Introduction to Statistics (MAT/SST 115.03 2008S)

Class 32: Topic 25: Inference for Two-Way Tables

Held: Friday, 18 April 2008

Summary: We consider yet another way in which we use samples to explore populations. We continue to emphasize categorical variables, this time considering pairs of categorical variables. We look at a more general version of the chi-square test from the previous topic.

Notes:

- Katherine will miss class today and Monday.
- I spent until midnight last night getting today's classes semi-prepared. My 151 students claim that you'll understand that the preparation meant that I did not get your exams graded.
- Extra credit for any *one* Pride Week event.
- Is anyone in Symphonic band? If so, EC for Sunday's concert.
- Due Monday: Mini-Project Memos.

- Handouts: R Notes for Topic 25.
- Due: 24-6, 24-7, 24-9, 24-15, 24-20.

Overview:

- Presentation Debriefing.
- Today's Topic: Inferences from Samples for Categorical Variables, Continued.
- Some R.

Debriefing on the Presentation

- What did you see as strengths?
- What did you see as weaknesses?
- What errors did you note?
  - I apologize that correcting errors may embarrass some of you. However, I think that it's important that you hear about errors.

What Samples Tell Us About Categorical Variables, Continued

- Context: Why do we call the topics of the first few weeks *descriptive statistics* and the more recent topics *inferential startistics*?
- Topic 21: Two populations, one binary categorical variable.
- Topic 24: One population, one non-binary categorical variable.
- Topic 25: One population, two categorical variables (binary or non-binary).
  - Can you think about the values for the explanatory variable as representing different populations?

- What might your null and alternate hypothesis be?

- Just as in topic 24, we answer questions about populations using a chi-square test.
  - That is, we compute expected values, subtract them from observed, square the difference, divide by the expected, and sum all of those values.

Doing Chi-Square Computations in R

So, how do we do a chi-squre computation in R? It ends up being fairly straightforward.

We start by making a table or data frame from our data.

```
> SC = fys[,c(1,3)]
> table(SC)
   CHOICE
SEX   1   2   3   4
  1   4  11  34 108
  2  15   7  35 127
> SCframe = data.frame(row.names=c("Below Third","Third","Second","First"),
+   Male=SCtable[1,],
+   Female=SCtable[2,])
> SCframe
            Male Female
Below Third    4     15
Third         11      7
Second        34     35
First        108    127
```

We can simple apply the `chisq.test` procedure to this table to get the important values.

```
> chisq.test(SCframe)
        Pearson's Chi-squared test
data:  SCframe
X-squared = 6.7122, df = 3, p-value = 0.08166
```

However, we will often want to look more carefuly at the differences between observed and expected values. The computation of the expected values is a strange formula that I don't expect you to understand. (I do expect that you could do an individual computation by hand, but this does all of them at once.)

```
> SCexpected = rowSums(SCframe) %o% colSums(SCframe)/sum(SCframe)
> SCexpected
                 Male    Female
Below Third   8.74780  10.25220
Third         8.28739   9.71261
Second       31.76833  37.23167
First       108.19648 126.80352
```

We can now compare directly.

```
> SCframe - SCexpected
                  Male     Female
Below Third -4.7478006  4.7478006
Third        2.7126100 -2.7126100
Second       2.2316716 -2.2316716
First       -0.1964809  0.1964809
```

As importantly, we can compute the deviations.

```
> (SCframe-SCexpected)^2/SCexpected
                    Male       Female
Below Third 2.5768317632 2.1987097110
Third       0.8878854292 0.7575978934
Second      0.1567711671 0.1337667023
First       0.0003568024 0.0003044455
```
