Practice: Using PROC FREQ to Perform Tests and Measures of Association

The insurance company wants to determine whether a vehicle's safety score is associated with either the region in which it was manufactured or the vehicle's size. The **stat1.safety** data set contains the data about vehicle safety.

1. Use PROC FREQ to create the crosstabulation of the variables **Region** by **Unsafe**. Along with the default output, generate the expected frequencies, the chi-square test of association, and the odds ratio. To clearly identify the values of **Unsafe**, create and apply a temporary format. Submit the code and view the results.

```
/*st107s01.sas*/  /*Part B*/
proc format;
   value safefmt 0='Average or Above'
                 1='Below Average';
run;

proc freq data=STAT1.safety;
   tables Region*Unsafe / expected chisq relrisk;
   format Unsafe safefmt.;
   title "Association between Unsafe and Region";
run;
```

   Here are the results.

2. For the cars made in Asia, what percentage had a Below Average safety score?

   Region is a row variable, so look at the Row Pct value in the *Below Average* cell of the *Asia* row. Of the cars made in Asia, 42.86% have a Below Average safety score.

3. For the cars with an Average or Above safety score, what percentage was made in North America?

   Look at the Col Pct value in the *Average or Above* cell of the *N America* row. Of the cars with an Average or Above safety score, 69.70% were made in North America.

4. Do you see a statistically significant (at the 0.05 level) association between **Region** and **Unsafe**?

   The association is not statistically significant at the 0.05 alpha level. The *p*-value is 0.0631.

5. What does the odds ratio compare? What does this suggest about the difference in odds between Asian and North American cars?

   The odds ratio compares the odds of Below Average safety for North America versus Asia. The odds ratio of 0.4348 means that cars made in North America have 56.52% lower odds for being unsafe than cars made in Asia.

   Note: Recall that the odds ratios in the Estimates of Relative Risk table are calculated by comparing row1/row2 for column1. In this problem, this comparison is Asia to N America and the outcome is Average or Above in safety. The value 0.4348 is interpreted as the odds of an Average or Above car made in Asia is 0.4348 times the odds for American-made cars. If you want to compare N America

to Asia, still using Average or Above for safety, the odds ratio would be the inverse of 0.4348, or approximately 2.3. This is interpreted as cars made in North America have 2.3 times the odds for being safe than cars made in Asia. This single inversion would also create the odds ratio for comparing Asia to N America but Below Average in safety. If you want to compare N America to Asia using Below Average in safety, you invert your odds ratio twice and return to the value 0.4348.

6. Write another PROC FREQ step to create the crosstabulation of the variables **Size** and **Unsafe**. Along with the default output, generate the measures of ordinal association. Format the values of **Unsafe**. Submit the code and view the results.

```
/*st107s01.sas*/  /*Part C*/
proc freq data=STAT1.safety;
   tables Size*Unsafe / chisq measures cl;
   format Unsafe safefmt.;
   title "Association between Unsafe and Size";
run;
```

Here are the results.

7. What statistic do you use to detect an ordinal association between **Size** and **Unsafe**?

The Mantel-Haenszel chi-square detects an ordinal association.

8. Do you reject or fail to reject the null hypothesis at the 0.05 level?

You reject the null hypothesis at the 0.05 level.

9. What is the strength of the ordinal association between **Size** and **Unsafe**?

The Spearman Correlation is -0.5425.

10. What is the 95% confidence interval around the statistic that measures the strength of the ordinal association?

The confidence interval is (-0.6932, -0.3917).

Hide Solution