Biostatistics 140.654
Fourth Term, 2021
May 15-17, 2021

Quiz 3

The purpose of this quiz is to assess your knowledge of the course materials covered during the third two weeks of class and covered in Problem Set 3.

Instructions:

- **This is an open book quiz; you may consult your course notes and handouts.**
- **You should not discuss this quiz with any other student during Monday May 3rd.**
- **This quiz is designed to be completed in 20-30 minutes.**
- **There are 4 questions on this quiz plus one BONUS question is a BONUS. Completing the bonus can only earn you extra points, i.e. if you choose to not answer this question, you will not lose any points.**
- **Questions 1 through 3 have a single best answer.**
- **You can use calculators or R on your computer for arithmetic.**
- **You may provide your solution by editing the word version of this quiz, annotating the pdf version of this quiz or writing your solution on paper and submitting a picture of your solution.**

Submit your quiz to the Quiz3 Dropbox by 5pm on May 17th.

By signing or typing your name below, you agree to abide by the Bloomberg School of Public Health Academic code and adhere to the above instructions.

Name (Print): ____

Signature: ____

Below find a set of times (in weeks) to hospitalization for persons with a diagnosis of schizophrenia who have been randomized to standard therapy (Trt=0) or a new drug treatment (Trt=1). A plus sign indicates censoring that we will assume to be independent of disease.

Trt=0: 6 8 11+ 13 16 16 19 21+ 22+ 28 28+ 29 31 35 40+ 41+ 41+ 59+ 86+ 132+

Trt=1: 6 9+ 9 10 11+ 12+ 13+ 17+ 18 19+ 19 20+ 22 24 28+ 31 43+ 48 51+ 57+

1. Assuming a constant hazard (i.e. risk of hospitalization) in weeks 11 through 20, the estimated hazard of hospitalization for the new drug group (Trt=1) during this period is:

a. 2/129 person-weeks == 10x9 + 1 + 2 + 3 + 7 + 8 + 9x2 = 129 person weeks and 2 events are at 18 and 19 weeks.
b. 5/129 person-weeks
c. 7/129 person-weeks
d. 2/169 person-weeks
e. 7/169 person-weeks

2. Suppose we bin the survival data obtaining the number of events and person-weeks for each 10-week interval separately for each treatment group. An estimator of the time-adjusted relative risk of hospitalization for persons on the new drug as compared to the standard drug is given by:

a. the treatment coefficient from a Poisson regression of the number of events per bin on: person time (offset), treatment, indicators for the time bins
b. the exponential of the treatment coefficient from a Poisson regression of the number of events per bin on: person time (offset), treatment, indicators for the time bins
c. the treatment coefficient from a Poisson regression of the number of events per bin on: log person time, treatment, indicators for the time bins
d. the exponential of the treatment coefficient from a Poisson regression of the number of events per bin on: log person time (offset), treatment, indicators for the time bins
e. the exponential of the interaction term for treatment with time from a a Poisson regression of the number of events on: log person time (offset), treatment, indicators for the time bin; and the interaction of treatment and time.

3. Suppose we bin time into the intervals: 1-10, 11-20, 21-30, 31-40, >40 weeks. In the space below, specify a null and extended model to conduct a likelihood ratio test of the null hypothesis that the relative rate of hospitalization for persons on the new drug as compared to the standard drug is the same within all follow-up time bins.

Null Model:

$$
\begin{aligned}
log[E(Y_i|TRT_i, midp_i)] \\
&= \log(py_i) + \beta_0 + \beta_1 TRT_i + \beta_2 I(bin = 11-20) + \beta_3 I(bin = 21-30) \\
&+ \beta_4 I(bin = 31-40) + \beta_5 I(bin => 40)
\end{aligned}
$$

Extended Model:

$$
\begin{aligned}
log[E(Y_i|TRT_i, midp_i)] \\
&= \log(py_i) + \beta_0 + \beta_1 TRT_i + \beta_2 I(bin = 11-20) + \beta_3 I(bin = 21-30) \\
&+ \beta_4 I(bin = 31-40) + \beta_5 I(bin => 40) + \beta_6 TRT_i x I(bin = 11-20) \\
&+ \beta_7 TRT_i x I(bin = 21-30) + \beta_8 TRT_i x I(bin = 31-40) + \beta_9 TRT_i x I(bin => 40)
\end{aligned}
$$

Degrees of freedom for LR test?

4 df to test $\beta_6, \beta_7, \beta_8, \beta_9$

4. There were a few key ideas in this course. Indicate true (T) or false (F) for each statement below.

_T_ models are never true; they can be useful tools to empirically address a scientific question

_T or F_ the most important assumption in regression analysis is that the observations are independent of one another: **we spoke a ton about this assumption; most important is having the "right model" and we said 2nd assumption was independence. So I will accept both here.**

_F_ well-trained biostatisticians can find the best model to address each scientific question

_T_ you should measure the quality of a prediction on data distinct from those data used to estimate the prediction model

BONUS question on the next page!

BONUS: Suppose we bin time into the intervals: 1-10, 11-20, 21-30, 31-40, >40 weeks. We fit an over-dispersed Poisson model of the number of events per bin on: logarithm of person time (offset), treatment, indicators for the time bins.

Provide estimates of and 95% confidence intervals for the expected number of hospitalizations during the first 10 weeks of follow-up among patients receiving the standard therapy (TRT=0) and new drug treatment (TRT=1). Show your work (I provided an extra blank page for space).

```
   BIN treat ptime  n event  midpoint
  1-10     0   194 20     2         5
 11-20     0   155 18     4        15
 21-30     0   108 13     2        25
 31-40     0    66  8     2        35
 41-50     0    32  5     0        45
   51+     0   127  3     0        55
  1-10     1   194 20     3         5
 11-20     1   129 16     2        15
 21-30     1    64  8     2        25
 31-40     1    41  5     1        35
 41-50     1    31  4     1        45
   51+     1     8  2     0        55
```

The variables midp15, midp25, midp35, midp45, midp55 are indicator variables for whether the midpoint of the time interval is equal to 15, 25, 35, 45, and 55, respectively.

```
fit = glm(event ~ treat + midp15 + midp25 + midp35 + midp45 + midp55,
       family = "quasipoisson", offset = log(ptime), data = pydat)

> summary(fit)$coeff
               Estimate   Std. Error       t value     Pr(>|t|)
(Intercept)  -4.4231023    0.3064846 -14.431725728 2.881554e-05
treat1        0.1382923    0.2771610   0.498960282 6.389810e-01
midp15        0.5006929    0.3635027   1.377411938 2.268492e-01
midp25        0.6081855    0.4040501   1.505222961 1.926072e-01
midp35        0.7936123    0.4393553   1.806311074 1.306915e-01
midp45        0.2095291    0.6572028   0.318819551 7.627514e-01
midp55      -19.4113881 4671.8186132  -0.004154996 9.968455e-01

> round(vcov(fit),3)
            (Intercept) treat1 midp15 midp25 midp35 midp45       midp55
(Intercept)       0.094 -0.041 -0.074 -0.077 -0.077 -0.072       -0.091
treat1           -0.041  0.077  0.004  0.010  0.009  0.001        0.036
midp15           -0.074  0.004  0.132  0.072  0.072  0.072        0.074
midp25           -0.077  0.010  0.072  0.163  0.073  0.072        0.077
midp35           -0.077  0.009  0.072  0.073  0.193  0.072        0.076
midp45           -0.072  0.001  0.072  0.072  0.072  0.432        0.072
midp55           -0.091  0.036  0.074  0.077  0.076  0.072 21825889.155
```

Solution:

$$log[E(Y_i|TRT_i, midp_i)] = \log(py_i) + \beta_0 + \beta_1 TRT_i + \beta_2 midp25 + \beta_3 midp35 + \beta_4 midp45 + \beta_5 midp55$$

The estimates of the log number of events in the first bin for control and treated patients are given by:

Control: $\log(194) + \hat{\beta}_0 = \log(194) - 4.42 = 0.85$

Treated: $\log(194) + \hat{\beta}_0 + \hat{\beta}_1 = \log(194) - 4.42 + 0.138 = 0.99$

The variance of the estimated log number of events in the first bin for control and treated patients are given by:

Control: $Var(\hat{\beta}_0) = 0.094$

Treated: $Var(\hat{\beta}_0) + Var(\hat{\beta}_1) + 2Cov(\hat{\beta}_0, \hat{\beta}_1)$=0.094 + 0.077 + 2 (-0.041) = 0.089

95% Confidence intervals for log number of events in the first bin for control and treated patients are given by:

0.85 +/- 1.96 sqrt(0.094): 0.25 to 1.45

0.99 +/- 1.96 sqrt(0.089): 0.41 to 1.57

Estimates of the expected number of events in the first bin for control and treated patients are given by:

Control: exp(0.85) = 2.34

Treated: exp(0.99) = 2.69

95% Confidence intervals for the number of events in the first bin for control and treated patients are given by:

Exp(0.25 to 1.45) -> 1.28 to 4.26

Exp(0.41 to 1.57) -> 1.51 to 4.81