Lab Seven: Categorical data analysis II

Lab Objectives
After today’s lab you should be able to:

1. Understand the function of a SAS MACRO.
2. Write and run a simple SAS MACRO.
3. Understand the use of stratification to correct for confounding.
4. Use PROC FREQ to calculate a (Cochran-) Mantel-Haenszel adjusted odds ratio.
5. Know when to combine stratum-specific OR’s and when to report them separately (in the presence of interaction, or effect modification).
6. Use PROC LOGISTIC for multivariate logistic regression.

Recommended reading in Walker: Chapters 19-20
LAB EXERCISE STEPS:

Follow along with the computer in front...

1. Double-click on to open the SAS editor file “data creation code” which should be saved in your stats210 folder from last week; run the libname statement:

   \texttt{libname stats210 'C:\Documents and Settings\mitl-pc.LANE-LIB\My Documents\Stats210';}

2. Using the Explorer Browser on the left hand side of your screen, double check that a stats210 library has been properly created, and that it contains the SAS dataset “runners”.

3. Recall the following output from lab six:

   Table of badedi by isirreg

<table>
<thead>
<tr>
<th>badedi</th>
<th>isirreg</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>27</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>9</td>
</tr>
</tbody>
</table>
   | Total  | 30      | 17        | 47

   The odds ratio is the cross product (so table organization doesn’t matter):

   \[
   OR = \frac{27 \times 9}{8 \times 3} = 10.125
   \]

   We can calculate either an odds ratio or a risk ratio here since the data are cross-sectional.

   Estimates of the Relative Risk (Row1/Row2)

<table>
<thead>
<tr>
<th>Type of Study</th>
<th>Value</th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case-Control (Odds Ratio)</td>
<td>10.125</td>
<td>2.2004</td>
</tr>
<tr>
<td>Cohort (Col1 Risk)</td>
<td>3.0857</td>
<td>1.1392</td>
</tr>
<tr>
<td>Cohort (Col2 Risk)</td>
<td>0.3048</td>
<td>0.1527</td>
</tr>
</tbody>
</table>

   Confidence limits show highly significant increase in risk of menstrual irregularity with high EDI score. Confidence limits are calculated as:

   \[
   95\% \text{ CI for } \ln(OR) = \ln(OR) \pm 1.96 \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}
   \]

   \[
   \ln(10.125) \pm 1.96 \sqrt{\frac{1}{3} + \frac{1}{8} + \frac{1}{9} + \frac{1}{27}} = (.78, 3.84)
   \]

   \[\begin{align*}
   &e^{.78} = 2.2 \\
   &e^{3.84} = 46.5
   \end{align*}\]
4. Generalize the above program into a SAS MACRO, such that you can enter any 2x2 table and get out any size confidence interval (90%, 95%, 99%, etc.)

/**MACRO to calculate XX% confidence limits for an odds ratio from the desired confidence and the 2x2 cell sizes: a,b,c,d**/

    data _null_;    
    OR=&a.*&d./(b.*c.);    
    lnOR=log(OR);    
    error=sqrt(1/a.+1/b.+1/c.+1/d.);    
    Z=-probit((1-confidence./100)/2); *gives left hand Z score, multiply by negative;    
    lower=exp(lnOR-Z*error);    
    upper=exp(lnOR+Z*error);    
    put lower;    
    put upper;    
    run;    
%mend oddsratio;

/**Invoke MACRO with the above data and ask for 95% confidence limit**/
%oddsratio(95,9,3,8,27);

DUE FROM LAB SIX AND SEVEN: Use the above MACRO to get a 99% confidence interval from a 2x2 table with a=20, b=20, c=10, d=20. Email me the resulting values.

5. Use PROC FREQ to determine whether there is a relationship between having had a stress fracture in the past (stressf=1) and having a stress fracture during the study (sf1=1).

    proc format;    
    value sf        
    1=fracture        
    0=no fracture;    
    run;    
    proc freq data=stats210.runners;    
    format sf1 sf.;    
    format stressf sf.;    
    tables stressf*sf1/measures nocol nopercent;    
    run;
6. You might wonder if this relationship can be entirely explained by low bone strength (that is, some women have low bone strength, and these women were more likely to fracture prior to the study, and are still more likely to fracture during the study). We can stratify by low bone strength and re-evaluate the relationship as follows:

```sas
class proc rank data=stats210.runners out=stats210.runners groups=4;
    ranks lowbone; *0 is lowest, 3 is highest;
    var bmc1;
run;
data stats210.runners;
    set stats210.runners;
    if lowbone=0 then lowbone=1; else lowbone=0;
run;
proc format;
    value quart
        0=higher
        1=lowest quartile;
run;
```
Examine the relationship between low bone strength and fracture prior to the study.

Examine the relationship between low bone strength and fracture during the study.

Stratify by lowbone and re-examine the relationship between stressf and sf1. Controlling for bone strength, what is the odds ratio between stressf and sf1? Recall, the crude OR was 5.2

40% in the lowest quartile of bone strength had a fracture prior to the study vs. 29% in the other quartiles of bone strength.

OR=19x4/(8x6)=1.5

null
27% in the lowest quartile of bone strength had a fracture during the study vs. 17% in the other quartiles of bone strength. OR=30x3/(8x6)=1.8

A higher percentage of runners with prior fracture went on to fracture during the study both in the low bone strength group (table 2: OR not calculated because of 0 cell) and the higher strength group (table 1: OR=1.8), though our small sample size makes stratification difficult (small numbers result making conclusions difficult!).
Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Alternative Hypothesis</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Nonzero Correlation</td>
<td>1</td>
<td>3.7275</td>
<td>0.0535</td>
</tr>
<tr>
<td>2</td>
<td>Row Mean Scores Differ</td>
<td>1</td>
<td>3.7275</td>
<td>0.0535</td>
</tr>
<tr>
<td>3</td>
<td>General Association</td>
<td>1</td>
<td>3.7275</td>
<td>0.0535</td>
</tr>
</tbody>
</table>

Estimates of the Common Relative Risk (Row1/Row2)

<table>
<thead>
<tr>
<th>Type of Study</th>
<th>Method</th>
<th>Value</th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case-Control</td>
<td>Mantel-Haenszel (Odds Ratio) Logit **</td>
<td>4.4778</td>
<td>21.5155</td>
</tr>
<tr>
<td>Cohort</td>
<td>Mantel-Haenszel (Col1 Risk) Logit</td>
<td>1.4808</td>
<td>20.9541</td>
</tr>
<tr>
<td>Cohort</td>
<td>Mantel-Haenszel (Col2 Risk) Logit **</td>
<td>0.2771</td>
<td>1.1704</td>
</tr>
</tbody>
</table>

** These logit estimators use a correction of 0.5 in every cell of those tables that contain a zero.

TEST OF INDEPENDENCE:
Null hypothesis: stressf and sf1 are independent after controlling for bone strength.

p=.05 here indicates that there is a relationship between stressf and sf1 even after controlling for bone strength.

TEST OF HOMOGENEITY:
Null hypothesis: odds/risk ratios are the same in all strata.

If the odds/risk ratios are not homogenous, it may not be a good idea to combine them into a single summary odds/risk ratio.

Here, there may be a borderline difference (p<.10) in OR/RR’s between table 1 and table 2, but the 0 cell in table 2 makes it difficult to evaluate this hypothesis.

MEASURE OF ASSOCIATION
After controlling for bone strength, the odds ratio between prior fracture and fracture during the study is 4.4, which is not much different than the crude (unadjusted) odds ratio of 5.2. Therefore, it does not appear that low bone strength explains much of the relationship between stressf and sf1.

Breslow-Day Test for Homogeneity of the Odds Ratios

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>3.2092</td>
<td>1</td>
<td>0.0732</td>
</tr>
</tbody>
</table>

Effective Sample Size = 37
Frequency Missing = 10

WARNING: 21% of the data are missing.
7. Another interesting question would be whether or not the effect of treatment differs depending on bone strength. That is, maybe treatment is helpful in women with lower bone strength, but has no effect in women with more “normal” bone strength.

```
proc format;
  value treat
  1=treatment
  0=control;
run;

proc freq data=stats210.runners; *does the effect of treatment differ depending on bone strength?;
  format sf1 sf.;
  format lowbone quart.;
  format treatr treat.;
  tables lowbone*treatr*sf1 /measures cmh nopercent nocol;
run;
```

OUTPUT:

Table 1 of treatr by sf1
Controlling for lowbone=higher

<table>
<thead>
<tr>
<th>treatr</th>
<th>sf1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no fracture</td>
</tr>
<tr>
<td></td>
<td>frequency</td>
</tr>
<tr>
<td>control</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>86.36</td>
</tr>
<tr>
<td>treatment</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>78.57</td>
</tr>
</tbody>
</table>

OR=1.7 for treated vs. control.

Table 2 of treatr by sf1
Controlling for lowbone=lowest quartile

<table>
<thead>
<tr>
<th>treatr</th>
<th>sf1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no fracture</td>
</tr>
<tr>
<td></td>
<td>frequency</td>
</tr>
<tr>
<td>control</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>60.00</td>
</tr>
<tr>
<td>treatment</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>83.33</td>
</tr>
</tbody>
</table>

OR=0.30 for treated vs. control.
The stratum-specific OR’s were: 1.7 and 0.30, obviously not homogenous. Because of small numbers, the Breslow-Day test of homogeneity does not reject at \( p<.05 \); but common sense tells us that combining the two OR’s might skew over separate effects at different levels of bone strength. We would need a larger sample to see if the trend (of protection when bone strength is low) bears out.

Combining an OR of 1.7 and an OR of 0.30 gives us an overall OR of 1.01 \((p=.98)\), indicating no relationship between treatment and stress fractures. In fact, it may be that treatment is only protective when bone density is low (effect modification, or interaction, by bone strength). However, we would need larger numbers to test this possible interaction.

8. Introduction to Logistic Regression: Type in the following code to show that proc freq and proc logistic give us the same OR for the relationship between sf1 and treatr. What happens when you adjust for bmc using the categorical variable lowbmc? Using the continuous variable bmc1?

```sas
proc freq data=stats210.runners; *unadjusted OR for treatr and sf1;
tables sf1*treatr/measures;
run;

proc logistic descending data=stats210.runners;
    model sf1= treatr /risklimits;
run;

proc logistic data=stats210.runners;
    model sf1= treatr lowbone /risklimits;
run;

proc logistic data=stats210.runners;
    model sf1= treatr bmc1 /risklimits;
run;
```