MAT111 REVIEW

Ch. 9

STATISTICS VS PARAMETERS

Statistics

- Sample Mean: $\bar{x}$
- Sample Standard Deviation: s
- Sample Proportion: $\hat{p}$

- Based on observed data

Parameters

- Population Mean: μ
- Population Standard Deviation: σ
- Population Proportion: p

- Typically unknown

SAMPLING DISTRIBUTIONS

"Distribution of all possible sample means that would result from drawing repeated random samples of a certain size from the population."

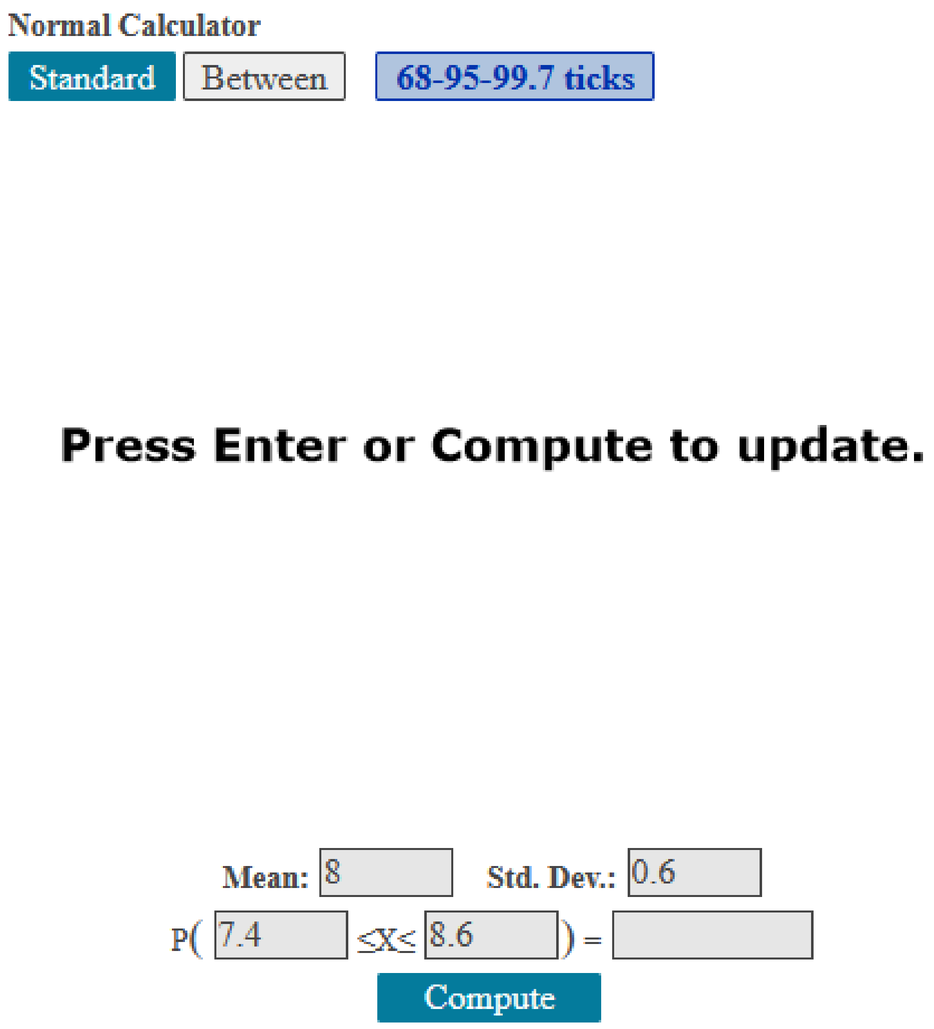
- The mean of the sampling distribution of the sample means is also the population mean.
- The Standard Deviation of the sampling distribution is the Standard Error.
- Standard Error is $\frac{\sigma}{\sqrt{n}}$ (which tells us the standard error is smaller for larger samples)
- Note: As sample size is increased, the spread of the sample means decreases.

EXAMPLE: BABY WEIGHT

Some sources report that the weights of full-term newborn babies in a certain town have a mean of 8 pounds and a standard deviation of 0.6 pounds and are Normally distributed.

- What is the probability that one newborn baby will have a weight within 0.6 pounds of the mean – that is, between 7.4 and 8.6 pounds, or within one standard deviation of the mean?

StatCrunch – Stat – Calculators – Normal
Fill in Mean, Standard Deviation, and weights



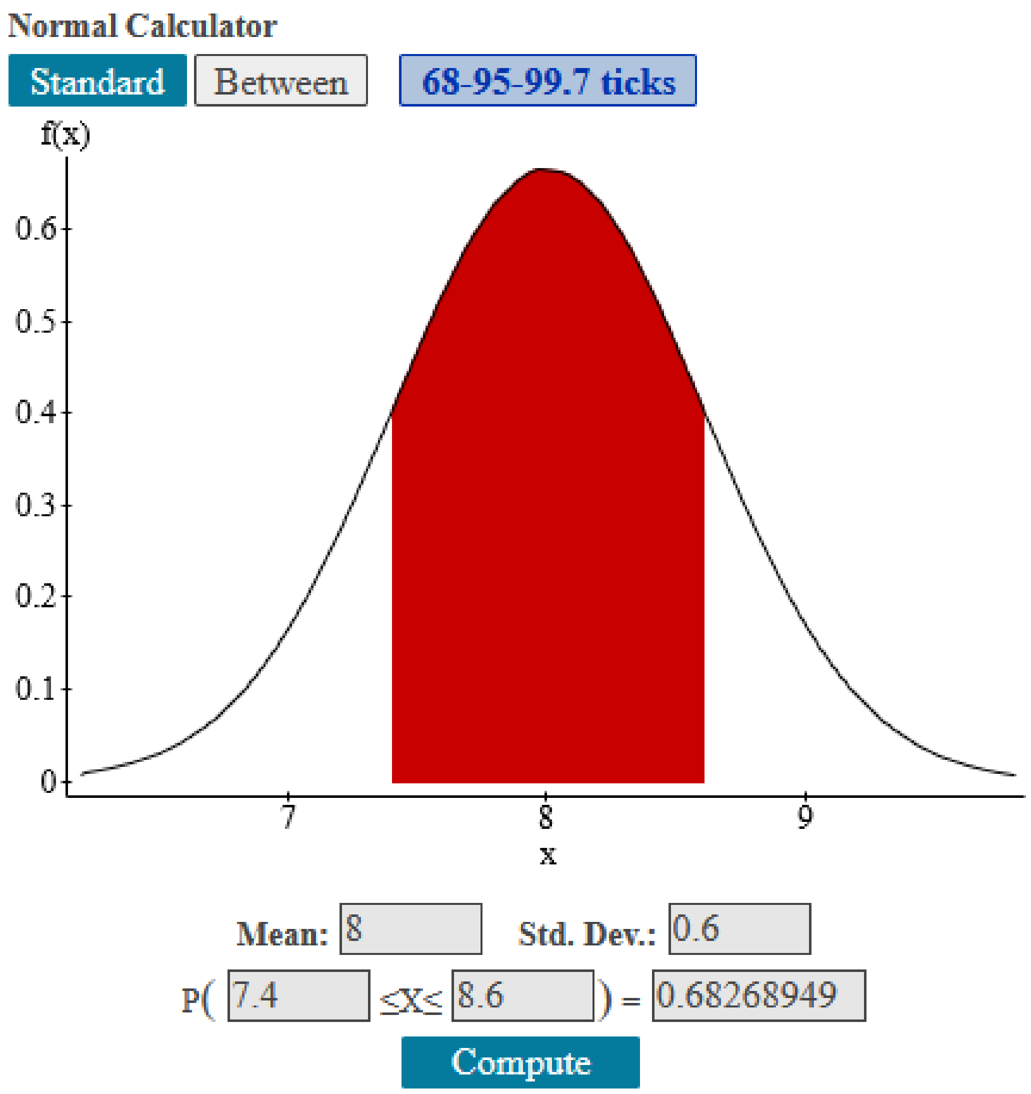
EXAMPLE: BABY WEIGHT

Some sources report that the weights of full-term newborn babies in a certain town have a mean of 8 pounds and a standard deviation of 0.6 pounds and are Normally distributed.

- What is the probability that one newborn baby will have a weight within 0.6 pounds of the mean – that is, between 7.4 and 8.6 pounds, or within one standard deviation of the mean?

StatCrunch – Stat – Calculators – Normal
Fill in Mean, Standard Deviation, and weights

The probability is 0.6827



EXAMPLE: BABY WEIGHT

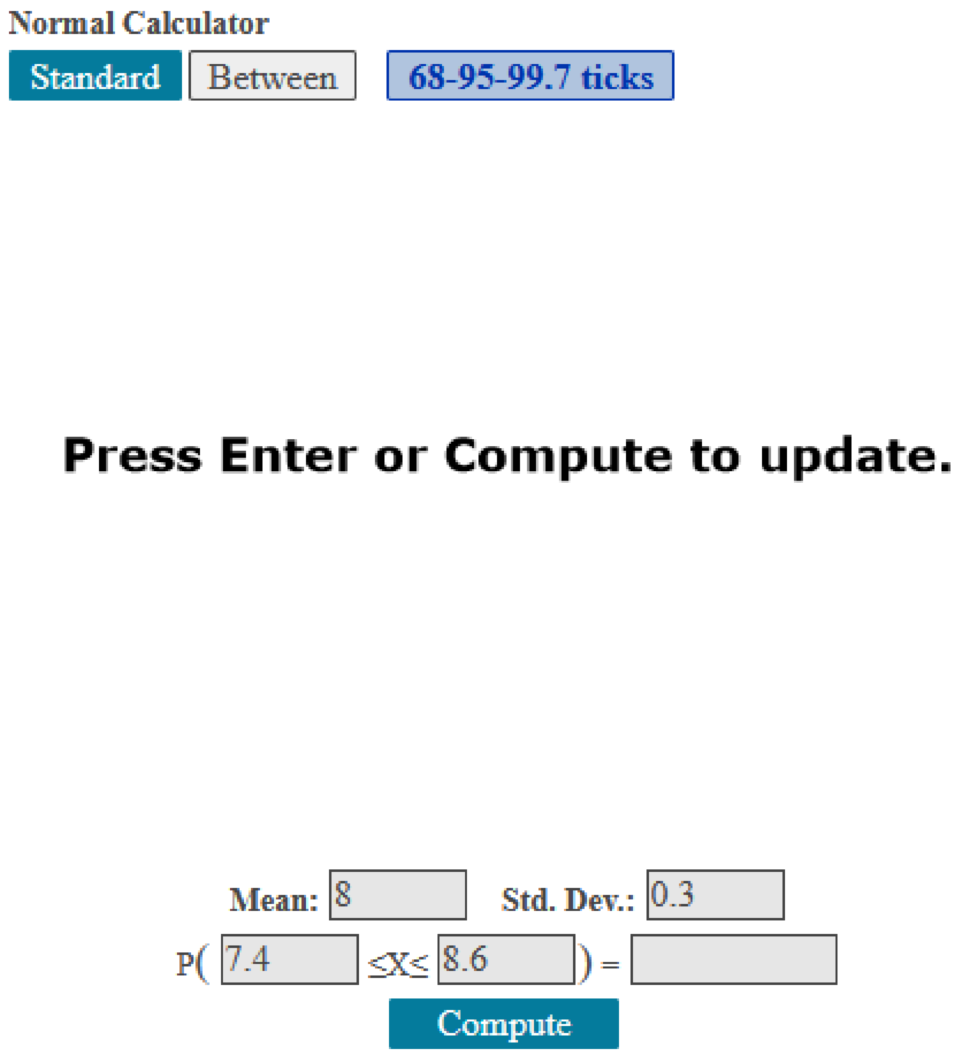
Some sources report that the weights of full-term newborn babies in a certain town have a mean of 8 pounds and a standard deviation of 0.6 pounds and are Normally distributed.

- What is the probability that the average of four babies' weights will be within 0.6 pounds of the mean; will be between 7.4 and 8.6 pounds?

StatCrunch – Stat – Calculators – Normal

Fill in Mean, Standard Error, and weights

$$Standard\ Error = \frac{0.6}{\sqrt{4}} = 0.3$$



EXAMPLE: BABY WEIGHT

Some sources report that the weights of full-term newborn babies in a certain town have a mean of 8 pounds and a standard deviation of 0.6 pounds and are Normally distributed.

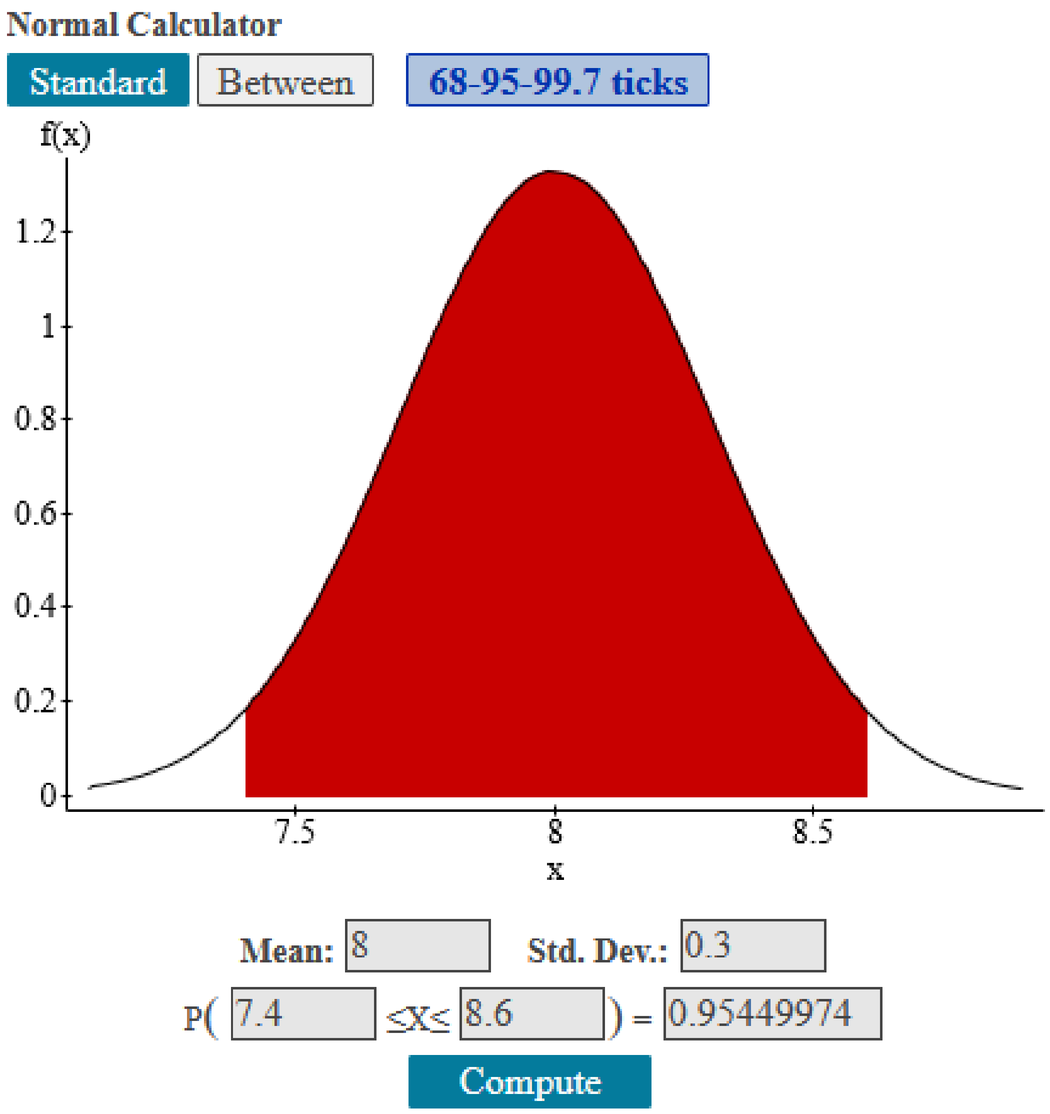
- What is the probability that the average of four babies' weights will be within 0.6 pounds of the mean; will be between 7.4 and 8.6 pounds?

StatCrunch – Stat – Calculators – Normal

Fill in Mean, Standard Error, and weights

$$Standard\ Error = \frac{0.6}{\sqrt{4}} = 0.3$$

The probability is 0.9545



EXAMPLE: BABY WEIGHT

Some sources report that the weights of full-term newborn babies in a certain town have a mean of 8 pounds and a standard deviation of 0.6 pounds and are Normally distributed.

- The distribution of means is **taller** and **narrower** than the original distribution. Therefore, the distribution of means will have **more** observations located closer to the center of the distribution.

CENTRAL LIMIT THEOREM

CLT assures us that no matter what the shape of the population distribution, if a sample is selected so that the following conditions are met, then the distribution of sample means follows an approximately Normal distribution

- Condition 1: Random Sample and Independence
  - Each observation is collected randomly from the population and observations are independent of each other.
- Condition 2: Large sample
  - The population distribution is normal or the sample size is large. ($n \geq 25$)
- Condition 3: Big Population
  - The population much be at least 10 times larger than the sample size.

EXAMPLE: CLT

The average income in a certain region in 2013 was $61,000 per person per year. Suppose the standard deviation is $26,000 and the distribution is right-skewed. Suppose we take a random sample of 100 residents of the region.

- Is the sample size large enough to use the Central Limit Theorem for means? Explain.

Yes! It is large enough because the sample size of 100 is greater than 25.

EXAMPLE: CLT

The average income in a certain region in 2013 was $61,000 per person per year. Suppose the standard deviation is $26,000 and the distribution is right-skewed. Suppose we take a random sample of 100 residents of the region.

- What are the mean and standard error of the sampling distribution?

The mean of the sampling distribution of the sample mean is also the population mean, so **the mean is $61,000.**

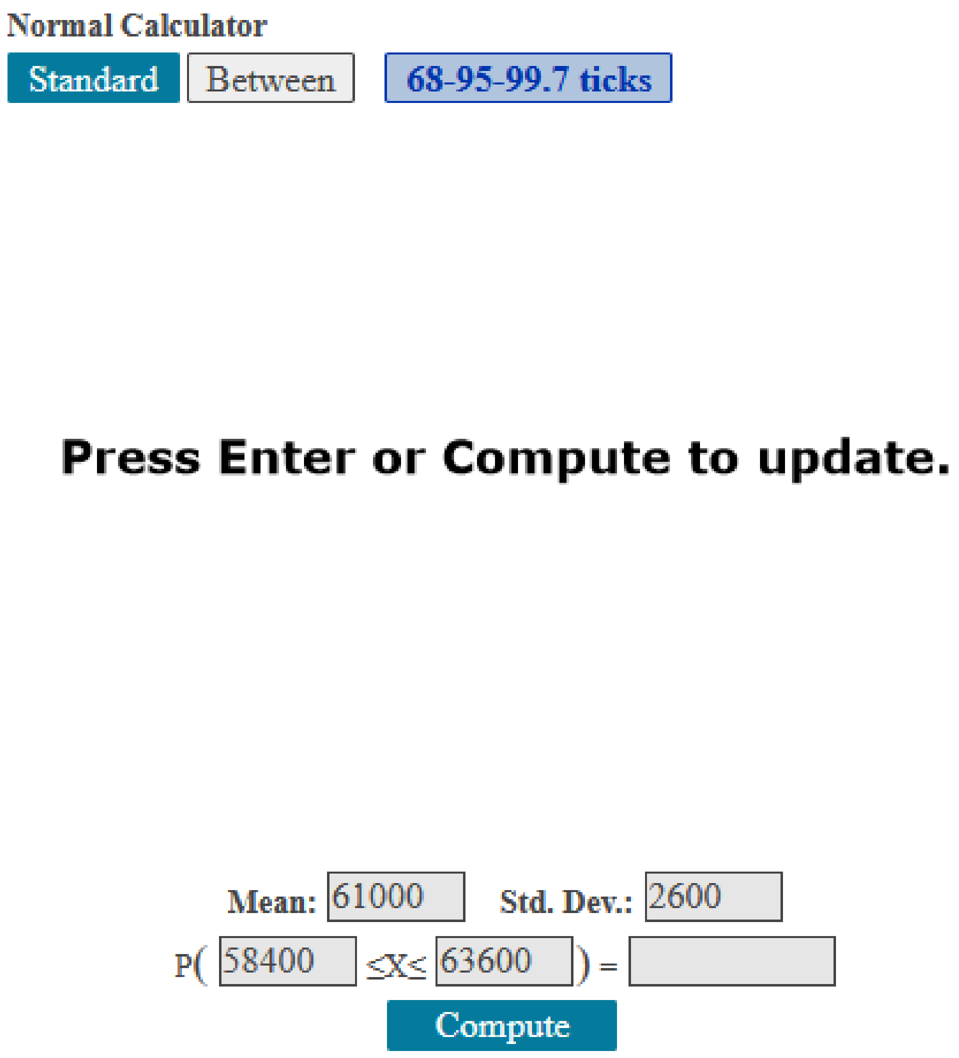
Remember that Standard Error is $\frac{\sigma}{\sqrt{n}} = \frac{26000}{\sqrt{100}} = 2600$, so **the standard error is $2,600**.

EXAMPLE: CLT

The average income in a certain region in 2013 was $61,000 per person per year. Suppose the standard deviation is $26,000 and the distribution is right-skewed. Suppose we take a random sample of 100 residents of the region.

- What is the probability that the sample mean will be more than $2,600 away from the population mean?

First, find the probability that the sample mean will be $2,600 or less away from the mean – that is, between $58,400 and $63,600.



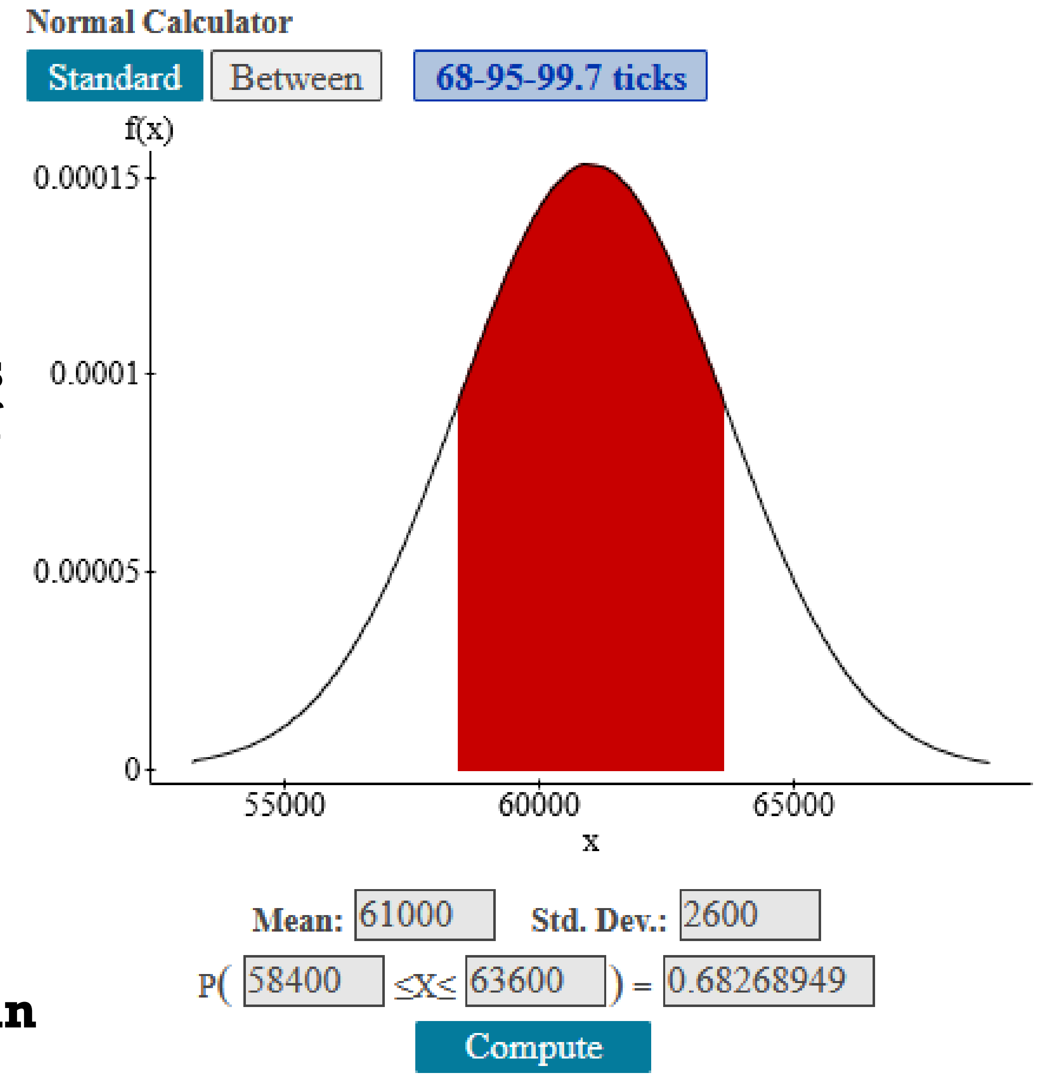
EXAMPLE: CLT

The average income in a certain region in 2013 was $61,000 per person per year. Suppose the standard deviation is $26,000 and the distribution is right-skewed. Suppose we take a random sample of 100 residents of the region.

- What is the probability that the sample mean will be more than $2,600 away from the population mean?

First, find the probability that the sample mean will be $2,600 or less away from the mean – that is, between $58,400 and $63,600.

The probability that the sample mean is less than $2,600 from the mean is 0.68268949.



EXAMPLE: CLT

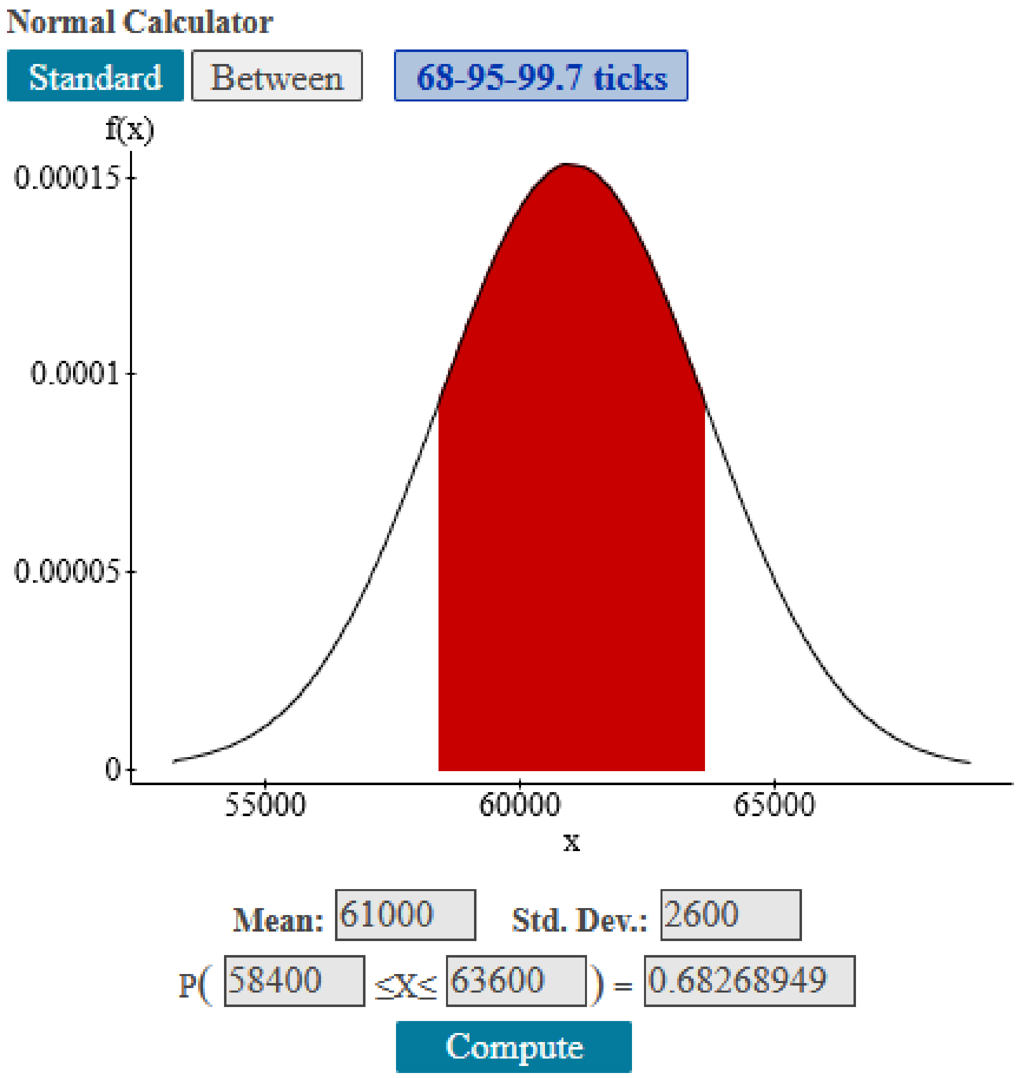
The average income in a certain region in 2013 was $61,000 per person per year. Suppose the standard deviation is $26,000 and the distribution is right-skewed. Suppose we take a random sample of 100 residents of the region.

- What is the probability that the sample mean will be more than $2,600 away from the population mean?

Then use the compliment to find the probability outside this range.

1 - 0.68268949=0.31731051

The probability that the sample mean is more than $2,600 away from the mean is 0.3173



CONFIDENCE INTERVALS

- Confidence Intervals are used to communicate an estimate of the mean and a measure of uncertainty.
- Confidence level (Percentage) tells us how confident we are that the range contains the population mean.

StatCrunch – Stat – T Stats – One Sample

EXAMPLE: INTERPRET CONFIDENCE INTERVALS

- A random sample of 30 colleges was taken. The mean debt after graduation was $18,212 with a margin of error of $1473. The distribution of debt is Normal.

- Choose the correct interpretation of the confidence interval below and fill in the blanks.
  - We are 95% confident that the sample mean is between $_____ and $_____.
  - We are 95% confident that the population mean is between $_____ and $_____.
  - We are 95% confident that the boundaries for the interval are $_____ and $_____.

EXAMPLE: INTERPRET CONFIDENCE INTERVALS

- A random sample of 30 colleges was taken. The mean debt after graduation was $18,212 with a margin of error of $1473. The distribution of debt is Normal.

- Choose the correct interpretation of the confidence interval below and fill in the blanks.
  - We are 95% confident that the sample mean is between $_____ and $_____.
  - **We are 95% confident that the population mean is between <u>$16,739</u> and <u>$19,685</u>.**
  - We are 95% confident that the boundaries for the interval are $_____ and $_____.

EXAMPLE: CONFIDENCE INTERVALS

In finding a confidence interval for a random sample of 35 students' GPAs, one interval was (2.60,3.20) and the other was (2.65, 3.15).

- a. One of them is a 95% interval and one is a 90% interval. Which is which, and how do you know?

  - **The interval (2.60, 3.25) is the 95% interval and (2.65, 3.15) is the 90% interval - a higher level of confidence results in a wider confidence interval.**

EXAMPLE: CONFIDENCE INTERVALS

In finding a confidence interval for a random sample of 35 students' GPAs, one interval was (2.60,3.20) and the other was (2.65, 3.15).

- b. If we used a larger sample size n=140 instead of n=35, would the 95% interval be wider or narrower than the one reported here?

  - **The 95% interval with n=140 will be narrower than the interval with n=35 because a larger sample size provides a smaller standard error, and this means a smaller margin of error at the same level of confidence.**

HYPOTHESIS TESTING

- Step 1: Hypothesize
  - State your hypothesis about the population parameter
- Step 2: Prepare
  - Choose a significance level and test statistic, check conditions and assumptions
- Step 3: Compare and Compute
  - Compute a test statistic and p-value
- Step 4: Interpret
  - Do you reject the null hypothesis or not? What does this mean?

StatCrunch – Stat – T Stats – One Sample

Important Formulas:

$$t = \frac{\bar{x} - \mu_0}{SE_{est}}$$

$$SE_{est} = \frac{s}{\sqrt{n}}$$

$$df = n - 1$$

REJECTING THE NULL

- If $p < \alpha$ then we reject the null hypothesis

EXAMPLE: HYPOTHESIS TESTING

The mean weight of all 20-year-old women in a certain region is 133 pounds. A random sample of 40 vegetarian women in the region who are 20 years old showed a sample mean of 129 pounds with a standard deviation of 14 pounds. The women's measurements were independent of each other.

- Determine whether the mean weight for 20-year old vegetarian women in the region is significantly less than 133, using a significance level of 0.05.
  - Step 1: State the null and alternate hypothesis.

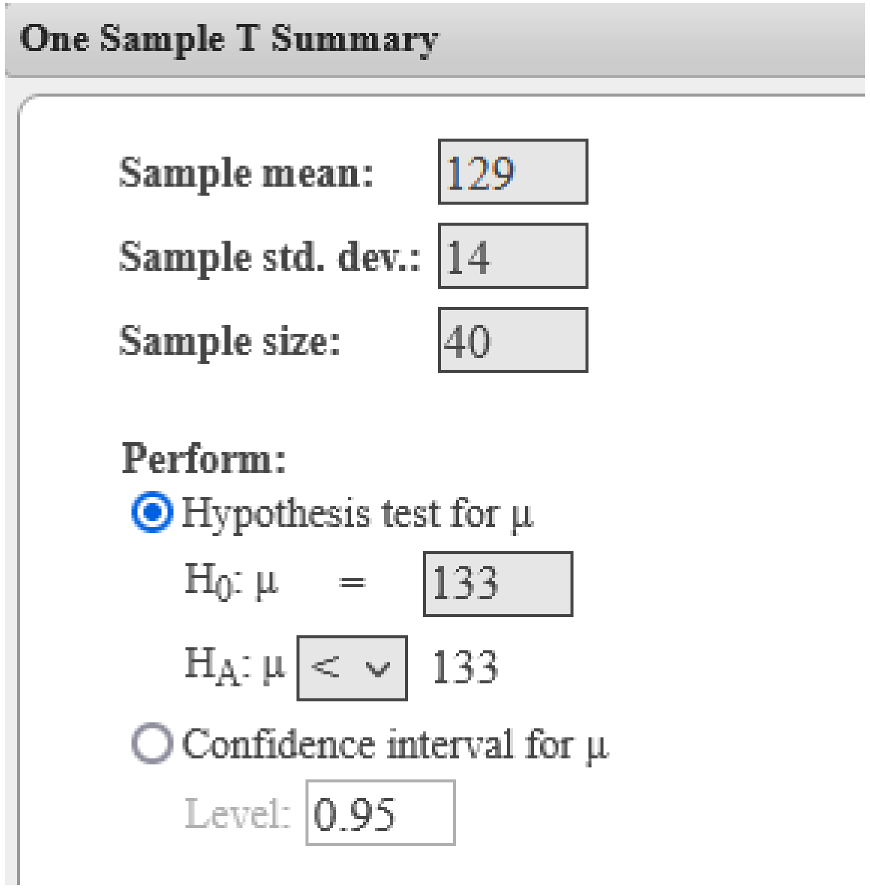
$$\boldsymbol{H_0: \mu = 133}$$
$$\boldsymbol{H_a: \mu < 133}$$

EXAMPLE: HYPOTHESIS TESTING

The mean weight of all 20-year-old women in a certain region is 133 pounds. A random sample of 40 vegetarian women in the region who are 20 years old showed a sample mean of 129 pounds with a standard deviation of 14 pounds. The women's measurements were independent of each other.

- Determine whether the mean weight for 20-year old vegetarian women in the region is significantly less than 133, using a significance level of 0.05.
  - Step 2: Find the test statistic and the p-value.

StatCrunch – Stat – T Stats – One Sample – With Summary



EXAMPLE: HYPOTHESIS TESTING

The mean weight of all 20-year-old women in a certain region is 133 pounds. A random sample of 40 vegetarian women in the region who are 20 years old showed a sample mean of 129 pounds with a standard deviation of 14 pounds. The women's measurements were independent of each other.

- Determine whether the mean weight for 20-year old vegetarian women in the region is significantly less than 133, using a significance level of 0.05.
  - Step 2: Find the test statistic and the p-value.

StatCrunch – Stat – T Stats – One Sample – With Summary

$$t = -1.807\,, \qquad p = 0.039$$

One sample T summary hypothesis test:
μ : Mean of population
$H_0 : \mu = 133$
$H_A : \mu < 133$

Hypothesis test results:

Mean	Sample Mean	Std. Err.	DF	T-Stat	P-value
μ	129	2.2135944	39	-1.8070158	0.0392

EXAMPLE: HYPOTHESIS TESTING

The mean weight of all 20-year-old women in a certain region is 133 pounds. A random sample of 40 vegetarian women in the region who are 20 years old showed a sample mean of 129 pounds with a standard deviation of 14 pounds. The women's measurements were independent of each other.

- Determine whether the mean weight for 20-year old vegetarian women in the region is significantly less than 133, using a significance level of 0.05.
  - Step 3: Reject or do not reject the null, H_0.

$0.039 < 0.05$ so we reject H_0.
There is reason to believe that the population mean is less than 133 pounds at a significance level of 0.05.

EXAMPLE: HYPOTHESIS TESTING

Thirty GPAs from a randomly selected sample of statistics students at a college are linked below. Assume that the population distribution is approximately Normal. The technician in charge of records claimed that the population mean GPA for the whole college is 2.82

- What is the sample mean? Is it higher or lower than the population mean of 2.82?

Student GPA Table

2.893.343.172.553.442.78
3.083.582.643.972.872.64
3.583.072.713.422.493.14
3.483.153.133.063.132.96
3.493.522.693.123.213.02

EXAMPLE: HYPOTHESIS TESTING

Thirty GPAs from a randomly selected sample of statistics students at a college are linked below. Assume that the population distribution is approximately Normal. The technician in charge of records claimed that the population mean GPA for the whole college is 2.82

- What is the sample mean? Is it higher or lower than the population mean of 2.82?

<u>Student GPA Table</u>

2.893.343.172.553.442.78
3.083.582.643.972.872.64
3.583.072.713.422.493.14
3.483.153.133.063.132.96
3.493.522.693.123.213.02

StatCrunch – Stat – Summary Stats – Columns - Mean

The sample mean is 3.11

EXAMPLE: HYPOTHESIS TESTING

Thirty GPAs from a randomly selected sample of statistics students at a college are linked below. Assume that the population distribution is approximately Normal. The technician in charge of records claimed that the population mean GPA for the whole college is 2.82

- The chair of the mathematics department claims that statistics students typically have higher GPAs than the typical college student. Use the four-step procedure and the data provided to test this claim. Use a significance level of 0.05.

Student GPA Table

2.893.343.172.553.442.78
3.083.582.643.972.872.64
3.583.072.713.422.493.14
3.483.153.133.063.132.96
3.493.522.693.123.213.02

EXAMPLE: HYPOTHESIS TESTING

Thirty GPAs from a randomly selected sample of statistics students at a college are linked below. Assume that the population distribution is approximately Normal. The technician in charge of records claimed that the population mean GPA for the whole college is 2.82

- The chair of the mathematics department claims that statistics students typically have higher GPAs than the typical college student. Use the four-step procedure and the data provided to test this claim. Use a significance level of 0.05.
  - Step 1: State the null and alternate hypothesis.

$$H_0: \mu = 2.82$$
$$H_a: \mu > 2.82$$

Student GPA Table

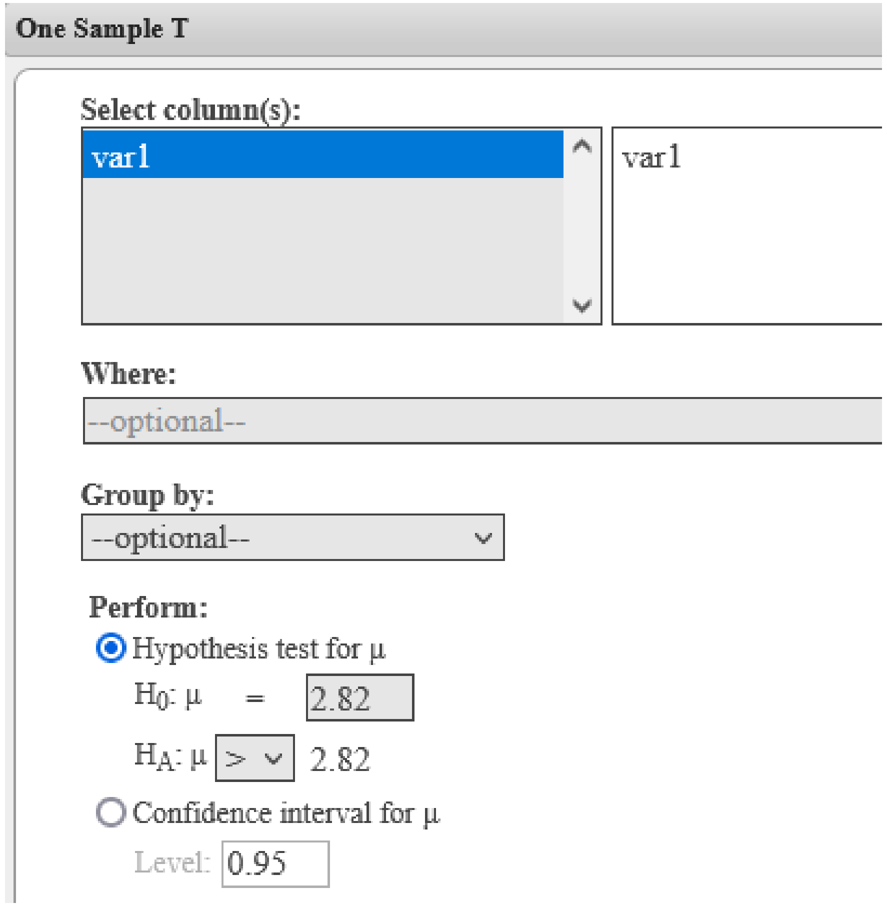
2.89	3.34	3.17	2.55	3.44	2.78
3.08	3.58	2.64	3.97	2.87	2.64
3.58	3.07	2.71	3.42	2.49	3.14
3.48	3.15	3.13	3.06	3.13	2.96
3.49	3.52	2.69	3.12	3.21	3.02

EXAMPLE: HYPOTHESIS TESTING

Thirty GPAs from a randomly selected sample of statistics students at a college are linked below. Assume that the population distribution is approximately Normal. The technician in charge of records claimed that the population mean GPA for the whole college is 2.82

- The chair of the mathematics department claims that statistics students typically have higher GPAs than the typical college student. Use the four-step procedure and the data provided to test this claim. Use a significance level of 0.05.
  - Step 2: Find the test statistic and the p-value.

StatCrunch – Stat – T Stats – One Sample – With Data



EXAMPLE: HYPOTHESIS TESTING

Thirty GPAs from a randomly selected sample of statistics students at a college are linked below. Assume that the population distribution is approximately Normal. The technician in charge of records claimed that the population mean GPA for the whole college is 2.82

- The chair of the mathematics department claims that statistics students typically have higher GPAs than the typical college student. Use the four-step procedure and the data provided to test this claim. Use a significance level of 0.05.
  - Step 2: Find the test statistic and the p-value.

StatCrunch – Stat – T Stats – One Sample – With Data

$$t = 4.498, \qquad p = 0.000$$

Student GPA Table

2.893.343.172.553.442.78
3.083.582.643.972.872.64
3.583.072.713.422.493.14
3.483.153.133.063.132.96
3.493.522.693.123.213.02

One sample T hypothesis test:

μ : Mean of variable

$H_0 : \mu = 2.82$

$H_A : \mu > 2.82$

Hypothesis test results:

Variable	Sample Mean	Std. Err.	DF	T-Stat	P-value
var1	3.1106667	0.064615455	29	4.4984078	<0.0001

EXAMPLE: HYPOTHESIS TESTING

Thirty GPAs from a randomly selected sample of statistics students at a college are linked below. Assume that the population distribution is approximately Normal. The technician in charge of records claimed that the population mean GPA for the whole college is 2.82

- The chair of the mathematics department claims that statistics students typically have higher GPAs than the typical college student. Use the four-step procedure and the data provided to test this claim. Use a significance level of 0.05.
  - Step 3: Reject or do not reject the null, H_0.

$0.000 < 0.05$ so we reject H_0.
The mean GPA for statistics students is significantly higher than 2.82.

Student GPA Table

2.89	3.34	3.17	2.55	3.44	2.78
3.08	3.58	2.64	3.97	2.87	2.64
3.58	3.07	2.71	3.42	2.49	3.14
3.48	3.15	3.13	3.06	3.13	2.96
3.49	3.52	2.69	3.12	3.21	3.02

COMPARING POPULATION MEANS

Independent

- Two different groups – same variable

StatCrunch – Stat - T Stats – Two Sample

Paired (Dependent)

- Same group – two different variables
  - Before and After comparisons
- Matched groups
  - Related objects/people

StatCrunch – Stat - T Stats – Paired

EXAMPLE: HYPOTHESIS TESTING 2 SAMPLES PAIRED

A statistics student collected data on the prices of the same items at a military commissary and a nearby corporate store. The items were matched for content, manufacturer, and size and were priced separately.

- Assume that they are a random sample of items, and use a significance level of 0.05 to test the hypothesis that the military commissary has a lower mean price. Assume that the population distribution of differences is approximately Normal.

Item	Corporate Stor	Navy Commiss
1	19.95	18.93
2	3.94	2.77
3	5.91	4.76
4	8.07	6.13
5	3.12	2.64
6	4.66	3.92
7	5.17	3.54
8	4.08	3.15
9	4.02	3.19
10	3.75	2.95
11	1.44	1.08
12	6.55	2.72
13	2.61	1.67
14	6.09	5.11
15	2.72	5.46
16	4.55	3.98
17	7.48	7.46
18	1.19	1.17
19	0.92	0.79
20	2.52	2.42
21	0.75	0.59
22	1.69	1.73
23	0.77	0.54
24	3.34	2.44
25	1.76	2.37
26	1.04	0.82
27	2.46	2.21
28	1.46	1.49
29	2.58	1.29
30	1.27	1.13
31	0.68	0.61
32	1.56	1.36

EXAMPLE: HYPOTHESIS TESTING 2 SAMPLES PAIRED

A statistics student collected data on the prices of the same items at a military commissary and a nearby corporate store. The items were matched for content, manufacturer, and size and were priced separately.

- Assume that they are a random sample of items, and use a significance level of 0.05 to test the hypothesis that the military commissary has a lower mean price. Assume that the population distribution of differences is approximately Normal.
  - Step 1: State the null and alternate hypothesis.

$$H_0: \mu_{difference} = 0$$
$$H_a: \mu_{difference} > 0$$

Item	Corporate Store	Navy Commiss
1	19.95	18.93
2	3.94	2.77
3	5.91	4.76
4	8.07	6.13
5	3.12	2.64
6	4.66	3.92
7	5.17	3.54
8	4.08	3.15
9	4.02	3.19
10	3.75	2.95
11	1.44	1.08
12	6.55	2.72
13	2.61	1.67
14	6.09	5.11
15	2.72	5.46
16	4.55	3.98
17	7.48	7.46
18	1.19	1.17
19	0.92	0.79
20	2.52	2.42
21	0.75	0.59
22	1.69	1.73
23	0.77	0.54
24	3.34	2.44
25	1.76	2.37
26	1.04	0.82
27	2.46	2.21
28	1.46	1.49
29	2.58	1.29
30	1.27	1.13
31	0.68	0.61
32	1.56	1.36

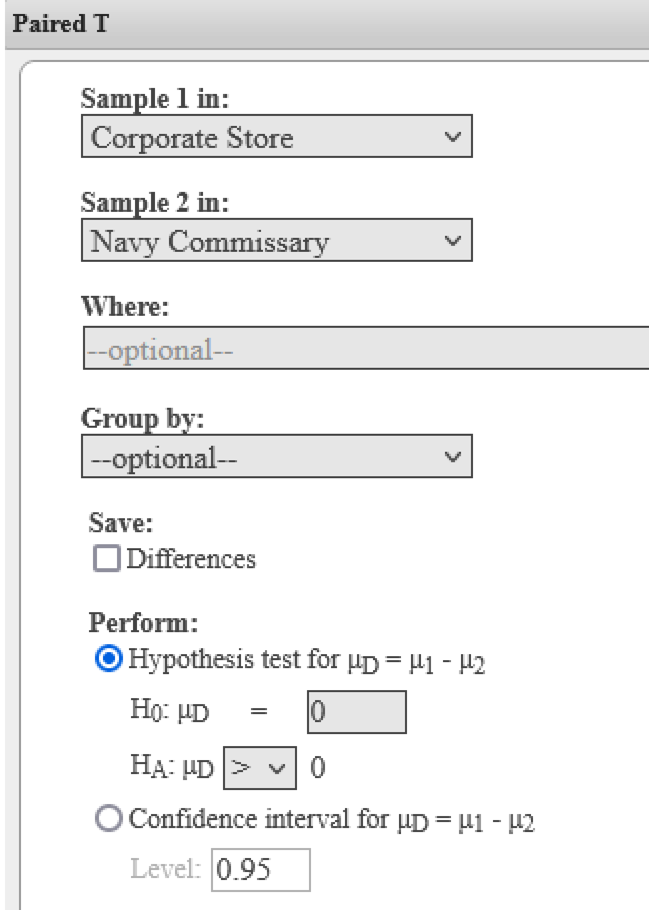
EXAMPLE: HYPOTHESIS TESTING 2 SAMPLES PAIRED

A statistics student collected data on the prices of the same items at a military commissary and a nearby corporate store. The items were matched for content, manufacturer, and size and were priced separately.

- Assume that they are a random sample of items, and use a significance level of 0.05 to test the hypothesis that the military commissary has a lower mean price. Assume that the population distribution of differences is approximately Normal.
  - Step 2: Find the test statistic and the p-value.

StatCrunch – Stat – T Stats – Paired

Item	Corporate Stor	Navy Commiss
1	19.95	18.93
2	3.94	2.77
3	5.91	4.76
4	8.07	6.13
5	3.12	2.64
6	4.66	3.92
7	5.17	3.54
8	4.08	3.15
9	4.02	3.19
10	3.75	2.95



EXAMPLE: HYPOTHESIS TESTING 2 SAMPLES PAIRED

Item	Corporate Store	Navy Commiss
1	19.95	18.93
2	3.94	2.77
3	5.91	4.76
4	8.07	6.13
5	3.12	2.64
6	4.66	3.92
7	5.17	3.54
8	4.08	3.15
9	4.02	3.19
10	3.75	2.95
11	1.44	1.08
12	6.55	2.72
13	2.61	1.67
14	6.09	5.11
15	2.72	5.46
16	4.55	3.98
17	7.48	7.46
18	1.19	1.17
19	0.92	0.79
20	2.52	2.42
21	0.75	0.59
22	1.69	1.73
23	0.77	0.54

A statistics student collected data on the prices of the same items at a military commissary and a nearby corporate store. The items were matched for content, manufacturer, and size and were priced separately.

- Assume that they are a random sample of items, and use a significance level of 0.05 to test the hypothesis that the military commissary has a lower mean price. Assume that the population distribution of differences is approximately Normal.
  - Step 2: Find the test statistic and the p-value.

StatCrunch – Stat – T Stats – Paired

$$t = 3.127, \qquad p = 0.002$$

Paired T hypothesis test:

$\mu_D = \mu_1 - \mu_2$: Mean of the difference between Corporate Store and Navy Commissary

$H_0 : \mu_D = 0$

$H_A : \mu_D > 0$

Hypothesis test results:

Difference	Mean	Std. Err.	DF	T-Stat	P-value
Corporate Store - Navy Commissary	0.5525	0.17666603	31	3.1273697	0.0019

EXAMPLE: HYPOTHESIS TESTING 2 SAMPLES PAIRED

A statistics student collected data on the prices of the same items at a military commissary and a nearby corporate store. The items were matched for content, manufacturer, and size and were priced separately.

- Assume that they are a random sample of items, and use a significance level of 0.05 to test the hypothesis that the military commissary has a lower mean price. Assume that the population distribution of differences is approximately Normal.
  - Step 3: Reject or do not reject the null, H_0.

$0.002 < 0.05$ so we reject H_0.
There is sufficient evidence to conclude that the military commissary has a lower means price.

Item	Corporate Stor	Navy Commiss
1	19.95	18.93
2	3.94	2.77
3	5.91	4.76
4	8.07	6.13
5	3.12	2.64
6	4.66	3.92
7	5.17	3.54
8	4.08	3.15
9	4.02	3.19
10	3.75	2.95
11	1.44	1.08
12	6.55	2.72
13	2.61	1.67
14	6.09	5.11
15	2.72	5.46
16	4.55	3.98
17	7.48	7.46
18	1.19	1.17
19	0.92	0.79
20	2.52	2.42
21	0.75	0.59
22	1.69	1.73
23	0.77	0.54
24	3.34	2.44
25	1.76	2.37
26	1.04	0.82
27	2.46	2.21
28	1.46	1.49
29	2.58	1.29
30	1.27	1.13
31	0.68	0.61
32	1.56	1.36

EXAMPLE: HYPOTHESIS TESTING 2 SAMPLES INDEPENDENT

A random sample of male college baseball players and a random sample of male college soccer players were obtained independently and weighed. The accompanying table shows the unstacked weights (in pounds). The distributions of both data sets suggest that the population distributions are roughly Normal.

- Determine whether the difference in means is significant, using a significance level of 0.05.

Baseball	Soccer
188	163
202	188
185	185
184	188
196	182
209	192
189	171
176	180
208	171
226	182
233	166
198	192
171	187
190	157
212	171
201	171
179	159
181	152
195	173
199	180
193	185
184	174
197	193
188	155
185	161

EXAMPLE: HYPOTHESIS TESTING 2 SAMPLES INDEPENDENT

A random sample of male college baseball players and a random sample of male college soccer players were obtained independently and weighed. The accompanying table shows the unstacked weights (in pounds). The distributions of both data sets suggest that the population distributions are roughly Normal.

- Determine whether the difference in means is significant, using a significance level of 0.05.
  - Step 1: State the null and alternate hypothesis.

$$H_0: \mu_{Baseball} - \mu_{Soccer} = 0$$
$$H_a: \mu_{Baseball} - \mu_{Soccer} \neq 0$$

Baseball	Soccer
188	163
202	188
185	185
184	188
196	182
209	192
189	171
176	180
208	171
226	182
233	166
198	192
171	187
190	157
212	171
201	171
179	159
181	152
195	173
199	180
193	185
184	174
197	193
188	155
185	161

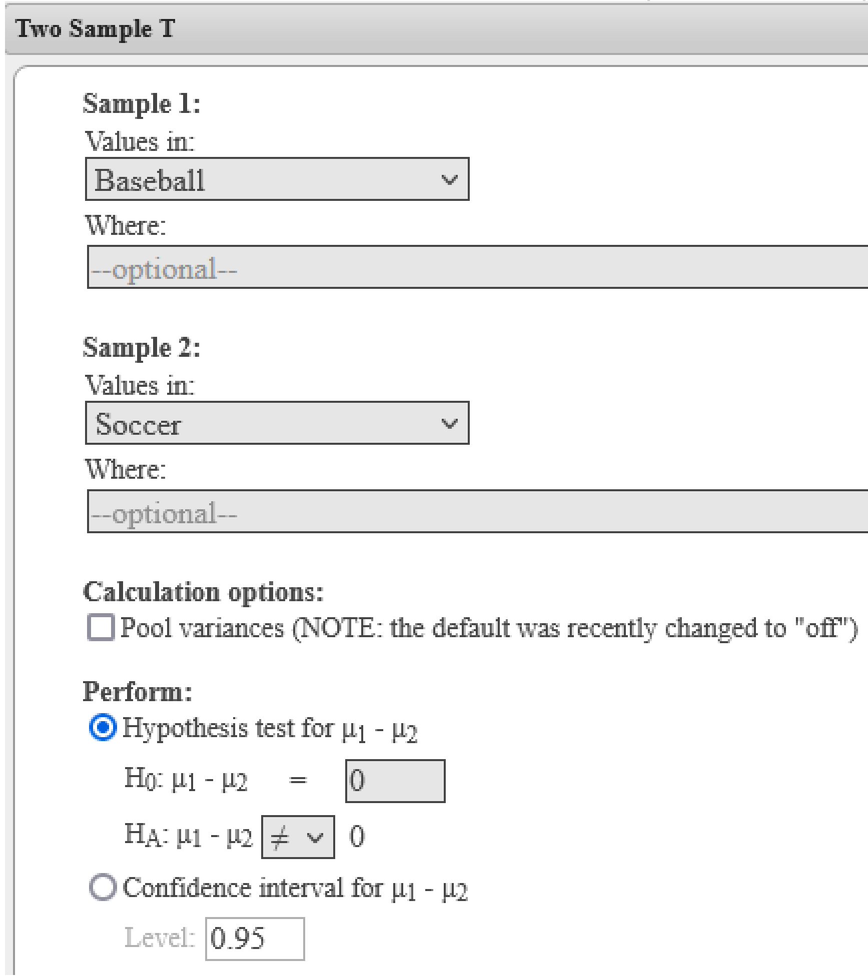
EXAMPLE: HYPOTHESIS TESTING 2 SAMPLES INDEPENDENT

Baseball	Soccer
188	163
202	188
185	185
184	188
196	182

A random sample of male college baseball players and a random sample of male college soccer players were obtained independently and weighed. The accompanying table shows the unstacked weights (in pounds). The distributions of both data sets suggest that the population distributions are roughly Normal.

- Determine whether the difference in means is significant, using a significance level of 0.05.
  - Step 2: Find the test statistic and the p-value.

StatCrunch – Stat – T Stats – Two Sample – With Data



EXAMPLE: HYPOTHESIS TESTING 2 SAMPLES INDEPENDENT

Baseball	Soccer
188	163
202	188
185	185
184	188
196	182
209	192
189	171
176	180
208	171
226	182
233	166
198	192
171	187
190	157
212	171
201	171

A random sample of male college baseball players and a random sample of male college soccer players were obtained independently and weighed. The accompanying table shows the unstacked weights (in pounds). The distributions of both data sets suggest that the population distributions are roughly Normal.

- Determine whether the difference in means is significant, using a significance level of 0.05.
  - Step 2: Find the test statistic and the p-value.

StatCrunch – Stat – T Stats – Two Sample – With Data

Two sample T hypothesis test:
μ_1 : Mean of Baseball
μ_2 : Mean of Soccer
$\mu_1 - \mu_2$: Difference between two means
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 \neq 0$
(without pooled variances)

Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	19.64	3.8508008	46.781493	5.1002379	<0.0001

$$t = 5.100\,, \qquad p = 0.000$$

EXAMPLE: HYPOTHESIS TESTING 2 SAMPLES INDEPENDENT

A random sample of male college baseball players and a random sample of male college soccer players were obtained independently and weighed. The accompanying table shows the unstacked weights (in pounds). The distributions of both data sets suggest that the population distributions are roughly Normal.

- Determine whether the difference in means is significant, using a significance level of 0.05.
  - Step 3: Reject or do not reject the null, H_0.

$0.000 < 0.05$ so we reject H_0.
There is sufficient evidence to conclude that the difference in mean weights is significant.

Baseball	Soccer
188	163
202	188
185	185
184	188
196	182
209	192
189	171
176	180
208	171
226	182
233	166
198	192
171	187
190	157
212	171
201	171
179	159
181	152
195	173
199	180
193	185
184	174
197	193
188	155
185	161

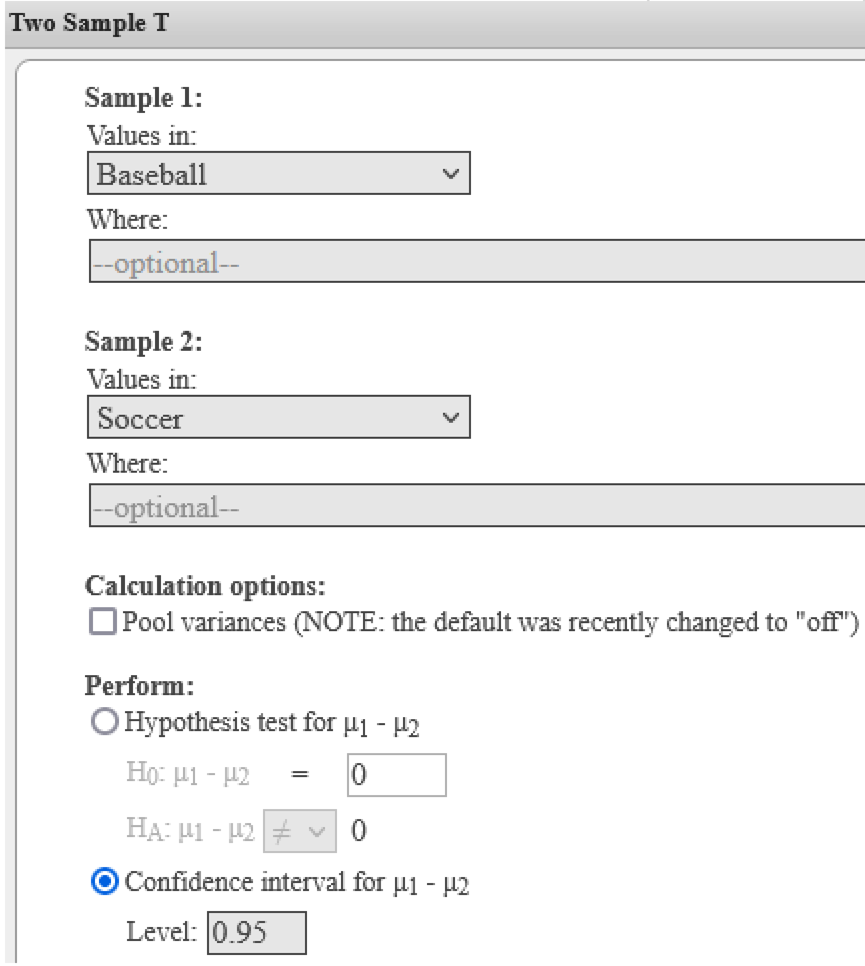
USING CONFIDENCE INTERVALS

Baseball	Soccer
194	171
204	197
195	190
188	191
201	189

A random sample of male college baseball players and a random sample of male college soccer players were obtained independently and weighed. The accompanying table shows the unstacked weights (in pounds). The distributions of both data sets suggest that the population distributions are roughly Normal.

- Find a 95% confidence interval for the difference between and explain what it shows.

StatCrunch – Stat – T Stats – Two Sample – With Data



USING CONFIDENCE INTERVALS

A random sample of male college baseball players and a random sample of male college soccer players were obtained independently and weighed. The accompanying table shows the unstacked weights (in pounds). The distributions of both data sets suggest that the population distributions are roughly Normal.

- Find a 95% confidence interval for the difference between and explain what it shows.

Baseball	Soccer
194	171
204	197
195	190
188	191
201	189
209	194
192	178
184	192
217	182
229	190
238	173
205	198
177	192
193	166
218	177
202	181

Two sample T confidence interval:

μ_1 : Mean of Baseball

μ_2 : Mean of Soccer

$\mu_1 - \mu_2$: Difference between two means

(without pooled variances)

95% confidence interval results:

Difference	Sample Diff.	Std. Err.	DF	L. Limit	U. Limit
$\mu_1 - \mu_2$	18.28	3.6061984	46.364091	11.022637	25.537363

The 95% interval for the difference in means is (11.02, 25.54).
Because the interval does not contain 0, it shows that the mean weights for soccer and baseball players are significantly different.

CENTER FOR ACADEMIC SUPPORT

Hearnes 213
816-271-4524