Print

Demo: Performing a Pearson Chi-Square Test of Association Using PROC FREQ

Filename: **st107d02.sas**

We know that there are possible associations between the binary response, Bonus, and the categorical predictors, Lot_Shape_2 and Fireplaces. Now let's run a formal test to determine whether the associations are significant.

PROC FREQ DATA=*SAS-data-set*;
TABLES *table-request(s)* < / *options*>;
< additional statements >
RUN;

1. Open program st107d02.sas.

```
/*st107d02.sas*/
ods graphics off;
proc freq data=STAT1.ameshousing3;
    tables (Lot_Shape_2 Fireplaces)*Bonus
          / chisq expected cellchi2 nocol nopercent
            relrisk;
    format Bonus bonusfmt.;
    title 'Associations with Bonus';
run;

ods graphics on;
```

   This PROC FREQ step requests a crosstabulation table for Lot_Shape_2 by Bonus and Fireplaces by Bonus. Notice the grouping syntax used here with parentheses around the predictor variables. This is just another way to request the crosstabulation tables. After the forward slash, the CHISQ option produces the Pearson chi-square test of association and the measures of association that are based on the chi-square statistic, as well. We also have some additional options related to measures of association. The EXPECTED option prints the expected cell counts, which are the cell counts we expect under the null hypothesis of no association. CELLCHI2 prints each cell's contribution to the total chi-square statistic. NOCOL suppresses the printing of the column percentages and NOPERCENT suppresses the printing of the cell percentages. Finally, we'll add the RELRISK, the relative risk option to print a table that contains risk ratios, or probability ratios, and the odds ratios.

2. Submit this program.

3. Review the output.

   The first cross-tabular frequency table shows the crosstabulation table for Lot_Shape_2 by Bonus. You can see how the options in the TABLES statement changed the statistics that appear in each cell. The actual frequency appears first. It seems that the cell for Lot_Shape_2, Irregular and Bonus, Bonus Eligible contributes the most to the chi-square statistic, with a Cell Chi-Square value of 21.905.

   Next is the table that shows the chi-square test and Cramer's V. Because the p-value for the Chi-Square statistic is less than .0001, you reject the null hypothesis at the 0.05 level and conclude that

there is evidence of an association between Lot_Shape_2 and Bonus. The Cramer's V value of -0.3531 indicates that the association detected with the chi-square test is relatively weak.

Exact tests are often useful when there are low cell counts. The chi-square test typically requires 20-25 total observations for a 2*2 table, with 80% of the table cells having counts greater than 5. In our case, we've met the requirements for the Bonus by Lot_Shape_2 crosstabulation. However, the next table, Fisher's Exact Test is provided by PROC FREQ when tests of association are requested for 2*2 tables, by default. Otherwise, the exact test must be requested using an EXACT statement.

In the Relative Risk Estimates table, the Odds Ratio and Relative Risk values show a measure of association strength. The Odds Ratio is shown in the first row of the table, along with the 95% confidence limits. To interpret the odds ratio, refer to the crosstabulation table at the beginning of the output. The top row (Irregular) is the numerator of the ratio, while the bottom row (Regular) is the denominator. The interpretation is stated in relation to the left column of the crosstabulation table (Not Bonus Eligible).

The value of 0.1347 says that an irregular lot has about 13.5% of the odds of not being bonus eligible, compared with a regular lot. This is equivalent to saying that a regular lot has about 13.5% of the odds of being bonus eligible, compared with an irregular lot. We can interpret the reciprocal of the odds ratio, 1/0.1347=7.423 similarly. The odds of being bonus eligible are more than seven times the odds for homes with irregular lot shapes than regular lot shapes.

It's often easier to report odds ratios by first transforming the decimal value to a percent-difference value. The formula for doing that is (Odds Ratio - 1) * 100. In other words, regular lots have 86.53% lower odds of being bonus eligible compared with irregular lots.

The 95% odds ratio confidence interval goes from 0.0664 to 0.2735, which doesn't include 1. This confirms the statistically significant result of the Pearson chi-square test of association. A confidence interval that includes the value of 1 would indicate equality of odds and would not be a significant result.

Relative risk estimates for each column are interpreted as probability ratios, rather than odds ratios. You have a choice of assessing probabilities of the left column (Column 1) or the right column (Column 2). For example, Relative Risk (Column 1) shows the ratio of the probabilities of irregular lots to regular lots being in the left column (66.67/93.69=0.7116).

The last two tables show the output from the Fireplaces by Bonus analysis. The output from the Chi-Square Tests show that there is also a statistically significant association between Fireplaces and Bonus, with a significant chi-square test statistic of 15.41 and a p-value 0.0004. However, Cramer's V for that association is 0.2267, indicating a relatively weak association.

Close