SOCI 385 - Social Statistics

Fall 2021

Assignment 6

This assignment is due via Canvas by 10:00 AM on Monday, November 8, 2021.

Please show all your code and write your responses in complete sentences as appropriate.

This week you will continue practicing with significance tests for difference in means and proportions. The assignment requires the GSS data you began to clean in class on Wednesday. You are free (and encouraged!) to use the R shortcuts we learned in class. ***Use an alpha level of .05 for all the tests.***

The variables are:

`year`: The year the survey was completed, either 2008 or 2018

`hrs1`: Self-reported hours worked last week, from 1-89

`health`: Self-reported condition of health, coded "Excellent" "Good", "Fair", or "Poor"

`class`: Self-reported social class, coded "Working class", "Lower class", "Middle class", or "Upper class"

`sex`: Self-reported sex, coded "Female" or "Male"

Questions

```
library(tidyverse)
library(gssr)
```

1. Finish cleaning the variables you downloaded from the GSS. The `notebook_gss_cleanup.Rmd` file should help you get started. Note that you have to clean the `health` and `class` variables on your own.

```
load("../../data/gss_cleanup_example.RData")
```

```
data(gss_doc)
```

```
gss_get_marginals(c("health", "class"), gss_doc)
```

```
## # A tibble: 17 x 6
##    variable percent     n value label         id
##    <chr>      <dbl> <int> <chr> <chr>         <chr>
##  1 health      29.8 14186 1     EXCELLENT     HEALTH
##  2 health      45.3 21559 2     GOOD          HEALTH
##  3 health      19.2  9123 3     FAIR          HEALTH
##  4 health       5.7  2722 4     POOR          HEALTH
##  5 health      NA   17099 0     IAP           HEALTH
##  6 health      NA      35 8     DK            HEALTH
##  7 health      NA      90 9     <NA>          HEALTH
##  8 health     100   64814 <NA>  Total         HEALTH
##  9 class        6.3  3872 1     LOWER CLASS   CLASS
## 10 class       45.6 27968 2     WORKING CLASS CLASS
## 11 class       44.9 27519 3     MIDDLE CLASS  CLASS
```

```
## 12 class        3.2  1971 4     UPPER CLASS   CLASS
## 13 class        0       1 5     NO CLASS      CLASS
## 14 class       NA    3064 0     IAP           CLASS
## 15 class       NA     185 8     DK            CLASS
## 16 class       NA     234 9     <NA>          CLASS
## 17 class      100   64814 <NA>  Total         CLASS
```

```
gss_subset <- gss_subset |>
  mutate(health = factor(health,
                         labels = c("Excellent", "Good",
                                    "Fair", "Poor")),
         class = factor(class,
                        labels = c("Lower", "Working",
                                   "Middle", "Upper")))
```

2. Is there a significant difference in mean hours worked last week between respondents identifying with the lower class and respondents identifying with the working class? Why or why not?

```
t.test(gss_subset$hrs1[gss_subset$class=="Lower"],
       gss_subset$hrs1[gss_subset$class=="Working"])
```

```
##
##  Welch Two Sample t-test
##
## data:  gss_subset$hrs1[gss_subset$class == "Lower"] and gss_subset$hrs1[gss_subset$class == "Working
## t = -4.0901, df = 138.41, p-value = 7.28e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.915424 -3.104682
## sample estimates:
## mean of x mean of y
##  36.28099  42.29104
```

3. Is there a significant difference in mean hours worked last week between respondents identifying with the lower class in the 2008 survey and respondents identifying with the lower class in the 2018 survey? Why or why not?

```
t.test(gss_subset$hrs1[gss_subset$class=="Lower" & gss_subset$year==2008],
       gss_subset$hrs1[gss_subset$class=="Lower" & gss_subset$year==2018])
```

```
##
##  Welch Two Sample t-test
##
## data:  gss_subset$hrs1[gss_subset$class == "Lower" & gss_subset$year == 2008] and gss_subset$hrs1[gss
## t = 0.54403, df = 110.52, p-value = 0.5875
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.026162  7.073315
## sample estimates:
## mean of x mean of y
##  37.21277  35.68919
```

4. Is there a significant difference in the 2018 survey in the proportion of female respondents who report having excellent or good health (vs fair or poor health) and the proportion of male respondents who report having excellent or good health (vs fair or poor health)? Why or why not?

```
q4 <- gss_subset |>
  filter(year == 2018) |>
  mutate(health_hi = ifelse(health == "Excellent" | health == "Good", 1, 0))
```

```
q4_table <- table(q4$sex, q4$health_hi)
prop.test(q4_table)
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  q4_table
## X-squared = 1.689, df = 1, p-value = 0.1937
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.01520217  0.07706626
## sample estimates:
##    prop 1    prop 2
## 0.2969871 0.2660550
```

5. Is there a significant difference in the 2018 survey in the proportion of working class respondents who report having excellent health (vs fair or poor health) and the proportion of middle class respondents who report having excellent health (vs fair or poor health)? Why or why not?

```
q5 <- q4 |>
  filter(class == "Working" | class == "Middle")  |>
  droplevels()
```

```
q5_table <- table(q5$class, q5$health_hi)
```

```
prop.test(q5_table)
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  q5_table
## X-squared = 9.2114, df = 1, p-value = 0.002405
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.0256697 0.1206125
## sample estimates:
##    prop 1    prop 2
## 0.2951977 0.2220566
```