Section 2F – Two-Population Mean & Proportion Confidence Intervals

Studying the differences between two populations is very common in statistics, however sampling variability makes it very difficult to determine. Think of it this way. We know that random samples are usually different from each other, so even if two populations were the same, the samples taken from those populations would be different. A key question to ask is why are the samples different? Are the samples different because the populations are different or are my samples different because of sampling variability? Here is another key question. Are my samples significantly different or only slightly different? Two-population confidence intervals are often used to answer these difficulty questions.

Before you can understand two-population confidence intervals, we have to take you back to arithmetic. A two-population confidence interval is the answer to a subtraction problem. Remember, the answer to a subtraction problem is often called the "difference". We need to understand how subtraction works and what a difference actually tells us.

Understanding Positive Differences

Suppose you subtract two numbers and the answer comes out positive. There is a positive difference. Is the first number bigger or smaller than the second number?   Let us look at an example.

$17 - 6 = +11$

What does this tell us? Since the difference comes out positive, we know that the first number (17) is larger than the second number (6). It actually tells us more than this. The answer of +11 tells us that the first number (17) is 11 units larger than the second number (6).

How does this translate to a two-population confidence interval?

A two-population confidence interval does not measure population 1 or population 2 individually. Instead, it measures the difference between the population parameters. Two-population mean confidence intervals measure $\mu_1 - \mu_2$ (the difference between the population means). Two-population proportion confidence intervals measure $\pi_1 - \pi_2$ or $p_1 - p_2$ (the difference between the population proportions). The key is that the confidence interval is the answer to a subtraction problem.

Example: We want to compare the population mean height of men and women. We used a random sample of 40 men's heights in inches and a random sample of 40 women's heights in inches. Putting the data into Statcato, we got the following two population mean confidence interval. Population 1 was men's heights and population 2 was women's heights. We will assume for now that this data did meet the assumptions to estimate the populations.

Confidence Intervals - Two population means: confidence level = 0.95
Samples of population 1 in C16 Men Ht (in)
Samples of population 2 in C2 Women Ht (in)

	N	Mean	Stdev
Population 1	40	68.335	3.020
Population 2	40	63.195	2.741

* Population standard deviations are unknown. *
DOF = 77
Margin of error = 1.284
95.0%CI = (3.8560, 6.4240)

First of all, 3.856 inches is NOT population 1 and 6.424 is NOT population 2. That is not how confidence intervals work. Remember this is an interval. It represents all of the numbers in between 3.856 and 6.424 inches and difference between the population means ($\mu_1 - \mu_2$) could be any of them. Think of the number line. Notice that all of the numbers between 3.856 and 6.424 are positive!
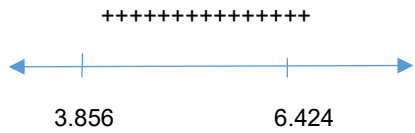
```
              ++++++++++++++
    <—————|—————————|—————>
        3.856         6.424
```

So while we do not know what the population difference is exactly, we do know that the difference is positive. Think back. Remember if the difference is positive, then the first number must be larger than the second. In this case, the mean average of population 1 (men's heights) is likely to be larger than the mean average of population 2 (women's heights). Remember the positive difference tells you how much larger.

Sentence to explain the confidence interval: We are 95% confident that the population mean average height of men (population 1) is between 3.856 and 6.424 inches <u>larger</u> than the population mean average height of women (population 2).

Note: You may also see the two-population confidence interval sentence written this way. We are 95% confident that the difference between the population mean average heights of men and women is between 3.856 and 6.424 inches. This can be a confusing way to explain the confidence interval though as people rarely understand the implications of that sentence.

Significance

Notice that the sample mean average height for the 40 men in the sample data was 68.335 inches and the sample mean average height for the 40 women in the sample data was 63.195 inches. Are these sample mean's significantly different? Yes. If both the upper and lower limits of your two population confidence interval are positive (+,+), then that does indicate that your sample statistic from group 1 is significantly higher than the sample statistic from group 2.

Positive Difference Two-population Confidence Intervals (+,+)

Sentence: "We are #% confident that the parameter from population 1 is between # and # larger than the parameter from population 2."

Significance: There is a significant difference between the two samples. The sample statistic for group 1 is significantly higher than group 2. This indicates that the parameter for population 1 might be higher than for population 2.

.

Example 2: Suppose we want to compare the percentage of statistics students that are democrat and the percentage of statistics students that are republican. We used the fall 2015 COC survey data to create the following confidence interval. For now, we will assume the problem met the assumptions for estimating the populations. Population 1 was democratic COC statistics students and population 2 was republican COC statistics students. We used a 90% confidence level and Statcato to calculate the following two-population proportion confidence interval.

## Confidence Interval - Two population proportions: confidence level = 0.9

	Number of Events	Number of trials	Proportion
Sample 1	110	328	0.335
Sample 2	63	328	0.192

Sample proportion difference = 0.143
Margin of error = 0.056
90.0%CI = (0.0874, 0.1991)

Notice that both of the numbers in the two-population confidence interval are positive. These proportions can be converted to their percentage equivalent (+8.74% , +19.91%). Again 8.74% is NOT population 1 and 19.91% is NOT population 2. That is not how two-population confidence intervals work. The difference between the population proportions ($\pi_1 - \pi_2$) could be any of the numbers between 0.0874 and 0.1991. Notice again that all of the numbers in between 0.0874 and 0.1991 are positive.

```
            +++++++++++++++
```



```
   0.0874              0.1991
```
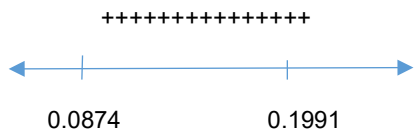
So the population proportion difference $\pi_1 - \pi_2$ is positive. This tells us that the population proportion (and percentage) of COC statistics students that are democrat (population 1) is likely to be larger than the population proportion or percentage of COC statistics students that are republican (population 2). Remember the positive difference tells you how much larger.

Sentence:

We are 90% confident that the population percentage of COC statistics students that are democratic (population 1) is between 8.74% and 19.91% <u>higher</u> than the percentage of COC statistics students that are republican (population 2).

OR

We are 90% confident that the population proportion of COC statistics students that are democratic (population 1) is between 0.0874 and 0.1991 <u>higher</u> than the population proportion of COC statistics students that are republican (population 2).

<u>Understanding Negative Differences</u>

Suppose you subtract two numbers and the answer comes out negative. (There is a negative difference.) Is the first number bigger or smaller than the second number? Let us look at an example.

5 – 13 = −8

What does this tell us? Since the difference comes out negative, we know that the first number (5) is smaller than the second number (13). It actually tells us more than this. The answer of −8 tells us that the first number (5) is eight

units smaller than the second number (13). Notice we did not say that the first number is −8 units smaller. The difference of −8 tells us that the first number is eight units smaller than the second number.

How does this translate to a two-population confidence interval?

Remember, a two-population confidence interval does not measure population 1 or population 2 individually. Instead, it measures the difference between the population parameters. Two-population mean confidence intervals measure $\mu_1 - \mu_2$ (the difference between the population means). Two-population proportion confidence intervals measure $\pi_1 - \pi_2$ or $p_1 - p_2$ (the difference between the population proportions).

Example: We want to compare the population mean weight of women and men. We used a random sample of 40 women's weights in pounds and a random sample of 40 men's weights in pounds. Putting the data into Statcato, we got the following two population mean confidence interval. Population 1 was women's weights and population 2 was men's weights. We will assume for now that this data did meet the assumptions to estimate the populations.

Confidence Intervals - Two population means: confidence level = 0.95
Samples of population 1 in C3 Women Wt (Lbs)
Samples of population 2 in C17 Men Wt (Lbs)

	N	Mean	Stdev
Population 1	40	146.220	37.621
Population 2	40	172.55	26.327

* Population standard deviations are unknown. *
DOF = 69
Margin of error = 14.484
95.0%CI = (-40.8135, -11.8465)

Remember, −40.8135 pounds is NOT population 1 and −11.8465 pounds is NOT population 2. That is not how two-population confidence intervals work. Remember this is an interval. It represents all of the numbers in between −40.8135 and −11.8465 pounds and difference between the population means ($\mu_1 - \mu_2$) could be any of them. Notice that the lower limit is now −40.8135 on the left and the upper limit is −11.8465 on the right. Many students are confused by this, but that is how the number line works. The more negative a number is, the smaller it is. Therefore, −40.8135 is smaller −11.8465.

How do we interpret this? Think again of the number line. Notice that all of the numbers between −40.8135 and −11.8465 are negative!
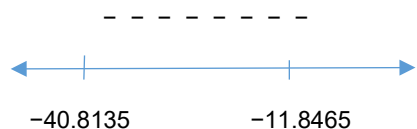
– – – – – – – –

<----+----------------+---->
−40.8135   −11.8465

So while we do not know what the population difference is exactly, we do know that the difference is negative. Remember if the difference is negative, then the first number must be smaller than the second number. In this case, the mean average of population 1 (women's weights) is likely to be smaller than the mean average of population 2 (men's weights). The negative difference tells you how much smaller.

Sentence to explain the confidence interval: We are 95% confident that the population mean average weight of women (population 1) is between 11.8465 pounds and 40.8135 pounds less than the population mean average height of men (population 2).

Significance

Notice that the sample mean average weight for the 40 women in the sample data was 146.220 pounds and the sample mean average height for the 40 men in the sample data was 172.55 pounds. Are these sample mean's significantly different? Yes. If both the upper and lower limits of your two population confidence interval are negative (−,−), then that does indicate that your sample statistic from group 1 is significantly lower than the sample statistic from group 2.

Negative Difference Two-population Confidence Intervals (−,−)

Sentence: "We are #% confident that the parameter from population 1 is between # and # lower than (or less than) the parameter from population 2."

Significance: There is a significant difference between the two samples. The sample statistic for group 1 is significantly lower than group 2. This indicates that parameter for population 1 is probably lower than for population 2.

Example 2: In a previous example, we compared the percentage of statistics students that are democrat and the percentage of statistics students that are republican. We assigned democrat to be population 1 and republican to be population 2. What would happen if we reverse that? Suppose we let population 1 to be republican COC statistics students and population 2 to be democrat COC statistics students. We used a 90% confidence level and Statcato to calculate the following two-population proportion confidence interval. Assume the problem met the assumptions for estimating the populations.

**Confidence Interval - Two population proportions: confidence level = 0.9**

	Number of Events	Number of trials	Proportion
Sample 1	63	328	0.192
Sample 2	110	328	0.335

Sample proportion difference = -0.143
Margin of error = 0.056
90.0%CI = (-0.1991, -0.0874)

Notice that the sample difference is now negative, but the margin of error is the same. Both of the numbers in the two-population confidence interval are now negative. These proportions can be converted to their percentage equivalent (−19.91%, −8.74%). Notice that these are the same percentages, but have opposite signs. Notice that the lower limit is now −19.91% on the left and the upper limit is −8.74% on the right. Remember, the more negative a number is, the smaller it is, so −19.91% is smaller −8.74%. The difference between the population proportions $(\pi_1 - \pi_2)$ could be any of the numbers between −0.1991 and −0.0874. Notice again that all of the numbers in between −0.1991 and −0.0874 are negative.

– – – – – – – –

−0.1991   −0.0874
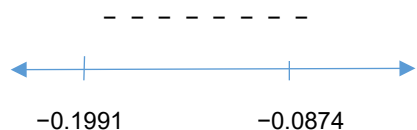
So the population proportion difference $\pi_1 - \pi_2$ is negative. This tells us that the population proportion (and percentage) of COC statistics students that are republican (population 1) is likely to be smaller than the population proportion or percentage of COC statistics students that are democrat (population 2). The confidence interval being negative tells you how much smaller.

Sentence:

We are 90% confident that the population percentage of COC statistics students that are republican (population 1) is between 8.74% and 19.91% lower than the percentage of COC statistics students that are democrat (population 2).

OR

We are 90% confident that the population proportion of COC statistics students that are republican (population 1) is between 0.0874 and 0.1991 lower than the population proportion of COC statistics students that are democrat (population 2).

Significance:

Since both the upper and lower limits of the confidence interval were negative, this suggests that the sample percentage for group 1 (republican) was significantly lower than the sample percentage for group 2 (democrat). This indicates that the population percentage for republican COC statistics students is likely to be lower than the percentage for democratic COC statistics students.

Zero Difference

If we subtract two numbers and the answer is zero, the two numbers must be the same. Look at the following example.

13 – 13 = 0

The zero difference tells us that the first number (13) is the same as the second number (13).

Example: Suppose 95% two-population proportion confidence interval came out to be (−0.028 , +0.017). Remember, −0.028 is NOT population 1 and +0.017 is NOT population 2. This confidence interval tells us that the difference between the population proportions $(\pi_1 - \pi_2)$ is somewhere between −0.028 and +0.017. Some people will write the sentence as follows.

Sentence: We are 95% confident that the population proportion difference is between −0.028 and +0.017.

What does that even mean? Is population 1 lower or higher than population 2? How much lower or higher? To answer these questions, we need to examine the number line between −0.028 and +0.017. Notice that there are many negative numbers in this interval, so population 1 may be lower than population 2. There are also many positive numbers in this interval, so population 1 may be higher than population 2. Zero is also in the interval, so it is also a possibility. Remember if the difference is zero, then population 1 and population 2 could be the same. This interval tells us that we really do not know which population is larger. When the upper and lower limits of a two-population confidence interval have opposite signs, this means there is no significant difference between the populations. The sample statistics for the two groups are so close, that we cannot tell if population 1 is lower or higher than population 2. They could be the same.
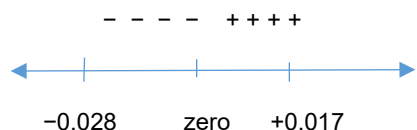

$-\ -\ -\ -\quad +\ +\ +\ +$

−0.028   zero   +0.017

Two-population Confidence Intervals (− , +)

Sentence: We are #% confident that there is no significant difference between the parameter for population 1 and parameter for population 2.

Significance: When the upper and lower limits for the two-population confidence interval have opposite signs, then that indicates that the sample statistics for the two groups are not significantly different.

Example: Let us compare the population mean average salary of people living in Northern Territory, Australia (μ_1) to people living in New South Wales, Australia (μ_2). We used StatKey and random sample data to create the following two-population mean bootstrap 90% confidence interval. Assume the data met all of the assumptions.

Original Sample

$\bar{x}_1 - \bar{x}_2 = -2.28$, $n_1 = 35$, $n_2 = 35$

Bootstrap Dotplot of $\bar{x}_1 - \bar{x}_2$
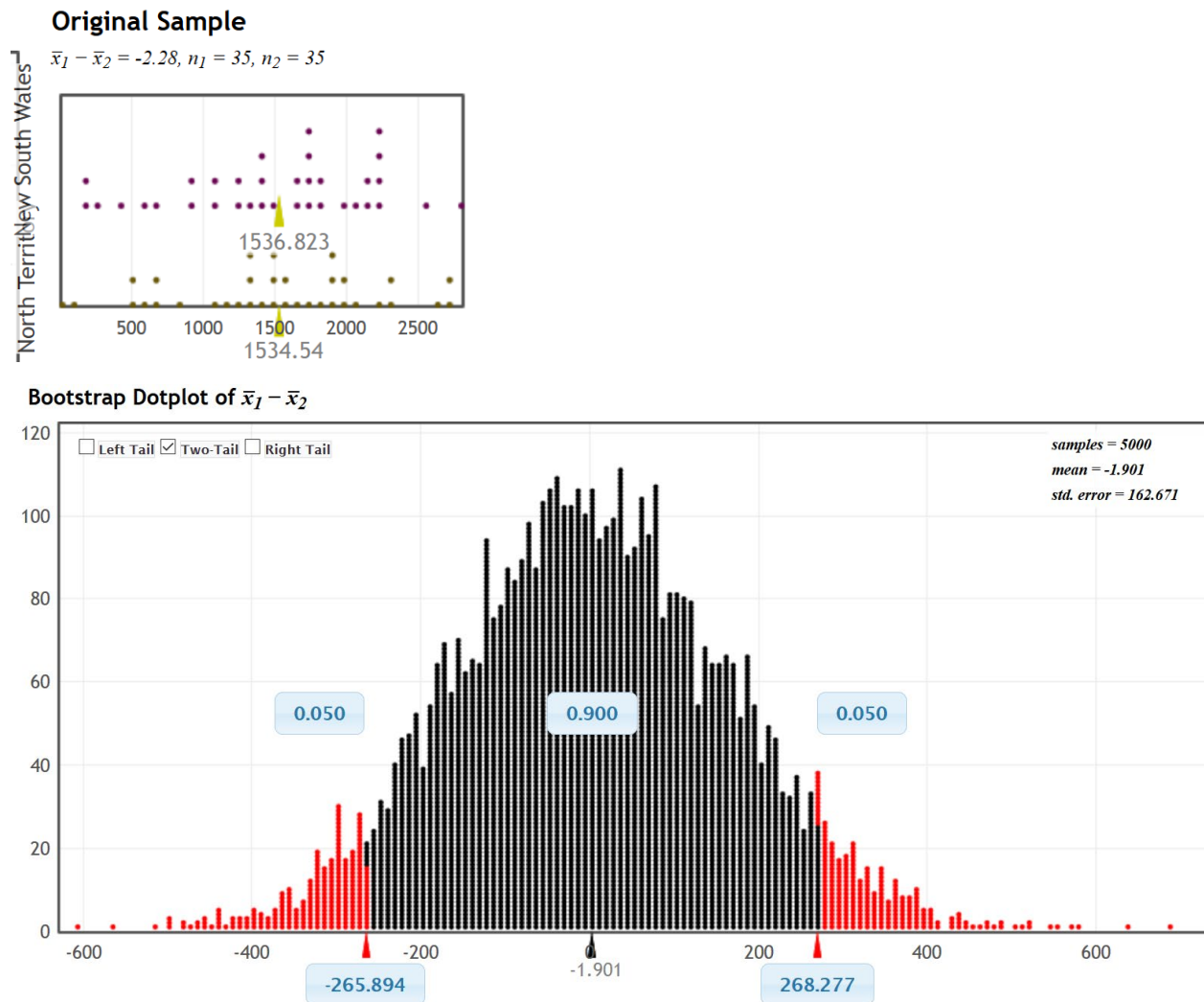
From the bootstrap, we see that the 90% confidence interval is (−265.894 , +268.277). Notice the upper and lower limit have opposite signs. This tells us that the sample mean average salary for people living in the Northern Territory ($1534.54) is not significantly different from the sample mean average salary for people living in New South Wales ($1536.82). Since our sample means are so close, we cannot tell which population has a higher population mean average salary.

Confidence Interval Sentence: We are 90% confident that the difference between the population mean average salary of people living in Northern Territory, Australia and those living in New South Wales, Australia is between −$265.894 and +$268.277. *(This sentence tends to be confusing.)*

Better Confidence Interval Sentence: We are 90% confident that there is no significant difference between the population mean average salary of people living in Northern Territory, Australia and those living in New South Wales, Australia.

Note: We could also have calculated the 90% confidence interval with Statcato. Notice the upper and lower limits of the bootstrap are similar to what Statcato calculated.

Confidence Intervals - Two population means: confidence level = 0.9
Samples of population 1 in C1 North Territory ...
Samples of population 2 in C2 New South Wales ...

	N	Mean	Stdev
Population 1	35	1534.540	701.525
Population 2	35	1536.823	677.140

* Population standard deviations are unknown. *
DOF = 67
Margin of error = 274.883
90.0%CI = (-277.1660, 272.5998)

Calculating Two-population Mean and Proportion Confidence Intervals

We will now discuss the formulas and calculations for two-population mean and proportion confidence intervals. It is important to understand the formulas and be able to explain them. However, no statistician or data scientist calculates these by hand with a formula. We virtually always use computer software to calculate any difficult calculations like confidence intervals.

Two-population Mean Confidence Intervals

There are two types of two-population mean confidence intervals, independent groups and matched pairs. Matched pair data is a one-to-one pairing between the two groups. Matched pair data usuallyfrom the same person measured twice. For example, the first number in the first data set comes from the same person as the first number in the second data set. The second numbers in each data set come from the same person and so on. Matched pairs do not have to be the same person measured twice. It could also be comparing husbands and wives, or sisters and brothers. You could be comparing two football teams and comparing the salary for each position: the starting quarterbacks, the starting running backs, the starting right guard, etc. Notice that in matched pairs, the sample sizes for the two groups are the same.

Use independent groups when you are comparing separate groups. For example, like comparing a random sample of men to a random sample of women or comparing a random sample of people from California to a random sample of people from Arizona.

Example 1 (Matched Pair): Let's use the random sample health data and a 99% confidence interval to compare the population mean systolic and diastolic blood pressure for men. Since these values come from the same 40 men, they are matched pairs.

Population 1: Men's Systolic Blood Pressure (mm of Hg)
Population 2: Men's Diastolic Blood Pressure (mm of Hg)

For independent groups, we calculate sample mean and sample standard deviation separately for each group and then subtract the sample means. For matched pair, we subtract the ordered pairs first, and then calculate the mean of the difference ($\overline{d}$) and the standard deviation of the difference (s_d).

Since the systolic and diastolic blood pressure for these 40 men were matched pairs, notice we subtracted each pair and created a new column of data called the "difference" column. A two-population mean matched pair confidence interval is calculating a one-population confidence interval using just the difference column.

Men Syst BP (mm of Hg)	Men Diast BP (mm of Hg)	Difference between pairs
125	78	47
107	54	53
126	81	45
110	68	42
110	66	44
107	83	24
113	71	42
126	72	54
137	85	52
110	71	39
109	65	44
153	87	66
112	77	35
119	81	38
113	82	31
125	76	49
131	80	51
121	75	46
132	81	51
112	44	68
121	65	56
116	64	52
95	58	37
110	70	40
110	66	44

Descriptive Statistics

Variable	Mean	Standard Deviation
C3 Difference between pairs	45.675	9.352

The sample mean of the difference ($\overline{d}$) is 45.675, the sample standard deviation of the difference (s_d) is 9.352 and the sample size (n) is 40. These are used in the confidence interval calculation. We will also need to look up the T critical value. In matched pair, the sample size is the number of pairs (40), so the degrees of freedom is 39. We can use the theoretical T distribution function in StatKey to look up the critical value.
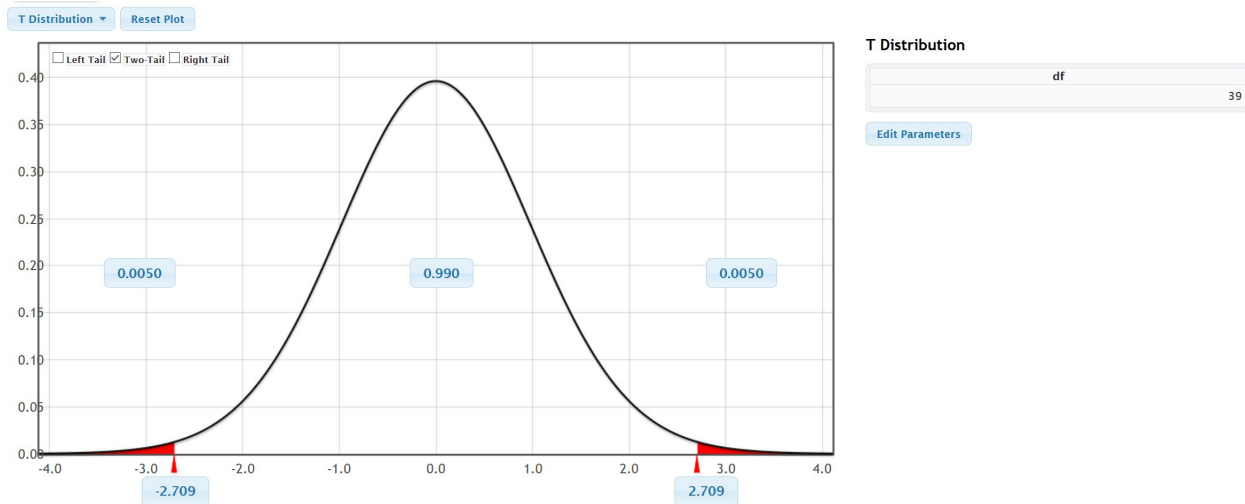
Notice that the T critical values are ±2.709. Here is the formula and calculation for the two-population mean matched pair confidence interval. Notice that the sample mean difference is 45.675 mm of Hg and the margin of error is 4.0057 mm of Hg. This gave us a confidence interval of

$$\overline{d} \pm T\frac{s_d}{\sqrt{n}}$$

$$45.675 \pm 2.709 \frac{9.352}{\sqrt{40}}$$

$$45.675 \pm 4.0057$$

(41.6693 , 49.6807)

Notice that the upper and lower limits of the confidence interval are both positive. This tells us that the population 1 (men's systolic blood pressure) is higher than population 2 (men's diastolic blood pressure).

Sentence: We are 99% confident that the population mean systolic blood pressure for men is between 41.67 mm of Hg and 49.68 mm of Hg higher than the population mean diastolic blood pressure for men.

We can use Statcato to calculate this for us. Just go to the "statistics" menu in Statcato, click on "confidence intervals" and then matched pair. You can put in the summary data (sample mean difference 45.675, sample standard deviation of the differences 9.352, and sample size 40). You can also copy and paste the two quantitative data sets and then click the "samples in two columns" button.

Statcato => Statistics => Confidence Intervals => Matched Pairs

Confidence Interval - Matched Pairs: confidence level = 0.99
Sample 1: C1 Men Syst BP (mm ...
Sample 2: C2 Men Diast BP (mm...
Difference of Matched Pairs C1 Men Syst BP (mm ... - C2 Men Diast BP (mm...

N	Mean	Stdev	Margin of Error	99.0%CI
40	45.675	9.352	4.004	(41.6710, 49.6790)

Notice that the confidence interval in Statcato is virtually the same as our formula calculation above.

We can also use bootstrapping in StatKey to calculate this confidence interval. Remember a matched pair is calculated as a one-population mean bootstrap from the differences between the pairs.   Let us start by calculating the difference column in Excel. Copy and pasted the two data sets into excel. In cell "C2" type in "=B2-C2" and push enter. Hold your curser on the bottom right corner until it turns into a "+". Double click and the formula will be applied to the rest of the data. You can also click and drag.

	A	B	C
1	Men Syst BP (mm of Hg)	Men Diast BP (mm of Hg)	Difference (Systolic - Diastolic)
2	125	78	47
3	107	54	53
4	126	81	45
5	110	68	42
6	110	66	44
7	107	83	24
8	113	71	42
9	126	72	54
10	137	85	52
11	110	71	39
12	109	65	44
13	153	87	66
14	112	77	35
15	119	81	38
16	113	82	31
17	125	76	49
18	131	80	51
19	121	75	46
20	132	81	51
21	112	44	68
22	121	65	56
23	116	64	52
24	95	58	37
25	110	70	40
26	110	66	44
27	125	82	43
28	124	79	45
29	131	69	62
30	109	64	45
31	112	79	33
32	127	72	55
33	132	74	58
34	116	81	35
35	125	84	41
36	112	77	35
37	125	77	48
38	120	83	37
39	118	68	50
40	115	75	40
41	115	65	50
42			

Cell reference: C2 fx =A2-B2

Open StatKey at www.lock5stat.com and click on "CI for Single Mean, Median, St.Dev." under the "bootstrap confidence interval" menu. Make sure the bootstrap dot plot is set to "mean". Click on edit data. Copy and paste the "difference" column only and push "Ok".

Edit data ✖

Difference (Systolic - Diastolic)
47
53
45
42
44
24
42
54
52
39
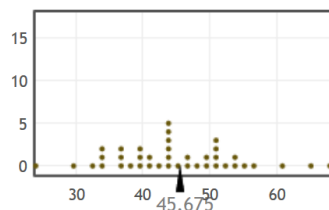44
66
35
38
31
49
51
46
51
68

☐ First column is identifier

☑ Data has header row

Manually edit the values above or paste a tab or comma seperated file into the box and click Ok. The file must have only one column (or two if there is an identifier).
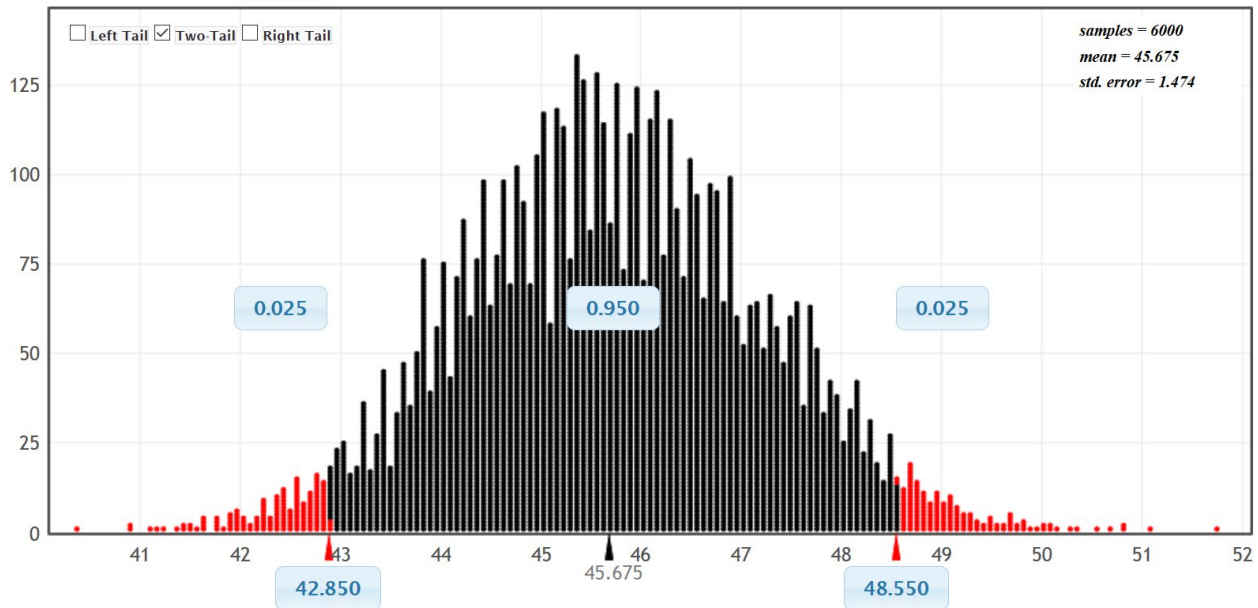
Ok

Original Sample

n = 40, mean = 45.675
median = 45, stdev = 9.352

Now create the bootstrap distribution by clicking on "generate 1000 samples" a few times and click on two-tail. The default is 95% but you can change the middle proportion to 0.90 or 0.99 if needed. Notice the 95% bootstrap confidence interval is (+42.85 , +48.55). This is similar to our formula calculations above.

Bootstrap Dotplot of [Mean ▾]

samples = 6000
mean = 45.675
std. error = 1.474

☐ Left Tail ☑ Two-Tail ☐ Right Tail

0.025   0.950   0.025

45.675

42.850   48.550

Important Notes about two-population bootstraps: Remember bootstrap confidence intervals will always come out slightly different because of sampling variability. Also that though we used a one-population bootstrap, this was not a one-population confidence interval. It measured the difference between the populations and must be interpreted accordingly. Remember to keep track of population 1 and population 2 and the signs of the confidence intervals.

Example 2 (Two-population mean from Independent Groups): Earlier we used the health data to calculate the following two-population confidence interval to compare the population mean average weight of women and men. Notice these groups are independent and not a one-to-one pairing. The upper and lower limits were negative, indicating that we are 95% confident that the population mean average weight of women is between 11.8465 pounds and 40.8135 pounds <u>less</u> than the population mean average weight of men.

Confidence Intervals - Two population means: confidence level = 0.95
Samples of population 1 in C3 Women Wt (Lbs)
Samples of population 2 in C17 Men Wt (Lbs)

	N	Mean	Stdev
Population 1	40	146.220	37.621
Population 2	40	172.55	26.327

* Population standard deviations are unknown. *
DOF = 69
Margin of error = 14.484
95.0%CI = (-40.8135, -11.8465)

Let us discuss how Statcato calculated this confidence interval. Let us start with the degrees of freedom. For independent groups, the degrees of freedom calculation is much more difficult. There are many free online

calculators for degrees of freedom. I like to use this one. You will need to enter the sample size and sample standard deviation for each of your two samples. Notice the degrees of freedom calculator gave 69.809. It is usually common to round down the degrees of freedom to account for possible greater variability. Notice Statcato rounded this degree of freedom down to 69 even though it was closer to 70.

http://web.utk.edu/~cwiek/TwoSampleDoF

Compute Degrees of Freedom for t-test comparing means of two independent samples

Enter in the sample sizes (n1, n2) and sample standard deviations (s1, s2) and click "Compute DF" to get the degrees of freedom describing the sampling distribution of the difference in sample means.

n1: 40   n2: 40   s1: 37.621   s2: 26.327   Compute DF   DF: 69.809

We can now look up the critical value T-scores for this confidence interval with the Theoretical Distribution T-score calculator in StatKey. Notice the critical value T-scores for 69 degrees of freedom are ± 1.995.
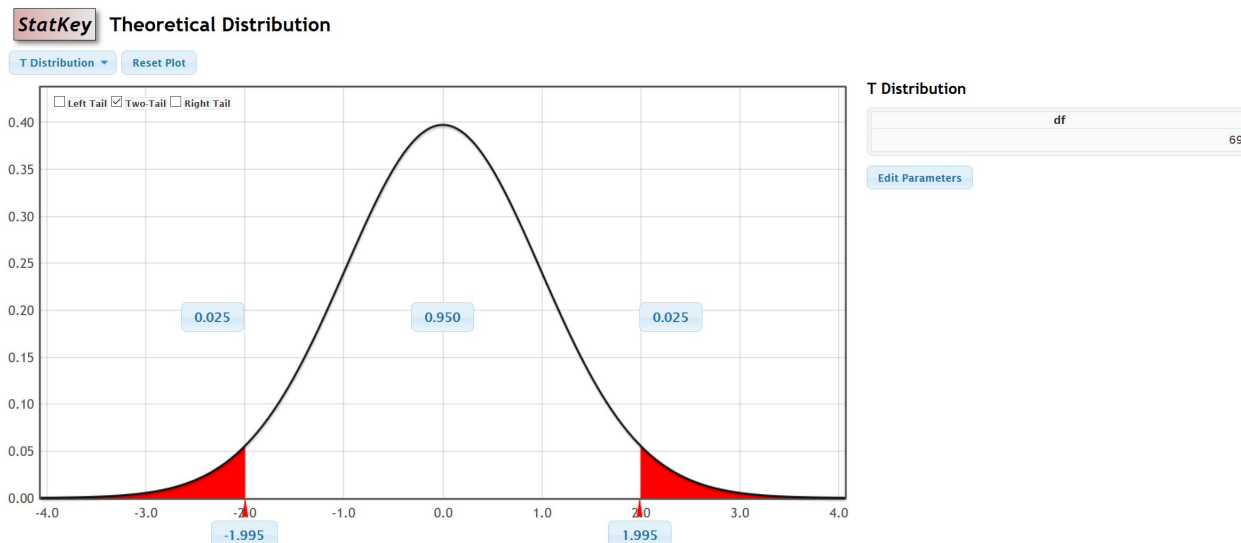
Here is the formula for the two-population mean confidence interval for independent groups. We see that the sample mean weight for the women ($\bar{x}_1$) was 146.220 pounds, the sample mean weight for the men ($\bar{x}_2$) was 172.55 pounds, the sample standard deviation for the women's weights (s_1) was 37.621 pounds and the sample standard deviation for the men's weights (s_2) is 26.327 pounds. While both sample sizes are 40 in this example, it is common for independent groups to have different sample sizes.

$$(\bar{x}_1 - \bar{x}_2) \pm T \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$$

$$(146.22 - 172.55) \pm 1.995 \sqrt{\left(\frac{37.621^2}{40} + \frac{26.327^2}{40}\right)}$$

−26.33 ± 1.995 (7.26)

−26.33 ± 14.48
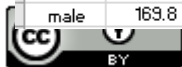
(−40.81 , −11.85)

Note about Pooling the Variances: Statisticians sometimes pool the variances when comparing the population means from two populations. It requires the population variances to be the same. For students new to stats, it is better not to pool the variances.

We can also calculate this confidence interval with bootstrapping. Go to www.lock5stat.com and click on StatKey. Under the "Bootstrap Confidence Intervals" menu, click on "CI for Difference in Means". This menu is for independent groups.   While Statcato prefers the data to be separated by group, StatKey prefers to have the categorical and quantitative data. Copy and paste the raw gender and weight data into a new excel spreadsheet first. They need to be next to each other. Now click on "Edit Data". Copy and paste the two columns into StatKey and push "Ok".

A	B
Gender	Weight (Lbs)
Female	114.8
Female	149.3
Female	107.8
Female	160.1
Female	127.1
Female	123.1
Female	111.7
Female	156.3
Female	218.8
Female	110.2
Female	188.3
Female	105.4
Female	136.1
Female	182.4
Female	238.4
Female	108.8
Female	119
Female	161.9
Female	174.1
Female	181.2
Female	124.3
Female	255.9
Female	106.7
Female	149.9
Female	163.1
Female	94.3
Female	159.7
Female	162.8
Female	130
Female	179.9
Female	147.8
Female	112.9
Female	195.6
Female	124.2
Female	135
Female	141.4
Female	123.9
Female	135.5
Female	130.4
Female	100.7
male	169.1
male	144.2
male	179.3
male	175.8
male	152.6
male	166.8
male	135
male	201.5
male	175.2
male	139
male	156.3
male	186.6
male	191.1
male	151.3
male	209.4
male	237.1
male	176.7
male	220.6
male	166.1
male	137.4
male	164.2
male	162.4
male	151.8
male	144.1
male	204.6
male	193.8
male	172.9
male	161.9
male	174.8
male	169.8

Edit data ✖

```
Gender    Weight (Lbs)
Female    114.8
Female    149.3
Female    107.8
Female    160.1
Female    127.1
Female    123.1
Female    111.7
Female    156.3
Female    218.8
Female    110.2
Female    188.3
Female    105.4
Female    136.1
Female    182.4
Female    238.4
Female    108.8
Female    119
Female    161.9
Female    174.1
Female    181.2
```
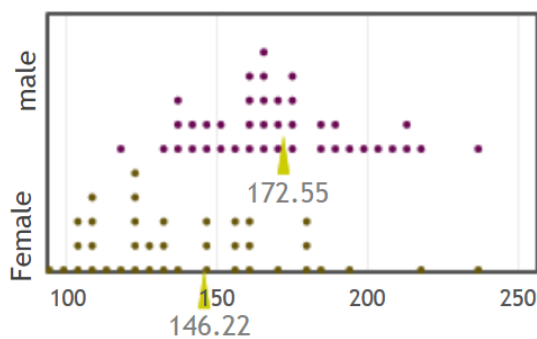
☑ Data has header row

Manually edit the values above or paste a tab or comma seperated file into the box and click Ok. The file must have only two columns where the first column is the categorical variable and the second is the quantitative.
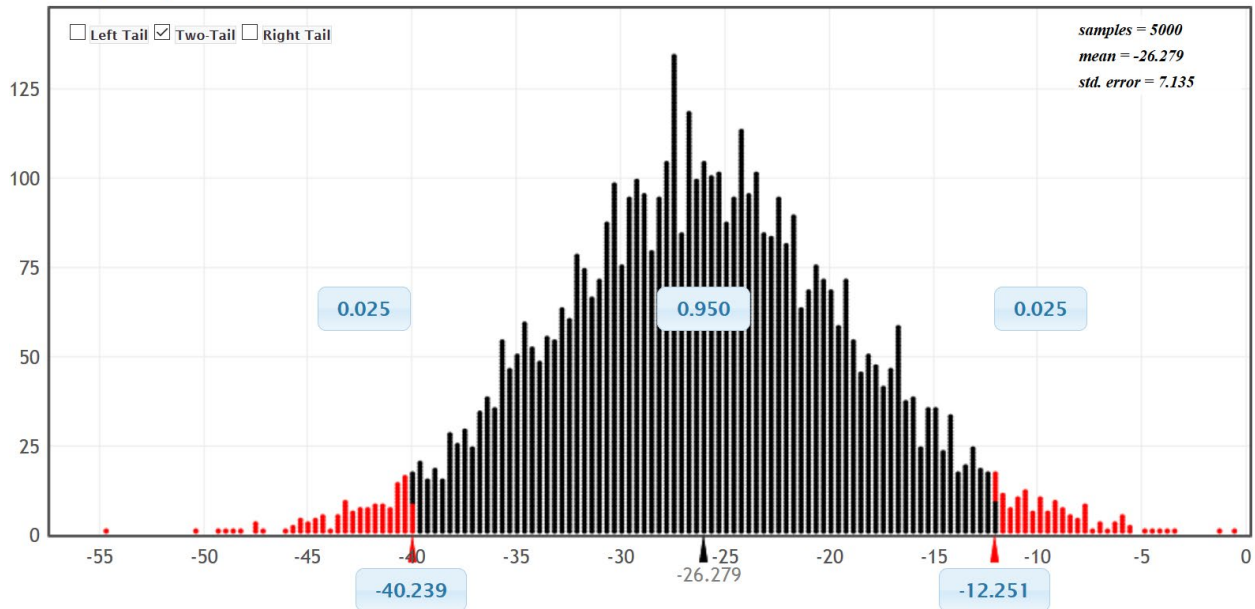
Ok

Original Sample

$\bar{x}_1 - \bar{x}_2 = -26.33,\ n_1 = 40,\ n_2 = 40$

Now create the bootstrap distribution by clicking on "generate 1000 samples" a few times and click on two-tail. The default is 95% but you can change the middle proportion to 0.90 or 0.99 if needed. Notice the 95% bootstrap confidence interval is (−40.239 , −12.251). This is similar to our formula calculations above.

Bootstrap Dotplot of $\bar{x}_1 - \bar{x}_2$

Important notes about two-population bootstraps: Remember bootstrap confidence intervals will always come out slightly different because of sampling variability. The two numbers at the bottom of the bootstrap distribution are the upper and lower limits of the confidence interval. For two-population, we need to keep track of population 1 and population 2 and the signs of the confidence intervals. In this case, population 1 was women's weights and population 2 was men's weights and the upper and lower limits were both negative.

Example 3 (Two-population proportion): Let's use the Math 140 survey data fall 2015 to compare the population percentage (proportion) of COC statistics students born in June (population 1) and the percentage (proportion) of COC statistics students born in December (population 2). We will assume the data met all of the assumptions for a two-population proportion confidence interval. The sample data showed that of the 336 total COC statistics students, 15 were born in June and 41 were born in December. We can use Statcato and a 90% confidence level to calculate the two-population proportion confidence interval. Just go to the "statistics" menu, and then click on "confidence intervals" and then "two-population proportion". Some computer programs will ask if you want to pool the samples. This means that you combine the two samples before calculating the standard error. Pooling is a technique used in hypothesis testing, but we do not pool the samples for two-population proportion confidence intervals.

Statcato => Statistics => Confidence Intervals => Two population Proportion

Note: Do not pool the sample proportions for confidence intervals.

Confidence Interval - Two population proportions: confidence level = 0.9

	Number of Events	Number of trials	Proportion
Sample 1	15	336	0.045
Sample 2	41	336	0.122

Sample proportion difference = -0.077
Margin of error = 0.035
90.0%CI = (-0.1121, -0.0427)

We see that the upper and lower limits of the confidence interval are both negative. This indicates that the proportion of COC statistics students born in June (population 1) is between 0.0427 and 0.1121 lower than the proportion of COC statistics students born in December (population 2).

Two-population proportion formula: Let us look at how this was calculated. Here is the two-population proportion formula. The sample proportion for group 1 $(\hat{p}_1)$ was 15 ÷ 336 ≈ 0.04464 and the sample proportion for group 2 $(\hat{p}_2)$ was 41 ÷ 336 ≈ 0.12202. Remember the famous Z-score critical value for 90% confidence is Z = ± 1.645

$$(\hat{p}_1 - \hat{p}_2) \pm Z \sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)}$$

$$(0.04464 - 0.12202) \pm 1.645 \sqrt{\frac{0.04464(1-0.4464)}{336} + \frac{0.12202(1-0.12202)}{336}}$$

$$-0.07738 \pm 0.034731$$

$$(-0.1121 , -0.04265)$$

Bootstrapping: We can also calculate this confidence interval with bootstrapping. Go to www.lock5stat.com and click on "StatKey". Under the "bootstrap confidence interval" section click on "difference in proportions". Click on the "Edit Data" button, and then enter the counts and sample sizes for both groups. Remember June was group 1 and December was group 2. Then push "Ok". Now generate a few thousand bootstrap samples, click two tail, and then put "0.90" for the middle proportion.

Edit data ✖

Please select values for two categories of count and sample size.

Group 1 count: 15
Group 1 sample size: 336
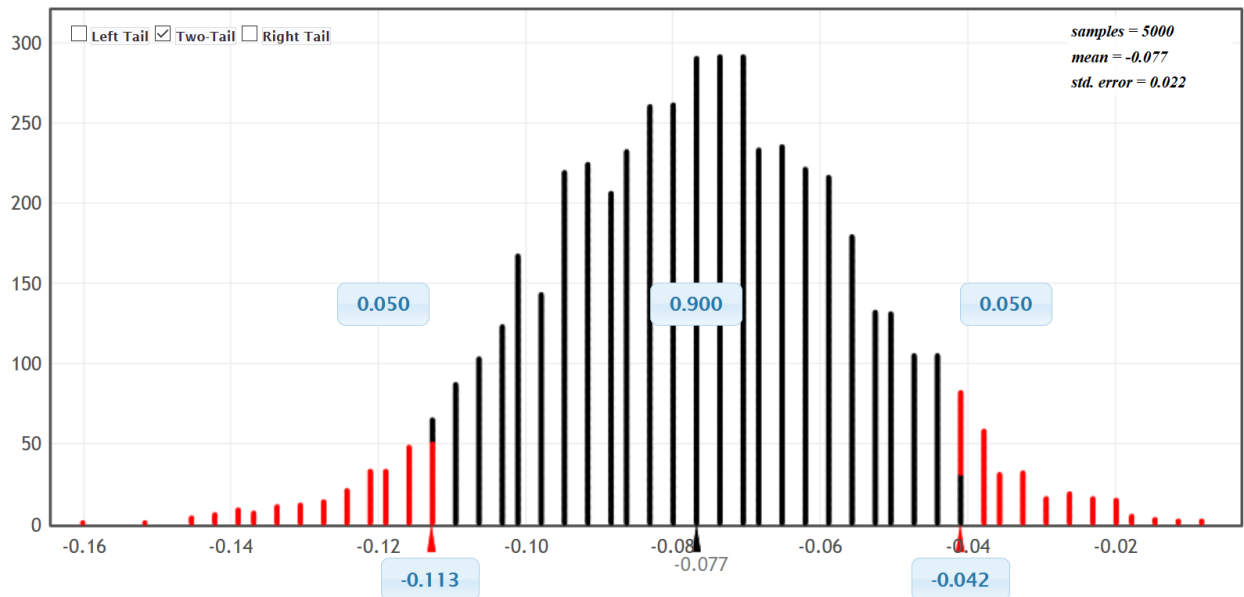Group 2 count: 41
Group 2 sample size: 336

Ok

Original Sample

Group	Count	Sample Size	Proportion
Group 1	15	336	0.045
Group 2	41	336	0.122
Group 1-Group 2	-26	n/a	-0.077

Bootstrap Dotplot of $\hat{p}_1 - \hat{p}_2$

Notice the bootstrap distribution is normal and centered close to the sample proportion difference of −0.077. It also indicates that the 90% confidence interval is (−0.113 , −0.042). This is close to what we got from the formula and Statcato.

Checking Assumptions

In order to compare populations, our sample data must be representative of the population. We usually require both samples to be large, random, and unbiased. The following assumptions are often used to check whether the sample data represents the population or not. Remember, if the sample data does not meet all of the assumptions, then we will not be able to draw any conclusions about the populations. It is also important to remember that these assumptions do not address all possible sources of bias.

Note about Independence:

- It is difficult to know for sure whether samples or individuals are indeed independent of each other. Individuals taken from two simple random samples from large populations will most likely be independent. A simple random sample of 50 people taken from a population of millions, will probably pass the individuals independent requirement. It is unlikely that we accidentally got people from the same family or people that work for the same company.

- Data collected conveniently or from voluntary response may fail the independence requirements. For example, if the sample data was collected from people in the same coffee shop or store, or on the same Facebook page, then they may be related or friends. This data would probably fail the independence requirements.

- Matched pair data means that there is a one-to-one pairing between the two samples. Usually it is the same people or objects measured twice. If the data is not matched pair, we often use the formulas for independent samples.

Notes about Bootstrapping:

- Bootstrapping does not require as many assumptions as traditional formula approaches and is often used when sample data fails the sample size requirements. However, bootstrapping does require the random and independence assumptions.

Notes about Experiments:

- Two-population confidence intervals can also be used in experimental design in order to prove cause and effect. In an experiment, the groups will not be random samples. They will need to be randomly assigned instead. Random assignment controls confounding variables.
- If the experiment uses random assignment, passes the assumptions, and shows a significant difference between the groups, then it indicates cause and effect.

Two-population Mean Assumptions (Matched Pair)

- The quantitative ordered pair sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- There should be at least thirty ordered pairs or the differences should have a nearly normal shape.

Two-population Mean Assumptions (Not Matched Pair)

- The two quantitative samples should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- Data values between the two samples should be independent of each other.
- The sample sizes should be at least 30 or have a nearly normal shape.

Two-population Proportion Assumptions

- The two categorical samples should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- Data values between the samples should be independent of each other.
- There should be at least ten successes and at least ten failures.

Two-population Bootstrap Assumptions

- The sample data should be collected randomly or be representative of the population.
- Data values within each sample should be independent of each other.
- If multiple samples were collected that were not matched pair, then the data values between the samples should be independent of each other.

Checking Assumptions Example 1: Earlier, we used the Math 140 survey data from fall 2015 to compare the population percentage (proportion) of COC statistics students born in June (population 1) and the percentage (proportion) of COC statistics students born in December (population 2). The sample data showed that of the 336 total COC statistics students, 15 were born in June and 41 were born in December. Would this data meet all of the assumptions for two-population confidence intervals with the traditional formula approach?

<u>Traditional Formula Assumptions for Comparing Two-Population Proportions (Percentages)</u>

- The two categorical samples should be collected randomly or be representative of the population.

  No. The month a person is born in is categorical; however, the sample data was not collected randomly. It was a census of all statistics students in the fall 2015 semester. Occasionally, data that is not collected randomly may still be representative. If we consider our population of interest as all statistics students from all semesters then this data may still be representative of the population of interest, even though it is not random.

- Both samples should have at least 10 successes and at least 10 failures.

  Yes. There were 15 students born in June and 41 in December. Both of these are greater than 10. There were 336 – 15 = 321 students NOT born in June and 336 – 41 = 295 NOT born in December. Both of these are greater than 10.

- Data values within each sample and between the samples should be independent of each other.

  No. This data likely fails the independence requirements. Many students came from the same statistics classes.

Checking Assumptions Example 2: earlier in this section, we used the randomly collected health data to compare the population mean weight of women to the population mean weight of men. Were these groups matched pair? Would this data pass the traditional formula assumptions for comparing the means? The sample data is given below.

Original Sample

$\bar{x}_1 - \bar{x}_2 = -26.33$, $n_1 = 40$, $n_2 = 40$
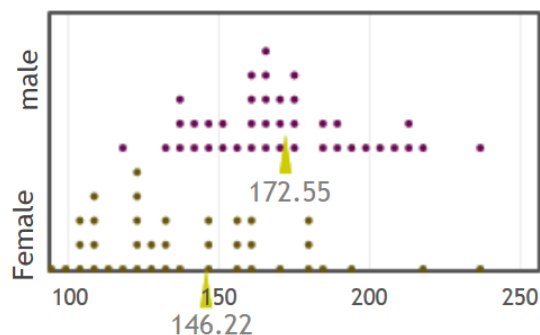
A random sample of men and women are not matched pair. They were not husband and wife or brother and sister. Therefore, we will proceed with checking the assumptions for independent groups.

Traditional Formula Assumptions for Comparing Two-Population Means from Independent Groups

- Both samples should be random and quantitative.

  Yes. Weights are quantitative and these were both random samples.

- Each sample should be Either Nearly Normal (almost bell shaped) OR have a Sample size of at least 30.

  Women: Yes. The dot plot for the women's weight data shows that it is skewed right. So our sample size must be over 30 for it to pass. The sample size is 40, which is greater than the 30 requirement. Even though the shape was not normal, it still passes the at least 30 or normal requirement.

  Men: Yes. The dot plot for the men's weight data shows that it is nearly normal. The sample size is 40, which is greater than the 30 requirement.

- Data values within the samples and between the samples should be independent of each other.

  Yes. A small random sample of 40 men and 40 women taken from millions of men and women in the population, are not likely to be related or know each other.