

Chapter 9: Two-Sample Inference

Chapter 7 discussed methods of hypothesis testing about one-population parameters. Chapter 8 discussed methods of estimating population parameters from one sample using confidence intervals. This chapter will look at methods of confidence intervals and hypothesis testing for two populations. Since there are two populations, there are two random variables, two means or proportions, and two samples (though with paired samples you usually consider there to be one sample with pairs collected). Examples of where you would do this are:

Testing and estimating the difference in testosterone levels of men before and after they had children (Gettler, McDade, Feranil & Kuzawa, 2011).

Testing the claim that a diet works by looking at the weight before and after subjects are on the diet.

Estimating the difference in proportion of those who approve of President Obama in the age group 18 to 26 year olds and the 55 and over age group.

All of these are examples of hypothesis tests or confidence intervals for two populations. The methods to conduct these hypothesis tests and confidence intervals will be explored in this method. As a reminder, all hypothesis tests are the same process. The only thing that changes is the formula that you use. Confidence intervals are also the same process, except that the formula is different.

Section 9.1 Two Proportions

There are times you want to test a claim about two population proportions or construct a confidence interval estimate of the difference between two population proportions. As with all other hypothesis tests and confidence intervals, the process is the same though the formulas and assumptions are different.

Hypothesis Test for Two Population Proportion (2-Prop Test)

1. State the random variables and the parameters in words.

x_1 = number of successes from group 1

x_2 = number of successes from group 2

p_1 = proportion of successes in group 1

p_2 = proportion of successes in group 2

2. State the null and alternative hypotheses and the level of significance

$H_o : p_1 = p_2$ or $H_o : p_1 - p_2 = 0$

$H_A : p_1 < p_2$ $H_A : p_1 - p_2 < 0$

$H_A : p_1 > p_2$ $H_A : p_1 - p_2 > 0$

$H_A : p_1 \neq p_2$ $H_A : p_1 - p_2 \neq 0$

Also, state your α level here.

3. State and check the assumptions for a hypothesis test
- A simple random sample of size n_1 is taken from population 1, and a simple random sample of size n_2 is taken from population 2.
 - The samples are independent.
 - The assumptions for the binomial distribution are satisfied for both populations.
 - To determine the sampling distribution of \hat{p}_1 , you need to show that $n_1 p_1 \geq 5$ and $n_1 q_1 \geq 5$, where $q_1 = 1 - p_1$. If this requirement is true, then the sampling distribution of \hat{p}_1 is well approximated by a normal curve. To determine the sampling distribution of \hat{p}_2 , you need to show that $n_2 p_2 \geq 5$ and $n_2 q_2 \geq 5$, where $q_2 = 1 - p_2$. If this requirement is true, then the sampling distribution of \hat{p}_2 is well approximated by a normal curve. However, you do not know p_1 and p_2 , so you need to use \hat{p}_1 and \hat{p}_2 instead. This is not perfect, but it is the best you can do. Since $n_1 \hat{p}_1 = n_1 \frac{x_1}{n_1} = x_1$ (and similar for the other calculations) you just need to make sure that x_1 , $n_1 - x_1$, x_2 , and $n_2 - x_2$ are all more than 5.

4. Find the sample statistics, test statistic, and p-value

Sample Proportion:

n_1 = size of sample 1

n_2 = size of sample 2

$$\hat{p}_1 = \frac{x_1}{n_1} \text{ (sample 1 proportion)}$$

$$\hat{p}_2 = \frac{x_2}{n_2} \text{ (sample 2 proportion)}$$

$$\hat{q}_1 = 1 - \hat{p}_1 \text{ (complement of } \hat{p}_1)$$

$$\hat{q}_2 = 1 - \hat{p}_2 \text{ (complement of } \hat{p}_2)$$

Pooled Sample Proportion, \bar{p} :

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\bar{q} = 1 - \bar{p}$$

Test Statistic:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

Usually $p_1 - p_2 = 0$, since $H_o: p_1 = p_2$

p-value:

On TI-83/84: use normalcdf(lower limit, upper limit, 0, 1)

(Note: if $H_A: p_1 < p_2$, then lower limit is $-1E99$ and upper limit is your test statistic. If $H_A: p_1 > p_2$, then lower limit is your test statistic and the upper limit is $1E99$. If $H_A: p_1 \neq p_2$, then find the p-value for

$H_A: p_1 < p_2$, and multiply by 2.)

On R: use `pnorm(z, 0, 1)`

(Note: if $H_A : p_1 < p_2$, then use `pnorm(z, 0, 1)`. If $H_A : p_1 > p_2$, then use `1 - pnorm(z, 0, 1)`. If $H_A : p_1 \neq p_2$, then find the p-value for $H_A : p_1 < p_2$, and multiply by 2.)

5. Conclusion

This is where you write reject H_o or fail to reject H_o . The rule is: if the p-value $< \alpha$, then reject H_o . If the p-value $\geq \alpha$, then fail to reject H_o .

6. Interpretation

This is where you interpret in real world terms the conclusion to the test. The conclusion for a hypothesis test is that you either have enough evidence to show H_A is true, or you do not have enough evidence to show H_A is true.

Confidence Interval for the Difference Between Two Population Proportion (2-Prop Interval)

The confidence interval for the difference in proportions has the same random variables and proportions and the same assumptions as the hypothesis test for two proportions. If you have already completed the hypothesis test, then you do not need to state them again. If you haven't completed the hypothesis test, then state the random variables and proportions and state and check the assumptions before completing the confidence interval step.

1. Find the sample statistics and the confidence interval

Sample Proportion:

n_1 = size of sample 1

n_2 = size of sample 2

$\hat{p}_1 = \frac{x_1}{n_1}$ (sample 1 proportion)

$\hat{p}_2 = \frac{x_2}{n_2}$ (sample 2 proportion)

$\hat{q}_1 = 1 - \hat{p}_1$ (complement of \hat{p}_1)

$\hat{q}_2 = 1 - \hat{p}_2$ (complement of \hat{p}_2)

Confidence Interval:

The confidence interval estimate of the difference $p_1 - p_2$ is

$$(\hat{p}_1 - \hat{p}_2) - E < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + E$$

where the margin of error E is given by $E = z_c \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$

z_c = critical value

2. Statistical Interpretation: In general this looks like, "there is a C% chance that $(\hat{p}_1 - \hat{p}_2) - E < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + E$ contains the true difference in proportions."
3. Real World Interpretation: This is where you state how much more (or less) the first proportion is from the second proportion.

The critical value is a value from the normal distribution. Since a confidence interval is found by adding and subtracting a margin of error amount from the sample proportion, and the interval has a probability of being true, then you can think of this as the statement $P((\hat{p}_1 - \hat{p}_2) - E < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + E) = C$. So you can use the `invNorm` command on the TI-83/84 calculator or `qnorm` on R to find the critical value. These are always the same value, so it is easier to just look at the table A.1 in the Appendix.

Example #9.1.1: Hypothesis Test for Two Population Proportions

Do husbands cheat on their wives more than wives cheat on their husbands ("Statistics brain," 2013)? Suppose you take a group of 1000 randomly selected husbands and find that 231 had cheated on their wives. Suppose in a group of 1200 randomly selected wives, 176 cheated on their husbands. Do the data show that the proportion of husbands who cheat on their wives are more than the proportion of wives who cheat on their husbands. Test at the 5% level.

Solution:

1. State the random variables and the parameters in words.
 x_1 = number of husbands who cheat on his wife
 x_2 = number of wives who cheat on her husband
 p_1 = proportion of husbands who cheat on his wife
 p_2 = proportion of wives who cheat on her husband
2. State the null and alternative hypotheses and the level of significance
 $H_o : p_1 = p_2$ or $H_o : p_1 - p_2 = 0$
 $H_A : p_1 > p_2$ $H_A : p_1 - p_2 > 0$
 $\alpha = 0.05$
3. State and check the assumptions for a hypothesis test
 - a. A simple random sample of 1000 responses about cheating from husbands is taken. This was stated in the problem. A simple random sample of 1200 responses about cheating from wives is taken. This was stated in the problem.
 - b. The samples are independent. This is true since the samples involved different genders.
 - c. The properties of the binomial distribution are satisfied in both populations. This is true since there are only two responses, there are a fixed number of trials, the probability of a success is the same, and the trials are independent.
 - d. The sampling distributions of \hat{p}_1 and \hat{p}_2 can be approximated with a normal distribution.
 $x_1 = 231$, $n_1 - x_1 = 1000 - 231 = 769$, $x_2 = 176$, and
 $n_2 - x_2 = 1200 - 176 = 1024$ are all greater than or equal to 5. So both sampling distributions of \hat{p}_1 and \hat{p}_2 can be approximated with a normal distribution.

4. Find the sample statistics, test statistic, and p-value

Sample Proportion:

$$n_1 = 1000$$

$$n_2 = 1200$$

$$\hat{p}_1 = \frac{231}{1000} = 0.231$$

$$\hat{p}_2 = \frac{176}{1200} \approx 0.1467$$

$$\hat{q}_1 = 1 - \frac{231}{1000} = \frac{769}{1000} = 0.769$$

$$\hat{q}_2 = 1 - \frac{176}{1200} = \frac{1024}{1200} \approx 0.8533$$

Pooled Sample Proportion, \bar{p} :

$$\bar{p} = \frac{231 + 176}{1000 + 1200} = \frac{407}{2200} = 0.185$$

$$\bar{q} = 1 - \frac{407}{2200} = \frac{1793}{2200} = 0.815$$

Test Statistic:

$$z = \frac{(0.231 - 0.1467) - 0}{\sqrt{\frac{0.185 * 0.815}{1000} + \frac{0.185 * 0.815}{1200}}} = 5.0704$$

p-value:

On TI-83/84: $\text{normalcdf}(5.0704, 1E99, 0, 1) = 1.988 \times 10^{-7}$

On R: $1 - \text{pnorm}(5.0704, 0, 1) = 1.988 \times 10^{-7}$

Figure #9.1.1: Setup for 2-PropZTest on TI-83/84 Calculator

```
2-PropZTest
x1:231
n1:1000
x2:176
n2:1200
P1:#P2 <P2 >P2
Calculate Draw
```

Figure #9.1.2: Results for 2-PropZTest on TI-83/84 Calculator

```
2-PropZTest
P1>P2
↑P1=.231
P2=.1466666667
P=.185
n1=1000
n2=1200
```

```
2-PropZTest
P1>P2
z=5.072404516
P=1.9674319E-7
P1=.231
P2=.1466666667
↓P=.185
```

On R: `prop.test(c(x1, x2), c(n1, n2), alternative = "less" or "greater")`. For this example, `prop.test(c(231, 176), c(1000, 1200), alternative="greater")`

2-sample test for equality of proportions with continuity correction

```
data: c(231, 176) out of c(1000, 1200)
X-squared = 25.173, df = 1, p-value = 2.621e-07
alternative hypothesis: greater
95 percent confidence interval:
 0.05579805 1.00000000
sample estimates:
 prop 1  prop 2
0.2310000 0.1466667
```

Note: the answer from R is the p-value. It is different from the formula or the TI-83/84 calculator due to a continuity correction that R does.

5. Conclusion
Reject H_0 , since the p-value is less than 5%.
6. Interpretation
This is enough evidence to show that the proportion of husbands having affairs is more than the proportion of wives having affairs.

Example #9.1.2: Confidence Interval for Two Population Proportions

Do more husbands cheat on their wives more than wives cheat on the husbands ("Statistics brain," 2013)? Suppose you take a group of 1000 randomly selected husbands and find that 231 had cheated on their wives. Suppose in a group of 1200 randomly selected wives, 176 cheated on their husbands. Estimate the difference in the proportion of husbands and wives who cheat on their spouses using a 95% confidence level.

Solution:

1. State the random variables and the parameters in words.
These were stated in example #9.3.1, but are reproduced here for reference.

x_1 = number of husbands who cheat on his wife
 x_2 = number of wives who cheat on her husband
 p_1 = proportion of husbands who cheat on his wife
 p_2 = proportion of wives who cheat on her husband

- State and check the assumptions for the confidence interval
The assumptions were stated and checked in example #9.1.1.

- Find the sample statistics and the confidence interval

Sample Proportion:

$$n_1 = 1000$$

$$n_2 = 1200$$

$$\hat{p}_1 = \frac{231}{1000} = 0.231$$

$$\hat{p}_2 = \frac{176}{1200} \approx 0.1467$$

$$\hat{q}_1 = 1 - \frac{231}{1000} = \frac{769}{1000} = 0.769 \quad \hat{q}_2 = 1 - \frac{176}{1200} = \frac{1024}{1200} \approx 0.8533$$

Confidence Interval:

$$z_c = 1.96$$

$$E = 1.96 \sqrt{\frac{0.231 * 0.769}{1000} + \frac{0.1467 * 0.8533}{1200}} = 0.033$$

The confidence interval estimate of the difference $p_1 - p_2$ is

$$(\hat{p}_1 - \hat{p}_2) - E < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + E$$

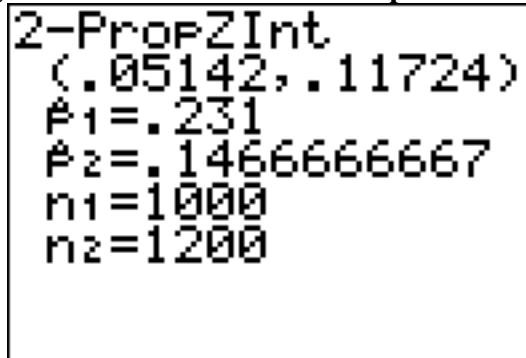
$$(0.231 - 0.1467) - 0.033 < p_1 - p_2 < (0.231 - 0.1467) + 0.033$$

$$0.0513 < p_1 - p_2 < 0.1173$$

Figure #9.1.3: Setup for 2-PropZInt on TI-83/84 Calculator

```

2-PropZInt
x1:231
n1:1000
x2:176
n2:1200
C-Level:.95
Calculate
  
```

Figure #9.1.4: Results for 2-PropZInt on TI-83/84 CalculatorA TI-83/84 calculator screen displaying the results of a 2-PropZInt test. The screen shows the following text: 2-PropZInt, (.05142, .11724), p1=.231, p2=.1466666667, n1=1000, n2=1200.

```
2-PropZInt
(.05142, .11724)
p1=.231
p2=.1466666667
n1=1000
n2=1200
```

On R: `prop.test(c(x1, x2), c(n1, n2), conf.level = C)`, where C is in decimal form. For this example, `prop.test(c(231, 176), c(1000, 1200), conf.level=0.95)`

2-sample test for equality of proportions with continuity correction

```
data: c(231, 176) out of c(1000, 1200)
X-squared = 25.173, df = 1, p-value = 5.241e-07
alternative hypothesis: two.sided
95 percent confidence interval:
 0.05050705 0.11815962
sample estimates:
   prop 1   prop 2 
0.2310000 0.1466667
```

Note: the answer from R is the confidence interval. It is different from the formula or the TI-83/84 calculator due to a continuity correction that R does.

4. Statistical Interpretation: There is a 95% chance that $0.0505 < p_1 - p_2 < 0.1182$ contains the true difference in proportions.
5. Real World Interpretation: The proportion of husbands who cheat is anywhere from 5.05% to 11.82% higher than the proportion of wives who cheat.

Section 9.1: Homework

In each problem show all steps of the hypothesis test or confidence interval. If some of the assumptions are not met, note that the results of the test or interval may not be correct and then continue the process of the hypothesis test or confidence interval.

- 1.) Many high school students take the AP tests in different subject areas. In 2007, of the 144,796 students who took the biology exam 84,199 of them were female. In that same year, of the 211,693 students who took the calculus AB exam 102,598 of them were female ("AP exam scores," 2013). Is there enough evidence to show that the proportion of female students taking the biology exam is higher than the proportion of female students taking the calculus AB exam? Test at the 5% level.
- 2.) Many high school students take the AP tests in different subject areas. In 2007, of the 144,796 students who took the biology exam 84,199 of them were female. In that same year, of the 211,693 students who took the calculus AB exam 102,598 of them were female ("AP exam scores," 2013). Estimate the difference in the proportion of female students taking the biology exam and female students taking the calculus AB exam using a 90% confidence level.
- 3.) Many high school students take the AP tests in different subject areas. In 2007, of the 211,693 students who took the calculus AB exam 102,598 of them were female and 109,095 of them were male ("AP exam scores," 2013). Is there enough evidence to show that the proportion of female students taking the calculus AB exam is different from the proportion of male students taking the calculus AB exam? Test at the 5% level.
- 4.) Many high school students take the AP tests in different subject areas. In 2007, of the 211,693 students who took the calculus AB exam 102,598 of them were female and 109,095 of them were male ("AP exam scores," 2013). Estimate using a 90% level the difference in proportion of female students taking the calculus AB exam versus male students taking the calculus AB exam.
- 5.) Are there more children diagnosed with Autism Spectrum Disorder (ASD) in states that have larger urban areas over states that are mostly rural? In the state of Pennsylvania, a fairly urban state, there are 245 eight year olds diagnosed with ASD out of 18,440 eight year olds evaluated. In the state of Utah, a fairly rural state, there are 45 eight year olds diagnosed with ASD out of 2,123 eight year olds evaluated ("Autism and developmental," 2008). Is there enough evidence to show that the proportion of children diagnosed with ASD in Pennsylvania is more than the proportion in Utah? Test at the 1% level.

- 6.) Are there more children diagnosed with Autism Spectrum Disorder (ASD) in states that have larger urban areas over states that are mostly rural? In the state of Pennsylvania, a fairly urban state, there are 245 eight year olds diagnosed with ASD out of 18,440 eight year olds evaluated. In the state of Utah, a fairly rural state, there are 45 eight year olds diagnosed with ASD out of 2,123 eight year olds evaluated ("Autism and developmental," 2008). Estimate the difference in proportion of children diagnosed with ASD between Pennsylvania and Utah. Use a 98% confidence level.
- 7.) A child dying from an accidental poisoning is a terrible incident. Is it more likely that a male child will get into poison than a female child? To find this out, data was collected that showed that out of 1830 children between the ages one and four who pass away from poisoning, 1031 were males and 799 were females (Flanagan, Rooney & Griffiths, 2005). Do the data show that there are more male children dying of poisoning than female children? Test at the 1% level.
- 8.) A child dying from an accidental poisoning is a terrible incident. Is it more likely that a male child will get into poison than a female child? To find this out, data was collected that showed that out of 1830 children between the ages one and four who pass away from poisoning, 1031 were males and 799 were females (Flanagan, Rooney & Griffiths, 2005). Compute a 99% confidence interval for the difference in proportions of poisoning deaths of male and female children ages one to four.

Section 9.2 Paired Samples for Two Means

Are two populations the same? Is the average height of men taller than the average height of women? Is the mean weight less after a diet than before?

You can compare populations by comparing their means. You take a sample from each population and compare the statistics.

Anytime you compare two populations you need to know if the samples are independent or dependent. The formulas you use are different for different types of samples.

If how you choose one sample has no effect on the way you choose the other sample, the two samples are **independent**. The way to think about it is that in independent samples, the individuals from one sample are overall different from the individuals from the other sample. This will mean that sample one has no effect on sample two. The sample values from one sample are not related or paired with values from the other sample.

If you choose the samples so that a measurement in one sample is paired with a measurement from the other sample, the samples are **dependent** or **matched** or **paired**. (Often a before and after situation.) You want to make sure there is a meaning for pairing data values from one sample with a specific data value from the other sample. One way to think about it is that in dependent samples, the individuals from one sample are the same individuals from the other sample, though there can be other reasons to pair values. This makes the sample values from each sample paired.

Example #9.2.1: Independent or Dependent Samples

Determine if the following are dependent or independent samples.

- a.) Randomly choose 5 men and 6 women and compare their heights

Solution:

Independent, since there is no reason that one value belongs to another. The individuals are not the same for both samples. The individuals are definitely different. A way to think about this is that the knowledge that a man is chosen in one sample does not give any information about any of the woman chosen in the other sample.

- b.) Choose 10 men and weigh them. Give them a new wonder diet drug and later weigh them again.

Solution:

Dependent, since each person's before weight can be matched with their after weight. The individuals are the same for both samples. A way to think about this is that the knowledge that a person weighs 400 pounds at the beginning will tell you something about their weight after the diet drug.

- c.) Take 10 people and measure the strength of their dominant arm and their non-dominant arm.

Solution:

Dependent, since you can match the two arm strengths. The individuals are the same for both samples. So the knowledge of one person's dominant arm strength will tell you something about the strength of their non-dominant arm.

To analyze data when there are matched or paired samples, called dependent samples, you conduct a paired t-test. Since the samples are matched, you can find the difference between the values of the two random variables.

Hypothesis Test for Two Sample Paired t-Test

1. State the random variables and the parameters in words.

x_1 = random variable 1

x_2 = random variable 2

μ_1 = mean of random variable 1

μ_2 = mean of random variable 2

2. State the null and alternative hypotheses and the level of significance

The usual hypotheses would be

$$H_o : \mu_1 = \mu_2 \quad \text{or} \quad H_o : \mu_1 - \mu_2 = 0$$

$$H_A : \mu_1 < \mu_2 \quad H_A : \mu_1 - \mu_2 < 0$$

$$H_A : \mu_1 > \mu_2 \quad H_A : \mu_1 - \mu_2 > 0$$

$$H_A : \mu_1 \neq \mu_2 \quad H_A : \mu_1 - \mu_2 \neq 0$$

However, since you are finding the differences, then you can actually think of $\mu_1 - \mu_2 = \mu_d$, μ_d = population mean value of the differences

So the hypotheses become

$$H_o : \mu_d = 0$$

$$H_1 : \mu_d < 0$$

$$H_A : \mu_d > 0$$

$$H_A : \mu_d \neq 0$$

Also, state your α level here.

3. State and check the assumptions for the hypothesis test

a. A random sample of n pairs is taken.

b. The population of the difference between random variables is normally distributed. In this case the population you are interested in has to do with the differences that you find. It does not matter if each random variable is normally distributed. It is only important if the differences you find are normally distributed. Just as before, the t-test is fairly robust to the assumption if the sample size is large. This means that if this assumption isn't

met, but your sample size is quite large (over 30), then the results of the t-test are valid.

4. Find the sample statistic, test statistic, and p-value

Sample Statistic:

Difference: $d = x_1 - x_2$ for each pair

Sample mean of the differences: $\bar{d} = \frac{\sum d}{n}$

Standard deviation of the differences: $s_d = \frac{\sum (d - \bar{d})^2}{n - 1}$

Number of pairs: n

Test Statistic:

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

with degrees of freedom = $df = n - 1$

Note: $\mu_d = 0$ in most cases.

p-value:

On TI-83/84: Use `tcdf(lower limit, upper limit, df)`

(Note: if $H_A : \mu_d < 0$, then lower limit is $-1E99$ and upper limit is your test statistic. If $H_A : \mu_d > 0$, then lower limit is your test statistic and the upper limit is $1E99$. If $H_A : \mu_d \neq 0$, then find the p-value for $H_A : \mu_d < 0$, and multiply by 2.)

On R: Use `pt(t, df)`

(Note: if $H_A : \mu_d < 0$, use `pt(t, df)`. If $H_A : \mu_d > 0$, use `1 - pt(t, df)`. If $H_A : \mu_d \neq 0$, then find the p-value for $H_A : \mu_d < 0$, and multiply by 2.)

5. Conclusion

This is where you write reject H_o or fail to reject H_o . The rule is: if the p-value $< \alpha$, then reject H_o . If the p-value $\geq \alpha$, then fail to reject H_o .

6. Interpretation

This is where you interpret in real world terms the conclusion to the test. The conclusion for a hypothesis test is that you either have enough evidence to show H_A is true, or you do not have enough evidence to show H_A is true.

Confidence Interval for Difference in Means from Paired Samples (t-Interval)

The confidence interval for the difference in means has the same random variables and means and the same assumptions as the hypothesis test for two paired samples. If you have already completed the hypothesis test, then you do not need to state them again. If you haven't completed the hypothesis test, then state the random variables and means, and state and check the assumptions before completing the confidence interval step.

1. Find the sample statistic and confidence interval

Sample Statistic:

Difference: $d = x_1 - x_2$

Sample mean of the differences: $\bar{d} = \frac{\sum d}{n}$

Standard deviation of the differences: $s_d = \frac{\sum (d - \bar{d})^2}{n - 1}$

Number of pairs: n

Confidence Interval:

The confidence interval estimate of the difference $\mu_d = \mu_1 - \mu_2$ is

$$\bar{d} - E < \mu_d < \bar{d} + E$$

$$E = t_c \frac{s_d}{\sqrt{n}}$$

t_c is the critical value where degrees of freedom $df = n - 1$

2. Statistical Interpretation: In general this looks like, "there is a C% chance that the statement $\bar{d} - E < \mu_d < \bar{d} + E$ contains the true mean difference."
3. Real World Interpretation: This is where you state what interval contains the true mean difference.

The critical value is a value from the Student's t-distribution. Since a confidence interval is found by adding and subtracting a margin of error amount from the sample mean, and the interval has a probability of containing the true mean difference, then you can think of this as the statement $P(\bar{d} - E < \mu_d < \bar{d} + E) = C$. To find the critical value, you use table A.2 in the Appendix.

How to check the assumptions of t-test and confidence interval:

In order for the t-test or confidence interval to be valid, the assumptions of the test must be met. So whenever you run a t-test or confidence interval, you must make sure the assumptions are met. So you need to check them. Here is how you do this:

1. For the assumption that the sample is a random sample, describe how you took the samples. Make sure your sampling technique is random and that the samples were dependent.
2. For the assumption that the population of the differences is normal, remember the process of assessing normality from chapter 6.

Example #9.2.2: Hypothesis Test for Paired Samples Using the Formula

A researcher wants to see if a weight loss program is effective. She measures the weight of 6 randomly selected women before and after the weight loss program (see table #9.2.1). Is there evidence that the weight loss program is effective? Test at the 5% level.

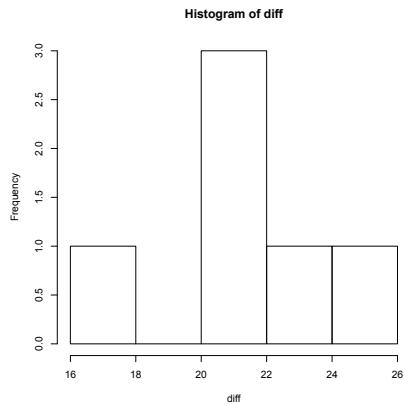
Table #9.2.1: Data of Before and After Weights

Person	1	2	3	4	5	6
Weight before	165	172	181	185	168	175
Weight after	143	151	156	161	152	154

Solution:

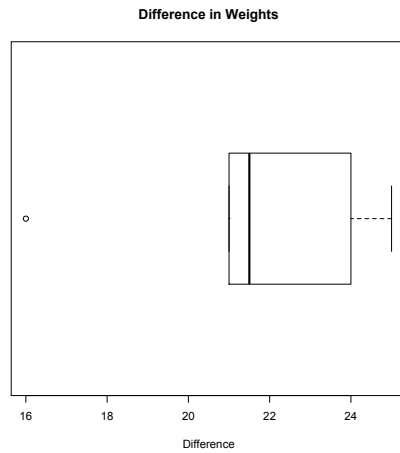
- State the random variables and the parameters in words.
 x_1 = weight of a woman after the weight loss program
 x_2 = weight of a woman before the weight loss program
 μ_1 = mean weight of a woman after the weight loss program
 μ_2 = mean weight of a woman before the weight loss program
- State the null and alternative hypotheses and the level of significance
 $H_o : \mu_d = 0$
 $H_A : \mu_d < 0$
 $\alpha = 0.05$
- State and check the assumptions for the hypothesis test
 - A random sample of 6 pairs of weights before and after was taken. This was stated in the problem, since the women were chosen randomly.
 - The population of the difference in after and before weights is normally distributed. To see if this is true, look at the histogram, number of outliers, and the normal probability plot. (If you wish, you can look at the normal probability plot first. If it doesn't look linear, then you may want to look at the histogram and number of outliers at this point.)

Figure #9.2.1: Histogram of Differences in Weights



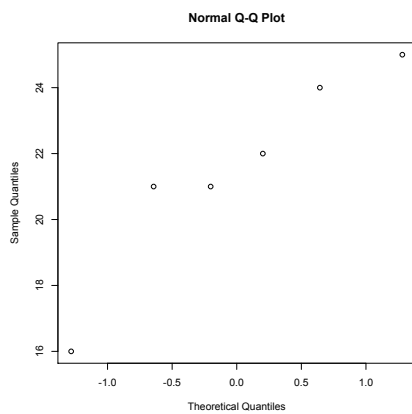
This histogram looks somewhat bell shaped.

Figure #9.2.2: Modified Box Plot of Differences in Weights



There is only one outlier in the difference data set.

Figure #9.2.3: Normal Quantile Plot of Differences in Weights



The probability plot on the differences looks somewhat linear.
So you can assume that the distribution of the difference in weights is normal.

4. Find the sample statistic, test statistic, and p-value

Sample Statistics:

Table #9.2.2: Differences Between Before and After Weights

Person	1	2	3	4	5	6
Weight after, x_1	143	151	156	161	152	154
Weight before, x_2	165	172	181	185	168	175
$d = x_1 - x_2$	-22	-21	-25	-24	-16	-21

The mean and standard deviation are

$$\bar{d} = -21.5$$

$$s_d = 3.15$$

Test Statistic:

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{-21.5 - 0}{3.15 / \sqrt{6}} = -16.779$$

p-value:

There are six pairs so the degrees of freedom are

$$df = n - 1 = 6 - 1 = 5$$

Since $H_1 : \mu_d < 0$, then p-value

$$\text{Using TI-83/84: } tcdf(-1E99, -16.779, 5) \approx 6.87 \times 10^{-6}$$

$$\text{Using R: } pt(-16.779, 5) \approx 6.87 \times 10^{-6}$$

5. Conclusion

Since the p-value < 0.05 , reject H_0 .

6. Interpretation

There is enough evidence to show that the weight loss program is effective.

Note: Just because the hypothesis test says the program is effective doesn't mean you should go out and use it right away. The program has statistical significance, but that doesn't mean it has practical significance. You need to see how much weight a person loses, and you need to look at how safe it is, how expensive, does it work in the long term, and other type questions. Remember to look at the practical significance in all situations. In this case, the average weight loss was 21.5 pounds, which is very practically significant. Do remember to look at the safety and expense of the drug also.

Example #9.2.3: Hypothesis Test for Paired Samples Using Technology

The New Zealand Air Force purchased a batch of flight helmets. They then found out that the helmets didn't fit. In order to make sure that they order the correct size helmets, they measured the head size of recruits. To save money, they wanted to use cardboard calipers, but were not sure if they will be accurate

enough. So they took 18 recruits and measured their heads with the cardboard calipers and also with metal calipers. The data in centimeters (cm) is in table #9.2.3 ("NZ helmet size," 2013). Do the data provide enough evidence to show that there is a difference in measurements between the cardboard and metal calipers? Use a 5% level of significance.

Table #9.2.3: Data for Head Measurements

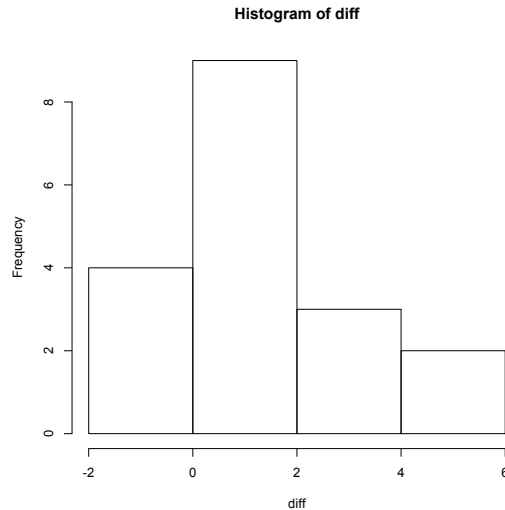
Cardboard	Metal
146	145
151	153
163	161
152	151
151	145
151	150
149	150
166	163
149	147
155	154
155	150
156	156
162	161
150	152
156	154
158	154
149	147
163	160

Solution:

1. State the random variables and the parameters in words.
 x_1 = head measurement of recruit using cardboard caliper
 x_2 = head measurement of recruit using metal caliper
 μ_1 = mean head measurement of recruit using cardboard caliper
 μ_2 = mean head measurement of recruit using metal caliper
2. State the null and alternative hypotheses and the level of significance
 $H_o : \mu_d = 0$
 $H_A : \mu_d \neq 0$
 $\alpha = 0.05$
3. State and check the assumptions for the hypothesis test
 - a. A random sample of 18 pairs of head measures of recruits with cardboard and metal caliper was taken. This was not stated, but probably could be safely assumed.

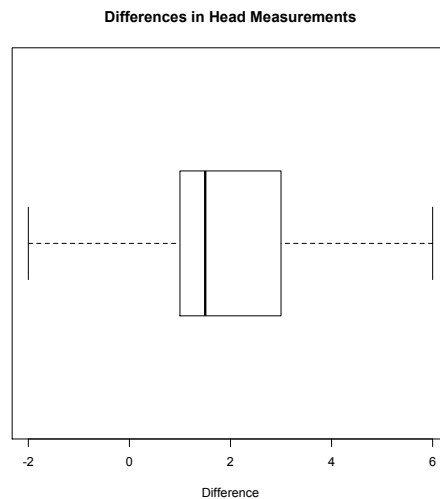
- b. The population of the difference in head measurements between cardboard and metal calipers is normally distributed. To see if this is true, look at the histogram, number of outliers, and the normal probability plot. (If you wish, you can look at the normal probability plot first. If it doesn't look linear, then you may want to look at the histogram and number of outliers at this point.)

Figure #9.2.4: Histogram of Differences in Head Measurements

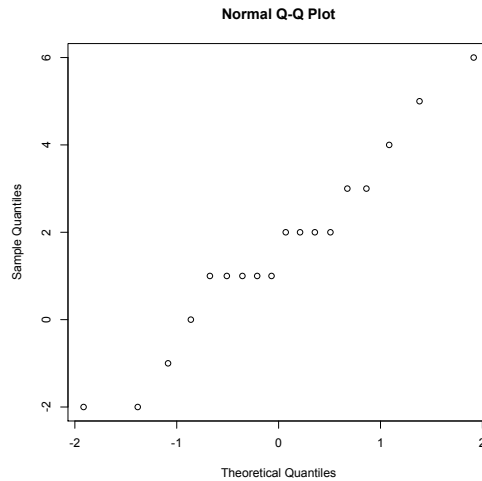


This histogram looks bell shaped.

Figure #9.2.5: Modified Box Plot of Differences in Head Measurements

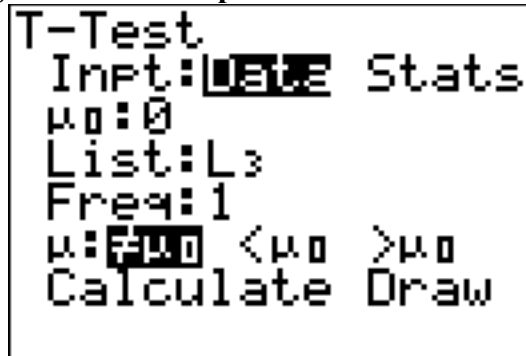


There are no outliers in the difference data set.

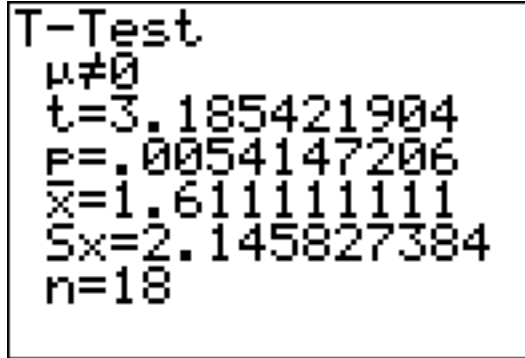
Figure #9.2.6: Normal Quantile Plot of Differences in Head Measurements

The probability plot on the differences looks somewhat linear. So you can assume that the distribution of the difference in weights is normal.

4. Find the sample statistic, test statistic, and p-value
Using the TI-83/84, put x_1 into L1 and x_2 into L2. Then go onto the name L3, and type $L1 - L2$. The calculator will calculate the differences for you and put them in L3. Now go into STAT and move over to TESTS. Choose T-Test. The setup for the calculator is in figure #9.2.7.

Figure #9.2.7: Setup for T-Test on Ti-83/84 Calculator

Once you press ENTER on Calculate you will see the result shown in figure #9.2.8.

Figure #9.2.8: Results of T-Test on TI-83/84 Calculator


```

T-Test
μ≠0
t=3.185421904
p=.0054147206
x̄=1.611111111
Sx=2.145827384
n=18

```

Using R: command is `t.test(variable1, variable2, paired = TRUE, alternative = "less" or "greater")`. For this example, the command would be `t.test(cardboard, metal, paired = TRUE)`

Paired t-test

```

data: cardboard and metal
t = 3.1854, df = 17, p-value = 0.005415
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.5440163 2.6782060
sample estimates:
mean of the differences
      1.611111

```

The $t = 3.185$ is the test statistic. The p-value is 0.0054147206.

5. Conclusion

Since the p-value < 0.05 , reject H_0 .

6. Interpretation

There is enough evidence to show that the mean head measurements using the cardboard calipers are not the same as when using the metal calipers. So it looks like the New Zealand Air Force shouldn't use the cardboard calipers.

Example #9.2.4: Confidence Interval for Paired Samples Using the Formula

A researcher wants to estimate the mean weight loss that people experience using a new program. She measures the weight of 6 randomly selected women before and after the weight loss program (see table #9.2.1). Find a 90% confidence interval for the mean the weight loss using the new program.

Solution:

1. State the random variables and the parameters in words.

These were stated in example #9.2.2, but are reproduced here for reference.

x_1 = weight of a woman after the weight loss program

x_2 = weight of a woman before the weight loss program

μ_1 = mean weight of a woman after the weight loss program

μ_2 = mean weight of a woman before the weight loss program

2. State and check the assumptions for the confidence interval
The assumptions were stated and checked in example #9.2.2.

3. Find the sample statistic and confidence interval

Sample Statistics:

From example #9.2.2

$$\bar{d} = -21.5$$

$$s_d = 3.15$$

The confidence level is 90%, so

$$C = 90\%$$

There are six pairs, so the degrees of freedom are

$$df = n - 1 = 6 - 1 = 5$$

Now look in table A.2. Go down the first column to 5, then over to the column headed with 90%.

$$t_c = 2.015$$

$$E = t_c \frac{s_d}{\sqrt{n}} = 2.015 \frac{3.15}{\sqrt{6}} \approx 2.6$$

$$\bar{d} - E < \mu_d < \bar{d} + E$$

$$-21.5 - 2.6 < \mu_d < -21.5 + 2.6$$

$$-24.1 \text{ pounds} < \mu_d < -18.9 \text{ pounds}$$

4. Statistical Interpretation: There is a 90% chance that $-24.1 \text{ pounds} < \mu_d < -18.9 \text{ pounds}$ contains the true mean difference in weight loss.
5. Real World Interpretation: The mean weight loss is between 18.9 and 24.1 pounds. (Note, the negative signs tell you that the first mean is less than the second mean, and thus a weight loss in this case.)

Example #9.2.5: Confidence Interval for Paired Samples Using Technology

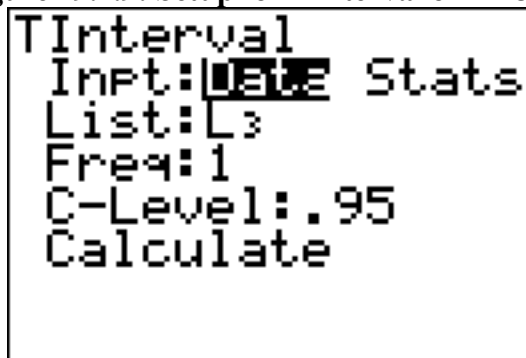
The New Zealand Air Force purchased a batch of flight helmets. They then found out that the helmets didn't fit. In order to make sure that they order the correct size helmets, they measured the head size of recruits. To save money, they wanted to use cardboard calipers, but were not sure if they will be accurate enough. So they took 18 recruits and measured their heads with the cardboard calipers and also with metal calipers. The data in centimeters (cm) is in table

#9.2.3 ("NZ helmet size," 2013). Estimate the mean difference in measurements between the cardboard and metal calipers using a 95% confidence interval.

Solution:

1. State the random variables and the parameters in words.
 These were stated in example #9.2.3, but are reproduced here for reference.
 x_1 = head measurement of recruit using cardboard caliper
 x_2 = head measurement of recruit using metal caliper
 μ_1 = mean head measurement of recruit using cardboard caliper
 μ_2 = mean head measurement of recruit using metal caliper
2. State and check the assumptions for the hypothesis test
 The assumptions were stated and checked in example #9.2.3.
3. Find the sample statistic and confidence interval
 Using the TI-83/84, put x_1 into L1 and x_2 into L2. Then go onto the name L3, and type $L1 - L2$. The calculator will now calculate the differences for you and put them in L3. Now go into STAT and move over to TESTS. Then chose TInterval. The setup for the calculator is in figure #9.2.9.

Figure #9.2.9: Setup for TInterval on Ti-83/84 Calculator

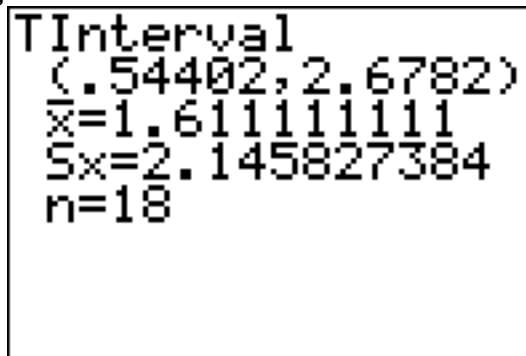


```

TInterval
Inpt: DATA Stats
List: L3
Freq: 1
C-Level: .95
Calculate
  
```

Once you press ENTER on Calculate you will see the result shown in figure #9.2.10.

Figure #9.2.10: Results of TInterval on TI-83/84 Calculator



```

TInterval
(.54402, 2.6782)
x̄=1.611111111
Sx=2.145827384
n=18
  
```

Using R: the command is `t.test(variable1, variable2, paired = TRUE, conf.level = C)`, where C is in decimal form. For this example the command would be `t.test(cardboard, metal, paired = TRUE, conf.level=0.95)`

Paired t-test

data: cardboard and metal
t = 3.1854, df = 17, p-value = 0.005415
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
0.5440163 2.6782060
sample estimates:
mean of the differences
1.611111

So
 $0.54 \text{ cm} < \mu_d < 2.68 \text{ cm}$

4. Statistical Interpretation: There is a 95% chance that $0.54 \text{ cm} < \mu_d < 2.68 \text{ cm}$ contains the true mean difference in head measurements between cardboard and metal calibers.
5. Real World Interpretation: The mean difference in head measurements between the cardboard and metal calibers is between 0.54 and 2.68 cm. This means that the cardboard calibers measure on average the head of a recruit to be between 0.54 and 2.68 cm more in diameter than the metal calibers. That makes it seem that the cardboard calibers are not measuring the same as the metal calibers. (The positive values on the confidence interval imply that the first mean is higher than the second mean.)

Examples #9.2.2 and #9.2.4 use the same data set, but one is conducting a hypothesis test and the other is conducting a confidence interval. Notice that the hypothesis test's conclusion was to reject H_0 and say that there was a difference in the means, and the confidence interval does not contain the number 0. If the confidence interval did contain the number 0, then that would mean that the two means could be the same. Since the interval did not contain 0, then you could say that the means are different just as in the hypothesis test. This means that the hypothesis test and the confidence interval can produce the same interpretation. Do be careful though, you can run a hypothesis test with a particular significance level and a confidence interval with a confidence level that is not compatible with your significance level. This will mean that the conclusion from the confidence interval would not be the same as with a hypothesis test. So if you want to estimate the mean difference, then conduct a confidence interval. If you want to show that the means are different, then conduct a hypothesis test.

Section 9.2: Homework

In each problem show all steps of the hypothesis test or confidence interval. If some of the assumptions are not met, note that the results of the test or interval may not be correct and then continue the process of the hypothesis test or confidence interval.

- 1.) The cholesterol level of patients who had heart attacks was measured two days after the heart attack and then again four days after the heart attack. The researchers want to see if the cholesterol level of patients who have heart attacks reduces as the time since their heart attack increases. The data is in table #9.2.4 ("Cholesterol levels after," 2013). Do the data show that the mean cholesterol level of patients that have had a heart attack reduces as the time increases since their heart attack? Test at the 1% level.

Table #9.2.4: Cholesterol Levels (in mg/dL) of Heart Attack Patients

Patient	Cholesterol Level Day 2	Cholesterol Level Day 4
1	270	218
2	236	234
3	210	214
4	142	116
5	280	200
6	272	276
7	160	146
8	220	182
9	226	238
10	242	288
11	186	190
12	266	236
13	206	244
14	318	258
15	294	240
16	282	294
17	234	220
18	224	200
19	276	220
20	282	186
21	360	352
22	310	202
23	280	218
24	278	248
25	288	278
26	288	248
27	244	270
28	236	242

- 2.) The cholesterol level of patients who had heart attacks was measured two days after the heart attack and then again four days after the heart attack. The researchers want to see if the cholesterol level of patients who have heart attacks reduces as the time since their heart attack increases. The data is in table #9.2.4 ("Cholesterol levels after," 2013). Calculate a 98% confidence interval for the mean difference in cholesterol levels from day two to day four.
- 3.) All Fresh Seafood is a wholesale fish company based on the east coast of the U.S. Catalina Offshore Products is a wholesale fish company based on the west coast of the U.S. Table #9.2.5 contains prices from both companies for specific fish types ("Seafood online," 2013) ("Buy sushi grade," 2013). Do the data provide enough evidence to show that a west coast fish wholesaler is more expensive than an east coast wholesaler? Test at the 5% level.

Table #9.2.5: Wholesale Prices of Fish in Dollars

Fish	All Fresh Seafood Prices	Catalina Offshore Products Prices
Cod	19.99	17.99
Tilapi	6.00	13.99
Farmed Salmon	19.99	22.99
Organic Salmon	24.99	24.99
Grouper Fillet	29.99	19.99
Tuna	28.99	31.99
Swordfish	23.99	23.99
Sea Bass	32.99	23.99
Striped Bass	29.99	14.99

- 4.) All Fresh Seafood is a wholesale fish company based on the east coast of the U.S. Catalina Offshore Products is a wholesale fish company based on the west coast of the U.S. Table #9.2.5 contains prices from both companies for specific fish types ("Seafood online," 2013) ("Buy sushi grade," 2013). Find a 95% confidence interval for the mean difference in wholesale price between the east coast and west coast suppliers.

- 5.) The British Department of Transportation studied to see if people avoid driving on Friday the 13th. They did a traffic count on a Friday and then again on a Friday the 13th at the same two locations ("Friday the 13th," 2013). The data for each location on the two different dates is in table #9.2.6. Do the data show that on average fewer people drive on Friday the 13th? Test at the 5% level.

Table #9.2.6: Traffic Count

Dates	6th	13th
1990, July	139246	138548
1990, July	134012	132908
1991, September	137055	136018
1991, September	133732	131843
1991, December	123552	121641
1991, December	121139	118723
1992, March	128293	125532
1992, March	124631	120249
1992, November	124609	122770
1992, November	117584	117263

- 6.) The British Department of Transportation studied to see if people avoid driving on Friday the 13th. They did a traffic count on a Friday and then again on a Friday the 13th at the same two locations ("Friday the 13th," 2013). The data for each location on the two different dates is in table #9.2.6. Estimate the mean difference in traffic count between the 6th and the 13th using a 90% level.

- 7.) To determine if Reiki is an effective method for treating pain, a pilot study was carried out where a certified second-degree Reiki therapist provided treatment on volunteers. Pain was measured using a visual analogue scale (VAS) immediately before and after the Reiki treatment (Olson & Hanson, 1997). The data is in table #9.2.7. Do the data show that Reiki treatment reduces pain? Test at the 5% level.

Table #9.2.7: Pain Measures Before and After Reiki Treatment

VAS before	VAS after
6	3
2	1
2	0
9	1
3	0
3	2
4	1
5	2
2	2
3	0
5	1
1	0
6	4
6	1
4	4
4	1
7	6
2	1
4	3
8	8

- 8.) To determine if Reiki is an effective method for treating pain, a pilot study was carried out where a certified second-degree Reiki therapist provided treatment on volunteers. Pain was measured using a visual analogue scale (VAS) immediately before and after the Reiki treatment (Olson & Hanson, 1997). The data is in table #9.2.7. Compute a 90% confidence level for the mean difference in VAS score from before and after Reiki treatment.

- 9.) The female labor force participation rates (FLFPR) of women in randomly selected countries in 1990 and latest years of the 1990s are in table #9.2.8 (Lim, 2002). Do the data show that the mean female labor force participation rate in 1990 is different from that in the latest years of the 1990s using a 5% level of significance?

Table #9.2.8: Female Labor Force Participation Rates

Region and country	FLFPR 25-54 1990	FLFPR 25-54 Latest year of 1990s
Iran	22.6	12.5
Morocco	41.4	34.5
Qatar	42.3	46.5
Syrian Arab Republic	25.6	19.5
United Arab Emirates	36.4	39.7
Cape Verde	46.7	50.9
Ghana	89.8	90.0
Kenya	82.1	82.6
Lesotho	51.9	68.0
South Africa	54.7	61.7
Bangladesh	73.5	60.6
Malaysia	49.0	50.2
Mongolia	84.7	71.3
Myanmar	72.1	72.3
Argentina	36.8	54
Belize	28.8	42.5
Bolivia	27.3	69.8
Brazil	51.1	63.2
Colombia	57.4	72.7
Ecuador	33.5	64
Nicaragua	50.1	42.5
Uruguay	59.5	71.5
Albania	77.4	78.8
Uzbekistan	79.6	82.8

- 10.) The female labor force participation rates of women in randomly selected countries in 1990 and latest years of the 1990s are in table #9.2.8 (Lim, 2002). Estimate the mean difference in the female labor force participation rate in 1990 to latest years of the 1990s using a 95% confidence level?

- 11.) Table #9.2.9 contains pulse rates collected from males, who are non-smokers but do drink alcohol ("Pulse rates before," 2013). The before pulse rate is before they exercised, and the after pulse rate was taken after the subject ran in place for one minute. Do the data indicate that the pulse rate before exercise is less than after exercise? Test at the 1% level.

Table#9.2.9: Pulse Rate of Males Before and After Exercise

Pulse before	Pulse after
76	88
56	110
64	126
50	90
49	83
68	136
68	125
88	150
80	146
78	168
59	92
60	104
65	82
76	150
145	155
84	140
78	141
85	131
78	132

- 12.) Table #9.2.9 contains pulse rates collected from males, who are non-smokers but do drink alcohol ("Pulse rates before," 2013). The before pulse rate is before they exercised, and the after pulse rate was taken after the subject ran in place for one minute. Compute a 98% confidence interval for the mean difference in pulse rates from before and after exercise.

Section 9.3 Independent Samples for Two Means

This section will look at how to analyze when two samples are collected that are independent. As with all other hypothesis tests and confidence intervals, the process is the same though the formulas and assumptions are different. The only difference with the independent t-test, as opposed to the other tests that have been done, is that there are actually two different formulas to use depending on if a particular assumption is met or not.

Hypothesis Test for Independent t-Test (2-Sample t-Test)

1. State the random variables and the parameters in words.

x_1 = random variable 1

x_2 = random variable 2

μ_1 = mean of random variable 1

μ_2 = mean of random variable 2

2. State the null and alternative hypotheses and the level of significance

The normal hypotheses would be

$$H_o : \mu_1 = \mu_2 \quad \text{or} \quad H_o : \mu_1 - \mu_2 = 0$$

$$H_A : \mu_1 < \mu_2 \quad H_A : \mu_1 - \mu_2 < 0$$

$$H_A : \mu_1 > \mu_2 \quad H_A : \mu_1 - \mu_2 > 0$$

$$H_A : \mu_1 \neq \mu_2 \quad H_A : \mu_1 - \mu_2 \neq 0$$

Also, state your α level here.

3. State and check the assumptions for the hypothesis test

- a. A random sample of size n_1 is taken from population 1. A random sample of size n_2 is taken from population 2. Note: the samples do not need to be the same size, but the test is more robust if they are.

- b. The two samples are independent.

- c. Population 1 is normally distributed. Population 2 is normally distributed. Just as before, the t-test is fairly robust to the assumption if the sample size is large. This means that if this assumption isn't met, but your sample sizes are quite large (over 30), then the results of the t-test are valid.

- d. The population variances are unknown and not assumed to be equal. The old assumption is that the variances are equal. However, this assumption is no longer an assumption that most statisticians use. This is because it isn't really realistic to assume that the variances are equal. So we will just assume the assumption of the variances being unknown and not assumed to be equal is true, and it will not be checked.

4. Find the sample statistic, test statistic, and p-value

Sample Statistic:

Calculate $\bar{x}_1, \bar{x}_2, s_1, s_2, n_1, n_2$

Test Statistic:

Since the assumption that $\sigma_1^2 = \sigma_2^2$ isn't being satisfied, then

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Usually $\mu_1 - \mu_2 = 0$, since $H_o: \mu_1 - \mu_2 = 0$

Degrees of freedom: (the Welch-Satterthwaite equation)

$$df = \frac{(A + B)^2}{\frac{A^2}{n_1 - 1} + \frac{B^2}{n_2 - 1}}$$

$$\text{where } A = \frac{s_1^2}{n_1} \text{ and } B = \frac{s_2^2}{n_2}$$

p-value:

Using the TI-83/84: tcdf(lower limit, upper limit, df)

(Note: if $H_A: \mu_1 - \mu_2 < 0$, then lower limit is $-1E99$ and upper limit is your test statistic. If $H_A: \mu_1 - \mu_2 > 0$, then lower limit is your test statistic and the upper limit is $1E99$. If $H_A: \mu_1 - \mu_2 \neq 0$, then find the p-value for $H_A: \mu_1 - \mu_2 < 0$, and multiply by 2.)

Using R: pt(t , df)

(Note: if $H_A: \mu_1 - \mu_2 < 0$, then use pt(t , df). If $H_A: \mu_1 - \mu_2 > 0$, then use $1 - \text{pt}(t, df)$. If $H_A: \mu_1 - \mu_2 \neq 0$, then find the p-value for $H_A: \mu_1 - \mu_2 < 0$, and multiply by 2.)

5. Conclusion

This is where you write reject H_o or fail to reject H_o . The rule is: if the p-value $< \alpha$, then reject H_o . If the p-value $\geq \alpha$, then fail to reject H_o .

6. Interpretation

This is where you interpret in real world terms the conclusion to the test. The conclusion for a hypothesis test is that you either have enough evidence to show H_A is true, or you do not have enough evidence to show H_A is true.

Confidence Interval for the Difference in Means from Two Independent Samples (2 Samp T-Int)

The confidence interval for the difference in means has the same random variables and means and the same assumptions as the hypothesis test for independent samples. If you have already completed the hypothesis test, then you do not need to state them again. If you haven't completed the hypothesis test, then state the random variables and means and state and check the assumptions before completing the confidence interval step.

1. Find the sample statistic and confidence interval

Sample Statistic:

Calculate $\bar{x}_1, \bar{x}_2, s_1, s_2, n_1, n_2$

Confidence Interval:

The confidence interval estimate of the difference $\mu_1 - \mu_2$ is

Since the assumption that $\sigma_1^2 = \sigma_2^2$ isn't being satisfied, then

$$(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$$

$$\text{where } E = t_c \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where t_c is the critical value with degrees of freedom:

Degrees of freedom: (the Welch–Satterthwaite equation)

$$df = \frac{(A + B)^2}{\frac{A^2}{n_1 - 1} + \frac{B^2}{n_2 - 1}}$$

$$\text{where } A = \frac{s_1^2}{n_1} \text{ and } B = \frac{s_2^2}{n_2}$$

2. Statistical Interpretation: In general this looks like, “there is a C% chance that $(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$ contains the true mean difference.”
3. Real World Interpretation: This is where you state what interval contains the true difference in means, though often you state how much more (or less) the first mean is from the second mean.

The critical value is a value from the Student's t-distribution. Since a confidence interval is found by adding and subtracting a margin of error amount from the difference in sample means, and the interval has a probability of containing the true difference in means, then you can think of this as the statement

$P((\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E) = C$. To find the critical value you use table A.2 in the Appendix.

How to check the assumptions of two sample t-test and confidence interval:

In order for the t-test or confidence interval to be valid, the assumptions of the test must be true. So whenever you run a t-test or confidence interval, you must make sure the assumptions are true. So you need to check them. Here is how you do this:

1. For the random sample assumption, describe how you took the two samples. Make sure your sampling technique is random for both samples.
2. For the independent assumption, describe how they are independent samples.
3. For the assumption about each population being normally distributed, remember the process of assessing normality from chapter 6. Make sure you assess each sample separately.

4. You do not need to check the equal variance assumption since it is not being assumed.

Example #9.3.1: Hypothesis Test for Two Means

The cholesterol level of patients who had heart attacks was measured two days after the heart attack. The researchers want to see if patients who have heart attacks have higher cholesterol levels over healthy people, so they also measured the cholesterol level of healthy adults who show no signs of heart disease. The data is in table #9.3.1 ("Cholesterol levels after," 2013). Do the data show that people who have had heart attacks have higher cholesterol levels over patients that have not had heart attacks? Test at the 1% level.

Table #9.3.1: Cholesterol Levels in mg/dL

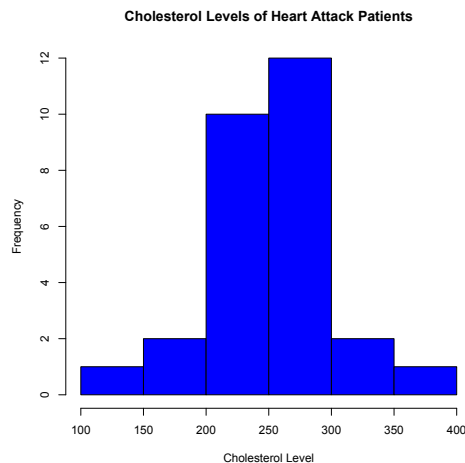
Cholesterol Level of Heart Attack Patients	Cholesterol Level of Healthy Individual
270	196
236	232
210	200
142	242
280	206
272	178
160	184
220	198
226	160
242	182
186	182
266	198
206	182
318	238
294	198
282	188
234	166
224	204
276	182
282	178
360	212
310	164
280	230
278	186
288	162
288	182
244	218
236	170
	200
	176

Solution:

- State the random variables and the parameters in words.
 x_1 = Cholesterol level of patients who had a heart attack
 x_2 = Cholesterol level of healthy individuals
 μ_1 = mean cholesterol level of patients who had a heart attack
 μ_2 = mean cholesterol level of healthy individuals
- State the null and alternative hypotheses and the level of significance
The normal hypotheses would be
 $H_o : \mu_1 = \mu_2$ or $H_o : \mu_1 - \mu_2 = 0$
 $H_A : \mu_1 > \mu_2$ $H_A : \mu_1 - \mu_2 > 0$
 $\alpha = 0.01$
- State and check the assumptions for the hypothesis test
 - A random sample of 28 cholesterol levels of patients who had a heart attack is taken. A random sample of 30 cholesterol levels of healthy individuals is taken. The problem does not state if either sample was randomly selected. So this assumption may not be valid.
 - The two samples are independent. This is because either they were dealing with patients who had heart attacks or healthy individuals.
 - Population of all cholesterol levels of patients who had a heart attack is normally distributed. Population of all cholesterol levels of healthy individuals is normally distributed.

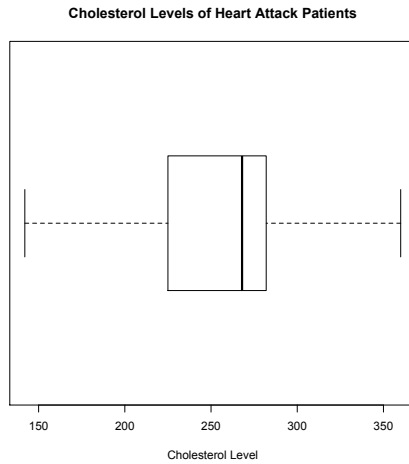
Patients who had heart attacks:

Figure #9.3.1: Histogram of Cholesterol Levels of Patients who had Heart Attacks



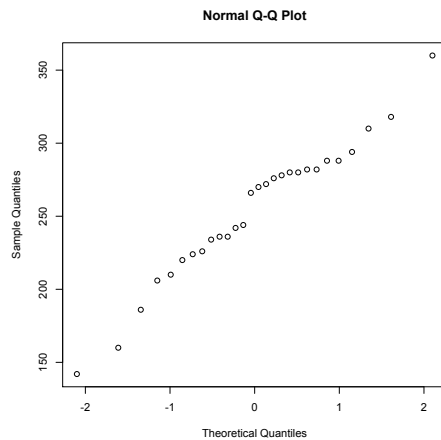
This looks somewhat bell shaped.

Figure #9.3.2: Modified Box Plot of Cholesterol Levels of Patients who had Heart Attacks



There are no outliers.

Figure #9.3.3: Normal Quantile Plot of Cholesterol Levels of Patients who had Heart Attacks

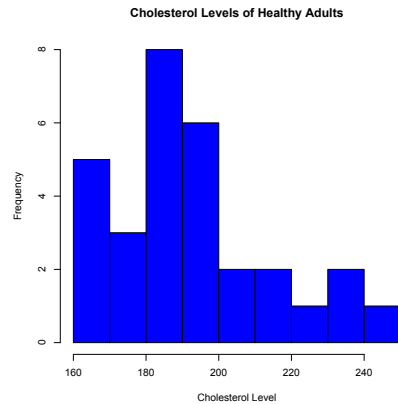


This looks somewhat linear.

So, the population of all cholesterol levels of patients who had heart attacks is probably somewhat normally distributed.

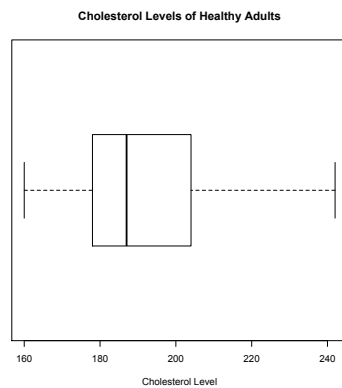
Healthy individuals:

Figure #9.3.4: Histogram of Cholesterol Levels of Healthy Individuals



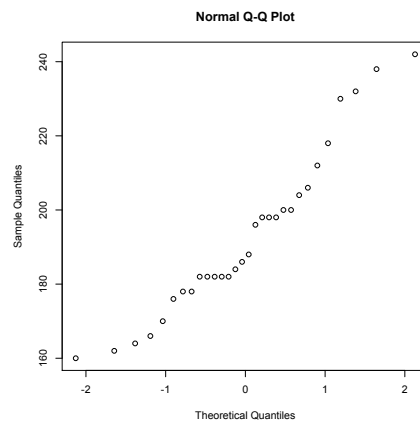
This does not look bell shaped.

Figure #9.3.5: Modified Box Plot of Cholesterol Levels of Healthy Individuals



There are no outliers.

Figure #9.3.6: Normal Quantile Plot of Cholesterol Levels of Healthy Individuals



This doesn't look linear.

So, the population of all cholesterol levels of healthy individuals is probably not normally distributed.

This assumption is not valid for the second sample. Since the sample is fairly large, and the t-test is robust, it may not be an issue. However, just realize that the conclusions of the test may not be valid.

4. Find the sample statistic, test statistic, and p-value

Sample Statistic:

$$\bar{x}_1 \approx 252.32, \bar{x}_2 \approx 193.13, s_1 \approx 47.0642, s_2 \approx 22.3000, n_1 = 28, n_2 = 30$$

Test Statistic:

$$\begin{aligned} t &= \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \\ &= \frac{(252.32 - 193.13) - 0}{\sqrt{\frac{47.0642^2}{28} + \frac{22.3000^2}{30}}} \\ &\approx 6.051 \end{aligned}$$

Degrees of freedom: (the Welch–Satterthwaite equation)

$$A = \frac{s_1^2}{n_1} = \frac{47.0642^2}{28} \approx 79.1085$$

$$B = \frac{s_2^2}{n_2} = \frac{22.3000^2}{30} \approx 16.5763$$

$$df = \frac{\frac{(A+B)^2}{\frac{A^2}{n_1-1} + \frac{B^2}{n_2-1}}}{\frac{A^2}{n_1-1} + \frac{B^2}{n_2-1}} = \frac{(79.1085 + 16.5763)^2}{\frac{79.1085^2}{28-1} + \frac{16.5763^2}{30-1}} \approx 37.9493$$


p-value:

$$\text{Using TI-83/84: } \text{tcdf}(6.051, 1E99, 37.9493) \approx 2.44 \times 10^{-7}$$

$$\text{Using R: } 1 - \text{pt}(6.051, 37.9493) \approx 2.44 \times 10^{-7}$$

Using Technology: Using the TI-83/84:

Figure #9.3.7: Setup for 2-SampTTest on TI-83/84 Calculator



```

2-SampTTest
Inpt: DATA Stats
List1: L1
List2: L2
Freq1: 1
Freq2: 1
mu1: #mu2 <mu2 >mu2
Pooled: NO Yes
  
```

Note: the Pooled question on the calculator is for whether you are assuming the variances are equal. Since this assumption is not being made, then the answer to this question is no. Pooled means that you assume the variances are equal and can pool the sample variances together.

Figure #9.3.8 Results for 2-SampTTest on TI-83/84 Calculator

NORMAL FLOAT AUTO REAL Radian MP	NORMAL FLOAT AUTO REAL Radian MP
2-SampTTest	2-SampTTest
$\mu_1 > \mu_2$	$\mu_1 > \mu_2$
$t = 6.14524511$	$\uparrow df = 37.674889$
$P = 1.8604395E-7$	$\bar{x}_1 = 253.9285714$
$df = 37.674889$	$\bar{x}_2 = 193.1333333$
$\bar{x}_1 = 253.9285714$	$Sx_1 = 47.7104916$
$\bar{x}_2 = 193.1333333$	$Sx_2 = 22.3000438$
$Sx_1 = 47.7104916$	$n_1 = 28$
$\downarrow Sx_2 = 22.3000438$	$n_2 = 30$

Using R: command in general: `t.test(variable1, variable2, alternative = "less" or "greater")`

For this example, the R command is:

`t.test(heartattack, healthy, alternative="greater")`

Welch Two Sample t-test

data: heartattack and healthy

$t = 6.1452$, $df = 37.675$, $p\text{-value} = 1.86e-07$

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

44.1124 Inf

sample estimates:

mean of x mean of y

253.9286 193.1333

The test statistic is $t = 6.1452$. The p-value is 1.86×10^{-7}

5. Conclusion

Reject H_0 since the $p\text{-value} < \alpha$.

6. Interpretation

This is enough evidence to show that patients who have had heart attacks have higher cholesterol level on average from healthy individuals. (Though do realize that some of assumptions are not valid, so this interpretation may be invalid.)

Example #9.3.2: Confidence Interval for $\mu_1 - \mu_2$

The cholesterol level of patients who had heart attacks was measured two days after the heart attack. The researchers want to see if patients who have heart attacks have higher cholesterol levels over healthy people, so they also measured the cholesterol level of healthy adults who show no signs of heart disease. The data is in table #9.3.1 ("Cholesterol levels after," 2013). Find a 99% confidence interval for the mean difference in cholesterol levels between heart attack patients and healthy individuals.

Solution:

1. State the random variables and the parameters in words.

These were stated in example #9.3.1, but are reproduced here for reference.

x_1 = Cholesterol level of patients who had a heart attack

x_2 = Cholesterol level of healthy individuals

μ_1 = mean cholesterol level of patients who had a heart attack

μ_2 = mean cholesterol level of healthy individuals

2. State and check the assumptions for the hypothesis test

The assumptions were stated and checked in example #9.3.1.

3. Find the sample statistic and confidence interval

Sample Statistic:

$$\bar{x}_1 \approx 252.32, \bar{x}_2 \approx 193.13, s_1 \approx 47.0642, s_2 \approx 22.3000, n_1 = 28, n_2 = 30$$

Test Statistic:

Degrees of freedom: (the Welch–Satterthwaite equation)

$$A = \frac{s_1^2}{n_1} = \frac{47.0642^2}{28} \approx 79.1085$$

$$B = \frac{s_2^2}{n_2} = \frac{22.3000^2}{30} \approx 16.5763$$

$$df = \frac{(A + B)^2}{\frac{A^2}{n_1 - 1} + \frac{B^2}{n_2 - 1}} = \frac{(79.1085 + 16.5763)^2}{\frac{79.1085^2}{28 - 1} + \frac{16.5763^2}{30 - 1}} \approx 37.9493$$

Since this df is not in the table, round to the nearest whole number.

$$t_c = 2.712$$

$$E = t_c \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 2.712 \sqrt{\frac{47.0642^2}{28} + \frac{22.3000^2}{30}} \approx 26.53$$

$$(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$$

$$(252.32 - 193.13) - 26.53 < \mu_1 - \mu_2 < (252.32 - 193.13) + 26.53$$

$$32.66 \text{ mg/dL} < \mu_1 - \mu_2 < 85.72 \text{ mg/dL}$$

Using Technology:

Using TI-83/84:

Figure #9.3.9: Setup for 2-SampTInt on TI-83/84 Calculator

```

2-SampTInt
Inpt: ☐ Data Stats
List1: L1
List2: L2
Freq1: 1
Freq2: 1
C-Level: .99
↓Pooled: ☐ Yes

```

Note: the Pooled question on the calculator is for whether you are assuming the variances are equal. Since this assumption is not being made, then the answer to this question is no. Pooled means that you assume the variances are equal and can pool the sample variances together.

Figure #9.3.10: Results for 2-SampTInt on TI-83/84 Calculator

```

2-SampTInt
(32.662, 85.714)
df=37.94930949
x̄1=252.3214286
x̄2=193.1333333
Sx1=47.0642221
↓Sx2=22.3000438
■

2-SampTInt
(32.662, 85.714)
↑x̄2=193.1333333
Sx1=47.0642221
Sx2=22.3000438
n1=28
n2=30

```

Using R: the command is `t.test(variable1, variable2, conf.level=C)`, where C is in decimal form.

For this example, the command is

`t.test(heartattack, healthy, conf.level=.99)`

Output:

Welch Two Sample t-test

data: heartattack and healthy

t = 6.1452, df = 37.675, p-value = 3.721e-07

alternative hypothesis: true difference in means is not equal to 0

99 percent confidence interval:

33.95750 87.63298

sample estimates:

mean of x mean of y

253.9286 193.1333

The confidence interval is $33.96 < \mu_1 - \mu_2 < 87.63$.

4. Statistical Interpretation: There is a 99% chance that $32.66 \text{ mg/dL} < \mu_1 - \mu_2 < 85.72 \text{ mg/dL}$ contains the true difference in means.
5. Real World Interpretation: The mean cholesterol level for patients who had heart attacks is anywhere from 32.66 mg/dL to 85.72 mg/dL more than the mean cholesterol level for healthy patients. (Though do realize that many of assumptions are not valid, so this interpretation may be invalid.)

If you do assume that the variances are equal, that is $\sigma_1^2 = \sigma_2^2$, then the test statistic is:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\text{where } s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}}$$

s_p = pooled standard deviation

The Degrees of Freedom is: $df = n_1 + n_2 - 2$

The confidence interval if you do assume that $\sigma_1^2 = \sigma_2^2$ has been met, is

$$(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$$

$$\text{where } E = t_c s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\text{and } s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}}$$

Degrees of Freedom: $df = n_1 + n_2 - 2$

t_c is the critical value where $C = 1 - \alpha$

To show that the variances are equal, just show that the ratio of your sample variances is not unusual (probability is greater than 0.05). In other words, make sure the following is true.

$P(F > s_1^2 / s_2^2) \geq 0.05$ (or $P(F > s_2^2 / s_1^2) \geq 0.05$ so that the larger variance is in the numerator). This probability is from an F-distribution. To find the probability on the TI-83/84 calculator use $\text{Fcdf}(s_1^2 / s_2^2, 1E99, n_1 - 1, n_2 - 1)$. To find the probability on R, use $1 - \text{pf}(s_1^2 / s_2^2, n_1 - 1, n_2 - 1)$.

Note: the F-distribution is very sensitive to the normal distribution. A better test for equal variances is Levene's test, though it is more complicated. It is best to do Levene's test when using statistical software (such as SPSS or Minitab) to perform the two-sample independent t-test.

Example #9.3.3: Hypothesis Test for Two Means

The amount of sodium in beef hotdogs was measured. In addition, the amount of sodium in poultry hotdogs was also measured ("SOCR 012708 id," 2013). The data is in table #9.3.2. Is there enough evidence to show that beef has less sodium on average than poultry hotdogs? Use a 5% level of significance.

Table #9.3.2: Hotdog Data

Sodium in Beef Hotdogs	Sodium in Poultry Hotdogs
495	430
477	375
425	396
322	383
482	387
587	542
370	359
322	357
479	528
375	513
330	426
300	513
386	358
401	581
645	588
440	522
317	545
319	430
298	375
253	396

Solution:

1. State the random variables and the parameters in words.

x_1 = sodium level in beef hotdogs

x_2 = sodium level in poultry hotdogs

μ_1 = mean sodium level in beef hotdogs

μ_2 = mean sodium level in poultry hotdogs

2. State the null and alternative hypotheses and the level of significance

The normal hypotheses would be

$$H_o : \mu_1 = \mu_2 \quad \text{or} \quad H_o : \mu_1 - \mu_2 = 0$$

$$H_A : \mu_1 < \mu_2 \quad \quad H_A : \mu_1 - \mu_2 < 0$$

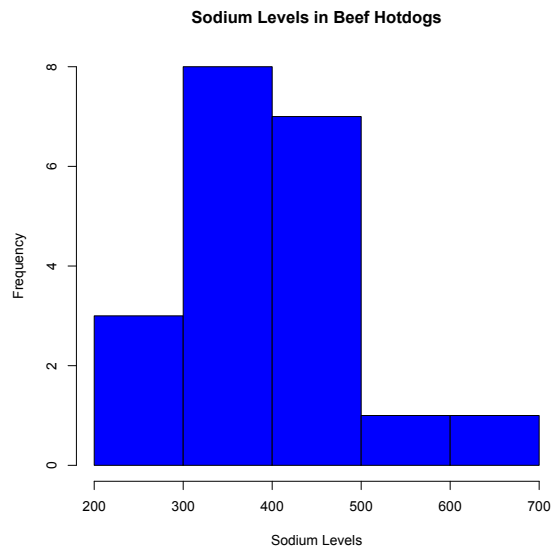
$$\alpha = 0.05$$

3. State and check the assumptions for the hypothesis test

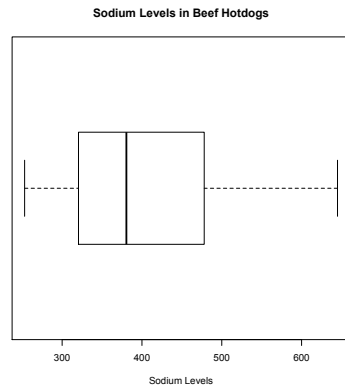
- A random sample of 20 sodium levels in beef hotdogs is taken. A random sample of 20 sodium levels in poultry hotdogs. The problem does not state if either sample was randomly selected. So this assumption may not be valid.
- The two samples are independent since these are different types of hotdogs.
- Population of all sodium levels in beef hotdogs is normally distributed.
Population of all sodium levels in poultry hotdogs is normally distributed.

Beef Hotdogs:

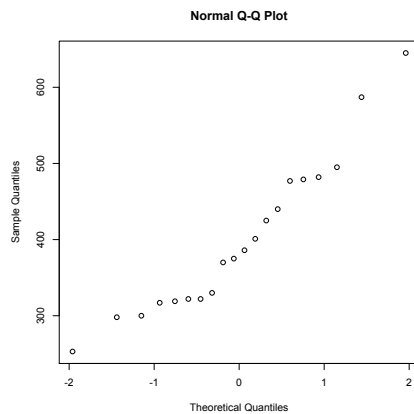
Figure #9.3.11: Histogram of Sodium Levels in Beef Hotdogs



This looks somewhat bell shaped.

Figure #9.3.12: Modified Box Plot of Sodium Levels in Beef Hotdogs

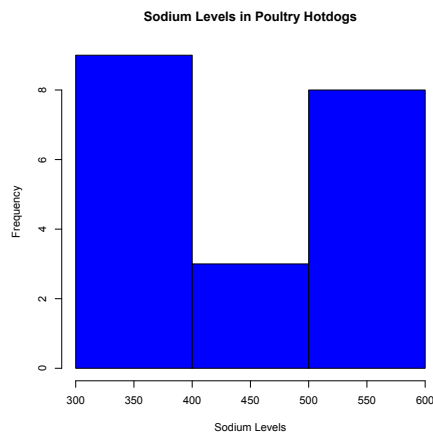
There are no outliers.

Figure #9.3.13: Normal Quantile Plot of Sodium Levels in Beef Hotdogs

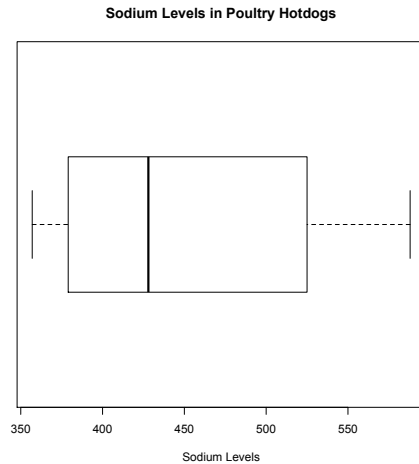
This looks somewhat linear.

So, the population of all sodium levels in beef hotdogs may be normally distributed.

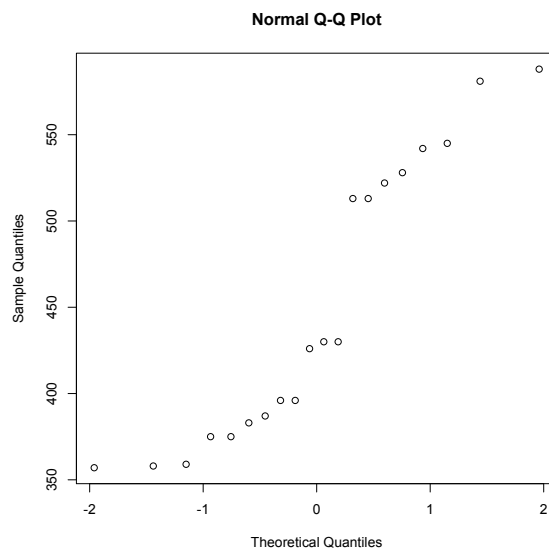
Poultry Hotdogs:

Figure #9.3.14: Histogram of Sodium Levels in Poultry Hotdogs

This does not look bell shaped.

Figure #9.3.15: Modified Box Plot of Sodium Levels in Poultry Hotdogs

There are no outliers.

Figure #9.3.16: Normal Quantile Plot of Sodium Levels in Poultry Hotdogs

This does not look linear.

So, the population of all sodium levels in poultry hotdogs is probably not normally distributed.

This assumption is not valid. Since the samples are fairly large, and the t-test is robust, it may not be a large issue. However, just realize that the conclusions of the test may not be valid.

- d. The population variances are equal, i.e. $\sigma_1^2 = \sigma_2^2$.

$$s_1 \approx 102.4347$$

$$s_2 \approx 81.1786$$

$$\frac{s_1^2}{s_2^2} = \frac{102.4347^2}{81.1786^2} \approx 1.592$$

Using TI-83/84: $\text{Fcdf}(1.592, 1E99, 19, 19) \approx 0.1597 \geq 0.05$

Using R: $1 - \text{pf}(1.592, 19, 19) \approx 0.1597 \geq 0.05$

So you can say that these variances are equal.

4. Find the sample statistic, test statistic, and p-value

Sample Statistic:

$$\bar{x}_1 = 401.15, \bar{x}_2 = 450.2, s_1 \approx 102.4347, s_2 \approx 81.1786, n_1 = 20, n_2 = 20$$

Test Statistic:

The assumption $\sigma_1^2 = \sigma_2^2$ has been met, so

$$\begin{aligned} s_p &= \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}} \\ &= \sqrt{\frac{102.4347^2 * 19 + 81.1786^2 * 19}{(20 - 1) + (20 - 1)}} \\ &\approx 92.4198 \end{aligned}$$

Though you should try to do the calculations in the problem so you don't create round off error.

$$\begin{aligned} t &= \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\ &= \frac{(401.15 - 450.2) - 0}{92.4198 \sqrt{\frac{1}{20} + \frac{1}{20}}} \\ &\approx -1.678 \\ df &= 20 + 20 - 2 = 38 \end{aligned}$$

p-value:

Using TI-83/84: $\text{tcdf}(-1E99, -1.678, 38) \approx 0.0508$

Using R: $\text{pt}(-1.678, 38) \approx 0.0508$

Using technology to find the t and p-value:
Using TI-83/84:

Figure #9.3.17: Setup for 2-SampTTest on TI-83/84 Calculator

```

2-SampTTest
Inpt: Data Stats
List1: L1
List2: L2
Freq1: 1
Freq2: 1
 $\mu_1$ :  $\neq \mu_2$   $< \mu_2$   $> \mu_2$ 
↓ Pooled: No Yes

```

Note: the Pooled question on the calculator is for whether you are using the pooled standard deviation or not. In this example, the pooled standard deviation was used since you are assuming the variances are equal. That is why the answer to the question is Yes.

Figure #9.3.18: Results for 2-SampTTest on TI-83/84 Calculator

```

2-SampTTest
 $\mu_1 < \mu_2$ 
t = -1.678317463
P = .0507460087
df = 38
 $\bar{x}_1$  = 401.15
↓  $\bar{x}_2$  = 450.2

2-SampTTest
 $\mu_1 < \mu_2$ 
↑ Sx1 = 102.434716
Sx2 = 81.1785556
SxP = 92.4197732
n1 = 20
n2 = 20

```

Using R: the command is `t.test(variable1, variable2, alternative="less" or "greater")`

For this example, the command is
`t.test(beef, poultry, alternative="less", equalvar=TRUE)`

Welch Two Sample t-test

data: beef and poultry

$t = -1.6783$, $df = 36.115$, $p\text{-value} = 0.05096$

alternative hypothesis: true difference in means is less than 0

95 percent confidence interval:

-Inf 0.2875363

sample estimates:

mean of x mean of y

401.15 450.20

The $t = -1.6783$ and the $p\text{-value} = 0.05096$.

5. Conclusion

Fail to reject H_o since the $p\text{-value} > \alpha$.

6. Interpretation

This is not enough evidence to show that beef hotdogs have less sodium than poultry hotdogs. (Though do realize that many of assumptions are not valid, so this interpretation may be invalid.)

Example #9.3.4: Confidence Interval for $\mu_1 - \mu_2$

The amount of sodium in beef hotdogs was measured. In addition, the amount of sodium in poultry hotdogs was also measured ("SOCR 012708 id," 2013). The data is in table #9.3.2. Find a 95% confidence interval for the mean difference in sodium levels between beef and poultry hotdogs.

Solution:

1. State the random variables and the parameters in words.

These were stated in example #9.3.1, but are reproduced here for reference.

x_1 = sodium level in beef hotdogs

x_2 = sodium level in poultry hotdogs

μ_1 = mean sodium level in beef hotdogs

μ_2 = mean sodium level in poultry hotdogs

2. State and check the assumptions for the hypothesis test

The assumptions were stated and checked in example #9.3.3.

3. Find the sample statistic and confidence interval

Sample Statistic:

$\bar{x}_1 = 401.15$, $\bar{x}_2 = 450.2$, $s_1 \approx 102.4347$, $s_2 \approx 81.1786$, $n_1 = 20$, $n_2 = 20$

Confidence Interval:

The confidence interval estimate of the difference $\mu_1 - \mu_2$ is

The assumption $\sigma_1^2 = \sigma_2^2$ has been met, so

$$\begin{aligned}
 s_p &= \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}} \\
 &= \sqrt{\frac{102.4347^2 * 19 + 81.1786^2 * 19}{(20 - 1) + (20 - 1)}} \\
 &\approx 92.4198
 \end{aligned}$$

Though you should try to do the calculations in the formula for E so you don't create round off error.

$$df = n_1 + n_2 - 2 = 20 + 20 - 2 = 38$$

$$t_c = 2.024$$

$$\begin{aligned}
 E &= t_c s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\
 &= 2.024(92.4198) \sqrt{\frac{1}{20} + \frac{1}{20}} \\
 &\approx 59.15
 \end{aligned}$$

$$(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$$

$$(401.15 - 450.2) - 59.15 < \mu_1 - \mu_2 < (401.15 - 450.2) + 59.15$$

$$-108.20 \text{ g} < \mu_1 - \mu_2 < 10.10 \text{ g}$$

Using technology:

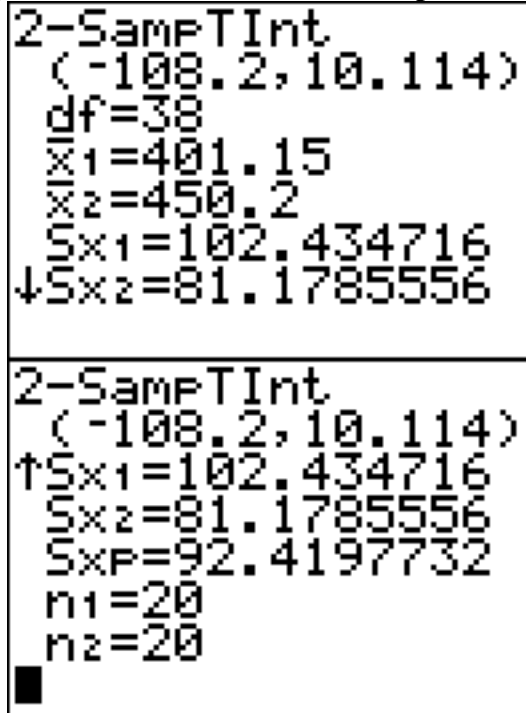
Using the TI-83/84:

Figure #9.3.19: Setup for 2-SampTInt on TI-83/84 Calculator



Note: the Pooled question on the calculator is for whether you are using the pooled standard deviation or not. In this example, the pooled standard deviation was used since you are assuming the variances are equal. That is why the answer to the question is Yes.

Figure #9.3.20: Results for 2-SampTInt on TI-83/84 Calculator



Using R: the command is `t.test(variable1, variable2, equalvar=TRUE, conf.level=C)`, where C is in decimal form.

For this example, the command is
`t.test(beef, poultry, conf.level=.95, equalvar=TRUE)`

Welch Two Sample t-test

data: beef and poultry
 $t = -1.6783$, $df = 36.115$, $p\text{-value} = 0.1019$
 alternative hypothesis: true difference in means is not equal to 0
 95 percent confidence interval:
 -108.31592 10.21592
 sample estimates:
 mean of x mean of y
 401.15 450.20

The confidence interval is $-108.32 < \mu_1 - \mu_2 < 10.22$.

4. Statistical Interpretation: There is a 95% chance that $-108.20 \text{ g} < \mu_1 - \mu_2 < 10.10 \text{ g}$ contains the true difference in means.
5. Real World Interpretation: The mean sodium level of beef hotdogs is anywhere from 108.20 g less than the mean sodium level of poultry hotdogs to 10.10 g more. (The negative sign on the lower limit implies that the first mean is less than

the second mean. The positive sign on the upper limit implies that the first mean is greater than the second mean.)

Realize that many of assumptions are not valid in this example, so the interpretation may be invalid.

Section 9.3: Homework

In each problem show all steps of the hypothesis test or confidence interval. If some of the assumptions are not met, note that the results of the test or interval may not be correct and then continue the process of the hypothesis test or confidence interval. Unless directed by your instructor, do not assume the variances are equal (except in problems 11 through 16).

- 1.) The income of males in each state of the United States, including the District of Columbia and Puerto Rico, are given in table #9.3.3, and the income of females is given in table #9.3.4 ("Median income of," 2013). Is there enough evidence to show that the mean income of males is more than of females? Test at the 1% level.

Table #9.3.3: Data of Income for Males

\$42,951	\$52,379	\$42,544	\$37,488	\$49,281	\$50,987	\$60,705
\$50,411	\$66,760	\$40,951	\$43,902	\$45,494	\$41,528	\$50,746
\$45,183	\$43,624	\$43,993	\$41,612	\$46,313	\$43,944	\$56,708
\$60,264	\$50,053	\$50,580	\$40,202	\$43,146	\$41,635	\$42,182
\$41,803	\$53,033	\$60,568	\$41,037	\$50,388	\$41,950	\$44,660
\$46,176	\$41,420	\$45,976	\$47,956	\$22,529	\$48,842	\$41,464
\$40,285	\$41,309	\$43,160	\$47,573	\$44,057	\$52,805	\$53,046
\$42,125	\$46,214	\$51,630				

Table #9.3.4: Data of Income for Females

\$31,862	\$40,550	\$36,048	\$30,752	\$41,817	\$40,236	\$47,476	\$40,500
\$60,332	\$33,823	\$35,438	\$37,242	\$31,238	\$39,150	\$34,023	\$33,745
\$33,269	\$32,684	\$31,844	\$34,599	\$48,748	\$46,185	\$36,931	\$40,416
\$29,548	\$33,865	\$31,067	\$33,424	\$35,484	\$41,021	\$47,155	\$32,316
\$42,113	\$33,459	\$32,462	\$35,746	\$31,274	\$36,027	\$37,089	\$22,117
\$41,412	\$31,330	\$31,329	\$33,184	\$35,301	\$32,843	\$38,177	\$40,969
\$40,993	\$29,688	\$35,890	\$34,381				

- 2.) The income of males in each state of the United States, including the District of Columbia and Puerto Rico, are given in table #9.3.3, and the income of females is given in table #9.3.4 ("Median income of," 2013). Compute a 99% confidence interval for the difference in incomes between males and females in the U.S.

- 3.) A study was conducted that measured the total brain volume (TBV) (in mm^3) of patients that had schizophrenia and patients that are considered normal. Table #9.3.5 contains the TBV of the normal patients and table #9.3.6 contains the TBV of schizophrenia patients ("SOCR data oct2009," 2013). Is there enough evidence to show that the patients with schizophrenia have less TBV on average than a patient that is considered normal? Test at the 10% level.

Table #9.3.5: Total Brain Volume (in mm^3) of Normal Patients

1663407	1583940	1299470	1535137	1431890	1578698
1453510	1650348	1288971	1366346	1326402	1503005
1474790	1317156	1441045	1463498	1650207	1523045
1441636	1432033	1420416	1480171	1360810	1410213
1574808	1502702	1203344	1319737	1688990	1292641
1512571	1635918				

Table #9.3.6: Total Brain Volume (in mm^3) of Schizophrenia Patients

1331777	1487886	1066075	1297327	1499983	1861991
1368378	1476891	1443775	1337827	1658258	1588132
1690182	1569413	1177002	1387893	1483763	1688950
1563593	1317885	1420249	1363859	1238979	1286638
1325525	1588573	1476254	1648209	1354054	1354649
1636119					

- 4.) A study was conducted that measured the total brain volume (TBV) (in mm^3) of patients that had schizophrenia and patients that are considered normal. Table #9.3.5 contains the TBV of the normal patients and table #9.3.6 contains the TBV of schizophrenia patients ("SOCR data oct2009," 2013). Compute a 90% confidence interval for the difference in TBV of normal patients and patients with Schizophrenia.
- 5.) The length of New Zealand (NZ) rivers that travel to the Pacific Ocean are given in table #9.3.7 and the lengths of NZ rivers that travel to the Tasman Sea are given in table #9.3.8 ("Length of NZ," 2013). Do the data provide enough evidence to show on average that the rivers that travel to the Pacific Ocean are longer than the rivers that travel to the Tasman Sea? Use a 5% level of significance.

Table #9.3.7: Lengths (in km) of NZ Rivers that Flow into the Pacific Ocean

209	48	169	138	64
97	161	95	145	90
121	80	56	64	209
64	72	288	322	

Table #9.3.8: Lengths (in km) of NZ Rivers that Flow into the Tasman Sea

76	64	68	64	37	32
32	51	56	40	64	56
80	121	177	56	80	35
72	72	108	48		

- 6.) The length of New Zealand (NZ) rivers that travel to the Pacific Ocean are given in table #9.3.7 and the lengths of NZ rivers that travel to the Tasman Sea are given in table #9.3.8 ("Length of NZ," 2013). Estimate the difference in mean lengths of rivers between rivers in NZ that travel to the Pacific Ocean and ones that travel to the Tasman Sea. Use a 95% confidence level.
- 7.) The number of cell phones per 100 residents in countries in Europe is given in table #9.3.9 for the year 2010. The number of cell phones per 100 residents in countries of the Americas is given in table #9.3.10 also for the year 2010 ("Population reference bureau," 2013). Is there enough evidence to show that the mean number of cell phones in countries of Europe is more than in countries of the Americas? Test at the 1% level.

Table #9.3.9: Number of Cell Phones per 100 Residents in Europe

100	76	100	130	75	84
112	84	138	133	118	134
126	188	129	93	64	128
124	122	109	121	127	152
96	63	99	95	151	147
123	95	67	67	118	125
110	115	140	115	141	77
98	102	102	112	118	118
54	23	121	126	47	

Table #9.3.10: Number of Cell Phones per 100 Residents in the Americas

158	117	106	159	53	50
78	66	88	92	42	3
150	72	86	113	50	58
70	109	37	32	85	101
75	69	55	115	95	73
86	157	100	119	81	113
87	105	96			

- 8.) The number of cell phones per 100 residents in countries in Europe is given in table #9.3.9 for the year 2010. The number of cell phones per 100 residents in countries of the Americas is given in table #9.3.10 also for the year 2010 ("Population reference bureau," 2013). Find the 98% confidence interval for the difference in mean number of cell phones per 100 residents in Europe and the Americas.

- 9.) A vitamin K shot is given to infants soon after birth. Nurses at Northbay Healthcare were involved in a study to see if how they handle the infants could reduce the pain the infants feel ("SOCR data nips," 2013). One of the measurements taken was how long, in seconds, the infant cried after being given the shot. A random sample was taken from the group that was given the shot using conventional methods (table #9.3.11), and a random sample was taken from the group that was given the shot where the mother held the infant prior to and during the shot (table #9.3.12). Is there enough evidence to show that infants cried less on average when they are held by their mothers than if held using conventional methods? Test at the 5% level.

Table #9.3.11: Crying Time of Infants Given Shots Using Conventional Methods

63	0	2	46	33	33
29	23	11	12	48	15
33	14	51	37	24	70
63	0	73	39	54	52
39	34	30	55	58	18

Table #9.3.12: Crying Time of Infants Given Shots Using New Methods

0	32	20	23	14	19
60	59	64	64	72	50
44	14	10	58	19	41
17	5	36	73	19	46
9	43	73	27	25	18

- 10.) A vitamin K shot is given to infants soon after birth. Nurses at Northbay Healthcare were involved in a study to see if how they handle the infants could reduce the pain the infants feel ("SOCR data nips," 2013). One of the measurements taken was how long, in seconds, the infant cried after being given the shot. A random sample was taken from the group that was given the shot using conventional methods (table #9.3.11), and a random sample was taken from the group that was given the shot where the mother held the infant prior to and during the shot (table #9.3.12). Calculate a 95% confidence interval for the mean difference in mean crying time after being given a vitamin K shot between infants held using conventional methods and infants held by their mothers.
- 11.) Redo problem 1 testing for the assumption of equal variances and then use the formula that utilizes the assumption of equal variances (follow the procedure in example 9.3.3).
- 12.) Redo problem 2 testing for the assumption of equal variances and then use the formula that utilizes the assumption of equal variances (follow the procedure in example 9.3.3).
- 13.) Redo problem 7 testing for the assumption of equal variances and then use the formula that utilizes the assumption of equal variances (follow the procedure in example 9.3.3).

- 14.) Redo problem 8 testing for the assumption of equal variances and then use the formula that utilizes the assumption of equal variances (follow the procedure in example 9.3.3).
- 15.) Redo problem 9 testing for the assumption of equal variances and then use the formula that utilizes the assumption of equal variances (follow the procedure in example 9.3.3).
- 16.) Redo problem 10 testing for the assumption of equal variances and then use the formula that utilizes the assumption of equal variances (follow the procedure in example 9.3.3).

Section 9.4 Which Analysis Should You Conduct?

One of the most important concept that you need to understand is deciding which analysis you should conduct for a particular situation. To help you to figure out the analysis to conduct, there are a series of questions you should ask yourself.

- 1.) Does the problem deal with mean or proportion?
Sometimes the problem states explicitly the words mean or proportion, but other times you have to figure it out based on the information you are given. If you counted number of individuals that responded in the affirmative to a question, then you are dealing with proportion. If you measured something, then you are dealing with mean.
- 2.) Does the problem have one or two samples?
So look to see if one group was measured or if two groups were measured. If you have the data sets, then it is usually easy to figure out if there is one or two samples, then there is either one data set or two data sets. If you don't have the data, then you need to decide if the problem describes collecting data from one group or from two groups.
- 3.) If you have two samples, then you need to determine if the samples are independent or dependent.
If the individuals are different for both samples, then most likely the samples are independent. If you can't tell, then determine if a data value from the first sample influences the data value in the second sample. In other words, can you pair data values together so you can find the difference, and that difference has meaning. If the answer is yes, then the samples are paired. Otherwise, the samples are independent.
- 4.) Does the situation involve a hypothesis test or a confidence interval?
If the problem talks about "do the data show", "is there evidence of", "test to see", then you are doing a hypothesis test. If the problem talks about "find the value", "estimate the" or "find the interval", then you are doing a confidence interval.

So if you have a situation that has two samples, independent samples, involving the mean, and is a hypothesis test, then you have a two-sample independent t-test. Now you look up the assumptions and the formula or technology process for doing this test. Every hypothesis test involves the same six steps, and you just have to use the correct assumptions and calculations. Every confidence interval has the same five steps, and again you just need to use the correct assumptions and calculations. So this is why it is so important to figure out what analysis you should conduct.

Data Sources:

AP exam scores. (2013, November 20). Retrieved from
http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data_Dinov_030708_APEXamScores

Buy sushi grade fish online. (2013, November 20). Retrieved from
<http://www.catalinaop.com/>

Center for Disease Control and Prevention, Prevalence of Autism Spectrum Disorders - Autism and Developmental Disabilities Monitoring Network. (2008). *Autism and developmental disabilities monitoring network-2012*. Retrieved from website:
<http://www.cdc.gov/ncbddd/autism/documents/ADDM-2012-Community-Report.pdf>

Cholesterol levels after heart attack. (2013, September 25). Retrieved from
<http://www.statsci.org/data/general/cholest.html>

Flanagan, R., Rooney, C., & Griffiths, C. (2005). Fatal poisoning in childhood, england & wales 1968-2000. *Forensic Science International*, 148:121-129, Retrieved from
http://www.cdc.gov/nchs/data/ice/fatal_poisoning_child.pdf

Friday the 13th datafile. (2013, November 25). Retrieved from
<http://lib.stat.cmu.edu/DASL/Datafiles/Fridaythe13th.html>

Gettler, L. T., McDade, T. W., Feranil, A. B., & Kuzawa, C. W. (2011). Longitudinal evidence that fatherhood decreases testosterone in human males. *The Proceedings of the National Academy of Sciences, PNAS 2011*, doi: 10.1073/pnas.1105403108
Length of NZ rivers. (2013, September 25). Retrieved from
<http://www.statsci.org/data/oz/nzrivers.html>

Lim, L. L. United Nations, International Labour Office. (2002). *Female labour-force participation*. Retrieved from website:
<http://www.un.org/esa/population/publications/completingfertility/RevisedLIMpaper.PDF>

Median income of males. (2013, October 9). Retrieved from
<http://www.prb.org/DataFinder/Topic/Rankings.aspx?ind=137>

Median income of males. (2013, October 9). Retrieved from
<http://www.prb.org/DataFinder/Topic/Rankings.aspx?ind=136>

NZ helmet size. (2013, September 25). Retrieved from
<http://www.statsci.org/data/oz/nzhelmet.html>

Olson, K., & Hanson, J. (1997). Using reiki to manage pain: a preliminary report. *Cancer Prev Control*, 1(2), 108-13. Retrieved from
<http://www.ncbi.nlm.nih.gov/pubmed/9765732>

Population reference bureau. (2013, October 8). Retrieved from <http://www.prb.org/DataFinder/Topic/Rankings.aspx?ind=25>

Seafood online. (2013, November 20). Retrieved from <http://www.allfreshseafood.com/>

SOCR 012708 id data hotdogs. (2013, November 13). Retrieved from http://wiki.stat.ucla.edu/socr/index.php/SOCR_012708_ID_Data_HotDogs

SOCR data nips infantvitK shotdata. (2013, November 16). Retrieved from http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data_NIPS_InfantVitK_ShotData

SOCR data Oct2009 id ni. (2013, November 16). Retrieved from http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data_Oct2009_ID_NI

Statistics brain. (2013, November 30). Retrieved from <http://www.statisticbrain.com/infidelity-statistics/>

Student t-distribution. (2013, November 25). Retrieved from <http://lib.stat.cmu.edu/DASL/Stories/student.html>

