Psy 5011: Categorical Data Analysis **(Due: 12/05/20)**

Homework #6

Instructor: Yao, Kaiping Grace *Name:* Hao-Cheng Lo, *Id:* D08227104

Problem 6.4: For the belief in an afterlife example, describe the gender effect by reporting and interpreting the estimated conditional odds ratio for the (a) *undecided* and *no* pair of response categories, (b) *yes* and *undecided* pair. Interpret.

```
> life <- read.table("http://www.stat.ufl.edu/~aa/cat/data/Afterlife.dat", header=TRUE)
> fit640 <- vglm(cbind(yes,undecided,no) ~ gender, family=multinomial(refLevel="no"), data=life)
> summary(fit640)
# Pearson residuals:
#   log(mu[,1]/mu[,3]) log(mu[,2]/mu[,3])
# 1             0.1991            0.03849
# 2             0.4982            0.21597
# 3            -0.4716           -0.09120
# 4            -1.4535           -0.63008
#
# Coefficients:
#              Estimate Std. Error z value Pr(>|z|)
# (Intercept):1   1.5867     0.1163  13.639   <2e-16 ***
# (Intercept):2  -0.4282     0.1688  -2.537   0.0112 *
# gendermale:1   -0.4008     0.1705  -2.350   0.0188 *
# gendermale:2   -0.0906     0.2457  -0.369   0.7123
#
# Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
#
# Residual deviance: 2.8481 on 4 degrees of freedom
# Log-likelihood: -20.7295 on 4 degrees of freedom
# Reference group is level  3  of the response
> fit641 <- vglm(cbind(yes,undecided,no) ~ gender, family=multinomial(refLevel="undecided"), data=
    life)
> summary(fit641)
# Pearson residuals:
#   log(mu[,1]/mu[,2]) log(mu[,3]/mu[,2])
# 1             0.1402            -0.1464
# 2             0.2864            -0.4613
# 3            -0.3323             0.3469
# 4            -0.8356             1.3459
#
# Coefficients:
#              Estimate Std. Error z value Pr(>|z|)
# (Intercept):1   2.0149     0.1398  14.414   <2e-16 ***
# (Intercept):2   0.4282     0.1688   2.537   0.0112 *
# gendermale:1   -0.3101     0.2078  -1.493   0.1355
# gendermale:2    0.0906     0.2457   0.369   0.7123
#
# Names of linear predictors: log(mu[,1]/mu[,2]), log(mu[,3]/mu[,2])
#
# Residual deviance: 2.8481 on 4 degrees of freedom
# Log-likelihood: -20.7295 on 4 degrees of freedom
# Reference group is level  2  of the response
```

For (a), with *no* as the baseline category for Y, the model is $\log(\frac{\pi_j}{\pi_3}) = \alpha_j + \beta_j x_G,\ j = 1, 2$. We can find that for the *undecided* and *no* pair of response categories, the estimated model is $\log(\frac{\pi_2}{\pi_3}) = \hat{\alpha}_2 + \hat{\beta}_2 x_G = -0.43 - 0.09 x_G$. Since $\hat{\beta}_2 = -0.09$, for males the estimated odds of response *undecided* rather than *no* on life after death are $e^{-0.09} = 0.91$ times those for females.

For (b), with *undecided* as the baseline category for Y, the model is $\log(\frac{\pi_j}{\pi_2}) = \alpha_j + \beta_j x_G,\ j = 1, 3$. We can find that for the *yes* and *undecided* pair of response categories, the estimated model is $\log(\frac{\pi_1}{\pi_2}) = \hat{\alpha}_1 + \hat{\beta}_1 x_G = 2.01 - 0.31 x_G$. Since $\hat{\beta}_1 = -0.31$, for males the estimated odds of response *yes* rather than *undecided* on life after death are $e^{-0.31} = 0.73$ times those for females.

Problem 6.4: *Conti.*

With no doubt, we can derive the results of (b) from (a) (here, I omit the symbol hat):

$$
\begin{aligned}
\log(\frac{\pi_1}{\pi_2}) &= \log(\frac{\pi_1/\pi_3}{\pi_2/\pi_3}) \\
&= \log(\frac{\pi_1}{\pi_3}) - \log(\frac{\pi_2}{\pi_3}) \\
&= \alpha_1 + \beta_1 x_G - \alpha_2 + \beta_2 x_G \\
&= (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2) x_G \\
&= (1.58 - (-0.43)) + (-0.40 - (-0.09)) x_G \\
&= 2.01 - 0.31 x_G
\end{aligned} \tag{1}
$$

Problem 6.14 (2 Ed.): Consider Exercise 6.4 on belief in an afterlife. Fit a model using (a) adjacent-categories logits, (b) alternative ordinal logits.

The alternative models for ordinal logits suit for the data with ordinal nature of Y. Hence, how levels of Y order will affect the results. In this data, we order Y = belief in an afterlife with $1 =$ *yes*, $2 =$ *undecided*, and $3 =$ *no*.

For (a), the model is:

$$\log(\frac{\pi_j}{\pi_{j+1}}) = \alpha_j + \beta x_G,\ j = 1, 2 \tag{2}$$

The R output:

```
# adjacent-categoriy ordinal logits
> f6140 <-vglm(cbind(yes,undecided,no) ~ gender, family=acat(parallel=T,reverse=T), data=life)
> summary(f6140)
# Pearson residuals:
#   loglink(P[Y=1]/P[Y=2]) loglink(P[Y=2]/P[Y=3])
# 1                0.34691               -0.09555
# 2                0.04194                0.73211
# 3               -0.24367               -0.45435
# 4               -0.91658               -1.24227
#
# Coefficients:
#               Estimate Std. Error z value Pr(>|z|)
# (Intercept):1  1.97057    0.10892  18.092  < 2e-16 ***
# (Intercept):2 -0.37278    0.13002  -2.867  0.00414 **
# gendermale    -0.21095    0.08277  -2.549  0.01081 *
#
# Names of linear predictors: loglink(P[Y=1]/P[Y=2]), loglink(P[Y=2]/P[Y=3])
#
# Residual deviance: 3.1183 on 5 degrees of freedom
# Log-likelihood: -20.8646 on 5 degrees of freedom
```

The gender effect is $\hat{\beta} = -0.21$. The estimated odds that male is in category j instead of $j+1$ are $e^{-0.21} = 0.81$ times the estimated odds for female. The estimated odds ratio for an arbitrary pair of $a < b$ equals $e^{\hat{\beta}(ba)}$. For example, the estimated odds that a male's belief in an afterlife is *yes* (category 1) instead of *no* (category 3) are $e^{-0.21(3-1)} = (0.81)^2 = 0.66$ times those for female. The model fit has deviance $G^2 = 3.12$ with $df = 5$, indicating an adequate fit.

Problem 6.14 (2 Ed.):*Conti.*

For (b), we take continuation-ratio logits (sequential logits):

$$\log(\frac{\pi_1}{\pi_2+\pi_3}) = \alpha_1 + \beta x_G,\ \log(\frac{\pi_2}{\pi_3}) = \alpha_2 + \beta x_G \tag{3}$$

The R output:

```
# seq ordinal logits
> f6141 <-vglm(cbind(yes,undecided,no) ~ gender, family=sratio(parallel=T), data=life)
> summary(f6141)
# Pearson residuals:
#   logitlink(P[Y=1|Y>=1]) logitlink(P[Y=2|Y>=2])
# 1                0.47955                -0.4275
# 2                0.09666                 0.9065
# 3               -0.30630                -0.4362
# 4               -1.39684                -0.7132
#
# Coefficients:
#              Estimate Std. Error z value Pr(>|z|)
# (Intercept):1  1.05413    0.08931  11.802   <2e-16 ***
# (Intercept):2 -0.33205    0.13470  -2.465   0.0137 *
# gendermale    -0.29712    0.12324  -2.411   0.0159 *
#
# Names of linear predictors: logitlink(P[Y=1|Y>=1]), logitlink(P[Y=2|Y>=2])
#
# Residual deviance: 3.7876 on 5 degrees of freedom
# Log-likelihood: -21.1992 on 5 degrees of freedom
```

For this model, the estimate of the gender effect is $\hat{\beta} = -0.29$ and $e^{-0.29} = 0.74$. For example, given that belief in an afterlife is *undecided*, the estimated odds for males having *undecided* belief in an afterlife rather than having *no* belief in an afterlife were 0.74 times the estimated odds for females. The model fits the data very well with deviance 3.79 ($df = 5$).

Problem 6.8: Table 6.8 results from a clinical trial for the treatment of small-cell lung cancer. Patients were randomly assigned to two treatment groups. The sequential therapy administered the same combination of chemotherapeutic agents in each treatment cycle. The alternating therapy used three different combinations, alternating from cycle to cycle. Fit a cumulative logit model with a proportional odds structure. Interpret the estimated treatment effect. Check whether a model allowing interaction provides a significantly better fit.

Table 6.8 Data for Exercise 6.8 on lung cancer treatment.

		Response to Chemotherapy			
Treatment Therapy	Gender	Progressive Disease	No Change	Partial Remission	Complete Remission
Sequential	Male	28	45	29	26
	Female	4	12	5	2
Alternating	Male	41	44	20	20
	Female	12	7	3	1

Source: Holtbrugge, W., and Schumacher, M., *Appl. Statist.* **40**: 249–259 (1991).

We consider x_1 = therapy and x_2 = gender as our explanatory variables. The Model 1 without interaction term is:

$$\text{logit}[P(Y \leq j)] = \alpha_j + \beta_1 x_1 + \beta_2 x_2, \ j = 1, 2$$

The Model 2 with interaction term is:

$$\text{logit}[P(Y \leq j)] = \alpha_j + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2, \ j = 1, 2$$

The R output:

```
# therapy + gender
> fit685 <- vglm(cbind(responseNo,responsePartial,responseComplete) ~ therapy + gender,
>                family=cumulative(parallel=TRUE),
>                data=lung)
> summary(fit685)
# Pearson residuals:
#   logitlink(P[Y<=1]) logitlink(P[Y<=2])
# 1            -0.1504            0.09284
# 2            -0.2016            0.59157
# 3             0.4295           -0.46548
# 4            -0.4865            0.45032
#
# Coefficients:
#               Estimate Std. Error z value Pr(>|z|)
# (Intercept):1   0.7526     0.4122   1.826   0.0678 .
# (Intercept):2   1.9649     0.4303   4.566 4.98e-06 ***
# therapyseq     -0.2152     0.2623  -0.820   0.4120
# genderM        -0.7150     0.4008  -1.784   0.0744 .
#
# Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2])
#
# Residual deviance: 1.2642 on 4 degrees of freedom
# Log-likelihood: -15.0043 on 4 degrees of freedom
#
# Exponentiated coefficients:
# therapyseq    genderM
#  0.8064017  0.4892116
# therapy + gender + interaction
> fit686 <- vglm(cbind(responseNo,responsePartial,responseComplete) ~ therapy + gender +  therapy:gender,
>                family=cumulative(parallel=TRUE),
>                data=lung)
> lrtest(fit686,fit685)
#   #Df  LogLik Df  Chisq Pr(>Chisq)
# 1   3 -14.973
# 2   4 -15.004  1 0.0621     0.8033
```

Problem 6.8: *Conti.*

By observing the results of Model 1, The estimated effect of therapy is $\hat{\beta}_1 = -0.21$ ($SE = 0.26$). Adjusting gender effect, for any fixed j, the estimated odds that a sequential therapy's response is in the no remission direction rather than the remission direction (i.e., $Y \leq j$ rather than $Y > j$) equal $e^{\hat{\beta}_1} = e^{-0.21} = 0.806$ times the estimated odds for alternating therapy. The estimated odds that a sequential therapy's response is in the remission direction rather than the no remission direction (i.e., $Y > j$ rather than $Y \leq j$) equal $e^{-\hat{\beta}_1} = e^{0.21} = 1.23$ times the estimated odds for alternating therapy. Therefore, sequential therapy tends to be much more therapeutic than alternating therapy. The model fits the data very well with deviance 1.26 ($df = 4$).

By observing the results of Model 2 and the LRT model comparison. We can find that after adding the interaction term, there provides no significantly better fit.

Problem 6.10: For the mental impairment example (Section 6.3.4), when we add an interaction term, we obtain the fit $\text{logit}[\hat{P}(Y \le j)] = \hat{\alpha_j} - 0.420x_1 + 0.371x_2 + 0.181x_1x_2$. The coefficient 0.181 of x_1x_2 has $SE = 0.238$. Find the estimated effect of life events for the low SES group ($x_2 = 0$) and for the high SES group ($x_2 = 1$). Explain why these suggest that the impact of life events may be more severe for the low SES group. Does the difference in effects seem to be significant?

Note that following the latent variable motivation, we can depict $\text{logit}[P(Y \le j)] = \alpha_j - \beta^T x$. When talking about the coefficients, we obey the formula above.

For the low SES group ($x_2 = 0$), the model is:

$$\text{logit}[\hat{P}(Y \le j)] = \hat{\alpha_j} - 0.420x_1$$

Therefore, the the estimated effect of life events for the low SES group ($x_2 = 0$) is $\hat{\beta}_1 = -(-0.420) = 0.420$, indicating that mental impairment tends to be worse with higher life events.

For the high SES group ($x_2 = 1$), the model is:

$$\text{logit}[\hat{P}(Y \le j)] = \hat{\alpha_j} - 0.420x_1 + 0.371 + 0.181x_1 = \hat{\alpha_j} + 0.371 - 0.239x_1$$

Therefore, the the estimated effect of life events for the high SES group ($x_2 = 1$) is $\hat{\beta}_1 = -(-0.239) = 0.239$, indicating that mental impairment tends to be worse with higher life events.

The reason why these results suggest that the impact of life events may be more severe for the low SES group is that $\hat{\beta}_{1|x_2=0} = 0.420 > \hat{\beta}_{1|x_2=1} = 0.239$. In fact, the difference in effects of life events may be not significant, because the z score of interaction term is $\dfrac{\hat{\beta}_3}{SE} = -0.76$ and P-value $= .4472$, indicating the effect of life events may not relate to SES.