Michael Leibert
Math 661
Homework 4

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.

Belief in heaven is ordinal, with categories (1=yes, 2 = unsure, 3 = no). The survey related $y$ =belief in heaven to two explanatory variables, gender $x_{i1}$ and race $x_{i1}$. The cumulative logit model of the proportional odds form with main effects has ML fit:

$$\text{logit}\left[\hat{P}\left(y_i \leq j\right)\right] = \hat{\alpha}_j + 0.76956x_{i1} + 1.01645x_{i2}$$

We interpret the race and gender effects.

$\hat{\alpha}_1 = 0.07631$: Is the estimated baseline log odds of falling into belief in heaven - yes versus all other categories for white male ($x_{i1} = x_{i2} = 0$).

$\hat{\alpha}_2 = 2.32238$: Is the estimated baseline log odds of falling into belief in heaven - unsure or belief in heaven - yes versus all other categories for white male ($x_{i1} = x_{i2} = 0$).

Note: `vglm` uses the parameterization $\alpha_j + \boldsymbol{x}_i^T\boldsymbol{\beta}$, so a positive $\beta_k$ indicates a greater likelihood of ending up into groups 1 and 2.

$\hat{\beta}_1 = 0.76956$. For a one unit increase in $x_{i1}$ (gender), controlling for $x_{i2}$, $\beta_1$ is the change in log-odds of falling into or below any category. The cumulative probability of falling into group $j$ or lower, is higher because $\beta_1 > 0$. That is controlling for race, females are more likely than males to fall into the categories: belief in heaven - yes or belief in heaven - unsure.

$\hat{\beta}_2 = 1.0165$. For a one unit increase in $x_{i2}$ (race), controlling for $x_{i1}$, $\beta_2$ is the change in log-odds of falling into or below any category. The cumulative probability of falling into group $j$ or lower, is higher because $\beta_2 > 0$. That is controlling for gender, blacks are more likely than whites to fall into the categories: belief in heaven - yes or belief in heaven - unsure.

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

```
> 1-pchisq(9.254,4)
[1] 0.05505495
```

The p-value, $P\left(\chi^2_4\right) = 0.05505495$, is $> .05$, so there's no evidence that the model does not fit the data well. The deviance goodness-of-fit test for the additive model suggests an okay but not great fit to the data.

A 95% confidence interval given for $\beta_1$ is given by:

$$\hat{\beta}_1 \pm z_{0.025}\ SE\left(\hat{\beta}_1\right) = 0.76956 \pm 1.96 \cdot 0.12253 = \big(0.5294012, 1.009719\big).$$

We are 95% confident that the log-odds for gender is expected to be at least 0.53 and at most 1.009 higher, on average, controlling for race. We note that the confidence interval does not contain zero, which shows evidence of the effect.

A 95% confidence interval given for $\beta_2$ is given by:

$$\hat{\beta}_2 \pm z_{0.025}\ SE\left(\hat{\beta}_2\right) = 1.01645 \pm 1.96 \cdot 0.21059 = \big(0.6036936, 1.429206\big).$$

We are 95% confident that the log-odds for race is expected to be at least 0.6 and at most 1.4 higher, on average, controlling for gender. We note that the confidence interval does not contain zero, which shows evidence of the effect.

A 95% confidence interval given for $\alpha_1$ is given by:

$$\hat{\alpha_1} \pm z_{0.025}\ SE\left(\hat{\alpha_1}\right) = 0.07631 \pm 1.96 \cdot 0.08963 = \big(-0.0993648, 0.2519848\big).$$

We are 95% confident that the baseline log-odds for gender is expected to be at least -0.099 and at most 0.252 higher, on average, controlling for race. We note that the confidence interval does contain zero, which does not show evidence of the effect.

A 95% confidence interval given for $\alpha_2$ is given by:

$$\hat{\alpha_2} \pm z_{0.025}\ SE\left(\hat{\alpha_2}\right) = 2.32238 \pm 1.96 \cdot 0.13522 = \big(2.057349, 2.587411\big).$$

We are 95% confident that the baseline log-odds for gender is expected to be at least -0.099 and at most 0.252 higher, on average, controlling for race. We note that the confidence interval does not contain zero, which does show evidence of the effect.

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

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

```
> mat
  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(formula = cbind(y1, y2, y3) ~ gender + race,
+ family = multinomial,data = mat) )

Call:
vglm(formula = cbind(y1, y2, y3) ~ gender + race, family = multinomial,
    data = mat)


Pearson residuals:
  log(mu[,1]/mu[,3]) log(mu[,2]/mu[,3])
1            -0.5504             1.5856
2             0.6302            -1.5095
3             0.1784            -0.4698
4            -0.2256             0.4294

Coefficients:
              Estimate Std. Error z value             Pr(>|z|)
(Intercept):1   1.7943     0.1675  10.712 < 0.0000000000000002 ***
(Intercept):2   1.5309     0.1717   8.918 < 0.0000000000000002 ***
gender:1        1.0339     0.2587   3.997            0.0000641 ***
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
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Number of linear predictors:  2

Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])

Residual deviance: 6.0748 on 2 degrees of freedom

Log-likelihood: -21.7917 on 2 degrees of freedom

Number of iterations: 4

No Hauck-Donner effect found in any of the estimates

Reference group is level  3  of the response
```

$$\log\frac{\pi_{i1}}{\pi_{i3}} = \log\frac{\frac{\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_1\right)}{\cancel{1+\sum_{h=1}^{c-1}\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_h\right)}}}{\frac{1}{\cancel{1+\sum_{h=1}^{c-1}\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_h\right)}}}$$
$$= \boldsymbol{x}_i^T\boldsymbol{\beta}_1$$
$$= 1.7943 + 1.0339x_{i1} + 0.6727x_{i2}$$

$$\log\frac{\pi_{i2}}{\pi_{i3}} = \log\frac{\frac{\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_2\right)}{\cancel{1+\sum_{h=1}^{c-1}\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_h\right)}}}{\frac{1}{\cancel{1+\sum_{h=1}^{c-1}\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_h\right)}}}$$
$$= \boldsymbol{x}_i^T\boldsymbol{\beta}_2$$
$$= 1.5309 + 0.3087x_{i1} - 0.4757x_{i2}$$

$$\begin{aligned}
\log\frac{\pi_{i1}}{\pi_{i2}} &= \log\left(\frac{\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_1\right)}{\cancel{1+\sum_{h=1}^{c-1}\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_h\right)}}\frac{\cancel{1+\sum_{h=1}^{c-1}\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_h\right)}}{\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_2\right)}\right)\\
&= \log\Big(\exp\left[\boldsymbol{x}_i^T\boldsymbol{\beta}_1\right]\Big) - \log\Big(\exp\left[\boldsymbol{x}_i^T\boldsymbol{\beta}_2\right]\Big)\\
&= \boldsymbol{x}_i^T\boldsymbol{\beta}_1 - \boldsymbol{x}_i^T\boldsymbol{\beta}_2\\
&= (1.7943 - 1.5309) + (1.0339 - 0.3087)\,x_{i1} + (0.6727 + 0.4757)\,x_{i2}\\
&= 0.2634469 + 0.7252119x_{i1} + 1.148419x_{i2}
\end{aligned}$$

(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 given for the odds ratio of $\beta_1^G$, the conditional gender effect, is given by:

$$\exp\left(\hat{\beta}_1^G \pm z_{0.025}\ SE\left(\hat{\beta}_1^G\right)\right) = \Big(\exp(0.526848)\,, \exp(1.540952)\Big) = \big(1.693586, 4.669033\big).$$

The odds of belief of an afterlife for females is $\exp(1.0339) = 2.8$ times higher than for males, controlling for race. The odds of belief of an afterlife are between 1.7 times and 4.7 times higher among females, adjusting for race. The 95% confidence interval does not contain 1, so there is evidence of the conditional gender effect, controlling for race.

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

$H_0$: opinion is independent of gender, given race

```
> heaven.fitt = vglm(cbind(y1,y2,y3)~race, family=multinomial, data=mat)
> heaven.fitt

Call:
vglm(formula = cbind(y1, y2, y3) ~ race, family = multinomial,     data = mat)


Coefficients:
(Intercept):1 (Intercept):2        race:1        race:2
    2.3057547     1.6559579     0.7041622    -0.4663739
```

```
Degrees of Freedom: 8 Total; 4 Residual
Residual deviance: 46.80651
Log-likelihood: -42.15753

This is a multinomial logit model with 3 levels
```

```
> 1-pchisq( 46.80651 -  6.0748, 4-2)
[1] 0.000000001429623
```

We reject $H_0$, and conclude there is evidence of a varying effect of gender across belief groups.