, interpret the conditional gender effect using a 95% confidence interval for the odds ratio.**

A 95% confidence interval for the odds ratio of a "yes" versus a "no" response in females versus males is

$$\exp(1.034 \pm 1.96 \times 0.259) = (1.694, 4.669)$$

Controlling for race, the odds of a "yes" rather than a "no" response for a female is between 1.7 and 4.7 times the odds for a male.

(c) **Conduct a likelihood ratio test of the hypothesis that opinion is independent of gender, given race. Interpret.**

We want to test

$$H_0: \beta_1^G = \beta_2^G = 0 \qquad \text{vs.} \qquad H_1: \text{ at least one } \beta_j^G \neq 0$$

Let's fit the model with only race effect:

```
               Estimate Std. Error z value Pr(>|z|)
(Intercept):1    2.3058     0.1321  17.452   <2e-16 ***
(Intercept):2    1.6560     0.1375  12.044   <2e-16 ***
race:1           0.7042     0.4091   1.721   0.0852 .
race:2          -0.4664     0.4530  -1.029   0.3032
---
Residual deviance: 46.8065 on 4 degrees of freedom
Log-likelihood: -42.1575 on 4 degrees of freedom
```

The likelihood ratio test is given by

$$-2(\log \hat{L}_{race} - \log \hat{L}_{gender,race}) = G^2_{race} - G^2_{gender,race} = 40.7316$$

with $df = 2$ which leads to a $p$-value $< 1.5 \times 10^{-9}$.

```
> 1-pchisq(-2*(-42.1575+21.7917),2)
[1] 1.429702e-09
```

Thus, we reject $H_0$. There is strong evidence that, given race, opinion is associated with gender.