Practical 3(a): Case Studies in Poisson Regression

Solutions

Dankmar Böhning
Southampton Statistical Sciences Research Institute
University of Southampton, UK

`Advanced Statistical Methods in Epidemiology`

1 Application of Poisson Regression to the BELCAP Study

questions: Using Poisson regression, calculate the

- **crude** relative risk for each intervention using the control school as reference
- and the relative risk for each intervention adjusting for initial dental status (DMFTB)
- and the relative risk for each intervention adjusting for initial dental status (DMFTB) and gender! Is the adjustment for gender necessary?

use as number of events the DMFT index at the end of study (DMFTE)!

solutions: first, since STATA uses the lowest value as reference we recode the value 3 to 0 for SCHOOL

```
. recode school (3 = 0)
```

for the first question we simply use DMFTE as dependent variable (outcome) and SCHOOL as independent variable (use i.school since it is a categorical variable)

the option `irr` will report risk ratios instead of log-relative risk

```
 xi: poisson dmfte i.school, irr
i.school          _Ischool_0-6        (naturally coded; _Ischool_0 omitted)


Poisson regression                                Number of obs   =        797
                                                  LR chi2(5)      =      48.92
                                                  Prob > chi2     =     0.0000
Log likelihood = -1476.1456                       Pseudo R2       =     0.0163

------------------------------------------------------------------------------
       dmfte |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  _Ischool_1 |   .7942158   .0686149    -2.67   0.008     .6705025    .9407552
  _Ischool_2 |   .5572534   .0533303    -6.11   0.000     .4619457    .6722247
  _Ischool_4 |   .9172604   .0748336    -1.06   0.290     .7817148    1.076309
  _Ischool_5 |   .7041359   .0590847    -4.18   0.000     .5973534    .8300067
  _Ischool_6 |   .7694778   .0672547    -3.00   0.003     .6483335    .9132585
------------------------------------------------------------------------------
```

for the second question we simply add DMFTB as independent variable, in addition to SCHOOL. No i. is required this time, since it is treated as a numerical variable.

```
. xi: poisson dmfte i.school  dmftb, irr
i.school          _Ischool_0-6        (naturally coded; _Ischool_0 omitted)


Poisson regression                                Number of obs   =        797
                                                  LR chi2(6)      =     468.57
                                                  Prob > chi2     =     0.0000
Log likelihood = -1266.3213                       Pseudo R2       =     0.1561

------------------------------------------------------------------------------
       dmfte |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  _Ischool_1 |   .7629138   .0659242    -3.13   0.002     .6440539    .9037092
  _Ischool_2 |   .6627532   .0635999    -4.29   0.000     .5491207    .7999002
  _Ischool_4 |   .9644138   .0786846    -0.44   0.657     .8218936    1.131648
  _Ischool_5 |   .7388284       .062    -3.61   0.000     .6267778    .8709106
  _Ischool_6 |   .8970524   .0786493    -1.24   0.215     .7554201    1.065239
       dmftb |   1.235216    .013063    19.98   0.000     1.209877    1.261086
------------------------------------------------------------------------------
```

evidently, the results change a lot, in particular for school 6 which is no longer having a significant intervention effect. The DMFT index at the beginning works as a confounder that needs to adjusted for. This becomes clear if we look at the average values of the DMFT at the beginning and end of study for the the 6 schools:

```
. tabstat dmftb dmfte, statistics( mean ) by(school) columns(variables)

Summary statistics: mean
  by categories of: school

  school |     dmftb     dmfte
---------+--------------------
       0 |  3.698529  2.345588
       1 |  3.903226  1.862903
       2 |   2.76378  1.307087
       4 |   3.30303  2.151515
       5 |  3.322581  1.651613
       6 |  2.926829  1.804878
---------+--------------------
   Total |  3.323714  1.854454
------------------------------
```

In school 6, the DMFT index at the beginning is already low, so that the low index at the end is less substantial.

For the final question, we simply add GENDER as a further independent variable. Since GENDER is a dummy (only 2 values), this can be done using the i. operator (categorical) or without it (numerical variable).

```
. xi: poisson dmfte i.school  dmftb gender, irr
i.school          _Ischool_0-6        (naturally coded; _Ischool_0 omitted)

Poisson regression                                Number of obs   =        797
                                                  LR chi2(7)      =     468.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -1266.3144                       Pseudo R2       =     0.1561

------------------------------------------------------------------------------
       dmfte |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  _Ischool_1 |   .7634049   .0660985    -3.12   0.002      .6442503    .9045971
  _Ischool_2 |   .6631546   .0637308    -4.27   0.000      .5493034    .8006033
  _Ischool_4 |   .9646178   .0787203    -0.44   0.659      .8220356    1.131931
  _Ischool_5 |   .7396435   .0624548    -3.57   0.000      .6268271    .8727646
  _Ischool_6 |   .8973438   .0787142    -1.23   0.217      .7556004    1.065677
       dmftb |   1.235055   .0131331    19.85   0.000      1.209581    1.261065
      gender |   1.006239   .0532403     0.12   0.906       .907118    1.116191
------------------------------------------------------------------------------
```

There is no significant gender effect, and the variable can be removed again.

2 alcohol and mortality

Shaper *et al.* (1988, *Lancet*) describe a cohort study of a random sample of 7729 middle-aged British men. Each man was asked, at baseline, his alcohol consumption. During the next 7.5 years death certificates were collected for any of the cohort who happened to die. The table given below was compiled.

Alcohol Consumption

	None	Occasional	Light	Moderate	Heavy
At Risk	466	1845	2544	2042	832
Deaths	41	142	143	116	62

alcohol and mortality: questions

use the Poisson regression for estimating relative risk of alcohol consumption, note that n = number at risk needs to be taken as offset!

1. fit the model where alcohol consumption is considered categorical!

```
. poisson events ib(0).alcohol, exposure(n) irr

Iteration 0:   log likelihood = -15.868351
Iteration 1:   log likelihood = -15.856443
Iteration 2:   log likelihood = -15.856441

Poisson regression Number of obs   = 5
LR chi2(4)      = 12.06
Prob > chi2     = 0.0169
Log likelihood = -15.856441 Pseudo R2       = 0.2755


events         IRR   Std. Err.      z P>z     [95% Conf. Interval]

alcohol
1     .8747703   .1550899    -0.75 0.450     .6179918 1.238241
2     .6650233    .117811    -2.30 0.021     .4699415 .9410874
3     .6456606   .1173095    -2.41 0.016     .4522205 .9218459
4     .8469746   .1704909    -0.83 0.409     .5708587 1.256644

_cons    .0879828   .0137406   -15.56 0.000     .0647832 .1194905
ln(n)           1  (exposure)
```

   every category of alcohol consumption is treated separately and has an associated relative risk (except 0 which is the reference)

2. fit the straight line model $\log E(Y) = \log(n) + \alpha + \beta x$ where x is alcohol consumption (x is treated as continuous)

```
. poisson events alcohol, exposure(n) irr
```

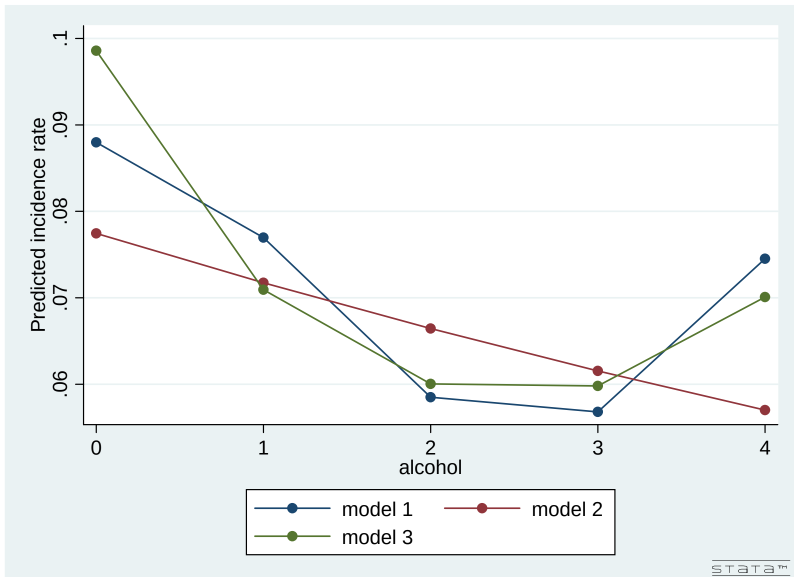
```
Iteration 0:   log likelihood = -20.147748
Iteration 1:   log likelihood = -20.147637
Iteration 2:   log likelihood = -20.147637

Poisson regression Number of obs   = 5
LR chi2(1)      = 3.48
Prob > chi2     = 0.0623
Log likelihood = -20.147637 Pseudo R2       = 0.0794


events         IRR   Std. Err.      z P>z     [95% Conf. Interval]

alcohol    .9262484   .0380792    -1.86 0.062     .8545422 1.003972
_cons    .0774572   .0073323   -27.02 0.000     .0643406 .0932478
ln(n)           1  (exposure)
```

We see that there is only a borderline preventive effect of alcohol consumption. As the comparison with the categorical model shows, this is not really the right model and misses the point that the relative risk frist declines with increasing alcohol consumption and then increases again.

3. fit the model with a curvature term: $\log E(Y) = \log(n) + \alpha + \beta x + \gamma x^2$

```
. poisson events alcohol  alc2, exposure(n) irr

Iteration 0:   log likelihood = -16.913617
Iteration 1:   log likelihood =  -16.91161
Iteration 2:   log likelihood =  -16.91161

Poisson regression Number of obs   = 5
LR chi2(2)      = 9.95
Prob > chi2     = 0.0069
Log likelihood =  -16.91161 Pseudo R2       = 0.2273


events         IRR   Std. Err.      z P>z     [95% Conf. Interval]

alcohol     .663327    .088861    -3.06 0.002     .5101508 .8624955
alc2    1.084694   .0339093     2.60 0.009     1.020228 1.153233
_cons    .0985836   .0125691   -18.17 0.000     .0767855 .1265699
ln(n)           1  (exposure)
```

evidently, this is the better model (although the fit at alcohol=0 is not great). see also the graph below.


.1
.09
.08
.07
.06
Predicted incidence rate
0
1
2
3
4
alcohol
model 1
model 2
model 3