Practice Session 5

In this practice section we will introduce the bootstrap distribution, bootstrap confidence interval, and hypothesis testing. You may use the functions: `sample()`, `do_it()` to generate the bootstrap distribution. `SDS1000::cnorm` and `qnorm` to find the critical value corresponding to a specific **confidence level**.

Part 1: Confidence interval concept

Practice 1:

True or False/ Confidence interval interpretation

A catalog sales company promises to deliver orders placed on the Internet within 3 days. Follow-up calls to a few randomly selected customers show that a 95% confidence interval for the proportion of all orders that arrive on time is 85% $\pm$ 5%.

1.) Between 80% and 90% of all orders arrive on time.

2.) 95% of all random samples of customers will show that 85% of orders arrive on time.

3.) The interval between 80% and 90% gives a plausible range of values for where the true population parameter lies since 95% of intervals created will contain the population proportion.

4.) For a given sample size, higher confidence means a larger margin of error.

5.) For a specified confidence level, smaller samples provide smaller margins of error.

Part 2: Construct Bootstrap Distribution

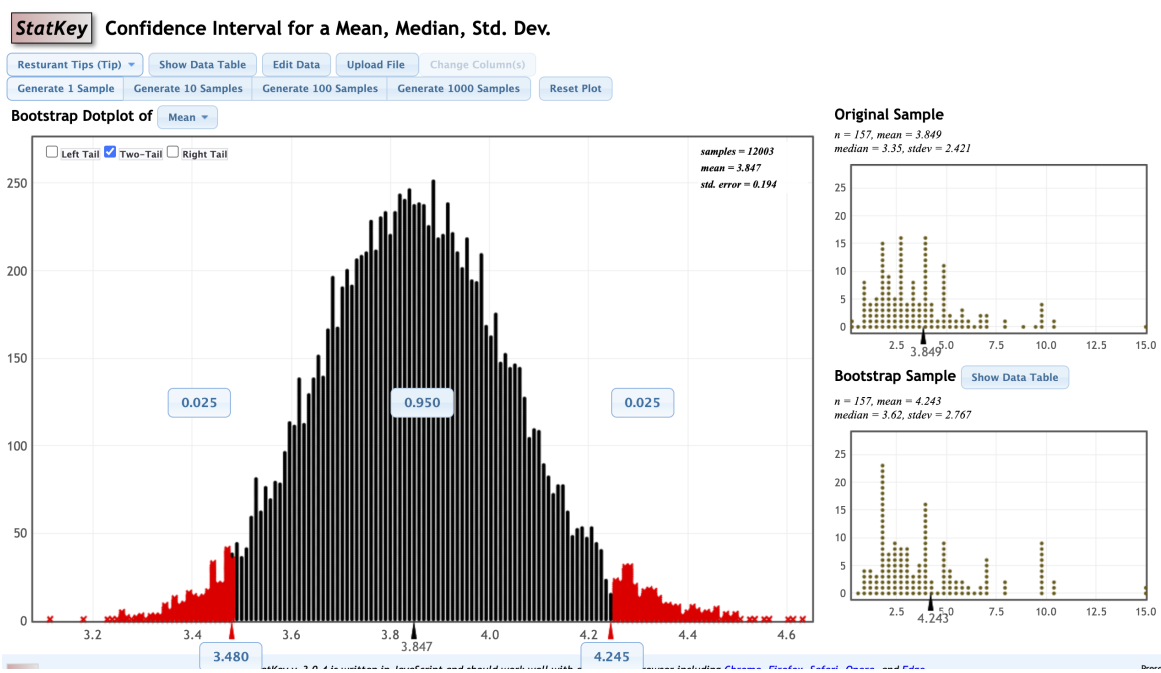
Here's the clever idea: We don't have the population, but we have a sample. Probably the sample is similar to the population in many ways.So let's sample from our sample. We'll call it `bootstrapping`. We want samples **the same size** as our original sample, so we will need to **sample with replacement**.This means that we may pick some members of the population more than once and others not at all. We'll replicate this many times.

Generating a Bootstrap Distribution:

- Generate bootstrap samples by **sampling with replacement** from the original sample, using the **same sample size**.
- Compute the **statistic of interest** (called a bootstrap statistic), for each of the bootstrap samples.
- Collect the **samples statistics** for many bootstrap samples to create a **bootstrap distribution**.

Example:

Using `StatKey`website link. Try to play with the creation of the bootstrap distribution from different data. The following picture shows the website and an example of bootstrap CI for the variable `tip` from the dataset `Restaurant tips` . `https://www.lock5stat.com/StatKey/bootstrap_1_quant/bootstrap_1_quant.html`



Practice 2:

The data `ExerciseHours` provide an in-class survey of statistics students asking them about the amount of exercise per week.

1.) **First**, create `histogram` of `Exercise`. What is the `sample size` of the data ?

```
library(Lock5Data)
library(SDS1000)
data(ExerciseHours)

# your code here #
```

2.) **Second**, create a one bootstrap sample from `Exercise` . You might use the functions `sample()`.

```
# your code here #
```

3.) **Third**, you might replicate this one sample `10000` times `with replacement` using the function `do_it()` to create a `bootstrap sampling ditribution`

```
# your code here #
```

4.) **Fourth**, create a `histogram` for the bootstrap distribution of the variable `Exercise`

```
# your code here #
```

Congrats! you created bootstrap sampling distribution from one sample !

Part 3: Construct Bootstrap Confidence Interval

Reminder:

The steps to Construct Bootstrap Confidence Interval are:

1) Compute the statistic from the original sample.
2) Create a bootstrap distribution by re-sampling from the sample.

- Same size samples as the original sample.
- With replacement.
- Compute the statistic for each sample.
- The distribution of these statistics is the bootstrap distribution.

3) Estimate the standard error SE by computing the standard deviation of the bootstrap distribution.
4) Create the 95% CI using the formula: $statistic \pm 2 * SE$.
5) Interpret the confidence interval within the context.

Practice 3:

From the `ExerciseHours` bootstrap sampling distribution you have created in the previous question, create a 95% CI for the the sample mean of `Exercise`.

1.) **First**, calculate the mean of `Exercise` from your original sample

```
# your code here #
```

2.) **Second**, create the bootstrap sampling distribution of the `Exercise`.

```
# your code here #
```

3). **Third**, calculate the standard error of our bootstrap sampling distribution of the `Exercise`.

```
# your code here #
```

4.) **Fourth**, calculate the 95% CI, which is based on the formula: $statistic \pm 2 * SE$

```
# your code here #
```

Congrats! you created a 95% CI with bootstrap distribution !

5.) **Fifth**. Interpret the confidence interval within the context.

Part 4: Extra Practice/ Create Bootstrap Confidence Interval with Different Confidence Levels

Practice 4:

Note: You might use the function `SDS1000::cnorm()` or ,`qnorm()` to help you find the critical values.

1.) Create a 90% CI of the mean `Exercise` in `R`.

```
## a 90\%  CI of the mean Exercise

# your code here #
```

2.) Calculate a 99% confidence interval of the mean `exercise` using the function `qnorm()`.

```
## a 99\% confidence interval of the mean Exercise


# your code here #
```

3.) Compare the three confidence intervals you have obtained from the three different confidence levels: 95% , 90%, and, 99% .

Part 5: Extra Practice/ Introduction to Hypothesis testing

Statistical Tests:

A statistical test is used to determine whether results from a sample are convincing enough to allow us to conclude something about the population.

We have two competing claims about the population, the **null hypothesis**, denoted by `H0`, and the **alternative hypothesis**, denoted by `Ha`.

Practice 5:

State the null and alternative hypotheses for the statistical test described:

1.) Testing to see if there is evidence that a mean is less than 50.

2.) Testing to see if there is evidence that a proportion is greater than 0.3.

3.) Testing to see if there is evidence that the mean of group A is not the same as the mean of group B.

4.) Testing to see if there is evidence that the correlation between two variables is positive.