
Subject: Guidance Needed on Weighting for Pooled DHS Data in Logistic Regression

Posted by [ykim127](#) on Sat, 02 Nov 2024 12:49:54 GMT

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

Hello,

I am reaching out with a question regarding data weighting. I am conducting research on "The Effect of Girls' Empowerment on Adolescent Pregnancy in Sub-Saharan Africa," aiming to investigate whether increased aged 15-19 girls' empowerment has a positive effect on reducing adolescent pregnancy rates in this region.

I plan to pool data from 27 Sub-Saharan African countries and will be using DHS-7 and DHS-8 data from the IR datasets of these countries. The explanatory variable will be women's empowerment, while the dependent variable will be the pregnancy status of adolescents aged 15-19. I intend to perform logistic regression analysis using Stata.

Since I need the overall pooled set weights that can represent Sub-Saharan Africa, I want to ensure that I am correctly calculating and applying these weights. I have read previous posts on this users forum and the "Note on DHS standard weight de-normalization" file, but I would appreciate your guidance to confirm my understanding and approach.

I have conducted weight de-normalization for each country using the formula: $V005 \times (\text{total females age 15-49 in the country at the time of the survey}) / (\text{number of women age 15-49 interviewed in the survey})$

I have extracted data only for married women aged 15-19 from each country.

I have used the "append" function to pool the data from the 27 countries into one dataset.

I want to apply weights when conducting logistic regression analysis, but I am unsure how to do so.

Please let me know if there are any mistakes in the sequence of these steps or in the weight de-normalization process. Additionally, I would greatly appreciate the exact Stata code for applying weights in the pooled dataset and conducting the logistic regression analysis.

I look forward to your response.

Thank you.
