**StatKey**

**Online Tools for Teaching a Modern Introductory Statistics Course**

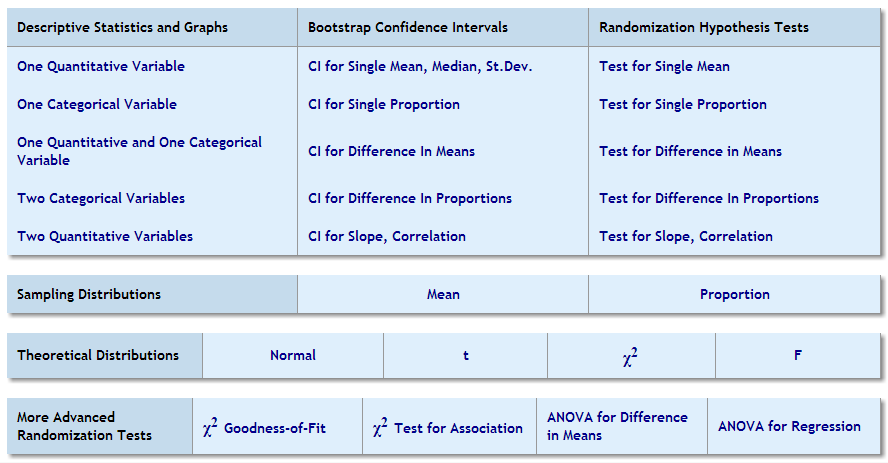
**What is it?**

A set of freely-available, user-friendly browser apps to support a simulation-based intro stat class with bootstrap distributions, randomization tests, sampling distributions, summary statistics/graphs, and graphical interfaces for standard distributions. The apps were created to accompany the text Statistics: Unlocking the Power of Data by Lock, Lock, Lock, Lock, and Lock.

**Where is it?**

*lock5stat.com/statkey* .

**What does it do?**



**Where are the data from?**

Drop-down menus provide a variety of data sets from the text.

**Can I enter my own data?**

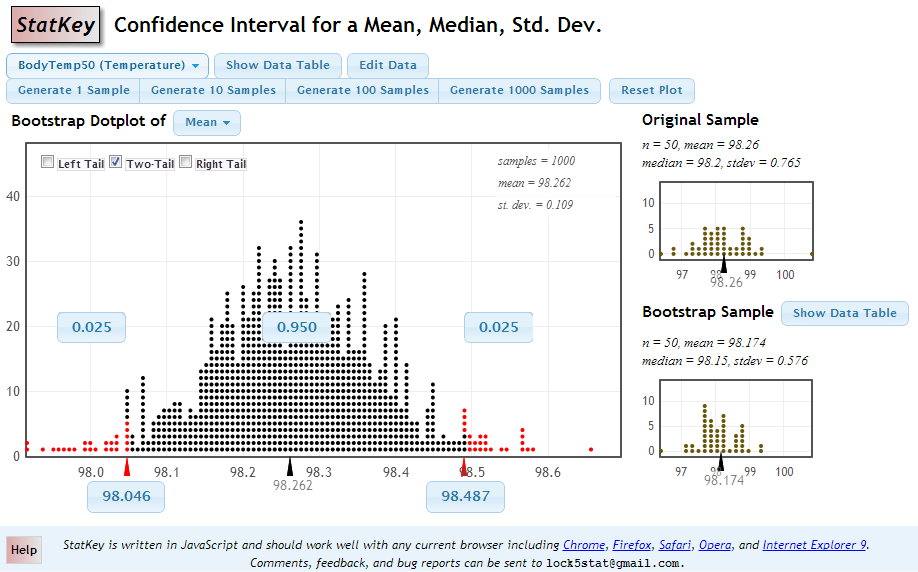
Yes. You can use the “Edit Data” button to modify the data or paste in your own data. Pasted data needs to follow the same format as the original data. For example, to compare means for two samples each row should have a group identifier in the first column and the quantitative value in the second column. Copy/paste from a spreadsheet usually works well if the column(s) are properly arranged.

**Is there help support?**

Yes. Click on the "Help" button to go to a page with support, including short videos for some common operations.

**Who created these apps?**

Designed by the Lock5 author team with computer implementation by Richard Sharp, Ed Harcourt, and Kevin Angstadt.



16

1

14

5

4

10

9

8

7

6

3

15

12

13

2

11

**Features:**

1. Display of the original sample.

2. Display of a particular bootstrap or randomization sample.

3. Dotplot of bootstrap/randomization statistics. Mouse over any dot to see the sample that produced it!

4. Choose one of the built-in datasets from the text.

5. Display or edit a table of the data.

6. Generate 1, 10, 100, or 1000 samples at a time.

7. Change the statistic, null hypothesis, or randomization method (when enabled).

8. Summary statistics for the original sample and a particular bootstrap/randomization sample.

9. Summary statistics for the bootstrap/randomization distribution.

10. Select regions in either or both tails of the bootstrap distribution. Two-tail gives equal proportions.

11. Editable values for proportions in different regions of the bootstrap/randomization distributions.

12. Editable endpoints for tail regions of the bootstrap/randomization distribution.

13. Mean for a bootstrap distribution, null parameter value for a randomization distribution.

14. Start over with a new simulation.

15. Get help (including videos) on StatKey features.

16. Return to the main StatKey menu.

**Your Turn - StatKey Confidence Intervals**

*What proportion of USCOTS participants use Google Chrome?*

For the \_\_\_\_\_\_\_\_ participants in this breakout session, how many use Google Chrome as a browser? \_\_\_\_\_\_\_\_\_\_

We'll assume this is a (random?) sample of USCOTS participants and compute a confidence interval for the proportion of all USCOTS participants who use Google Chrome.

* ***Get into StatKey (***[***www.lock5stat.com/statkey***](http://www.lock5stat.com/statkey)***) and choose the CI for a Single Proportion under Bootstrap CI's.***

Obviously, we'll need to enter the data for this particular sample.

* ***Choose the "Edit Data" button and enter the count (# of Chrome users) and sample size.***

Based on the information in the original sample, what is the sample proportion of Google Chrome users? \_\_\_\_\_\_\_\_\_\_

* ***Click on "Generate One Sample" button to create one bootstrap sample.***

This will select one sample (with replacement) from the original "sample" of yes/no responses, using the same sample size. Give the count, sample size, and proportion for this first bootstrap sample.

count = \_\_\_\_\_\_\_\_ (simulated Chrome users) sample size = \_\_\_\_\_\_\_\_ = \_\_\_\_\_\_\_\_\_\_\_\_

* ***Generate a few more bootstrap samples.***

Generate five more bootstrap samples and record the sample proportion for each in the spaces below

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Bootstrap proportions |  |  |  |  |  |

This should start to give you an idea about how proportions for this sample size can vary, but we'll get a lot better idea if we generate a lot of bootstrap samples.

* ***Generate a bootstrap distribution with 1,000 (or more) proportions for simulated samples.***

Draw a very rough graph showing the general shape of this distribution. Include a numeric scale for the horizontal axis.

-------------------------------------------------------------------------------------------------------------

* ***Find the standard deviation of the bootstrap proportions and use it to find a confidence interval.***

The standard error of the proportions is SE = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Use this with the proportion from the original sample to find an interval using .

* ***Check the "Two-tail" option to find the percentiles for a 95% confidence interval.***

Write down the lower and upper bound for the interval.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ to \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Write a sentence that interprets the interval in the context of this problem.

How does this interval based on percentiles of the bootstrap distribution compare to the one based on SE?

*CI for Slope: Price vs. Mileage for Used Mustangs*

We have prices (in $1,000) for a sample of n=25 used mustang cars and also know the mileage (in 1,000's) for each car.

* ***Go back to the main StatKey menu. (i.e. click on the word StatKey)***
* ***Choose CI for Slope, Correlation and use the drop down menu to find the Mustang Price data.***

Looking at the information for the original sample, what is the least squares line for predicting Price based on Mileage?

* ***Generate one bootstrap sample.***

The slope for this bootstrap sample is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* ***Generate a bootstrap distribution (with at least 1000 samples) for the slope of the regression line.***

Use the percentiles of the bootstrap distribution to find a 98% confidence interval for the slope of the line to predict Mustang *Price* based on *Mileage*.

If you are done early, try to find a 90% confidence interval for the *standard deviation* of the Mustang prices.

**Your Turn - StatKey Randomization Tests**

*Split or Steal*

A popular British TV show called *Golden Balls* features a final round where two contestants each make a decision to either split or steal the final jackpot. If both choose “split” they share the prize, but if one chooses “split” and the other picks “steal” the whole prize goes to the player who steals. If both choose “steal”, they both win nothing. Some researchers[[1]](#footnote-1) collected data from 287 episodes, each with two participants, to give 574 “split” or “steal” decisions. Some results are displayed in the table below, broken down by the age of the participant.

|  |  |  |  |
| --- | --- | --- | --- |
| Age group | Split | Steal | Total |
| Under 40 | 187 | 195 | 382 |
| Over 40 | 116 | 76 | 192 |
| Total | 303 | 271 | 574 |

We use these data to test if there is a significant difference in the proportions who choose “split” between younger and older players. H0: p1=p2 vs. Ha: p1≠ p2 where p1 and p2 are the proportions choosing split in the respective age groups.

* ***Open the Test for Difference in Proportions application in StatKey.***
* ***Use the ‘Edit Data’ button to input the "split "counts and sample sizes for both groups.***

What is the difference in sample proportions, in the original data? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* ***Generate a randomization distribution (at least 1000 values) for the difference in proportions.***

Each dot represents the difference in proportions when the 574 choices in the original sample are split randomly to put 382 in the "Under 40" group and the other 192 in the "Over 40" group. Since this is done at random, there is no preference for a particular split or steal response to go into either age group (thus satisfying H0).

Hover over a few of the dots in the randomization distribution to see the randomized counts that produced that particular difference.

* ***Compute a p-value from the randomization distribution.***

We are interested in how unusual it is to see a difference as large as that observed in the sample, when H0 is true.

Since this is a two-tail alternative (Ha: p1≠ p2), choose the “Two-tail” option. By default, StatKey places 5% in the tails. To find the proportion of randomized differences more extreme than the observed difference, click on the *endpoint* (not the proportion) in the appropriate tail to change its value to match the original sample difference in proportions. The p-value is then the proportion in both tails (which is twice the proportion in one tail).

p-value = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Can we conclude that the proportion of participants who choose to “split” is different between younger and older players?

*Uniform Malevolence and Penalty Yards for NFL Teams*

*Is there a significant positive relationship between the perceived malevolence of NFL (football) uniforms and penalty yards called against teams?*\*

|  |  |  |
| --- | --- | --- |
| NFL\_Team | NFL\_Malevolence | ZPenYds |
| LA Raiders | 5.10 | 1.19 |
| Pittsburgh | 5.00 | 0.48 |
| Cincinnati | 4.97 | 0.27 |
| New Orleans | 4.83 | 0.10 |
| Chicago | 4.68 | 0.29 |
| Kansas City | 4.58 | -0.19 |
| Washington | 4.40 | -0.07 |
| St. Louis | 4.27 | -0.01 |
| NY Jets | 4.12 | 0.01 |
| LA Rams | 4.10 | -0.09 |
| Cleveland | 4.05 | 0.44 |
| San Diego | 4.05 | 0.27 |
| Green Bay | 4.00 | -0.73 |
| Philadelphia | 3.97 | -0.49 |
| Minnesota | 3.90 | -0.81 |
| Atlanta | 3.87 | 0.30 |
| Indianapolis | 3.83 | -0.19 |
| San Francisco | 3.83 | 0.09 |
| Seattle | 3.82 | 0.02 |
| Denver | 3.80 | 0.24 |
| Tampa Bay | 3.77 | -0.41 |
| New England | 3.60 | -0.18 |
| Buffalo | 3.53 | 0.63 |
| Detroit | 3.38 | 0.04 |
| NY Giants | 3.27 | -0.32 |
| Dallas | 3.15 | 0.23 |
| Houston | 2.88 | 0.38 |
| Miami | 2.80 | -1.60 |

The key variables are *NFL\_Malevolence* as rated by subjects viewing the team uniforms (higher values are more malevolent) and *ZPenYds*, a scaled measure (z-score) of the number of penalty yards for the team.

If we let ρ be the correlation between malevolence and penalties, the hypotheses are H0: ρ=0 v. Ha: ρ>0.

* ***Find the NFL Malevolence data in the Test for Correlation.***

What is the correlation between NFL\_Malevolence and ZPenYds for the original sample?

*r =* \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Whether or not this is large enough that we can conclude there is a positive association between these variables is what the test should assess.

* ***Create one randomization sample***

Record its correlation in the first spot of the table below. Also click on “Show Data Table” for the sample to see how the *ZPenYds* values have been scrambled among the teams to match the null hypothesis.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Randomization correlations |  |  |  |  |  |

Click on “Generate 1 sample” a few more times and write down the correlations for the randomization samples in the table above. Did you get any correlations as large as *r* from the original sample?

* ***Create a randomization distribution with at least 1000 correlations***

Where is the randomization distribution of sample correlations centered? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* ***Find an approximate p-value***

Since this is an upper tail alternative (Ha: ρ>0), choose the “Right tail” option. We are interested in how unusual it is to see a correlation as large (or larger) than *r*=0.43 (when H0 is true). Click on the endpoint for the right tail and change the correlation to 0.43 to match what was observed in the original sample. The p-value is the proportion in the right tail beyond 0.43.

p-value = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

What does this p-value indicate about the hypotheses in this situation?

\* Frank and Gilovich, “The Dark side of Self- and Social Perception: Black Uniforms and Aggression in Professional Sports”, *Journal of Personality and Social Psychology* Vol. 54, No. 1 p. 74-85 (1988)

**StatKey- More Advanced Randomization Tests**

*Do Ants Have Preferences for Different Types of Sandwiches?*

Some students did an experiment to count the number of ants attracted to three different types of sandwiches. Use the drop-down menu to find the data in the "ANOVA for Difference in Means" option in the Advanced Randomization tests in StatKey.

* ***Verify the table below for the mean number of ants attracted to each sandwich type***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Vegemite | Peanut Butter | Ham & Pickles | Overall |
| Sample size | 8 | 8 | 8 | 24 |
| Mean | 30.8 | 34.0 | 49.3 | 38.0 |
| Std. dev. | 9.3 | 14.6 | 10.8 | 13.9 |

You can also click on the "Show Data" button to see the counts for each of the 24 original samples.

* ***Click on "ANOVA Table" next to the original sample to show the ANOVA table.***

What is the value of the F-statistic for the ANOVA table? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Note that this value is also displayed above the table of summary statistics for the original sample.)

Would this be an unusual value to see if there really were no difference between the attractiveness of the three sandwich types? That is what the randomization test will investigate.

* ***Generate one randomization sample.***

The ant counts are scrambled and randomly assigned to the sandwich types (8 each) to create a sample that might occur under a null hypothesis that the sandwich type doesn't matter.

Which sandwich type had the largest mean in the randomization sample? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ mean = \_\_\_\_\_\_\_

What is the value of the F-statistic for the randomization ANOVA table? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* ***Generate lots of randomization samples to create a randomization distribution of F-statistics***

What shape do you see for the distribution of randomization F-statistics for this sample?

* ***Estimate a p-value***

Use the “Right Tail” option in StatKey to find the proportion of randomization F-statistics at (or beyond) the F-statistic for the original sample (click on the endpoint to adjust the value to match the original sample).

p-value ≈ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Based on this p-value, what decision would you make for a 5% test?

* ***Compare to the F-distribution***

Use the F-distribution in StatKey with 2 and 21 degrees of freedom to find the area beyond F=5.627.

p-value = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Is the p-value similar to what you obtained via simulation? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

How does the shape of the theoretical F-distribution compare to the randomization distribution of simulated

F-statistics?

*Does temperature matter in hatch rate of python eggs?*

The two way table below arises from an experiment where 187 python eggs were randomly assigned to either a cool, neutral (room temperature), or warm (incubator) environment, with expected values shown in parentheses.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Hatch? | |  |
|  |  | Yes | No | Total |
| Temperature | Cool | 16 (18.6) | 11 (8.4) | 27 |
| Neutral | 38 (38.6) | 18 (17.4) | 56 |
| Warm | 75 (71.7) | 29 (32.3) | 104 |
|  | Total | 129 | 58 | n=187 |

Ho: Hatch rate is not related to temperature

Ha: Hatch rate is related to temperature

The value of the χ2-statistic is

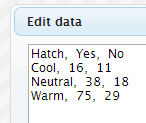
The p-value using the upper tail of a χ2-distribution with 2 d.f. is 0.427, so we do not have enough evidence to conclude that hatch rate depends on temperature.

------------------------------------------------------------------------------------------------------------------------------

**How can we do this as a randomization test?**

Under the null hypothesis (temperature doesn’t matter) we’ll assume that the 129 eggs that hatched would hatch no matter which temperature they were assigned to and same for the 58 eggs that didn’t hatch.

Using StatKey, we can randomly re-assign the eggs to the temperature groups (keeping 27 cool, 56 neutral and 104 warm), find a new two-way table, and compute its χ2-statistic. We repeat this many times to get a randomization distribution of chi-square statistics, all done when the null hypothesis is true.



* ***Choose the “χ2-Test for Association” in StatKey.***

The python egg data is not one of the existing datasets, so you will need to edit the data to enter the results from the two-way table, using the “Edit Data” button, deleting what is there, and then entering the data as shown on the right.

Check that the counts and χ2-statistic match the original sample. (You can also click on “Show Details” to confirm the expected counts above.)

* ***Generate one randomization sample.***

Click on “Generate 1 Sample” to randomly re-assign the eggs to the temperature groups. Note that the row and column totals are the same for the new sample as the original one, but the yes/no values have been re-distributed among the temperatures.

What is the χ2-statistic for your first randomization sample? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* ***Generate lots of randomization samples***

You know the routine now…

Draw a rough sketch of the randomization distribution of χ2-statistics (when H0 is true) below.

* ***Find the p-value***

What proportion of your randomization samples have a χ2-statistic that is as (or more) extreme than the original data? (Show where these are on the sketch above.)

* ***Check the theoretical p-value***

Use the χ2-distribution in StatKey (with 2 d.f.) to verify the p-value given above for χ2=1.70.

Does the result appear to match the randomization distribution well?

1. Van den Assem, Martijn J., Van Dolder, Dennie and Thaler, Richard H., Split or Steal? Cooperative Behavior When the Stakes Are Large (February 19, 2011). Available at SSRN: http://ssrn.com/abstract=1592456 [↑](#footnote-ref-1)