Lab 13: Two-way tables
STT 421: Summer, 2004
Vince Melfi

We've already used **proc freq** to help in analyzing the output of simulations. Today we'll see that **proc freq** also performs tests for two-way tables as described in Chapter 9 of the text.

13.1 Chi-square tests

Exercise 9.31 in the text presents data collected to determine whether contacting people by phone or letter before sending them a survey will increase the response rate. Specifically, one group of people received a letter before getting the survey; one group received a phone call before receiving the survey; and one group didn't receive any information before the survey arrived. For this study, a response was defined as returning the survey within 2 weeks. The data are contained in the file `U:\msu\course\stt\421\summer04\survey.dat`. For reference, the contents of the file are given next.

```
01 yes letter 171
02 no  letter 220
03 yes phone  146
04 no  phone   68
05 yes none   118
06 no  none   455
```

First comes the record number (1 through 6), then response status (yes or no) then contact method (letter, phone, none), then the number in the group. For example, there were 171 people who responded and received a preliminary letter; there were 220 people who didn't respond and received a preliminary letter.

The following SAS code reads in the data, drops the useless variable `record` and prints the raw data.

```
data survey;
  infile 'U:\msu\course\stt\421\summer04\survey.dat';
  input record response $ contact $ number;
  drop record;

proc print data = survey;

run;
```

Since the data contains the numbers in each group rather than a separate observation for each subject, we'll need to use the **weight** statement in **proc freq**. For example, the following SAS code will compute the frequencies for each variable `response` and `contact`.

```
proc freq data = survey;
  weight number;
```

```
run;
```

1. What proportion of subjects responded?

2. What proportion of subjects received a phone call?

We want to create a two-way table and see whether contacting the subject before the survey arrives is related to whether the subject responds. To create the table we use the **tables** statement. To peform the chi-square test of association we use the **chisq** option.

```
proc freq data = survey;
  weight number;
  tables contact * response / chisq;

run;
```

13.1.1 Explanation and questions

1. Each cell of the two-way table contains the count in the cell, the percentage of the whole dataset in the cell, the percentage of that row's data in the cell, and the percentage of that column's data in the cell. For example, the upper left cell represents those who received a letter but didn't respond. The number 220 tells us that there were 220 people in this category. The percentage 18.68 tells us that these 220 people are 18.68% of all people in the study. The percentage 56.27 tells us that 56.27% of those who received a letter didn't respond. The percentage 29.61 tells us that 29.61% of those who didn't respond received a letter.

2. What percentage of those who were not contacted responded? What proportion of those who were contacted by phone responded? Does it seem that contacting the subjects increases response?

3. After the table comes the output from the statistical tests. We'll concentrate on the first line labeled "Chi-Square." The other lines present other statistical methods for analyzing these data. What is the value of the Chi-Square statistic? Would you reject the null hypothesis of independence at the level $\alpha = 0.01$?

13.2 Confidence intervals for proportions

It's also possible to get confidence intervals for proportions using **proc freq**. We'll do this in the context of Exercise 8.19. This exercise reports the results of a marathon coin-tossing experiment performed by a mathematician named John Kerrich. He tossed a coin 10000 times and obtained 5067 heads. We'll get a 95% confidence interval for the probability that the coin comes up HEADS using the **binomial** and **alpha** options.

```
data coin;
  input result $ count;
  cards;
  HEADS 5067
  TAILS 4933
  ;

proc freq data = coin;
  weight count;
  tables result / binomial alpha = .05;

run;
```

You should see a freqeuency table for the result (HEADS or TAILS) and then various statistics related to the proportion of HEADS. We'll only concentrate on the first set of statistics, which give the proportion of heads, the standard error (called ASE), and the lower and upper confidence limits.

1. Compute the 95% confidence interval by hand, and verify that you get the same result as reported in SAS.