

Stata Program Notes

Biostatistics: A Guide to Design, Analysis, and Discovery Second Edition

by Ronald N. Forthofer, Eun Sul Lee, Mike Hernandez

Chapter 10: Analysis of Categorical Data

Program Note 10.1 – Chi-square Test for a 2 by 2 Contingency Table

In Example 10.3, we wish to analyze the education and iron status data presented in Table 10.6 to determine if there was a statistically significant relationship between these variables. The **tabulate** command with the **chi2** option was used to test the null hypothesis of no relationship. The test statistic used was the uncorrected Pearson chi-square statistic. We also specified the **exact** option to obtain a p-value from Fisher's exact test. Because the data are already presented in a 2 by 2 table, we used the command **tabi** which is the command required for immediate entry of the table's cell frequencies. Notice that cell frequencies along rows are separated by a backslash (“\”). So, the cell frequencies: 4 and 26 correspond to values on the first row, and 4 and 66 correspond to values on the second row.

Stata Command:

```
tabi 4 26 \ 4 66, chi2 exact
```

Stata Output:

row	col	Total	
	1	2	
1	4	26	30
2	4	66	70
Total	8	92	100

Pearson chi2(1) = 1.6563 Pr = 0.198
Fisher's exact = 0.236
1-sided Fisher's exact = 0.185

Program Note 10.2 – Chi-square Test for an r by c Contingency Table

In Example 10.7, the (2 by 3) contingency table resulted from the cross-tabulation of whether a woman knew someone with breast cancer with her opinion about a mammography. Notice here that there are three values in the first row followed by the “\” and the three values in the second row.

Stata Command:

```
tabi 120 45 28 \ 77 15 8
```

Stata Output:

row	1	2	3	Total
1	120	45	28	193
2	77	15	8	100
Total	197	60	36	293

Pearson chi2(2) = 6.6479 Pr = 0.036

Program Note 10.3 – Trend Test

Stata has an **epitab** package that provides for analyses of epidemiological data. Stata provides a score test for trends by using the Stata command **tabodds** which is a component of the **epitab** package. In Example 10.8, a test of the null hypothesis of no linear trend in the opinion about mammography is desired. The alternative hypothesis is that there is a linear trend.

Program Note 10.4 – The Mantel-Haenszel Odds Ratio

Stata provides a Mantel-Haenszel odds ratio that can be computed with the command **mhodds** which is also a part of Stata's **epitab** package.