Re: Averaging probabilities?

Source: http://sci.tech−archive.net/Archive/sci.stat.math/2008−03/msg00182.html

- *From*: rocketD <darahx@xxxxxxxxx>
- *Date*: Thu, 20 Mar 2008 10:11:53 −0700 (PDT)

On Mar 19, 5:26 pm, Bruce Weaver <bwea...@xxxxxxxxxxxx> wrote:

> On Mar 19, 2:57 pm, rocketD <dar...@xxxxxxxxx> wrote:
>
> > Hi all.
> >
> > Background: I'm working on a study looking at neonatal
> > hypoglycemia as a binary outcome, using infants who have neonatal
> > hypoglycemia (cases) matched for comparison with infants who don't
> > (controls). I've run a univariate model using conditional logistic
> > regression in Stata 9, and for each observation, have predicted the
> > conditional probability of developing neonatal hypoglycemia given
> > gestational age. The probability value output by Stata is called
> >
> > P(Neonatal Hypoglycemia | Single outcome w/in group)
> >
> > and I want to plot this against gestational age to show that there is
> > a steep negative slope up to a certain gestational age, after which
> > point it levels out. (Translation: risk of hypoglycemia decreases as
> > gestational age increases, to a point, then becomes stable.) I'd like
> > to do this cleanly, with one summary probability, like an average,
> > [that is, one y−value] for each gestational age (x−value), rather than
> > the 30 or so I have for each category now.
>
> −−− snip −−−
>
> Does this mean that you treated age as a continuous variable in your

> model, but now you want to use age categories in the plot? If age was
> treated as a continuous variable, why not treat it that way in your
> plot too. I.e., make a scatter−plot with X = age (selected values),
> and Y = the fitted probability (or log−odds) of the outcome? If your
> model treated age as continuous (and did not contain age−squared or
> any other higher order terms), then such a plot with Y = log−odds of
> the outcome should give you a straight line. With Y = the fitted
> probability of the outcome, it will be curvilinear, as you said.
>
> −−
> Bruce Weaver
> bwea...@xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx/wv/bwhomedir
> "When all else fails, RTFM."− Hide quoted text −
>
> − Show quoted text −

Hi Bruce and Paul,

Thanks for your replies. I agree on the boxplots, but wanted a concise graph for publication to illustrate one of our points visually. We chose our sample from the entire population of babies born at a certain hospital that fit very specific criteria, and babies of this nature are generally considered similar anywhere. Of course that generalization is an assumption that we'll state in the paper.

Regarding the nature of the variable, I used gestational age both as a continuous and as indicator. Technically, it is continuous, but it is measured in weeks and for our population, there are only 6 possible values (37−42). Indicator variables show similar results in the predicted conditional probabilities, but there is much less power for each of the individual levels when treating the variable as an indicator.

The plot Bruce describes is exactly what I've done, but I can't use logodds because they are not output from a conditional logistic regression − instead, the program will only predict the probability of outcome given one case within a group (that means that for each group of case & controls, the probabilities on the individuals will sum to 1.0). I *might* be able to get the logodds in SAS but haven't the time needed to figure it out. Anyway, I calculated the average probability for all observations falling within each age (37, 38, 39, etc) and overlaid it on the graph, and it does come out the way I expected, with the curve evident at least through the points.

Thanks for jogging my memory, though, I had to really think about each post which helped me understand the data and the mechanics of the analysis a lot better.

Dara
.