

Lab 4

Sydney Stitt

Math 241, Week 4

```
# Put all necessary libraries here  
library(tidyverse)  
library(readr)
```

Problem 1: COVID survey - interpretation

Based on the plot, the majority of respondents have a positive opinion of the vaccine. For example, for “all” respondents, the majority believe the vaccine will make them feel safer, have confidence in the scientific vetting process, trust the related information, have and would recommend the vaccine to their family. The 10th and 90th percentile error bars for those four response variables show a small range with answers concentrated at a likert score of 1 or 2, meaning respondents vary from somewhat agree to strongly agree. I don’t think using 10 and 90 as percentile limits for error bars is the most practical, as typically 25 and 75 are the most conventionally used percentiles as part of the interquartile range. I think the 25 and 75 percentiles give a better understanding of the distribution of responses than 10 and 90. This makes sense, as vaccines are typically well accepted in public opinion, and people tend to trust medical researchers to produce effective vaccines.

There is a large variation in responses for whether respondents believe the vaccine is safe and are concerned with the safety and side effects. The error bars range from 1 to 5 for both responses, showing large variability as that is the entire response range. This makes sense, as there is a lot of news media circulating on the side effects and safety that can cause fear mongering among the general public.

The y axis divides respondents into groups based on their demographics, such as ethnicity, race, gender, age, profession, etc. This is to see if demographics play an impact in whether one would trust a vaccine and what influences their opinions on vaccines. Interestingly, the profession group is split between nursing and medical professionals, and intuitively it is strange that there is more variation in medical professional’s confidence in the scientific vetting process (average is roughly 1.5, error bars at 1 and 3) than the variation in all responses (average is roughly 1.5, error bars at 1 and 3). Intuitively, I would think that medical professionals would have more trust in the scientific process, as many medical professionals are likely more knowledgeable of the process than the general public.

Problem 2: COVID survey - reconstruct

```
#Reading the survey data  
covid_survey <- read_csv("/Users/sydneystitt/Desktop/MATH241/Class Repo/labs/lab04/data/covid-survey.csv")  
print(dim(covid_survey))
```

```
## [1] 1121 14
```

```
#Tidying up survey data
covid_survey <- covid_survey %>% filter(if_all(everything(), ~ !is.na(.)))
print(dim(covid_survey))
```

```
## [1] 926 14
```

```
#Recoding observation values
covid_survey <- covid_survey %>%
  mutate(exp_already_vax = recode(exp_already_vax, "0" = "No",
                                   "1" = "Yes"))
covid_survey <- covid_survey %>%
  mutate(exp_flu_vax = recode(exp_flu_vax, "0" = "No",
                                   "1" = "Yes"))
covid_survey <- covid_survey %>%
  mutate(exp_profession = recode(exp_profession, "0" = "Medical",
                                   "1" = "Nursing"))
covid_survey <- covid_survey %>%
  mutate(exp_gender = recode(exp_gender, "0" = "Male",
                                   "1" = "Female",
                                   "3" = "Non-binary third gender",
                                   "4" = "Prefer not to say"))
covid_survey <- covid_survey %>%
  mutate(exp_race = recode(exp_race, "1" = "American Indian/Alaskan Native",
                                   "2" = "Asian",
                                   "3" = "Black/African American",
                                   "4" = "Native Hawaiian/Other Pacific Islander",
                                   "5" = "White"))
covid_survey <- covid_survey %>%
  mutate(exp_ethnicity = recode(exp_ethnicity, "1" = "Hispanic/Latino",
                                   "2" = "Non-Hispanic/Non-Latino"))
covid_survey <- covid_survey %>%
  mutate(exp_age_bin = recode(exp_age_bin, "0" = "<20",
                                   "20" = "21-25",
                                   "25" = "26-30",
                                   "30" = ">30"))
print(dim(covid_survey))
```

```
## [1] 926 14
```

```
#Pivoting the data
covid_survey_longer <- covid_survey %>%
  pivot_longer(
    cols = starts_with("exp_"),
    names_to = "explanatory",
    values_to = "explanatory_value"
  ) %>%
  filter(!is.na(explanatory_value)) %>%
  pivot_longer(
    cols = starts_with("resp_"),
    names_to = "response",
    values_to = "response_value"
  )
```

The `pivot_longer` statement is consolidating the explanatory and response variables into a single explanatory and single response variable with the original variable titles being the new values for the new explanatory and new response variables. Thus this combines the many variables into one variable that is easier to work with with the old variables representing the new levels.

```
#Making the summary statistics table by group
covid_survey_summary_stats_by_group <- covid_survey_longer %>%
  group_by(explanatory, explanatory_value, response) %>%
  summarize(mean = mean(response_value),
             low = quantile(response_value,
                             probs = 0.1,
                             na.rm = TRUE),
             high = quantile(response_value,
                              probs = 0.9,
                              na.rm = TRUE))

#Making summary statistics by response
covid_survey_summary_stats_all <- covid_survey_longer %>%
  group_by(response) %>%
  summarize(mean = mean(response_value),
             low = quantile(response_value,
                             probs = 0.1,
                             na.rm = TRUE),
             high = quantile(response_value,
                              probs = 0.9,
                              na.rm = TRUE))

explanatory <- c("All", "All", "All", "All", "All", "All")
explanatory_value <- c("~", "~", "~", "~", "~", "~")
covid_survey_summary_stats_all$explanatory <- explanatory
covid_survey_summary_stats_all$explanatory_value <- explanatory_value

#Merging the two tables into one
covid_survey_summary_stats <- full_join(covid_survey_summary_stats_by_group, covid_survey_summary_stats_all)
covid_survey_summary_stats
```

```
## # A tibble: 132 x 6
## # Groups:   explanatory, explanatory_value [22]
##   explanatory explanatory_value response      mean    low  high
##   <chr>         <chr>          <chr>    <dbl> <dbl> <dbl>
## 1 exp_age_bin 21-25      resp_concern_safety 3.32     2     5
## 2 exp_age_bin 21-25      resp_confidence_science 1.30     1     2
## 3 exp_age_bin 21-25      resp_feel_safe_at_work 1.18     1     2
## 4 exp_age_bin 21-25      resp_safety 1.97     1     5
## 5 exp_age_bin 21-25      resp_trust_info 1.29     1     2
## 6 exp_age_bin 21-25      resp_will_recommend 1.09     1     1
## 7 exp_age_bin 26-30      resp_concern_safety 3.32     1     5
## 8 exp_age_bin 26-30      resp_confidence_science 1.39     1     2
## 9 exp_age_bin 26-30      resp_feel_safe_at_work 1.27     1     2
## 10 exp_age_bin 26-30      resp_safety 2.17     1     5
## # i 122 more rows
```

```
#Creating labels
response_labels <- c(
```

```

resp_safety = "Based on my \nunderstanding, I \nbelieve the vaccine \nis safe",
resp_confidence_science = "I am confident in \nthe scientific \nvetting process for \nthe new COVID \n",
resp_feel_safe_at_work = "Getting the vaccine \nwill make me feel \nsafer at work",
resp_will_recommend = "I will recommend \nthe vaccine to \nfamily, friends, \nand community \nmembers",
resp_trust_info = "I trust the \ninformation that I \nhave received about \nthe vaccines",
resp_concern_safety = "I am concerned \nabout the safety \nand side effects of \nthe vaccine"
)
explanatory_labels <- c(
  exp_age_bin = "Age",
  exp_gender = "Gender",
  exp_race = "Race",
  exp_ethnicity = "Ethnicity",
  exp_profession = "Profession",
  exp_already_vax = "Had COVID vaccine",
  exp_flu_vax = "Had flu vaccine this year",
  All = "All"
)
facet_labels <- labeller(explanatory = explanatory_labels,
  response = response_labels)

```

```

#Recreating the graph
ggplot(covid_survey_summary_stats, aes(x = mean, y = explanatory_value)) +
  geom_line() +
  geom_errorbar(aes(xmin = low, xmax = high)) +
  geom_point() +
  theme(axis.line = element_line(color = "white"),
    strip.text.y.right = element_text(angle = 0)) +
  facet_grid(explanatory ~ response,
    scales = "free",
    space = "free",
    labeller = facet_labels,
    drop = TRUE) +
  theme_classic(base_size = 7) +
  labs(x = "Mean Likert Score \n(Error bars range from 10th to 90th percentile)",
    y = "")

```

