Matt Isaac

Rates and Proportions

Homework 2 – Due 10/2/17

1. Data from problem 7.4
   1. Fitting the log-linear model with all 2-way but no 3-way interactions resulted in a deviance of 0.3007 and 1 degree of freedom. P(X21 ≥ 0.3007) = 0.5834. This p-value is not statistically significant, indicating that this model adequately describes these data.
   2. Gender\*Health\_Opinion was the least significant 2-way interaction (p = 0.1490), and was the first term to be removed. The resulting model gave a deviance of 2.3831 on 2 degrees of freedom. P(X22 ≥ 2.3831) = 0.3038. Removing Gender\*Health\_Opinion gives a model that still adequately describes these data.

Gender\*Inf\_Opinion is the least significant, although it is borderline (p = 0.0737). I removed it to see if the resulting model still adequately described the data. Removing this term gave the new model a deviance of 5.5810 on 3 degrees of freedom. P(X23 ≥ 5.5810) = 0.1339. This model still adequately describes the data.

The remaining terms in the model (Gender, Inf\_Opinion, Health\_Opinion, and Inf\_Opinion\*Health\_Opinion) are all highly statistically significant. As stated previously, this model adequately describes the data (p = 0.1339). The table below displays the terms left in the final model.

| **LR Statistics For Type 3 Analysis** | | | |
| --- | --- | --- | --- |
| **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** |
| **Gender** | 1 | 12.23 | 0.0005 |
| **Inf\_Opinion** | 1 | 343.09 | <.0001 |
| **Health\_Opinion** | 1 | 66.48 | <.0001 |
| **Inf\_Opini\*Health\_Opi** | 1 | 10.74 | 0.0011 |

* 1. Independence Graph:

Gender

Inf\_Opinion

Health\_Opinion

As shown in the independence graph, Gender and Inf\_Opinion are conditionally independent. Gender and Health\_Opinion are also conditionally independent. Inf\_Opinion and Health\_Opinion, however, are not conditionally independent. There is some relationship or dependence between people supporting government information program and supporting government funded healthcare for AIDS patients.

* 1. Are persons who support government information program more likely or less likely to support the government paying for health care costs of AIDS patients? By how much?

Inf\_Opinion (support) Health\_Opinion(support)

OR = exp{βsupport, support - βsupport, oppose – βoppose, support + βoppose,oppose}

OR = exp{0 – 0 – 0 + 0.8724}

OR = exp{0.8724}

OR = 2.3926

Those who support the government information program were more likely to support the government paying for healthcare of AIDS patients than those who did not support the government information program. Specifically, people who support the government information program were 2.3926 times as likely to also support the government paying for the healthcare of AIDS patients as those who did not support the government information program (and vise-versa).

1. Data from problem 7.6
   1. The model with all three way interactions and no four way interactions does not adequately describe the model. The deviance is 7.0963 on 1 degree of freedom. This gives us P(X21 ≥ 7.0963) = 0.0077 for a p-value.
   2. Strangely enough, removing all the three way interactions and fitting a model with only the two way interactions yields a model that is adequate to describe the data. We get a deviance of 10.1617 and a p-value of P(X25 ≥ 10.1617) = 0.0708.

From the model with all two-way interactions, I removed ei\*jp and ei\*tf one at a time. This left us with a final model. All terms in the model are statistically significant except for jp (p-value = 0.4749). The term jp was left in the model because it is involved in two highly statistically significant interactions: sn\*jp and tf\*jp (both have p-value < 0.0001). Our final model consists of the following terms: ei, sn, tf, jp, ei\*sn, sn\*tf, sn\*jp, tf\*jp. See table below.

| **LR Statistics For Type 3 Analysis** | | | |
| --- | --- | --- | --- |
| **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** |
| **ei** | 1 | 8.63 | 0.0033 |
| **sn** | 1 | 178.91 | <.0001 |
| **tf** | 1 | 73.70 | <.0001 |
| **jp** | 1 | 0.51 | 0.4749 |
| **ei\*sn** | 1 | 5.60 | 0.0180 |
| **sn\*tf** | 1 | 7.91 | 0.0049 |
| **sn\*jp** | 1 | 74.82 | <.0001 |
| **tf\*jp** | 1 | 17.32 | <.0001 |

ei

tf

jp

sn

The term ei is conditionally independent from tf and jp. This model has a lot of interactions. There is a significant dependence/relationship between sn and tf; sn and ei; sn and jp; and tf and jp.

* 1. Odds Ratios:

ei (i) and sn (s)

OR = exp{βi,s – βi,n – βe,s + βe,n}

OR = exp{0 – 0 – 0 + 0.3219}

OR = exp{0.3219}

OR = 1.380

People who are type “i” are 1.380 times as likely to also be of type “s” as people who are not of type “i”. People who are not of type “i” are 1.380 times as likely to not be type “s” as people who are of type “i”.

tf (f) and jp (j)

OR = exp{βf,j – βf,p – βt,j + βt,p}

OR = exp{-0.5585 – 0 – 0 + 0 }

OR = exp{-0.5585}

OR = 0.5721

People who are of type “f” are only 0.5721 times as likely to also be of type “j” as people who are of type “t”. Likewise, people who are not of type “f” are 0.5721 times as likely to not be of type “j” as people who are of type “f”.

1. Data from problem 7.13
   1. I fit a log-linear model with all three-way but no four-way interactions. I saw the following warning about the Hessian matrix: “Warning: Negative of Hessian not positive definite. The convergence is questionable.” In the SAS LOG, there was also a warning that said that the validity of this model is questionable. Our model is still valid, but this warning occurs because sometimes with higher order models, the interactions may be collinear with the variables.
   2. The deviance for the model with all three-way interactions is 8.5237. Assuming that the value for deviance is valid, our p-value is P(X216 ≥ 8.5237) = 0.9317. This indicates that the model does adequately describe these data.
   3. The model with all two way interactions (homogeneous associations model) adequately describes these data. The deviance is 31.6695 on 48 degrees of freedom. This gives a p-value of P(X248 ≥ 31.6695) = 0.9667.
   4. After removing the two least significant interaction terms (environment\*law, health\*cities) we are left with a final model that adequately describes the data. This model includes the following terms: environment, health, cities, law, environment\*health, environment\*cities, health\*law, cities\*law.

I stopped removing terms at this point because all the remaining terms are statistically significant.

| **LR Statistics For Type 3 Analysis** | | | |
| --- | --- | --- | --- |
| **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** |
| **environment** | 2 | 104.01 | <.0001 |
| **health** | 2 | 65.84 | <.0001 |
| **cities** | 2 | 37.38 | <.0001 |
| **law** | 2 | 94.39 | <.0001 |
| **environment\*health** | 4 | 23.36 | 0.0001 |
| **environment\*cities** | 4 | 17.38 | 0.0016 |
| **health\*law** | 4 | 27.93 | <.0001 |
| **cities\*law** | 4 | 13.82 | 0.0079 |

Environment

Cities

Law Enforcement

Health

Opinion on government spending on the environment and opinion on spending on law enforcement are conditionally independent. Opinion on spending on healthcare and spending on large cities are also conditionally independent. There is some relationship and dependence between opinion on spending on environment and healthcare; environment and cities; law enforcement and healthcare; and law enforcement and cities.

* 1. From looking at the parameter estimates, it looks to me like people who believe there is too little government spending in one area are more likely to believe that there is too little spending in another area as well. When looking at the parameter estimates, the largest estimated parameter values are for those responses where a person answered “too little” (1) two or three times. The parameter values for answering “too little” zero or one time are almost always smaller.
  2. Compute the following odds ratios:

Environment (3) and Cities (3)

OR = exp{βEnv(3),Cities(3)– βEnv(3),Cities(2) – βEnv(2),Cities(3) + βEnv(2),Cities(2)}

OR = exp{0 – 0 – 0 + 1.4275 }

OR = exp{1.4275 }

OR = 4.1683

People who think that the government is spending too much on the environment are 4.1683 times as likely to think that the government is spending too much on the big cities as compared to people who don’t think the government is spending too much on the environment (and vice-versa).

Environment (1) and Cities (1)

OR = exp{βEnv(1),Cities(1)– βEnv(1),Cities(2) – βEnv(2),Cities(1) + βEnv(2),Cities(2)}

OR = exp{1.1597 – 1.4889 – 0.6443 + 1.4275}

OR = exp{0.454 }

OR = 1.5746

People who think the government is spending too little on the environment are 1.5746 times as likely to think that the government is also spending too little on large cities (and vise versa). People who think the government is spending just the right amount on the environment are 1.5746 times as likely to think that the government is spending just the right amount on large cities.

Health(3) and Cities(3)

This last odds ratio requested is 1. I know this without doing any calculations because opinion on healthcare spending and opinion on large city spending are conditionally independent. Thus the parameters will all be equal to 0, and exp{0} = 1.