Chapter 4

Statistical Inference Using Simulation

The Role of Probability in Statistical Inference

Probability: In the discussion so far, did you notice that we cheated? We always assumed that probabilities and distributions of r.v.'s were *known*.

For example "Let X be the number of cells with [...]. Assume $X \sim \text{Bin}(n, p)$, with $p = 0.6$. Find $\Pr(X \geq 100)$."

In a real experiment, e.g., carried out in a lab, you *never know* p! Or the rate λ of the service time for a customer in the post office etc. This is where statistics comes in.

Statistics: We ask "Using an observed value x for $X \sim \text{Bin}(n, p)$, can we guess what p could be?", or "Could p be greater 0.5?"

That is, we change perspective. Assuming we have a good description of the experimental data as a r.v.'s, we try to report inference on the parameters, like p etc.

The change of perspective between probability and statistics naturally gives rise to some different vocabulary in statistics.

- *Sample*: a set of r.v.'s $X = \{X_i, i = 1, \ldots, n\}$, $X_i \sim F$, i.i.d. that are actually observed. We also just refer to X as the *data*.

- *Statistic*: any function of the observed data, $S = f(X_1, \ldots, X_n)$. In probability we would have simply said S is another r.v. Of course, it is both!

- *Parameters:* This is tricky. In some pars of statistics (Bayesian statistics), parameters are r.v.'s that are not observed.
  In other contexts they index (describe) the distribution of the data. We usually use greek letters, like λ, μ, well, or p.

- *Hypothesis:* a hypothesis is simply an event for the parameters, like $A = \{\mu > 0\}$ (and we might use different names like H_0 etc.)

In summary, probability describes uncertainty, whereas statistics is about decisions in the face of such uncertainty.

Transition

Probability

We describe uncertainty.

Statistics:

We make decisions
in the face of uncertainty.

Example: Do We Like to Think? 1

A psychologist did an experiment where college students were asked to sit for 10 minutes in a room alone, just thinking, with no cellphones, books, etc.) Then they were asked whether the experience was pleasant or unpleasant.

13 of 25 said that

the experience was unpleasant.

Two different types of questions can be explored:

Example: Do We Like to Think? 2

1. **Estimate the population proportion** of college students at that university who would rate such an experience unpleasant.

2. **Test the claim** that the population proportion of college students at this university who would who find such an experience unpleasant is greater than 40%

Using Probability to Address These

- How can we answer this using probability calculations and/or statistical methods?
- What methods?
- What additional assumptions are needed about the way the study was conducted to use those methods?
- How "robust" is our answer against slight deviations from each of these conditions? (And how do we measure "robust" and "slight?")

statistic

confidence intervals,
hypothesis testing

Estimating a Parameter

- **Standard method**:
  Form and interpret a confidence interval for the parameter.
- Pick some confidence level that is large enough for the user to consider it adequate to make their decision.

Testing a Claim: Method

- **Standard method**:
  - State a claim of interest.
  - Decide on a counter-claim that is, more or less, “status quo.”
  - Do a probability calculation:
    Find the p-value: the probability of obtaining data that is as extreme or more extreme than the data that we have, if the”status quo” is really true.
  - Decide how to act,
    based on that probability.

Assumptions

- What assumptions?
  - Sample has to be “large enough.” (Depends on the statistical method.)
  - Possible actions:
    - Generalize to some population. For this, the sampling method must be a probability-based method (simple random sampling, stratified sampling, etc.)
    - Consider this evidence for causality. For this, we must have an experiment where the values of the explanatory variable are randomly assigned to the subjects.

Obtain Information about the study

- How was it designed?
- How was it implemented?
- If the report doesn’t make all that clear, how much do you know about the source of the study, and the reputation of the source?

How can we ever do anything with statistics?

- Be aware of the assumptions.
- Be aware that they are seldom “exactly” met.
- Typical statistical practice includes many guidelines that help restrict application of the methods to situations where they are “robust” against reasonable deviations from them.

The Value of Deeper Understanding

- A good understanding of the basics for the statistical calculations is very valuable in understanding HOW the particular results depend on the assumptions.
- This is particularly useful in thinking about robustness of the results despite some deviations from some of the required assumptions.

Sampling Distributions

Sampling Distribution of a Statistic

The sampling dist’n of the statistic for individual samples of size *n* for a particular population is constructed by

- Taking all possible samples of size *n* from that population
- Computing the value of the statistic from each of those samples
- Collecting all of those values of the statistic to comprise the sampling dist’n of the statistic from that population.

Using the Sampling Dist’n of a Statistic

From here, generally one would

- compute summary statistics of that sampling dist’n

  and / or

- graph it.

Then, make inferences about the parameter based on that information.

Our Data and Summary Statistic

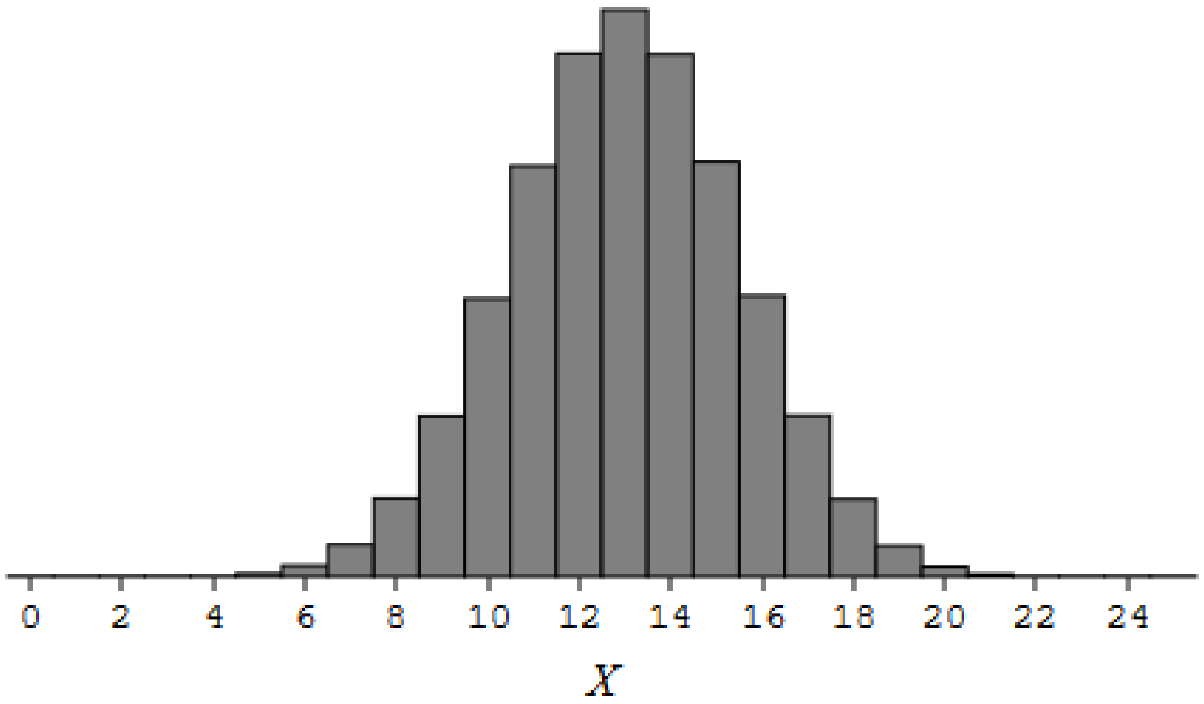
- Data from 25 individual students, whose results can be considered Bernoulli random variables.
- Total "successes" can be considered to be Binomial(25,p)
  - What assumptions are we making about the data to do this? (same p, independent)
- Statistics: making decisions. . . .
  What does that mean here?

Sampling dist’n of the sample proportion

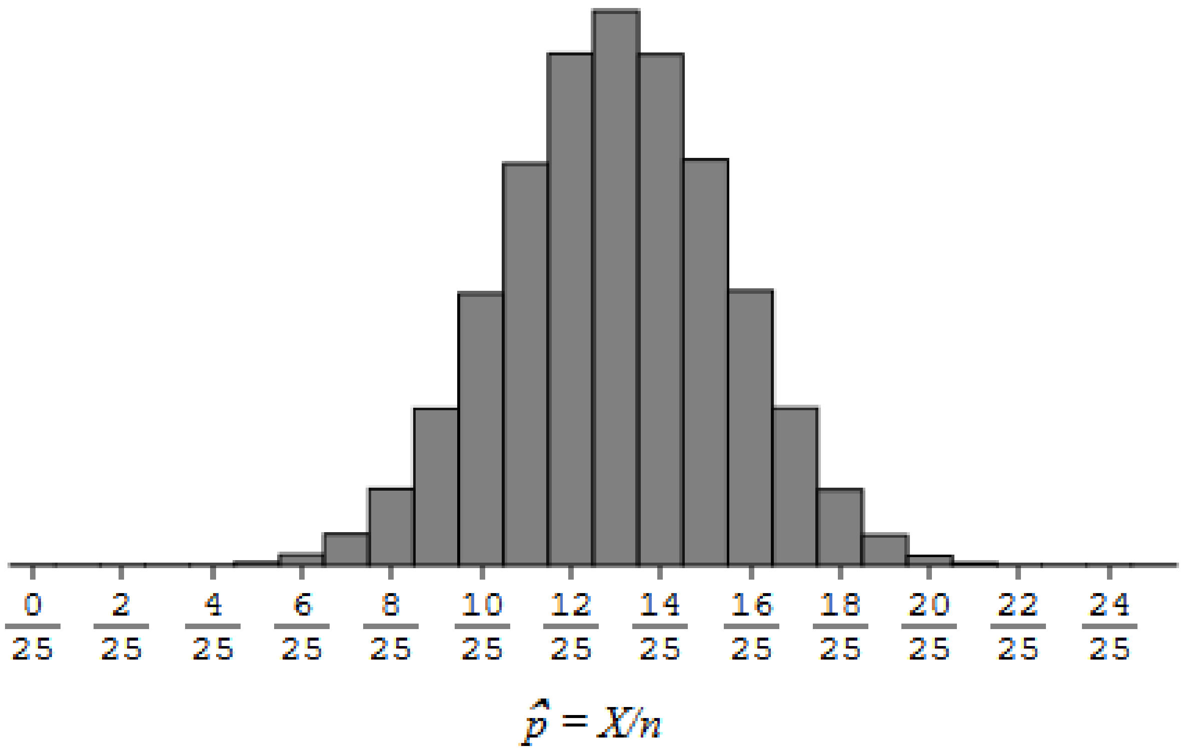
- On the next two slides we are reminded that the actual probabilities of the distribution are the same whether the random variable is called
- The number of “ successes”

  or
- The sample proportion of “successes”

Sampling Dist’n for Count



Sampling Dist’n for sample proportion



Which summary statistic do we use?

- To estimate the pop'n parameter *p*, it makes sense to use a summary statistic that gives us good numbers to "plug in" for *p*.
- We could "plug in" p-hat values for *p*, since the values of p-hat are "good approximations" of *p*.
  (In Math Stat terms, "p-hat is an unbiased estimator of *p*.")

Exercise

Stop the video to review probability work.

1. What are the mean and variance of X?
2. Derive the mean and variance of p-hat = X/n .

3. After you have done these, resume the video to see the solutions.

Exercise Solutions

What are the mean and variance of X?

Mean of X = np   Variance of X = np(1-p)

Derive the mean and variance of p-hat = X/n .

Mean of X/n = np/n = p

and Variance of X/n = $(1/n^2)np(1-p)$ = p(1-p)/n

Remember that **these are quite different formulas**, despite the fact that the useful information in the two dist'ns is the same.

Will it always be this easy?

- No.
- Because the dist’n of the sum of the individual independent r.v.’s (Bernoulli) was a known dist’n (Binomial)
  and
- the dist’n of the statistic we want is a constant multiple of the sum,
  then
- we had all we needed to find the sampling dist’n of the statistic.

What do we do next?

- There are very few situations in which finding the sampling dist’n of the statistic we want is as easy as that.

- In those situations, we go back to “basics” to think of how we might use computing power to simulate a sampling distribution of a statistic.

How strong is the evidence?

Back to basics: Simulation

Yes, we can "see" sampling distributions with a very general and useful method.

To do this, we will use a collection of tools that will always be freely available on the web.

StatKey www.lock5stat.com

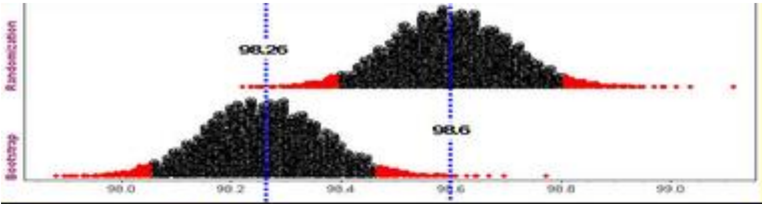
(Also has many datasets, with descriptions, and instructions of using it in "Videos.")

Learn to use StatKey

Ideally, replicate the StatKey work as you watch the videos.

If something isn't clear, stop and write a question.   Keep track of your questions as you go through this. (And any answers to them you figure out!)

Ask your questions at least three times a week.



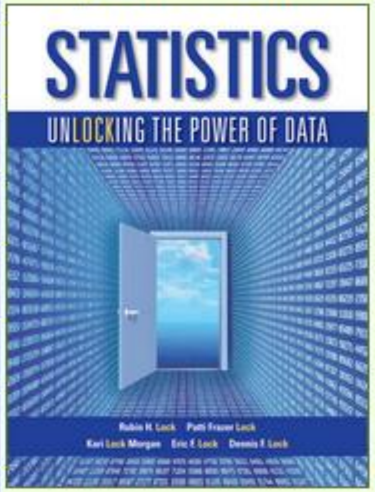
Statistics: Unlocking the Power of Data

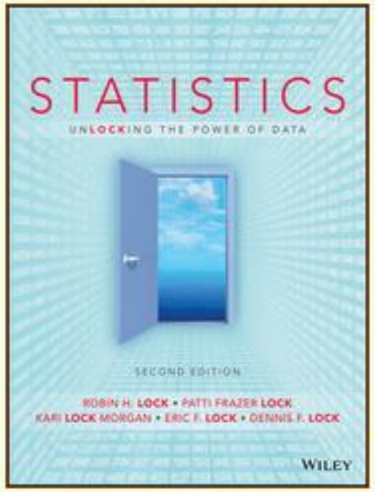
by Lock, Lock, Lock, Lock, and Lock

October 18, 2020

- Home
- Authors
- Overview (pdf)
- Table of Contents
- Datasets
- StatKey
- StatKey help
- Videos
- Presentations
- Other Resources

Companion Materials for





First Edition (2013)
Follow this link for more information at the Wiley site.

Second Edition (2017)
Follow this link for more information at the Wiley site

... or contact Michael MacDougald at Wiley.

Find an interactive unit from the text and more information about the approach at the Wiley showcase site.

Robin Lock
St. Lawrence University

Patti Frazer Lock
St. Lawrence University.

Kari Lock Morgan
Penn State University

Eric F. Lock
University of Minnesota

Dennis F. Lock
Iowa State University and Buffalo Bills

StatKey to accompany *Statistics: Unlocking the Power of Data*
by Lock, Lock, Lock, Lock, and Lock

Descriptive Statistics and Graphs	Bootstrap Confidence Intervals	Randomization Hypothesis Tests
One Quantitative Variable	CI for Single Mean, Median, St.Dev.	Test for Single Mean
One Categorical Variable	CI for Single Proportion	Test for Single Proportion
One Quantitative and One Categorical Variable	CI for Difference In Means	Test for Difference in Means
Two Categorical Variables	CI for Difference In Proportions	Test for Difference In Proportions
Two Quantitative Variables	CI for Slope, Correlation	Test for Slope, Correlation

Sampling Distributions	Mean	Proportion

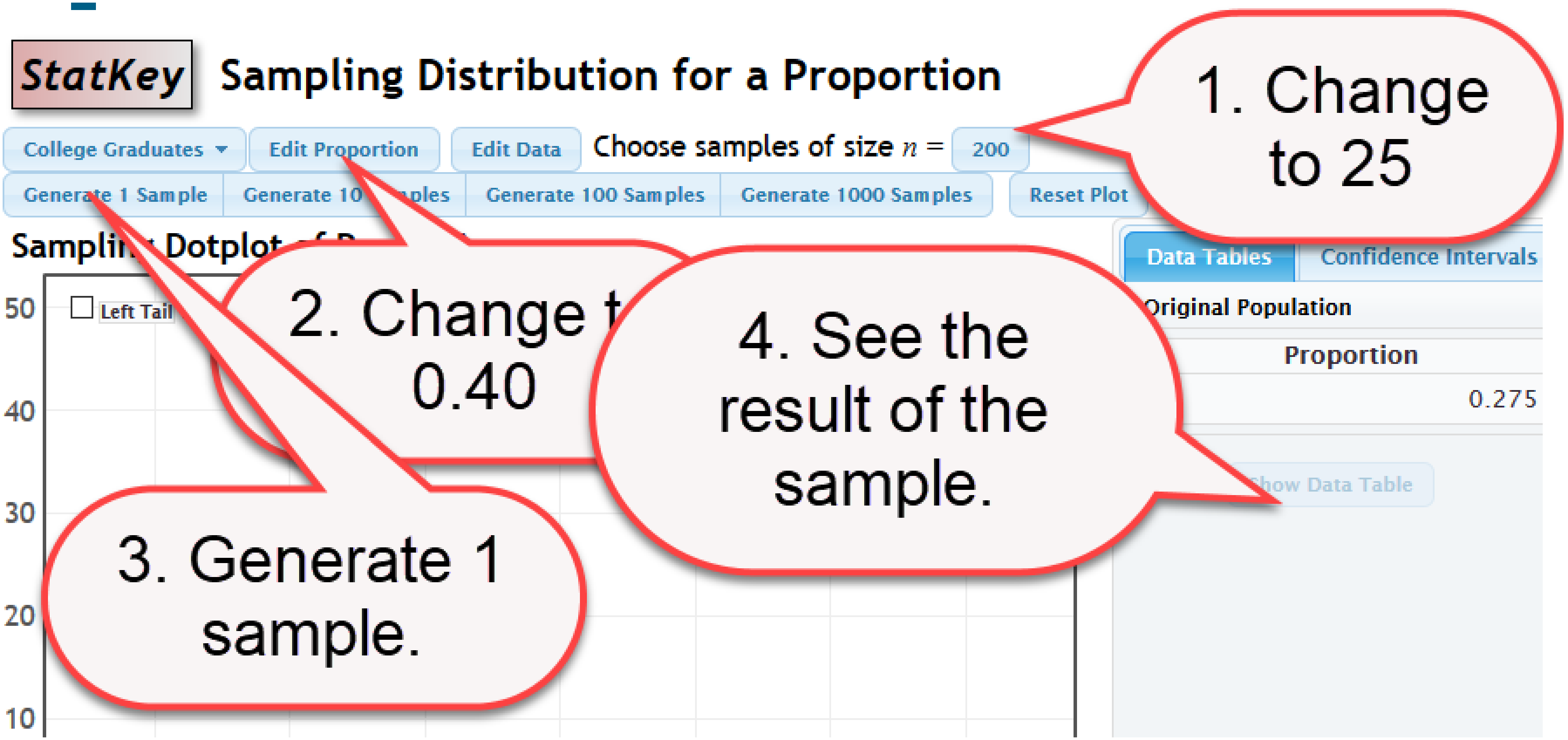
Theoretical Distributions	Normal	t	χ^2	F

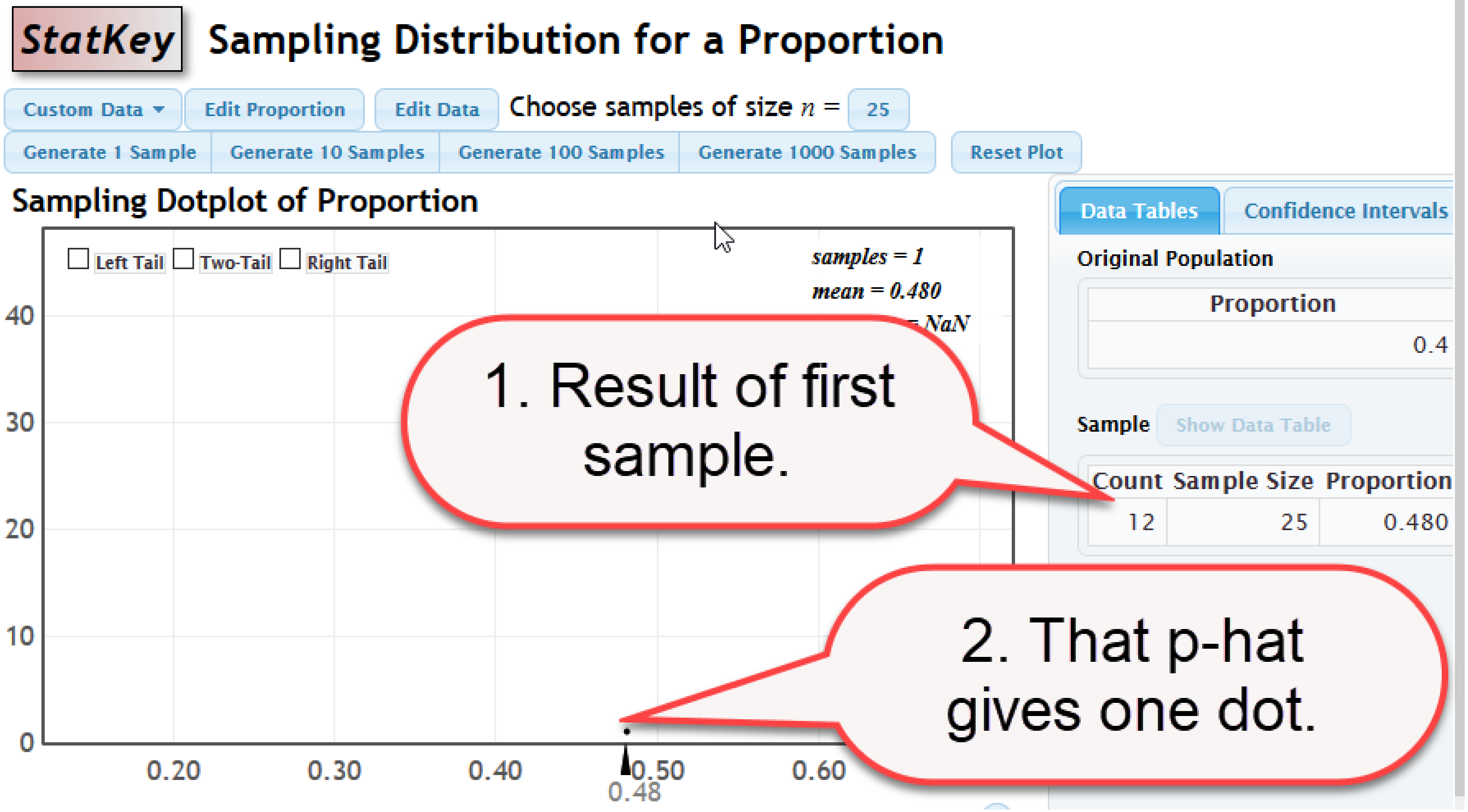
More Advanced Randomization Tests	χ^2 Goodness-of-Fit	χ^2 Test for Association	ANOVA for Difference in Means	ANOVA for Regression

Simulate a sampling dist'n

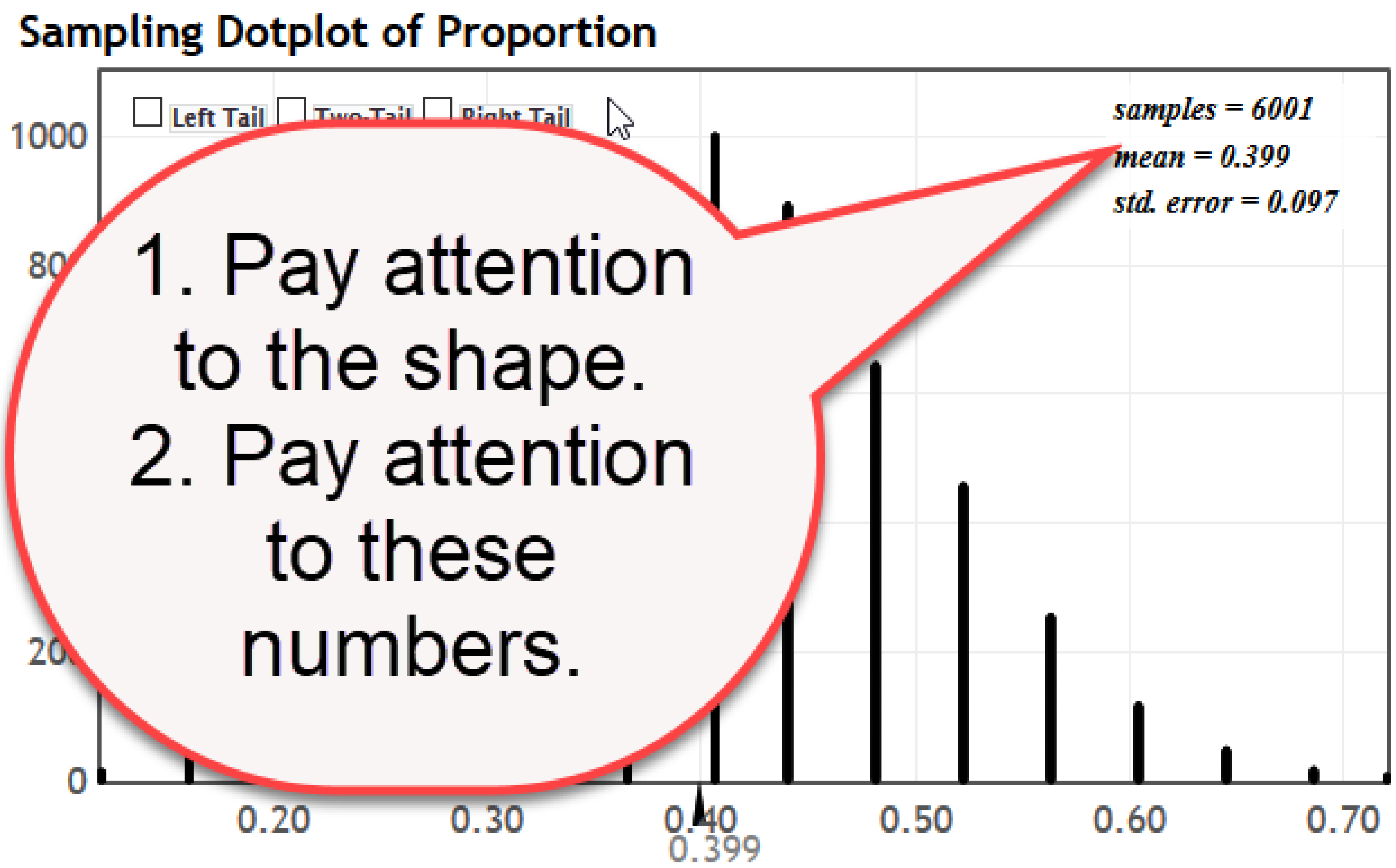
Let's simulate a sampling dist'n for the statistic p-hat, when the pop'n we're interested in is Bernoulli(0.40)

- Population   Bernoulli(p) where $p = 0.40$
- Statistic: p-hat
- Type of sampling: simple random sampling
- Size of the sample we're interested in: $n = 25$
- How many "replications" we'll do: maybe 6000

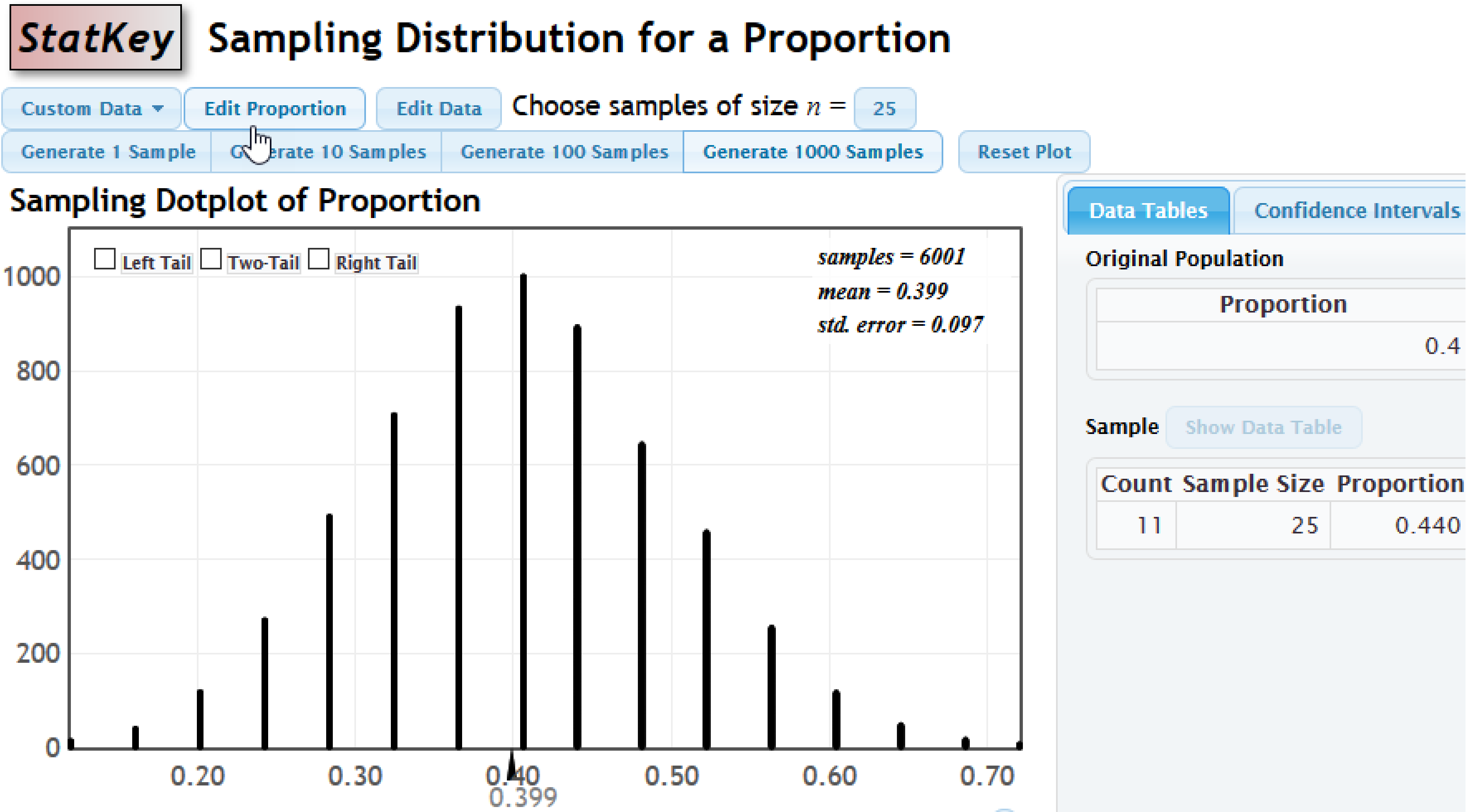

StatKey
Sampling Distribution for a Proportion
College Graduates
Edit Proportion
Edit Data
Choose samples of size n = 200
Generate 1 Sample
Generate 100 Samples
Generate 1000 Samples
Reset Plot
Left Tail
50
40
30
20
10
Data Tables
Confidence Intervals
Original Population
Proportion
0.275
1. Change to 25
2. Change t
0.40
3. Generate 1 sample.
4. See the result of the sample.


StatKey
Sampling Distribution for a Proportion
Custom Data
Edit Proportion
Edit Data
Choose samples of size n = 25
Generate 1 Sample
Generate 10 Samples
Generate 100 Samples
Generate 1000 Samples
Reset Plot
Sampling Dotplot of Proportion
Left Tail
Two-Tail
Right Tail
samples = 1
mean = 0.480
NaN
40
30
20
10
0
0.20
0.30
0.40
0.50
0.60
0.48
Data Tables
Confidence Intervals
Original Population
Proportion
0.4
Sample
Show Data Table
Count Sample Size Proportion
12 25 0.480
1. Result of first sample.
2. That p-hat gives one dot.

Generate MANY more replications.



Simulated sampling dist'n of p-hat



Comments

- Looks VERY much like the theoretical dist’n we graphed earlier for p-hat. A good approximation!
- Made 6000 replications by generating them with the 1000 button just above the graph of the dist’n.
- 6000 is an arbitrary choice.
- We didn’t choose a type of sampling. StatKey only does simple random sampling for its simulations.

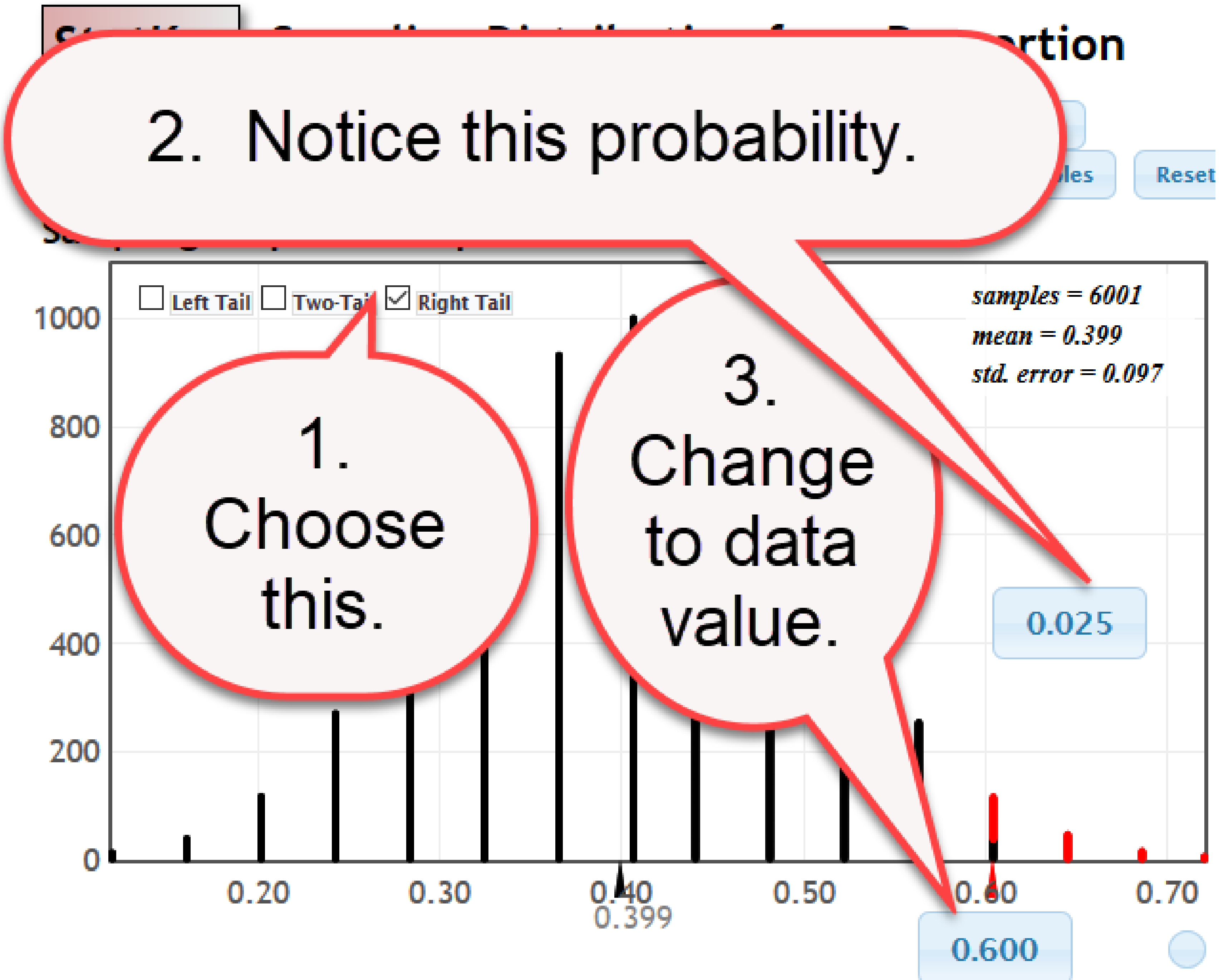
- New section begins

How strong is the evidence for….?

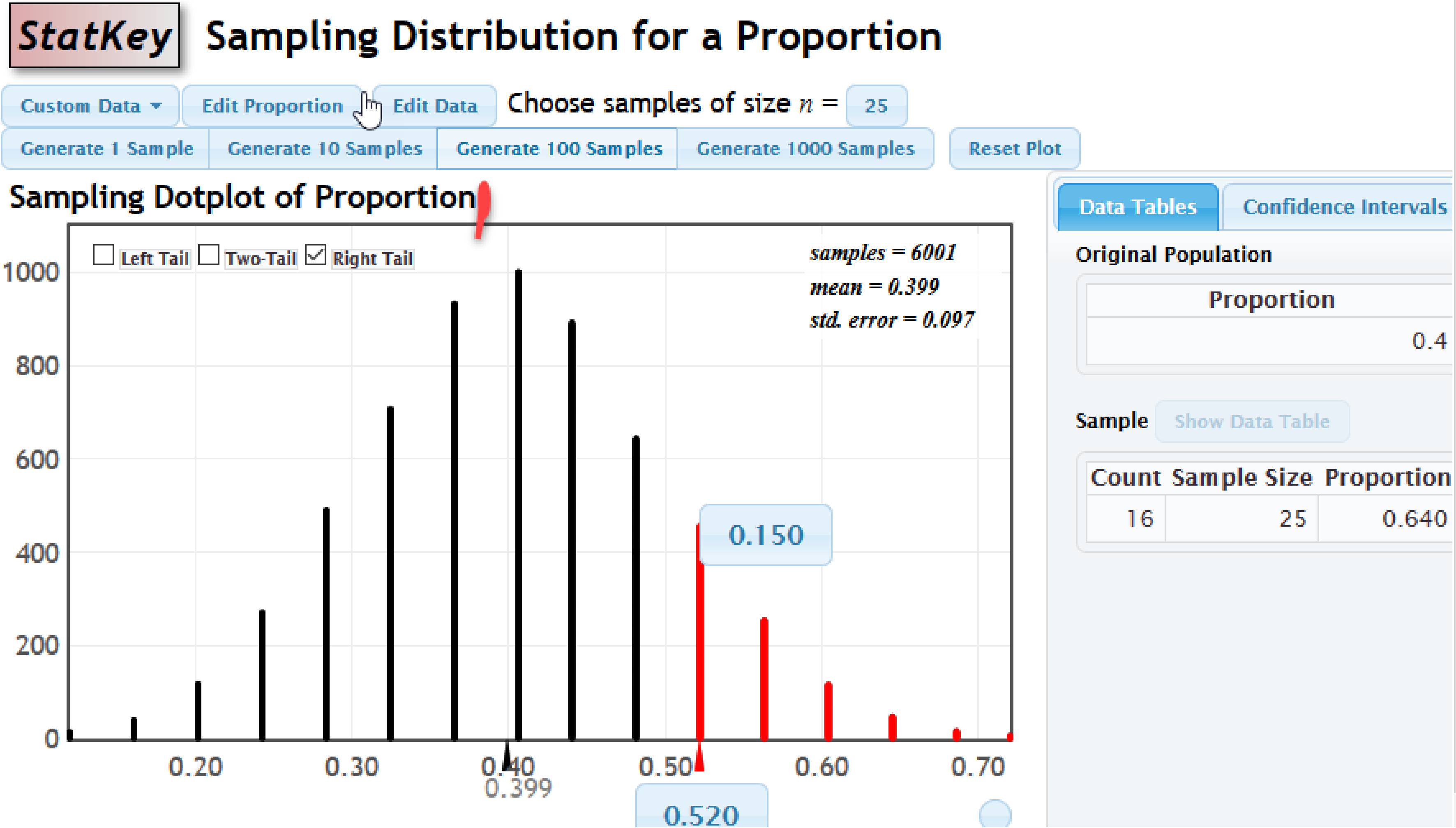
Back to one of our two example questions:

Test the claim that the population proportion of college students at this university who would who find such an experience unpleasant is greater than 40%.

So, let's look at a (simulated) sampling dist'n of the sample proportion
where p = 0.40 (the base value in our claim)
and see where our sample proportion
13/25 = 0.52 falls.



rtion
2. Notice this probability.
Reset
Left Tail
Two-Tai
Right Tail
samples = 6001
mean = 0.399
std. error = 0.097
1. Choose this.
3. Change to data value.
1000
800
600
400
200
0
0.20
0.30
0.40
0.399
0.50
0.60
0.70
0.025
0.600


StatKey
Sampling Distribution for a Proportion
Custom Data
Edit Proportion
Edit Data
Choose samples of size n = 25
Generate 1 Sample
Generate 10 Samples
Generate 100 Samples
Generate 1000 Samples
Reset Plot
Sampling Dotplot of Proportion
Left Tail
Two-Tail
Right Tail
samples = 6001
mean = 0.399
std. error = 0.097
0.150
0.520
0.399
Data Tables
Confidence Intervals
Original Population
Proportion
0.4
Sample
Show Data Table
Count Sample Size Proportion
16 25 0.640

How strong is the evidence that $p > 0.40$?

Our evidence (p-hat= 0.52) seems much larger than 0.40. But then, the sample size isn't very large.

In terms of probability, about 15% of the values in the (simulated) sampling dist'n of the sample proportion, when $p = 0.40$ are that large or larger.

There is some evidence, but weak evidence, that the population proportion is larger than 0.40.

Do you remember hypothesis testing?

Ho: p =0.40 versus Ha: p > 0.40

The sampling distribution we just used is the distribution to use for this hypothesis test, and the p-value of the data we found is 0.150.

On the following page is the output for this problem from the StatKey tool for a hypothesis test of one proportion. It shows a p-value of 0.156.

This is, of course, consistent with our previous answer.

(Not exactly the same, but, of course, this is a simulation.)

In terms of probability, about 15% of the values in the (simulated) sampling dist'n of the sample proportion, when p = 0.40 are that large or larger.

There is some evidence, but weak evidence, that the population proportion is larger than 0.40.

StatKey to accompany *Statistics: Unlocking the Power of Data*
by Lock, Lock, Lock, Lock, and Lock

Descriptive Statistics and Graphs	Bootstrap Confidence Intervals	Randomization Hypothesis Tests
One Quantitative Variable	CI for Single Mean, Median, St.Dev.	Test for Single Mean
One Categorical Variable	CI for Single Proportion	Test for Single Proportion
One Quantitative and One Categorical Variable	CI for Difference In Means	Test for Difference in Means
Two Categorical Variables	CI for Difference In Proportions	Test for Difference In Proportions
Two Quantitative Variables	CI for Slope, Correlation	Test for Slope, Correlation

Sampling Distributions	Mean	Proportion

Theoretical Distributions	Normal	t	χ^2	F

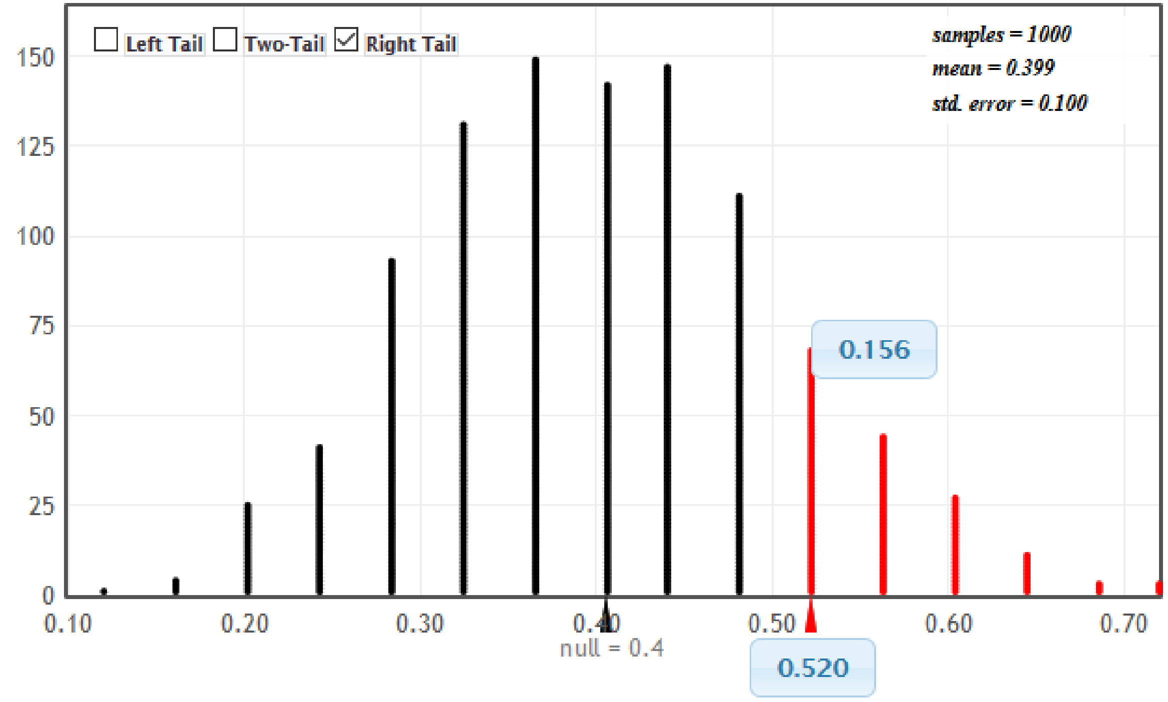
More Advanced Randomization Tests	χ^2 Goodness-of-Fit	χ^2 Test for Association	ANOVA for Difference in Means	ANOVA for Regression

StatKey Randomization Test for a Proportion

Custom Data | Edit Data

Generate 1 Sample | Generate 10 Samples | Generate 100 Samples | Generate 1000 Samples | Reset Plot

Randomization Dotplot of Proportion Null hypothesis: p = 0.4



Original Sample

Count	Sample Size	Proportion
13	25	0.520

Randomization Sample

Count	Sample Size	Proportion
13	25	0.520

StatKey Sampling Dist’n Tools

StatKey only provides a sampling dist’n tool for two statistics: sample mean and sample proportion.

It provides tools for confidence intervals and hypothesis tests for the other statistics commonly used in standard applied statistics courses.

Going Forward: Questions

Here are some questions you might be thinking about:

1. How will I learn to use StatKey (or will I use something else?)
2. Will we review hypothesis testing more?
3. What sampling dist'n can we use for a confidence interval, since we don't know the pop'n parameter to use for it?

Going Forward: Remarks

1. Yes, you may use other simulation software (or write some yourself.) In the course materials, I will provide documents and links to materials to help with StatKey.
2. Yes, I will demonstrate using StatKey to do a hypothesis test for a difference of two proportions.
3. Immediately after this slide, we will address confidence intervals using simulation.

Estimating a Parameter

XXXSteps: Hypothesis Testing in StatKey

1. In the right-hand column (Randomization Tests,) choose the appropriate parameters.
2. Put in the data (for proportions, use Edit Data and put in the count and the sample size.)
3. Put in the Ho value.
4. Generate many replications.
5. Choose the appropriate tails.
6. Put the test statistic value on the horizontal axis.
7. Read the p-value from the graph.

Estimating a parameter

Estimating a Parameter

Now, let's shift our attention to the other question:

Estimate the population proportion of college students at that university who would rate such an experience unpleasant.

Estimate a population proportion

- Our single-point estimate of the population proportion is our sample proportion.

- We know that our sample with sample proportion 13/25 = 0.52 could easily have come from a population with population proportion of 0.53 or 0.50 or other values close to it.
  - How close?
  - And with what probability?

Estimate a population parameter

- In statistical inference, we want to estimate parameters with an interval estimator. That is, we want to give a conclusion like this:
- "I have 90% confidence that the true population parameter is between __ and ___."
- So we need to go some way on either side of our sample proportion.
  - How far?
  - And what does 90% have to do with how far?
  - Should we always go the same distance in either direction? Why or why not?

How to we DO it?

- Generally speaking, we reinterpret that as:

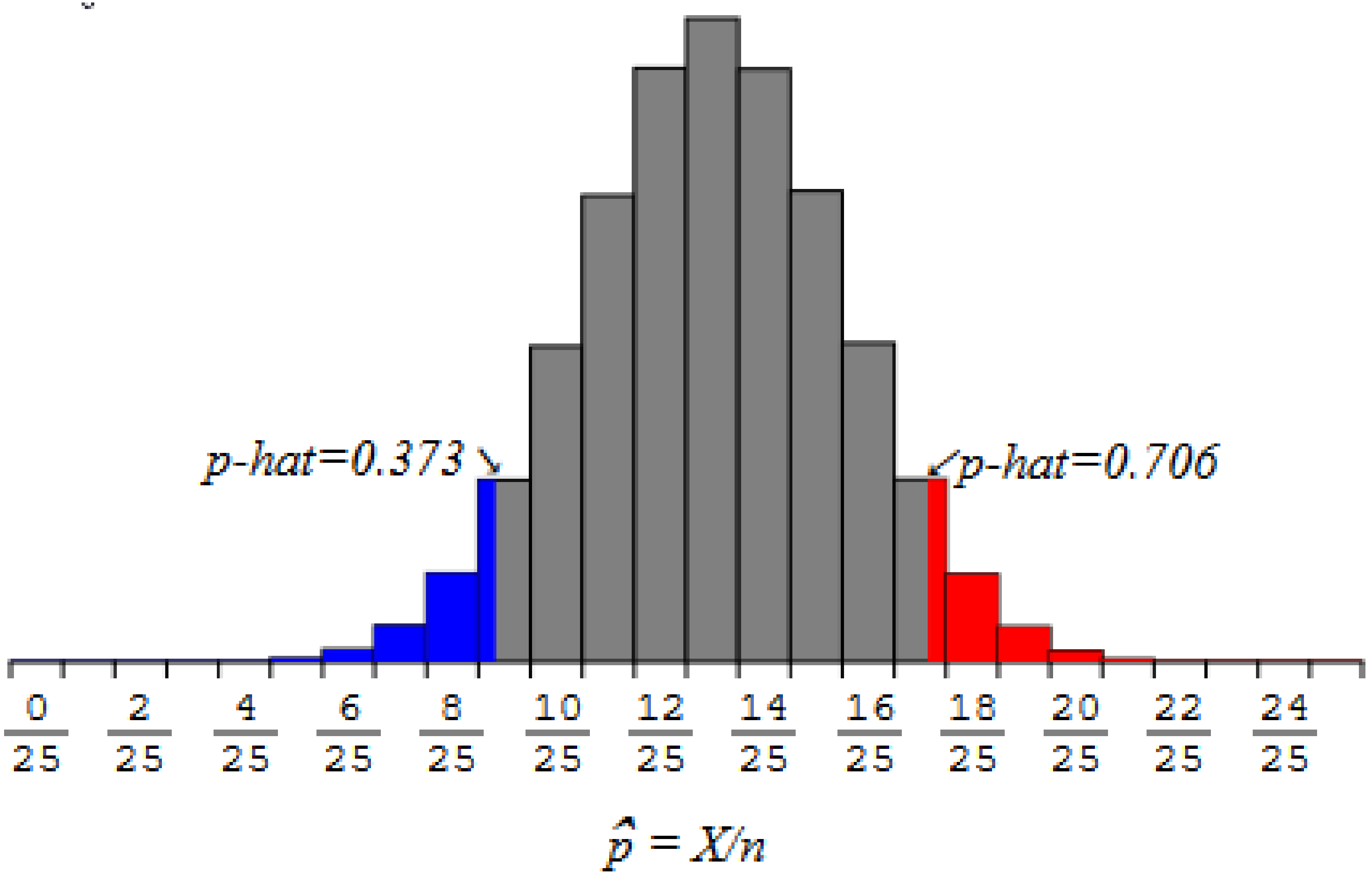
  What are the middle 90% of the values of the sampling dist’n of *p*,
  when we use our sample proportion (rather than our theoretical idea of what the proportion is)
  to create the sampling dist’n?

- We call that our 90% confidence interval for the population proportion, based on this sample.

How do we DO this in a theoretical dist’n?

- Since our sample proportion is 0.52 and our sample size is 25,
- we simply create a Bin(25,0.52) distribution
- and pick some endpoints that cut off the outer 10% of the values,
- which we would usually do by cutting off 5% on each end.
- The following graph has 90% of the area in the middle and 5% cut from each end.

- The 90% confidence interval is 0.373 to 0.706.



p-hat=0.373
p-hat=0.706
0/25
2/25
4/25
6/25
8/25
10/25
12/25
14/25
16/25
18/25
20/25
22/25
24/25
$\hat{p} = X/n$

Why were some of the bars split?

- In this discrete distribution, you saw that, to get 5% on each end, we had to “split” some bars, which seems a bit strange.
- On the other hand, it is also strange to limit our view of possible endpoints to just the values along the horizontal axis that outline the “bars” indicating the probabilities of the various integer values of X.
- The difficulty is, of course, that our population parameter p is defined on the entire interval from 0 to 1, but our sample proportion in this distribution only takes on 26 different values.

Sidelight: What we REALLY do in a theoretical dist’n

- In practice, our theoretical approach is to avoid this issue: **approximate this with a continuous distribution and compute probabilities in it**. (If the conditions for that approximation are met.)
- As you may remember from a previous statistics course, we approximate this with a normal dist’n, matching the mean and variance of the normal dist’n to the mean and variance of this discrete dist’n.

How do we DO it using simulation? 1

- We use an idea called “Bootstrapping.”
- (Do you know the saying “Pull yourself up by your bootstraps”? It was a 19th century saying to describe doing something by your own efforts that might have seemed impossible.)

- Let’s start a “picture” of how we do it. Here are the 25 values from our sample:

0, 0 ,0, 0, 0, 0 ,0, 0, 0, 0 ,0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

0, 0 ,0, 0, 0, 0 ,0, 0, 0, 0 ,0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

0, 0 ,0, 0, 0, 0 ,0, 0, 0, 0 ,0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

0, 0 ,0, 0, 0, 0 ,0, 0, 0, 0 ,0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

0, 0 ,0, 0, 0, 0 ,0, 0, 0, 0 ,0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

0, 0 ,0, 0, 0, 0 ,0, 0, 0, 0 ,0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

0, 0 ,0, 0, 0, 0 ,0, 0, 0, 0 ,0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

0, 0 ,0, 0, 0, 0 ,0, 0, 0, 0 ,0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

0, 0 ,0, 0, 0, 0 ,0, 0, 0, 0 ,0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

0, 0 ,0, 0, 0, 0 ,0, 0, 0, 0 ,0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

0, 0 ,0, 0, 0, 0 ,0, 0, 0, 0 ,0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

0, 0 ,0, 0, 0, 0 ,0, 0, 0, 0 ,0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

0, 0 ,0, 0, 0, 0 ,0, 0, 0, 0 ,0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

How do we think about it using simulation?

- We think of our sample as "just like" the population. (After all, it's all the information we really have to use about our population.)
- So we, essentially, duplicate our sample over and over until we get MANY values – enough to be as big as the population.
- And then we sample from that –
  - taking 25 values at random
  - to compute **one** sample statistic as our first value for the bootstrap dist'n.
- We sample from that MANY times until we get enough (what do you think? 1000, 6000, 9000?) to feel comfortable that it's a large enough bootstrap distribution

  to **simulate the variability** in the of the sampling dist'n of the sample proportion.

How do we ACTUALLY do it?

- To actually DO, this, we simply take simple random samples of size 25, **with replacement**, from the original sample. (Of course, we have computer software do this!)
- And, each of those bootstrap samples has a sample proportion. We collect all of those values into what we call a bootstrap distribution for the sample proportion.
- In that distribution, we look at the middle 90% and find the endpoints to form the 90% confidence interval.

StatKey to accompany *Statistics: Unlocking the Power of Data*
by Lock, Lock, Lock, Lock, and Lock

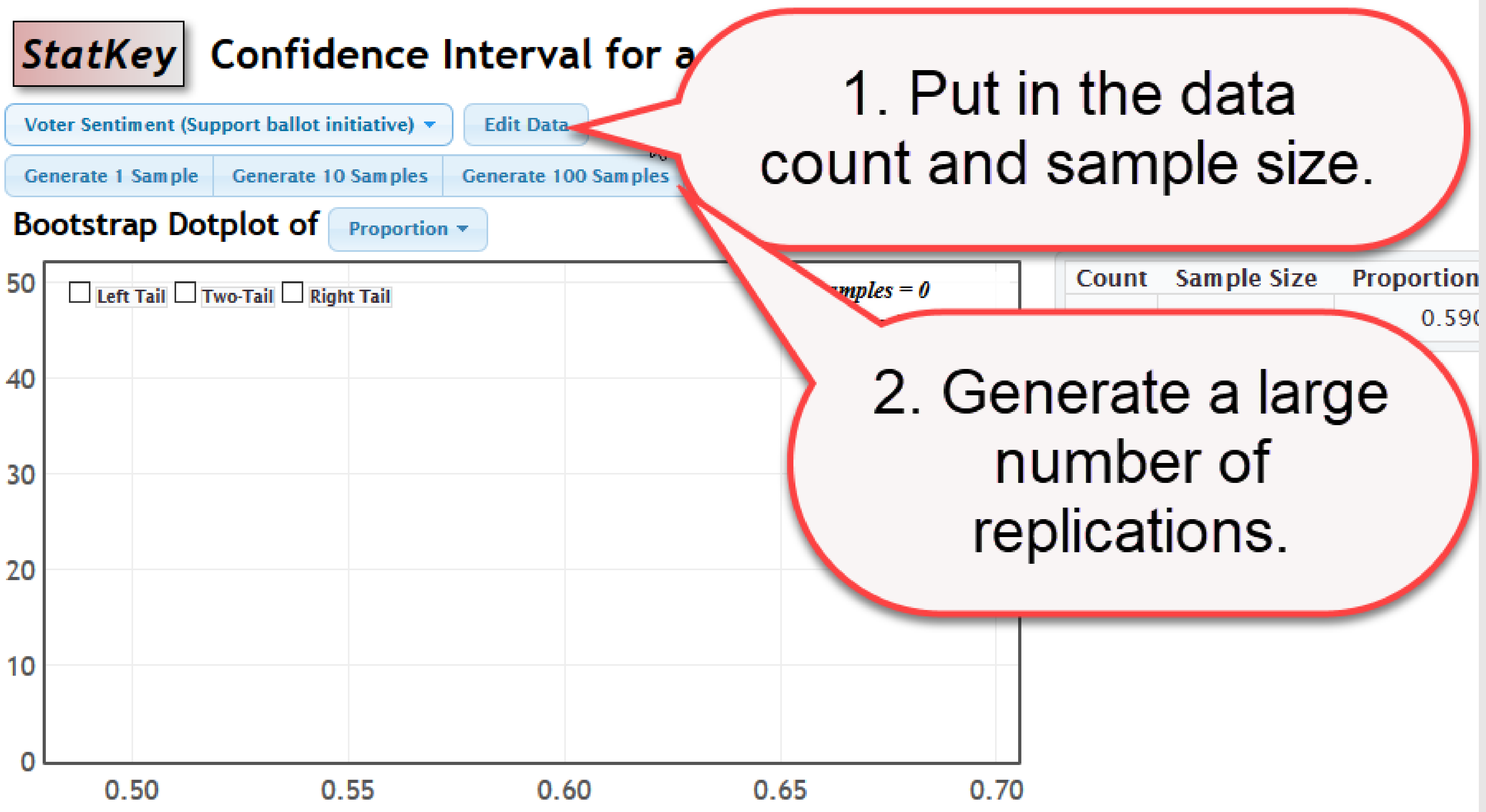
Choose this.

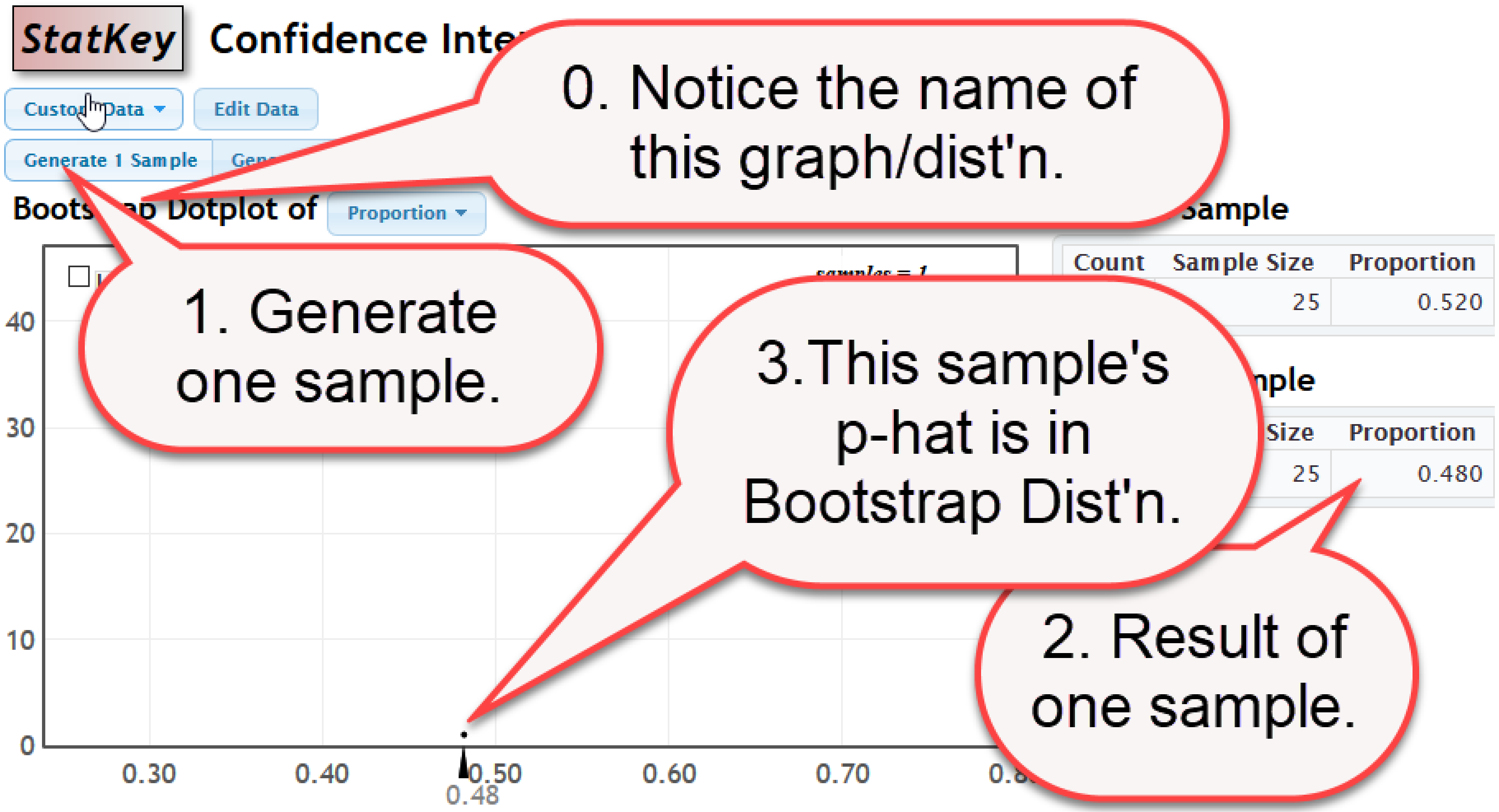
Descriptive Statistics and Graphs	Bootstrap Confidence Intervals	ypothesis Tests
One Quantitative Variable	CI for Single Mean, Median, St	ean
One Categorical Variable	CI for Single Proportion	Test for Single Proportion
One Quantitative and One Categorical Variable	CI for Difference In Means	Test for Difference in Means
Two Categorical Variables	CI for Difference In Proportions	Test for Difference In Proportions
Two Quantitative Variables	CI for Slope, Correlation	Test for Slope, Correlation

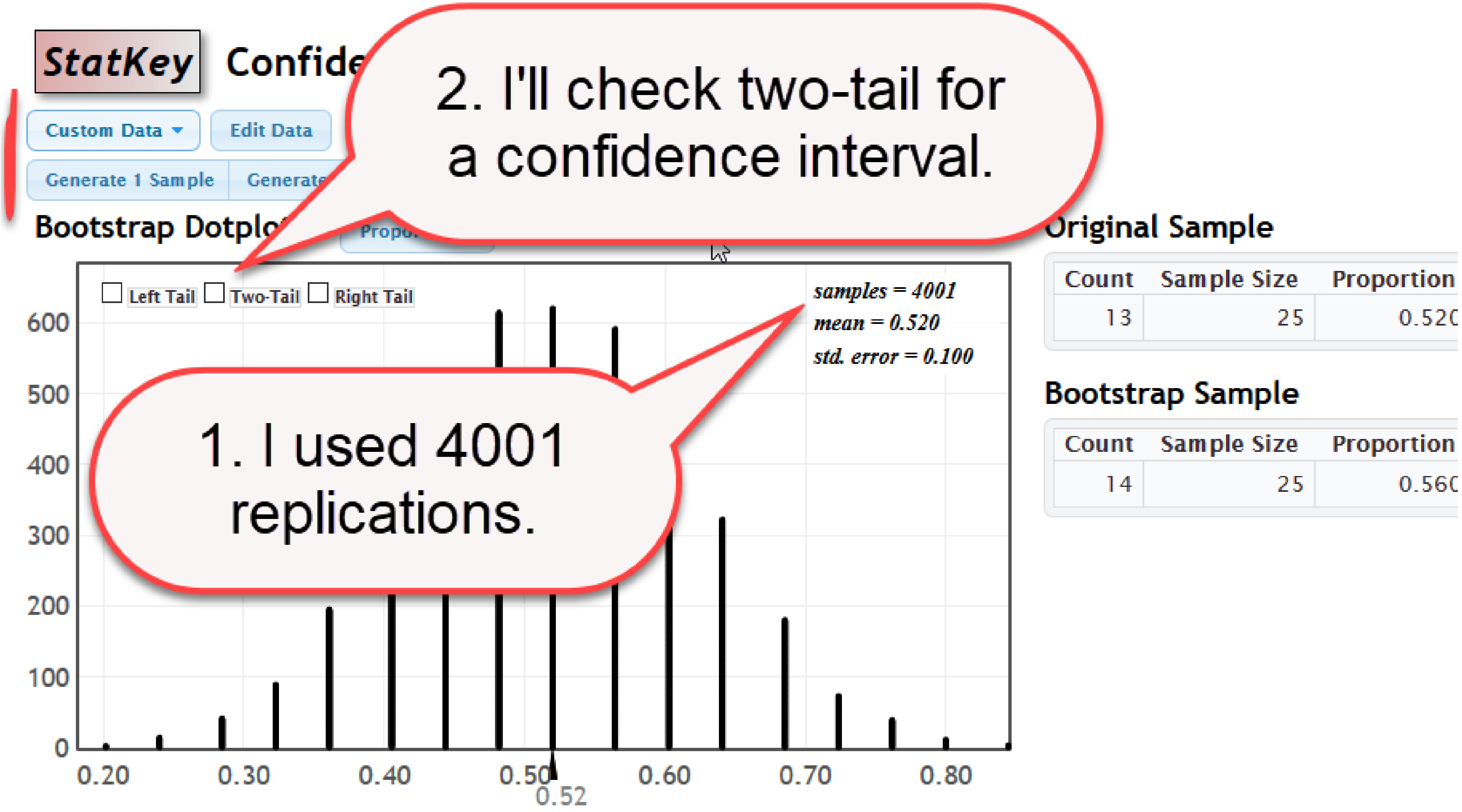
Sampling Distributions	Mean	Proportion

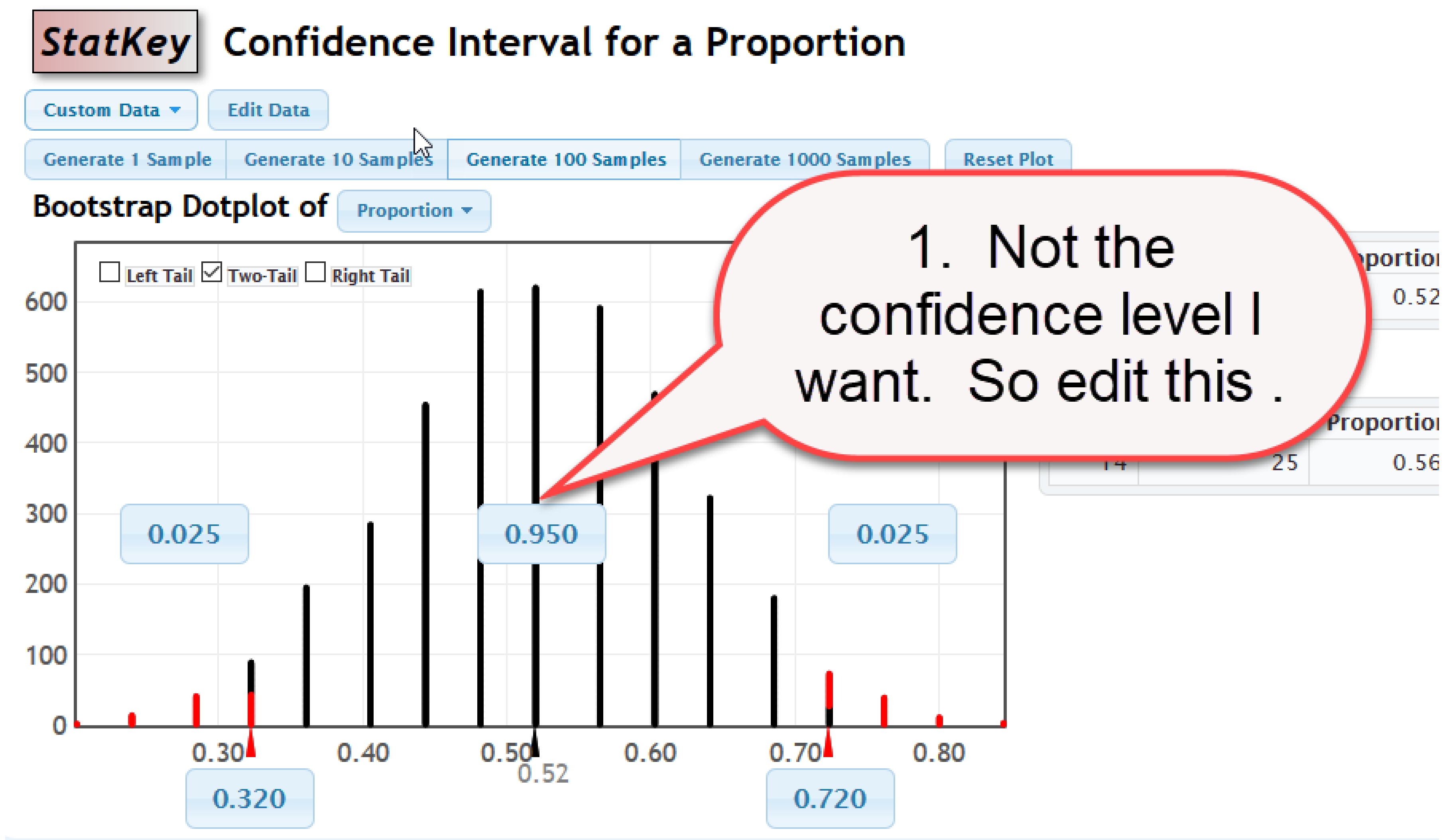
Theoretical Distributions	Normal	t	χ^2	F

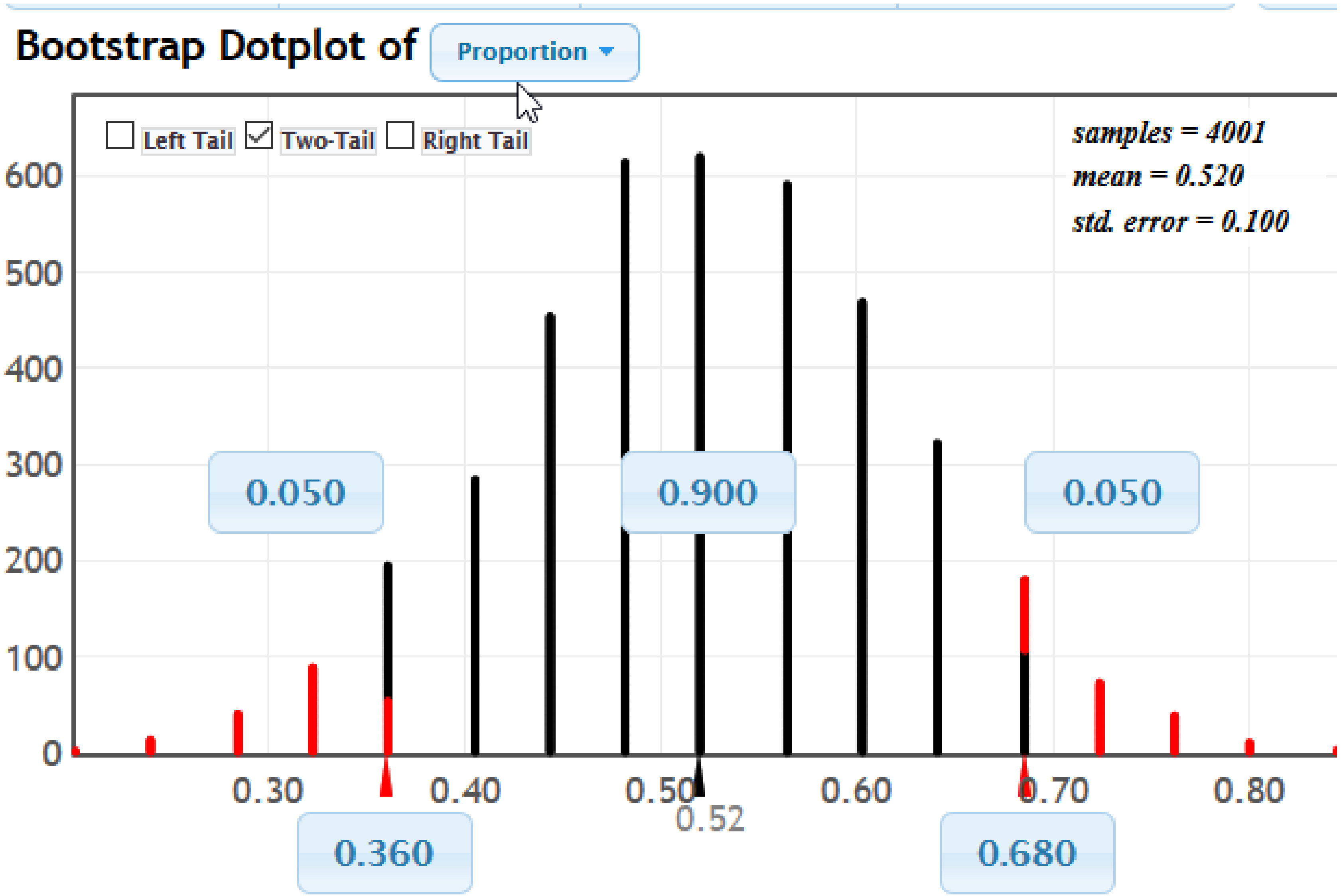
More Advanced Randomization Tests	χ^2 Goodness-of-Fit	χ^2 Test for Association	ANOVA for Difference in Means	ANOVA for Regression


StatKey
Confidence Interval for a
Voter Sentiment (Support ballot initiative)
Edit Data
Generate 1 Sample
Generate 10 Samples
Generate 100 Samples
Bootstrap Dotplot of
Proportion
Left Tail
Two-Tail
Right Tail
Count
Sample Size
Proportion
1. Put in the data count and sample size.
2. Generate a large number of replications.


StatKey
Confidence Inte
Edit Data
Generate 1 Sample
Proportion
sample
Count
Sample Size
Proportion
25
0.520
Size
Proportion
25
0.480
0. Notice the name of this graph/dist'n.
1. Generate one sample.
3.This sample's p-hat is in Bootstrap Dist'n.
2. Result of one sample.
40
30
20
10
0
0.30
0.40
0.50
0.60
0.70
0.48


StatKey
Confide
2. I'll check two-tail for a confidence interval.
Custom Data
Edit Data
Generate 1 Sample
Generate
Bootstrap Dotplo
Left Tail
Two-Tail
Right Tail
samples = 4001
mean = 0.520
std. error = 0.100
1. I used 4001 replications.
600
500
400
300
200
100
0
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.52
Original Sample
Count
Sample Size
Proportion
13
25
0.520
Bootstrap Sample
Count
Sample Size
Proportion
14
25
0.560


StatKey
Confidence Interval for a Proportion
Custom Data
Edit Data
Generate 1 Sample
Generate 10 Samples
Generate 100 Samples
Generate 1000 Samples
Reset Plot
Bootstrap Dotplot of
Proportion
Left Tail
Two-Tail
Right Tail
600
500
400
300
200
100
0
0.025
0.950
0.025
0.30
0.40
0.50
0.60
0.70
0.80
0.52
0.320
0.720
1. Not the confidence level I want. So edit this .
25
0.52
0.56


Bootstrap Dotplot of
Proportion
Left Tail
Two-Tail
Right Tail
samples = 4001
mean = 0.520
std. error = 0.100
600
500
400
300
200
100
0
0.050
0.900
0.050
0.30
0.40
0.50
0.60
0.70
0.80
0.52
0.360
0.680

Conclusion for estimating a parameter

- Estimate the population proportion of college students at that university who would rate such an experience unpleasant.
- **Answer:** I have 90% confidence that the population proportion of students who are uncomfortable in this situation is between 0.36 and 0.68.
- Assumptions we're making for this conclusion:
  - The students in the study are, essentially, a random sample of students at that university.
  - The study was designed appropriately (e.g. all the students were treated in the same way, etc.)

Meaning of a Confidence Interval

Possible misunderstanding

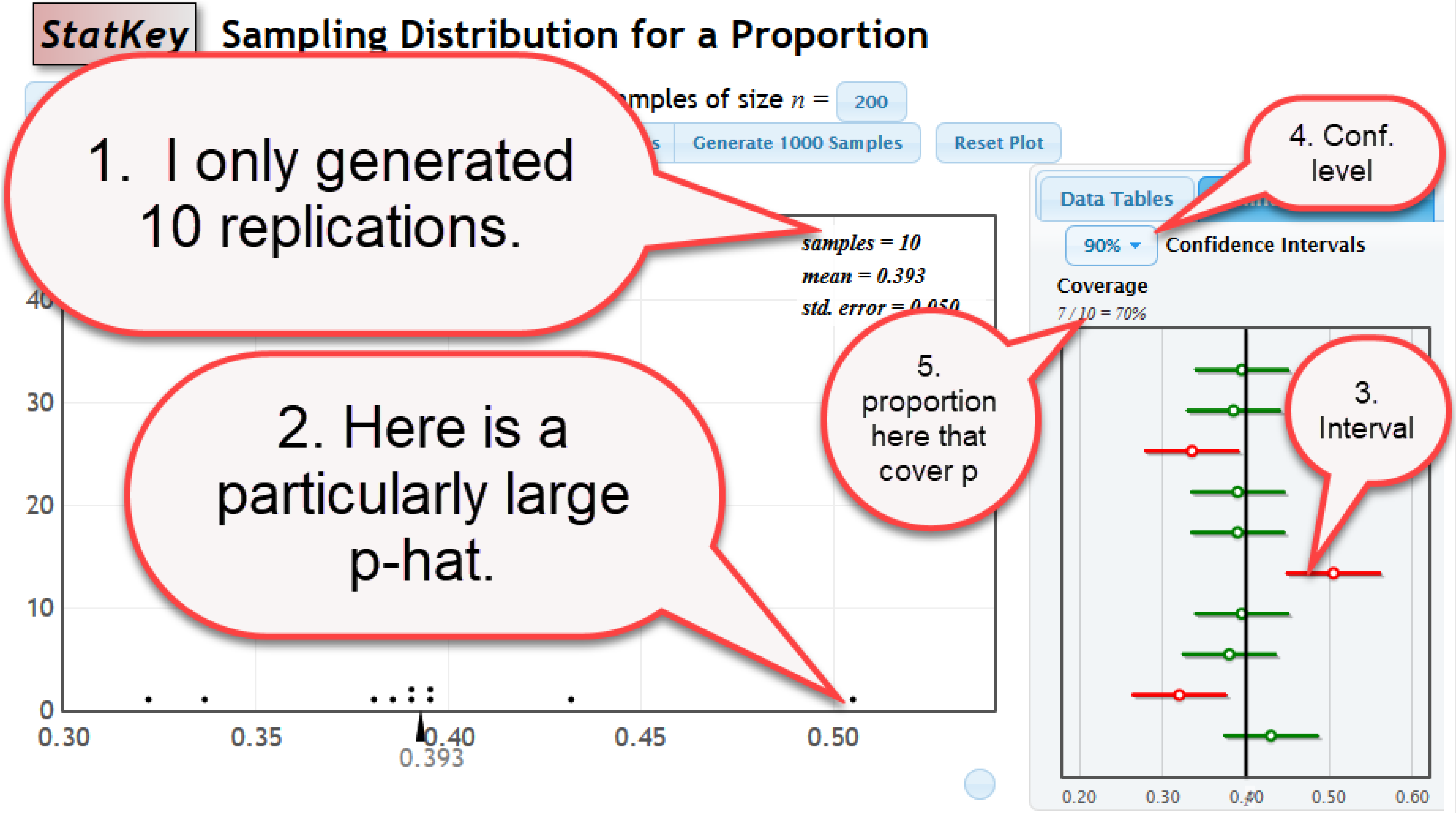
It is important not to get too focused on the actual endpoints of the confidence interval.

The 90% confidence in our result is, essentially, a confidence in the process giving us an interval that, when the process is repeated, the probability that the interval contains the true population parameter is 90%.

Following are illustrations of the meaning of this.

Illustrating the meaning of a CI.

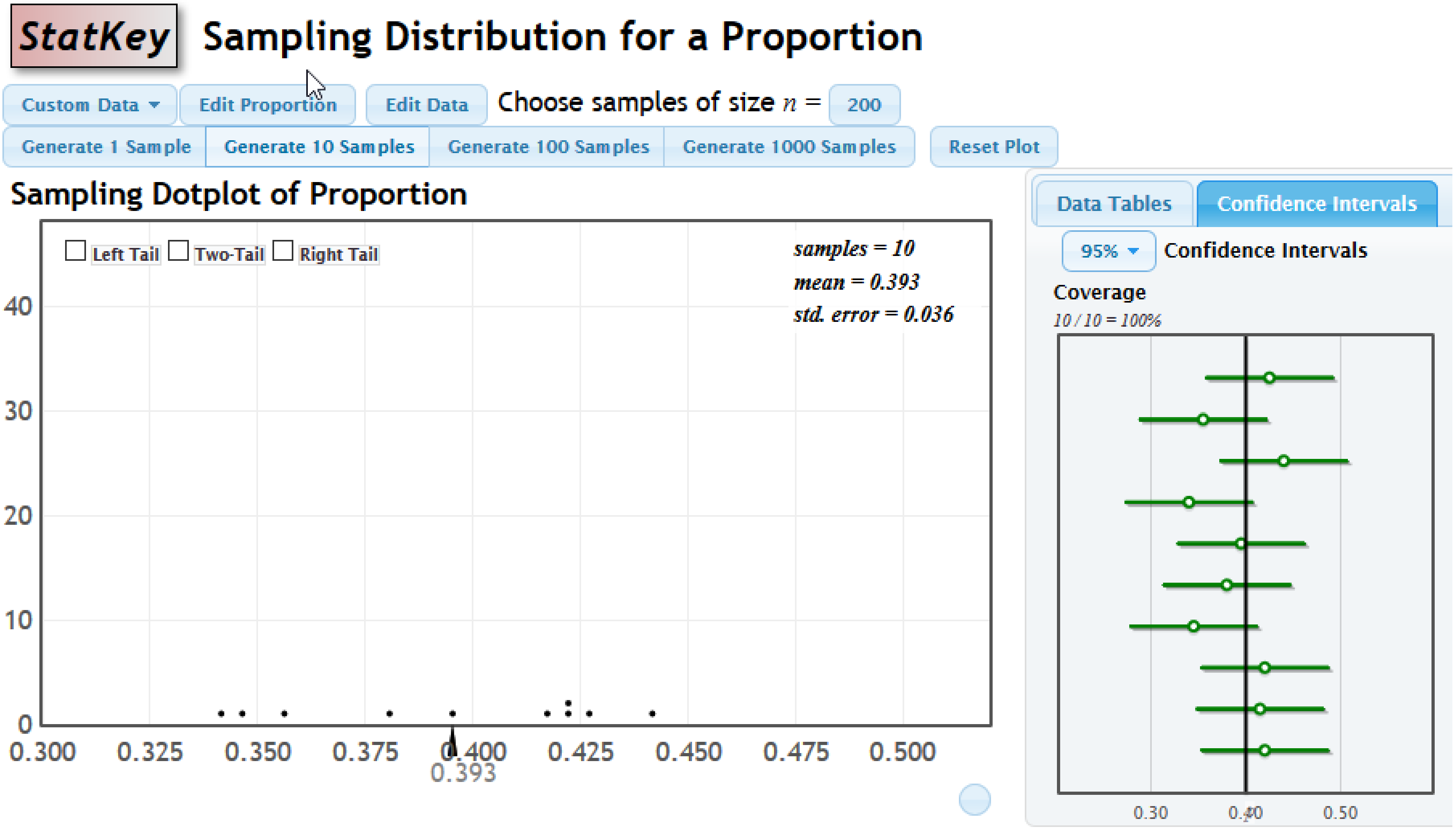
- Notice that EVERY value we generate in a sampling distribution is a value that might have occurred for our dataset.
- Thus every value in the sampling distribution could be the center of a confidence interval for the parameter.
- Some of those values are far out in the tail of the sampling dist’n, so confidence intervals from them **do not contain** the actual population proportion.


StatKey Sampling Distribution for a Proportion
mples of size n = 200
Generate 1000 Samples
Reset Plot
1. I only generated 10 replications.
2. Here is a particularly large p-hat.
samples = 10
mean = 0.393
std. error = 0.050
5. proportion here that cover p
4. Conf. level
3. Interval
Data Tables
90%
Confidence Intervals
Coverage
7 / 10 = 70%
0.30 0.35 0.40 0.45 0.50
0.393
0.20 0.30 0.40 0.50 0.60

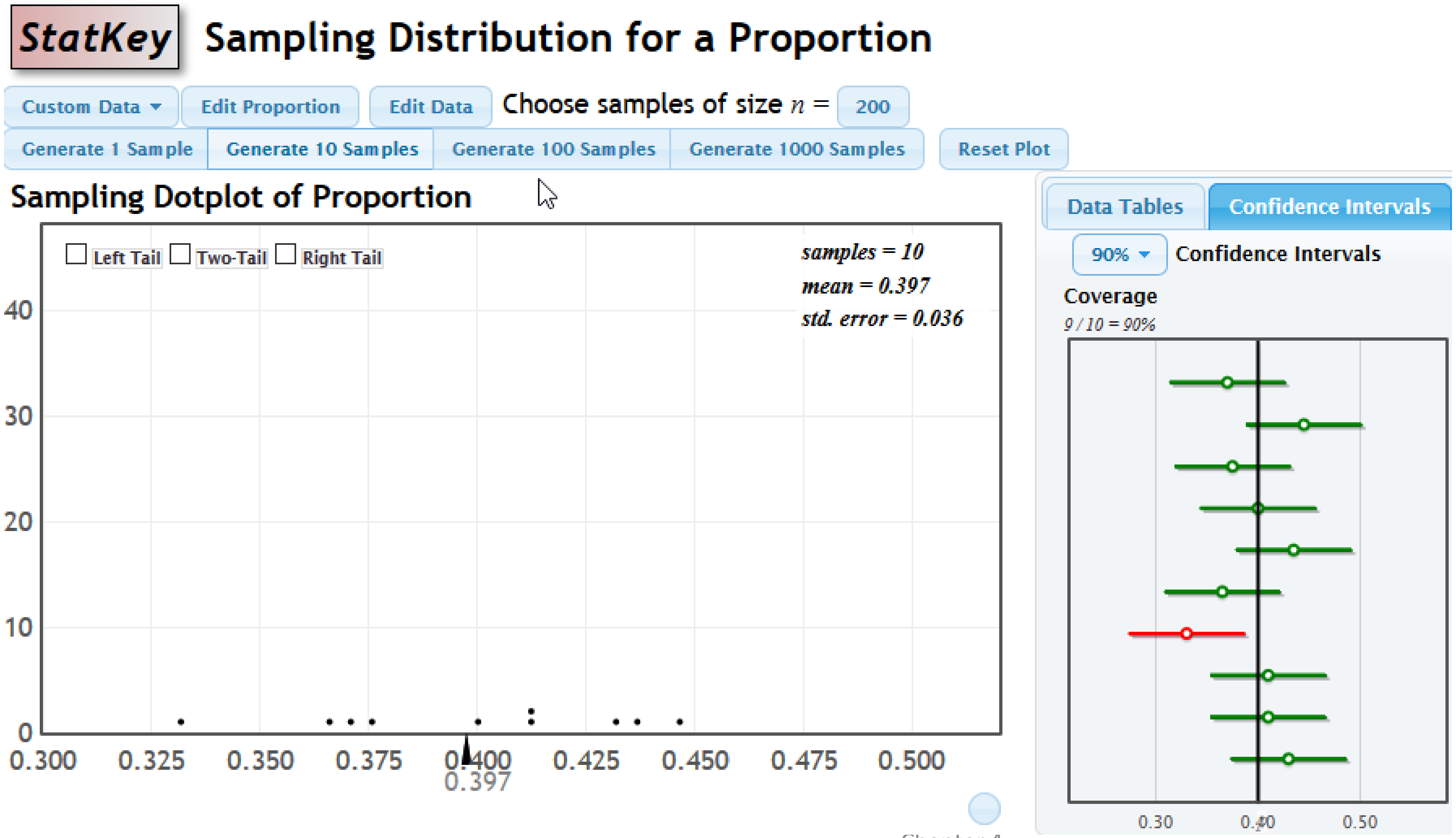
Specific Illustration of Meaning of a CI

- I generated only ten replications from this sampling distribution.
- I identified a particular one of those replications – the one with the largest value.
- I chose to display the confidence intervals (right top instead of the data table.)
- I looked at the confidence interval that has the largest values in it, because that comes from the largest value in the displayed values in the sampling dist’n.
- I chose to display 90% confidence intervals.
- I noticed that 70% of these ten intervals contain the population parameter (in a tiny display box) and obviously 3 out of 10 do not, when I look at the picture.

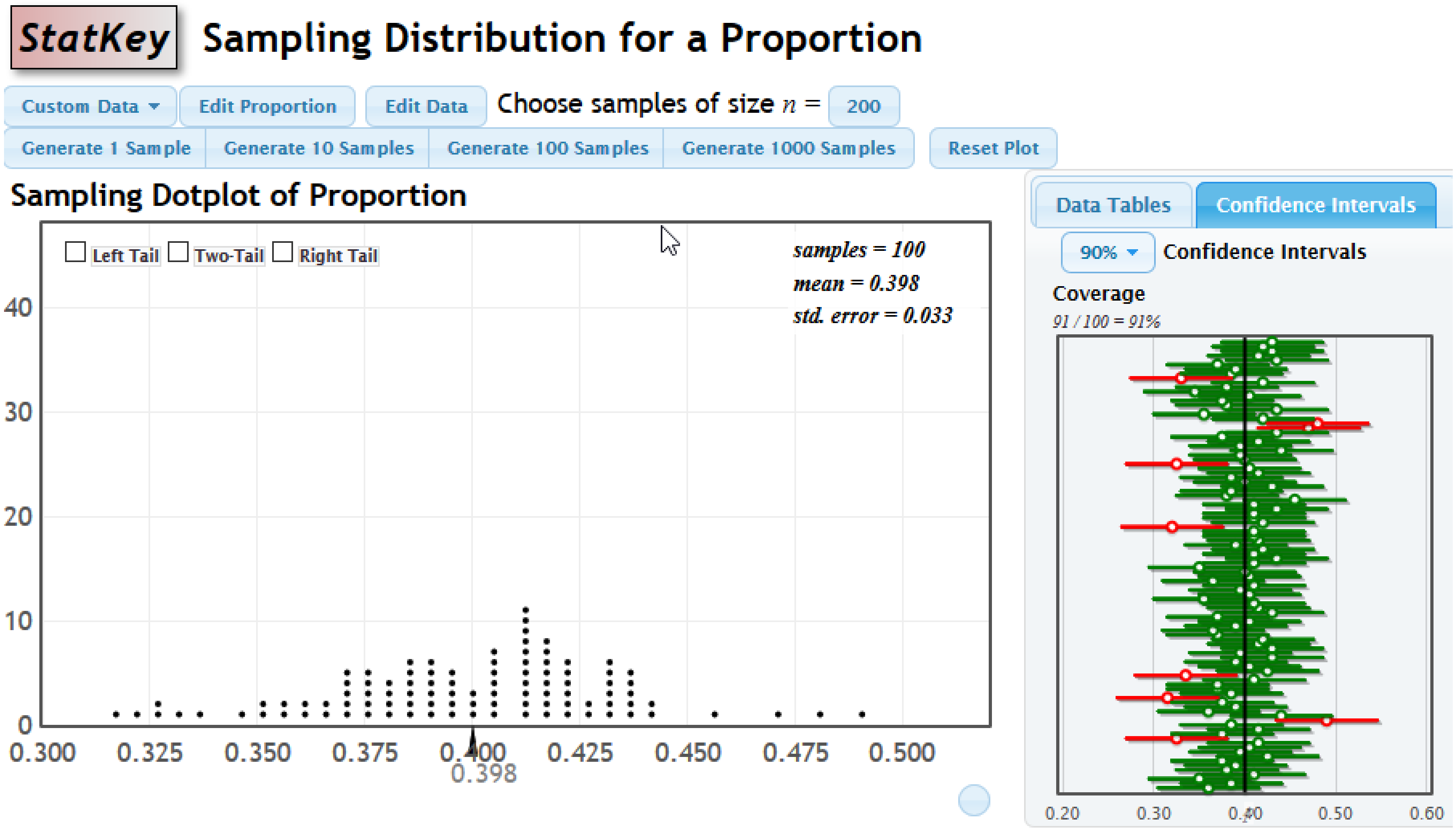
Another illustration of the same.



A third illustration of the same



A fourth illustration with 100 CIs



Meaning of a Confidence Interval

- If we carry out the process of taking a sample from the pop’n, computing the statistic value from it, and forming a 90% confidence interval with that value in the center

Then

- On the average, 90% of the resulting intervals will contain the actual pop’n parameter.

Summary of Chapter A

Two types of statistical questions: Estimation and Testing a claim

Main idea: Each is answered in the sampling dist'n of a statistic

Technique: We can simulate the appropriate sampling dist'ns

Testing a claim: Uses the sampling dist'n of the statistic if our claim about the parameter is true. (Center is at the claimed parameter value)

Estimating a parameter: Bootstrapping gives us an approximation of the population from which we generate the bootstrap distribution of the statistic.
We use that to approximate the shape and variability of the sampling distribution.
We note that it holds no additional information about the center of the sampling dist'n than the sample statistic has. (Center is at the value of the sample statistic.)

Meaning of a confidence interval: The "90% claim" is about the process of generating values in the dist'n, not about the individual values of the endpoints of the interval.