Creating NHANES Insurance Groups

Thomas E. Love, Ph.D.

Table of contents

1 R packages

```
knitr::opts_chunk$set(comment = NA)
library(janitor)
library(nhanesA)
library(tidyverse)
```

2 Pull in some NHANES 2017-18 Data

- ID number (SEQN) plus insurance information, only.

```
demo_raw <- nhanes('DEMO_J', translated = FALSE) |> tibble()
hiq_raw <- nhanes('HIQ_J', translated = FALSE) |> tibble()

combo <- left_join(demo_raw, hiq_raw, by = "SEQN")
```

```
combo <- combo |>
  mutate(SEQN = as.character(SEQN)) |>
  select(SEQN, HIQ011, HIQ031A, HIQ031B, HIQ031D)
```

3 Creating a four-level INSUR variable

Goal A new variable, called INSUR, which is a factor with four levels, in this order: Medicare, Commercial, Medicaid, Uninsured, and with some (inevitable) missing values.

Current Variables	Description
SEQN	Subject ID
HIQ011	1 if subject is covered by insurance, 2 if not, 7 = refused, 9 = don't know
HIQ031A	14 if covered by Commercial (Private) insurance
HIQ031B	15 if covered by Medicare
HIQ031D	17 if covered by Medicaid

```
combo <- combo |>
  mutate(INSUR = factor(
    case_when(
      HIQ011 == 2 ~ "Uninsured",
      HIQ011 == 1 & HIQ031D == 17 ~ "Medicaid",
      HIQ011 == 1 & HIQ031B == 15 ~ "Medicare",
      HIQ011 == 1 & HIQ031A == 14 ~ "Commercial")
    ),
    INSUR = fct_relevel(INSUR, "Medicare"))

combo |> tabyl(INSUR) |> adorn_pct_formatting()
```

```
      INSUR    n percent valid_percent
   Medicare 1320   14.3%         15.9%
 Commercial 3392   36.7%         40.8%
   Medicaid 2527   27.3%         30.4%
  Uninsured 1072   11.6%         12.9%
       <NA>  943   10.2%             -
```

4 What if we use 2017-2020 NHANES data instead?

- It's `P_DEMO` and `P_HIQ`, not `DEMO_J` and `HIQ_J`.
- The variables within HIQ have different names.
- The levels (codes) within each insurance type are different, too.

Resulting table turns out to be (across all ages - remember you will prune to adults ages 18-79 or perhaps to children):

```
combo2 |> tabyl(INSUR) |> adorn_pct_formatting()
```

```
      INSUR    n percent valid_percent
   Medicare 2153   13.8%         15.4%
 Commercial 5769   37.1%         41.3%
   Medicaid 4179   26.9%         30.0%
  Uninsured 1852   11.9%         13.3%
       <NA> 1607   10.3%             -
```

5 What if we use 8/21-8/23 NHANES data instead?

- It's `DEMO_L` and `HIQ_L`, not `DEMO_J` and `HIQ_J`.
- The variables within HIQ have different names.
- The levels (codes) within each insurance type are different, too.

Resulting table turns out to be (across all ages - remember you will prune to adults ages 18-79 or perhaps to children):

```
combo3 |> tabyl(INSUR) |> adorn_pct_formatting()
```

```
      INSUR    n percent valid_percent
   Medicare 2424   20.3%         23.3%
 Commercial 4539   38.0%         43.7%
   Medicaid 2565   21.5%         24.7%
  Uninsured  864    7.2%          8.3%
       <NA> 1541   12.9%             -
```