Subject: weighting data in regression analysis
Posted by Hejie Wang on Wed, 30 Oct 2024 15:15:43 GMT
View Forum Message <> Reply to Message

I want to explore the main determinants affecting childhood anemia, using variables from the KR document. I mainly use R for analysis, and the code is as follows:
DHS_data$wt<-DHS_data$v005/100000
model <- glm(formula, data = DHS_data, family = binomial, weights = wt)
Warning message:
In eval(family$initialize, rho) : non-integer #successes in a binomial glm!
As you can see, there's always a warning. But when I don't do the weighting, the warning goes away. So I want to know how to set my weights correctly. Another question I would like to ask is whether it is reasonable for me to take the cluster and country of the research object as random items when conducting multi-level logistic regression analysis. In addition, I use the lme4 package for multilevel analysis, but it always takes a lot of time to run a model, because there are about 400,000 samples included, so I wonder if there is any way to run my code faster

Subject: Re: weighting data in regression analysis
Posted by Bridgette-DHS on Thu, 31 Oct 2024 12:13:51 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS staff member, Tom Pullum:

You have combined too many questions, and these are mainly questions about R syntax.

It appears that you have pooled many surveys into a single file and you are treating "survey" or "country" as a random effect. There are profound differences in anemia between countries, and changes within countries over time. I recommend that you analyze each survey or country separately.

Perhaps other users can help.