Math/Stat 244 Statistics II Course Packet 24807

Lead Instructor: Cara Lee

Contents

- Lecture Notes and Practice Pages
- Midterm Practice
- Final Practice

Blank page

Lesson 1: Review and Preview for Stat 244

Major Themes for Statistics I and II

- *Data are not neutral or objective. They reflect the lens through which they were gathered. All data are socially, culturally, and historically situated.*
- *There's variation in everything.*
- *All calculations in statistics are estimates/imperfect. (All models are wrong, but some are useful.)*
- *Two main methods for inferential statistics:*
    - *When simulating, the simulation must match the real-life situation for the results to be valid.*
    - *When using a theoretical model, the conditions must be met, or the results are not meaningful.*

Statistical Process

- Examine impacts and implicit bias. Who is harmed, who is served?
- Center nondominant voices and impacted communities. "Nothing about us, without us," -*South African disability rights movement*

1. Identify a question about a target population.
2. Design a study, check conditions, and collect data from a sample.
3. Analyze data and draw inferences.
4. Form a conclusion, review, and repeat.

Descriptive Statistics

Summarizing a set of data.

Inferential Statistics

- Sampling Distributions: Foundation for the Central Limit Theorem and inference, used to understand the distribution of random samples.
- Confidence Intervals (CI): Using a sample to estimate a parameter.
- Hypothesis Tests (HT): Comparing a sample with a null distribution to determine statistical significance.

Preview of Math/Stat 244 - Inference

- Categorical data: CI and HT for single proportions and the difference of two proportions.
- Categorical data in a contingency table: tests for goodness of fit, homogeneity and independence.
- Numerical data: CI and HT for single means, two independent means, and matched pairs.
- Numerical data: tests for more than two groups: ANOVA (analysis of variance).
- Two numerical variable data: Regression inference.

- Three Methods: simulations, theoretical (using formulas) and theoretical calculators
- Distributions: Normal, Chi-squared, T, F

Hypothesis Test Example with Simulation – Detecting Parkinson's

As reported by the Washington Post, Joy Milne of Perth, UK, smelled a "subtle musky odor" on her husband Les that she had never smelled before. At first, Joy thought maybe it was just from the sweat after long hours of work as an anesthesiologist. But when Les was diagnosed with Parkinson's Disease 6 years later, Joy suspected the odor might be a result of the disease.

Scientists were intrigued by Joy's claim and designed an experiment to test her ability to smell Parkinson's. Joy was presented with 12 different shirts, each worn by a different person, 6 of whom had Parkinson's and 6 of whom did not. The shirts were given to Joy in a random order, and she had to smell each shirt and decide whether each shirt was worn by a Parkinson's patient or not.

What type of variable is this?

Let's write the hypotheses for this test. What proportion would she get right if she was guessing?

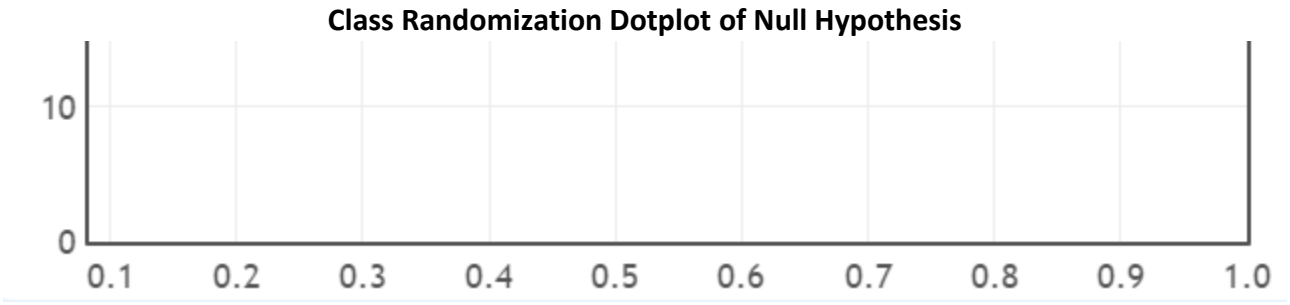
Next, we will do a physical simulation to build a null distribution (guessing randomly) for a sample size of size n=12. Each group will get 12 shirt cards. Mix up your shirt cards without looking at the backs and randomly divide them into 2 groups using the boxes below.

Has Parkinson's	Does not have Parkinson's

Now look on the back of each shirt and count how many were correctly identified. In the table below, use the first column for tally marks, then write the number correct and the proportion correct. Repeat this until you have 3 trials. Mark your proportions on the graph below and the class graph on the document camera.

Your group results:

Trial	Tally marks for correct identifications	Number Correct	Proportion Correct
1			
2			
3			



Proportion Correct in Samples of 12 when Guessing Randomly

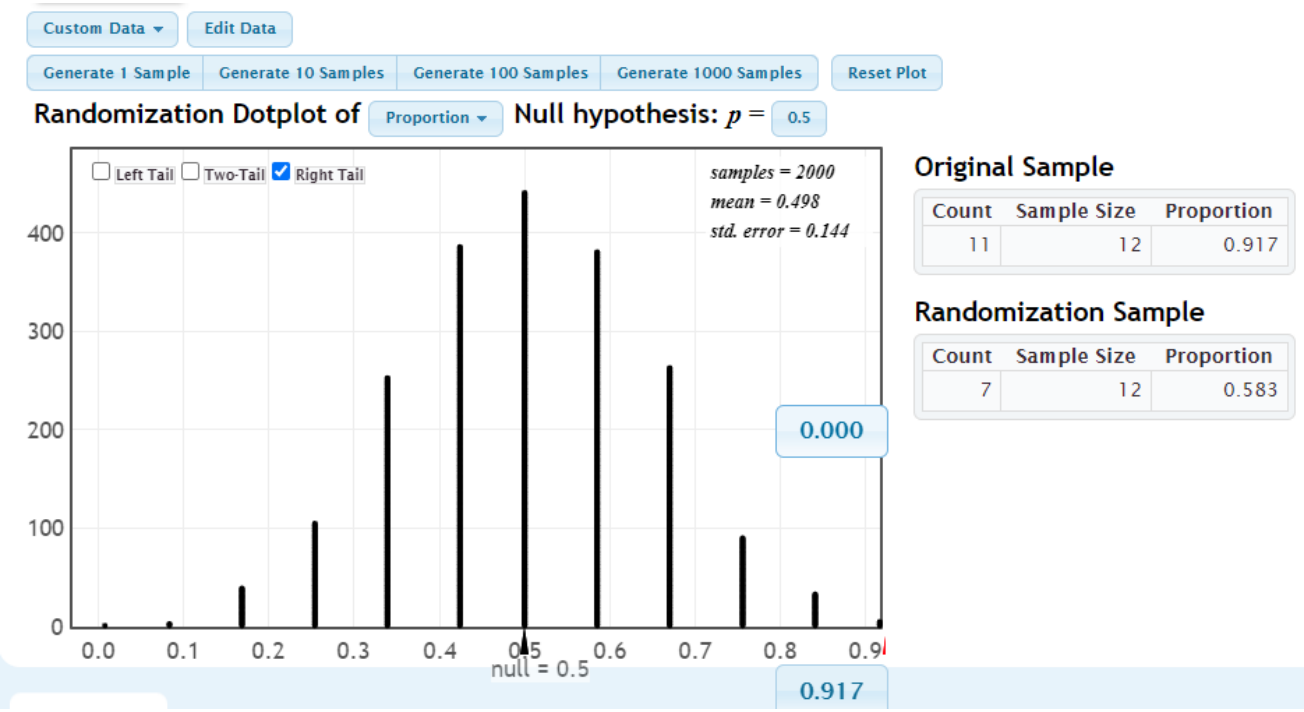
This would take a long time to build a full randomization distribution, so we will use StatKey to speed it up. Select the StatKey Applet under **Randomization Hypothesis Tests** for Single Proportion.

In the experiment, Joy guessed 11 out of the 12 shirts correctly.

Write the observed statistic (also called sample statistic):

In StatKey, select Edit Data and enter 11 out of 12. The Null hypothesis is already set to 0.5. Generate 1 sample a few times so you see what the computer is doing. Then generate approximately 2000 samples.

This is a randomization distribution of the null hypothesis:



Proportion of Correct Guesses out of 12 Tries

Write the p-value from your simulation. These vary each time you do the simulation.

What does the p-value represent?

Write the conclusion of this hypothesis test in context, including the p-value.

Lesson 2: Sampling Distributions and Theoretical Hypothesis Test for One Proportion

Last time we did a simulated hypothesis test and this time we will review the theoretical test. First let's write down the hypotheses and the observed statistic. Then we will review sampling distributions and the central limit theorem.

According to the Oregon Health Authority, 96% of Oregonians had some type of health insurance coverage in 2024. We want to test whether this percentage is lower among PCC students. A random sample of 300 PCC students shows that 284 had health insurance coverage. Is this evidence that the proportion is lower among PCC students than the general Oregon population?

Hypotheses:

Observed statistic:

Comparing the two methods for finding a p-value in hypothesis testing:

- Simulation – we can use simulation in any situation as long as we can imitate the situation.
- Theoretical – we need a mathematical model based on the central limit theorem. The conditions must be met in order for the model to be valid. This can be checked on a theoretical calculator.

A **sampling distribution** is the distribution of a statistic calculated from many random samples of the same size.

Figure 1. Normal Population

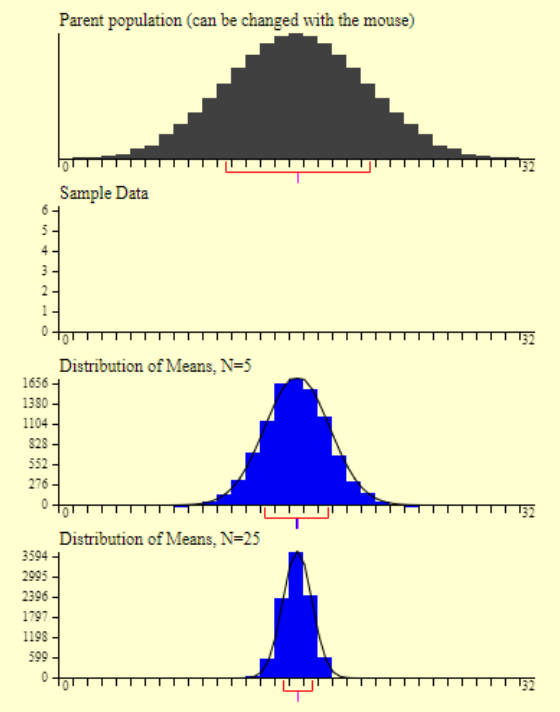


Figure 2. Uniform Population

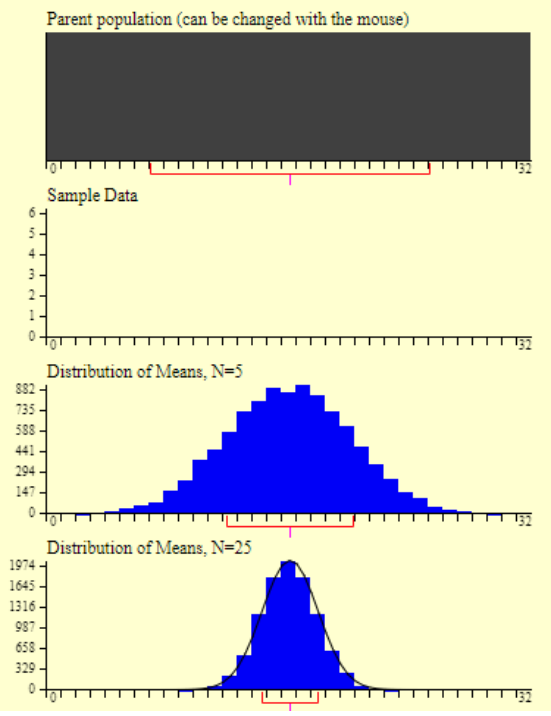


Figure 3. Skewed Population

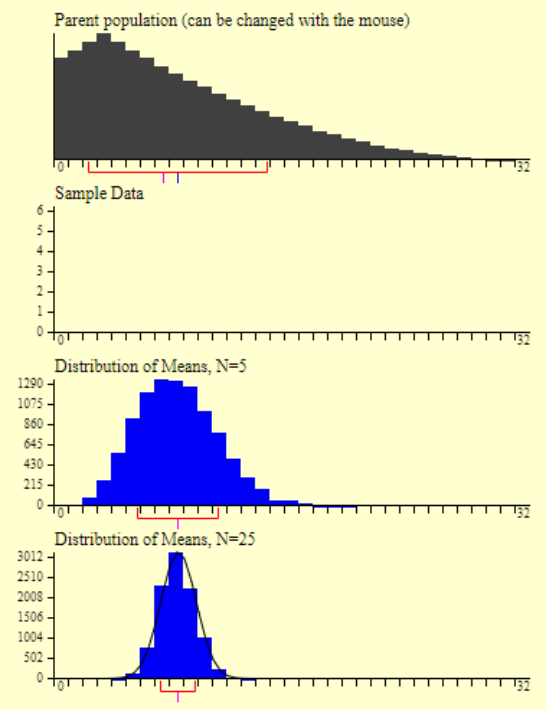
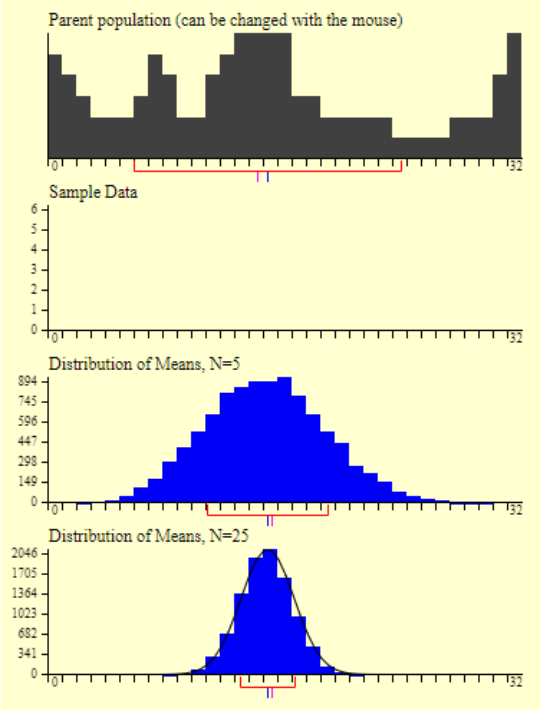


Figure 4. Multimodal Population



Sampling Distribution for a Proportion - Conditions and Model

We have seen in our simulations that the distribution of a statistic for many random samples of the same size from the same population is close to a normal model. We've also seen that the mean stays the same but the standard error gets smaller the larger the sample size. Let's summarize the Central Limit Theorem for the sampling distributions of a proportion.

The Central Limit Theorem Conditions and Model for a Proportion, $\hat{p}$

When using a theoretical model, we need to check the conditions or assumptions to ensure the model applies to the situation.

1. **Independence**: The individuals or items must be independent of each other regarding the variable measured.
2. **Randomization**: The samples need to be randomly chosen, or it's not safe to assume independence.
3. **Large Population**: If sampling without replacement, we must be sampling from a large population to consider the individuals independent of each other. Sometimes this is called the 10% Condition or said that the population must be at least 10 times our sample size.
4. **Success/Failure Condition**: You should expect to have at least 10 successes and 10 failures in your data to ensure a normal distribution. Check that $np \geq 10$ and $nq \geq 10$.

When the conditions are met, the sampling distribution for a proportion, $\hat{p}$, is modeled by a normal distribution with the following parameters:

$$\hat{p} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$$

Theoretical Hypothesis Tests for Proportions

The theoretical hypothesis test for a proportion uses the normal model that comes from the Central Limit Theorem. The steps are similar to using a randomization distribution, but we calculate a test statistic and find the p-value using the normal distribution.

Steps for a Hypothesis Test with the Normal Distribution (Z Test of a Proportion)

a. Write the null and alternate hypotheses.

b. Check the conditions to use the Central Limit Theorem: Independence, Randomization, Large population, and Success/Failure condition using p_0.

c. Write the observed statistic, $\hat{p}$, and calculate the test statistic, $Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$.

   Use a normal calculator with $Z \sim N(0,1)$ to find the p-value.

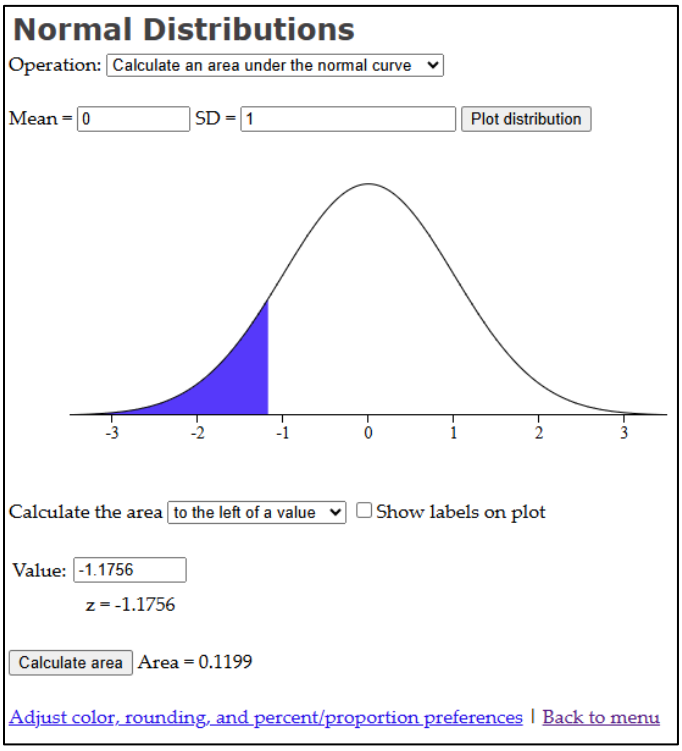
d. Compare the p-value with the significance level, α, and determine whether the result is statistically significant.

e. State the conclusion in context, including the p-value and whether we reject or fail to reject the null hypothesis.

Next, let's write out the conditions for the health insurance test.

Calculate the test statistic. What is a z-score and what does it represent?

Find the p-value using the Normal Distribution on Stapplet:



[Stapplet Normal Calculator]

In the Stapplet Normal Calculator, make sure the Operation Menu is set to:
Calculate an area under the Normal Curve.

Type in 0 for the mean and 1 for the standard deviation and click Plot Distribution.

Then select the type of probability you want: between two values, to the left, or right and enter your value(s) and click on Calculate area. Notice that Stapplet will give you the Z-score of the value and label the point if you select show labels on plot.

You can use the link at the bottom to switch between percentages and proportions and change the number of decimal places given.

Find the theoretical p-value:

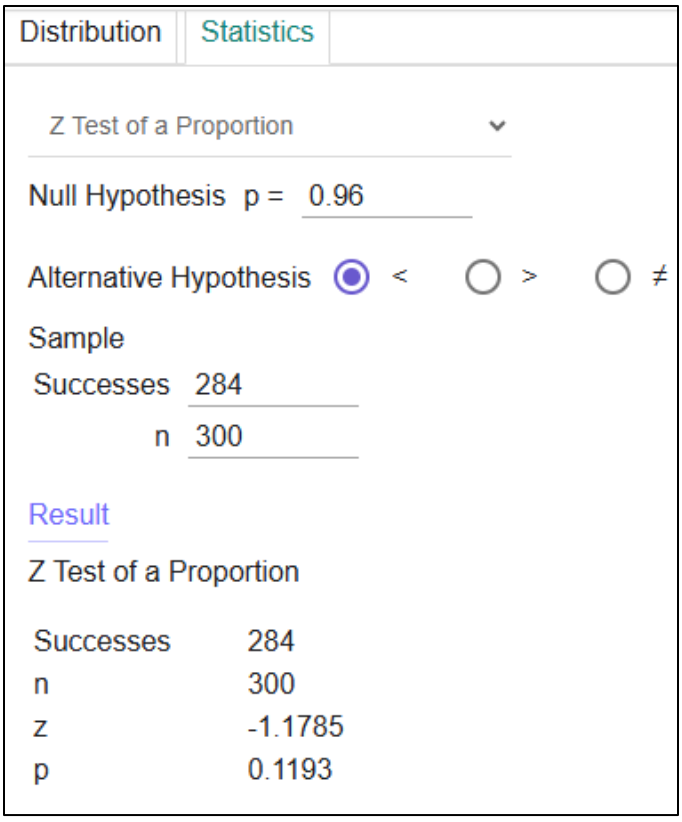
Check with the Theoretical HT Calculator on GeoGebra:

[GeoGebra Classic Probability Calculator]

Select the Statistics Tab

For a confidence interval of a proportion, select **Z Test of a Proportion**.

Enter the null hypothesis and sample proportion and the result appears below.



Write the conclusion of the hypothesis test:

Why the Large Population Condition?

Small vs. Large Populations

Imagine you have one fun size bag of m&m's as your population. If you take out one m&m, what happens to the probability of selecting each color?

If you have a huge box of m&m's as your population, what happens to the probability of selecting each color when you take out one?

Figure 5. A 0.53-ounce bag



Figure 6. A 25-pound box



As long as we sample less than 10% of the population, we can say the individuals will be independent and will meet this condition. Some books say the population should be 10-20 times the sample size or more.

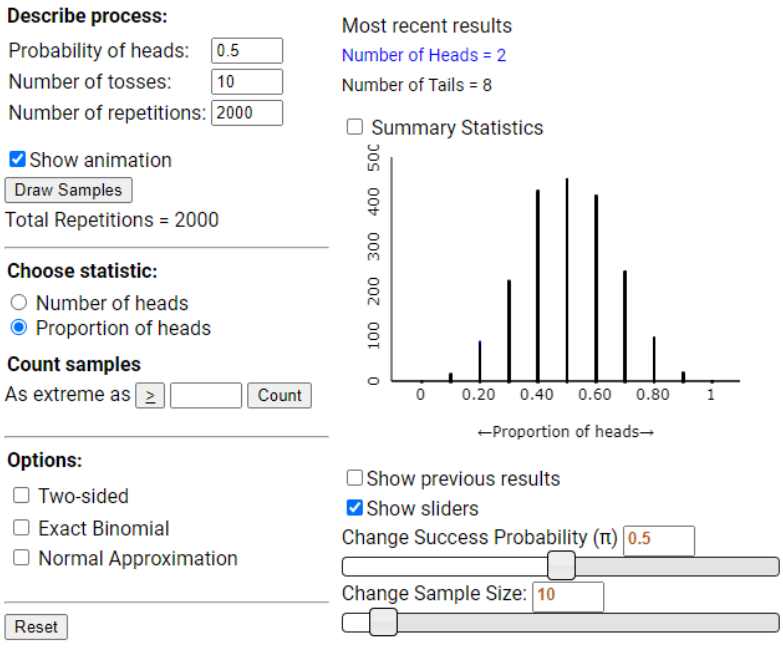
There are methods for small samples and small populations that are beyond the scope of this course.

Why the Success/Failure Condition?

The sample size condition for proportions depends on both n and p, rather than just n in the case of means. A good place to explore this is the Rossman/Chance One Proportion Inference Applet.

Simulate 2000 repetitions of size n=10 with a 0.50 probability of heads. Then click on show sliders. Move the slider for success probability and notice what happens.

Move the slider for n and notice what happens.



Conditions Worksheet for a Single Proportion

For the practice problem on the next page, identify each of these elements and use them to write out the conditions.

Subjects	
Variable	
What factors might influence the variable for these subjects?	
Population	
Sample size	
p_0 and/or $\hat{p}$	

Conditions for a single proportion

1. Independence: The ***subjects*** must be independent of each other in terms of the ***variable***. List some ***factors*** to consider when choosing a sampling method.

2. Randomization: We need a random sample.

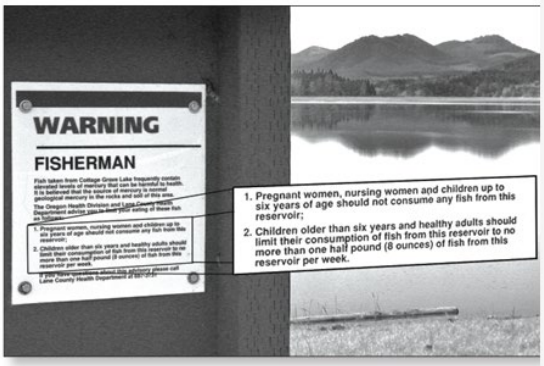
3. Large Population: The ***population*** must be at least 10x***n*** = _____.

5. Success/Failure Condition: Check that $\boldsymbol{np} \geq 10$ and $\boldsymbol{nq} \geq 10$. Use $\boldsymbol{p_0}$ for hypothesis tests and $\boldsymbol{\hat{p}}$ for confidence intervals.

Lessons 1-2 Practice Problem

It has been estimated that fish in 20% of freshwater streams in the US have levels of mercury that exceed the EPA standards. A report by the USGS in 2014[1] estimated that fish in a quarter of 300 streams across the US have levels of methylmercury that exceed the EPA standards.

[Note: The average mercury concentrations in most commercial fish and seafood fall within levels considered safe for people who consume average amounts of fish.]



A posted advisory communicates the risk of eating fish that contain high concentrations of mercury. (*Photograph by Dennis A. Wentz.*)

a. Test the claim that the true proportion of streams that have fish with high levels of mercury is different than originally estimated. Write out the conditions and steps for a theoretical test.
b. Check your results on a theoretical calculator.
c. Repeat the test using a simulation and compare your results.

[1] Wentz, D.A., Brigham, M.E., Chasar, L.C., Lutz, M.A., and Krabbenhoft, D.P., 2014, Mercury in the Nation's streams— Levels, trends, and implications: U.S. Geological Survey Circular 1395, 90 p., http://dx.doi.org/10.3133/cir1395.

Lesson 3: Theoretical and Simulated Confidence Intervals for One Proportion

A Pew Research Center poll asked 1453 U.S. teens, aged 13 to 17 about their views on social media use. Of the teens surveyed, 668 said they supported requiring parental consent for minors to create a social media account. Create a 95% confidence interval for the true proportion of teens who support parental consent to create a social media account. Source: Pew Research Study.

What is the point estimate?

What is the confidence level?

Steps for a Confidence Interval with the Normal Distribution (Z Estimate of a Proportion)

a. Calculate the point estimate and find the critical value, z^*, for the confidence level.

b. Check the conditions to use the Central Limit Theorem: Independence, Randomization, Large population, and Success/Failure condition using $\hat{p}$.

c. Calculate the interval using the formula:

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

d. Write the interpretation in context.

Note, the standard error is given by $SE = \sqrt{\frac{\hat{p}\hat{q}}{n}}$ and the margin of error is given by $ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$.

Check the conditions needed to use the Central Limit Theorem.

Calculate the 95% confidence interval using $z^* = 1.96$.

Write the interpretation in context.

What does this confidence interval represent?

Finding Critical Z-Values for Confidence Intervals with Stapplet:

Stapplet Normal Calculator

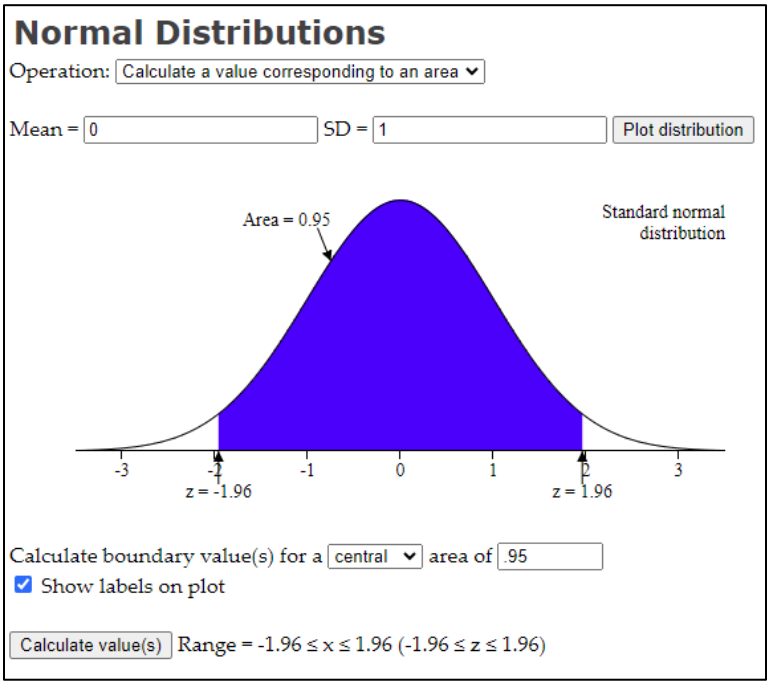
A critical z-value is a cutoff value, so we'll use the inverse normal function on Stapplet.

Select *Calculate a value corresponding to an area.*

Then enter 0 for the mean and 1 for the standard deviation and click on Plot distribution.

For a 95% confidence interval, we select central and enter 0.95.

Do this for 80%, 90%, 95% and 99%. These are the most commonly used Z-values, but you can find another one if needed.



Confidence Level	Critical Z-value
80%	
90%	
95%	
99%	

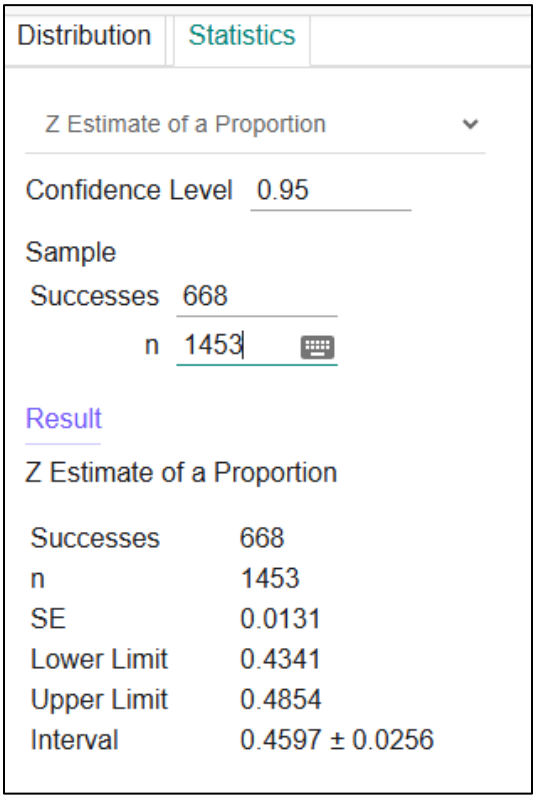
Theoretical Confidence Interval Calculator on GeoGebra:

[GeoGebra Classic Probability Calculator]

Select the Statistics Tab.

For a confidence interval of a proportion, select Z Estimate of a proportion and enter your confidence level as a decimal.

Enter the sample proportion and the result appears below.

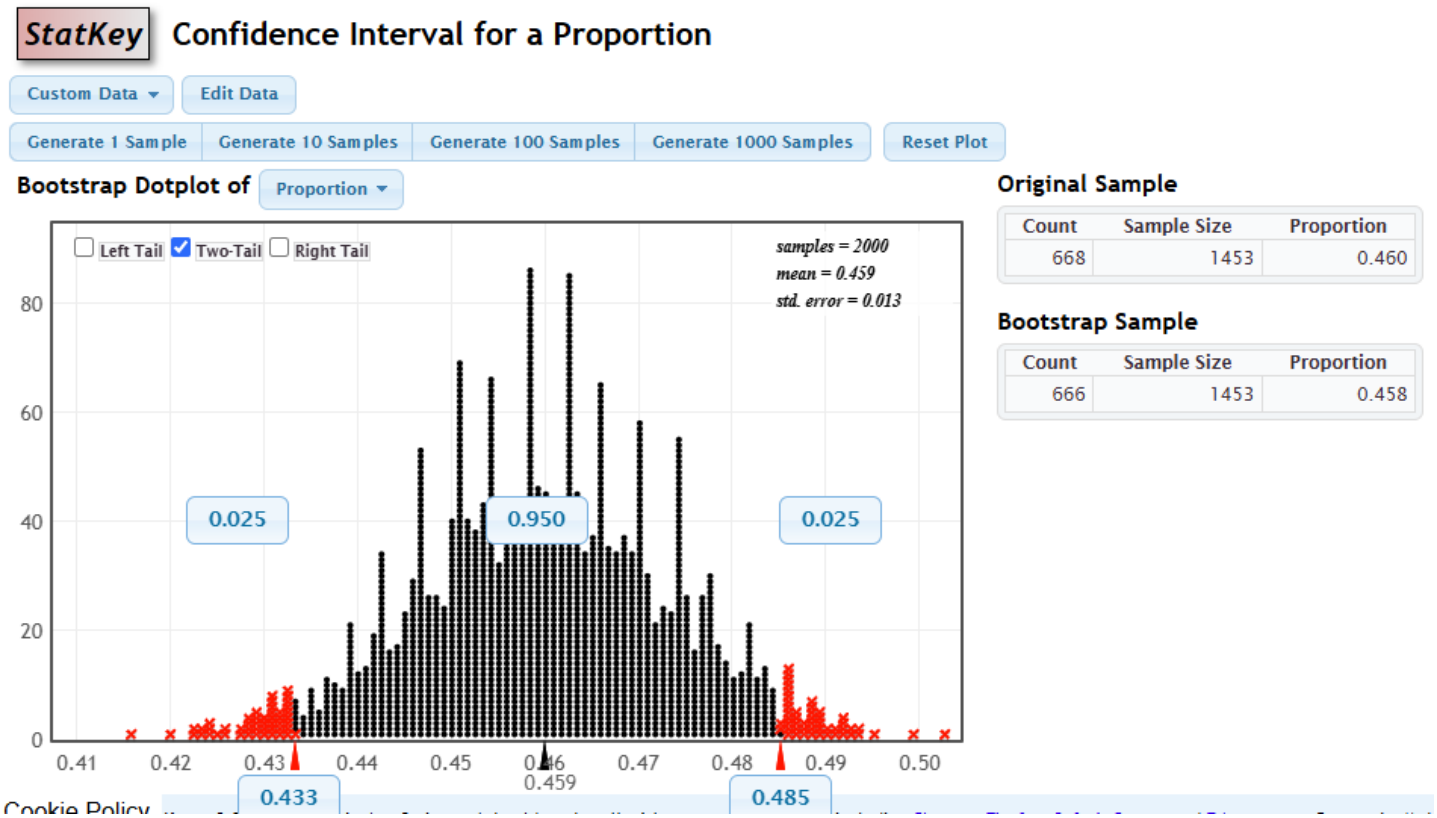


Bootstrap Confidence Intervals

Now let's find a confidence interval for the same situation using simulation. Confidence intervals are simulated using the bootstrapping method.

Select the StatKey Applet under **Bootstrap Confidence Intervals** for CI for Single Proportion.

Click on Edit Data and enter 668 out of 1453. Click on Ok, then generate 2000 samples.



Check the **Two-Tail** box to find the 95% confidence interval:

Edit the upper middle box to find a 90% confidence interval:

Edit the upper middle box to find an 80% confidence interval:

What happens to the interval as the confidence level gets larger? Why?

Finding an Ideal Sample Size

When given a desired margin of error, ME, we can find the sample size, n, needed using this formula.

$$n = \hat{p} \cdot \hat{q} \left(\frac{z^*}{\text{ME}} \right)^2$$

This is found by solving the margin of error formula for n: $ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$.

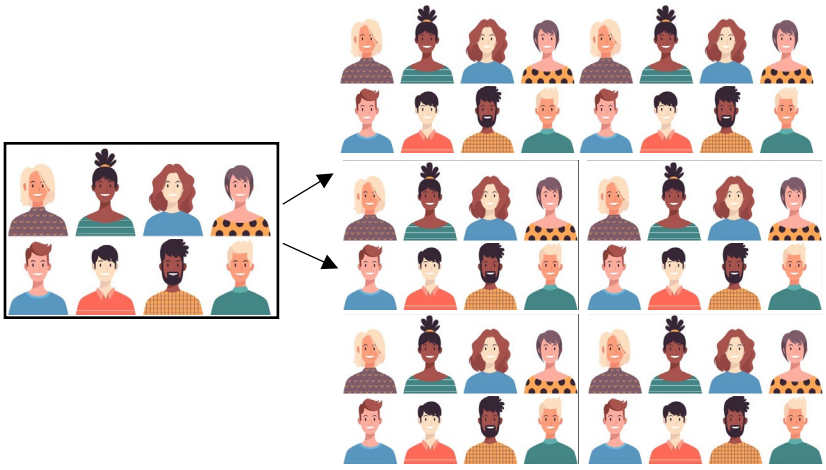
Sample size must be a whole number. This is a minimum value, so we always round up, even if the decimal is under 0.5. If we were to round down that would make the margin of error larger than the desired value.

Continuing the example above, calculate the sample size needed to reduce the margin of error to 1.5%.

Bootstrap Confidence Intervals - Explanation

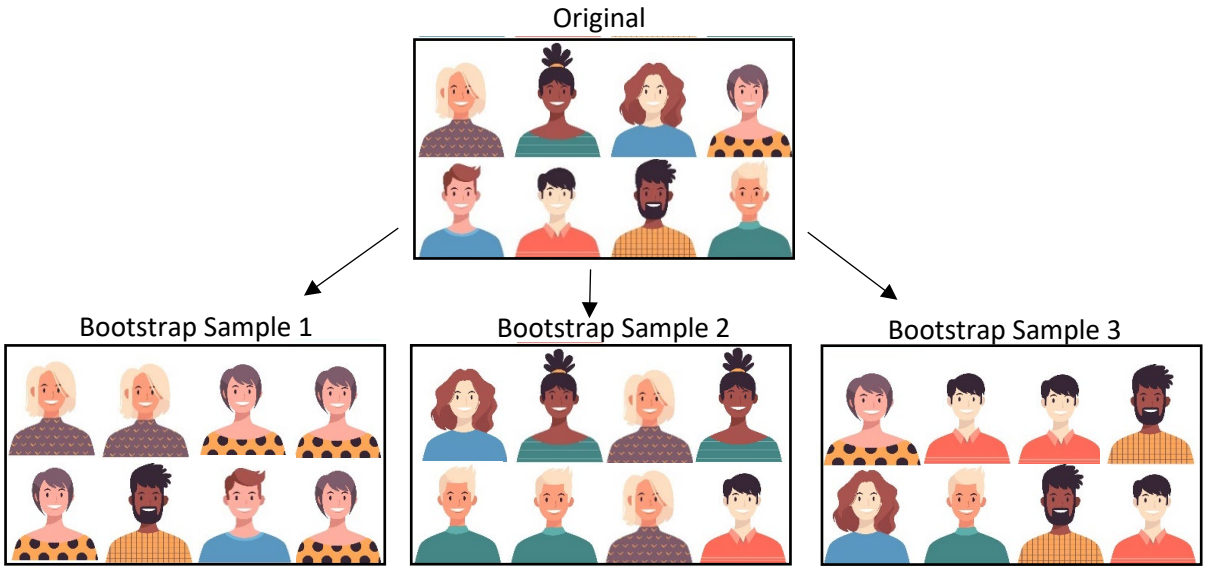
Bootstrap Samples

In real life, we take one sample from our population. If we had more resources to do another sample, it would be better to just take a larger sample. We don't have the information for the full population, but, if our sample is representative, we can think of the population as many copies of our sample. If we sample many times from that we can estimate our sample-to-sample variation.



Bootstrapping is the idea that a population is many copies of a representative sample, so we take many samples from that "population." Another way to think about it is called sampling with replacement. That means each time we choose a subject; they go back in the hat and can be chosen again.

Just like with a sampling distribution, we use the distribution of random samples of the same size to estimate the sample-to-sample variation.



Generating a Bootstrap Distribution

In the same way we simulated sampling distributions, we will take repeated samples from one sample using specific applets to form a distribution. It has been shown that the sample-to-sample variation of the bootstrap distribution is very close to that of the sampling distribution.

How Can This Work?

If you are skeptical, I was, too. There is an optional video you can watch in D2L showing that bootstrap distributions have a very similar standard error as their corresponding sampling distributions.

Conditions Worksheet for a Single Proportion

For the practice problem on the next page, identify each of these elements and use them to write out the conditions.

Subjects	
Variable	
What factors might influence the variable for these subjects?	
Population	
Sample size	
p_0 and/or $\hat{p}$	

Conditions for a single proportion

1. Independence: The ***subjects*** must be independent of each other in terms of the ***variable***. List some ***factors*** to consider when choosing a sampling method.

2. Randomization: We need a random sample.

3. Large Population: The ***population*** must be at least 10x***n*** = _____.

4. Success/Failure Condition: Check that $\boldsymbol{np} \geq 10$ and $\boldsymbol{nq} \geq 10$. Use $\boldsymbol{p_0}$ for hypothesis tests and $\boldsymbol{\hat{p}}$ for confidence intervals.

Lesson 3 Practice Problem

A poll of 500 voters in the Portland-metro area asked questions on the direction of the city. When asked an open-ended question about the region's top concern, 40% named homelessness as the biggest issue facing Portland. Source: OPB Article.

a. Calculate and interpret a 90% confidence interval for the true proportion of Portland-metro voters who think homelessness is the biggest issue facing Portland. Check the conditions, use the formula and check with a theoretical calculator. Then do a simulation and see how it compares.
b. If the study were to be repeated using the same confidence level, what sample size would be needed for a margin of error of 2%?

Lesson 4: Comparing HT with CI, Error Types and Power

How are confidence intervals and hypothesis tests similar and different?

Confidence Intervals	Hypothesis Tests
Determines a range of plausible values for a parameter based on a sample.	Determines statistical significance (yes/no) based on a sample.
Determines the size of the effect.	Determines the strength of evidence against the null hypothesis.
2-sided, using a confidence level, C (central area).	Can be 1-or 2-tailed, using a significance level, α (compared with p-value).
Centered at $\hat{p}$, SE uses $\hat{p}$.	Centered at p_0., SE uses p_0.

Visual Comparison

Figure 1. A 1-tailed test

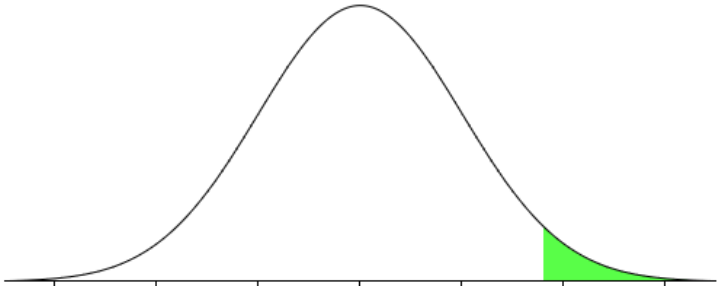
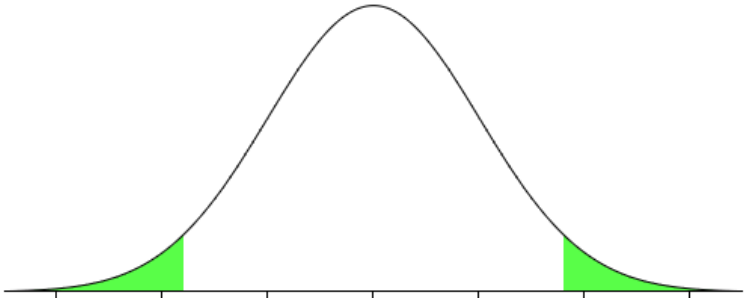


Figure 2. A 2-tailed test



- If the null value, p_0, falls within the confidence interval, that supports the null hypothesis, and we fail to reject H_0.
- If the interval does not contain the null value, p_0, and the interval is in the direction of the claim, that is evidence against the null hypothesis, and we reject H_0.

Example: Use the Lesson 3 confidence interval for the true proportion of US teens who supported parental consent for social media (.4341, .4854) to test the hypothesis that fewer than 50% of teens are in favor of parental support. Run a simulated test with a significance level of .05 and see whether it agrees. $\hat{p} = \frac{668}{1453}$.

The Relationship Between the Confidence Level (C) and Significance Level (α)

Draw a normal curve for each situation to calculate the significance level that matches the confidence level or vice versa.

Two-tail test	**One-tail test**
If $\alpha = 0.10$, find C	If $\alpha = 0.05$, find C
If $\alpha = 0.05$, find C	If $\alpha = 0.025$, find C
If C=99%, find α	If C=99%, find α
General Formula: $\alpha = 1 - C$ or $C = 1 - \alpha$	General Formula: $\alpha = \frac{1-C}{2}$ or $C = 1 - 2\alpha$ (Or draw it out, that's what I do!)

Type I and II Errors

Even if we do the statistical procedures correctly, there is still an element of chance and we may not make the correct decision.

Example: A medical test

H_0: The test is negative

H_A: The test is positive

	H_0 is really true	H_0 is not really true
We reject H_0		
We fail to reject H_0		

A Type I error is…

A Type II error is…

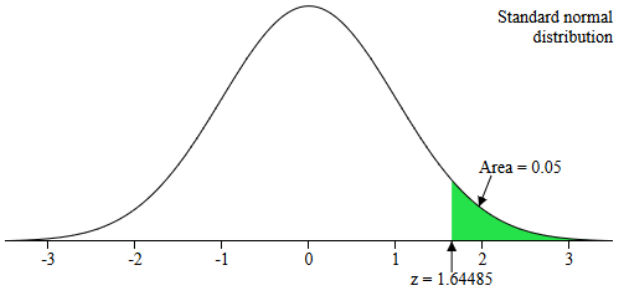
Example: Write out the hypotheses and determine what a Type I and II error would be in this situation. What are the possible consequences of each error type and which one do you think is worse?

a. A person tests to see if they are pregnant.

b. A person is on trial for theft.

Choosing the Significance Level for a Test

Let's say the null hypothesis is true, and we have a significance level of 0.05. This is the threshold for rejecting the null, and saying the result did not happen by chance.
So we are saying we are willing to reject the null if there is less than a 5% probability that the observed result happened by random chance.



The probability of making a Type I error is α.

The probability of making a Type II error is called β.

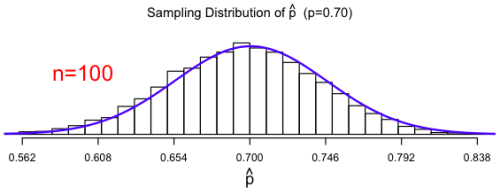
There is a trade-off between α and β. If we make α smaller then β gets larger, and vice versa. The neutral level for α is 0.05.

If making a Type I error is more dangerous or costly, we should make α _____.

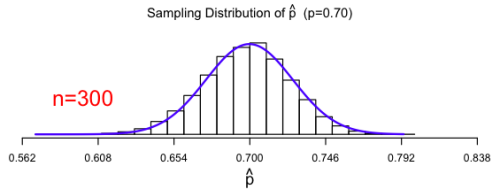
If making a Type II error is more dangerous or costly, we should make α _____.

The Power of a Test

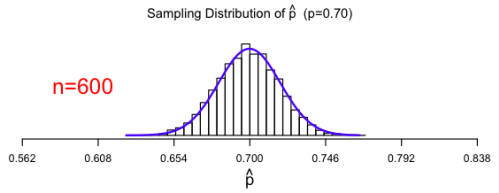
We need to be concerned about the power of a test in case we fail to reject the null hypothesis. Was this result correct or was our sample size not big enough to detect it? A smaller sample size gives a larger standard error, so we would need a larger difference to be significant.



The power of a test is the chance we will correctly identify a significant difference and reject the null if it is actually false. It is also the chance of **not** making a Type II error.



The power of a test is $1-\beta$, which ranges from 0 to 1, with 0.80 as a baseline. If a statistician calculates a power of 0.80 or greater they will usually continue with the test.



Factors affecting power

- If we increase α, then β decreases and the power $(1-\beta)$ goes up
- If we decrease α, then β increases and the power $(1-\beta)$ goes down
- Increasing the sample size will generally increase the power and decrease the chances of both Type I and Type II errors. (It decreases the SE)

Lesson 4 Practice Problem 1: Comparing HT and CI Results

Hypothesis Test: Chevy brand cars got recalled due to the heavy keys bouncing out of the ignition while the car was running. It was determined that approximately 20% of all Chevy cars had this problem. After fixes were made, Chevy hired 3 independent testing firms to each test a representative sample of cars to see if the situation was improved.

Hypotheses:

Assume the conditions are met and use the GeoGebra Theoretical calculator to find each p-value and confidence interval, then determine whether they agree and the conclusion.

	Results of HT, $\alpha = .01$	Results of 98% CI	Conclusion
Sample 1: $\hat{p} = \frac{23}{200}$			
Sample 2: $\hat{p} = \frac{19}{150}$			
Sample 3: $\hat{p} = \frac{15}{100}$			

Why is it important to do a hypothesis test and confidence interval together?

Lesson 4 Practice Problem 2

To get a voter initiative on a state ballot, petitions that contain at least 250,000 valid voter signatures must be filed with the Elections Commission. The board then has 60 days to certify the petition. A group wanting to create a statewide system of universal health insurance filed a petition with a total of 304,266 signatures. As a first step in the process, the board checks a random sample of 200 signatures and 173 are valid.

a. What proportion of the signatures collected need to be valid in order to meet the requirement?
b. Test the claim that they meet the requirement based on the sample taken using the theoretical method using $\alpha = 0.05$. Check this on a theoretical calculator and find a simulated p-value as well.
c. Describe what a Type I error would be in this situation.
d. Describe what a Type II error would be in this situation.
e. Would the result change if instead they had 346 valid signatures out of 400? What does this tell us about power?

Lesson 5: Confidence Intervals for 2 Proportions

Juvenile Diabetes (Type I) is usually diagnosed in children and has a genetic component. A study enrolled 76 children at high risk for the disease. The kids were randomly assigned to either a 2-week intravenous course of the drug teplizumab or a control group. The study ended when 42 kids developed the disease, and the results are shown in the table.

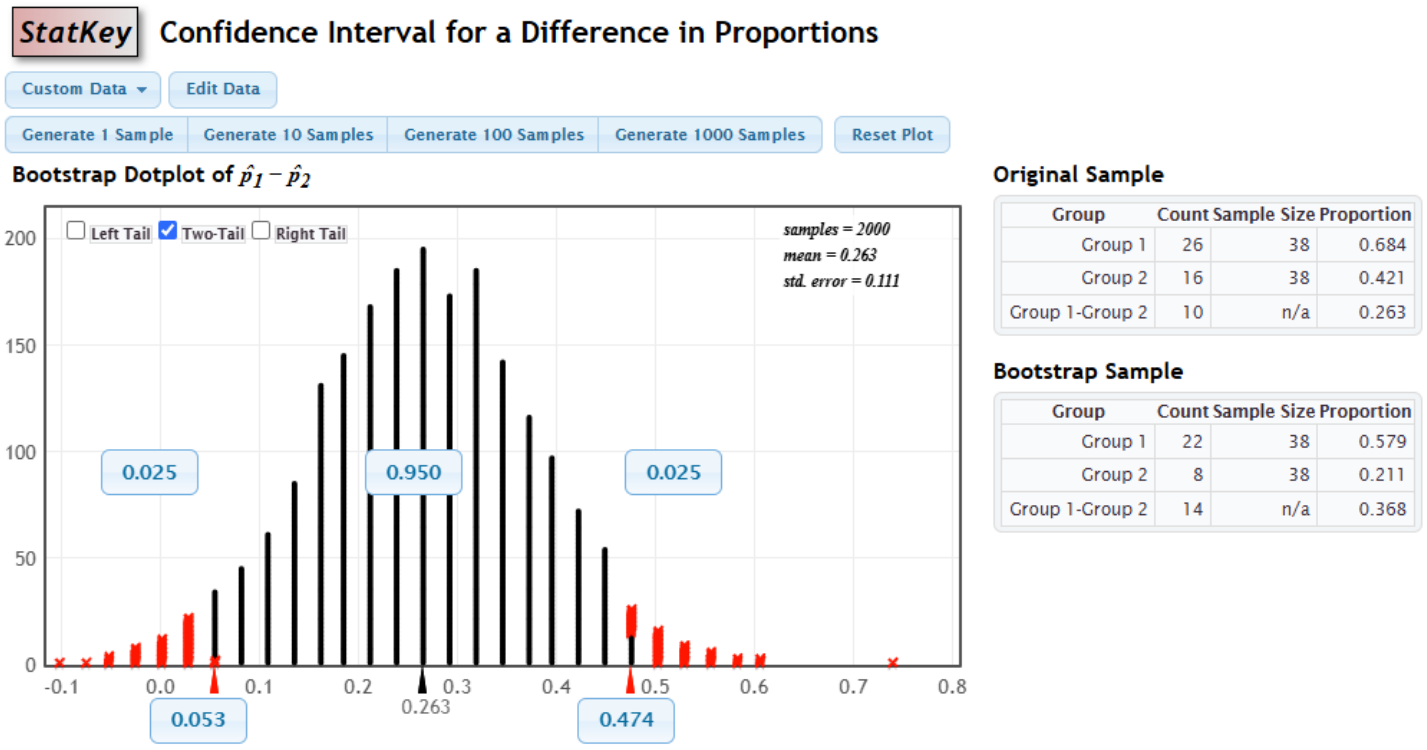
	Diabetes	No Diabetes	Total
Control	26	12	38
Teplizumab	16	22	38
Total	42	34	76

a. Find the proportion in the control group that developed diabetes, $\hat{p}_C$.

b. Find the proportion in the treatment group that developed diabetes, $\hat{p}_T$.

c. Find the difference in the sample proportions, $\hat{p}_C - \hat{p}_T$ (Control - Teplizumab).

d. Find a bootstrap 95% confidence interval for the difference and write the interpretation.



e. Can we be confident at the 99% level that the drug teplizumab is more effective than the control? Find the bootstrap confidence interval and explain.

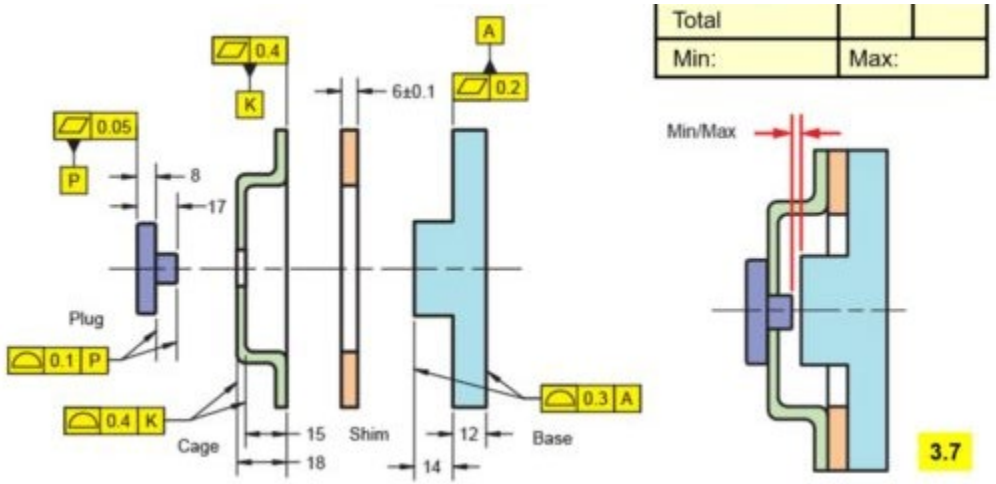
How to Write Interpretations

Confidence Interval	Result	Interpretation
(positive, positive)		With C% confidence, the true proportion of _____ is __% to __% higher for group 1 than group 2.
(negative, negative)		With C% confidence, the true proportion of _____ is __% to __% lower for group 1 than group 2. (reverse the values and make positive)
(negative, positive)		With C% confidence, there is not a significant difference in the true proportion of _____ for group 1 and 2.

Developing a Model for 2 Proportions

Before we define a model, let's look at what happens when we have more than one thing that varies.

Tolerance stack in manufacturing - multiple variables



When we add or subtract variables, it increases the variation. If the variables are independent, we can add the variances. $Var = SD^2$, or the value before you take the square root.

$$Var(A - B) = Var(A) + Var(B)$$

$$SD(A - B) = \sqrt{Var(A) + Var(B)}$$

or

$$SD(A - B) = \sqrt{SD(A)^2 + SD(B)^2}$$

Using this rule to add the variances, we get this model, if the conditions are met:

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right)$$

Confidence Interval for 2 Proportions (Z Estimate, Difference of Proportions)

Conditions

1. **Independence**: The individuals in each group and between groups must be independent of each other regarding the variable measured.
2. **Randomization**: The data are from 2 independent random samples or 2 randomly assigned treatments.
3. **Large Population for each group**: If sampling without replacement, the population must be 10 times the size of each group.
4. **Success/Failure Condition for each group:** $n_1\hat{p}_1$ and $n_1\hat{q}_1 \geq 10$ and $n_2\hat{p}_2$ and $n_2\hat{q}_2 \geq 10$. If the data are given in a table, the 4 interior numbers are the success and failures.

Formula

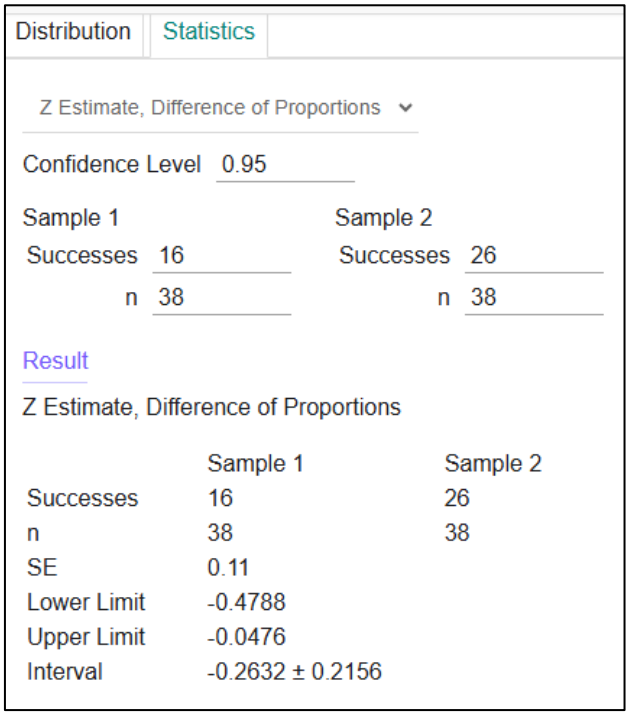
$$\hat{p}_1 - \hat{p}_2 \pm z^* \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

The critical z-values are the same as for 1 proportion:

Confidence Level	Critical Z-value
80%	1.2816
90%	1.6449
95%	1.96
99%	2.5758

a. Write the conditions needed to find a theoretical confidence interval for difference in the true proportions of kids who develop Type I diabetes, $p_T - p_C$ (Teplizumab - Control). Note: The order is reversed so we can see what happens.

b. Use the formula to calculate and interpret the theoretical 95% confidence interval for the difference in the true proportion of kids who develop Type I diabetes, $p_T - p_C$ (Teplizumab - Control). Check with the GeoGebra theoretical calculator.



Conditions Worksheet for 2 Proportions

For the practice problem, identify each of these elements and use them to write out the conditions.

Subjects and groups	
Variable	
What factors might influence the variable for these subjects/groups?	
Population and Randomization: Is there a single population with 2 randomly assigned treatments or 2 populations with independent random samples?	
Sample sizes	
CI: Number of successes and failures HT: $\hat{p}_C$	

Conditions for the difference of 2 proportions

1. Independence: The ***subjects*** in each group and between groups must be independent of each other in terms of the ***variable***. List some ***factors*** to consider when choosing a sampling method.

2. Randomization: State whether we have 2 independent random samples or 2 randomly assigned groups.

3. Large Population: The ***population*** must be at least 10x$\boldsymbol{n_1}$= _____and 10x$\boldsymbol{n_2}$= _____, or ***population 1*** must be at least 10x$\boldsymbol{n_1}$= _____and ***population 2*** must be at least 10x$\boldsymbol{n_2}$= _____.

4. Success/Failure Condition:
   CI: Check that the successes and failures are all greater than 10 (or $n_1\hat{p}_1$ and $n_1\hat{q}_1$ and $n_2\hat{p}_2$ and $n_2\hat{q}_2 \geq 10$).
   HT: Check that $n_1\hat{p}_C$ and $n_1\hat{q}_C$ and $n_2\hat{p}_C$ and $n_2\hat{q}_C \geq 10$.

Lesson 5 Practice Problem 1

How detrimental are metal tags to penguins? A study looked at the 10-year survival rate of penguins tagged either with a metal tag or an electronic tag. The results are shown below. Check the conditions and calculate the theoretical 90% confidence interval for the difference in survival rate (metal – electronic). Check your result with a StatKey simulation and on GeoGebra.

	Survived	**Died**	**Total**
Metal Band	33	134	167
Electronic Tag	68	121	189
Total	101	255	356

Lesson 5 Practice Problem 2

Errors in medical prescriptions occur, and a study examined whether electronic prescribing would help reduce errors. Two groups of doctors used written prescriptions and had a similar error rate before the study. One group switched to e-prescriptions while the other continued to use written prescriptions, and error rates were measured a year later. The results are given in the table. Check the conditions and calculate the theoretical 95% confidence interval for the difference in proportion of errors between the two groups (written – electronic). Check your result with a StatKey simulation and on GeoGebra.

	Error	**No Error**	**Total**
Written	1478	2370	3848
Electronic	254	3594	3848
Total	1732	5964	7696

Lesson 6: Hypothesis Testing for 2 Proportions

Dolphin Therapy Experiment. Researchers wanted to know whether swimming with dolphins would be therapeutic for patients with clinical depression. They recruited 30 subjects aged 18-65 with a clinical diagnosis of mild to moderate depression. Other treatments were discontinued 4 weeks prior and during the experiment. The subjects went to an island near Honduras where they were randomly assigned to two treatment groups. All other conditions were identical, except for the presence of dolphins.

- Dolphin group – engaged in one hour of swimming and snorkeling with dolphins each day.
- Control group – engaged in one hour of swimming and snorkeling each day without dolphins.

Identify the response variable:

Identify the explanatory variable:

	Dolphin group	Control group	Total
Improved	10	3	13
Not Improved	5	12	17
Total	15	15	30

Write the hypotheses:

Physical Simulation

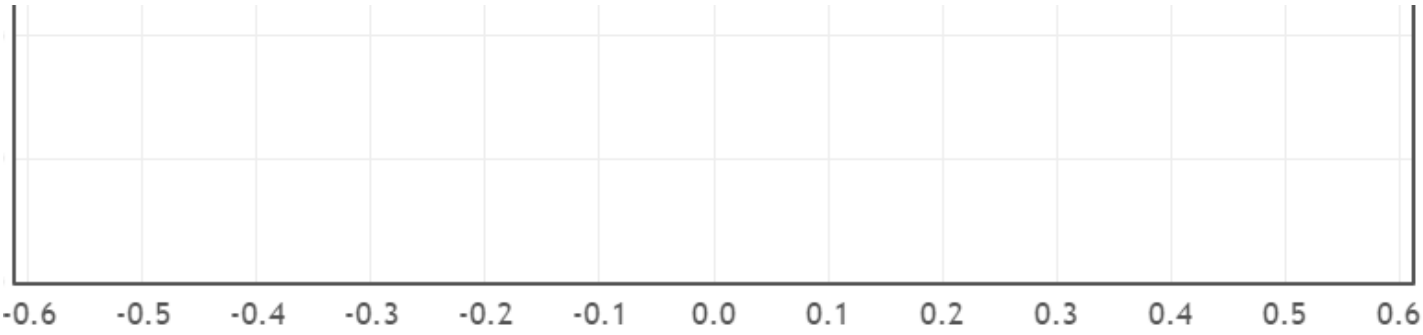
Get a set of response cards: 13 improvers cards and 17 non-improvers. Shuffle them and divide them randomly into two groups of 15 using the boxes below, because the explanatory groups were equal in size.

Explanatory Variable

Dolphin Therapy Group (15 people)	Control Group (15 people)

Trial	Improvers in Dolphin Group Proportion	Improvers in Control Group Proportion	Difference in the proportions
1			
2			
3			

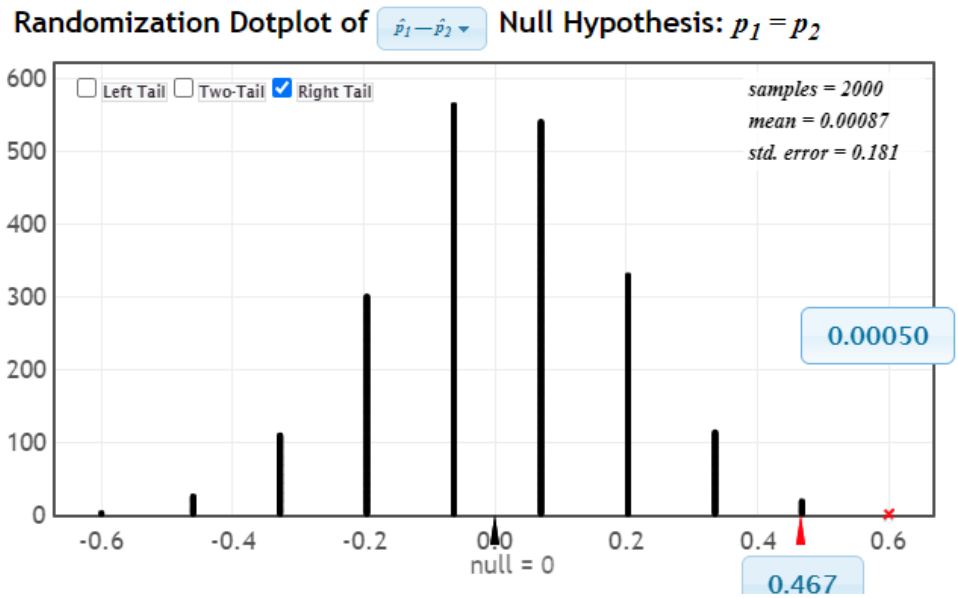
Class Randomization Dotplot of Null Hypothesis (no difference in the treatment)



What do you notice so far in the graph?

Now to do more simulations, go to StatKey test for difference in proportions. Enter the observed proportions for the response variable: improved in each group.

Write down your simulated p-value with your conclusion in context.



Original Sample

Group	Count	Sample Size	Proportion
Group 1	10	15	0.667
Group 2	3	15	0.200
Group 1-Group 2	7	n/a	0.467

Randomization Sample

Group	Count	Sample Size	Proportion
Group 1	8	15	0.533
Group 2	5	15	0.333
Group 1-Group 2	3	n/a	0.200

Steps for a Theoretical Hypothesis Test comparing 2 proportions

Since the null is $p_1 = p_2$, no difference in the groups, we combine both groups together to get $\hat{p}_c$, the pooled or combined sample proportion:

$$\hat{p}_c = \frac{\text{number of successes}}{\text{number of cases}} = \frac{x_1 + x_2}{n_1 + n_2} \text{ or } \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

For a theoretical hypothesis test we get our p-value from the Normal distribution, so the conditions have to be met.

Steps for a Hypothesis Test with the Normal Distribution (Z Test, Difference of Proportions)

a. Write the null and alternate hypotheses.

b. Calculate the pooled/combined $\hat{p}_C$. Check the conditions to use the Central Limit Theorem (below).

c. Write the observed statistic and calculate the test statistic, $Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}_C\hat{q}_C\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$.

   Use a normal calculator with $Z \sim N(0,1)$ to find the p-value.

d. Compare the p-value with the significance level, α, and determine whether the result is statistically significant.

e. State the conclusion in context, including the p-value and whether we reject or fail to reject the null hypothesis.

Conditions

1. **Independence**: The individuals in each group and between groups must be independent of each other regarding the variable measured.
2. **Randomization**: The data are from 2 independent random samples or 2 randomly assigned groups.
3. **Large Population**: If sampling without replacement, the population must be 10 times the size of each group.
4. **Success/Failure Condition**: You should expect to have at least 10 successes and 10 failures in your data under the null hypothesis (no difference) to ensure a normal distribution. Check that $n_1\hat{p}_C$ and $n_1\hat{q}_C \geq 10$ and $n_2\hat{p}_C$ and $n_2\hat{q}_C \geq 10$ using the pooled/combined $\hat{p}_C$.

Let's go through the theoretical test for the dolphin therapy experiment. We will also check the hypothesis test and confidence interval using GeoGebra and see if the hypothesis test and confidence interval have the same conclusion.

Conditions Worksheet for 2 Proportions

For the practice problem, identify each of these elements and use them to write out the conditions.

Subjects and groups	
Variable	
What factors might influence the variable for these subjects/groups?	
Population and Randomization: Is there a single population with 2 randomly assigned treatments or 2 populations with independent random samples?	
Sample sizes	
CI: Number of successes and failures HT: $\hat{p}_C$	

Conditions for the difference of 2 proportions

1. Independence: The ***subjects*** in each group and between groups must be independent of each other in terms of the ***variable***. List some ***factors*** to consider when choosing a sampling method.

2. Randomization: State whether we have 2 independent random samples or 2 randomly assigned groups.

3. Large Population: The ***population*** must be at least 10x$\boldsymbol{n_1}$= _____and 10x$\boldsymbol{n_2}$= _____, or ***population 1*** must be at least 10x$\boldsymbol{n_1}$= _____and ***population 2*** must be at least 10x$\boldsymbol{n_2}$= _____.

4. Success/Failure Condition:
   CI: Check that the successes and failures are all greater than 10 (or $n_1\hat{p}_1$ and $n_1\hat{q}_1$ and $n_2\hat{p}_2$ and $n_2\hat{q}_2 \geq 10$).
   HT: Check that $n_1\hat{p}_C$ and $n_1\hat{q}_C$ and $n_2\hat{p}_C$ and $n_2\hat{q}_C \geq 10$.

Lesson 6 Practice Problem 1: Percentages Given Instead of Counts

Test the claim that people with a specific genetic marker are more likely to have clinical depression than people without the genetic marker. In a sample of 42 people with the genetic marker, 38% have had clinical depression. In a sample of 758 people without the genetic marker, 12% have had clinical depression. Use $\alpha = .05$. Check on StatKey and GeoGebra and compare with a confidence interval.

Lesson 6 Practice Problem 2

Participants in a study to evaluate the accuracy of lie detectors were divided into two groups, with one group reading true material and the other group reading false material, while connected to a lie detector. Both groups received electric shocks to add stress. The two-way table indicates whether the participants were lying or telling the truth and whether the lie detector indicated they were lying or not.

Test whether there is a difference in the proportion of times the lie detector says the person is lying, depending on whether they are lying or telling the truth. Use $\alpha = .05$. Check on StatKey and GeoGebra and compare with a confidence interval.

	Detector says lying	Detector says not lying	Total
Person reading false material	31	17	48
Person reading true material	27	21	48
Total	58	38	96

Lesson 7: Goodness of Fit for 1 Categorical Variable

How can we test whether a sample is representative of a population? Let's look at an example of race and ethnicity in jury pools. The American Civil Liberties Union (ACLU) conducted an analysis of the jury pools for a sample of 10 trials in Alameda County, California.

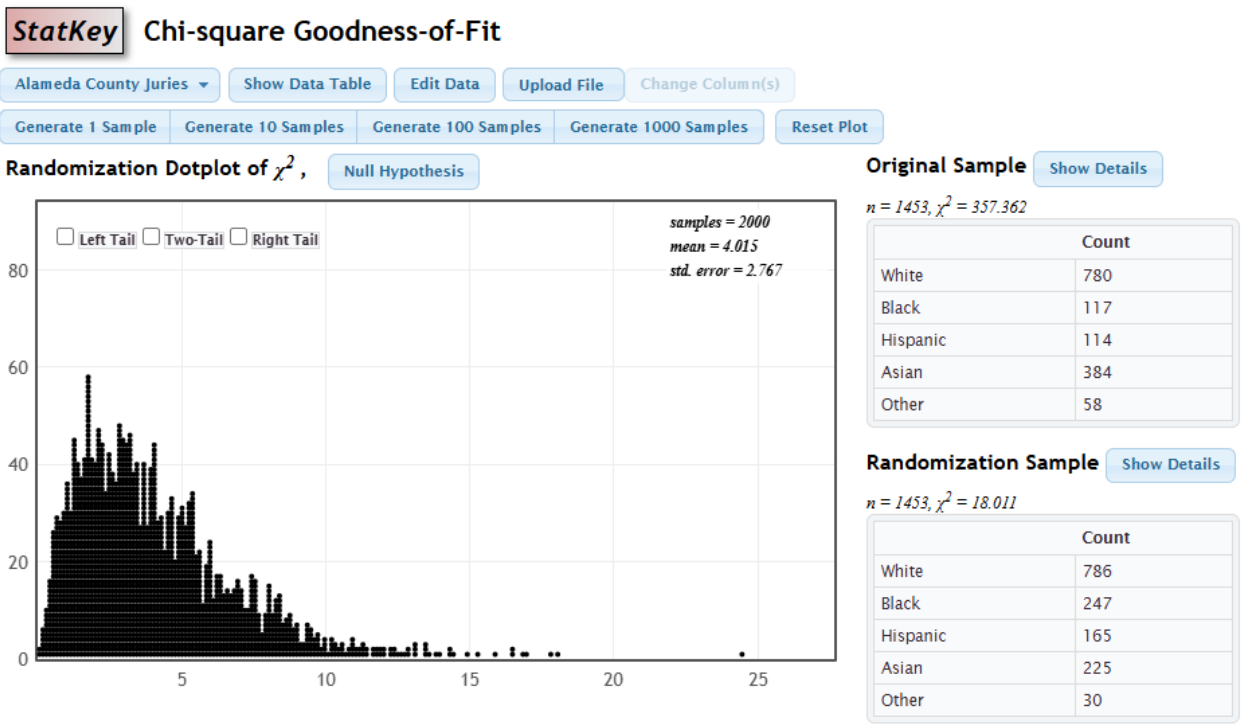
The frequency table below gives the racial makeup of the 1453 people in the sample, and the census demographics for all eligible jurors in the county.

Race/Ethnicity	Asian	Black	Hispanic	White	Other	Total
Number in jury pools	384	117	114	780	58	1453
All eligible jurors from the census data	15%	18%	12%	54%	1%	100%

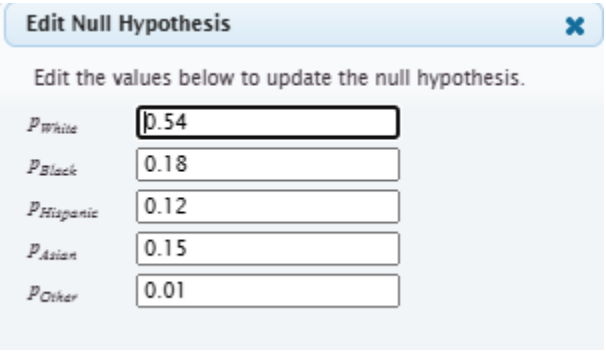
How do we show evidence that the jury pools either are or are not representative of the population? Here we have a single categorical variable, with 5 proportions. The null hypothesis can be that the proportions match a certain distribution like in this situation, or that they are all equal or uniform. For example, $H_0: p_i = .20$.

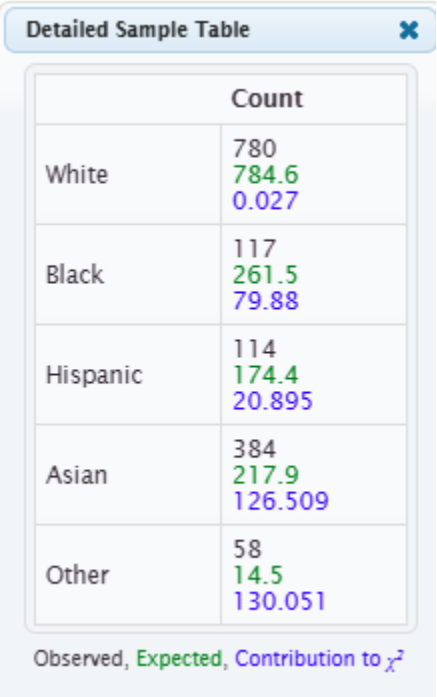
Write the hypotheses for this test:

Let's simulate this situation in StatKey. Select Goodness of Fit Test at the bottom and select the Alameda County Juries data set. Click on Edit Data and you should see the same data from the table above. Click on the Null Hypothesis to verify that it also matches the census distribution. We will generate one sample at a time to see what is happening and then generate 2000 more.



Click on Show Details for the Original Sample to see the Expected Counts and the Contribution to the Chi-squared statistic.





Expected Counts and Differences

Race/Ethnicity of Eligible Jurors	Asian (15%)	Black (18%)	Hispanic (12%)	White (54%)	Other (1%)
Observed in the sample out of 1453	384	117	114	780	58
Expected count in a sample of 1453					

$$\mathcal{X}^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} =$$

Entering the Data in GeoGebra

Distribution | Statistics

Goodness of Fit Test

Rows 5

☐ Column %

	Observed Count	Expected Count
Asian	384	217.95
Black	117	261.54
Hispanic	114	174.36
White	780	784.62
Other	58	14.53
	1453	1453

Result

Goodness of Fit Test

df	4
X^2	357.3625
p	0

Conclusion:

Finding $\mathcal{X}^2$ Probabilities with Stapplet:

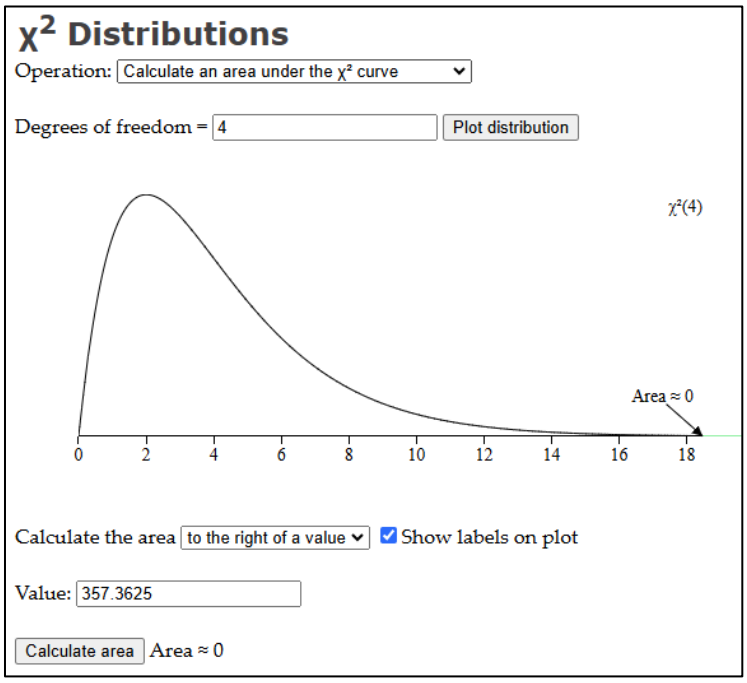
[Stapplet Chi-Squared Distribution]

In the Stapplet $\mathcal{X}^2$ Calculator, make sure the Operation Menu is set to:
Calculate an area under the Curve.

Type in the degrees of freedom and click Plot Distribution.

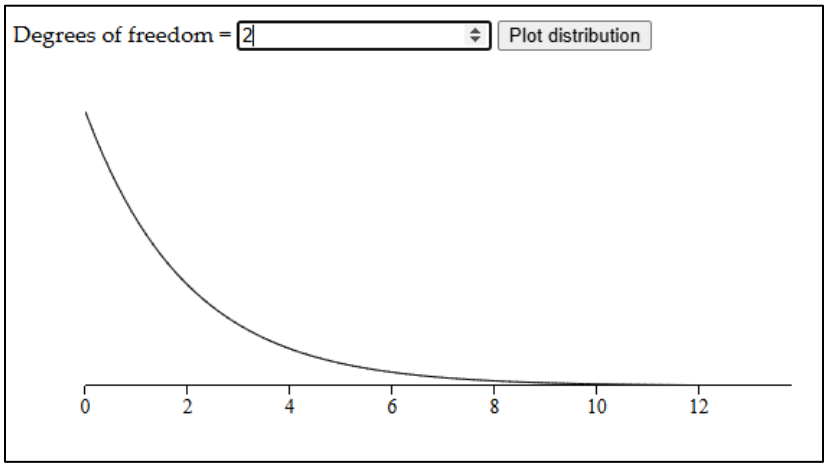
Then select the type of probability you want: This is always a right tail test.

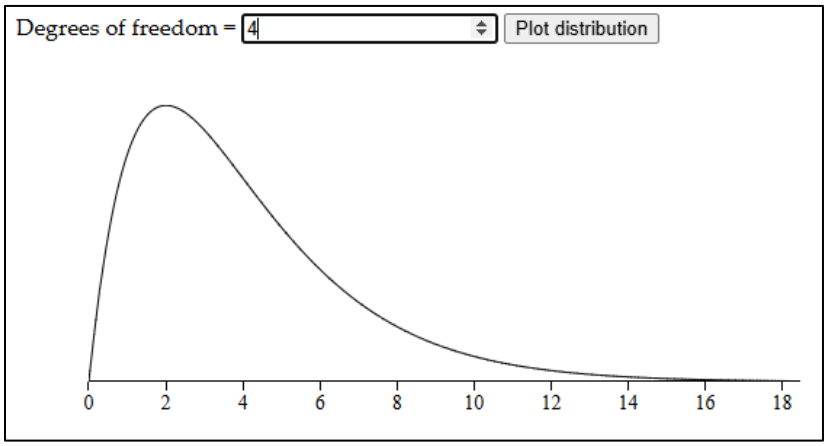
You can use the link at the bottom to switch between percentages and proportions and change the number of decimal places given.

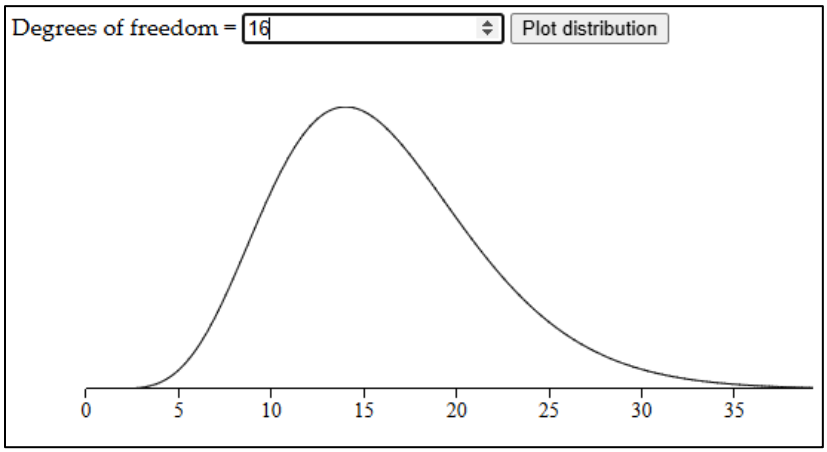


The X^2 Distribution $X \sim X^2(df)$

This distribution has only one input: degrees of freedom. This gives us a sense of how many categories there are. The mean = df and the mode is at $df - 2$.







Steps for a Goodness of Fit Test

The null is that the counts match the given distribution. Our test statistic is a measure of how far the counts are off from the expected, relative to each expected group size.

Steps for a Goodness of Fit Hypothesis Test with the Chi-Squared Distribution (Goodness of Fit Test)

a. Write the null and alternate hypotheses. The null is uniform or matching a distribution. The alternate is at least one proportion is different.

b. Check the conditions to use this test (below).

Calculate the test statistic, $\mathcal{X}^2 = \sum \frac{(\text{observed}-\text{expected})^2}{\text{expected}}$, and the degrees of freedom,

$df = \#\text{ of categories} - 1.$

Use a Chi-squared calculator to find the p-value. This is always a right-tail test.

c. Compare the p-value with the significance level, α, and determine whether the result is statistically significant.

d. State the conclusion in context, including the p-value and whether we reject or fail to reject the null hypothesis.

Conditions

1. **Counted Data:** The data must be counts for the categories of a categorical variable.
2. **Independence**: The individuals counted must be independent of each other regarding the variable measured.
3. **Randomization**: The data are from a random sample.
4. **Large Population**: If sampling without replacement, the population must be 10 times the size of the sample.
5. **Expected Counts:** The expected counts should all be at least 5.

Standardized Residuals

When we reject the null hypothesis, we still don't know which category or categories were off. In order to find out, we can calculate and compare the standardized residuals.

$$c = \frac{\text{observed} - \text{expected}}{\sqrt{\text{expected}}}$$

Race/Ethnicity	Asian	Black	Hispanic	White	Other
Standard Residual					

Lesson 7 Practice Problem: Uniform Hypothesis

Many standardized tests use multiple choice questions. For example, the Advanced Placement (AP) exams for various subjects frequently use multiple choice questions for a portion of the exam, with five options, labeled A, B, C, D, and E. Are certain choices more or less likely to occur than others as correct answers? The table below shows the frequencies of the letters for correct answers in a sample of 400 questions selected from previously released AP exams. Perform a goodness of fit test that the distribution is uniform, using $\alpha = .05$. Check your results on StatKey and GeoGebra.

Answer	A	B	C	D	E	Total
Frequency	85	90	79	78	68	400

Calculate the Standard Residuals

Answer	A	B	C	D	E
Standard Residual					

Lesson 8: Tests of Homogeneity and Independence

3 Tests for Multiple Proportions: The null is always independent/no association or the same distribution.

- **Goodness of Fit** – Compares one categorical variable with a claimed distribution (could be uniform).
- **Homogeneity (Sameness)** – Compares one categorical variable for several subpopulations, usually a random sample from each subpopulation.
- **Independence** – Compares two categorical variables to see whether the variables are independent or not independent, usually from a single random sample and sorted into all the categories.

Determine which of the 3 tests should be used

a. We want to see if there is a relationship between gender and health care coverage (covered/not covered).

b. The program dean wants to compare the final grade distributions for three different instructors to see if they are the same.

c. A trading card company claims that 5% of cards are rare, 20% are uncommon, and 75% are common. We want to check this claim.

Example

A random sample of adults in the U.S. was taken in March 2012, and each was asked whether they approved or disapproved of the job being done by President Obama, Democrats in Congress, and Republicans in Congress. The results are shown in the table. We would like to determine if the three groups and the approval ratings are associated.

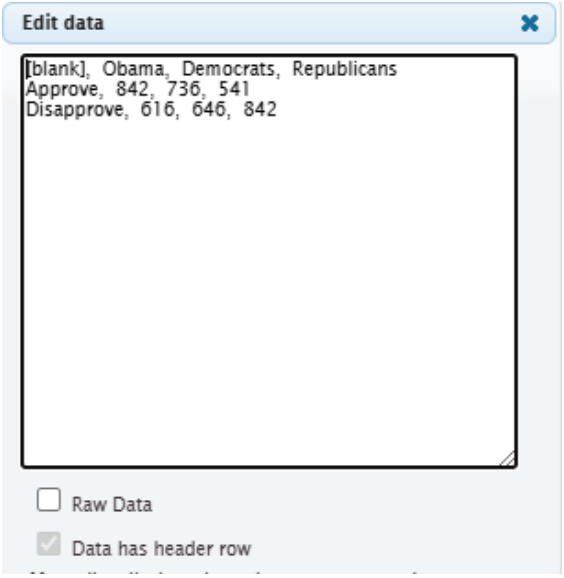
	President Obama	Democrats in Congress	Republicans in Congress	Total
Approve	842	736	541	2119
Disapprove	616	646	842	2104
Total	1458	1382	1383	4223

Which type of test is this? Write the hypotheses for the test:

Let's first use StatKey to calculate the expected counts for each cell in the table, the $\mathcal{X}^2$ test statistic and the contribution from each cell. Select $\mathcal{X}^2$ Tests for Association from the menu at the bottom. Notice the null hypothesis is set at no association. Select Edit Data, and enter the data like this or copy/paste:

[blank], Obama, Democrats, Republicans
Approve, 842, 736, 541
Disapprove, 616, 646, 842

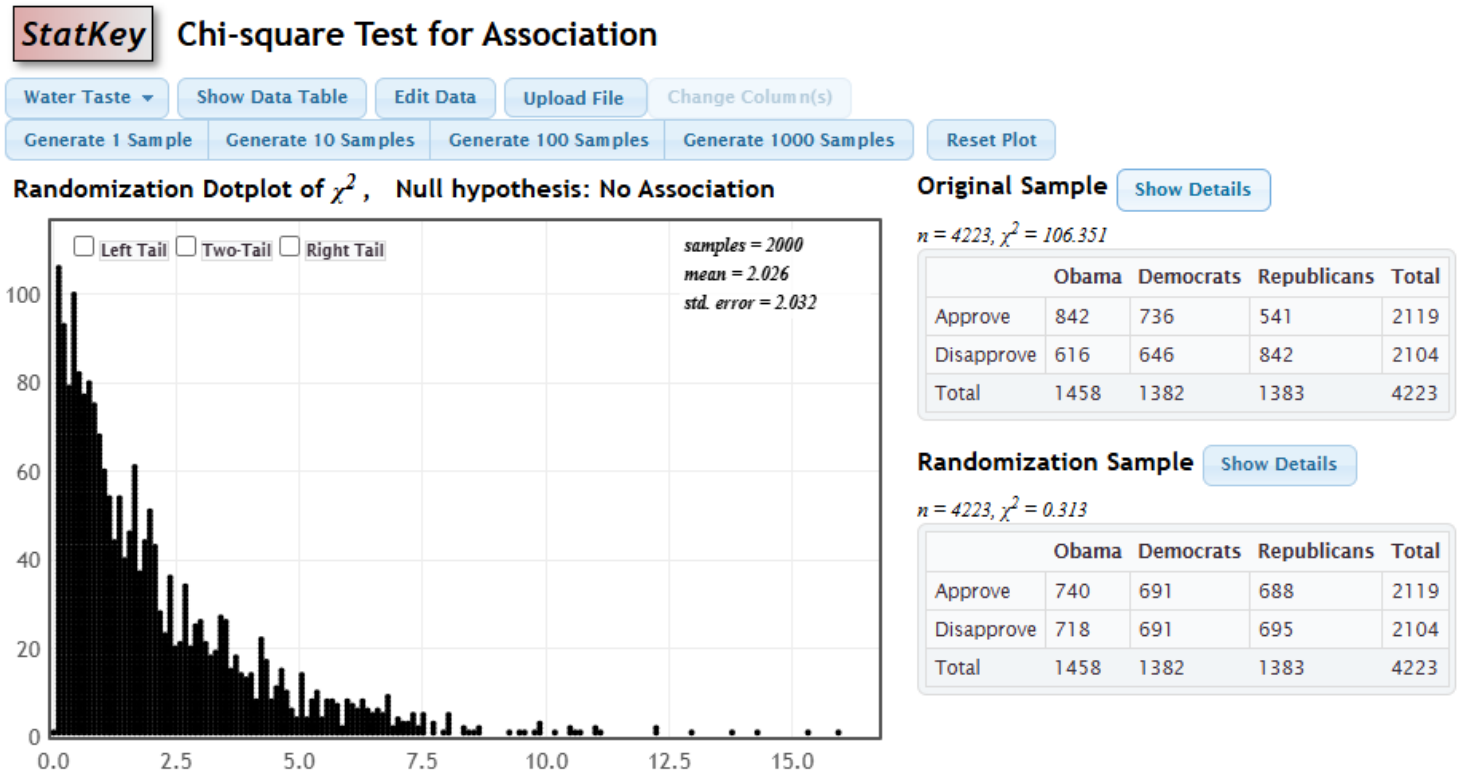
Don't enter the totals or it will think that is another category. Check that the totals match after you hit ok.



Detailed Sample Table

	Obama	Democrats	Republicans	Total
Approve	842 731.6 16.663	736 693.5 2.61	541 694 33.713	2119
Disapprove	616 726.4 16.782	646 688.5 2.629	842 689 33.954	2104
Total	1458	1382	1383	4223

Observed, Expected, Contribution to χ^2



Find the Expected Counts and Differences

$$\text{Expected} = \frac{\text{row total} \cdot \text{column total}}{\text{grand total}}$$

	President Obama	Democrats in Congress	Republicans in Congress	Total
Approve				2119
Disapprove				2104
Total	1458	1382	1383	4223

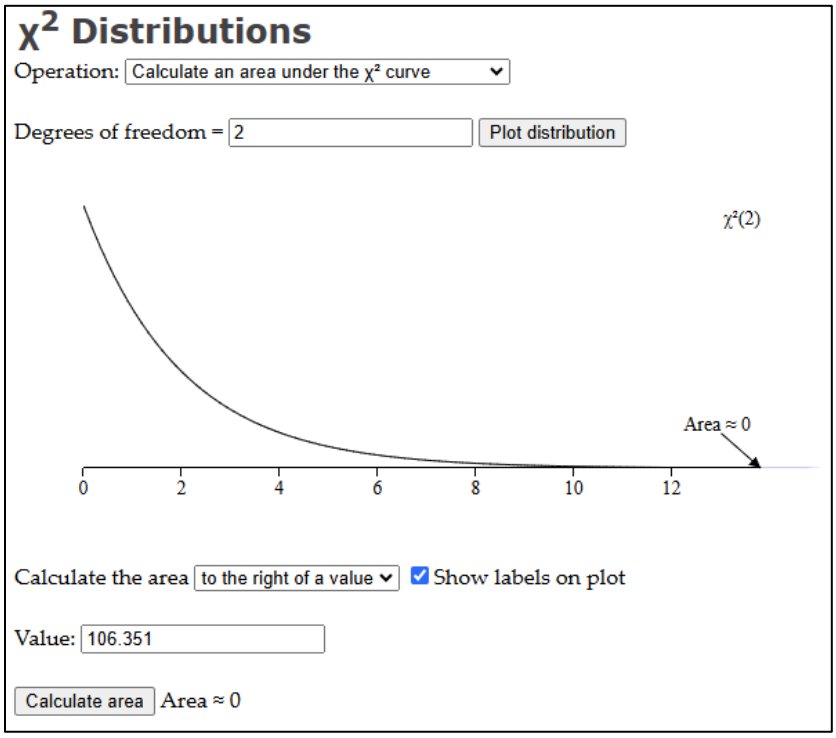
$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} =$$

$$df = (\#\, rows - 1)(\#\, cols - 1) =$$

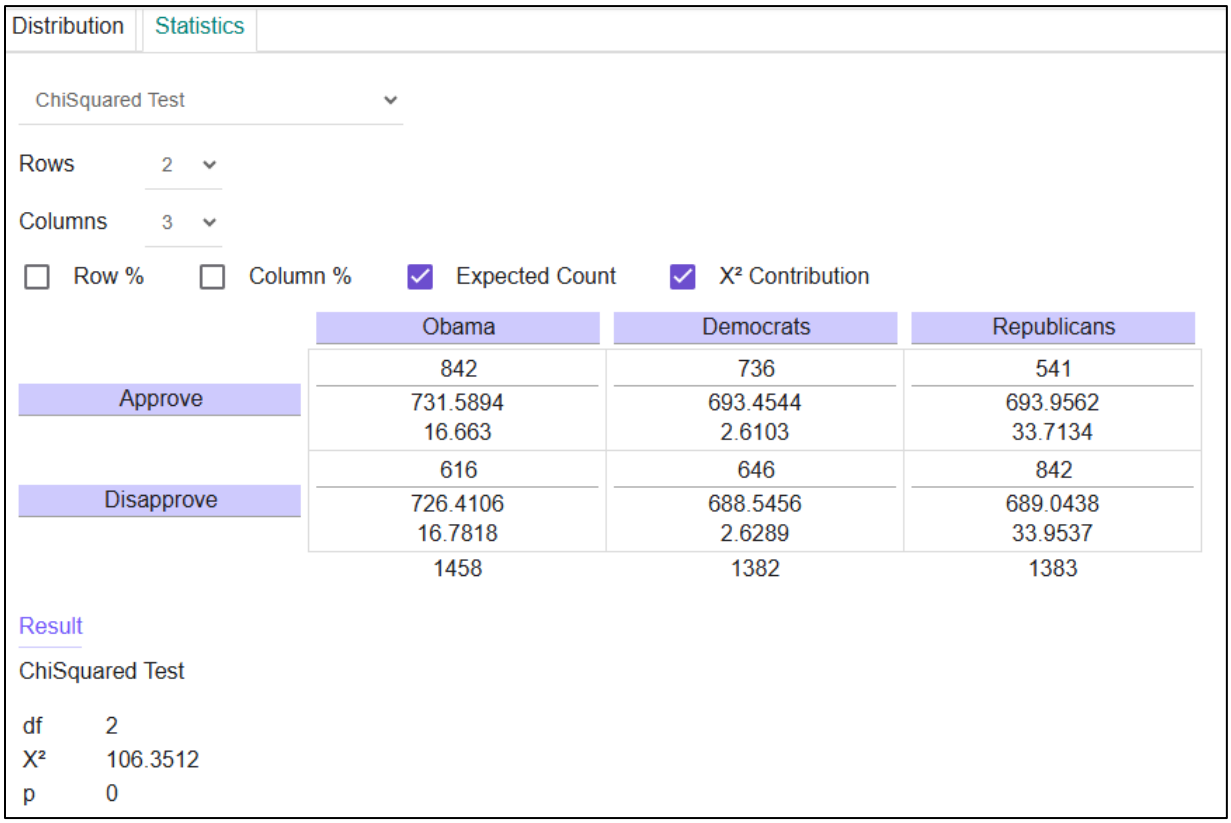
Find the p-value and write the conclusion for this test:

Finding $\mathcal{X}^2$ Probabilities with Stapplet:

[Stapplet Chi-Squared Distribution]



Entering the Data in GeoGebra



Steps for Tests of Homogeneity and Independence

The test statistic is a measure of how far the counts are off from the expected, relative to each expected group size.

Steps for Tests of Homogeneity and Independence (ChiSquared Test)

a. Write the hypotheses. The null is always independent/no association or the same distribution.

b. Check the conditions to use this test (below).

Calculate the test statistic, $\mathcal{X}^2 = \sum \frac{(\text{observed}-\text{expected})^2}{\text{expected}}$, and the degrees of freedom,

$df = (\#\, rows - 1)(\#\, cols - 1)$.

Use a Chi-squared calculator to find the p-value. This is always a right-tail test.

c. Compare the p-value with the significance level, α, and determine whether the result is statistically significant.

d. State the conclusion in context, including the p-value and whether we reject or fail to reject the null hypothesis.

Conditions

1. **Counted Data:** The data must be counts for the categories of one or two categorical variables.
2. **Independence**: The individuals counted must be independent of each other regarding the variable(s) measured.
3. **Randomization**: The data are from a single random sample or multiple random samples.
4. **Large Population**: If sampling without replacement, the population must be 10 times the size of each sample.
5. **Expected Counts:** The expected counts should all be at least 5.

Standardized Residuals

When we reject the null hypothesis, we still don't know which category or categories were off. In order to find out, we can calculate and compare the standardized residuals.

$$c = \frac{\text{observed} - \text{expected}}{\sqrt{\text{expected}}}$$

	Obama	Democrats	Republicans
Approve			
Disapprove			

Notes on Tests of Independence and Homogeneity

- Be careful with the word dependent, it might be confused with causation. Better to say not independent or associated.
- Beware of large sample sizes – there will always be some differences and large sample sizes can make the chi-squared statistic large. The degrees of freedom don't increase for sample size.
- For a 2x2 table, use the 2-proportion method. There is a relationship between the two methods. See the example below.

Lesson 8 Practice Problem 1: Comparison with the 2-proportion method for a 2x2 table

Use theoretical calculators to compare a 2-proportion test with a $\mathcal{X}^2$ test for independence for the lie detector test in Lesson 6. What do you notice? Compare the conditions as well.

Test whether there is a difference in the proportion of times the lie detector says the person is lying, depending on whether they are lying or telling the truth. Use $\alpha = .05$.

	Detector says lying	Detector says not lying	Total
Person lying	31	17	48
Person not lying	27	21	48
Total	58	38	96

Lesson 8 Practice Problem 2

The National Sleep Foundation conducted a survey on the sleep habits of randomly sampled groups of transportation workers and non-transportation workers. A summary of their results is in the table below. Determine whether the data indicate that the amount of sleep received is distributed in the same way for each profession at the 5% significance level.

	Type of professional					
Sleep	Non-transportation	Pilot	Truck driver	Train operator	Bus/taxi/limo driver	Total
Less than 6 hours	35	19	35	29	21	139
6 to 8 hours	193	132	117	119	131	692
Greater than 8 hours	64	51	51	32	58	256
Total	292	202	203	180	210	1087

Lesson 9: Simulated and Theoretical Confidence Intervals for a Mean

How many credits do Stats 244 students at PCC take? A fictional random sample of 15 students is shown below. Let's explore this data using StatKey and create a confidence interval using bootstrapping.

Credit Hours
12
18
13
15
15
14
18
12
8
10
16
15
16
10
21

Notation Refresher for Means

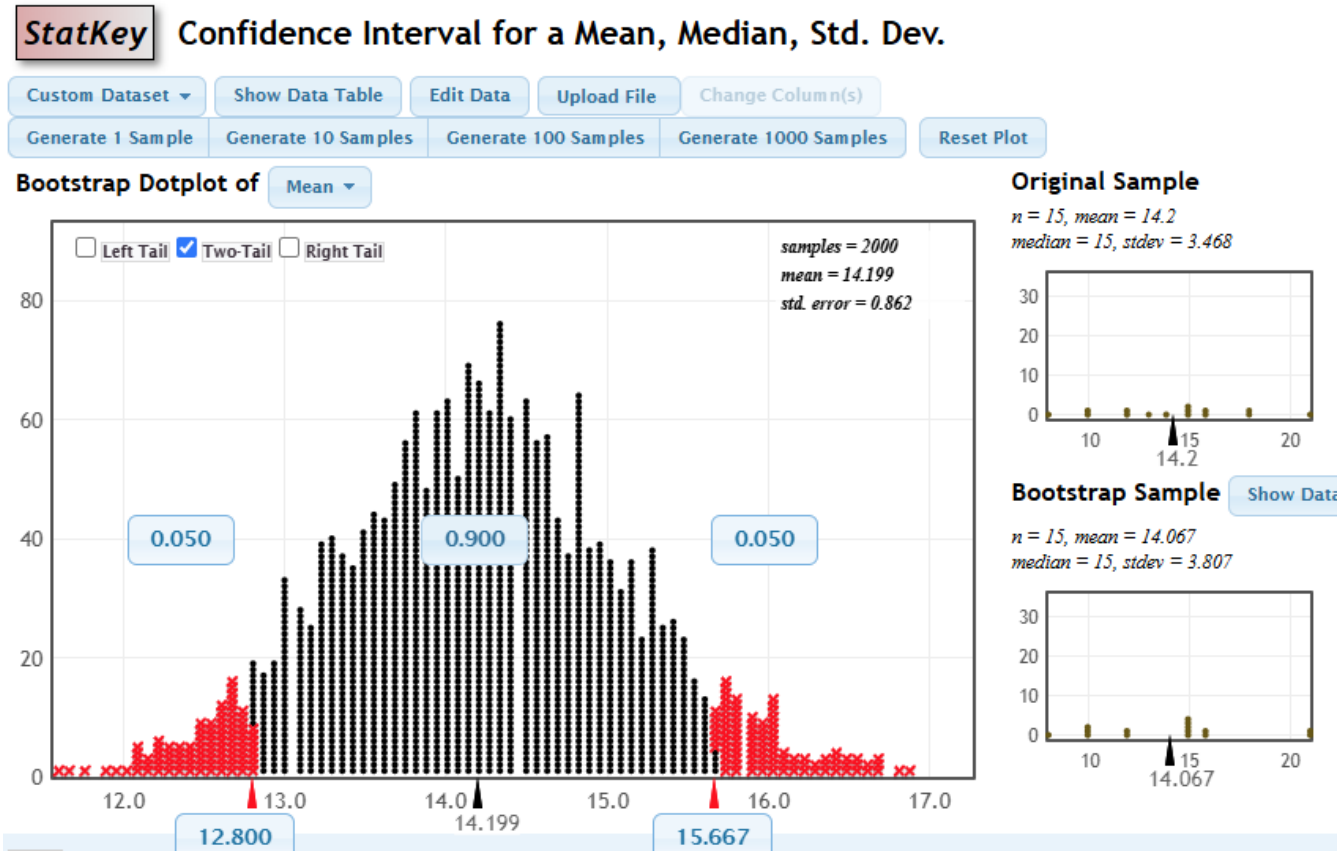
Numerical variables almost always have units, so remember to write them after each statistic.

- Population mean: μ
- Sample mean: $\bar{x}$
- Population standard deviation: σ
- Sample standard deviation: s

What is the best point estimate for the number of credits Stat 244 students are taking?

Write and interpret the 90% bootstrap confidence interval for the true population mean for credit hours that PCC Stat 244 students take.

Bootstrap Simulated Confidence Interval for a Mean



Sampling Distribution for a Mean - Conditions and Model

We have seen in our simulations that the distribution of a statistic for many random samples of the same size from the same population is close to a normal model. We've also seen that the mean stays the same but the standard error gets smaller the larger the sample size. Let's summarize the Central Limit Theorem for the sampling distributions of a mean.

The Central Limit Theorem Conditions and Model for a Mean, $\bar{x}$

When using a theoretical model, we need to check the conditions or assumptions to ensure the model applies to the situation.

1. **Independence**: The individuals or items must be independent of each other regarding the variable measured.
2. **Randomization**: The samples need to be randomly chosen, or it's not safe to assume independence.
3. **Large Population**: If sampling without replacement, we must be sampling from a large population to consider the individuals independent of each other. Sometimes this is called the 10% Condition or said that the population must be at least 10 times our sample size.
4. **Sample Size:** The population is normal, or the sample is nearly normal, or $n \geq 30$. If there are outliers or a strong skew we cannot use this method (check the histogram and boxplot).

When the conditions are met, the sampling distribution for a mean, is modeled by a normal distribution with the following parameters:

$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Note: For means there are 2 unknowns: the population mean and standard deviation, which increases the variation. If we know the population standard deviation or have a large sample size, we can use the normal distribution, but this is unrealistic so we will always use the t-distribution in this class. The t-distribution is also called Student's t-distribution and works better for small sample sizes.

Steps for a Theoretical Confidence Interval for a Mean (T Estimate of a Mean)

a. Calculate the point estimate and find the critical value, t^*, for the confidence level, using $df = n - 1$.

b. Check the conditions to use the Central Limit Theorem: Independence, Randomization, Large population, and Sample size.

c. Calculate the interval using the formula:

$$\bar{x} \pm t^*_{n-1} \frac{s}{\sqrt{n}}$$

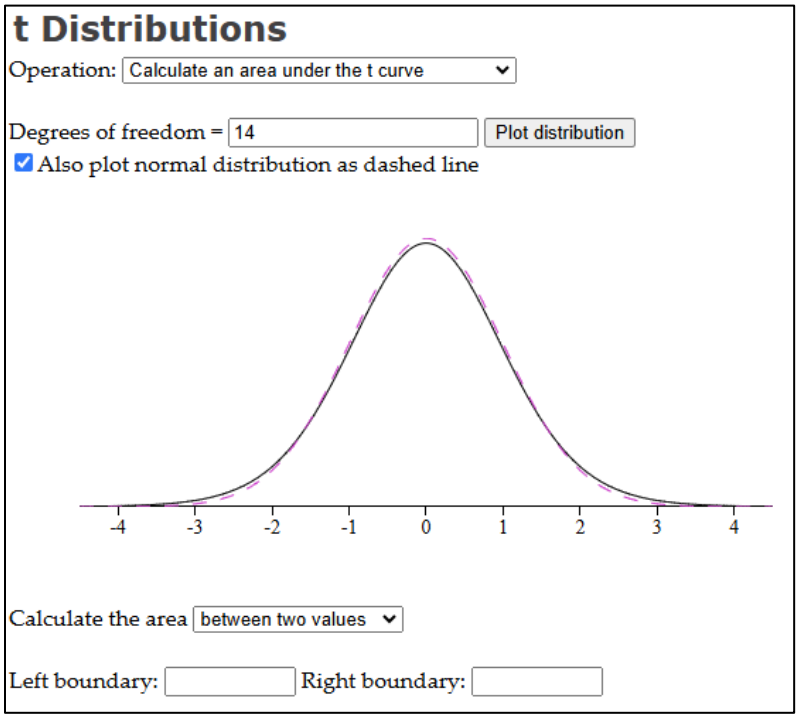
d. Write the interpretation in context.

Note: the standard error is given by $SE = \frac{s}{\sqrt{n}}$ and the margin of error is given by $ME = t^*_{n-1} \frac{s}{\sqrt{n}}$.

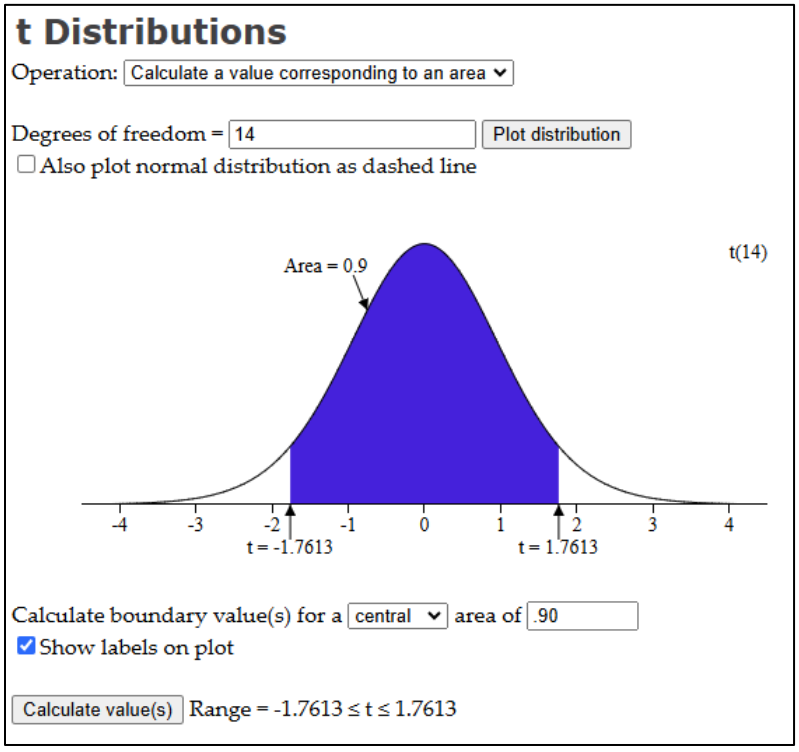
Why do we need a new distribution? Video: The Brewer Who Secretly Revolutionized Statistics

[Stapplet t-Distribution]

Finding an Area



Finding a Critical t* Value – We have to find this for each sample size and confidence level

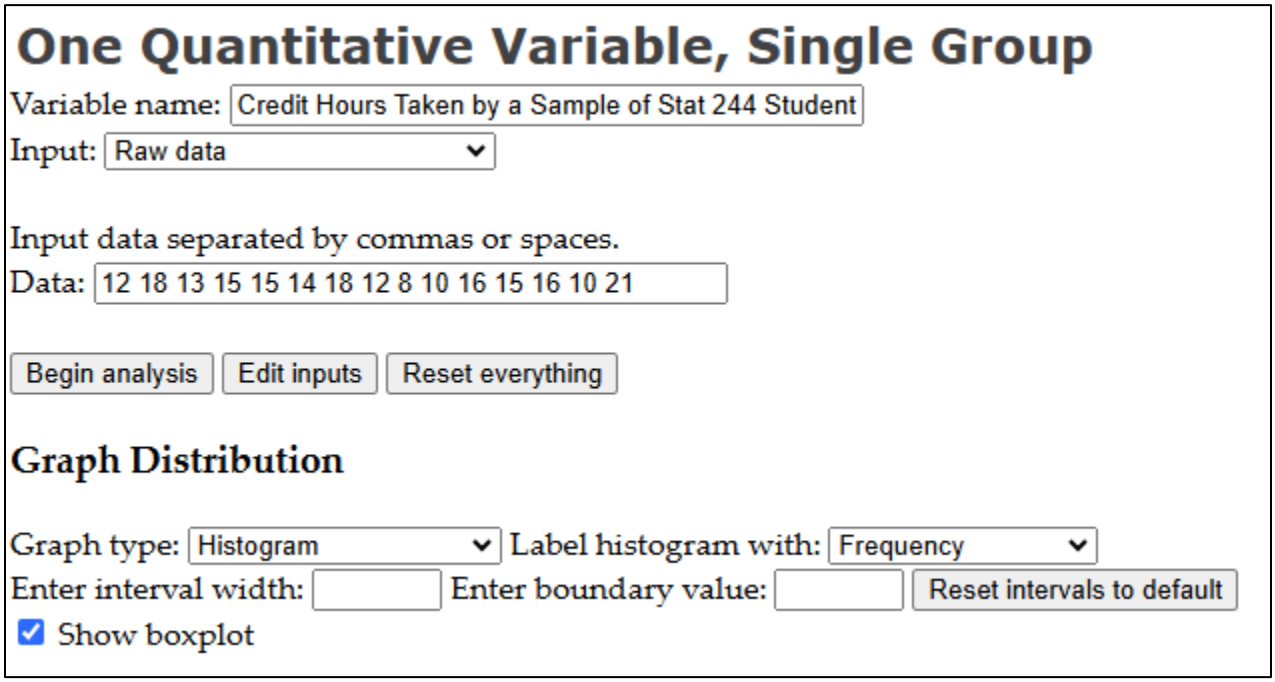


Let's continue with our credit hour data. Check the conditions needed to use the Central Limit Theorem.

How to check for a nearly normal sample? Stapplet Descriptive Statistics (1 quantitative, single group)

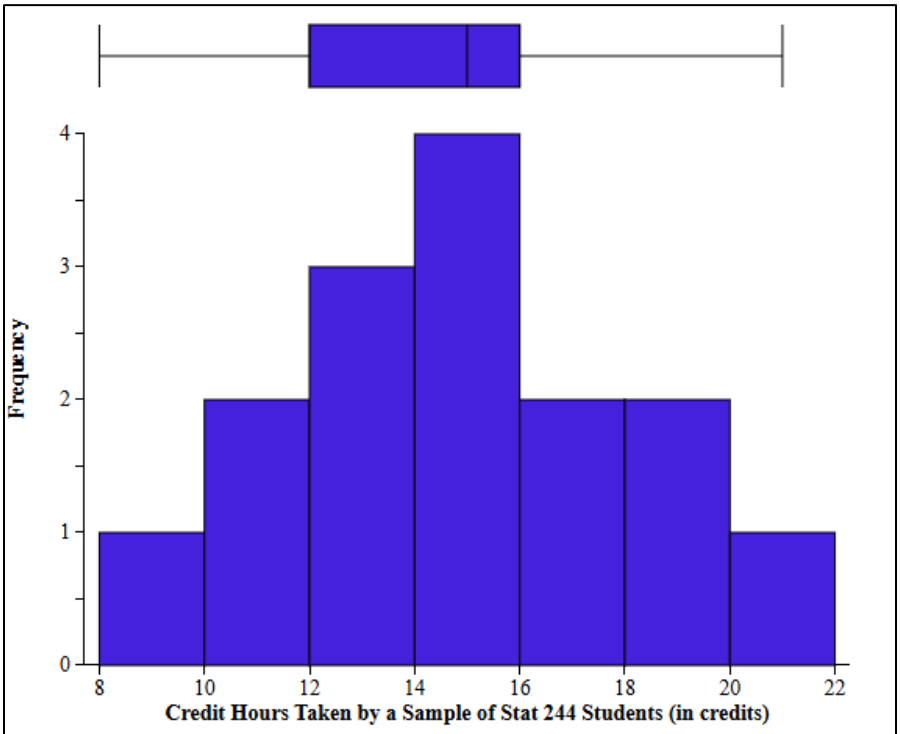
Make a stacked histogram and boxplot of the data and calculate the summary statistics.

Stapplet 1 Quantitative, Single Group



What is the shape of the data?

How can you use the mean and median to decide or confirm the shape?



Summary Statistics

n	mean	SD	min	Q_1	med	Q_3	max
15	14.2	3.4682	8	12	15	16	21

Let's check our class sample – does it meet the nearly normal condition?

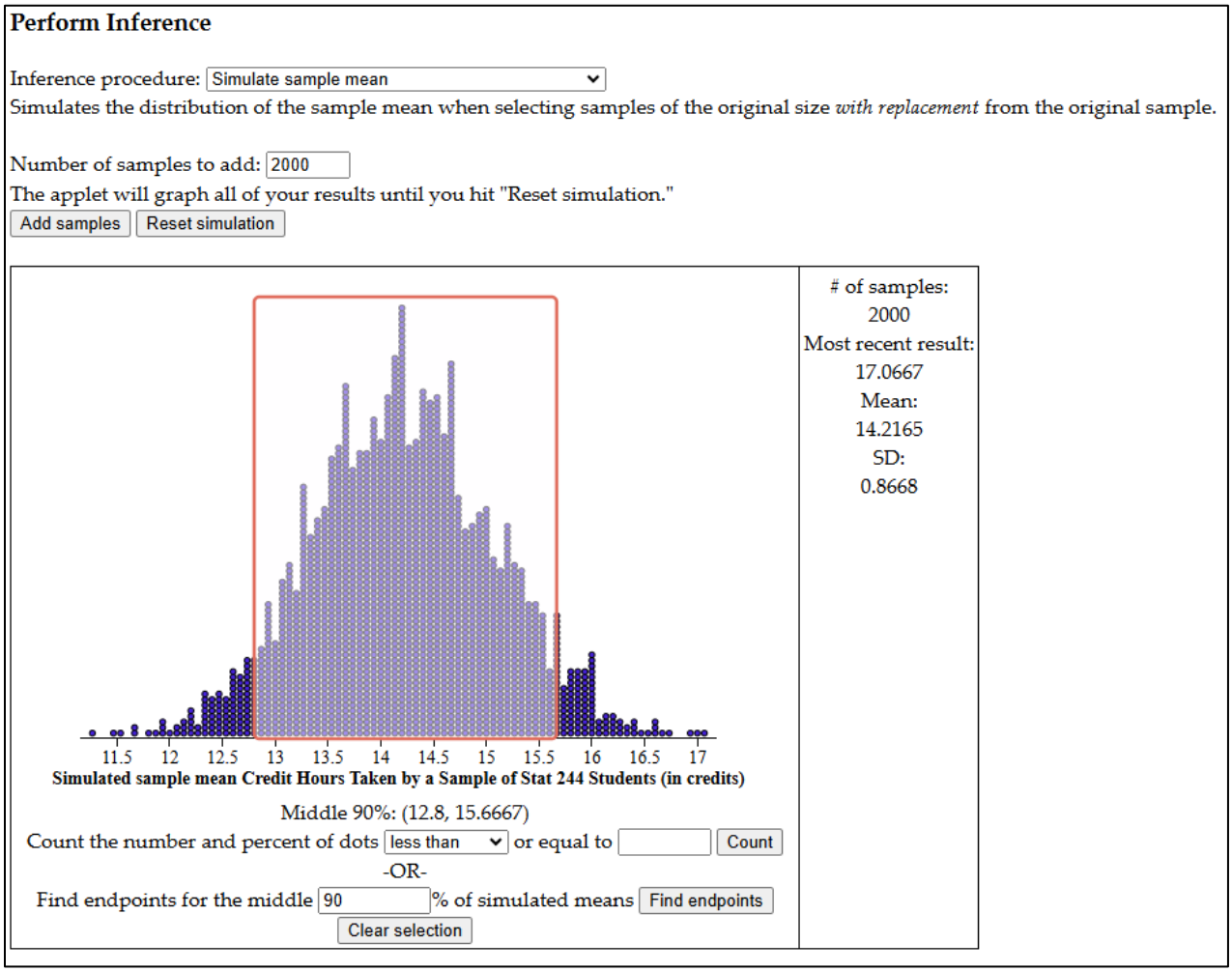
Now we can calculate our theoretical 90% confidence interval for the original data. We can also calculate one using our class data if it is nearly normal.

Bootstrapping and Theoretical Calculator in Stapplet

Stapplet also does inference and simulation. Since our data are already in there, we can continue.

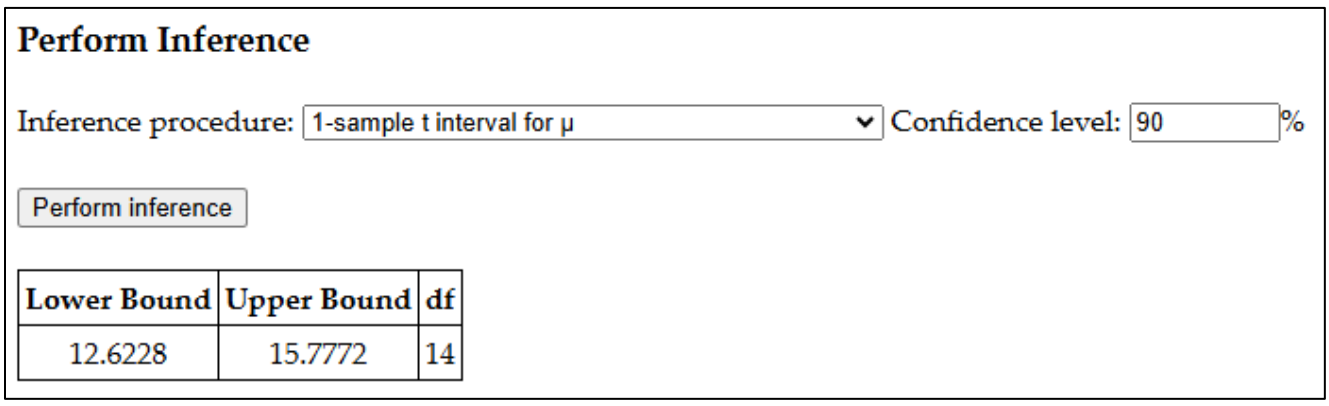
Use the same applet – keep scrolling down to **Perform Inference**. Select **Simulate sample mean**.

Enter 2000 samples and click Add samples. Enter the confidence level and select Find endpoints



Theoretical Confidence Interval

Select 1-sample t interval for μ and enter the confidence level.



GeoGebra has inputs for the mean and standard deviation, but it doesn't take raw data.

- If a problem gives you the sample mean and standard deviation, use GeoGebra or Stapplet (select mean and standard deviation instead of raw data)
- If a problem gives you raw data, use Stapplet

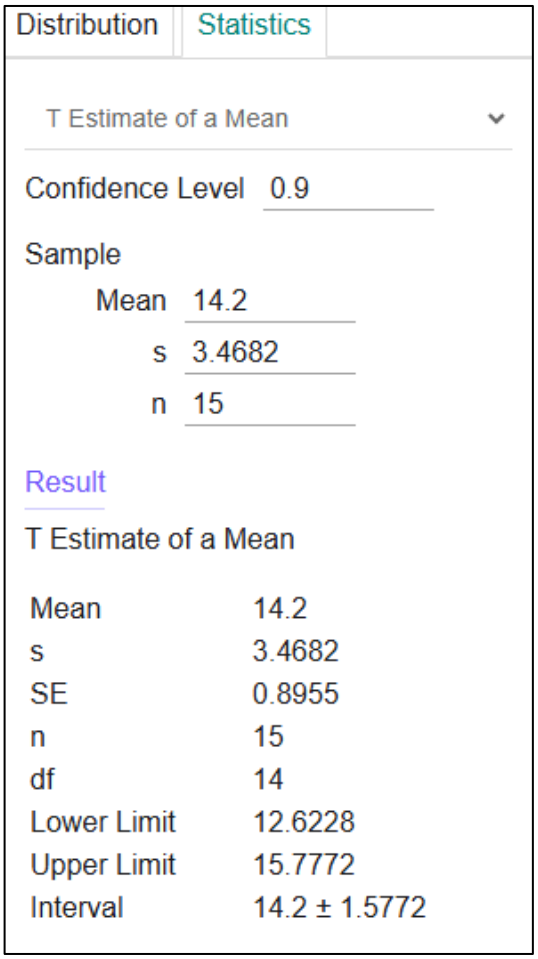
GeoGebra Theoretical Calculator

GeoGebra Classic Probability Calculator

Select the Statistics Tab.

For a confidence interval of a mean, select **T Estimate of a mean** and enter your confidence level as a decimal.

Enter the sample mean, standard deviation and sample size and the result appears below.



Finding an Ideal Sample Size

Just like we did with proportions, when given a desired margin of error, ME, we can find the sample size, n, by solving the margin of error equation for n. If we do this for ME = $t^*_{n-1}\frac{s}{\sqrt{n}}$, we get $n = \left(\frac{t^*_{n-1}\cdot s}{\text{ME}}\right)^2$.

But there's a problem. How do we know what t^*_{n-1} is without knowing what n is? The degrees of freedom depend on n. We could use an iterative process, but for $n \geq 60$, z^* is a good substitute.

So we will use this formula, using the same z^* values that we used for proportions.

$$n = \left(\frac{z^* s}{\text{ME}}\right)^2$$

Sample size must be a whole number. This is a minimum value, so we always round up, even if the decimal is under 0.5. If we were to round down that would make the margin of error larger than the desired value.

Continuing the example above, calculate the sample size needed to reduce the margin of error to 0.5 credits at 90% confidence. Note, the ME is not usually a percentage now, so we don't move the decimal place. (If our numerical data are proportions, we may need to move the decimal place.)

Lesson 9 Practice Problem 1 (With only a mean and SD given)

A nursing union wants to study pay rates vs. patient costs. To assess patient costs, they take a random sample of 50 patients and find they have an average hospital bill of $3200 per night with a sample standard deviation of $700 per night.

a. Check the conditions, calculate and interpret the theoretical 95% confidence interval for the population mean cost per night at the hospital. Check this on the GeoGebra and/or Stapplet theoretical calculator.
b. Why can't we do a simulation using Stapplet or StatKey in this case?
c. If you were going to do another study with 95% confidence, what sample size would be needed to get a margin of error of $100?

Lesson 9 Practice Problem 2

The data table shows driving speeds from a random sample of cars selected on a California highway at about 3:30 pm on a weekday. The speed limit for the highway is 65 mph. Use the sample data to construct an 80% confidence interval for the mean driving speed. How does this relate to the speed limit?

a. Check the conditions, calculate and interpret the theoretical 80% confidence interval for the population mean traffic speed on the highway. Check this on the Stapplet theoretical calculator.
b. Find a bootstrap 80% confidence interval using Stapplet or StatKey
c. If you were going to collect more data to do another study with 95% confidence, what sample size would be needed to get a margin of error of 1 mph?

Highway Speeds (mph)
62
61
61
57
61
54
59
58
59
69
60
67

Lesson 10: Simulated and Theoretical Hypothesis Testing for a Mean

A tire manufacturer is considering a newly designed tread pattern for its all-weather tires. Tests have indicated that these tires will provide better gas mileage and longer tread life. The last remaining test is for braking effectiveness. The company hopes the tread pattern will allow a car traveling at 60 mph to come to a complete stop within an average of 125 feet after the brakes are applied. They will adopt the new tread pattern unless there is strong evidence that the tires do not meet this objective. The distances (in feet) for 10 stops on a test track are shown in the table to the right.

Stopping Distance (feet)
129
128
130
132
135
123
121
125
128
130

You are the statistical analysis team hired by this company. What hypotheses and level of significance would you use? Test the appropriate hypothesis and state your conclusion. Verify this with the corresponding confidence interval. Would you recommend this company adopt the new tread pattern? Why or why not?

Write the hypotheses for this test.

What would a Type I and Type II error be in this context. Which has worse consequences? Use this to set your significance level.

When would we use simulation vs. a theoretical test in real life?

Theoretical Hypothesis Testing for a Single Mean

We will continue with the case that the population standard deviation is not known, so we will use the t-distribution.

Steps for a Hypothesis Test of a Mean with σ unknown (T Test of a Mean)

a. Write the null and alternate hypotheses.

b. Check the conditions to use the Central Limit Theorem: Independence, Randomization, Large population, and Sample Size.

c. Write the observed statistic, $\bar{x}$, and calculate the test statistic, $t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$.

   Use a t-distribution calculator with $df = n - 1$ to find the p-value.

d. Compare the p-value with the significance level, α, and determine whether the result is statistically significant.

e. State the conclusion in context, including the p-value and whether we reject or fail to reject the null hypothesis.

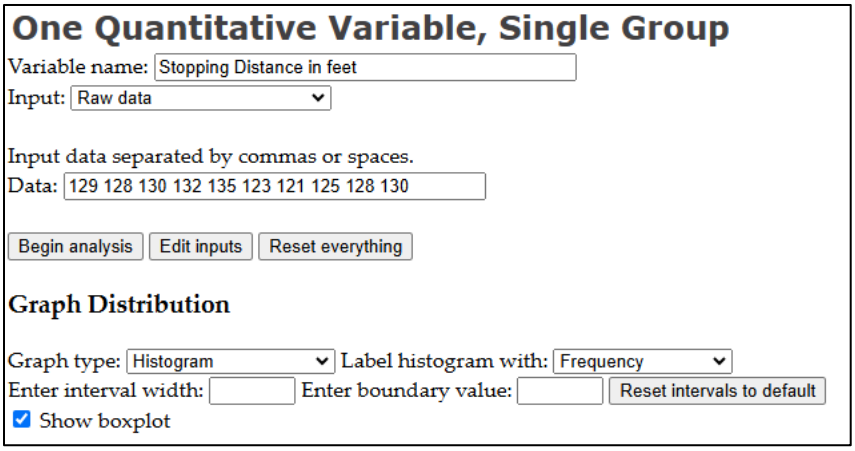
Conditions needed to use the t-distribution for a mean

1. **Independence**: The individuals or items must be independent of each other regarding the variable measured.
2. **Randomization**: The samples need to be randomly chosen, or it's not safe to assume independence.
3. **Large Population**: If sampling without replacement, we must be sampling from a large population to consider the individuals independent of each other. Sometimes this is called the 10% Condition or said that the population must be at least 10 times our sample size.
4. **Sample Size:** The population is normal, or the sample is nearly normal, or $n \geq 30$. If there are outliers or a strong skew we cannot use this method (check the histogram and boxplot).

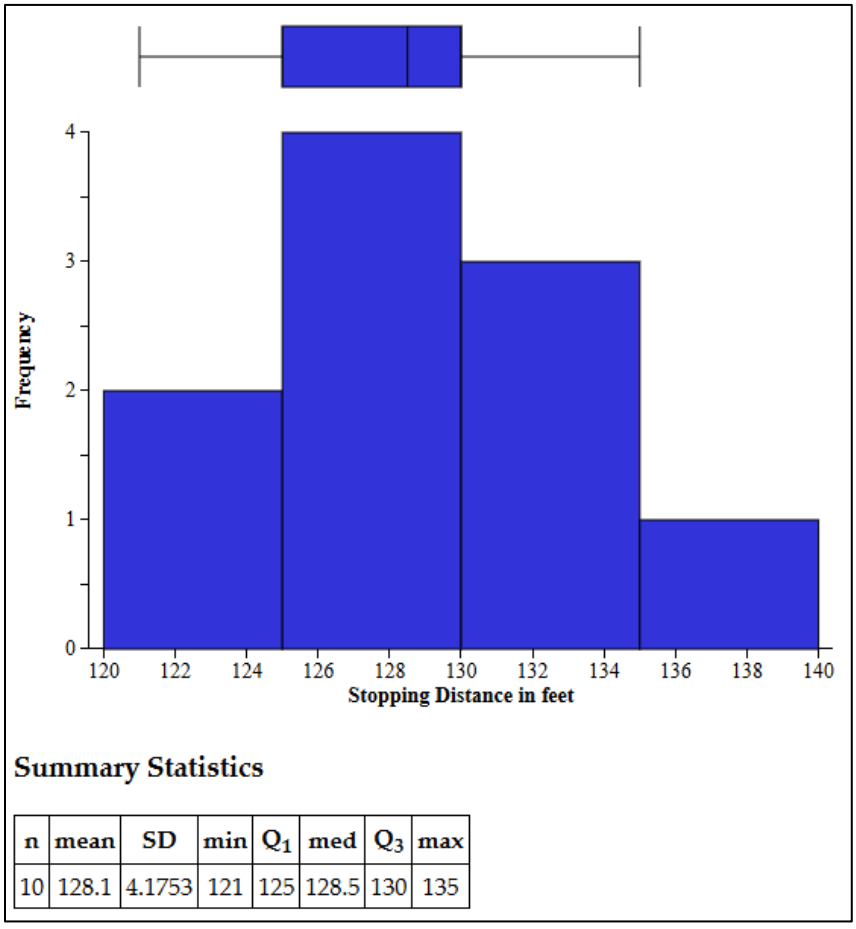
Write out the conditions for a theoretical test to see if they are met.

Checking the Nearly Normal Condition on Stapplet

Enter your variable and units, data, and select histogram and show boxplot



Does the histogram and boxplot show a nearly normal shape? What if you enter an interval width of 1, 2, 3, 4, 5, 6, or 7?



Conduct the theoretical test and write your conclusion.

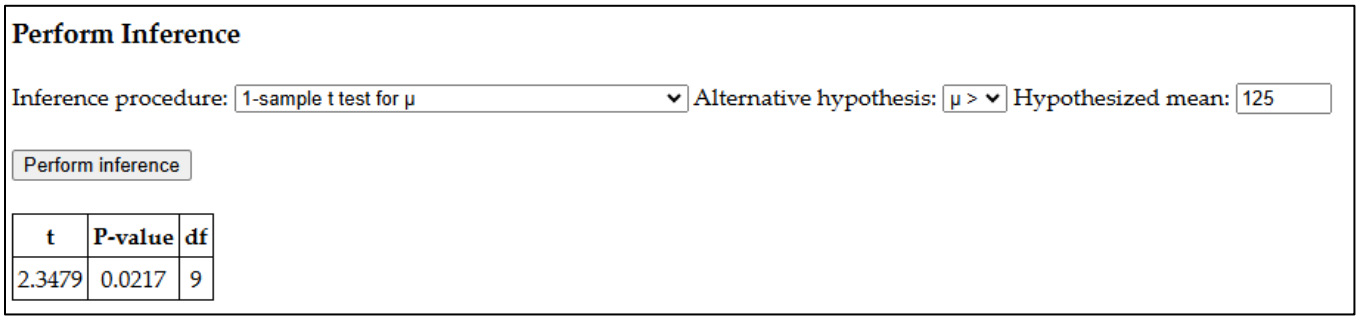
Calculate and interpret the corresponding confidence interval. Would you recommend this company adopt the new tread pattern? Why or why not?

What would you do if there was an 11^{th} data point of 102 feet?

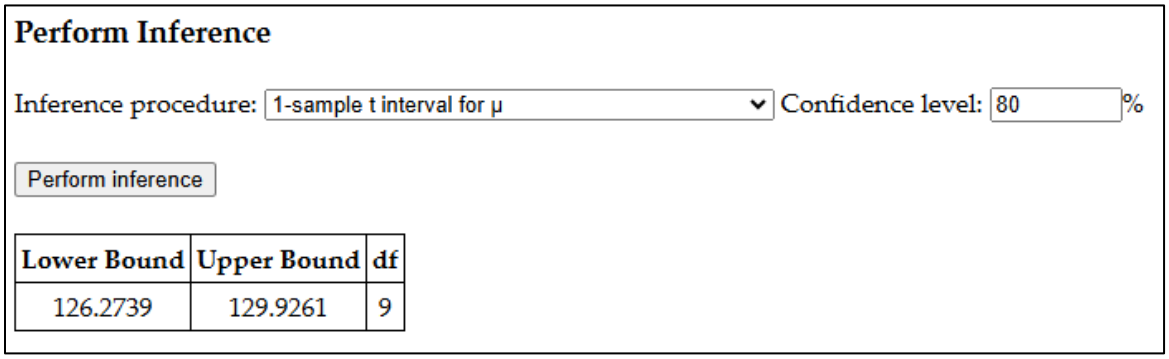
What would you do if the 11^{th} data point was 150 feet?

Theoretical Calculators on Stapplet

Hypothesis Test: 1-sample t test for μ.



Confidence Interval: 1-sample t interval for μ.



Which Theoretical Calculator to Use

GeoGebra has inputs for the mean and standard deviation, but it doesn't take raw data.

- If a problem gives you raw data, use Stapplet (can't use for simulation of a 1 mean test)
- If a problem gives you the sample mean and standard deviation, use GeoGebra or Stapplet (select mean and standard deviation instead of raw data)

Simulation for One Mean

StatKey will do a randomized hypothesis test of a mean but Stapplet will not. There is no place to enter the null value.

Lesson 10 Practice Problem 1

Twelve different video games showing alcohol use were observed. The duration times of alcohol use were recorded, with the times in seconds listed in the table. These data are from the study "Content and Ratings of Teen-Rated Video Games," by Haninger and Thompson, *Journal of the American Medical Association*, Vol. 291, No. 7.

a. Check the conditions to see whether we can use a theoretical test to see if the sample provides evidence that the mean duration of alcohol use is greater than 90 seconds at the .10 significance level.
b. If the conditions are met, do the theoretical test. If they are not met, do a simulated test on StatKey.
c. Calculate the corresponding confidence level for the test above. Find the theoretical or simulated confidence interval and compare your results. Do the hypothesis test and confidence interval give the same result? Explain why or why not.

Duration of Alcohol Use (seconds)
84
14
583
50
0
57
207
43
178
0
2
57

Lesson 10 Practice Problem 2 (With only summary statistics given)

In the 2010-11 National Hockey League (NHL) regular season, the number of penalty minutes per game for each of the 30 teams ranged from a low of 8.8 minutes for the Florida Panthers to a high of 18.0 minutes for the New York Islanders. The mean for all 30 teams in that season was 12.2 penalty minutes per game with a standard deviation of 2.25 penalty minutes.

a. Check the conditions and use a theoretical test to see if this sample provides evidence that the mean number of penalty minutes per game in the hockey league is less than 13 minutes per game at the .05 significance level.
b. Calculate the corresponding confidence level for the test above. Calculate the theoretical confidence interval and compare your results. Do the hypothesis test and confidence interval give the same result? Explain why or why not.

Lesson 11: Inference for Matched Pairs

CAOS Exam (Comprehensive Assessment of Outcomes in Statistics). This exam is an online multiple-choice test on concepts covered in a typical introductory statistics course. Students take one version before the start of the course and another version after the course ends. Before and after scores for a random sample of 10 students are shown in the table and provided in a google sheet in D2L. Each exam consists of 40 multiple choice questions and the score is the percentage correct.

Table 1. CAOS Scores for a random sample of 10 students

Student	Score Before Course (%)	Score After Course (%)	
A	42.5	60	
B	40	45	
C	47.5	55	
D	65	80	
E	60	65	
F	47.5	72.5	
G	42.5	57.5	
H	37.5	55	
I	42.5	57.5	
J	37.5	55	

Are these 2 groups independent? Why or why not?

Let's do a hypothesis test at the 5% significance level. Does it appear there is a significant difference between the before and after scores?

Write the hypotheses:

Find the differences using the Google Sheet, then enter the data in Stapplet – One Quantitative, Single Group.

Use the Summary Statistics to write the observed statistic/point estimate:

Simulated Hypothesis Test and Confidence Interval

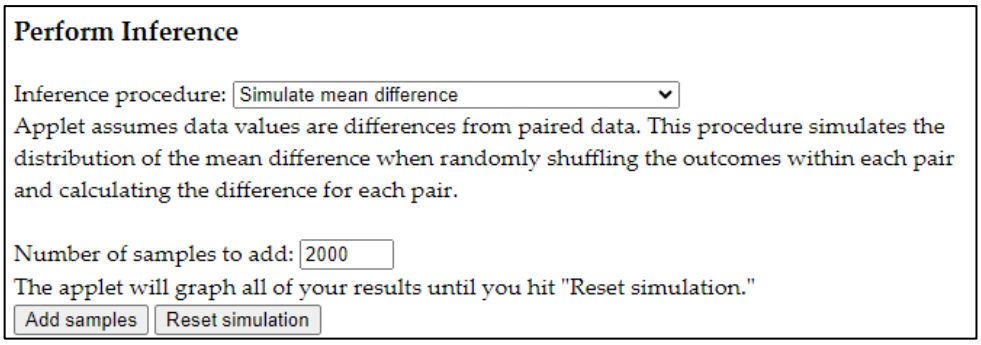
First, we'll do a simulation because conditions don't need to be met. Stapplet (1 Quantitative, Single Group) will do the simulation for us. Scroll down to Perform Inference, select Simulation mean difference and read how it does the simulation.

Which Simulation to Use

- Use **Simulate mean difference** for a hypothesis test. (Centered at 0)
- Use **Simulate sample mean** for a confidence interval (Centered at $\bar{x}$). Bootstrapping resamples the differences with replacement

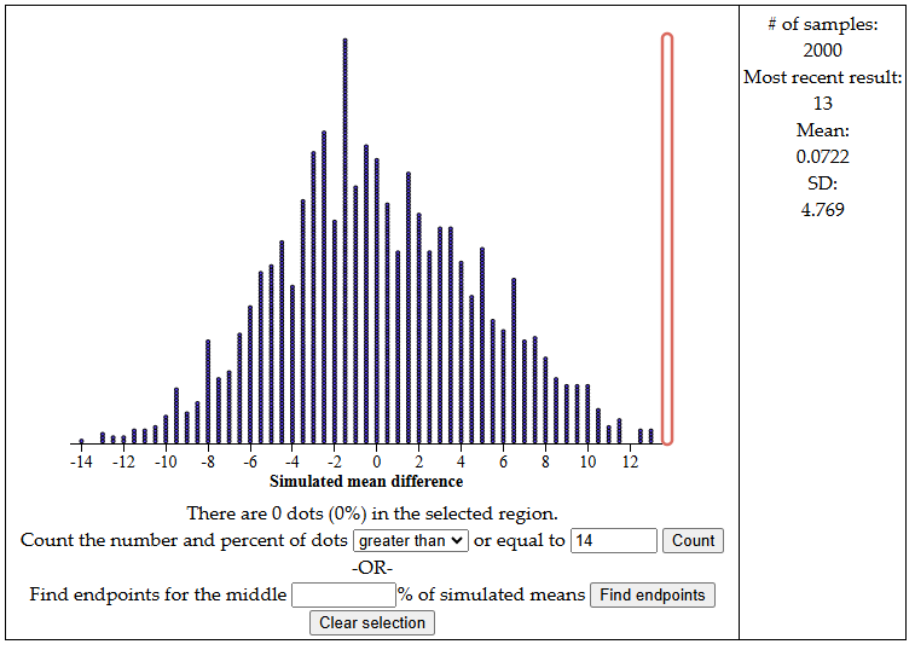
For the hypothesis test we want to know how likely it is to get a difference as large as our observed statistic (or further from 0) in our simulation of the null hypothesis (no difference). Use a significance level of 0.05.

Write down your simulated p-value.



For the confidence interval, select simulate sample mean, add 2000 simulations, and enter the confidence level. Write down your simulated confidence interval with units.

Write your conclusion in context, talking about the mean difference.



Conditions and Formulas for HT and CI For Matched Pairs (Treat like a Single Mean)

For a theoretical hypothesis test and confidence interval, we get our p-value and t^* from the t-distribution, so the conditions need to be met.

Conditions

1. **Paired Data:** The data are matched pairs from the same subjects.
2. **Independence**: The individuals must be independent from each other.
3. **Randomization**: The paired data are from a random sample or matched pairs experiment.
4. **Large Population**: If sampling without replacement, we must be sampling from a large population to consider the individuals independent of each other. The population must be 10 times the sample size.
5. **Sample Size Condition**: $n_{diff} \geq 30$, or population of differences normally distributed, or the sample differences are nearly normal (check histogram in Stapplet).

Degrees of Freedom

$$df = n_{diff} - 1$$

Hypothesis Test (T Test of a Mean)

Test Statistic: $$T_{df} = \frac{\bar{x}_{diff} - \mu_0}{\frac{s_{diff}}{\sqrt{n_{diff}}}}$$

Confidence Interval (T Estimate of a Mean)

$$\bar{x}_{diff} \pm t^*_{df} \frac{s_{diff}}{\sqrt{n_{diff}}}$$

Check the conditions to use the theoretical method.

Complete the theoretical test at the 0.05 significance level.

Find the corresponding confidence level. Construct and interpret the theoretical confidence interval for the true mean of the differences. Do your results agree? Explain why or why not.

Check both of these with the Stapplet and/or GeoGebra theoretical calculator.

Lesson 11 Practice Problem 1

A local high school track coach wants to track improvement for the team on average. To do this, they randomly select 12 runners and record their mile times at the start and end of the season. Do the data show evidence that they have improved on average? How much improvement if any? The data are in the Lesson 11 Google sheet.

Table 2. High School Track 1-mile times for 12 randomly selected runners

a. Use the data spreadsheet to calculate the differences and put them into Stapplet.
b. Test whether the runners have improved significantly at the 0.05 level of significance using the simulation in Stapplet.
c. Check the conditions and do the theoretical test.
d. Find the corresponding bootstrap and theoretical confidence intervals to find the estimate of the improvement if any.

Runner	Start of Season time (min)	End of Season time (min)	
A	8.75	8.40	
B	7.30	6.30	
C	12.50	14.60	
D	6.40	6.00	
E	7.15	8.18	
F	9.50	8.58	
G	10.68	9.30	
H	9.32	7.22	
I	11.15	9.09	
J	13.20	10.00	
K	8.76	6.48	
L	7.43	6.52	

Lesson 11 Practice Problem 2

Heilbronn, Smith, Martin, Anton and Ravussin[1] conducted a study with 16 nonobese subjects to determine the effect of fasting every other day. The subjects participated in the experiment for 22 days. One variable that was measured was the weight of the subjects before and after the experiment was conducted.

The table below contains the weight (in pounds) before and after the study for the 16 subjects. The mean weight before and after the study was 164.19 and 160.10, respectively.

Table 3. Weights in pounds before and after the study

a. Use the data spreadsheet to calculate the differences and put them into Stapplet (Before – After).
b. Test whether the subjects lost weight on average at the 0.01 level of significance using the simulation in Stapplet.
c. Check the conditions and do the theoretical test.
d. Find the corresponding bootstrap and theoretical confidence intervals to find the estimate of the weight loss if it is significant.

ID	Before	After	Weight Loss Before - After
1	112	109	
2	118	118.5	
3	121	117.1	
4	149	144.7	
5	149	145.2	
6	153	147.8	
7	163	158.7	
8	163	157.5	
9	168	163.9	
10	170	166.3	
11	178	173.6	
12	179	175.1	
13	187	182.2	
14	191	186.7	
15	211	206.1	
16	215	209.2	

[1] Heilbronn, L. K., Smith, S. R., Martin, C. K., Anton, S. D. & Ravussin, E. (2005). Alternative-day fasting in nonobese subjects: Effects on body weight, body composition, and energy metabolism. *The American Journal of Clinical Nutrition, 81*(1), pp. 69-73.

Lesson 12: Inference for 2 Independent Means

Researchers did a study with 4-year-old children on whether watching 9 minutes of a fast-paced cartoon would affect self-control, the ability to delay gratification and working memory. Sixty kids were randomly assigned to one of 3 groups that did one of the following activities for 9 minutes:

- Watch a fast-paced cartoon
- Watch a slower paced educational program
- Color with markers and crayons

The children selected for the study had no differences in attention or amount of TV usually watched. We have the data on the time waited to push a button for a treat (in seconds). See the Google Sheet in D2L.

Let's do a hypothesis test. Do you think there would be a significant difference between the time waited for a snack between the 2 groups?

Write the hypotheses:

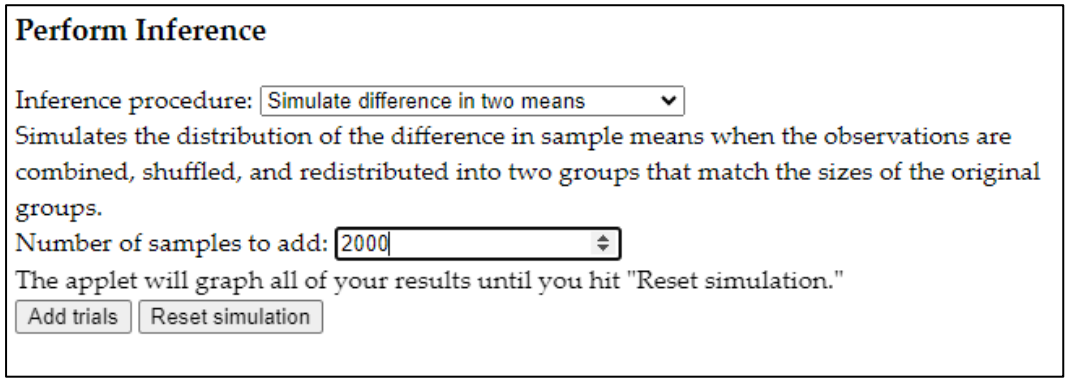
Enter the data in Stapplet – One Quantitative, Multiple Groups, labeling the groups and including units.

Use the Summary Statistics to calculate the observed statistic/point estimate:

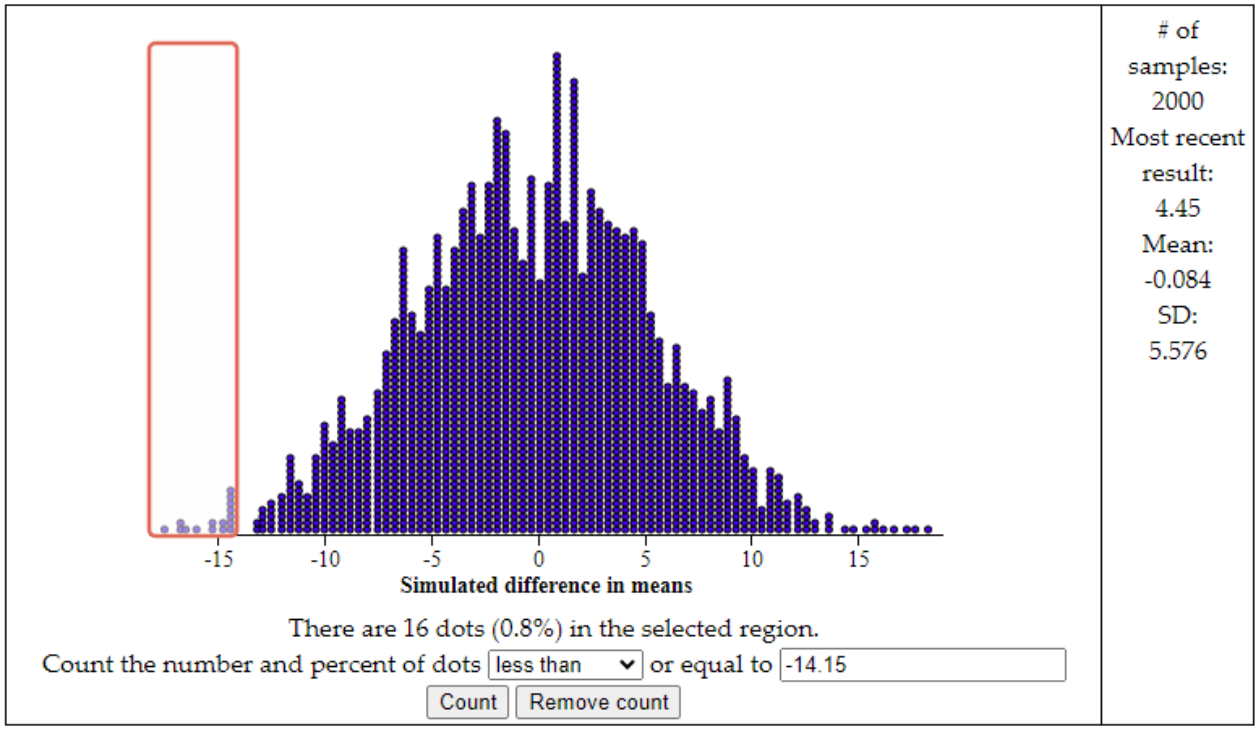
Simulated Hypothesis Test

First, we'll do a simulation because conditions don't need to be met. Just like in the dolphin experiment, if we were to do this physically, we would put each of the 40 data values on a card, mix them up and then divide them back into 2 groups to simulate the null distribution. We would calculate the difference between the 2 averages and plot it on a dotplot.

Stapplet will do the simulation for us. Scroll down to Perform Inference and add 2000 trials.



Now we want to know how likely it is to get a difference as large as our observed statistic (or further from 0) in our simulation of the null hypothesis (no difference). Use a significance level of .05.



Write down your simulated p-value. Remember to multiply by 2 for a 2-tail test.

Write your conclusion in context, talking about the difference in the two means.

Note: Stapplet does not have a bootstrapping option for 2 means. StatKey does, but the data would need to be formatted differently.

Conditions and Formulas for HT and CI Comparing 2 Independent Means

For a theoretical hypothesis test and confidence interval, we get our p-value and t^* from the t-distribution, so the conditions need to be met.

Conditions

1. **Independence**: The individuals in each group and between groups must be independent of each other regarding the variable measured.
2. **Randomization**: The data are from 2 independent random samples or 2 randomly assigned treatments.
3. **Large Population**: If sampling without replacement, we must be sampling from a large population to consider the individuals independent of each other. The population must be 10 times each group's size.
4. **Sample Size Condition**: $n_1 \text{ and } n_2 \geq 30$, or population is normally distributed, or samples are nearly normal.

Degrees of Freedom

The combined df calculation is a complex formula. We will use technology to find the degrees of freedom or use the **estimated/conservative method of the smaller of $\boldsymbol{n_1 - 1}$ or $\boldsymbol{n_2 - 1}$.**

Calculated (we will use technology for this) $df = \dfrac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\left(\frac{1}{n_1-1}\right)\left(\frac{s_1^2}{n_1}\right)^2+\left(\frac{1}{n_2-1}\right)\left(\frac{s_2^2}{n_2}\right)^2}$

Hypothesis Test (T Test, Difference of means)

Test Statistic: $T_{df} = \dfrac{(\bar{x}_1-\bar{x}_2)-0}{\sqrt{\frac{{s_1}^2}{n_1}+\frac{{s_2}^2}{n_2}}}$

Confidence Interval (T Estimate, Difference of means)

$$\bar{x}_1 - \bar{x}_2 \pm t^*_{df}\sqrt{\frac{{s_1}^2}{n_1}+\frac{{s_2}^2}{n_2}}$$

Pooled vs. Unpooled T-tests

Remember how we pooled the proportions for the null to get $\hat{p}_C$ for the 2 proportions test? There is a pooled t-test, but that assumes the variances of each population are the same, so we are using the unpooled t-test. Assuming the means are the same in the null hypothesis doesn't tell us anything about the variances.

Check the conditions and write out the theoretical test at the .05 significance level. Use the estimated degrees of freedom and see how it compares to the calculated, df=28.44. Then construct the equivalent confidence interval for the true mean difference between the two groups. Summarize all your results. Check both of these using the Stapplet theoretical calculator.

Continued…

Matched Pairs or 2 Independent Means?

For each of the following examples determine whether we should be using matched pairs or 2 independent means. Also give the degrees of freedom. Use the estimated method for 2 independent means.

a. We want to compare the age gap between two people in a relationship. We study 60 people (30 couples).

b. We want to compare the average time it takes to get to work for two different routes. We randomly select 30 days for each route and record the data.

c. We want to compare the number of credits taken by students at the Southeast and Cascade campuses of PCC. We sample 70 SE students and 80 CA students.

d. We want to compare how long it takes to complete a crossword puzzle with and without coffee. We measure the amount of time for 20 people without coffee and then measure the same 20 people on a similar puzzle with coffee.

e. We want to compare how long it takes to complete a wordle puzzle with and without coffee. We measure the amount of time for 20 people without coffee and then measure a different group of 20 people on the same puzzle with coffee.

Lesson 12 Practice Problem 1

Marijuana Users Executive Functioning. Lacerda et al.,[1] studied marijuana users who started using before the age of 15. Volunteers for the study included marijuana users (n = 49) and non-marijuana users (n = 44). They were compared to each other using a neuropsychological task (the Stroop Test).

There are many versions of the Stroop Test. This study used 3 different versions. The first test showed the participant's blocks of colors, and they had to name the colors as fast as they could. The second listed multiple nouns written in different colors and again the participants had to name the colors as fast as they could. The last test listed multiple colors (in word form) that were shown in different colors. The participant would have to say the color that the word was written in. For example, the participant would see the word "red" but it was written in blue. The participant would have to name the color blue.

To demonstrate this last test, your instructor will now show a test on the document camera and you will name the color of the words as quickly as you can with the class.

For this study, each of the volunteers took each of the 3 Stroop Tests. For each test they were given a score. Those 3 scores were then added to create a single measure, the total response time spent on the 3 Stroop Tests. The data file for this activity is in the Lesson 12 Google Sheet in D2L.

a. Check the conditions and use the theoretical method to test the claim that the marijuana users take longer on the Stroop Tests. Use a significance level of 0.05. The calculated df = 72.865.
b. Determine the corresponding confidence level and calculate the confidence interval.
c. Use Stapplet to find a simulated p-value and check the theoretical results and summarize your results.
d. Can you generalize these results to all marijuana users and non-marijuana users? Why or why not?
e. Can you make cause-and-effect conclusions? Why or why not?

[1] Fontes, M. A., Bolla, K. I., Cunha, P. J., Almeida, P. P., Jungerman, F., Laranjeira, R. R., Bressan, R. A., & Lacerda, A. L. T. (2011). Cannabis use before age 15 and subsequent executive functioning. *The British Journal of Psychiatry, 198*, pp. 442-447. doi: 10.1192/bjp.bp.110.077479

Practice Problem 1 continued.

Lesson 12 Practice Problem 2

Do consumers spend more on a trip to Walmart or Target? Suppose researchers interested in this question collected a systematic sample from 85 Walmart customers and 80 Target customers by asking customers for their purchase amount as they left the stores. The data collected are summarized in the table. The calculated df = 162.94. Check the conditions, then test the claim that there is a difference in the purchase amounts at the 0.05 significance level. Find the corresponding confidence interval and summarize all your results.

	Walmart	Target
n	85	80
$\bar{x}$	\$135	\$151
s	\$26	\$24

Lesson 13: ANOVA (ANalysis Of VAriance)

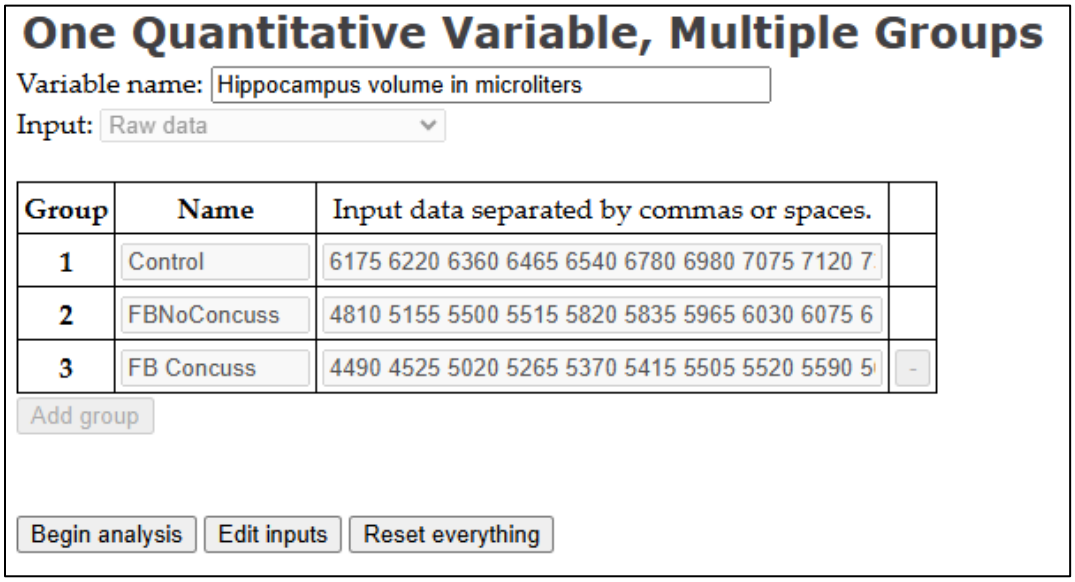
We have compared multiple counts or proportions using the chi-squared tests for goodness of fit, homogeneity, and independence. Now we can use ANOVA to compare multiple means, and we will learn a new distribution.

A study examined the relationship between playing football and concussions on hippocampus volume in the brain. The study included three groups with n=25 in each group. The data for the study are shown in the Lesson 13 spreadsheet and give the brain volume in microliters, μL.

- Group 1 were healthy control subjects who had never played football
- Group 2 were football players with no history of concussions
- Group 3 were football players with a history of concussions.

Do you think there might be a difference in brain size between the three groups? Let's write our hypotheses.

Enter the data in Stapplet – One Quantitative, Multiple Groups, labeling the groups and including units with the variable name.



Click on Begin analysis and select histograms first. Check whether these are nearly normal. Then select boxplot to compare the boxplots side-by-side.

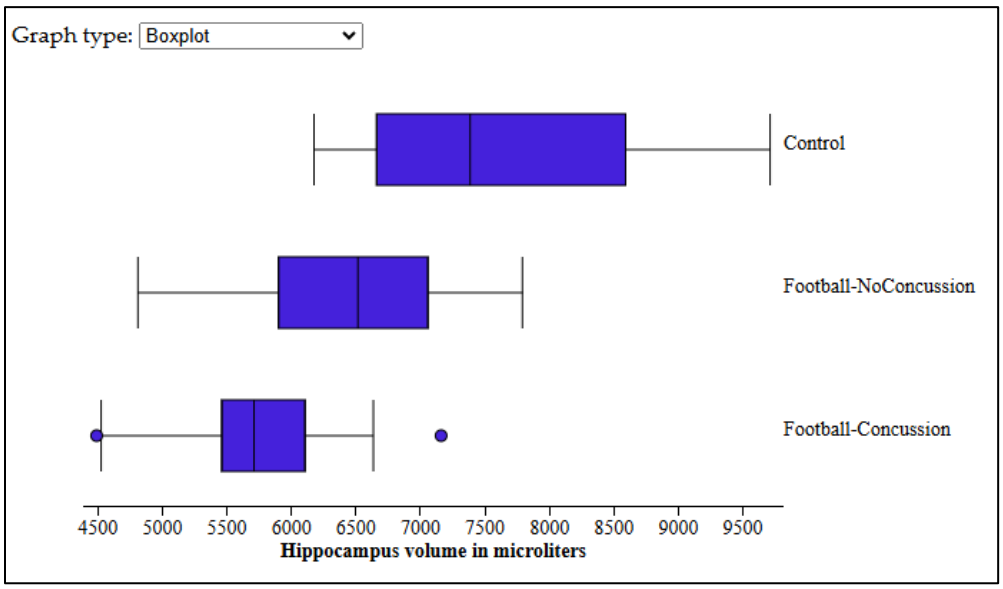


Figure 1. Summary Statistics from Stapplet

Group Name	n	mean	SD	min	Q1	med	Q3	max
1: Control	25	7602.6	1074.023	6175	6660	7385	8590	9710
2: Football-NoConcussion	25	6459.2	779.7391	4810	5900	6515	7057.5	7790
3: Football-Concussion	25	5734.6	593.3976	4490	5460	5710	6105	7160

Under Inference procedure, select One-way ANOVA.

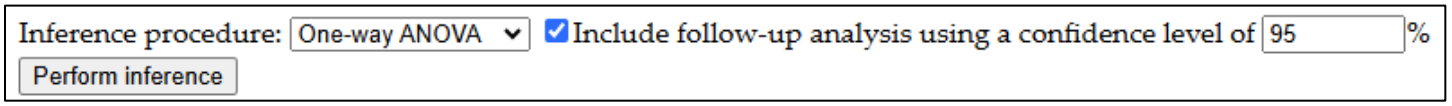


Figure 2. ANOVA output table from Stapplet

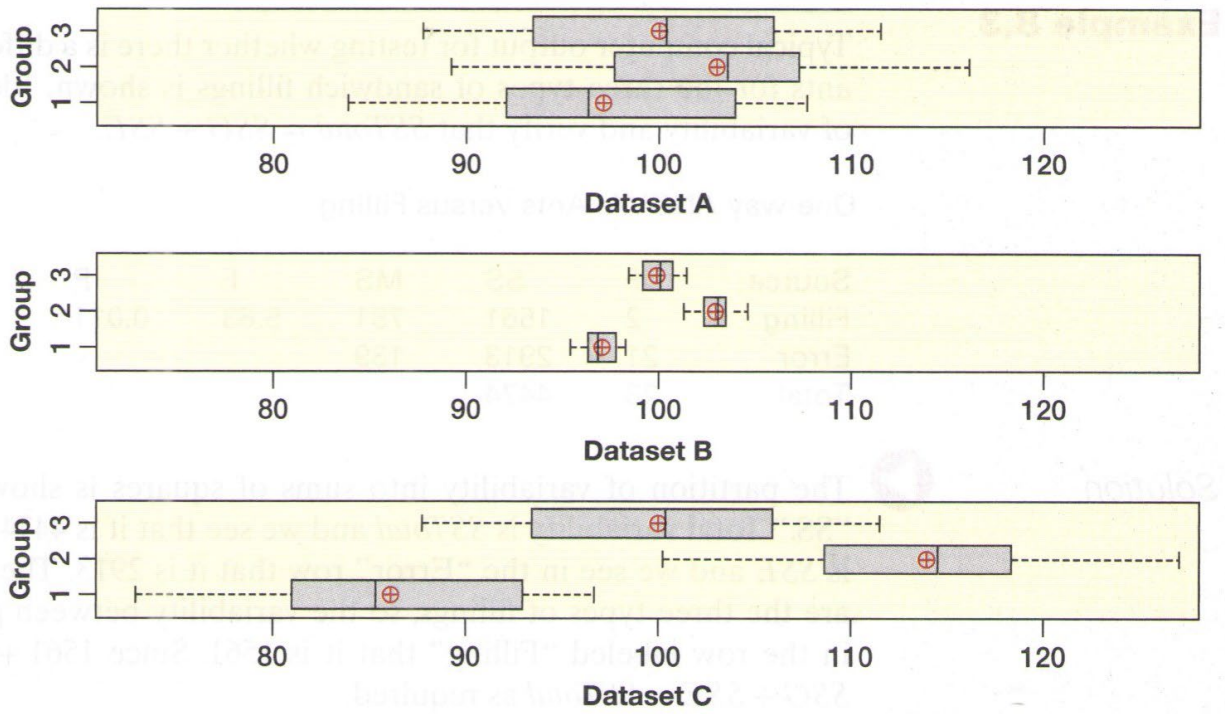
Source	df	Sum of Squares	Mean Square	F-value	P-value
Group	2	44348606	22174303	31.4732	<0.001
Error	72	50727336	704546.3333		
Total	74	95075942			

Figure 3. Confidence Interval Output from Stapplet

Groups	Fisher's LSD Interval
Group 1 - Group 2	(670.131, 1616.669)
Group 1 - Group 3	(1394.731, 2341.269)
Group 2 - Group 3	(251.331, 1197.869)

The F-Statistic: Comparing Means and Variation

Look at the 3 sets of side-by-side boxplots, A, B and C. The location of the group mean is shown in each boxplot. Datasets A and B have the same group means. Datasets A and C have the same spreads. Between Datasets A and B, which has the strongest evidence of a difference in means? What do you think about Dataset C?



Comparing Datasets A and B shows us that we need to consider the variation along with the differences in means. Just knowing the sample means is not enough. If the values within each sample are very close, then small differences may be significant as in Dataset B. If there is more variation in the samples, then we need the group means to be further apart to be significant as in Dataset C.

This leads us to consider a ratio of the two types of variation. We want to compare the differences in the means relative to the "average spread." So we will compute 2 quantities and divide them. The F-ratio is roughly the proportion of allowed variation used by the differences in the means.

$$\text{F-ratio} = \frac{\text{"difference in means"}}{\text{"average spread"}}$$

$$\text{F-ratio} = \frac{\text{Variation Between Groups}}{\text{Variation Within Groups}}$$

Finding the Overall Mean

Use the spreadsheet to find the mean of all 75 subjects, $\bar{X}$.

Variation Between Groups (MSG – mean square group/ treatment, also called MST)

We need a measure of how far away the means are from the overall mean for k groups. $df = k - 1$.

$$\text{Var Between} = \frac{\sum n_i(\bar{x_i} - \bar{X})^2}{k - 1}$$

Variation Within Groups (MSE – mean square error)

We want to get a sense of the average spread of the groups for our baseline. N is the total sample size. The degrees of freedom for the k groups add up to $df = N - k$.

$$\text{Var Within } = \frac{\sum df_i(s_i)^2}{N-k}$$

Conditions and Formulas for ANOVA – Multiple Means

For a theoretical hypothesis test, we get our p-value from the F-distribution, so the conditions need to be met.

Conditions

1. **Independence**: The individuals in each group and between groups must be independent of each other regarding the variable measured.
2. **Randomization**: The data are from independent random samples or randomly assigned treatments.
3. **Large Population**: If sampling without replacement, we must be sampling from a large population to consider the individuals independent of each other. The population must be 10 times each group's size.
4. **Sample Size Condition**: $n_i \geq 30$, or population is normally distributed, or samples are nearly normal. (check histograms)
5. **Approximately Equal Spreads**: The groups have approximately the same spread. Rough guideline: the largest standard deviation should not be more than twice the smallest.

Degrees of Freedom

Variation between groups: $df = k - 1$. (# groups – 1)

Variation within groups: $df = N - k$. (Total sample size – # groups)

Hypotheses

H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$ or there is no difference in the means.
H_A: At least one mean is different.

Hypothesis Test (ANOVA, F Test)

Test Statistic: $F = \frac{Var\ Between}{Var\ Within} = \frac{MSG}{MSE}$. This is a positive ratio, so this is always a right tail test.

$$MSG = \frac{\sum n_i(\bar{x}_i - \bar{X})^2}{k - 1}$$

$$MSE = \frac{\sum df_i(s_i)^2}{N - k}$$

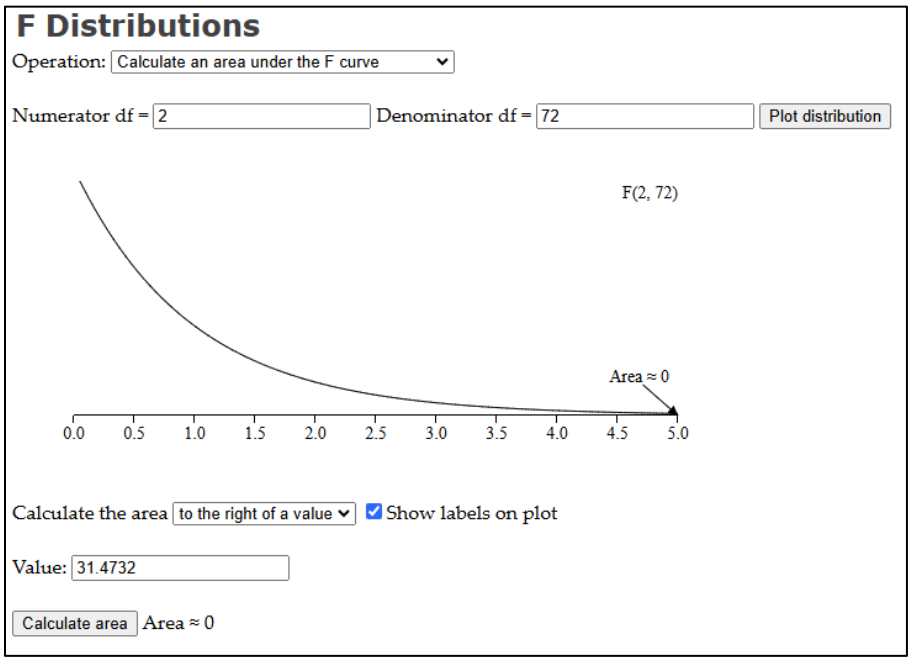
Check the conditions needed to use this model.

Write the F-statistic for this test.

Find the p-value with the F-distribution

The F distribution has 2 inputs, the degrees of freedom for the numerator and denominator.

[Stapplet F Distribution]



- If the F-ratio is small, there is not a large difference in the means compared to the spread, which is evidence for the null hypothesis. (Should lead to a large p-value which fails to reject the null)
- If the F-ratio is large, there is a lot of difference in the means compared to the spread, which is evidence for the alternate. (Should lead to a small p-value which rejects the null)

Patterns in the ANOVA Table and Common Abbreviations

The number of groups = k
The total sample size = N

Source	df	Sum of Squares	Mean Square	F-value	p-value
Group					
Error					
Total					

Finding the overall mean without raw data

If we don't have the raw data and all the group sample sizes are the same, we can average the group means. If the group sizes are different, we can find the overall mean with this formula.

$$\bar{X} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots + n_k\bar{x}_k}{N}$$

Multiple Comparisons

When we reject the null that tells us that one of the means is different, but it doesn't tell us which one. We can use Stapplet to find the Fisher's Least Significant Difference (LSD) intervals and determine which mean(s) are significantly different.

This method uses the square root of MSE from the ANOVA as the pooled standard deviation. We can pool these because of the approximately equal spread condition. It is a bit stronger test because we are using the pooled standard deviation for all the groups rather than just two groups and we are using $df = N - k$.

Least Significant Difference Intervals

$$\boldsymbol{\bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{MSE\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

When comparing many confidence intervals there's an issue because the chance of a type I error for multiple comparisons would end up being larger than our original significance level. In order to fix this, we can adjust our confidence level by dividing the significance level by the number of pairwise comparisons we are making. When doing this with the LSD intervals it is call the MSD or minimum significant difference.

Bonferroni Correction

Let $\boldsymbol{\alpha^* = \frac{\alpha}{\#\text{ of pairs}}}$, then $\boldsymbol{C^* = 1 - \alpha^*}$

Lesson 13 Practice Problem 1

A pharmaceutical company tested three formulations of pain relief medicine for migraine headache sufferers. For the experiment, 27 volunteers were selected and 9 were randomly assigned to one of three drug formulations. The subjects were instructed to take the drug during their next migraine headache episode and to report their pain 30 minutes after taking the drug on a scale of 1 = no pain to 10 = extreme pain. The data for the study are shown in the Lesson 13 spreadsheet.

Write the hypotheses and calculate the F-statistic to see whether there is a difference in the effectiveness of the three drugs.

Lesson 13 Practice Problem 2 (Only summary statistics given)

A nursing student investigated how effective hand washing with soap is in eliminating bacteria. To do this they tested four different methods – washing with water only, washing with regular soap, washing with antibacterial soap, and spraying hands with antibacterial spray containing 65% ethanol. They generated random numbers to determine the order of the 4 treatments, then placed their right hand on a sterile media plate designed to encourage bacteria growth. They incubated each plate for 2 days at 36℃ and then counted the bacteria colonies. This procedure was done 8 times for each treatment. The data are shown in the table below. Run an ANOVA analysis to test whether there is a difference in the methods and follow up with MSD intervals to see which methods are different, if any.

Group	Mean	Standard Deviation	n
Alcohol Spray	37.5	26.56	8
Antibacterial Soap	92.5	41.96	8
Soap	106.0	46.96	8
Water	117.0	31.13	8

ANOVA Output

Figure 4. ANOVA Table output from Stapplet

Source	df	Sum of Squares	Mean Square	F-value	P-value
Group	3	29882	9960.6667	7.0638	0.0011
Error	28	39482.7559	1410.0984		
Total	31	69364.7559			

Figure 5. LSD Intervals (95% confidence each)

Groups	Fisher's LSD Interval
Group 1 - Group 2	(-93.4602, -16.5398)
Group 1 - Group 3	(-106.9602, -30.0398)
Group 1 - Group 4	(-117.9602, -41.0398)
Group 2 - Group 3	(-51.9602, 24.9602)
Group 2 - Group 4	(-62.9602, 13.9602)
Group 3 - Group 4	(-49.4602, 27.4602)

Figure 6. MSD Intervals (Total of 95% confidence)

Groups	Fisher's LSD Interval
Group 1 - Group 2	(-108.3338, -1.6662)
Group 1 - Group 3	(-121.8338, -15.1662)
Group 1 - Group 4	(-132.8338, -26.1662)
Group 2 - Group 3	(-66.8338, 39.8338)
Group 2 - Group 4	(-77.8338, 28.8338)
Group 3 - Group 4	(-64.3338, 42.3338)

Lesson 14: Inference for Linear Regression

Is there a relationship between family income and financial aid award at Elmhurst University? Elmhurst is a private university outside of Chicago, Illinois. Open the Elmhurst University Data Set (Google Sheet in D2L) and enter the family income as the explanatory variable (x) and financial gift aid as the response variable (y) in [Stapplet – 2 Quantitative Variables.]

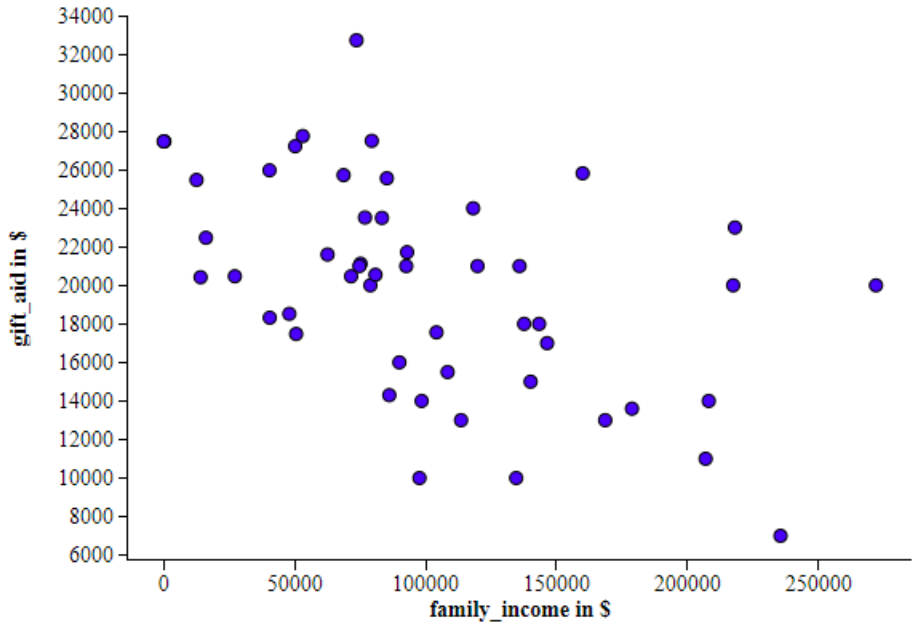
Who are the subjects? Explore the output and inference options and then we'll go through each part.

Two Quantitative Variables

Variable	Name	Observations (separated by commas or spaces) *Keep individuals in the same order.*
Explanatory	family_income in	92922 250 53092 50200 137613 47957 113534 168579 208115
Response	gift_aid in $	21720 27470 27750 27220 18000 18520 13000 13000 14000 25

Begin analysis | Edit inputs | Reset everything

Financial Aid Gift Award vs. Family Income for a Random Sample of Elmhurst University Students, n=50



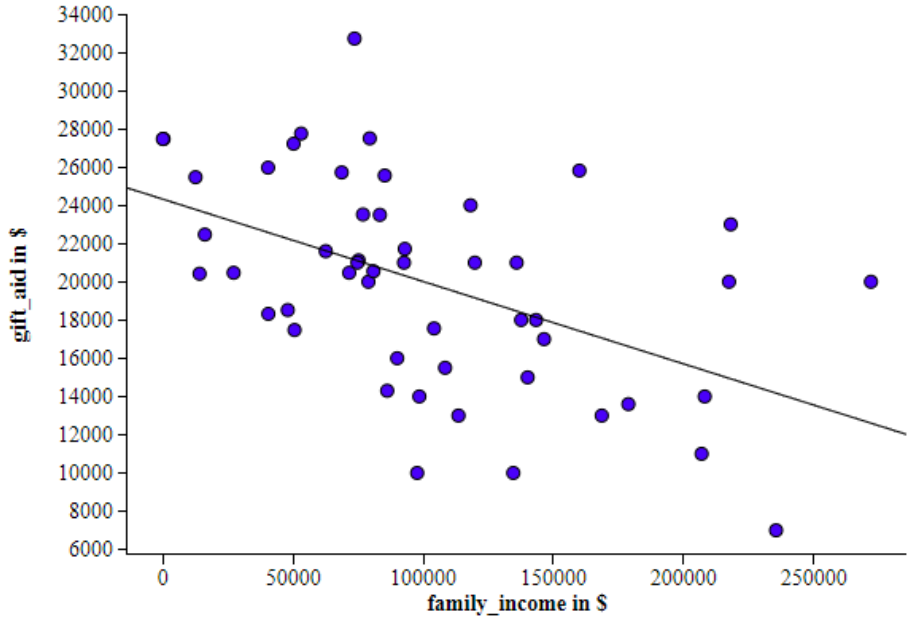
Does it look like there's a relationship between family income and gift award amounts?

Describe four features of the association between two variables

- Direction (positive, negative, no direction)
- Form (linear, curved, no pattern)
- Strength, measured by the correlation coefficient (strong, moderate, weak)
- Unusual Features (groupings or outliers)

After we find the correlation, we'll describe this association.

Scatterplot



Calculate Correlation

Calculate correlation r = -0.4986

The **linear correlation coefficient, r,** measures the strength of the linear correlation between the paired quantitative x- and y-values in a sample.

Caution: Even if it appears that y can be "predicted" from x, it does not follow that x **causes** y.

Characteristics of r

- The value of r only makes sense for linear relationships.
- The value of r is between -1 and 1, 0 is no correlation, -1 or 1 are perfect correlations.
- The sign of r is the direction of the association, same as the sign of the slope.
- r has no units.



-1 **0** **1**

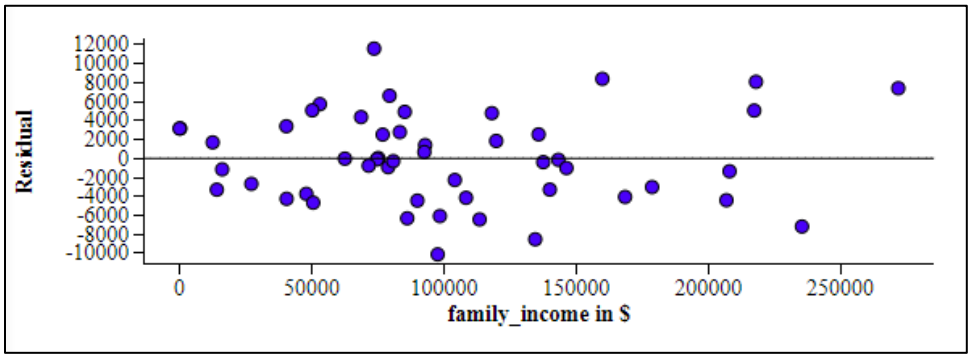
Describe the association between family income and gift aid award at Elmhurst University.

Residual/Error plot

Residuals are the errors in the model. They follow a t-distribution with $df = n - 2$.

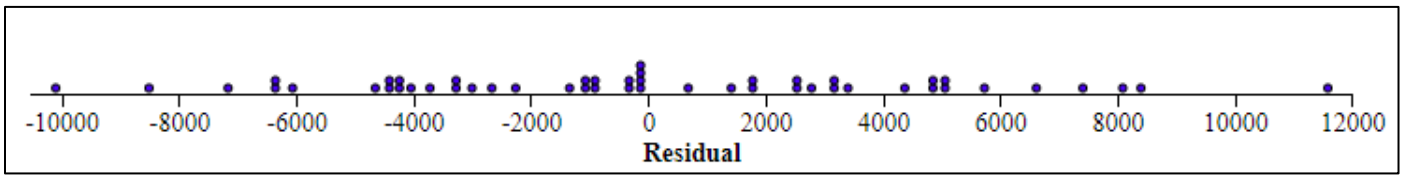
Residual $= y - \hat{y}$ (observed – predicted)

Look for even scatter in the residual scatterplot (no patterns or variation in spread).



Residual distribution (dot plot)

Look for normality in the residual distribution.



Least Squares Regression Model

Equation for the population regression line: $\boldsymbol{\mu_y = \alpha + \beta x}$

Least Squares estimate from the sample: $\boldsymbol{\hat{y} = a + bx}$

Equation	n	s	r^2
$\hat{y} = 24319.329007 - 0.04307x$	50	4782.5989	0.2486

$$\text{Slope} = \frac{\text{rise}}{\text{run}} = \frac{\text{change in } y}{\text{change in } x}$$

For the Elmhurst data, what is the sample slope, intercept and standard deviation of the residuals? What do they mean?

R^2, the Coefficient of Determination

The percentage of variation in the [y-variable] that can be explained by the [x-variable]. Describe this for our data set in context.

Conditions and Formulas for Linear Regression Inference

The residuals follow a t-distribution.

Conditions

1. **Linearity:** Check the scatterplot for a linear pattern.
2. **Independence:** The observations are independent of each other, be careful of time series plots.
3. **Randomization:** The data should be from a random sample or randomized experiment.
4. **Sample size/Nearly normal residuals:** $n \geq 30$ or check the residual dotplot for nearly normal residuals.
5. **Constant variability of the residuals:** check the residual scatterplot to make sure there are no obvious patterns.

Degrees of Freedom

$$df = n - 2$$

Hypotheses

H_0: $\beta = 0$ (There is no real trend in the population. Any apparent trend is due to data selection by chance)

H_A: $\beta \neq 0,\ \beta > 0,\ or\ \beta < 0$ (There is a real trend in the population)

Hypothesis Test (T Test for slope)

Test Statistic: $T = \frac{\text{observed statistic - null value}}{\text{SE of estimate}} = \frac{b-0}{SE}$

Confidence Interval for the Slope

$$b \pm t^*_{df} \cdot SE$$

Check the conditions to perform inference on the Elmhurst slope.

Test whether the slope is negative with a significance level of .05. Form the corresponding confidence interval for the slope and give the interpretation.

Inference Options on Stapplet (skip simulation)

- t interval for slope
- t test for slope
- Prediction interval for value of response variable (given explanatory value)
- Confidence Interval for mean of response variable (given an explanatory value)

Perform Inference

Inference procedure: t test for slope Alternative hypothesis: β < 0

Perform inference

t	P-value	df
-3.9846	<0.001	48

Inference procedure: t interval for slope Confidence level: 90 %

Perform inference

Lower Bound	Upper Bound	df
-0.0612	-0.0249	48

More Detailed Regression Output

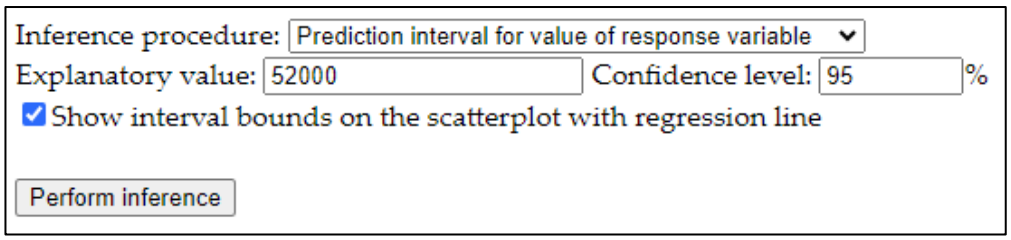
Predictor	Coef	SE Coef	T	P
Constant	24319.329	1291.4503	18.831	<0.001
family_income in $	-0.0431	0.0108	-3.9846	<0.001

Find with Stapplet: Prediction interval for individual value of response variable (given an explanatory value)

A student named Cam is planning to attend Elmhurst University and they want to estimate about how much financial aid they are likely to receive based on their family's income of $52,000.

Find the best estimate for the predicted value using the linear model:

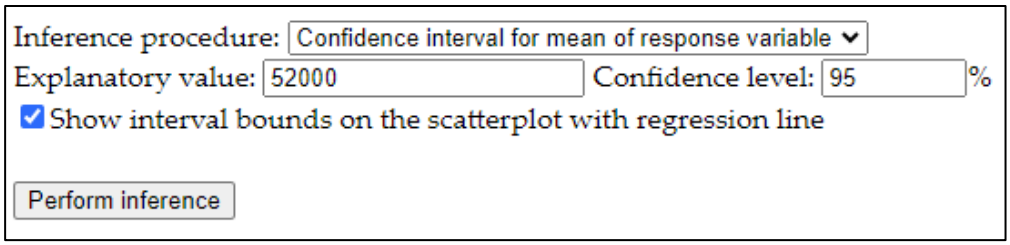
Use Stapplet to find the prediction interval for an x-value of $52,000.



Lower Bound	Upper Bound	df
12307.7845	31851.4217	48

Confidence Interval for mean of response variable (given an explanatory value)

Suppose that Cam would like to estimate the mean financial aid award for all prospective Elmhurst students whose families earn $52,000 per year. Stapplet will calculate a confidence interval for the mean response value.



Lower Bound	Upper Bound	df
20341.8346	23817.3716	48

It is more reliable to predict a mean than an individual value, so the confidence interval for a mean value is much narrower than for an individual value.

Lesson 14 Practice Problem

Listed below are costs (in dollars) of air fares for different airlines from New York City (JFK) to San Francisco (SFO). The costs are based on tickets purchased 30 days in advance and one day in advance, and the airlines are US Air, Continental, Delta, United, American, Alaska, and Northwest. Is there sufficient evidence to conclude that there is a linear correlation between costs of tickets purchased 30 days in advance and those purchased one day in advance?

30 Days in Advance	244	260	264	264	278	318	280
One Day in Advance	456	614	567	943	628	1088	536

Enter the data into Stapplet – 2 Quantitative Variables and complete the linear regression analysis.

a. What is the explanatory variable and what is the response variable?
b. Make a scatterplot, find the correlation and describe the association.
c. Write the regression equation.
d. Write the value of R^2 and explain what it means.
e. Look at the residual plots and check the conditions for inference.

f. Test whether the slope is significantly different from 0 with a significance level of .10.
g. Form the corresponding confidence interval for the slope.
h. What does the model predict for the cost of a flight purchased the day before if the price is $300 30 days before the flight? Form a 90% prediction interval for this value as well.
i. Form a 90% confidence interval for the mean price of all flights the day before if the cost is $300 30 days before.

Midterm Practice: Identification

Circle, underline or highlight the key words and phrases and identify the type of test or confidence interval.

Types of Scenarios: Single proportion CI and/or HT, difference of two proportions CI and/or HT, test of goodness of fit, test of independence, or test of homogeneity.

1. A campus club holds a raffle with prizes labeled as small, medium, or large. They claim the distribution of prize sizes is 20% small, 50% medium, and 30% large. After several raffles, another student collects data to test whether the observed frequencies match the claimed distribution.

2. A student leader claims that 25% of undergraduates attend at least one campus event per week. A statistics class surveys students to see if the proportion differs from the claimed 25%.

3. The campus health center wants to know if the proportion of students who exercise at least three times a week is different between first-year and second-year students and how different.

4. A survey of students at a community college records their employment status (not working, part-time, full-time) and whether they report being satisfied with their work–school balance (yes/no). Researchers want to test if satisfaction is associated with employment status.

5. Four community college campuses survey their students to ask if they prefer online, hybrid, or in-person classes. The district wants to know if the distributions of preferences are the same across campuses.

6. A researcher compares the proportion of students who regularly drink coffee between those living off campus and those living on campus. They are curious if more students who live off campus regularly drink coffee.

7. The cafeteria manager claims that students choose pizza, sandwiches, salads, and hot meals in equal proportions. A statistics class records the number of each meal type sold during lunch to test the claim.

8. A community college administrator wants to estimate what proportion of its students are enrolled part-time. They take a random sample of students and record their enrollment status.

Midterm Sample Problems

To practice fully, for each problem write out the conditions, show your steps using the theoretical method and check with a theoretical calculator and simulation where appropriate.

1. A car maker would like to determine if more than 40% of potential customers would consider a hybrid car. In a random sample of 500 potential customers, 212 said they would be interested in purchasing one.

    a. Test this claim at $\alpha = .05$. Show all steps and check your result on the GeoGebra theoretical calculator and StatKey simulation.

    b. Find the confidence level that corresponds to the value for α in part a for a confidence interval. Construct that confidence interval. Does your result agree with the hypothesis test? Show your steps and check your result on GeoGebra and StatKey.

    c. Let's say the car maker would like to repeat the confidence interval at the same confidence level but get a 2% margin of error. How many potential customers would they need to survey?

2. To investigate whether giving chest-compression-only (CC) instructions rather than standard cardiopulmonary resuscitation (CPR) instructions to the witness of a heart attack changes the chance of survival, researchers combined the results of three randomized experiments. In each experiment, the emergency services dispatcher randomly assigned either CC or CPR instructions to the bystander of a person experiencing a heart attack. The data for survival rates are shown in the table below.

   a. Test to see whether there is a difference in the survival rate depending on whether CC or CPR was given. Use a significance level of 0.05. Show all the steps of the test. Check your work on GeoGebra and StatKey.

	CC	CPR	Total
Survived	211	178	389
Did not survive	1289	1353	2642
Total	1500	1531	3031

   b. Explain what your p-value means as if you're helping a fellow student.

c. Compute the corresponding confidence interval and interpret the results. Do the results agree with the hypothesis test?

3. Write the hypotheses and identify the Type I Error and the Type II Error that corresponds to the claim: The percentage of college students who consume alcohol is greater than 70%.

4. Last school year, in the school of Business Administration, 30% were Accounting majors, 24% Management majors, 26% Marketing majors, and 20% Economics majors. A sample of 300 students taken from this year's students at the school showed the following number of students in each major:

Major	Students
Accounting	80
Management	68
Marketing	85
Economics	67
Total	300

Test at α = 0.05 to determine if there has been any significant change in the percentage of students in each major between the last school year and this school year. Show at least the first term of your test statistic and use GeoGebra or StatKey.

5. In an article published in Lancet (2001), researchers shared their findings from a study where they followed 6,272 Swedish cisgender men for 30 years to see whether there was an association between the amount of fish in their diet and the likelihood of prostate cancer. The results are presented in the following 2-way table.

   a. Test, at $\alpha = 0.05$, whether there is an association between the amount of fish in the diet and prostate cancer. Show at least the first term of your test statistic and use GeoGebra or StatKey.

	Large	Moderate	Small	None	Total
Prostate Cancer	42	209	201	14	466
No Prostate Cancer	507	2769	2420	110	5806
Total	549	2978	2621	124	6272

   b. If the result is or was significant, could we conclude that eating more fish causes prostate cancer? Explain why or why not.

Final Review: Which Method to Use

State which statistical method would be most appropriate for each question or situation. Highlight or underline the key words and phrases you used in each description and explain your reasoning.

Interval for a single proportion	Test for a single proportion	Chi-square test for goodness of fit
Interval for a single mean	Test for a single mean	Chi-square test for independence
Interval for a difference in proportions	Test for a difference in proportions	Chi-square test for homogeneity
Interval for a difference in independent means	Test for a difference in independent means	ANOVA
Interval for a difference of matched pairs	Test for a difference of matched pairs	Linear regression test of slope

1. Anthropologists have found two burial mounds in the same region. They know that several different tribes lived in the region and that the tribes have been classified according to different lengths of skulls. They measure the skulls found in each burial mound and wish to determine if the two mounds were made by different tribes.

2. Researchers were commissioned by the Violence In Children's Television Investigative Monitors (VICTIM) to study the frequency of depictions of violent acts in Saturday morning TV shows. They selected a random sample of 40 shows which aired during this time period over a 12-week period. Suppose that 28 of the 40 shows in the same were judged to contain scenes depicting overtly violent acts. How should they use this information to make a statement about the population of all Saturday morning TV shows?

3. The Career Planning Office is interested in seniors' plans and how they might relate to their majors. A large number of students are surveyed and classified according to their major (Natural Science, Social Science, Humanities) and future plans (Graduate School, Job, Undecided). Are the type of major and future plans related?

4. How many times a day do humans urinate, on average?

5. In one of his adventures, Sherlock Holmes found footprints made by the criminal at the scene of a crime and measured the distance between them. After sampling many people, measuring their height and length of stride, he confidently announced that he could predict the height of the suspect. How?

6. Is there an association between whether or not a person is smiling and whether or not the sun is shining?

7. What percentage of Americans support same-sex marriage?

8. Is the percentage of the national budget spent on health care associated with life expectancy for countries?

9. People were recruited for a study on weight loss, in which participants were randomly assigned to one of two groups. Group 1 was given exercise instructions and group 2 was given no exercise instructions. The researchers are interested in estimating how much more weight was lost on average by people who were given exercise instructions, as opposed to those who weren't.

10. A professional figure skater was interested in which of two jumps she landed more consistently. She did 50 double loops (landed 47 successfully) and 50 double flips (landed 48 successfully) and wanted to determine whether this was enough evidence to conclude that she had a higher success rate with one jump than another.

Final Sample Problems

To practice fully, for each problem write out the conditions making graphs where needed, show your steps using the theoretical method and check with a theoretical calculator and simulation where appropriate.

1. A professor at a large university believes that students take an average of 15 credit hours per term. A random sample of 24 students in their class of 250 reported the following number of credit hours that they were taking that term.

   Credit hours: 12, 13, 14, 14, 15, 15, 15, 16, 16, 16, 16, 16, 17, 17, 17, 18, 18, 18, 18, 19, 19, 19, 20, 21

   a. Does this sample indicate that students are taking more credit hours than the professor believes? Test an appropriate hypothesis and state your conclusion. Use alpha = .05.
   b. Form and interpret the confidence interval with the corresponding confidence level.
   c. Explain whether your 2 methods agree.
   d. If the professor wanted to lower their margin of error to 0.5 credits, how many students would they need to survey?

2. A current treatment for eczema has a 62% success rate. A new treatment method is being studied and in a sample of 40 randomly assigned subjects, 31 had success.
    a. Does this sample indicate that the new treatment is more effective than the current one? Test an appropriate hypothesis and state your conclusion. Use alpha = .02.
    b. Form and interpret the confidence interval with the corresponding confidence level.
    c. Explain whether your 2 methods agree.

3. Every year favorite songs compete to be on a Top 200 list based upon sales and rankings by experts in the music industry. A DJ wondered if there's a difference in the tempo, measured in beats per minute, of dance songs and songs in the Top 200. Two random samples of songs are shown below. None of the songs selected were on both lists.

   Dance songs (bpm): 121, 122, 117, 120, 120, 121, 118, 120, 117, 119, 121, 122, 119, 119, 118, 120, 123, 118

   Top 200 song (bpm): 122, 121, 121, 122, 121, 119, 120, 119, 120, 118, 121, 123, 118, 120, 124

   a. Test whether there is a significant difference in the tempos at a 5% level of significance. [df=30.05]
   b. Form and interpret the confidence interval with the corresponding confidence level.
   c. Explain whether your 2 methods agree.

4. Two types of medication for hives are being tested to determine whether there is a difference in the effects. The data are given in the table.

	Improved in 30 min	Did not Improve in 30 min	Total
Medication A	180	20	200
Medication B	188	12	200
Total	368	32	400

   a. Test whether there is a significant difference in the improvements at a 1% level of significance.
   b. Form and interpret the confidence interval with the corresponding confidence level.
   c. Explain whether your 2 methods agree.

5. A medical researcher wants to compare the pulse rate (beats per minute) of smokers to the pulse rate of nonsmokers. Independent samples of 100 smokers and 100 nonsmokers are selected. The results are shown below.

	Smokers	Nonsmokers
$\bar{x}$	90.0	88.0
s	5.0	6.0
n	100	100

   a. Test at $\alpha = 0.05$, if smokers have, on average, higher pulse rates than nonsmokers. [df=191.76]
   b. Form and interpret the confidence interval with the corresponding confidence level.
   c. Explain whether your 2 methods agree.

6. Before you took statistics you may have heard bad stories about statistics classes. To test whether taking a statistics class decreases students' anxieties about statistics, an instructor gave a test to rate student anxiety at the beginning and end of their course. Anxiety levels were measured on a scale of 0-10. The data are shown from 16 randomly chosen students from a class of 180 students.

    a. Test whether there is a significant decrease in the anxiety level at the end of the course at a 5% level of significance.
    b. Form and interpret the confidence interval with the corresponding confidence level.
    c. Explain whether your 2 methods agree and why.

Pre-course anxiety level	Post-course anxiety level
7	4
6	3
9	7
5	3
6	4
7	5
5	4
7	6
6	5
4	3
3	2
2	2
1	1
3	3
4	4
2	3

7. A statistics student works at a movie theater and decides to do their project on whether there is a relationship between the type of movie a party sees and whether or not they get snacks. They took a random sample of 600 customers and the data are shown in the table.

Type of Movie	Snacks	No Snacks	Total
Action	50	75	125
Comedy	125	175	300
Family	90	30	120
Horror	45	10	55
Total	310	290	600

   a. Test whether there is a relationship between the type of movie and getting snacks at the 5% level of significance.
   b. The student plans that if there is a relationship, they could make signs in the genres that don't purchase as many snacks to encourage them to buy more. Describe a type I and type II error in this context.

8. Information regarding the ACT scores of samples of students in three different majors are given below.

   a. Test with 5% significance whether there is a difference in the average ACT scores by major. Show all work & calculations.
   b. Use Stapplet with the Bonferroni correction to make MSD intervals to verify the results and if applicable, determine which group means are different.

Management	Finance	Accounting
28	22	29
26	23	27
25	24	26
27	22	28
21	24	25
19	26	26
27	27	28
17	29	20
17	28	20
23	24	
28		
28		
29		

9. The following table gives the ages (in years, x) and prices (in hundreds of dollars, y) for eight used cars of a specific model.

Age in years	Price in hundred $
8	16
3	74
6	38
9	19
2	102
5	36
6	33
3	69

   a. Determine the Least Squares Regression Line, correlation and describe the association.
   b. Calculate the predicted selling price for a 7-year-old vehicle of this model. Express your answer as a dollar amount.
   c. Test at 5% significance, if the slope of the model is statistically significant.
   d. Form and interpret the confidence interval with the corresponding confidence level.
   e. Explain whether your 2 methods agree and why.
   f. Use Stapplet to find the prediction interval if you are trying to sell a 7-year old vehicle of this model with 95% confidence.
   g. Use Stapplet to find a prediction interval for the mean of all 7-year old cars of this model for sale with 95% confidence.