Data Analysis Workshop for PERI

[Zahid Asghar](https://zahedasghar.netlify.app)

19-10-2023

Summary

## Objectives

* Delineate variables among various types, like nominal, ordinal and interval.
* Exporting, importing and wrangling data to get insight with data.
* Understand important elements of data, summarizing, visualization and tabulating data as per best practices
* Creating awesome graphs and tables for report writing and getting insight
* Exploring MICS/PDHS/ASER/Punjab ST other data sets
* Learning to use R for serving data analysis instead of learning coding
* All slides, data , data codes and related stuff will be handed over to reproduce all training material and prepare reports whenever required after the training.

## Main Topics

### Defining and Measuring Concepts

Taxonomy of data

### Measuring and Describing Variables with Real Data

constant/variable, nominal/ordinal/interval variable, dummy variable, Likert item, when can you treat an ordinal variable as interval/continuous?, mode/median/mean.

### Correlation and Linear Regression

correlation, scatterplot, Pearson’s r, multicollinearity, (multiple) regression, regression coefficient, standard error, prediction error, “ordinary least squares”, partial effects, interactive effects.

## Daywise detail

|  |  |  |
| --- | --- | --- |
| Day-1 | Topics | Sources |
| Session-1 | Introduction to R, RStudio, POSIT Cloud , gituhub , Taxonomy of data | [POSIT](p%20os%20it%20%20%20.cloud) |
| Session-2 | Reading /writing STATA/ SPSS/ csv/xlsx webAPI , pdf data into R | Hands-on practice with MICS/PDHS, Punjab STax data |
| Session-3 | Data Visualization (when to make which graph ) bar, scatter, line and histograms with educational/other case study data in ggplot2 | graphics for categorical and continuous data using R | ggplot2, WSJ/Economist/538 style graphs |
| Day-2 |  |  |
| Session-1 | Data Tables, Measure of central tendency, dispersion | Learning Data Tables with R as published in world leading outlets |
| Session-2 |  | Hands-on workshop on making Tables by exploring real data sets like GST and others to produce tables like 538, NYT, WSJ, FT… |
| Session-3 | Basics of Regression and Correlation |  |
|  | * Parameterized and automating reporting using Quarto | Hands on practice for writing html/pdf/word document/revealjs slides |

## Also New ways of having data and analyzing will also be explored

### Analyzing Google Trends, Google News API, Media API

Discussion of various sources from where one can read all data directly from web API. - Google trends - Google news API - Media API - UN votes data from 1948-49 on 6 key issues directly from API till to-date (Comparing Pakistan and India ) - Participants feedback for way forward