BIOS 6244 Analysis of Categorical Data
Assignment 4 Solutions

1. Work Exercise 3.8, p. 68. Perform the CMH test by hand and test the hypothesis of conditional independence using the approximate χ^2 test. Interpret the results in the context of the applied problem. Please show all your work.

   Solution

   To perform the CMH test of conditional independence, we must first calculate μ_{iik} & $\text{Var}(n_{11k})$ for each layer. These values are calculated as follows:

$$\mu_{111} = \frac{(21+82)(21+73)}{21+82+73+188} = \frac{(103)(94)}{364} = 26.599$$

$$\text{Var}(n_{111}) = \frac{(103)(94)(73+188)(82+188)}{(364)^2(364-1)} = \frac{682290540}{48096048} = 14.186$$

$$\mu_{112} = \frac{(5+16)(5+19)}{5+16+19+38} = \frac{(21)(24)}{78} = 6.462$$

$$\text{Var}(n_{112}) = \frac{(21)(24)(19+38)(16+38)}{(78)^2(78-1)} = \frac{1551312}{468468} = 3.311$$

$$\mu_{113} = \frac{(71+249)(71+137)}{71+249+137+363} = \frac{(320)(208)}{820} = 81.171$$

$$\text{Var}(n_{113}) = \frac{(320)(208)(137+363)(249+363)}{(820)^2(820-1)} = \frac{20367360000}{550695600} = 36.985$$

Thus, $\sum_k n_{11k} = 97$, $\sum_k \mu_{11k} = 114.232$, $\sum_k \text{Var}(n_{11k}) = 54.482$

The CMH test statistic is given by

$$\text{CMH} = \frac{\left(\sum_k n_{11k} - \sum_k \mu_{11k}\right)^2}{\sum_k \text{Var}(n_{11k})} = \frac{(97-114.232)^2}{54.482} = \frac{296.942}{54.482} = 5.450\text{, df = 1.}$$

The approximate p-value using χ^2 is given by Pr(CMH ≥ 5.45 | df = 1) = .020 by SimCalc. Therefore, based on the CMH test, we conclude that we do not have conditional independence. In terms of the applied problem, our conclusion is that there is a significant association between passive smoking and lung cancer among

non-smoking women married to smokers, after controlling for the effect of country.

2. Consider Exercise 3.9, p. 68. Write the SAS code, including the DATA step, to perform the approximate CMH test and the approximate test for homogeneous association (with adjustment), and to calculate an approximate 95% CI for the true overall conditional odds ratio.

   Solution

   The following SAS code will perform the requested analyses:

```
data lung_ca;
input country passive case count @@;
cards;
1 1 1 21  1 1 2 82   1 2 1 73  1 2 2 188
2 1 1 5   2 1 2 16   2 2 1 19  2 2 2 38
3 1 1 71  3 1 2 249  3 2 1 137 3 2 2 363
;

proc freq order = data; weight count;
   tables  country * passive * case / nocol nopct cmh bdt relrisk
   chisq;
   exact or chisq ;
   title 'Exercise 3.8';
   title2 'Cochran-Mantel-Haenszel Analyses';
   run;
```

   Note: The RELRISK option in the TABLES statement and the EXACT statement are not required to perform the analyses requested in Exercise 2. However, they ARE required to generate the output in Exercise 3 below.

3. Use the SAS output below to answer the following questions. When giving the results for an hypothesis test, be sure to provide the value of the test statistic, the d.f. (if any), and the p-value. **Be sure to indicate on the printout where you obtained your answers.**

   (a) Perform the exact version of the χ^2 test for independence and calculate an exact 95% CI(OR) for each of the conditional odds ratios.

   Solution

   The following SAS output provided with the assignment is used to answer this question:

Statistics for Table 1 of passive by case
Controlling for country=1

Pearson Chi-Square Test

Chi-Square		2.2159
DF		1
Exact	Pr >= ChiSq	0.1458

Odds Ratio (Case-Control Study)

Odds Ratio	0.6595

Exact Conf Limits

95% Lower Conf Limit	0.3607
95% Upper Conf Limit	1.1721

Statistics for Table 2 of passive by case
Controlling for country=2

Pearson Chi-Square Test

Chi-Square		0.6534
DF		1
Exact	Pr >= ChiSq	0.5816

Odds Ratio (Case-Control Study)

Odds Ratio	0.6250

Exact Conf Limits

95% Lower Conf Limit	0.1560
95% Upper Conf Limit	2.1643

Statistics for Table 3 of passive by case
Controlling for country=3

Pearson Chi-Square Test

Chi-Square		2.8003
DF		1
Exact	Pr >= ChiSq	0.1003

Odds Ratio (Case-Control Study)

Odds Ratio	0.7555

Exact Conf Limits

95% Lower Conf Limit	0.5351
95% Upper Conf Limit	1.0618

These results are summarized in the table on the following page:

	Exact χ^2 Test				
Country	**X^2**	**Df**	**p-value**	**OR**	**Exact 95% CI(OR)**
Japan	2.22	1	.146	.66	(.36, 1.17)
G.B.	0.65	1	.582	.63	(.16, 2.16)
U.S.	2.80	1	.100	.76	(.54, 1.06)

(b) Does the use of the Cochran-Mantel-Haenszel methodology appear to be justified for these data? Why or why not?

Solution

Yes, the conditional OR's were all in the same direction (negative) and of the same order of magnitude (.6 - .8), so it is appropriate to use the CMH procedure.

(c) Perform the approximate version of the CMH test.

Solution

The following SAS output provided with the assignment is used to answer this question:

Summary Statistics for passive by case
Controlling for country

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	5.4497	0.0196

Thus, CMH = 5.45, df = 1, p = .020. Thus, we reject the null hypothesis of conditional independence and conclude that the conditional OR's are not all equal to 1.

(d) Perform the test for homogeneous association with adjustment.

Solution

The following SAS output provided with the assignment is used to answer this question:

Breslow-Day-Tarone Test for Homogeneity of the Odds Ratios	
Chi-Square	0.2381
DF	2
Pr > ChiSq	0.8878

Thus, $X^2 = .24$, df = 2, p= .888. We conclude that the conditional OR's appear to be equal across layers.

(e) Calculate the M-H estimate of the overall conditional OR, along with an approximate 95% CI. Interpret the results in the context of the applied problem.

Solution

The following SAS output provided with the assignment is used to answer this question:

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95% Confidence Limits	
Case-Control	Mantel-Haenszel	0.7218	0.5489	0.9492

Therefore, $\widehat{OR}_{XY} = .72$, 95% CI(OR_{XY}) = (.55, .95). This tells us the odds of lung cancer for non-smoking women married to non-smokers are decreased by a factor of .72 when compared to non-smoking women married to smokers. The true overall conditional OR could be as small as .55 or as large as .95. All of these results are adjusted for the effect of the country in which the study was conducted.

(f) Compare the results in parts (a) and (e) above. What benefit(s) does the M-H approach provide relative to the separate analyses in each layer?

Solution

The stratified analysis in Part (a) above indicated no significant associations between passive smoking and lung cancer for any of the 3 countries individually. However, when the results are combined across

countries using the CMH approach, statistical significance is achieved since the 95% CI(OR_{XY}) = (.55, .95) does not contain 1.

(g) The marginal analysis for this problem yielded $\widehat{OR}$ = 0.72, with an exact 95% CI(OR) = (0.54, 0.95), and a p-value for the exact χ^2 test of independence = 0.018. Does it appear that "country" is a significant confounder in this study? Why or why not?

Solution

There is a negligible difference between the results of the CMH analysis and the marginal analysis in terms of the test of (conditional) independence (p = .020 vs. .018) and the 95% CI for the overall (conditional) OR [(.55, .95) vs. (.54, .95)]. Therefore, it appears that country is not a clinically significant counfounder in this study and that it is not necessary to adjust for it.