

P8130 Assignment 1

Emil Hafeez (eh2928)

9/18/2020

Assignment 1 for Biostatistics P8130, using a dataset called Antibodies, practices data manipulation and descriptive statistics (including table, histogram, and boxplot generation). The dataset contains information on demographic variables, IgM antibodies, and self-reported smell loss for patients diagnosed with COVID-19 (via the PCR gold-standard).

Let's load the appropriate packages first.

```
library(arsenal)
library(dplyr)
library(ggplot2)
library(tidyverse)
```

NOTE LETS TRY THIS TO SHORTEN THE CODE RUNOFFS IN THE KNITTED PDF LATER

```
opts_chunk$set(tidy.opts=list(width.cutoff=80),tidy=TRUE) #####
```

Question 1

Question 1, Part 1, Section a)

First, read the CSV and examine the basics.

```
#Read the CSV data into a dataframe
antibodies_df <- read.csv("/Users/emilhafeez/Google Drive/Columbia/Fall 2020/Classes/Biostatistics 1/As
```

I examine the basics of the data, to become familiar. Results are hidden for the sake of brevity.

```
names(antibodies_df)
nrow(antibodies_df)
ncol(antibodies_df)
head(antibodies_df)
tail(antibodies_df)
anyNA(antibodies_df)
str(antibodies_df)
```

Now, we can provide descriptive statistics for all variables of interest. We can see that some variable types need manipulation, too. I suppress the code output here.

```
summary(antibodies_df)

#Make AgeCategory, Smell, and Gender to show N (%). Requires ensuring all unique values, too.
unique(antibodies_df$AgeCategory)
antibodies_df <- antibodies_df %>%
  mutate(AgeCategory = factor(AgeCategory, labels = c("18-30", "31-50", "51+") ))
unique(antibodies_df$Smell)
antibodies_df <- antibodies_df %>%
```

```

mutate(Smell = factor(Smell, labels = c("Normal", "Altered", "Other")))
unique(antibodies_df$Gender)
antibodies_df <- antibodies_df %>%
  mutate(Gender = factor(Gender, labels = c("Male", "Female")))

#To prepare a tidy table, let's change variable names/labels by creating a my_labels variable that will
my_labels <- list(AgeCategory = "Age Range (Years)", Antibody_IgM = "Antibody Levels(IgM)", Smell = "Smell Changes (Self Report)")

#Clean the output by creating a my_controls variable that will modify the tableby() command
my_controls <- tableby.control(
  total = TRUE,
  test = FALSE,
  numeric.stats = c("meansd", "medianq1q3", "range", "Nmiss2"),
  cat.stats = c("countpct", "Nmiss2"),
  stats.labels = list(
    meansd = "Mean (SD)",
    medianq1q3 = "Median (Q1, Q3)",
    range = "Min - Max",
    Nmiss2 = "Missing",
    countpct = "N (%)")

#Making a table
table_1 <- tableby( ~ AgeCategory + Antibody_IgM + Smell + Gender, data = antibodies_df, control = my_controls,
summary(table_1, title = "Descriptive Statistics: Antibodies & Smell Data", labelTranslations = my_labels)

```

```

##
## Table: Descriptive Statistics: Antibodies & Smell Data
##
## | | Overall (N=1491) |
## |-----|:-----:|
## |Age Range (Years) |
## |- 18-30 | 318 (21.3%) |
## |- 31-50 | 810 (54.3%) |
## |- 51+ | 363 (24.3%) |
## |- Missing | 0 |
## |Antibody Levels(IgM) |
## |- Mean (SD) | 0.124 (0.110) |
## |- Median (Q1, Q3) | 0.091 (0.069, 0.129) |
## |- Min - Max | 0.048 - 1.048 |
## |- Missing | 1224 |
## |Smell Changes (Self Report) |
## |- Normal | 1047 (70.2%) |
## |- Altered | 410 (27.5%) |
## |- Other | 34 (2.3%) |
## |- Missing | 0 |
## |Gender (M/F) |
## |- Male | 981 (65.8%) |
## |- Female | 510 (34.2%) |
## |- Missing | 0 |
table_1

```

```

## tableby Object
##
## Function Call:

```

Question 1, Part 1, Section b)

Question 1, Part 2, Section a)

```
table_2 <- tableby( Smell ~ AgeCategory + Antibody_IgM + Smell + Gender, data = antibodies_df, control = list(
summary(table_2, title = "Descriptive Statistics: Antibodies & Smell Data", labelTranslations = my_label_translations)

```

3