
Subject: Re: Predciting Low Birth Weight Prediction with Random Forest in R
Posted by [Janet-DHS](#) on Tue, 27 Aug 2024 18:17:02 GMT

[View Forum Message](#) <> [Reply to Message](#)

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

I suggest that you tabulate the responses (m19 in the KR file). You will see that 82% of the children were either "not weighed" or have a "don't know" code. For children with numerical values, most are heaped at multiples of 500 grams, with a substantial bump right at 2500 grams, the boundary for LBW. There is a lot of omission and measurement error. You are further losing information by collapsing an interval-level variable into a binary variable.

But even if you had no omission or measurement error, it would be very hard to predict (or fit) birthweight, especially using the variables in DHS surveys. For example, the surveys do not include good estimates of gestational age, which has a clear causal relationship with birthweight.

There is a huge literature on this topic. DHS staff cannot help. Perhaps other users can offer suggestions.
