960-584– Biostatistics I– Fall, 2003

Homework 4 Solutions, 10 Nov 2003

1. The following are counts of deaths by falls in a certain cohort, by month, presented by Ryan and Joiner (1994) and originally published in the World Almanac and Book of Facts (1984).

Month	Falls	Days in Month	Ice Score
Jan	1150	31	1
Feb	1034	28	1
Mar	1080	31	2
Apr	1126	30	0
May	1142	31	0
Jun	1100	30	0
Jul	1112	31	0
Aug	1099	31	0
Sep	1114	30	0
Oct	1079	31	0
Nov	999	30	2
Dec	1181	31	2

The final column above contains a score for how icy each particular month is. This data may also be found at `http://www.stat.rutgers.edu/~kolassa/960-584/falls.dat`.

a. Fit a Poisson regression model to these data, allowing the rate of death from falls to vary by what category of icyness is present in the month. Make sure you account for the differing number of days in the month as well. Assume a constant rate of death from falls within each month.

Here are the SAS commands to do the job:

```
data falls; infile 'falls.dat'; input falls days ice;
  l=log(days); run;
proc genmod data=falls; class ice ;
  model falls=ice/dist=poi offset=l; run;
```

and here are the results:

```
              Standard    Wald 95%      Chi-  Pr >
Parameter  DF Estimate  Error Confidence Limits  Square  ChiSq
Intercept   1   3.5677 0.0175  3.5334   3.6020 41494.8 <.0001
ice     0   1   0.0246 0.0209 -0.0163   0.0655   1.39  0.2382
ice     1   1   0.0437 0.0277 -0.0105   0.0979   2.50  0.1142
ice     2   0   0.0000 0.0000  0.0000   0.0000    .    .
```

To get credit you need to give estimates and standard errors.

b. Assume that the model from part (a) fits well. Test whether the relative risk of the level 1 icyness time to the level 0 icyness time is the same as the relative risk of the level 2 icyness time to the level 1 icyness time.

Three solutions are given. Only one is required for full credit. Let Z_j *be the iciness score for month* j *. The model from part (a) is* $\log(\lambda_j) = \beta_0 + \beta_1 X_j + \beta_2 W_j$ *, with*

$X_j = \begin{cases} 1 & \text{if } Z_j = 0 \\ 0 & \text{otherwise} \end{cases}$, *and* $W_j = \begin{cases} 1 & \text{if } Z_j = 1 \\ 0 & \text{otherwise} \end{cases}$. *This model satisfies the requirement of part (b) if and only if the ratio of rates for ice score 1 to the rate for ice score 0 is the same as the ratio of rates for ice score 2 to the rate for ice score 1. This holds if* $[\exp(\beta_0 + \beta_2)/\exp(\beta_0 + \beta_1)] = [\exp(\beta_0)/\exp(\beta_0 + \beta_2)]$ *, or equivalently* $\beta_2 = 2\beta_1$. *The model for part (b) is* $\log(\lambda_j) = \beta_0 + \beta_1 X_j + 2\beta_1 W_j = \beta_0 + \beta_1(X_j + 2W_j) = \beta_0 + \beta_1(2 - Z_j)$. *Equivalently,*

$$\log(\lambda_j) = \gamma_0 + \gamma_1 Z_j \qquad (*)$$

for $\gamma_1 = -\beta_1$ *and* $\gamma_0 = \beta_0 + 2\beta_1$. *Hence*

```
proc genmod data=falls;
  model falls=ice/dist=poi offset=l; run;
```

fits this model; this model differs from the previous model in that ice is used as a continuous variable here. We need to test whether this model is adequate; use the likelihood ratio test. The two log likelihoods are 79357.3920 and 79356.5290 *; twice the difference is* 1.726 *, which fails to exceed the* χ^2_1 *critical value of* 1.96^2 . *Do not reject the null hypothesis. The second model is adequate.*

Alternatively, we could also have performed this analysis by creating an additional covariate which when added to the model (*) reproduces the model of part (a). That is, $\log(\lambda_j) = \gamma_0 + \gamma_1 Z_j + \gamma_2 X_j = \gamma_0 + \gamma_1(2 - X_j - 2W_j) + \gamma_2 X_j = \gamma_0 + 2\gamma_1 - 2\gamma_1 W_j + (\gamma_2 - \gamma_1)X_j$. *Hence this model agrees with the model of part (a) with* $\beta_0 = \gamma_0 + 2\gamma_1$, $\beta_1 = \gamma_2 - \gamma_1$ *, and* $\beta_2 = -2\gamma_1$. *The present hypothesis corresponds to* $\gamma_2 = 0$. *This may be tested using*

```
data falls; set falls; x=0; if ice=1 then x=1; run;
proc genmod data=falls;
  model falls=ice x/dist=poi offset=l; run;
```

Tests of coefficients are given by

```
                Analysis Of Parameter Estimates
              Standard Wald 95% Confidence Chi-   Pr >
Parameter DF Estimate Error      Limits       Square ChiSq
Intercept  1   3.5923 0.0113   3.5701    3.6145  100295 <.0001
ice        1  -0.0123 0.0104  -0.0328    0.0081    1.39 0.2382
x          1   0.0314 0.0238  -0.0153    0.0780    1.74 0.1875
```

The null hypothesis of $\gamma_2 = 0$ *is not rejected; the second model is adequate.*

Again alternatively, we might test this null hypothesis by directly manipulating the rate estimates. Let κ_j *be the log of the rate for iciness score* j . *Then under the null hypothesis,* $\kappa_1 - \kappa_0 = \kappa_2 - \kappa_1$ *, or* $\kappa_2 - 2\kappa_1 + \kappa_0 = 0$. *Hence we can test this hypothesis by dividing* $\hat{\kappa}_2 - 2\hat{\kappa}_1 + \hat{\kappa}_0$ *by its standard error and comparing to a standard normal distribution. The standard error is the square of the sum of the inverse number of falls. To do this in SAS, do*

```
proc sort data=falls; by ice; run;
proc means data=falls sum noprint; by ice; var falls days;
  output out=byice sum=; run;
data byice; set byice; drop _type_ _freq_;
  lograte=log(falls/days); mult=1; if ice=1 then mult=-2;
  top=mult*lograte; bot=mult**2/falls; run;
proc means data=byice sum noprint;
  output out=test sum=; run;
data test; set test; test=top/sqrt(bot);
  keep test top bot; run;
proc print noobs data=byice; run;
proc print noobs data=test; run;
```

to obtain

```
ice  falls days  lograte   mult    top         bot
0   7772  214  3.59231    1   3.59231   .000128667
1   2184   59  3.61138   -2  -7.22275   .001831502
2   3260   92  3.56769    1   3.56769   .000306748
```

and

```
              top          bot    test
           -0.062751   .002266917   -1.31796
```

Again, this test statistic fails to exceed the Z statistic cutoff of ± 1.96 *; do not reject the null hypothesis.*

c. Does the model in part (a) fit well?

Output from the commands in (a) show that the Pearson goodness of fit statistic, divided by its degrees of freedom, is 1.6936. No, it doesn't seem to fit well.

d. Fit the same model of part (a) to the slightly collapsed data set

Month	Falls	Days in Month	Ice Score
Jan–Feb	2184	59	1
Mar	1080	31	2
Apr–Oct	7772	214	0
Nov–Dec	2180	61	2

Compare the parameter estimates, standard errors, and goodness of fit statistics that you see. Explain any similarities and differences.

Coefficients and standard errors are identical. This is generally the case for generalized linear models, when data lines with identical covariates are collapsed. Goodness of fit measures are much different.

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 below:

Lab	1		2		3		4		5		6	
	C	*T*	*C*	*T*	*C*	*T*	*C*	*T*	*C*	*T*	*C*	*T*
Healthy Chicks	5	6	7	6	8	8	9	9	8	8	9	9
Unhealthy Chicks	3	4	2	3	2	2	1	1	2	1	1	1

These data may also be found at `http://www.stat.rutgers.edu/~kolassa/960-584/hen.dat`. Note that most of the labs have fewer than 20 chicks classified, because some of the chicks were lost to causes unrelated to the presence or absence of radiation during the experiment. This data is a subset of that collected by Berman, *et. al.* (1990). Perform these calculations using logistic regression.

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

This entire question may be done either using logistic regression, conditional logistic regression, or contingency table methods. You need do ONLY ONE OF THESE. These SAS commands will do the job:

```
data hen; infile 'hen.dat'; input lab sick treat count;
  healthy=1-sick; run;
proc genmod data=hen descending;
  model sick=treat/dist=bin; freq count; run;
```

The important part of the output is:

```
        Analysis Of Parameter Estimates
Parameter  DF  Estimate    Std Err  ChiSquare  Pr>Chi
INTERCEPT  1   -1.4307     0.3356    18.1719  0.0001
TREAT      1    0.0870     0.4666     0.0348  0.8521
SCALE      0    1.0000     0.0000          .       .
```

Hence there is no evidence at all that radiation influences deformities. You might also have performed these calculations using conditional logistic regression; recall that phreg models the probability that the response is zero, and so we use healthy as the response:

```
proc phreg data=hen; model healthy*healthy(1)=treat/risklimits ties=discrete;
  freq count; run;
```

and we see

```
          Analysis of Maximum Likelihood Estimates
          Parameter   Standard                   Hazard
Variable  DF   Estimate     Error  Chi-Square  Pr > ChiSq    Ratio
treat     1    0.08621    0.46457     0.0344      0.8528    1.090
```

Again there is no evidence at all that radiation influences deformities. Finally, the standard contingency table approach is also acceptable; here we use

```
proc freq data=hen; table sick*treat/chisq relrisk; weight count; run;
```

to obtain

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

Again 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 interval is $\exp(0.0870 \pm 1.96 \times 0.4666) = (0.437, 2.722)$. *The answers from the conditional logistic regression*

```
        Analysis of Maximum Likelihood Estimates
                Hazard    95% Hazard Ratio
     Variable    Ratio    Confidence Limits
     treat      1.090     0.439     2.709
```

and from the contingency table analysis

```
Type of Study            Value     95% Confidence Limits
-----------------------------------------------------------------
Case-Control (Odds Ratio)   1.0909     0.4371     2.7225
```

are also acceptable.

c. Test the null hypothesis that chick deformities are unrelated to radiation, accounting for the fact that the data were collected from different labs.
Here are the sas commands to do the job:

```
proc genmod data=hen descending; class lab;
  model sick=treat lab/dist=bin; freq count; run;
```

Here is the important part of the output:

```
INTERCEPT      1   -2.2196    0.7848    7.9980 0.0047
TREAT          1    0.0443    0.4829    0.0084 0.9269
LAB      1     1    1.7429    0.8888    3.8456 0.0499
LAB      2     1    1.2418    0.9124    1.8523 0.1735
LAB      3     1    0.8110    0.9317    0.7576 0.3841
LAB      4     1    0.0000    1.0541    0.0000 1.0000
LAB      5     1    0.5244    0.9755    0.2890 0.5908
LAB      6     0    0.0000    0.0000         .      .
```

Here we see results that are qualitatively the same as for the analysis ignoring lab. Again, do not reject the null hypothesis; there is no evidence that radiation impacts deformities. We could also have done this analysis with conditional logistic regression or via a contingency table:

```
proc phreg data=hen; model healthy*healthy(1)=treat/risklimits
ties=discrete;
  strata=lab; freq count; run;
proc freq data=hen; table lab*sick*treat/cmh relrisk; weight count; run;
```

and found

```
          Analysis of Maximum Likelihood Estimates
            Parameter    Standard
Variable   DF    Estimate      Error   Chi-Square   Pr > ChiSq
treat      1     0.08621     0.46457     0.0344      0.8528
```

and

```
          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
```

respectively, and also found 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).

Again, the confidence interval is given by $\exp(0.0443 \pm 1.96 \times 0.4829) = (0.406, 2.693)$ *. The width is almost exactly the same. The conditional logistic regression and contingency table confidence intervals are*

```
          Analysis of Maximum Likelihood Estimates
                  Hazard    95% Hazard Ratio
        Variable    Ratio    Confidence Limits
        treat       1.090    0.439     2.709
```

and

```
  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

1	*2*	*3*	*4*	*5*	*6*
1.11111	*1.75000*	*1.00000*	*1.00000*	*0.50000*	*1.00000*

.

These calculations might be done two ways: Using `proc freq`*, the commands*

```
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;
```

Using `proc genmod`*, the command*

```
proc genmod data=hen; class lab;
  model sick=treat lab lab*treat/dist=bin; freq count; run;
```

gives (with lab effects deleted):

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT	1	-2.1972	1.0541	4.3450	0.0371
TREAT	1	0.0000	1.4907	0.0000	1.0000
TREAT*LAB 1	1	0.1054	1.7811	0.0035	0.9528
TREAT*LAB 2	1	0.5596	1.8344	0.0931	0.7603
TREAT*LAB 3	1	-0.0000	1.8634	0.0000	1.0000
TREAT*LAB 4	1	-0.0000	2.1082	0.0000	1.0000
TREAT*LAB 5	1	-0.6931	1.9930	0.1210	0.7280
TREAT*LAB 6	0	0.0000	0.0000	.	.

Then lab 6 is treated as the baseline, and the estimated odds ratio is 1. Hence all of the other estimates of comparisons with lab 6 are the odds ratios for those labs as well. Exponentiating gives the above results. 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.