

Chapters covered:

1. Ch 9.3 Hypothesis test for mean (one sample t-test)
2. Ch 10.3 Independent two sample t-test

Assume that the **Fall2022survey** is a random sample drawn from the population of all U of M students. Use the following code to access data from the survey.

```
survey_f22 <- read.csv("http://users.stat.umn.edu/~parky/Fall2022Survey.csv", header=TRUE)
```

Problem 1 (Same as Lab 6 Problem 2)

We want to test if the average time students at the U spend per day on social media is less than 4 hours. We will be working with the variable `HoursSpentOnSocialMedia`.

- a) Let μ denote the average time U of M students spend per day on social media. Test claim B. Write down all five steps.
- b) Now repeat the analysis in R using the `t.test()` function. You only need to show the R commands and the output. Does the result agree with what you have calculated previously?
- c) What type of error have you possibly made? Explain.
- d) Construct a 95% confidence interval for μ . Can you use this confidence interval to draw a conclusion for the hypothesis test? What if we want to test whether or not μ is equal to 4 (meaning that $H_a : \mu \neq 4$)?

Problem 2

Now, we want to test if there is any significant difference between mean hours spent on social media for freshmen and for sophomore students.

- a) Use R command below to extract freshmen and sophomore's social media hours.

```
social.fr<-survey_f22$HoursSpentOnSocialMedia[survey_f22$Year=="Freshman"]
#extracting Freshmen's social media hours
social.so<-survey_f22$HoursSpentOnSocialMedia[survey_f22$Year=="Sophomore"]
```

- b) Use R command below to construct a side-by-side boxplot. Do you think Freshmen and Senior mean hours spent on social media are different?

```
#boxplot(group1 quantitative variable, group2 quantitative variable)
boxplot(social.fr, social.so, names=c("Freshmen", "Sophomore"), main ="Hours spent on social media")
```

- c) Are two samples are independent or matched pairs?
- d) Find the standard error of difference between sample means, $SE_{\bar{X}_1 - \bar{X}_2}$. Explain what this number means.

- e) Use `t.test()` command to test if there is any significant difference between mean hours spent on social media for freshmen and for sophomore students at $\alpha = 0.05$. Use p-value and draw a conclusion.

```
t.test(x=_____, y=_____, conf.level=_____, alternative="_____")
```

- f) Use the same R output from the previous part and interpret the 95% confidence interval. Does it agree with the conclusion from the hypothesis test?