

Examples: Bootstrap Confidence Intervals

USCOTS, May 2019

Example 1: Atlanta Commutes

The American Housing Survey (AHS) regularly collects data from samples of people in different metropolitan areas. A survey conducted in Atlanta asked about the distance (in miles) and time (in minutes) for daily commutes. We use a sample of 500 Atlanta commuters to find a 95% confidence interval for mean commute time for all Atlanta commuters.

- Find StatKey at www.lock5stat.com/statkey.
- Choose the CI for Single Mean option in the Bootstrap Confidence Intervals column.
- Get the Atlanta Commute (Time) data from the drop down menu of datasets.

What is the mean commute time for the original sample? _____

- Generate a single bootstrap sample using the “Generate 1 Sample” button.

What is the mean for this simulated bootstrap sample? _____

There are two very large commute times in the original sample. Are either or both in the bootstrap sample you generated? _____

- Generate a full bootstrap distribution using the “Generate 1000 Samples” button several times.

Use the “Two-Tail” option to construct a 95% confidence interval for the mean time of all Atlanta commutes:

Lower bound: _____ Upper bound: _____

What is the Standard Error from the bootstrap distribution? _____

Use the mean from the original sample and the standard error from the bootstrap distribution, and the formula “Statistic \pm 2 · SE,” as an alternate way to construct the 95% confidence interval:

Lower bound: _____ Upper bound: _____

(The two answers should be very similar.)

Example 2: How do you prefer to get your news?

“Whether offline or online, do you prefer to get your news by watching it, reading it, or listening to it?” Several months ago, the Pew Research Organization asked this question of 3,425 US adults, and 1,164 of them said they preferred reading it. We use this sample to estimate a 90% confidence interval for the proportion of all US adults who prefer reading the news.

- **Click on the “StatKey” button to return to the main StatKey menu.**
- **Choose the “CI for Single Proportion” option in the Bootstrap Confidence Intervals column.**
- **This data is not in the drop-down menu. Select “Edit Data” and put in the data (a count of 1164 in a sample size of 3425).**

What is the proportion for the original sample? _____

- **Generate a single bootstrap sample using the “Generate 1 Sample” button.**

What is the proportion for this simulated bootstrap sample? _____

- **Generate a full bootstrap distribution using the “Generate 1000 Samples” button several times.**

Use the “Two-Tail” option, and then change 0.950 to 0.90, to construct a 90% confidence interval for the proportion of all US adults who prefer to read their news:

Lower bound: _____ Upper bound: _____

Example 3: Does Diet Cola leach calcium from the body?

In a study to investigate the effect of diet cola on calcium levels in the body, sixteen women were randomly assigned to drink 24 ounces of either a diet cola or water. The amount of calcium (in mg) in their urine was measured for the next three hours. Use the results of this experiment to find a 99% confidence interval for the difference in mean amount of calcium excreted between women who drink diet cola and women who drink water. (*The data are in StatKey as “Cola and Calcium Excretion.”*)

The 99% confidence interval for the difference in mean amount of calcium excreted is:

Lower bound: _____ Upper bound: _____

Bonus Question:

Use StatKey to find a 95% confidence interval for the *standard deviation* of prices for used Mustangs offered for sale online (based on the same sample of 25 cars that we used to estimate mean price).

Examples: Randomization Hypothesis Tests

USCOTS, May 2019

Example 1: Light at Night

Is exposure to light at night harmful to health? In one experiment, eighteen mice were randomly assigned to either a normal light/dark cycle or an environment with a light on around the clock. Many measurements of health were made, and we look here specifically at body mass gain (in grams) over the several weeks of the experiment. We are interested in whether there is evidence that mice with a light on at night have greater mean body mass gain.

- **Find the “Mice Wgt Gain 2e” data under Test for a Difference in Means in StatKey.**

What is the difference in sample means, $\bar{x}_1 - \bar{x}_2$, in the original data: _____

- **Generate one randomization sample (based on a null hypothesis of no difference).**

What is the difference in means for this simulated sample? _____

- **Generate a randomization distribution for the difference in means.**

Notice the shape and center of your randomization distribution.

Find the approximate location of the difference in sample means from the original data.

Using the “Right-Tail” option, and the difference in sample means from the study:

What is the p-value? _____

What can you conclude?

Example 2: Malevolent Uniforms

Uniforms for all teams in the National Football League were rated on a malevolence score, and a z-score was calculated from total penalty yards over several seasons for each team. We are interested in whether these results provide evidence that the true correlation between these two variables is positive.

- **Find the “Malevolent Uniforms” data under Test for a Correlation in StatKey.**

What is the sample correlation? _____

- **Generate a randomization distribution for the correlation (based on a null hypothesis of a correlation of zero).**

What is the p-value? _____

What can you conclude?

Example 3: 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 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
40 and over	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.

- ***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 proportion split for each age group? Under 40 _____ 40 and over _____

What is the difference in sample proportions in the original data? _____

- ***Generate a randomization distribution for the difference in proportions.***

Based on your randomization distribution, what is the p-value? _____

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

Bonus Question: Is the average body temperature 98.6°F?

The Test for a Single Mean procedure in StatKey has a dataset on body temperatures measured for a sample of 50 college students. Does this sample provide evidence that the mean body temperature differs from the traditional 98.6°F? (For example, perhaps global warming is also leading to people warming!)