

22S:105
Statistical Methods and Computing

**Contingency Tables and the
 Chi-Square Test
 Introduction to ANOVA**

Lecture 21
 Apr. 14, 2008

Kate Cowles
 374 SH, 335-0727
 kcowles@stat.uiowa.edu

- death certificate incorrect and required re-coding of underlying cause of death
- Question of interest: Are there differences between the two hospitals with respect to practices in completing death certificates
- One way to address the question: Test null hypothesis that, within each category of death certificate status, the proportions of death certificates coming from Hospital A are the same.

The Chi-square test for differences among more than 2 proportions

We are interested in the *independent samples* case.

Example:

- A study investigated the accuracy of death certificates by comparing the results of 575 autopsies to the causes of death listed on the certificates.
- Two hospitals participated in the study.
 - community hospital, labeled A
 - university ospital, labeled B
- Three possible cases
 - death certificate confirmed accurate
 - death certificate contained inaccuracies but did not require correction of underlying cause of death

Another multiple comparisons problem!

$$H_0 : p_c = p_i = p_r$$

$$H_a : p_c \neq p_i \text{ or } p_c \neq p_r \text{ or } p_i \neq p_r$$

- We will *first* test whether there are *any* significant differences.
- Only if we reject H_0 in the overall test will we do pairwise tests to find out *which* population proportions are different.

Results

	Hospital A	Hospital B	Total
Confirmed accurate	157	268	425
Inacc, no recoding	18	44	62
Incorrect, recoding	54	34	88
Total	229	346	575

The overall sample proportion of death certificates from hospital A is

$$\frac{229}{575} = 0.398$$

If H_0 is true, we would expect this same proportion of hospital A certificates in all three categories.

According to Table E, the .05 cutoff under a Chi-square distribution with 2 d.f. is 5.99.

We can reject H_0 because $21.62 > 5.99$. The p -value < 0.001 .

We conclude that the proportions of death certificates from Hospital A are not the same for the three different categories of certificate status.

Observed and expected counts

	Hospital A	Hospital B	Hospital A	Hospital B
Accurate	157	268	169.3	255.7
Incorrect	18	44	24.7	37.3
Recode	54	34	35.0	53.0

The Chi-square statistic is

$$X^2 = 21.62$$

- $r = 3$ rows
- $c = 2$ columns
- So the degrees of freedom is $(r - 1)(c - 1) = 2(1) = 2$

This Chi-square test in SAS

```
options linesize = 72 ;

data dthcert ;
input hosp $ status $ count ;
datalines ;
A C 157
A I 18
A R 54
B C 268
B I 44
B R 34
;

proc freq data = dthcert ;
tables status * hosp / expected ;
weight count ;
run ;

proc freq data = dthcert ;
tables status * hosp / chisq ;
weight count ;
run ;
```

TABLE OF STATUS BY HOSP

STATUS	HOSP		Total
	A	B	
Frequency			
Expected			
Percent			
Row Pct			
Col Pct	A	B	Total
-----+-----+-----+			
C	157	268	425
	169.26	255.74	
	27.30	46.61	73.91
	36.94	63.06	
	68.56	77.46	
-----+-----+-----+			
I	18	44	62
	24.692	37.308	
	3.13	7.65	10.78
	29.03	70.97	
	7.86	12.72	
-----+-----+-----+			
R	54	34	88
	35.047	52.953	
	9.39	5.91	15.30
	61.36	38.64	
	23.58	9.83	
-----+-----+-----+			

Total	229	346	575
	39.83	60.17	100.00

TABLE OF STATUS BY HOSP

STATUS	HOSP		Total
	A	B	
Frequency			
Percent			
Row Pct			
Col Pct	A	B	Total
-----+-----+-----+			
C	157	268	425
	27.30	46.61	73.91
	36.94	63.06	
	68.56	77.46	
-----+-----+-----+			
I	18	44	62
	3.13	7.65	10.78
	29.03	70.97	
	7.86	12.72	
-----+-----+-----+			
R	54	34	88
	9.39	5.91	15.30
	61.36	38.64	
	23.58	9.83	
-----+-----+-----+			
Total	229	346	575
	39.83	60.17	100.00

STATISTICS FOR TABLE OF STATUS BY HOSP

Statistic	DF	Value	Prob
Chi-Square	2	21.523	0.001
Likelihood Ratio Chi-Square	2	21.189	0.001
Mantel-Haenszel Chi-Square	1	12.864	0.001
Phi Coefficient		0.193	
Contingency Coefficient		0.190	
Cramer's V		0.193	

Sample Size = 575