Dropouts

July 13, 2023

```
# Load packages
library(tidyverse)
library(finalfit)
library(flextable)

p_to_star <- function(p.values) {
  star <- p.values |>
    symnum(corr = FALSE, na = FALSE,
           cutpoints = c(0, 0.001, 0.01, 0.05, 0.1, 1),
           symbols = c("***", "**", "*", "+", " ")) |>
    unclass()
  glue::glue("{scales::pvalue(p.values)}{star}")
}

summarize_NA_numeric <- function(col) {
  y <- W131415[[col]]
  g <- is.na(W131415$Dropouts)

  test <- t.test(y ~ g, var.equal = TRUE) |>
    parameters::model_parameters() |>
    mutate(t = scales::number(t, 0.01),
           df_error = scales::number(df_error, 1),
           p = p_to_star(p),
           Test = glue::glue("t({df_error})={t}")) |>
    select(Test, p)

  data.frame(g, y) |>
    drop_na() |>
    group_by(g) |>
    summarise(across(y, list(M = mean, S = sd), .names = "{.fn}")) |>
```

```
    mutate(across(M:S, scales::number_format(0.01)),
           value = glue::glue("{M}, ({S})")) |>
    select(g, value) |>
    arrange(g) |>
    pivot_wider(names_from = g, values_from = value) |>
    rename("Not Missing" = 1, "Missing" = 2) |>
    mutate(Variable = col, .before = 1) |>
    mutate(test)
}

summarize_NA_factor <- function(col) {
  y <- W131415[[col]]
  g <- is.na(W131415$Dropouts)

  tab <- table(y, g)
  tab_prop <- proportions(tab, margin = 1)

  Test <- chisq.test(tab) |> parameters::model_parameters() |>
    as.data.frame() |>
    mutate(Chi2 = scales::number(Chi2, 0.01),
           df = scales::number(df, 1),
           p = p_to_star(p),
           Test = glue::glue("Chi2({df})={Chi2}")) |>
    select(Test, p) |>
    reframe(across(everything(), \(x) c(x, "")))

  as.data.frame(tab) |>
    left_join(as.data.frame(tab_prop), by = join_by(y, g)) |>
    mutate(
      Freq.y = scales::number(Freq.y, scale = 100),
      value = glue::glue("{Freq.x} ({Freq.y})")
    ) |>
    select(g, y, value) |>
    arrange(g, y) |>
    pivot_wider(names_from = g, values_from = value) |>
    rename("Not Missing" = 2, "Missing" = 3) |>
    mutate(Variable = c(col, ""), .before = 1) |>
    mutate(Test)
}

print_pp_table <- function(df) {
```

```
  df |>
    flextable() |> autofit() |>
    footnote(
      j = 1, i = 1:5, ref_symbols = c("a"),
      value = as_paragraph(c("Values are Mean (SD)."))
    ) |>
    footnote(
      j = 1, i = c(6, 8, 10), ref_symbols = c("b"),
      value = as_paragraph(c("Values are N (row %)."))
    ) |>
    footnote(j = 6, part = "header", ref_symbols = c("c"),
             value = as_paragraph("+p<.1, *p<.05, **p<.01, ***p<.001"))
}
```

```
# Tidy data
W131415 <- haven::read_sav("data-raw/dropouts for MBS.sav") |>
  mutate(
    Dropouts = na_if(Dropouts, 1), # recode 1 to missing
    across(c(sexbin, EDUBIN, RELIDBIN), factor)
  )
```

```r
X_num <- c("age", "StateAnx", "WorriedSpread", "DEqsc1", "traitanx")
X_fac <- c("EDUBIN", "RELIDBIN", "sexbin")

results <- c(map(X_num, summarize_NA_numeric),
             map(X_fac, summarize_NA_factor)) |>
  bind_rows() |>
  relocate(y, .after = 1) |>
  rename(" " = y)


print_pp_table(results)
```

Variable		Not Missing	Missing	Test	p[c]
age[a]		47.70, (15.18)	37.94, (14.61)	t(989)=10.32	<0.001***
StateAnx[a]		2.09, (1.04)	2.18, (1.08)	t(989)=-1.36	0.174
WorriedSpread[a]		3.57, (1.84)	3.77, (1.84)	t(983)=-1.68	0.094+
DEqsc1[a]		4.49, (1.62)	4.43, (1.63)	t(978)=0.60	0.551
traitanx[a]		2.78, (1.68)	2.93, (1.74)	t(978)=-1.39	0.166
EDUBIN[b]	0	263 (45.8)	311 (54.2)	Chi2(1)=4.08	0.043*
	1	218 (52.5)	197 (47.5)		
RELIDBIN[b]	0	374 (49.5)	382 (50.5)	Chi2(1)=0.75	0.386
	1	108 (46.0)	127 (54.0)		
sexbin[b]	0	237 (49.07)	246 (50.93)	Chi2(1)=0.04	0.841
	1	245 (48.23)	263 (51.77)		

[a]Values are Mean (SD).

[b]Values are N (row %).

[c]+p<.1, *p<.05, **p<.01, ***p<.001