QMH Statistics Club
Department of Statistics, University of Dhaka

Data Analysis Competition
Topic: Data Management and Data Summarization

Instructions

- **Participating Years** : Batch-68 & Batch-69, Statistics, DU
- **Group** : Consists of 3 members (maximum) within each batch (Mix-up between batches is not allowed)
- **Top page** : Should contains name, email ID and phone number of each member of the group. It is mandatory to mention the academic year of the participants.
- **Submission** : By email – jafar@du.ac.bd
  Email subject: Data analysis competition (level-I)
- **Due** : **Feb 08, 2022**
- **Winner** : One from each batch
- **Presentation** : Each winning team will present their analysis result on the prize giving day.

The dataset has been selected from Bangladesh Demographic and Health Survey (BDHS), 2017-18. The respondents were women of reproductive age (15 – 49 years) who had one or more live birth in the five years preceding the survey.

Prepare data for analysis to addresses all the questions. Use the data to answer the questions Q1 to Q4, given in the next page. You should provide data description (specific variable(s) you have selected for answering a specific question).

Q1:

A respondent is said to be exposed to media if she reads newspaper/magazine, or listens radio, or watches TV at least once in a week. Discuss the behavior of the respondents regarding "media exposure".

Q2:

Generate a variable named "economic_status" by using the "wealth index factor score" in such a way that divide the ranked "wealth index factor score" scores in to three equal quintiles (e.g., 33.33 percentile each) with categories: *poor, middle and rich*. Describe the variable "economic_status".

Q3:

Compute a suitable measure of central tendency for "preceding birth interval" and interpret it. Note that, preceding birth interval is calculated as the difference in months between the current birth and the previous birth, counting twins as one birth i.e., preceding birth interval is computed for all births except the first birth and its twins.

Q4:

Compute a quantitative variable named "age_of_child". The variable takes value equal to "current age" of the children who are alive. Alternatively, it takes value equal to "age at death" for those children who had died before the interview date. Now compute a dichotomous variable "survival_status" which takes value "Survived" if the child is alive and "Died" if the child is not alive. Finally, compute a suitable measures of central tendency of "age_of_child" for the children who are not alive and describe your result.

Note that, a value of CMC for an "*event*" indicates that the particular "*event*" occurred at the number of months since the beginning of 1900. For example, a date of interview CMC is 1295 indicates "the interview happed after 1295 months from 1900".