1. The file *crash.dat* was obtained from a national database of automobile crashes. Data were selected from serious accidents in New Jersey in 1999. The data file has one line for every individual involved in the crashes. Individuals in a given car are numbered, and given in the variable *pn*. Since individuals in the same car are likely to have related outcomes, in order to obtain an set of independent observations, the code below retains only the subject labeled 1. Autos in a crash are numbered, and the number is contained in the variable *vh*. Again, in order to obtain an set of independent observations, retain only the vehicles labeled 1. The following code reads in crash data, and creates categorical variables reflecting car body type, restraint use, and whether the crash was a fatality.
data crash ; infile "crash.dat";
    input id state case vh pn sl bt v1 v2 v3 rs is;
    if vh>1 then delete; if pn>1 then delete;
    fatal="Not fatal"; if is=4 then fatal="Fatal";
    vt="Car "; if bt>10 then vt="Light truck";
    if bt>20 then vt="Van"; if bt>30 then vt="Truck";
    beltuse="Improper or no restraint use";
    beltuse="Proper restraint use";
    if rs> 12 then beltuse="Improper or no restraint use";
    if rs=0 then beltuse="Improper or no restraint use";
    keep vt fatal beltuse;
run;
This code and the crash data set can be found on the course web page.

a. Test whether the tables stratified on body type have a common odds ratio not equal to 1.

Running the command

```
proc freq data=crash; table vt*fatal*beltuse/cmh ; run;
```

gives a p-value of .001. Reject the null hypothesis of no association.

b. Estimate a common odds ratio stratified based on car body type. Give a confidence interval for this value.

From the output generated by the command in the previous answer,

```
<table>
<thead>
<tr>
<th>Type of Study Method</th>
<th>Value 95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case-Control Mantel-Haenszel</td>
<td>4.2872 3.1757 5.7877</td>
</tr>
<tr>
<td>(Odds Ratio) Logit **</td>
<td>4.4388 3.2422 6.0770</td>
</tr>
</tbody>
</table>
```

c. What assumption underlies parts (a) and (b)?

Both (a) and (b) assume that the tables have a common odds ratio. If for example, vans had no association between belt use and fatalities, but trucks had association, then the assumption would be violated.

d. The variables `vh` and `pn` represent the number of the person and vehicle in the accident. Why did I remove those whose value for either is greater than one? I am looking for a statistical, rather than computational, issue here. Distributional assumptions about the fitted odds ratio rely on independence between individuals in the data set. Individuals involved in the same accident are likely to share conditions linked to fatality. For example, if the driver of one car is killed, it is likely that the accident was physically more violent, and others involved are more likely to be killed as well. I have selected individuals so only one person per accident contributes to the odds ratio.

Some students noted that position in the car might be a confounder, and removing everyone from the study except the driver gets rid of this confounder. This is right; I hadn’t thought of it.

2. In each of six labs, twenty chicks were randomly divided into a treatment group and a control group. The treatment (T) group were exposed
to pulsed electro-magnetic radiation, and the control (C) chicks were placed in the presence of a similar apparatus which was not turned on. The chicks were examined for deformities, and the results were tabulated. These data may be found at

http://www.stat.rutgers.edu/~kolassa/960-553/hen.dat. Please use a method covered in class so far.

a. Ignoring the fact that these data were collected in different labs, test the null hypothesis that chick deformities are unrelated to radiation.

Use
data hen; infile 'hen.dat'; input lab sick treat count; run;
proc freq data=hen;
table sick*treat/chisq relrisk norow nocol nopercent;
exact or;
weight count; run;

to obtain

<table>
<thead>
<tr>
<th>sick</th>
<th>treat</th>
<th>Frequency</th>
<th>0</th>
<th>1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>46</td>
<td>46</td>
<td>92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>12</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>57</td>
<td>58</td>
<td>115</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statistic DF Value Prob
----------------------------------------
Chi-Square 1 0.0348 0.8520

There is no evidence at all that radiation influences deformities.

b. Again ignoring the fact that these data were collected in different labs, calculate a 95% confidence interval for the odds ratio associating deformities with radiation.

The estimate is \( \frac{12 \times 46}{11 \times 46} = \frac{12}{11} = 1.090909 \), and its log is
log(1.090909) = 0.0870. The standard error of the log odds ratio is
\[ \sqrt{\frac{1}{46} + \frac{1}{46} + \frac{1}{11} + \frac{1}{12}} = 0.4666. \]

The 95% confidence interval for the odds ratio is \( \exp(0.0870 \pm 1.96 \times 0.4666) = (0.437, 2.722) \). This corresponds to the confidence intervals given by SAS.

c. Test the null hypothesis that chick deformities are unrelated to radiation, accounting for the fact that the data were collected from different labs.

Use the commands
```
proc freq data=hen; table lab*sick*treat/cmh relrisk;
  weight count; run;
```

to find
```
Summary Statistics for sick by treat
Controlling for lab
Cochran-Mantel-Haenszel Statistics (Based on Table Scores)
Statistic Alternative Hypothesis DF Value Prob
---------------------------------------------------------------
1 Nonzero Correlation 1 0.0080 0.9288
```

Hence we find no impact of treatment on deformities.

d. Allowing for the fact that these data were collected in different labs, calculate a 95% confidence interval for the odds ratio associating deformities with radiation. Compare the width of the interval with that obtained in part (b).

The contingency table confidence intervals computed from the previous input are
```
Type of Study Method 95% Confidence Limits
-------------------------------------------------------------
Case-Control Mantel-Haenszel 0.4063 2.6885
(Odds Ratio) Logit 0.4034 2.7324
```

respectively. Again, there is little difference in widths.

e. Estimate odds ratios for each table separately. Describe what you see.
Odds ratios for the individual labs are

\[
\begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
1.11111 & 1.75000 & 1.00000 & 1.00000 & 0.50000 & 1.00000 \\
\end{array}
\]

You could have done this using the commands

```latex
proc sort data=hen; by lab; run;
proc freq data=hen noprint; output out=outset relrisk;
  by lab; table treat*sick/relrisk; weight count; run;
proc print data=outset noobs; var LAB _RROR_; run;
```

Hence most of the labs show no effect; lab 1 shows a small effects indicating treatment damages chicks, lab 2 shows a larger effect in the same direction, and lab 5 shows a larger protective effect of treatment. None of these are significant.