

BIOS 6244 Analysis of Categorical Data
October 10, 2005
Computer Lab

Comparing Two Binomial Proportions

Using PROC FREQ in SAS (Stokes et al., pp. 29-38)

PROC FREQ can be used to perform many of the statistical procedures for categorical data that we will cover in this course.

The following SAS code (available on the course website) creates a 2x2 table for the data from our example that examined the association between exposure to certain chemicals and occurrence of a particular tumor among lab animals.

```
data cancer;  
input chemical tumor count @@;  
cards;  
1 1 37 1 2 40  
2 1 212 2 2 138  
;
```

To calculate an approximate p-value for the test of $H_0: \pi_1 - \pi_2 = 0$ and to find an approximate 95% CI ($\pi_1 - \pi_2$), use the following code:

```
proc freq order=data; weight count;  
  tables chemical*tumor / chisq riskdiff;  
  title 'Carcinogenicity Example';  
  title2 'Difference in Proportions';  
run;
```

The relevant output produced by this code is as follows:

Statistics for Table of chemical by tumor

Statistic	DF	Value	Prob	
Chi-Square	1	4.0695	0.0437	(This is the p-value.)
Likelihood Ratio Chi-Square	1	4.0207	0.0449	
Continuity Adj. Chi-Square	1	3.5708	0.0588	
Mantel-Haenszel Chi-Square	1	4.0600	0.0439	
Phi Coefficient		-0.0976		
Contingency Coefficient		0.0972		
Cramer's V		-0.0976		

Column 1 Risk Estimates

	Risk	ASE	(Asymptotic) 95% Confidence Limits		(Exact) 95% Confidence Limits	
Row 1	0.4805	0.0569	0.3689	0.5921	0.3652	0.5974
Row 2	0.6057	0.0261	0.5545	0.6569	0.5524	0.6573
Total	0.5831	0.0239	0.5364	0.6299	0.5348	0.6304
Difference	-0.1252	0.0626	-0.2480	-0.0024	[This is the CI($\pi_1 - \pi_2$).]	

Difference is (Row 1 - Row 2)

To find an approximate 95% CI $\left(\frac{\pi_1}{\pi_2} \right)$, use the following SAS code:

```
proc freq order=data; weight count;
  tables chemical*tumor / relrisk;
  title 'Carcinogenicity Example';
  title2 'Relative Risk';
run;
```

The relevant output is as follows:

```

Statistics for Table of chemical by tumor

Estimates of the Relative Risk (Row1/Row2)

Type of Study          Value          95% Confidence Limits
-----
Case-Control (Odds Ratio)  0.6021          0.3667          0.9886
Cohort (Col1 Risk)       0.7933          0.6196          1.0157 [This is the estimate of RR & a 95% CI(RR).]
Cohort (Col2 Risk)       1.3175          1.0250          1.6935

Sample Size = 427

```

The current version of SAS does not provide exact confidence intervals for the difference or the ratio of two proportions. A statistical package called StatXact, which is available on the research computer in Room 1124 of the UNO Tec Building, can provide these analyses. The following table contains a comparison of the approximate and asymptotic results for our example.

Parameter (θ)	Approximate 95% CI(θ)	Exact 95% CI(θ)
Risk difference	(-.248, -.002)	(-.248, -.001)
Relative risk	(.620, 1.016)	(.604, .998)

Note that the exact procedure gives a statistically significant result for the RR, whereas the approximate method does not.

Using SPSS

The statistical package SPSS can also be used to find a p-value for the risk difference and to find a p-value and CI for the relative risk. Screen shots for doing this are given on the course website, along with the SPSS database, syntax, and output files for the classroom example.