Re: Single−Factor−Cox−Regression

Source: http://sci.tech−archive.net/Archive/sci.stat.math/2007−02/msg00400.html

- *From*: David Winsemius <doe_snot@xxxxxxxxxxx>
- *Date*: Sun, 25 Feb 2007 15:43:03 −0600

Richard Ulrich <Rich.Ulrich@xxxxxxxxxxx> wrote in
news:grh1u25qmjbgaphi8d7ousfbrh6i8jitmf@xxxxxxx:

> On 23 Feb 2007 12:08:09 −0800, patrick.ringsmoeller@xxxxxxxx wrote:
>
> > Hello everybody,
> >
> > I would like to test the influence of a clinical parameter upon
> > patient survival by Cox−Regression. But I am not sure, whether this is
> > allowed (using only a single factor (covariate) within Cox−Regression)
> > or if Cox−Regression always needs a certain amount of different
> > covariates to be correct in a statistical sense. Moreover, I wondered
> > if there exists a rule of thumb how many covariates can be included in
> > a Cox−Regression at maximum for a given number of patients under
> > consideration.
>
> The Cox proportionate hazard regression can is surely
> safer with one variable than with several.
>
> I have not seen mention of limits for covariates.
> I have run a number of them, with fairly small N, and at some
> point, it can fail to converge, or give bad numbers. Run some
> with your own data −− Then try it with partial samples −− Does it
> seem to work?
>
> Like the logistic regression, which it is sort−of an extension
> of, it is probably best used as a large sample procedure.
> That is, there is a dichotomous criterion at each of the
> multiple periods, and there is a further assumption that the
> rates for two groups have the same Odds Ratio at each
> of the several periods.

Cox regression models hazards and hazard ratios. No odds ratios. The goal is to model probability of an event per unit time. Cox models do have

linear predictors inside exponential functions, but they are not extensions of logistic models.

> One conservative guideline that has been mentioned for Logistic regression is to require at least 20 more cases in the smaller category, for each additional covariate. (I think that requiring "20" is being conservative; you are almost assured of having no problems with an N that large, but you might do pretty well with half that many... or, you might not.)
>
> Logistic regression has one outcome; Cox regression has outcomes at many periods. So it will need a bigger N.

I rather doubt that you can provide a citation or a mathematical justification for that assertion.

Logistic regression has no representation of time or censoring.

Cox regression models time between events and calculates rates. It should have more power, rather than less power to properly characterize risk in cohorts.

> I *think* that the Cox regressions are implemented around discrete periods, so that you need cases in both groups at every period. Thus, your N for Cox regression is probably a multiple of what you need for Logistic regression, depending on the number of periods.

No. Time is _not_ modeled as a discrete variable in Cox regression. There are no "numbers of periods" that would affect the statistical power of the procedure. You may be confusing Poisson regression with Cox regression. In Poisson regression the assumption of constant hazards during pre−specified intervals is often used. If you only measure time of event to the nearest year, then that it your choice, not something imposed by the Cox model. I suggest the OP completely ignore your unsubstantiated hand−waving arguments.

Although I have not looked at it, not having access from home, I suspect that persons with academic accounts may find this article interesting: "Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression" by Eric Vittinghoff and Charles E. McCulloch:

Abstract:
The rule of thumb that logistic and Cox models should be used with a minimum of 10 outcome events per predictor variable (EPV), based on two simulation studies, may be too conservative. The authors conducted a large simulation study of other influences on confidence interval

coverage, type I error, relative bias, and other model performance measures. They found a range of circumstances in which coverage and bias were within acceptable levels despite less than 10 EPV, as well as other factors that were as influential as or more influential than EPV. They conclude that this rule can be relaxed, in particular for sensitivity analyses undertaken to demonstrate adequate control of confounding. <http://aje.oxfordjournals.org/cgi/reprint/kwk052v1>

> If anyone knows something more definite about that, I will be happy to hear it.

−−
David Winsemius

.