Logistic Regression Homework Assignment #1

*Homework Team 7: Diana Shaw, Bright Small, Andrew Williams, Logan Corbett, Rob Causey*

[Insert Summary]

## Examining Flu Association with Gender and Income

[Text]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table of Flu Cases by Gender** | | | | |
| **Gender** | | **Flu** | | **Total** |
| **No** | **Yes** |
| **Female** | **Frequency** | 148.00 | 45.00 | 193.00 |
| **Percent** | 42.65 | 12.97 | 55.62 |
| **Row Pct** | 76.68 | 23.32 |  |
| **Col Pct** | 63.25 | 39.82 |  |
| **Male** | **Frequency** | 86.00 | 68.00 | 154.00 |
| **Percent** | 24.78 | 19.60 | 44.38 |
| **Row Pct** | 55.84 | 44.16 |  |
| **Col Pct** | 36.75 | 60.18 |  |
| **Total** | **Frequency** | 234.00 | 113.00 | 347.00 |
| **Percent** | 67.44 | 32.56 | 100.00 |

The odds ratio is 2.6 with a 95% confidence interval of 1.6 - 4.1. Since the confidence interval does not include 1, the odds ratio indicates that there is a significant relationship between the variables Flu and Gender.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table of Flu Cases by Income Level** | | | | |
| **Income Level** | | **Flu** | | **Total** |
| **No** | **Yes** |
| **High** | **Frequency** | 34.00 | 9.00 | 43.00 |
| **Percent** | 9.80 | 2.59 | 12.39 |
| **Row Pct** | 79.07 | 20.93 |  |
| **Col Pct** | 14.53 | 7.96 |  |
| **Medium** | **Frequency** | 130.00 | 80.00 | 210.00 |
| **Percent** | 37.46 | 23.05 | 60.52 |
| **Row Pct** | 61.90 | 38.10 |  |
| **Col Pct** | 55.56 | 70.80 |  |
| **Low** | **Frequency** | 70.00 | 24.00 | 94.00 |
| **Percent** | 20.17 | 6.92 | 27.09 |
| **Row Pct** | 74.47 | 25.53 |  |
| **Col Pct** | 29.91 | 21.24 |  |
| **Total** | **Frequency** | 234.00 | 113.00 | 347.00 |
| **Percent** | 67.44 | 32.56 | 100.00 |

Since income level is an ordinal variable and flu is a binary variable, the Mantel-Haenszel statistic is used to test for association. The Mantel-Haenszel statistic has a p-value of 0.76, meaning that there is not sufficient evidence to indicate a relationship between flu cases and income level. This can also be seen in the table above, which shows that the Medium income level has the highest row percentage of flu cases.

The Spearman Correlation coefficient between Flu and Income is 0.0263, meaning that even if there is an association between income level and flu it is only a very weak one.

## Testing for Confounding Between Gender and Income

[Insert Text]

…

…

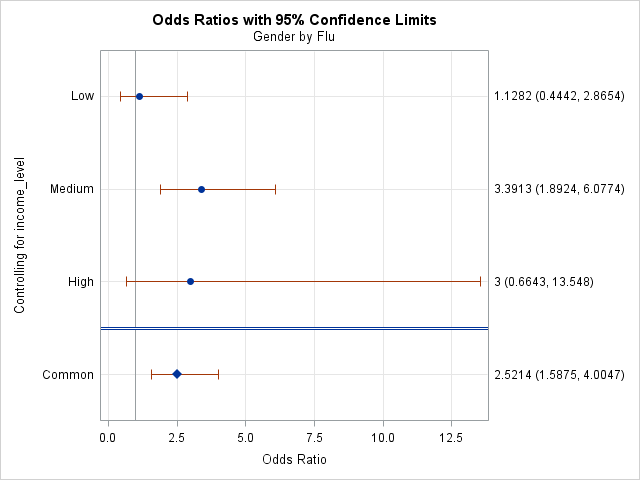
…

…

When controlling for Income, there is evidence of confounding because the odds ratio for gender when controlling for low income is outside of the original confidence limits.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Controlling for Income = Low** | | | | |
|  | | **Flu** | | **Total** |
| **No** | **Yes** |
| **Gender** |  | 40 | 13 | 53 |
| **Female** | **Frequency** |
| **Row Pct** | 75.47 | 24.53 |  |
| **Male** | **Frequency** | 30 | 11 | 41 |
| **Row Pct** | 73.17 | 26.83 |  |
| **Total** | **Frequency** | 70 | 24 | 94 |
| **Percent** | 74.47 | 25.53 | 100.00 |
| **Adjusted Odds Ratio** | | **Value** | **95% Confidence Limits** | |
| 1.1282 | 0.4442 | 2.8654 |
| **Controlling for Income = Medium** | | | | |
|  | | **Flu** | | **Total** |
| **No** | **Yes** |
| **Gender** |  | 84 | 28 | 112 |
| **Female** | **Frequency** |
| **Row Pct** | 75.00 | 25.00 |  |
| **Male** | **Frequency** | 46 | 52 | 98 |
| **Row Pct** | 46.94 | 53.06 |  |
| **Total** | **Frequency** | 130 | 80 | 210 |
| **Percent** | 61.90 | 38.10 | 100.00 |
| **Adjusted Odds Ratio** | | **Value** | **95% Confidence Limits** | |
| 3.3913 | 1.8924 | 6.0774 |
| **Controlling for Income = High** | | | | |
|  | | **Flu** | | **Total** |
| **No** | **Yes** |
| **Gender** |  | 24 | 4 | 28 |
| **Female** | **Frequency** |
| **Row Pct** | 85.71 | 14.29 |  |
| **Male** | **Frequency** | 10 | 5 | 15 |
| **Row Pct** | 66.67 | 33.33 |  |
| **Total** | **Frequency** | 34 | 9 | 43 |
| **Percent** | 79.07 | 20.93 | 100.00 |
| **Adjusted Odds Ratio** | | **Value** | **95% Confidence Limits** | |
| 3.0000 | 0.6643 | 13.5477 |

[Text]



## Testing for Interactions Between Gender and Income

[Text]

| **Tests for Homogeneity of Odds Ratios** | |
| --- | --- |
| **Breslow-Day-Tarone Chi-Square** | 3.9681 |
| DF | 2 |
| Pr > ChiSq | 0.1375 |
|  |  |
| **Zelen's Exact Test (P)** | 0.0099 |
| Exact Pr <= P | 0.1380 |

Based on the results of the calculations of the Breslow-Day-Tarone statistic and the Zelen’s exact test, there is not significant evidence of an interaction between income and gender.

## Predictive Model Using Logistic Regression

[Text]

*Note: Note sure what alpha we want to use in this section. I don’t think it changes the result, be we are asked to state it*

| **Testing Global Null Hypothesis: BETA=0** | | | |
| --- | --- | --- | --- |
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| Likelihood Ratio | 32.8588 | 10 | 0.0003 |
| Score | 31.4880 | 10 | 0.0005 |
| Wald | 28.0824 | 10 | 0.0018 |

According to the Likelihood Ratio test, the model containing all variables is overall statistically significant.

| **Type 3 Analysis of Effects** | | | |
| --- | --- | --- | --- |
| **Effect** | **DF** | **Wald Chi-Square** | **Pr > ChiSq** |
| Gender | 1 | 14.4226 | 0.0001 |
| Age | 1 | 0.9601 | 0.3272 |
| Distance | 1 | 0.0155 | 0.9011 |
| Income | 2 | 5.8659 | 0.0532 |
| Previous | 1 | 0.2345 | 0.6282 |
| Race | 3 | 6.0396 | 0.1097 |
| Visits | 1 | 0.2966 | 0.5860 |

According the Analysis of Effects table, Gender is the only statistically significant predictor variable.

Variables Gender and Income are selected for the final model because Gender is statistically significant and there is evidence of confounding between Gender and Income.

| **Type 3 Analysis of Effects** | | | |
| --- | --- | --- | --- |
| **Effect** | **DF** | **Wald Chi-Square** | **Pr > ChiSq** |
| Gender | 1 | 15.4844 | <.0001 |
| Income | 2 | 6.5806 | 0.0372 |

*\*\*Note: When other variables are dropped from the model, the p-value for income is now less than 0.05, but based on Peter’s comment I don’t know if that’s the right alpha.*

| **Testing Global Null Hypothesis: BETA=0** | | | |
| --- | --- | --- | --- |
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| Likelihood Ratio | 23.7876 | 3 | <.0001 |
| Score | 23.2552 | 3 | <.0001 |
| Wald | 22.1157 | 3 | <.0001 |

According to the Likelihood Ratio test, the final model containing Gender and Income as predictor variables is overall statistically significant.

## Summary of Findings

[Text]