
Subject: Re: Calculating Under-5 Mortality Rate for Regression Analysis Using DHS 2022 Data in R-Studio

Posted by [n.borgmann](#) on Mon, 05 Aug 2024 08:51:04 GMT

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

Hi everyone

I've made some initial progress and have performed a linear regression with fixed effects, where the dependent variable is the U5MR and the independent variable is the mother's years of schooling. However, I am seeking feedback on my methodology and results to ensure the robustness and accuracy of my analysis.

I have compiled my preliminary findings and the regression results in an HTML document, which I am attaching here. I would greatly appreciate it if anyone with experience in this area could take a look and provide some feedback.

I am aware that this is not a particularly precise calculation, but for a bachelor's thesis, a simple calculation is sufficient. I am curious if I have made any significant errors or if the calculations are somewhat correct. I would be very grateful for any tips and help.

Here is the link: `file:///Users/nilsborgmannprivate/Desktop/Bachelor%20Arbeit%20R%20Auswertung/analysis3_KEBR8BFL_CM.html`.

Some specific questions I have:

Are my assumptions and steps in the data preparation and analysis correct?

Is the fixed effects model I used appropriate for this type of analysis?

Any suggestions for improving the accuracy and interpretability of my results?

Thank you very much for your time and assistance.

Best regards,
Nils Borgmann

I have written a description of my approach here, as some parts are in German. It is sufficient to just read through this to check if my approach is correct.

Description of the Linear Regression

Goal of the Analysis:

The goal of this analysis is to investigate the relationship between the number of years of schooling a mother has and the under-5 mortality rate (U5MR) in Kenya. Additionally, the analysis accounts for regional differences by using fixed effects for the various regions.

Data Preparation:

Loading and Filtering Data:

The data was loaded from a DHS database containing relevant information on births, child mortality, educational levels, and regions.

Only the relevant variables were selected: CASEID (Case Identification), B3 (Date of Birth), B7 (Age at Death in Months), V008 (Date of Interview), V106 (Educational Level), V107 (Number of Years of Schooling), and V024 (Region).

Calculating Under-5 Mortality (death_before_5):

A new binary variable death_before_5 was created, which is set to 1 if a child died before their fifth birthday ($B7 < 60$), and 0 if the child reached at least five years old or is still alive.

Ensuring Region is Treated as a Factor:

The variable V024 (Region) was converted into a factor to be used as a fixed effect in the model.

Performing the Linear Regression:

Creating the Model Formula:

The formula for the model is: death_before_5 ~ V107 | V024

death_before_5: Dependent variable indicating whether a child died before the age of five.

V107: Independent variable indicating the number of years of schooling of the mother.

V024: Fixed effect for the region to control for regional differences.

Executing the Fixed Effects Model:

The model was estimated using the felm function from the lfe package.

The model controls for regional differences by including regional fixed effects (V024).

File Attachments

1) [analysis3_KEBR8BFL_CM.html](#), downloaded 214 times
