STAT 477/STAT 577
HW 9 - Module 3: Sections 3 through 6

1. In the 1998 General Social Survey, respondents were classified according to their sex, race, and belief in an afterlife. The data are given in the table below. The data file for analysis (**afterlife.csv**) is located with this assignment in Canvas.

		Belief in Afterlife		
Race	Sex	Yes	Undecided	No
White	Female	371	49	74
	Male	250	45	71
Black	Female	64	9	15
	Male	25	5	13

   (a) Give the equation for predicting the probability for each of the three categories (using Belief in Afterlife = No as the baseline category) from a person's race and sex using a nominal regression model.

   Read in the data.

```
life.data<- read.csv(file.choose(), header = T)
```

   Set the baseline category for the variable Afterlife to No and fit the nominal response model.

```
life.data$Afterlife<- factor(life.data$Afterlife, levels =
                                c("No", "Undecided", "Yes"))
life.model<- multinom(Afterlife ~ Gender + Race,
                      data = life.data)

## # weights:  12 (6 variable)
## initial  value 1088.724778
## iter  10 value 773.755197
## final  value 773.726510
## converged
```

   The model is

```
summary(life.model)

## Call:
## multinom(formula = Afterlife ~ Gender + Race, data = life.data)
##
## Coefficients:
##             (Intercept) GenderMale RaceWhite
## Undecided     -0.652994 -0.1050684 0.2710218
## Yes            1.301614 -0.4185635 0.3417684
```

```
##
## Std. Errors:
##             (Intercept) GenderMale RaceWhite
## Undecided    0.3404666  0.2465102 0.3541320
## Yes          0.2264952  0.1712549 0.2370373
##
## Residual Deviance: 1547.453
## AIC: 1559.453
```

(b) Use the model from part (a) to estimate the probabilities for each of the three categories for belief in afterlife for all four combinations of sex and race.

Below are the four conditional logit values for undecided versus no, one for each of the four combinations of sex and race.

```
male.white.U<- -0.652994 - 0.1050684 + 0.2710218
female.white.U<- -0.652994 + 0.2710218
male.black.U<- -0.652994 - 0.1050684
female.black.U<- -0.652994
```

Below are the four conditional logit values for yes versus no, one for each of the four combinations of sex and race.

```
male.white.Y<- 1.301614 - 0.4185635 + 0.3417684
female.white.Y<- 1.301614 + 0.3417684
male.black.Y<- 1.301614 - 0.4185635
female.black.Y<- 1.301614
```

Now, we can calculate the probabilities of the three categories for afterlife for each of the four combinations of sex and race. For white males:

```
#Yes
exp(male.white.Y)/(1 + exp(male.white.Y) + exp(male.white.U))

## [1] 0.6782693

#Undecided
exp(male.white.U)/(1 + exp(male.white.Y) + exp(male.white.U))

## [1] 0.1224478

#No
1/(1 + exp(male.white.Y) + exp(male.white.U))

## [1] 0.1992829
```

For white females:

```r
#Yes
exp(female.white.Y)/(1 + exp(female.white.Y) + exp(female.white.U))

## [1] 0.754562

#Undecided
exp(female.white.U)/(1 + exp(female.white.Y) + exp(female.white.U))

## [1] 0.09956223

#No
1/(1 + exp(female.white.Y) + exp(female.white.U))

## [1] 0.1458757
```

For black males:

```r
#Yes
exp(male.black.Y)/(1 + exp(male.black.Y) + exp(male.black.U))

## [1] 0.6221676

#Undecided
exp(male.black.U)/(1 + exp(male.black.Y) + exp(male.black.U))

## [1] 0.1205539

#No
1/(1 + exp(male.black.Y) + exp(male.black.U))

## [1] 0.2572785
```

For black females:

```r
#Yes
exp(female.black.Y)/(1 + exp(female.black.Y) + exp(female.black.U))

## [1] 0.7073575

#Undecided
exp(female.black.U)/(1 + exp(female.black.Y) + exp(female.black.U))

## [1] 0.100176

#No
1/(1 + exp(female.black.Y) + exp(female.black.U))

## [1] 0.1924665
```

(c) Test for the significance of sex in the model for predicting belief in the afterlife.

To test for the significance of sex in the model, we will need to run the model with just race.

```
life.model.race<- multinom(Afterlife ~ Race, data = life.data)

## # weights:  9 (4 variable)
## initial  value 1088.724778
## final  value 777.322834
## converged
```

We will then compare the deviances of the two models, the one with both Race and Sex and the one with just Race. This is

```
anova(life.model.race, life.model, test = "Chisq")

##             Model Resid. df Resid. Dev   Test    Df LR stat.    Pr(Chi)
## 1            Race      1978   1554.646           NA       NA         NA
## 2 Gender + Race      1976   1547.453 1 vs 2     2 7.192648 0.02742434
```

The test statistic is 7.1926 and the p-value is 0.0274. We will conclude we have moderate evidence a person's sex helps to explain their belief in afterlife in the model that includes Race.

(d) Test for the significance of race in the model for predicting belief in the afterlife.

To test for the significance of race in the model, we will need to run the model with just sex.

```
life.model.sex<- multinom(Afterlife ~ Gender, data = life.data)

## # weights:  9 (4 variable)
## initial  value 1088.724778
## iter  10 value 774.723632
## iter  10 value 774.723631
## final  value 774.723631
## converged
```

We will then compare the deviances of the two models, the one with both Race and Sex and the one with just Sex. This is

```
anova(life.model.sex, life.model, test = "Chisq")

##             Model Resid. df Resid. Dev   Test    Df LR stat. Pr(Chi)
## 1          Gender      1978   1549.447           NA       NA      NA
## 2 Gender + Race      1976   1547.453 1 vs 2     2 1.994243 0.36894
```

The test statistic is 1.9942 and p-value is 0.3689. We will conclude we have no evidence a person's Race helps to explain their belief in afterlife in the model that includes Sex.

2. The following table of counts was obtained from a random sample of 1,397 respondents from the population of adults (more then 18 years old) in the United States in 1982. Each respondent was cross-classified with respect to opinions regarding gun registration as a part of comprehensive gun control legislation and imposing the death penalty on adults convicted of certain violent acts. The data file for analysis (**gunpenalty.csv**) is located with this assignment in Canvas.

	Death Penalty	
Gun Registration	Favor	Oppose
Favor	784	236
Oppose	311	66

(a) Fit the log linear model for the counts in the contingency table under the assumption of independence. Set the baseline categories for both variables to Oppose. Use the estimated coefficients to calculate the four expected counts in the table. Does this model fit the data? Explain your answer.

To begin, we will read in the data from the file **Opinion.csv**.

```
gp.data<- read.csv(file.choose(), header = T)
```

Since the data is currently in long format, we can move to setting the baseline categories for `Gun.Registration` and `Death.Penalty`.

```
gp.data$Gun.Register<- factor(gp.data$Gun.Register,
                              levels = c("Oppose", "Favor"))
gp.data$Death.Penalty<- factor(gp.data$Death.Penalty,
                              levels = c("Oppose", "Favor"))
```

Here is the command to fit the independence model to these data.

```
gpI.model<- glm(Count ~ Gun.Register + Death.Penalty,
                            data = gp.data,
                            family = poisson(link = "log"))
```

The coefficients in the model are:

```
gpI.model$coefficients

##        (Intercept)  Gun.RegisterFavor Death.PenaltyFavor
##          4.4005898          0.9953127          1.2880826
```

The expected counts for the four rows in the data table are:

```
##Gun.Register = Favor, Death.Penalty = Favor
exp(gpI.model$coefficients[1]
    + gpI.model$coefficients[2]
    + gpI.model$coefficients[3])
```

```
## (Intercept)
##    799.4989

##Gun.Register = Favor, Death.Penalty = Oppose
exp(gpI.model$coefficients[1]
    + gpI.model$coefficients[2])

## (Intercept)
##    220.5011

##Gun.Register = Oppose, Death.Penalty = Favor
exp(gpI.model$coefficients[1]
    + gpI.model$coefficients[3])

## (Intercept)
##    295.5011

##Gun.Register = Oppose, Death.Penalty = Oppose
exp(gpI.model$coefficients[1])

## (Intercept)
##    81.49893
```

You can check your values using the `fitted.values` output from the model.

```
gpI.model$fitted.values

##         1         2         3         4
## 799.49893 220.50107 295.50107  81.49893
```

To look at the goodness of fit of this model, we will need the model deviance and its p-value. This is

```
gpI.model$deviance

## [1] 5.32065

pchisq(gpI.model$deviance, 1, lower.tail = F)

## [1] 0.02107415
```

We have moderately strong evidence the independence model does not fit the data.

(b) Fit the saturated log linear model for the counts in the contingency table, again using the category Oppose as the baseline category. Use the model to obtain the odds ratio for the table and interpret its value.

The saturated log-linear model for the counts is

```r
gpS.model<- glm(Count ~ Gun.Register + Death.Penalty +
                  Gun.Register:Death.Penalty, data = gp.data,
                  family = poisson(link = "log"))
```

The 4th coefficient in the model will be the interaction term between the variables `Gun.Register` and `Death.Penalty`. The predicted odds ratio for these two variables is:

```r
exp(gpS.model$coefficients[4])

## Gun.RegisterFavor:Death.PenaltyFavor
##                            0.7049975
```

Since the value of the predicted odds ratio is less than 1, I will interpret the reciprocal value. There are two ways to interpret this odds ratio:

- The predicted odds a person was in favor of the death penalty if they opposed gun registration as a part of gun control is $1/0.7050 = 1.4184$ times the predicted odds a person was in favor of the death penalty if they were in favor of gun registration as a part of gun control.
- The predicted odds a person was in favor of gun registration as a part of gun control if they opposed the death penalty is $1/0.7050 = 1.4184$ times the predicted odds a person was in favor of gun registration as a part of gun control if they were in favor of the death penalty.

3. During the semester, we have looked at survey data on shower habits collected from students in STAT 101 at Iowa State University. In the survey data, we asked students to specify which direction they face (either toward the shower or away from the shower) when wetting, lathering, and rinsing their hair. The three-way cross-classification table for the three variables, Wet, Lather, and Rinse is given below. The data file for analysis (**shower.csv**) is located with this assignment in Canvas. Set the baseline categories for all three variables to the category Toward.

		Rinse	
Wet	Lather	Toward	Away
Toward	Toward	198	44
	Away	273	72
Away	Toward	26	515
	Away	74	621

Find the log-linear model that best fits the counts in the data table. Then, give an interpretation of all interaction terms in this model.

To begin, we will read in the data from the file **Shower.csv**.

```r
shower.data<- read.csv(file.choose(), header = T)
```

Since the data is already in long format, we just need to set the baseline categories for the three variables `Wet` and `Lather`, and `Rinse`.

```r
shower.data$Wet<- factor(shower.data$Wet,
                         levels = c("Toward", "Away"))
shower.data$Lather<- factor(shower.data$Lather,
                            levels = c("Toward", "Away"))
shower.data$Rinse<- factor(shower.data$Rinse,
                           levels = c("Toward", "Away"))
```

To begin, we will look at the fit of the homogeneous association model. If this model fits the data, then we will look at simpler models. If this model does not fit the data, then we know we must use the saturation model. The homogeneous association model includes all two-way interaction terms, but not the three-way interaction term. For these data, the model is

```r
showerHA.model<-
  glm(Count ~ Wet + Lather + Rinse +
      Wet:Lather + Wet:Rinse + Lather:Rinse,
      data = shower.data, family = poisson(link = "log"))
```

The deviance and p-value for the goodness of fit for this model is:

```r
showerHA.model$deviance

## [1] 11.02128

pchisq(showerHA.model$deviance,
           showerHA.model$df.residual, lower.tail = F)

## [1] 0.0009007176
```

The p-value is very small so we have very strong evidence the homogeneous association model does not fit the data.

So, at this point, we will fit the saturation model to obtain the coefficients for the two-way and three-way interactions. This model is

```r
showerS.model<-
  glm(Count ~ Wet + Lather + Rinse +
      Wet:Lather + Wet:Rinse + Lather:Rinse + Wet:Lather:Rinse,
      data = shower.data, family = poisson(link = "log"))
```

The coefficients in this model are:

```
showerS.model$coefficients

##                 (Intercept)                      WetAway
##                   5.2882670                   -2.0301705
##                  LatherAway                    RinseAway
##                   0.3212048                   -1.5040774
##          WetAway:LatherAway            WetAway:RinseAway
##                   0.7247638                    4.4901478
##        LatherAway:RinseAway WetAway:LatherAway:RinseAway
##                   0.1712717                   -1.0300761
```

The odds ratios can be found using coefficients 5 through 8 in the model. Here are the values and interpretations.

- **Wet and Lather**
  - When Rinse is Toward

    ```
    exp(showerS.model$coefficient[5])

    ## WetAway:LatherAway
    ##           2.064243
    ```

    For a student who faces toward the shower when they rinse their hair, the predicted odds they face away from the shower when they wet their hair if they also face away from the shower when they lather their hair is 2.0642 times the predicted odds they face Away from the shower when they wet their hair if they face toward the shower when they lather their hair.
  - When Rinse is Away

    ```
    exp(showerS.model$coefficient[5] +
          showerS.model$coefficient[8])

    ## WetAway:LatherAway
    ##          0.7368932
    ```

    For a student who faces away from the shower when rinsing their hair, the predicted odds they face away from the shower when they wet their hair if they face toward the shower when they lather their hair is $1/0.7369 = 1.357$ times the predicted odds they face away from the shower when they wet their hair if they away from the shower when they lather their hair.
- **Wet and Rinse**
  - When Lather is Toward

    ```
    exp(showerS.model$coefficient[6])

    ## WetAway:RinseAway
    ##          89.13462
    ```

For a student who faces toward the shower when lathering their hair, the predicted odds they face away from the shower when they wet their hair if they also face away from the shower when they rinse their hair is 89.1346 times the predicted odds they face away from the shower when they wet their hair if they face toward the shower when the rinse their hair.

- When Lather is Away

```
exp(showerS.model$coefficient[6] +
      showerS.model$coefficient[8])
```

```
## WetAway:RinseAway
##          31.81926
```

For a student who faces away from the shower when lathering their hair, the predicted odds they face away from the shower when they wet their hair if they also face away from the shower when they rinse their hair is 31.8193 times the predicted odds they face away from the shower when they wet their hair if they face toward the shower when the rinse their hair.

- **Lather and Rinse**
  - When Wet is Toward

```
exp(showerS.model$coefficient[7])
```

```
## LatherAway:RinseAway
##             1.186813
```

For a student who faces toward the shower when they wet their hair, the predicted odds they face away from the shower when they lather their hair if they face away from the shower when they rinse their hair is 1.1868 times the predicted odds they face away from the shower when they lather their hair if they face toward the shower when they rinse their hair.

- When Wet is Away

```
exp(showerS.model$coefficient[7] +
      showerS.model$coefficient[8])
```

```
## LatherAway:RinseAway
##            0.4236683
```

For a student who faces away from the shower when they wet their hair, the predicted odds they face away from the shower when they lather their hair if they face toward the shower when they rinse their hair is $1/0.4237 = 2.36016$ times the predicted odds they face away from the shower when they lather their hair if they face away from the shower when they rinse their hair.