Descriptive analysis (Frequency and %)

Ikramul H Khan

2025-07-19

loading libraries

```
library(haven)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(survey)
```

```
## Loading required package: grid

## Loading required package: Matrix

## Loading required package: survival

##
## Attaching package: 'survey'

## The following object is masked from 'package:graphics':
##
##     dotchart
```

```
library(srvyr)
```

```
##
## Attaching package: 'srvyr'
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
library(glmmTMB)
library(lme4)
library(officer)
library(flextable)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.5
## v ggplot2   3.5.2     v stringr   1.5.1
## v lubridate 1.9.4     v tibble    3.3.0
## v purrr     1.0.4     v tidyr     1.3.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x purrr::compose() masks flextable::compose()
## x tidyr::expand()  masks Matrix::expand()
## x srvyr::filter()  masks dplyr::filter(), stats::filter()
## x dplyr::lag()     masks stats::lag()
## x tidyr::pack()    masks Matrix::pack()
## x tidyr::unpack()  masks Matrix::unpack()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

Preparing Dataset

```
# Open the dataset
BDHS <- read_sav("adolescent fertility new_1.SAV")

#Recode the adolescent fertility using new variable V213 as currenly pregnant status

BDHS <- BDHS %>%
  mutate(adol_fertility = ifelse(V201 >= 1 | V213 == 1, 1, 0))

# Prepare variables

BDHS <- BDHS %>%
rename(
    education = V106,
    partner_education = V701,
    division = V024,
    residence = V025,
    religion = V130,
    wealth = V190,
    age = V012,
    age_gap = Age_Gap,
    contraceptive_status = V312New,
    WomenEmpowerment = WomenEmpowerment
  ) %>%
```

```
  mutate(
    weight = V005 / 1000000
  )
```

Create survey design object

```
bdhs_design <- BDHS %>%
  as_survey_design(
    ids = V001,
    strata = V023,
    weights = weight,
    nest = TRUE
  )
```

Final Descriptive output

```
# For strata with single PSU
options(survey.lonely.psu = "adjust")
# Variables to summarize
vars_to_summarize <- c(
  "education", "partner_education", "division", "residence", "religion",
  "wealth", "contraceptive_status", "age", "age_gap", "WomenEmpowerment"
)

# Function to summarize each variable
get_summary <- function(var) {
  bdhs_design %>%
    group_by(value = .data[[var]]) %>%
    summarise(
      n = unweighted(n()),
      percent = survey_mean(proportion = TRUE, na.rm = TRUE) * 100,
      .groups = "drop"
    ) %>%
    mutate(variable = var) %>%
    select(variable, category = value, n, percent)
}

# Apply and combine
Descriptive_Table <- bind_rows(lapply(vars_to_summarize, get_summary))
```

```
## Warning: There was 1 warning in `dplyr::summarise()`.
## i In argument: `percent = survey_mean(proportion = TRUE, na.rm = TRUE) * 100`.
## i In group 1: `value = 0`.
## Caused by warning:
## ! na.rm argument has no effect on survey_mean when calculating grouped proportions.
## This warning is displayed once per session.
```

```
## Warning: `..1$category` and `..3$category` have conflicting value labels.
## i Labels for these values will be taken from `..1$category`.
## x Values: 1, 2, and 3

## Warning: `..1$category` and `..4$category` have conflicting value labels.
## i Labels for these values will be taken from `..1$category`.
## x Values: 1 and 2

## Warning: `..1$category` and `..5$category` have conflicting value labels.
## i Labels for these values will be taken from `..1$category`.
## x Values: 1 and 2

## Warning: `..1$category` and `..6$category` have conflicting value labels.
## i Labels for these values will be taken from `..1$category`.
## x Values: 1, 2, and 3

## Warning: `..1$category` and `..7$category` have conflicting value labels.
## i Labels for these values will be taken from `..1$category`.
## x Values: 0 and 1

## Warning: `..1$category` and `..9$category` have conflicting value labels.
## i Labels for these values will be taken from `..1$category`.
## x Values: 1, 2, and 3

## Warning: `..1$category` and `..10$category` have conflicting value labels.
## i Labels for these values will be taken from `..1$category`.
## x Values: 0 and 1
```

```r
# View table
print(Descriptive_Table)
```

```
## # A tibble: 35 x 4
##    variable          category             n percent
##    <chr>             <dbl+lbl>        <int>   <dbl>
##  1 education         0 [No education]    40    1.37
##  2 education         1 [Primary]        339   14.6
##  3 education         2 [Secondary]     1879   76.9
##  4 education         3 [Higher]         191    7.11
##  5 partner_education 0 [No education]   117    4.85
##  6 partner_education 1 [Primary]        429   17.9
##  7 partner_education 2 [Secondary]     1541   62.9
##  8 partner_education 3 [Higher]         362   14.3
##  9 division          1 [Primary]        271    2.98
## 10 division          2 [Secondary]      340    5.36
## # i 25 more rows
```

```r
# Create Word document and add formatted table
doc <- read_docx() %>%
  body_add_par("Descriptive Table: Weighted Percentage and Unweighted Frequency", style = "heading 1") %
  body_add_flextable(flextable(Descriptive_Table))


# Save the document
print(doc, target = "Descriptive_Table_Output.docx")
```