Example R code to replicate NCHS Data Brief No. 303, Figure 1

# Figure 1. Percentage of persons aged 20 and over with depression, by age and sex: United States, 2013-2016

Brody DJ, Pratt LA, Hughes J. Prevalence of depression among adults aged 20 and over: United States, 2013-2016. NCHS Data Brief, no 303. Hyattsville, MD: National Center for Health Statistics. 2018.

# Load survey and dplyr packages

library(dplyr)
library(survey)

options(survey.lonely.psu='adjust')

# Display Version Information
cat("R package versions:\n")

## R package versions:

for (p in c("base", "survey","dplyr")) {
 cat(p, ": ", as.character(packageVersion(p)), "\n")
}

## base : 3.5.2
## survey : 3.36
## dplyr : 0.8.0.1

# Data preparation

# Download & Read SAS Transport Files
# Demographic (DEMO)
download.file("https://wwwn.cdc.gov/nchs/nhanes/2013-2014/DEMO\_H.XPT", tf <- tempfile(), mode="wb")
DEMO\_H <- foreign::read.xport(tf)[,c("SEQN","RIAGENDR","RIDAGEYR","SDMVSTRA","SDMVPSU","WTMEC2YR")]
download.file("https://wwwn.cdc.gov/nchs/nhanes/2015-2016/DEMO\_I.XPT", tf <- tempfile(), mode="wb")
DEMO\_I <- foreign::read.xport(tf)[,c("SEQN","RIAGENDR","RIDAGEYR","SDMVSTRA","SDMVPSU","WTMEC2YR")]

# Mental Health - Depression Screener (DPQ)
download.file("http://wwwn.cdc.gov/nchs/nhanes/2013-2014/DPQ\_H.XPT", tf <- tempfile(), mode="wb")
DPQ\_H <- foreign::read.xport(tf)
download.file("http://wwwn.cdc.gov/nchs/nhanes/2015-2016/DPQ\_I.XPT", tf <- tempfile(), mode="wb")
DPQ\_I <- foreign::read.xport(tf)

# Append Files
DEMO <- bind\_rows(DEMO\_H, DEMO\_I)
DPQ <- bind\_rows(DPQ\_H, DPQ\_I)

# Merge DEMO and DPQ files and create derived variables

One <- left\_join(DEMO, DPQ, by="SEQN") %>%
 # Set 7=Refused and 9=Don't Know To Missing for variables DPQ010 thru DPQ090 ##
 mutate\_at(vars(DPQ010:DPQ090), ~ifelse(. >=7, NA, .)) %>%
 mutate(. ,
 # create indicator for overall summary
 one = 1,
 # Create depression score as sum of variables DPQ010 -- DPQ090
 Depression.Score = rowSums(select(. , DPQ010:DPQ090)),
 # Create depression indicator as binary 0/100 variable. (is missing if Depression.Score is missing)
 Depression= ifelse(Depression.Score >=10, 100, 0),
 # Create factor variables
 Gender = factor(RIAGENDR, labels=c("Men", "Women")),
 Age.Group = cut(RIDAGEYR, breaks=c(-Inf,19,39,59,Inf),labels=c("Under 20", "20-39","40-59","60 and over")),
 # Generate 4-year MEC weight (Divide weight by 2 because we are appending 2 survey cycles)
 # Note: using the MEC Exam Weights (WTMEC2YR), per the analytic notes on the
 # Mental Health - Depression Screener (DPQ\_H) documentation
 WTMEC4YR = WTMEC2YR/2 ,
 # Define indicator for analysis population of interest: adults aged 20 and over with a valid depression score
 inAnalysis= (RIDAGEYR >= 20 & !is.na(Depression.Score))
 ) %>%
 # drop DPQ variables
 select(., -starts\_with("DPQ"))

## Define survey design

# Define survey design for overall dataset
NHANES\_all <- svydesign(data=One, id=~SDMVPSU, strata=~SDMVSTRA, weights=~WTMEC4YR, nest=TRUE)

# Create a survey design object for the subset of interest: adults aged 20 and over with a valid depression score
# Subsetting the original survey design object ensures we keep the design information about the number of clusters and strata
NHANES <- subset(NHANES\_all, inAnalysis)

## Analysis

# Define a function to call svymean and unweighted count
getSummary <- function(varformula, byformula, design){
 # Get mean, stderr, and unweighted sample size
 c <- svyby(varformula, byformula, design, unwtd.count )
 p <- svyby(varformula, byformula, design, svymean )
 outSum <- left\_join(select(c,-se), p)
 outSum
}

### Calculate prevalence of depression overall, by gender, by age group, and by age and gender

Adults

getSummary(~Depression, ~one, NHANES)

## Joining, by = "one"

## one counts Depression se
## 1 1 9942 8.056844 0.3599894

By sex

getSummary(~Depression, ~Gender, NHANES)

## Joining, by = "Gender"

## Gender counts Depression se
## 1 Men 4821 5.549344 0.4293217
## 2 Women 5121 10.427654 0.5658239

By age

getSummary(~Depression, ~Age.Group, NHANES)

## Joining, by = "Age.Group"

## Age.Group counts Depression se
## 1 20-39 3328 7.744613 0.5236944
## 2 40-59 3307 8.429826 0.6164284
## 3 60 and over 3307 7.971216 0.7797954

By sex and age

getSummary(~Depression, ~Gender + Age.Group, NHANES)

## Joining, by = c("Gender", "Age.Group")

## Gender Age.Group counts Depression se
## 1 Men 20-39 1654 5.513778 0.6461045
## 2 Women 20-39 1674 10.050321 0.8036891
## 3 Men 40-59 1556 5.222060 0.7699895
## 4 Women 40-59 1751 11.477238 1.2011361
## 5 Men 60 and over 1611 6.052782 0.8295114
## 6 Women 60 and over 1696 9.579923 1.0534115

### Compare Prevalence Between Men And Women

svyttest(Depression~Gender, NHANES)$p.value %>% as.numeric

## [1] 1.706236e-07

svyttest(Depression~Gender, subset(NHANES, Age.Group=="20-39"))$p.value %>% as.numeric

## [1] 0.0001131167

svyttest(Depression~Gender, subset(NHANES, Age.Group=="40-59"))$p.value %>% as.numeric

## [1] 0.0005705859

svyttest(Depression~Gender, subset(NHANES, Age.Group=="60 and over"))$p.value %>% as.numeric

## [1] 0.003983706

### Pairwise t-testing by age groups for total, men, and women

Differences by age group, among all adults

# Full output from svyttest command
svyttest(Depression~Age.Group, subset(NHANES, Age.Group=="20-39" | Age.Group=="40-59"))

##
## Design-based t-test
##
## data: Depression ~ Age.Group
## t = 0.79398, df = 29, p-value = 0.4337
## alternative hypothesis: true difference in mean is not equal to 0
## 95 percent confidence interval:
## -1.006251 2.376677
## sample estimates:
## difference in mean
## 0.6852129

# Displaying p-values only
svyttest(Depression~Age.Group, subset(NHANES, Age.Group=="20-39" | Age.Group=="40-59"))$p.value %>% as.numeric

## [1] 0.433655

svyttest(Depression~Age.Group, subset(NHANES, Age.Group=="20-39" | Age.Group=="60 and over"))$p.value %>% as.numeric

## [1] 0.8245736

svyttest(Depression~Age.Group, subset(NHANES, Age.Group=="40-59" | Age.Group=="60 and over"))$p.value %>% as.numeric

## [1] 0.6001652

Differences by age group, among men

svyttest(Depression~Age.Group, subset(NHANES, Gender=="Men" & (Age.Group=="20-39" | Age.Group=="40-59")))$p.value %>% as.numeric

## [1] 0.7927599

svyttest(Depression~Age.Group, subset(NHANES, Gender=="Men" & (Age.Group=="20-39" | Age.Group=="60 and over")))$p.value %>% as.numeric

## [1] 0.5938032

svyttest(Depression~Age.Group, subset(NHANES, Gender=="Men" & (Age.Group=="40-59" | Age.Group=="60 and over")))$p.value %>% as.numeric

## [1] 0.4339905

Differences by age group, among women

svyttest(Depression~Age.Group, subset(NHANES, Gender=="Women" & (Age.Group=="20-39" | Age.Group=="40-59")))$p.value %>% as.numeric

## [1] 0.3508035

svyttest(Depression~Age.Group, subset(NHANES, Gender=="Women" & (Age.Group=="20-39" | Age.Group=="60 and over")))$p.value %>% as.numeric

## [1] 0.7530381

svyttest(Depression~Age.Group, subset(NHANES, Gender=="Women" & (Age.Group=="40-59" | Age.Group=="60 and over")))$p.value %>% as.numeric

## [1] 0.2201892