Final exercise

Xin Xia

2024-04-12 (updated: 2023-04-13)

Contents

```
# run the install command if did not do it before
# install.packages('remotes')
# remotes::install_github("Bolin-Wu/workshopr", subdir = "rpackage", force = TRUE)

# load the package
library(workshopr)
library(tidyverse)
library(here)
```

Introduction

Generate a html/word report based on the simulated data set "fake_data" using (1) the rmarkdown templates created by Bolin Wu and (2) tidyverse/loop/apply functions. Feel free to build on the syntaxes Bolin Wu and Ashley Tate wrote.

Optional tasks

The report can include the following contents:

- Summary statistics of the data set (e.g., mean and standard deviation for continuous variables, count and percentage for categorical variables, the number of missing values in each variable).

Hints: use "mutate" (from tidyverse) to define variable classes; use "select(where())" (from tidyverse) to select variables; use "summarise()" (from tidyverse) or "sapply()" to summarise.

- Summary statistics of the data set, stratified by sex (e.g., mean and standard deviation for continuous variables, count and percentage for categorical variables, the number of missing values in each variable).

Hints: use "tapply()" to summarise across sex strata; write loops to summarise multiple variables.

- Make any graph you want to make using the simulated data. If you don't know what to draw, try to make a spaghetti plot of MMSE changes over time.

Hint: use "pivot_longer()" to reshape data from a wide format to a long format.

- In a table, summarise the results of a regression. If you don't know which model to build, try summarising the output of this model: glm(formula = mmse_wave1~age_base+sex+education, family = "gaussian",data = fake_data).

Hints: first, check the elements in the output of the model summary by typing summary(model_name)$, confint(model_name), etc; build an empty data frame with names of quantities of interest (e.g., coefficients); use loops and "rbind" to add model results into the empty data frame.

Alternatives

- Using the simulated dataset, create a report and decide the contents of the report on your own.
- Create a report using your own data.

Solutions

You can find solutions from Xin Xia by

```
workshopr::get_solution_2023(name = "practice_solution")
```