HW #4: t-tests

Eddie Chapman

10/1/19

1. A psychologist would like to know how many casual friends are in the average person's social network. She interviews a random sample of people and determines for each the number of friends or social acquaintances they see or talk to at least once a year. The data are as follows: 5, 11, 15, 9, 7, 13, 23, 8, 12, 7, 10, 11, 21, 20, 13. (To save you time, $\bar{X} = 12.33$ and $s = 5.38$.)

   a. Find the 95% confidence interval for the mean (μ) of the population.

$$CI_{.95} = \bar{X} \pm (t_{0.5/2})(s_{\bar{X}}) = \bar{X} \pm (t_{.05/2})(\frac{s}{\sqrt{n}})$$

$$CI_{.95} = 12.33 \pm 2.15(\frac{5.38}{\sqrt{15}}) = 12.33 \pm 2.98$$

$$\hat{\mu}_{lower} = 9.35 \quad \hat{\mu}_{upper} = 15.31$$

   b. If a previous researcher had predicted the population mean of 10 casual friends per person, could that prediction be rejected as a hypothesis at the .05 level, two-tailed? How do you know?

   Our confidence interval quantifies the variability around the point estimate. We can be fairly certain that the range of values between the upper and lower bounds contains the true population mean. Because the predicted point estimate of 10 is located within our confidence interval, it cannot be rejected as a hypothesis at the .05 level.

2. Use `R` to check your answer to #9(a). Include your code and output, and indicate where you can find the relevant values.

```r
friend_counts <- c(5, 11, 15, 9, 7, 13, 23, 8, 12, 7, 10, 11, 21, 20, 13)
t.test(friend_counts, mu = mean(friend_counts), conf.level = .95)
```

```
##
##  One Sample t-test
##
## data:  friend_counts
## t = 0, df = 14, p-value = 1
## alternative hypothesis: true mean is not equal to 12.33333
## 95 percent confidence interval:
##   9.353578 15.313089
## sample estimates:
## mean of x
##  12.33333
```

3. The data file "speed.csv" contains the speeds (in miles per hour) for 25 individuals driving on the freeway. Test the null hypothesis at the $\alpha = .05$ level (two-tailed) that the mean of the population from which these data were drawn is 55 mph. What are the point estimate and 95% confidence interval of the population mean? Use `R` to answer this question, including relevant code and output.

```r
data <- read.csv('speed.csv', header = T)
t.test(data$Speed, mu = 55, conf.level = .95)
```

```
##
##  One Sample t-test
##
## data:  data$Speed
## t = 3.3104, df = 24, p-value = 0.002937
## alternative hypothesis: true mean is not equal to 55
## 95 percent confidence interval:
##  57.83154 67.20846
## sample estimates:
## mean of x
##     62.52
```

We can reject the null hypothesis that μ is 55 mph because it does not fall within the 95 percent confidence interval of 57.83-67.21. The point estimate for the population mean is 62.52 mph.

4. An educational psychologist is interested in whether a test prep program increases students' verbal GRE scores. Students take the verbal GRE before the test prep program, and then they take the test again after the test prep program. The data are recorded below:

Verbal GRE (1)	Verbal GRE (2)
540	570
510	520
580	600
550	530
520	520

a. Why is a related samples t-test appropriate to address this research question?

The two sets of scores are not independent of each other. They can be grouped into matched pairs in order to test the difference between them using a related samples t-test.

b. By hand, conduct a related samples t-test to determine whether the test prep program increases verbal GRE scores. (You may use `R` to calculate the mean and variance of the difference scores.) Use $\alpha = .05$ and do a **one-tailed** test.

```r
gre1 <- c(540, 510, 580, 550, 520)
gre2 <- c(570, 520, 600, 530, 520)
diff <- gre2 - gre1
n <- length(diff)
df <- n - 1
diff_mean <- mean(diff)
sd <- sd(diff)
std_err <- sd / sqrt(n)
```

$\bar{D} = \texttt{diff\_mean} = 8$
$s_{\bar{D}} = \texttt{std\_err} = 8.6023253$
$n = \texttt{n} = 5$
$df = \texttt{df} = 4$
$H_0 : \mu_D = 0, \quad H_1 : \mu_D \neq 0$
$\mu_{\bar{D}} = 0$
$t_{crit}(4) = 2.132$
$t = \frac{\bar{D} - \mu_{\bar{D}}}{s_{\bar{D}}} = \texttt{(diff\_mean - 0)/std\_err} = 0.9299811$

Our test statistic (0.93) does not exceed the critical value (2.132) so we cannot reject the null hypothesis that the test prep program does has no significant effect on the distribution of verbal GRE scores.

c. Construct a 95% confidence interval for $\hat{\mu}_D$. (Use the two-tailed critical value.)

$$CI_{.95} = \bar{D} \pm (t_{0.5/2})(s_{\bar{D}}) = \bar{D} \pm (t_{.05/2})(\frac{s}{\sqrt{n}})$$

$$CI_{.95} = 8 \pm 2.776 * 8.6 = 8 \pm 23.87$$

$$\hat{\mu}_{lower} = -15.87 \quad \hat{\mu}_{upper} = 31.87$$

d. Regardless of the outcome of the hypothesis test, if you were to compute Cohen's d, what value would make the most sense to use as the denominator (i.e., the standard deviation)? Why?

It would be best to use the standard deviation of the pre-treatment scores to provide a baseline to compare against the effect of the treatment.

5. Double check your hand calculations from #4 by inputting the data into `R` and conducting the appropriate test. Include any relevant code and output.

```r
gre1 <- c(540, 510, 580, 550, 520)
gre2 <- c(570, 520, 600, 530, 520)
t.test(gre2, gre1, paired = T, conf.level = .95)
```

```
##
##  Paired t-test
##
## data:  gre2 and gre1
## t = 0.92998, df = 4, p-value = 0.405
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -15.88388  31.88388
## sample estimates:
## mean of the differences
##                       8
```

6. A psychologist studying the dynamics of marraige wanted to know how many hours per week the average American couple spends discussing marital problems. The sample mean ($\bar{X}$) of 155 randomly selected couples turned out to be 2.6 hours, with $s = 1.8$.

a. Find the 95% confidence interval for $\hat{\mu}$.

$$CI_{.95} = \bar{X} \pm (z_{0.5/2})(s_{\bar{X}}) = \bar{X} \pm (z_{.05/2})(\frac{s}{\sqrt{N}})$$

$$CI_{.95} = 2.6 \pm 1.96 * (\frac{1.8}{155}) = 2.6 \pm 0.28$$

$$\hat{\mu}_{lower} = 2.3166 \quad \hat{\mu}_{upper} = 2.8834$$

b. A European study had already estimated the population mean to be 3 hours per week for European couples. Are the American couples significantly different from the European couples at the $\alpha = .05$ level? Explain how your answer to part (a) makes it easy to answer part (b) without needing to do any additional calculations.

Yes, the American couples are significantly different from the European couples because the sample mean of the European couples (3) falls outside of the 95% confidence interval calculated around the estimated population mean (2.32 - 2.88).