Lab Handout Week 11

Problem 1: What value of $z_{\alpha/2}$ or $t_{\alpha/2,df}$ is used to construct:

a. a 92% confidence interval to estimate $p_1 - p_2$ if the number of successes is 100 and 30 and the number of failures is 70 and 30 in each random sample.

b. a 95% confidence interval to estimate $\mu_1 - \mu_2$, difference in two population means and two samples are independent. The sample size for sample 1 is 23 and for sample 2 is 41. (Assume random sample assumption and normal population distribution assumption are met.)

c. a 93% confidence interval to estimate μ_D (mean of difference within pairs). There are 30 matched pairs.

d. a 98% confidence interval to estimate μ if the sample size is 1982. (Assume random sample assumption is met.)

Problem 2. Time spent on social network

For this problem, load and attach the **Getting To know you Survey** for Fall 2019 using the following code:

```
NoU<-read.csv ("http://users.stat.umn.edu/~parky/SurveyFall2019.csv")
#Check names of NoU
names(NoU)
```

We assume that the survey represents a random sample from the population of all U of M students. Consider the problem of constructing a 98% confidence interval for μ, the population mean hours per day spent on social media for Freshmen students. Use the following R command to define a new variable `social.Fr`. (social network hours for Freshmen only).

```
social.Fr<-NoU$hours.social.networks[NoU$year=="Freshman"]
```

1. Explore the distribution of `social.networks.hrs` graphically using either the `hist()` command. Describe the shape of the data.

```
hist(social.Fr, xlim=c(0, 18), breaks=20,
  main="Histogram of hours per day spent on social media for Freshmen", xlab="Hours per
    day")
```

In the R command above, option `xlim=c(0, 18)` sets the x-axis limit from 0 to 18 hours. `breaks=20` specifies the number of bars in the histograms. `main` gives the title of the plot and `xlab` gives the x-axis title.

2. Construct boxplot and find the 5 number summary to identify the interval where the middle 50% of the distribution falls within.

```
boxplot(social.Fr, ylim=c(0, 20))
summary(social.Fr)
```

3. Construct Q-Q plot using `qqnorm()` `qqline()` as well. Q-Q plot compares the data distribution to standard normal distribution. If most of dots are along the straight line, we can conclude that the sample comes from a normal population distribution.

```
qqnorm(social.Fr) ##plot
qqline(social.Fr) ##add a straight line over the existing plot
```

4. Based on your plots, does `social.Fr` appear to follow a normal distribution? If not, briefly explain why you can still construct the confidence interval for μ using large sample confidence interval method.

5. Construct a 98% confidence interval to estimate μ. Interpret the result. (Do not use t.test)

6. Use `t.test()` to construct the 98% confidence interval for μ. Compare the result with your answer from (h). Use `?t.test` in console to learn more details.

```
t.test(social.Fr, conf.level = 0.98, alternative="two.sided")
```

7. Now, we want to compare mean hours spent on social network for Freshmen and for Senior students. Use R command below to construct a side-by-side boxplot. Do you think Freshmen and Senior mean hours spent on social network are different? Do you think two population have equal variance?

```
social.Sr<- NoU$hours.social.networks[NoU$year == "Senior"]
boxplot(social.Fr, social.Sr, names=c("Freshmen", "Senior"), main="Time spent on social
    network")
```

8. Use the following command below to estimate $\mu_{Fr} - \mu_{Sr}$ where μ_{Fr} represents the mean hours spent on social media for Freshmen and μ_{Sr} represents the mean hours spent on social media for Senior students. Interpret the interval.

```
t.test(x=social.Fr, y=social.Sr, conf.level=0.98, alternative="two.sided")
```