Final Exam, Question 4 - R

Stats 506, Fall 2019

12/16/2019

```
# Final Exam, Question 4 - R
# Stats 506, Fall 2019
#
# This R script reads in NHANES demographic data (DEMO_D) and
# audiometry data (AUX_D) and merges them together. It then uses the merged
# data to compute t-statistics comparing the audible volume at which
# younger people (25 years old or younger) and older people can hear at various
# sounds frequencies or pitches.
#
# Updated: December 15, 2019
# Author: James Henderson
# 80: ------------------------------------------------------------------------

# libraries: -----------------------------------------------------------------
library(tidyverse)

# read in data: --------------------------------------------------------------
aux = haven::read_xpt('./AUX_D.XPT')
demo = haven::read_xpt('./DEMO_D.XPT')

# for easier typing
names(aux) = stringr::str_to_lower(names(aux))
names(demo) = stringr::str_to_lower(names(demo))

# merge, select vars, and drop second reading at 1k hz: ----------------------
aud = left_join(aux, demo, by = 'seqn') %>%
  select(seqn, riagendr, ridageyr, starts_with('auxu')) %>%
  select(-contains('1k2'))

# decode missing values: -----------------------------------------------------
aud = aud %>% mutate_all( function(x) ifelse( x %in% c(888, 666), NA, x))

# age groups: ----------------------------------------------------------------
aud = aud %>%
  mutate( young = 1*{ridageyr <= 25}, old = 1*{ridageyr >= 50})

# reshape to long and filter people with all missing values: -----------------
aud_long =
  pivot_longer(aud, cols = starts_with('auxu'),
               values_to = 'auxu', names_to = 'thresh',
               names_prefix = 'auxu' # this removes this from the varname
  ) %>%
  filter( !is.na(auxu) )
```

```
# Task A, use regex to create "left" and "freq" from thresh: -------------------




# Task B, compute nominal confidence intervals: --------------------------------
#   comparing young to old for each ear/frequency
# Hint: the easiest way to do this, is not reshape but
#        to use an additional summarize









# Task C, clean for presentation: ----------------------------------------------
# For the solution, it is sufficient to have the right columns with clean names,
# they need not be in the correct order, as long as the order is logical.
```