For Lab 11 and 12, you will collect your own data. For Lab 11, we will analyze the colors of Starburst. For Lab 12, we will analyze the personality types of statistics students, but we will collect the data now.

To collect the personality type data, please open the following link in your browser window: http://www.16personalities.com/free-personality-test, and complete the 10-15 minute personality test. Once you complete the test, your personality will be classified into one of 16 "types," but we will consider only the 4 main personality categories: Analysts, Diplomats, Sentinels, and Explorers. The figure below shows someone classified as an "Architect," which is one of the Analyst types as shown by the highlighting. Note that you may have to click on "Start Reading" to find this type. While you are waiting for the rest of the class to finish, feel free to peruse your "results," or to click on the Personality Types tab to look at all of the possible types.
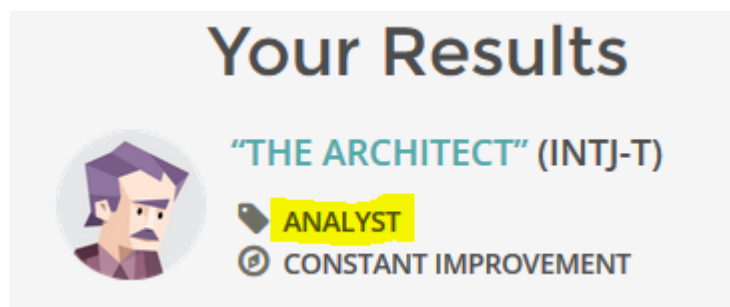
Question #1 Once you obtain your individual results, complete the Questionnaire on Titanium for your personality type, major, and the number of Starbursts in each color you received.

In this lab, we will test the claim that all four main colors are equally likely in the population of all Starburst.

Question #2 Write out the null hypothesis for this goodness-of-fit test.

H_0: p_{Red} = 0.25, p_{Orange} = 0.25, p_{Pink} = 0.25, p_{Yellow} = 0.25

It's okay to say all four colors are equally likely, but remember that we want to create a model, so we should specify the actual values of all four proportions.

Question #3 If the null hypothesis is true, how many of each color Starburst would we expect to see in our sample?
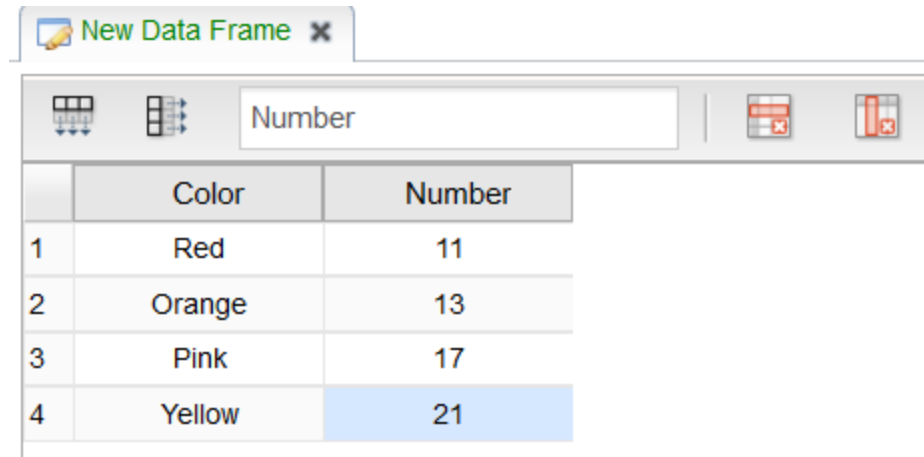
Our sample had 88 candies in it, so if the null hypothesis is true, we expect 22 red Starburst, 22 orange Starburst, 22 pink, and 22 yellow.

Question #4 If the sample size assumptions are met, what would be the sampling distribution of your test statistic (i.e., what is the type of distribution and the degrees of freedom)?

If the sample size assumptions are met, our test statistic would have a χ² distribution with 3 degrees of freedom.

Let's put in our data and running the goodness-of-fit test in Rguroo.

Click on *Data* ✉ *Data Import* ✉ *Create New Data Frame*. Create the data frame as shown in the screenshot below, except put in the correct counts as indicated by the Titanium survey.
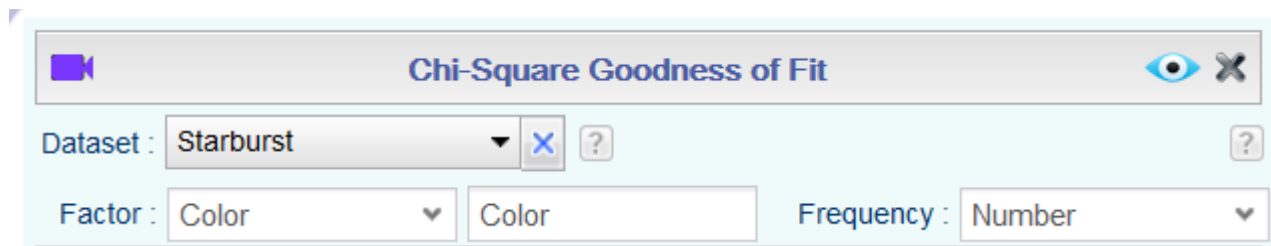

Save the dataset as the Rguroo dataset *Starburst*.

Question #5 Are the sample size assumptions are met for a chi-square goodness of fit test? Show how you checked the assumptions.

Yes, the sample size assumptions are met. The expected counts in all categories are 22 > 5.

Click on *Analytics* ✉ *Goodness of Fit* and fill out the dialog as shown below.

The observed counts should now automatically pop up in the table in the dialog. Fill in the expected probability of each color.

If the sample size assumptions are met, check the *Chi-Square* box. Otherwise, we will create a <u>simulated</u> chi-square distribution; check the *Simulation* box.

Question #6 Copy the table output by Rguroo below.

Chi-Squared Goodness of Fit Test

Research Hypothesis Ha: Population proportions of Color are different from the expected distribution

Observed Test Statistic	Degrees of Freedom	p-value
7.45455	3	0.05874

Test is not significant at the 5% significance level

Question #7 What is the value of the chi-square test statistic as computed by Rguroo?

7.45455 (you should probably round this to 7.45 or 7.455)

Question #8 What is the p-value for this test?

0.05874 (it's okay to round this to 0.059 or 0.06)

Question #9 Using a 5% significance level, what can you conclude about the distribution of Starburst?

We fail to reject the null hypothesis. We conclude that our model (which assumes each color has a 25% chance of showing up) is a reasonable approximation of the true/real distribution of Starburst colors.

Question #10 Do you believe that your conclusion (from **Question #9**) actually applies to the population of <u>all</u> Starburst? (HINT: Think about the sample we used and the way we collected the data)

Answers will vary. Essentially, I'm looking for one of two things:
1) **No, the conclusion does not apply, for any number of reasons why the bag of Starburst might not actually represent a "random sample" of Starburst (the bagging process may not be random; different manufacturing plants may have different distributions; etc.)**
2) **Yes, the conclusion does apply, because there is no obvious way in which this sample of 88 Starburst would look meaningfully different from any other sample of 88 Starburst (except because of random luck).**

Lab Assignment #12

The two-way table below shows the result of the survey asking you for your personality type and major in two sections of Math 338.

OBSERVED COUNTS	Biology	Computer Science	Other	**TOTAL**
Analyst	2	5	0	**7**
Diplomat	13	10	5	**28**
Sentinel	3	1	2	**6**
Explorer	3	4	0	**7**
TOTAL	**21**	**20**	**7**	**48**

Let's investigate whether personality type affects the choice of major.

Question #1 Write the null hypothesis for this test of independence.

H_0: the personality type does not affect the choice of major. OR, there is no relationship/association between the personality type and the major.

Question #2 If the sample size assumptions are met (all expected counts ≥ 5), what would be the sampling distribution of your test statistic (i.e., what is the type of distribution and the degrees of freedom)?

Our test statistic would come from a χ^2 distribution with (r-1)(c-1) = (4 – 1)(3 – 1) = 6 df.

Question #3 If the null hypothesis is true, calculate the <u>expected</u> number of analysts who are Computer Science majors. If your number is not an integer, round it to at least one decimal place.

If the null hypothesis is true, we would expect (7)(20)/48 = 2.92 analysts who are Computer Science majors.

Question #4 If the null hypothesis is true, calculate the <u>Pearson residual</u> and <u>contribution</u> to the chi-squared statistic for analysts who are Computer Science majors.

Pearson residual: $\dfrac{5-2.92}{\sqrt{2.92}} = 1.22$

Contribution to chi-squared statistic: $(1.22)^2$ = 1.49

Now, let's put in our data and run the test of independence in Rguroo. Import the majors.csv data file from Titanium into Rguroo.

Click on *Analytics* ✉ *Contingency Table*, select your newly imported dataset, and fill out the dialog as shown below.

* **Factor 1** :	Personality.Type	⌄	Personality.Type
* **Factor 2** :	Major	⌄	Major
Frequency :	Num. Variable...	⌄	

Question #5 To obtain the p-value, can we use the sampling distribution from **Question #2**, or do we have to simulate a sampling distribution? Explain your reasoning. (HINT: look at your answer to **Question #3**)

Clearly we do not have 5 expected counts in every cell of the table. Therefore we cannot use a χ^2 distribution with 6 df to estimate our p-value, and we must simulate the sampling distribution.

If you answered "yes" to **Question #5**, check the *Chi-Square* box. Otherwise, we will create a <u>simulated</u> chi-square distribution; check the *Simulation* box.

Question #6 Copy the table with a p-value below.

Chi-Squared Test of Independence by Simulation

Random generator seed = 100

Observed Test Statistic	p-value	Number of Simulations
7.19184	0.3142	10000

Test is not significant at the 5% significance level

Question #7 What is the value of the observed chi-square test statistic as computed by Rguroo?

7.19184 (best to round to 7.19)

Question #8 What is the p-value for this test?

0.3142 (can round to 0.31)

Question #9 Using a 5% significance level, can you conclude that people's personality type affects their choice of major?

Since 0.31 > 0.05, we fail to reject our hypothesis (H$_0$). It appears reasonable that people's personality type does not affect their choice of major.

Question #10 Do you believe that your answer (from **Question #9**) applies to <u>all</u> students at Cal State Fullerton? (HINT: Think about the sample we used and the way we collected the data)

If we wanted to model all CSUF students, we would have problems with both the sample we used and the data collection method. First, we only looked at (mostly) biology and computer science majors, so it's hard to say whether our results would apply to people in other majors. Second, these personality tests can give pretty different results even for the same person taking the test an hour later, so it's not clear whether people were classified accurately.

Assume that squirrel weights are approximately normally distributed with $\mu = 1.7$ lbs and $\sigma = 0.5$ lbs.

Question #1 Using the 68-95-99.7 rule, the middle 68% of squirrels should fall within what values? What range corresponds to the lowest 16% of squirrel weights? What range corresponds to the highest 16% of squirrel weights?

The middle 68% of squirrels should fall between 1.7 – 0.5 = 1.2 and 1.7 + 0.5 = 2.2 lbs.

The lowest 16% of squirrel weights are below 1.2 lbs.

The highest 16% of squirrel weights are above 2.2 lbs.

Question #2 Using the 68-95-99.7 rule, the middle 95% of squirrels should fall within what values? What range corresponds to the lowest 2.5% of squirrel weights? What range corresponds to the highest 2.5% of squirrel weights?

The middle 95% of squirrels should fall between 1.7 – 2(0.5) = 0.7 and 1.7 + 2(0.5) = 2.7 lbs.

The lowest 2.5% of squirrel weights are below 0.7 lbs.

The highest 2.5% of squirrel weights are above 2.7 lbs.

Question #3 Suppose we want to find the proportion of squirrels that weigh between 0.8 and 1.2 lbs. Standardize the weights of 0.8 lbs and 1.2 lbs. Write the z-scores corresponding to the two weights below.

For 0.8 lbs: $z = \dfrac{x - \mu}{\sigma} = \dfrac{0.8 - 1.7}{0.5} = -1.8$

For 1.2 lbs: $z = \dfrac{x - \mu}{\sigma} = \dfrac{1.2 - 1.7}{0.5} = -1.0$

In Rguroo, click on *Probability* ✉ *Calculator* ✉ *Continuous*. Make sure that the *Values* ✉ *Probability* button is selected, and that the *Distribution* is *Normal*. The default values are $\mu = 0$ and $\sigma = 1$, which corresponds to a standard normal distribution. Make sure the drop-down menu below those values reads *Below*. For each z-score you calculated in **Question #3**, type in the z-score and press Enter. The cumulative proportion corresponding to the z-score appears in the gray box.

Question #4 What are the cumulative proportions that correspond to your z-scores?

A z-score of -1.8 corresponds to a cumulative proportion of 0.03593

A z-score of -1.0 corresponds to a cumulative proportion of 0.15866

Question #5 By subtracting the two cumulative proportions, find the proportion of squirrels that weigh between 0.8 and 1.2 lbs.

0.15866 – 0.03953 = 0.12273

Approximately 12.3% of squirrels weigh between 0.8 and 1.2 lbs

Now we will find this proportion in Rguroo without using z-scores. Input the correct population mean and standard deviation in the appropriate text box. Change the drop-down menu from *Below* to *Between*, and type in the appropriate (original) values. Now check the *Graph* box and preview the result.

Question #6 Copy and paste the graph from the output. Do you get the same answer as **Question #5**?
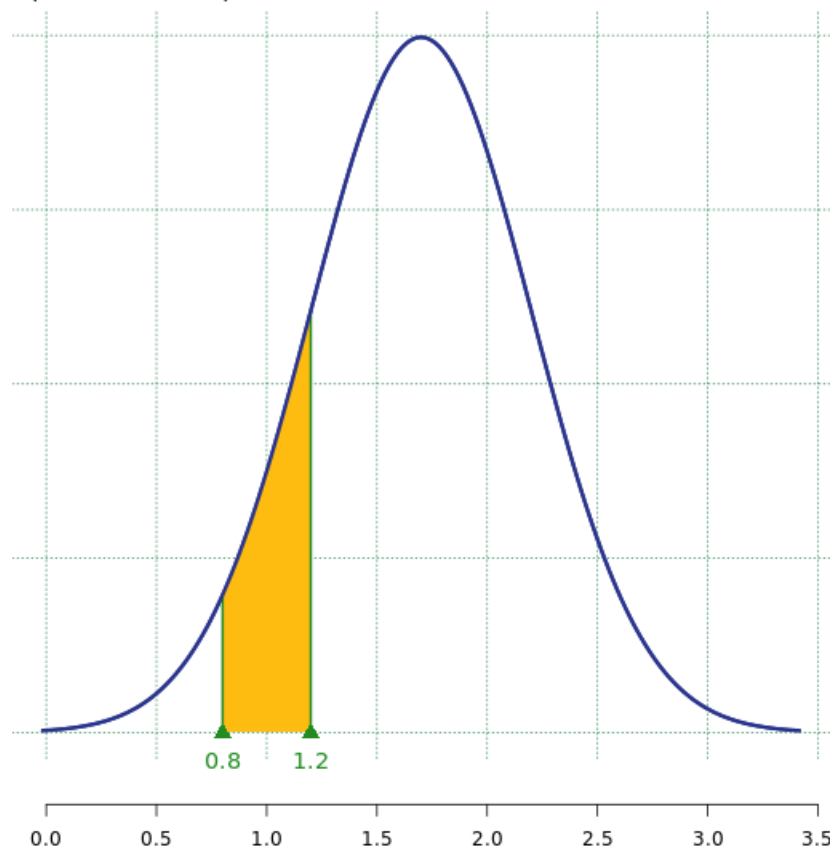
The probability of the weight being between 0.8 and 1.2 lbs is 0.1227, which matches our answer to Question 5.

Suppose now we want to find the weight of the 8% heaviest squirrels.

Question #7 What is the cumulative proportion corresponding to the 8% heaviest squirrels?

The cumulative proportion is 100% - 8% = 92% or 0.92

Reset the mean to 0 and the standard deviation to 1. Click the *Probability* ✉ *Values* button, since we have a range of proportions and want to find a range of values. Since we want the heaviest squirrels, we want the weights of squirrels in the *upper* tail of the distribution, so we change the drop-down menu to *Upper Tail* and type in 8% as a decimal (0.08) in the box to the right of *Upper Tail*.

Question #8 What is the z-score that corresponds to this cumulative proportion?

The z-score corresponding to 8% in the upper tail is 1.4051.

Question #9 Convert this z-score back to the original scale. What range of weights corresponds to the 8% heaviest squirrels?

$$x = \sigma z + \mu = 0.5(1.4051) + 1.7 = 2.40255$$

The 8% heaviest squirrels weigh about 2.40 lbs or heavier.

Now, again, change the mean and standard deviation to the actual population mean and standard deviation. Check the *Graph* box and preview the result.

Question #10 Copy and paste the graph from the output. Do you get the same answer as **Question #9**?

Based on the graph below we can confirm that the 8% heaviest squirrels weigh 2.403 lbs or greater.

X ~ Normal(μ = 1.7, σ = 0.5)
P(X ≥ 2.403) = 0.08

Lab Assignment #14
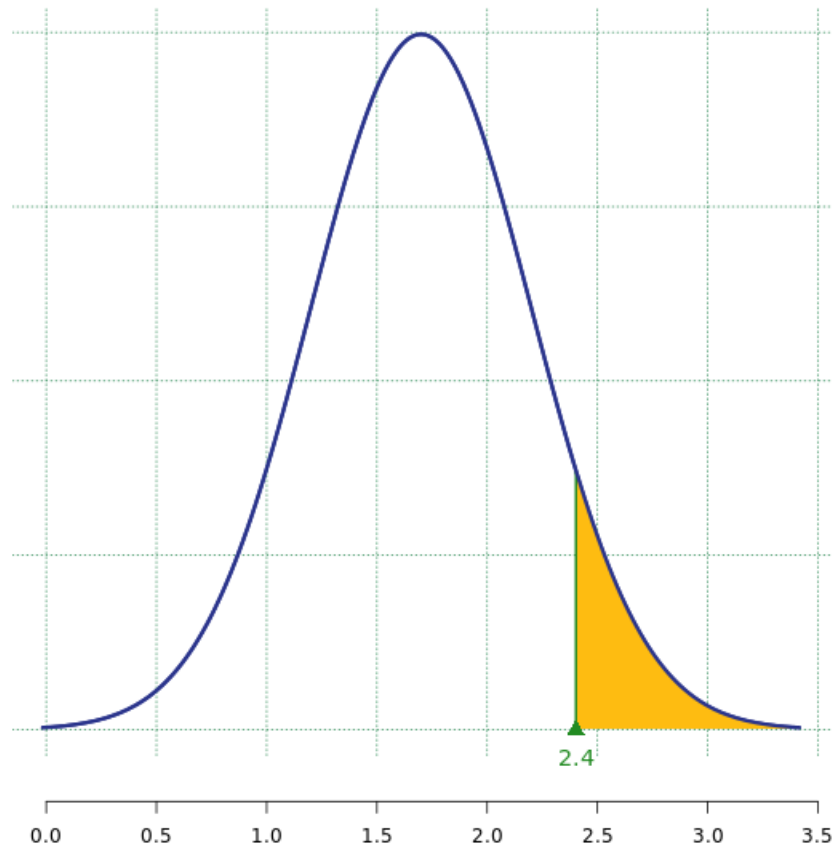
In this lab we are going to investigate some properties of numerical data. We will use a dataset already available in Rguroo. In the Data section, click on *Data Import* ✉ *Repository* to bring up the Rguroo data repository. Select the *R Datasets* repository. In the text box just above the list of datasets, type `trees` to find just the *trees* dataset we are interested in. Click the information button to the very right of the row to view information about the dataset.

Repository Dataset Import 🎥

| Filter Repository... | ✖ | ○ User Defined | ⬤ Public |

Repository	Description
boot	R Package boot
car	R Package car
cluster	R Package cluster
COUNT	R Package COUNT
R datasets	Base R

trees ✖

Dataset	Title	Rows	Columns	Info
trees	Girth, Height and Volume for Black Cherry Trees	31	3	ⓘ

[Import]   [Close]

Question #1 What are the cases in these data, and how many are there?
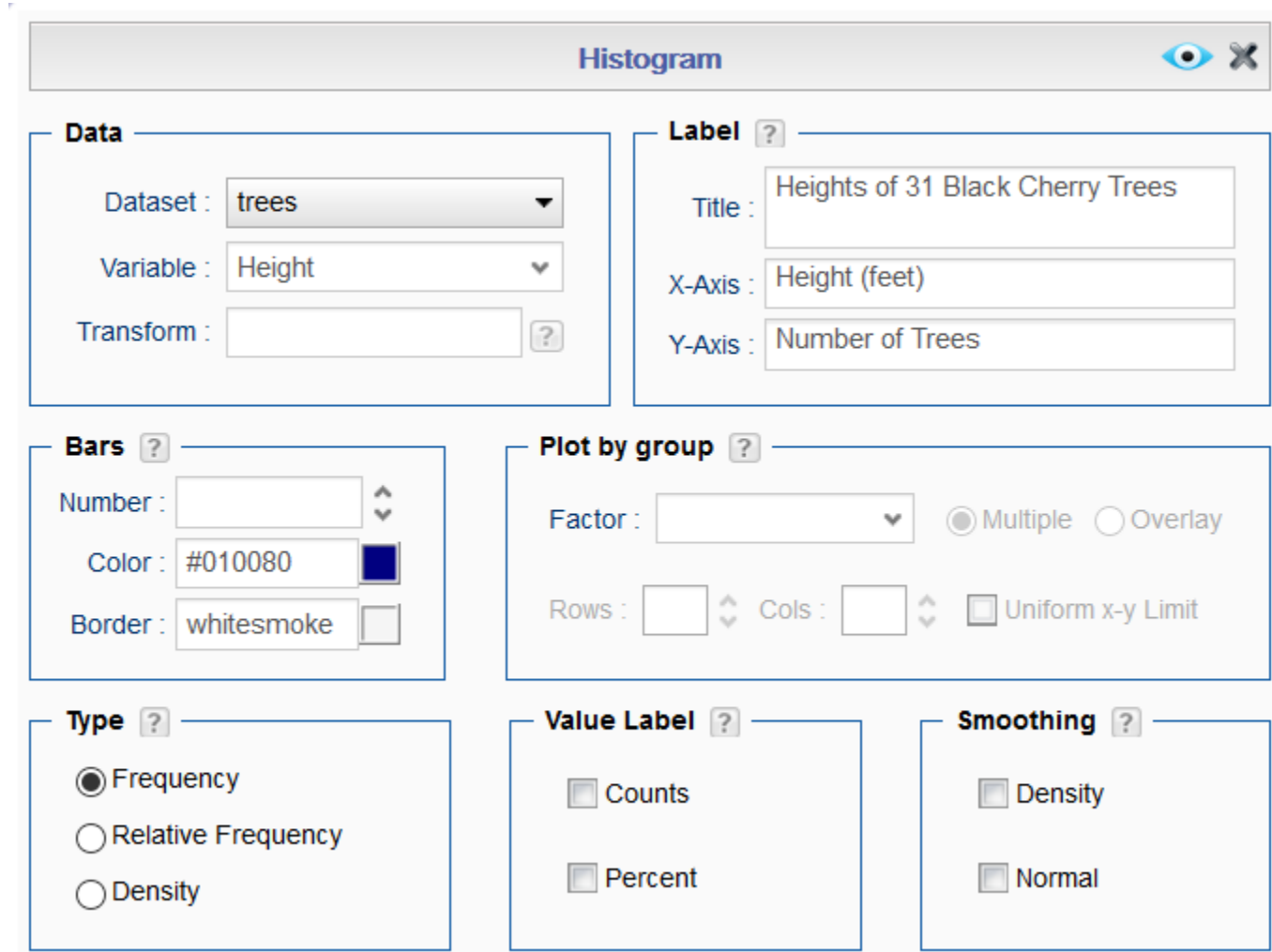
The cases in these data are black cherry trees. There are 31 cases.

Question #2 Name the three numerical variables in the dataset. Include the units of measurement.

The numerical variables are Girth (in inches), Height (in feet), and Volume (in cubic feet).

Once you are done viewing the information, click the *Import* button at the bottom to import the dataset.

Now, let's obtain a histogram of these data by going to *Create Plot* and selecting *Histogram*. Fill in the selections as indicated below, and then press the eye icon.

Histogram

Data

Dataset : trees
Variable : Height
Transform : [?]

Label [?]

Title : Heights of 31 Black Cherry Trees
X-Axis : Height (feet)
Y-Axis : Number of Trees

Bars [?]

Number :
Color : #010080
Border : whitesmoke

Plot by group [?]

Factor : ● Multiple ○ Overlay

Rows : ↕ Cols : ↕ ☐ Uniform x-y Limit

Type [?]

● Frequency
○ Relative Frequency
○ Density

Value Label [?]

☐ Counts

☐ Percent

Smoothing [?]

☐ Density

☐ Normal

Question #3 Copy the histogram and paste it below.

Heights of 31 Black Cherry Trees
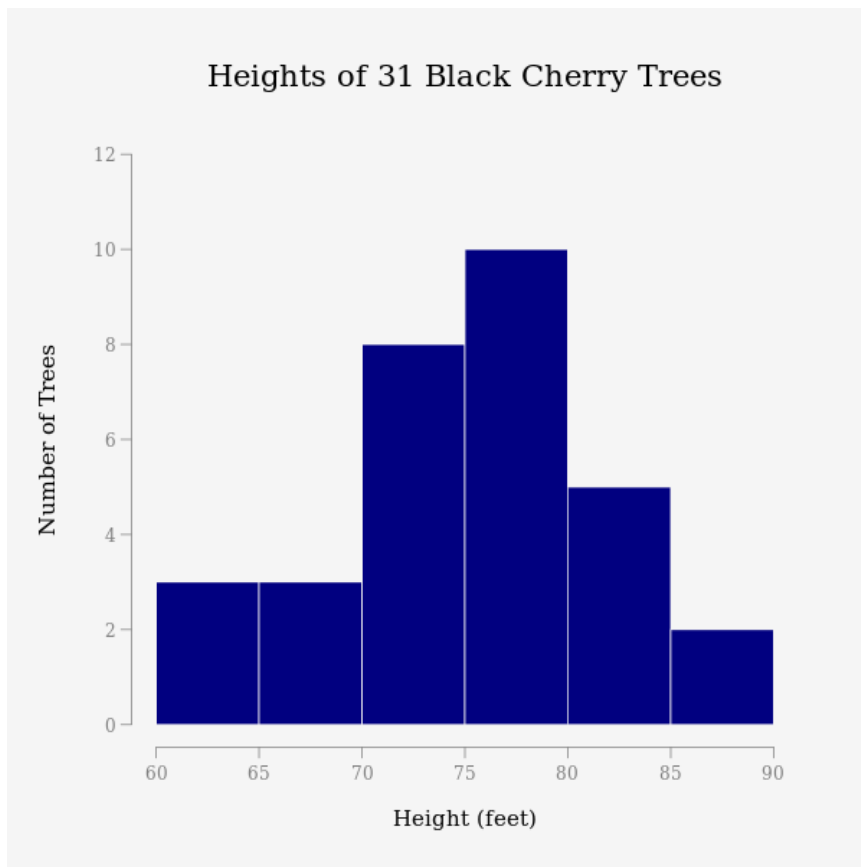
Question #4 Is the distribution of height symmetric or skewed? If it is skewed, state the direction. Do you see any potential outliers on the histogram?

The distribution of height is slightly skewed left (you can see this if you overlay the density curve), so I'd accept a description of symmetric. There do not appear to be any potential outliers.

Go back to the *Data* section, right-click the *trees* dataset, and select *Summary*. You should see a table similar to that automatically output when a dataset is uploaded from a file.

Question #5 According to the table, what is the mean height of these trees? What is the standard deviation of the height?

The trees have a mean height of 76 feet. The standard deviation is 6.37 feet.

Question #6 Using the table, report the five-number summary of the height of the trees.

The five-number summary is:
Minimum: 63 feet
First Quartile: 72 feet
Median: 76 feet
Third Quartile: 80 feet
Maximum: 87 feet
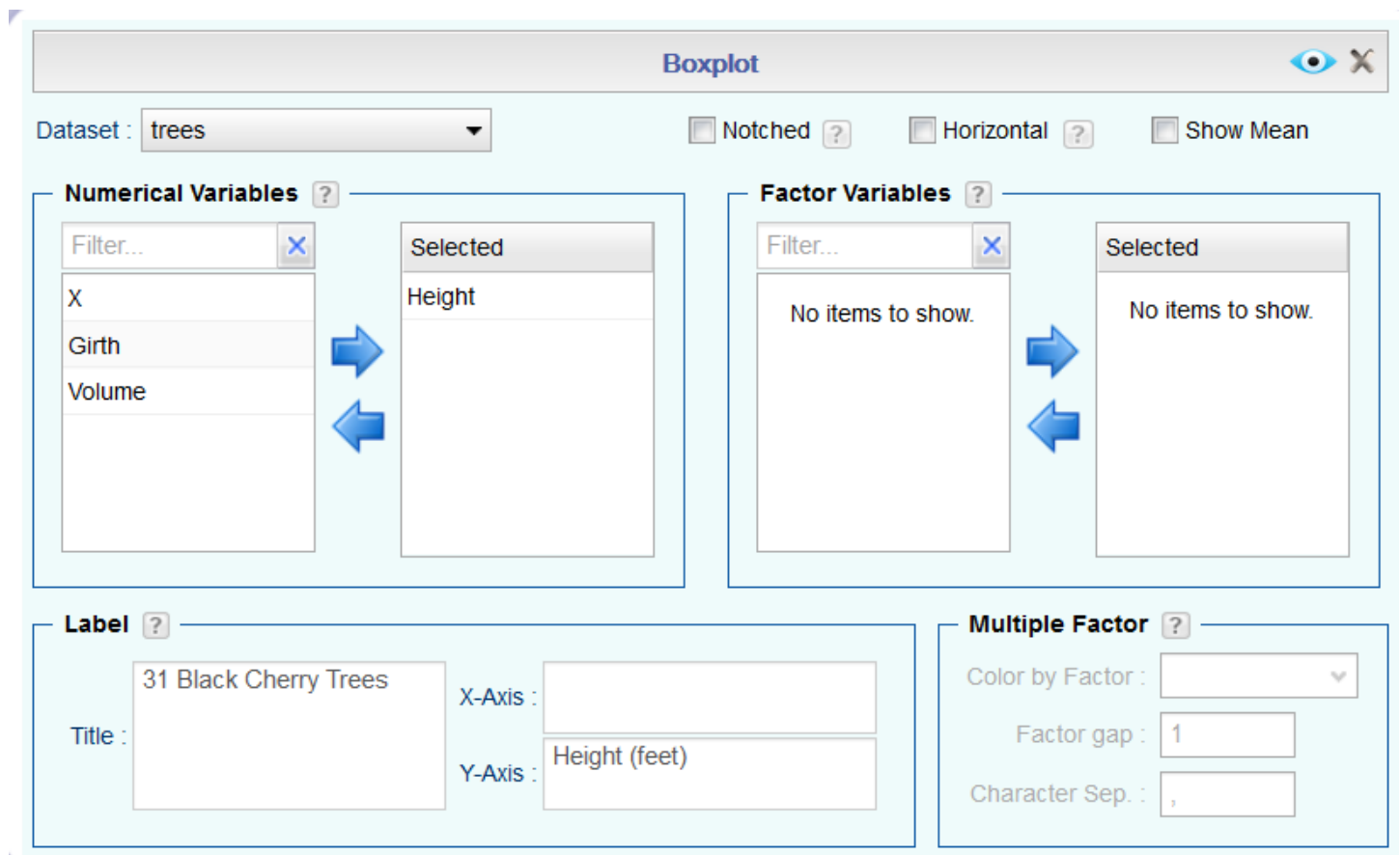
Question #7 According to our 1.5 IQR rule of thumb, does this dataset contain any outlier heights? If so, which data points are outliers? Show your work.

Lower fence: Q_1 − 1.5 IQR = 72 − (1.5 x 8) = 60 feet
Upper fence: Q_3 − 1.5 IQR = 80 + (1.5 x 8) = 92 feet

Since all heights lie within the fences, this dataset does not contain any outliers.

Now, let's obtain a boxplot of these data by going to *Plots* ✉ *Create Plot* and selecting *Boxplot*. Fill in the selections as indicated below, and then press the eye icon.

Boxplot

Dataset : trees   ☐ Notched [?]   ☐ Horizontal [?]   ☐ Show Mean

Numerical Variables [?]

Filter...	✕	Selected
X		Height
Girth		
Volume		

Factor Variables [?]

Filter...	✕	Selected
No items to show.		No items to show.

Label [?]

Title : 31 Black Cherry Trees
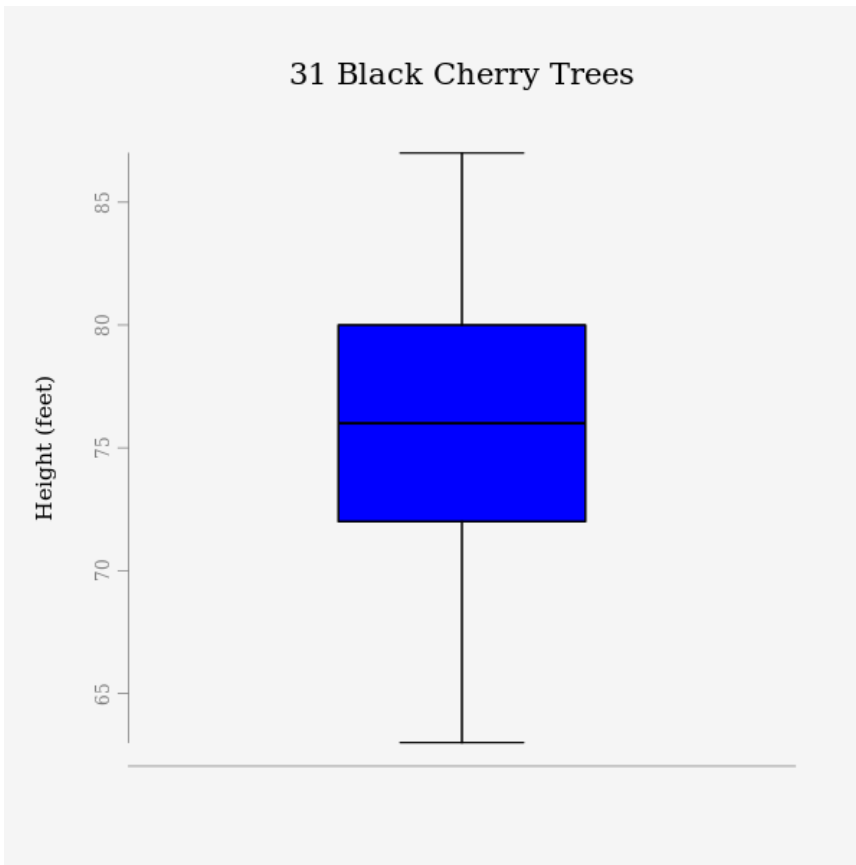
X-Axis :

Y-Axis : Height (feet)

Multiple Factor [?]

Color by Factor :

Factor gap : 1

Character Sep. : ,

Question #8 Copy the boxplot and paste it below. Did Rguroo identify any outliers based on the 1.5 IQR rule?

31 Black Cherry Trees

Height (feet)

Rguroo did not identify any outliers according to the 1.5 IQR rule of thumb.

Before the first exam, we discussed how linear transformations ($x_{\text{new}} = a + bx$) affect the mean and standard deviation of a random variable. In the last part of the lab, we will show that it works with actual data too. We will convert the measurements to meters (1 foot = 0.3048 meters) and discount the first 10m of the measurement. This results in the variable $Height.new = 0.3048 \times Height - 10$.

Question #9 Without doing anything in Rguroo, use the transformation rules and the values from Questions **#5** to compute the values of the mean and standard deviation of *Height.new*. Show all steps in your work.

The new mean should be 0.3048 x 76 – 10 = 13.16 meters.

The new standard deviation should be 0.3048 x 6.37 = 1.94 meters.

Now we will do the actual transformation in Rguroo. Right-click the *trees* dataset and select *Functions* ✉ *Transform*. The variables currently in the dataset will show up in the *Returned Variable* section on the right of the menu. To create a new variable, click the + sign in the *Variable* box on the left and type the name of the new variable (in this case *Height.new*). To create the transformation, double-click the variable *Height* to copy it into the *Transformation* box and modify the formula to read 0.3048*Height - 10. Finally, get rid of all the variables except *Height* and *Height.new* by dragging them to the *Excluded Variable* box. If you've done everything correctly so far, your screen should look like the figure below:

Click the eye icon to preview the new dataset, then save it as *trees_transform*.

Go back to the *Data* section, right-click the *trees_transform* dataset, and select *Summary*. Now you should see a summary of only the original heights and the transformed heights (*Height.new*).
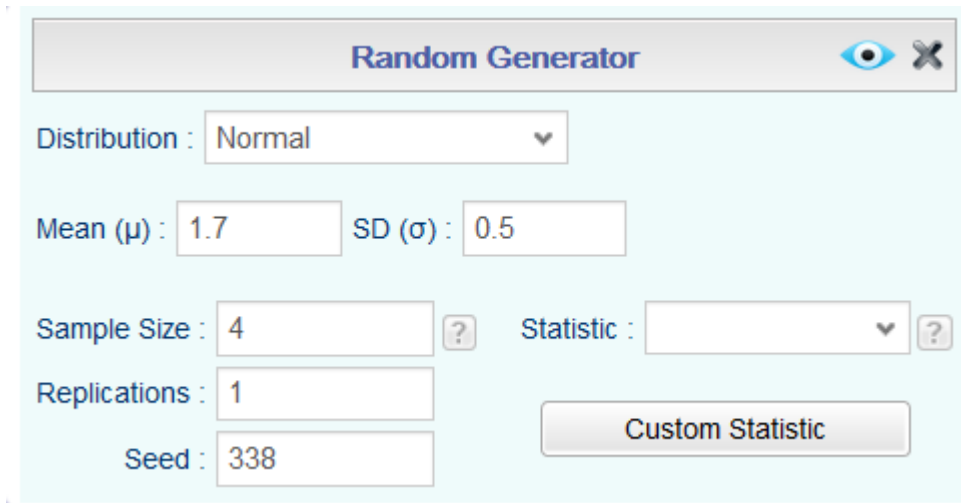
Question #10 Copy your summary table and paste it below. Did you get the same mean and standard deviation for your transformed height as Rguroo computed?

Numerical Variables

Variable	No. read	No. observed	No. missing	Min	Q1	Q2	Q3	Max	Mean	Std. deviation	Variance	SE of mean
Height.new	31	31	0	9.20240	11.9456	13.1648	14.3840	16.5176	13.1648	1.94213	3.77186	0.348817
Height	31	31	0	63	72	76	80	87	76	6.37181	40.6000	1.14441

The values output by Rguroo for the mean and standard deviation are the same as we computed.

Lab Assignment #15

Assume that squirrel weights are approximately normally distributed with $\mu=1.7$ lbs and $\sigma=0.5$ lbs. Let's simulate a sample of 4 squirrels. Click on *Probability-Simulation* ✉ *Probability* ✉ *Random Generator* and fill in the dialog as shown in the figure below. We'll use seed 338 so that everyone gets the same results.

Question #1 Record the weights of the four simulated squirrels below. Round each weight to three decimal places.

The simulated squirrels have weights of 1.529, 2.141, 2.903, and 1.725 lbs.

The next step in our simulation is to automatically compute a statistic that summarizes the sample of four squirrels. In this lab we'll compute both the mean and standard deviation. In the *Statistic* drop-down menu, check both the *Mean* and *Standard Deviation* boxes, as shown below.

Question #2 What are the mean and standard deviation of the weights of these four squirrels? Are they exactly equal to the population mean and standard deviation? Are they close?

These four squirrels have a sample mean of 2.075 lbs and a sample standard deviation of 0.609 lbs. These values are not equal to the population mean and standard deviation and are only somewhat close (both are about 22% higher).

Finally, we will simulate this process many times (1000 to be exact) to obtain 1000 sample means and 1000 sample standard deviations.

Question #3 Explain the difference between the *Sample Size* and *Replications*. Which one should we use as our *n* in the formula for computing the standard deviation of the sampling distribution?

Question #4 Compute the theoretical mean and standard deviation of the sampling distribution of the mean squirrel weight in samples of size *n* (where *n* is the number you identified in **Question #3**).

The theoretical mean of the sampling distribution of the mean squirrel weight in samples of size 4 is the population mean, $\mu=1.7$ lbs.

The theoretical standard deviation of the sampling distribution of the mean squirrel weight in samples of size 4 is $\dfrac{\sigma}{\sqrt{n}}=\dfrac{0.5}{\sqrt{4}}=0.25$ lbs.

Save the new dataset as *sim_squirrels*. You may also want to save the parameters of the random generator for later reference.

Question #5 What are the actual mean and standard deviation of your sample means? Are they close to what you calculated in **Question #4**? (Hint: right-click the *sim_squirrels* dataset and select *Summary*)

The actual mean of the sample means is 1.697. The actual standard deviation of the sample means is 0.247. These values are pretty close to what was calculated in #4.

Be sure that you don't confuse the standard deviation of the sample means with the mean of the sample standard deviations (which is 0.465)!

Question #6 Using the *Continuous Distribution Calculator*, find the approximate probability of obtaining a sample mean weight between 0.8 and 1.2 lbs (assuming the sampling distribution is normal).

The probability of obtaining a sample mean weight between 0.8 and 1.2 lbs is approximately 0.0226 or 2.26%.

Question #7 In Lab 13, we computed the probability of obtaining a squirrel weight between 0.8 and 1.2 lbs to be 0.1227. Why do you get a different answer to **Question #6**?

In Question #6 we are looking at the sampling distribution of sample mean weights, whereas in Lab 13 we were looking at the population distribution of squirrel weights. They are centered at

the same mean, but the sampling distribution has a lower standard deviation, meaning that individual values are more likely than sample means to be far away from the population mean.

Question #8 Create a density histogram of the sample means of squirrel weights. Overlay the normal density curve. Paste the graph below.

Simulated Samples of Size 4
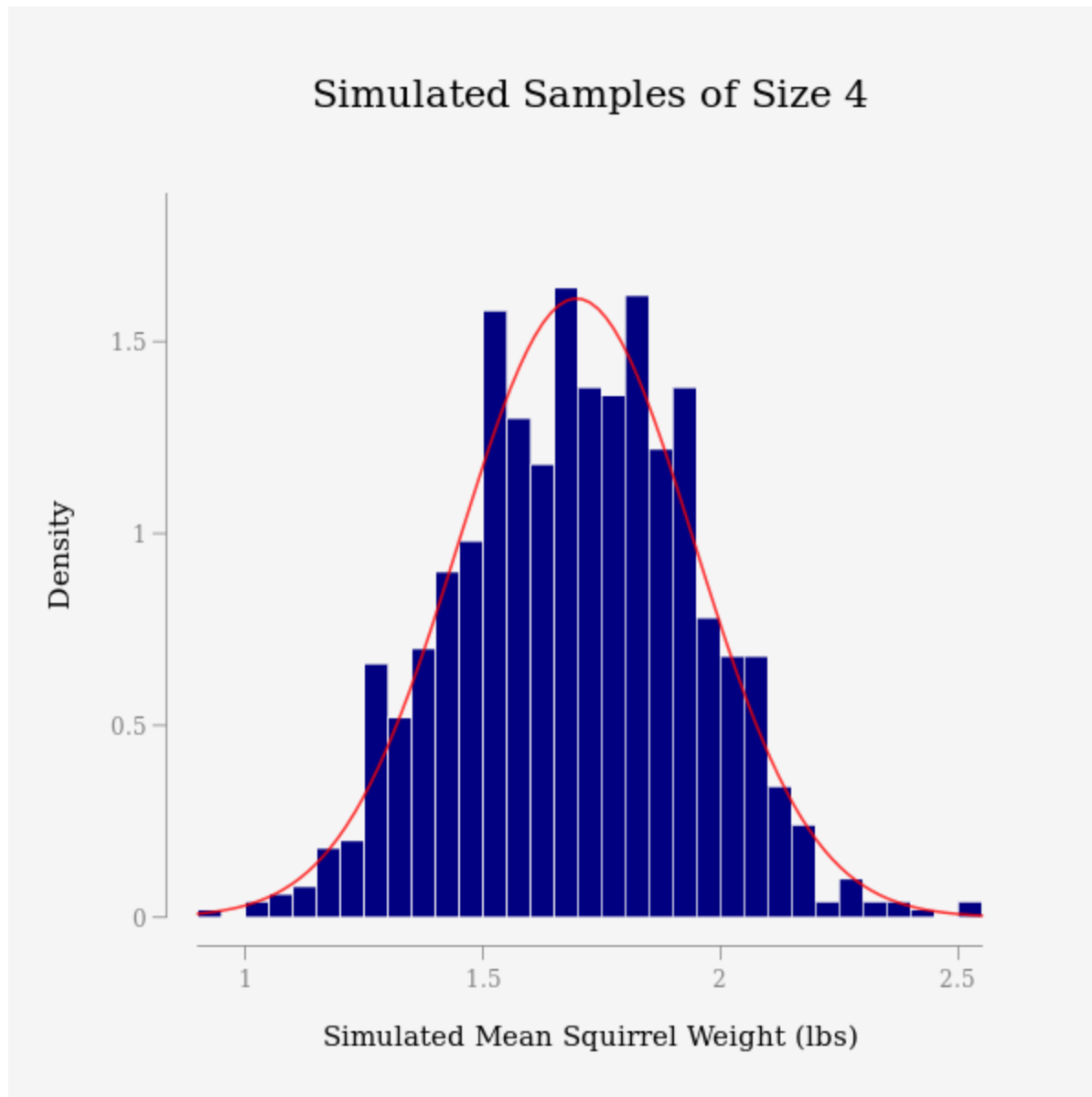
Right-click the dataset and select *Functions* ✉ *Transform*. Create a new variable, **mean_range**, and in the *Transformation* box, type the R code shown in the figure below. This code finds the number of simulated sample means that are greater than 0.8 but less than 1.2.
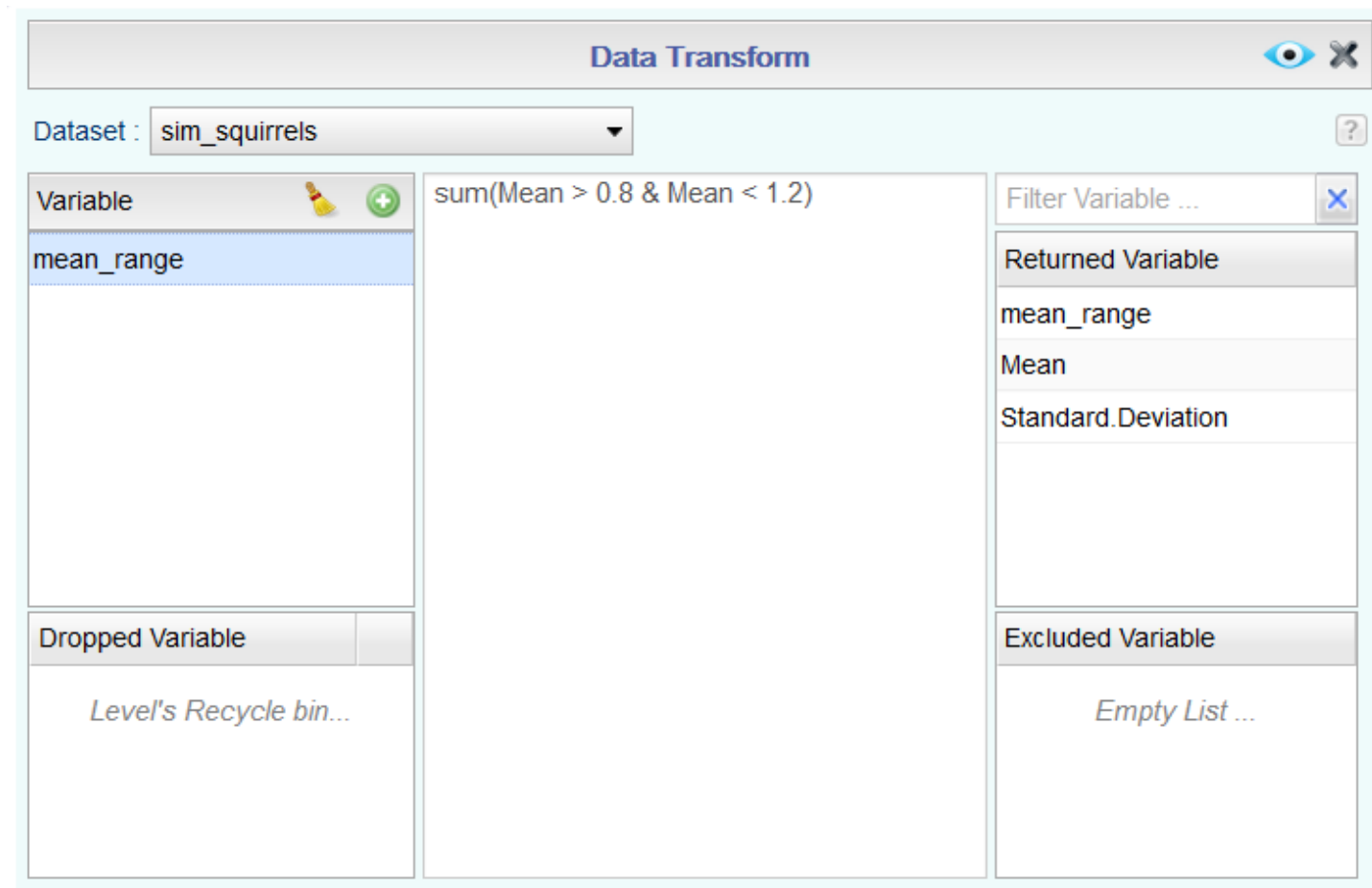
Question #9 How many of your sample means were between 0.8 and 1.2 lbs? What proportion is that? Is the proportion close to what you calculated in **Question #6**?

19 out of the 1000 simulated sample means, or 1.9%, were between 0.8 and 1.2 lbs. This is also somewhat close to what was calculated in #6 but not exact (16% too low).
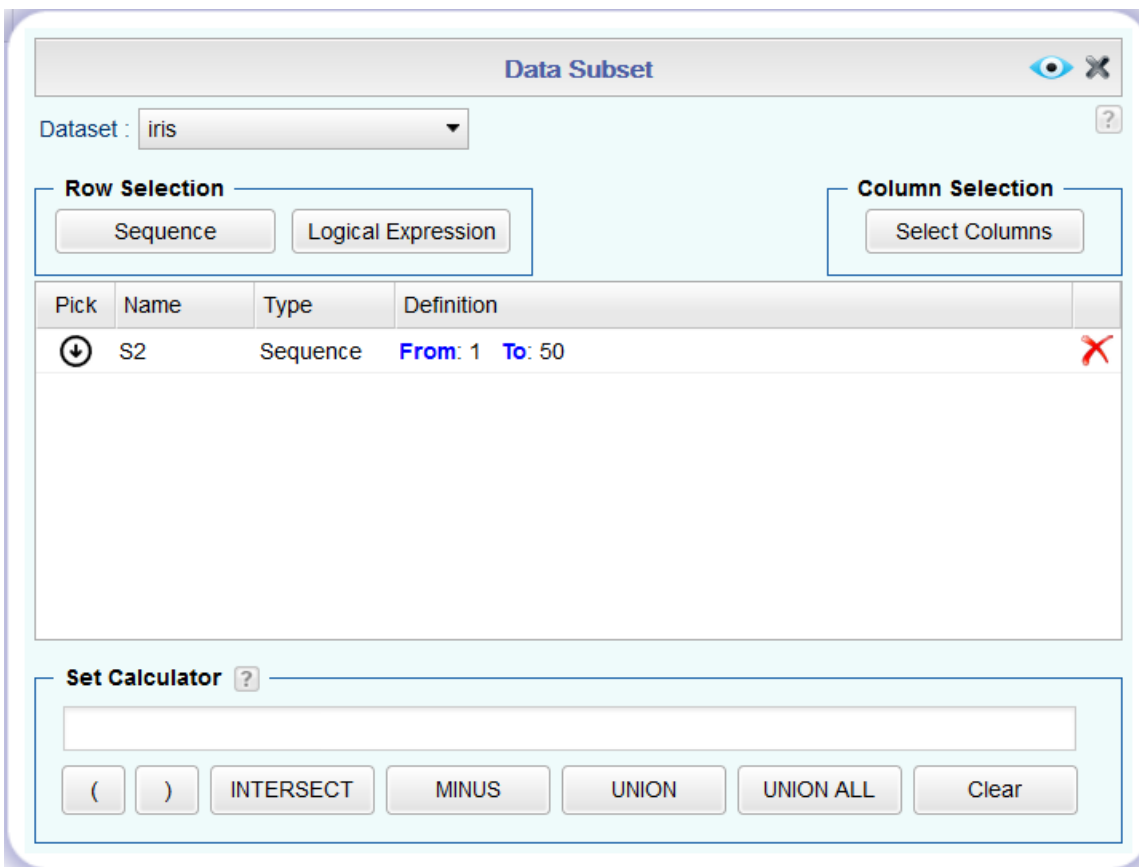
Question #10 Was the first simulated sample of 4 squirrels unusually heavy? Explain your reasoning. (Hint: Use the *Continuous Distribution Calculator* to find the probability of simulating a heavier sample mean of 4 squirrels than the sample mean from **Question #2**.)

Assuming that the sampling distribution of sample mean weights is roughly N(1.7, 0.25), the probability of simulating a sample with a heavier sample mean than we actually got (2.075) is approximately 0.067. Roughly 6.7% of the time we will simulate a heavier sample of size 4.

This doesn't seem particularly unusual, but you might have a different threshold for "unusually heavy."

In this lab, we will be working with a subset of a much larger data set commonly known as Fisher's or Anderson's Iris data set, see Fisher, R.A. (1936), *The use of multiple measurements in taxonomic problems*. The data set describes the petal length, petal width, sepal length, and sepal width of 50 *setosa*, 50 *versicolor*, and 50 *virginica* iris flowers. It can be found in the Rguroo Data Repository (*Data* ✉ *Data Import* ✉ *Repository*) by selecting the *R Datasets* Repository and then scrolling down to find the *iris* dataset. Import the dataset into Rguroo and *View* it.

We will only be working with the petal length for the *setosa* species, which is given by rows 1:50 and the column labeled Petal.Length. Let's extract that data from the original set by clicking on *Data* ✉ *Functions* and selecting *Subset*. Select the *iris* dataset from the *Dataset* dropdown menu. Now, under *Row Selection* click on *Sequence* and input **1** and **50** in the *From* and *To* boxes, then click *Done*. See the figure below for an example of what you should see at this point.

Now, click the *Select Columns* button, and in the dialog that pops up, move the variable *Petal.Length* to the right box ("Selected Columns"). Click *Done*, preview the new subset, and save it as *Setosa.Petal.Length*.

Question #1 Create a histogram of Petal.Length for the *setosa* species and insert it below.

Petal Length of Setosa Species
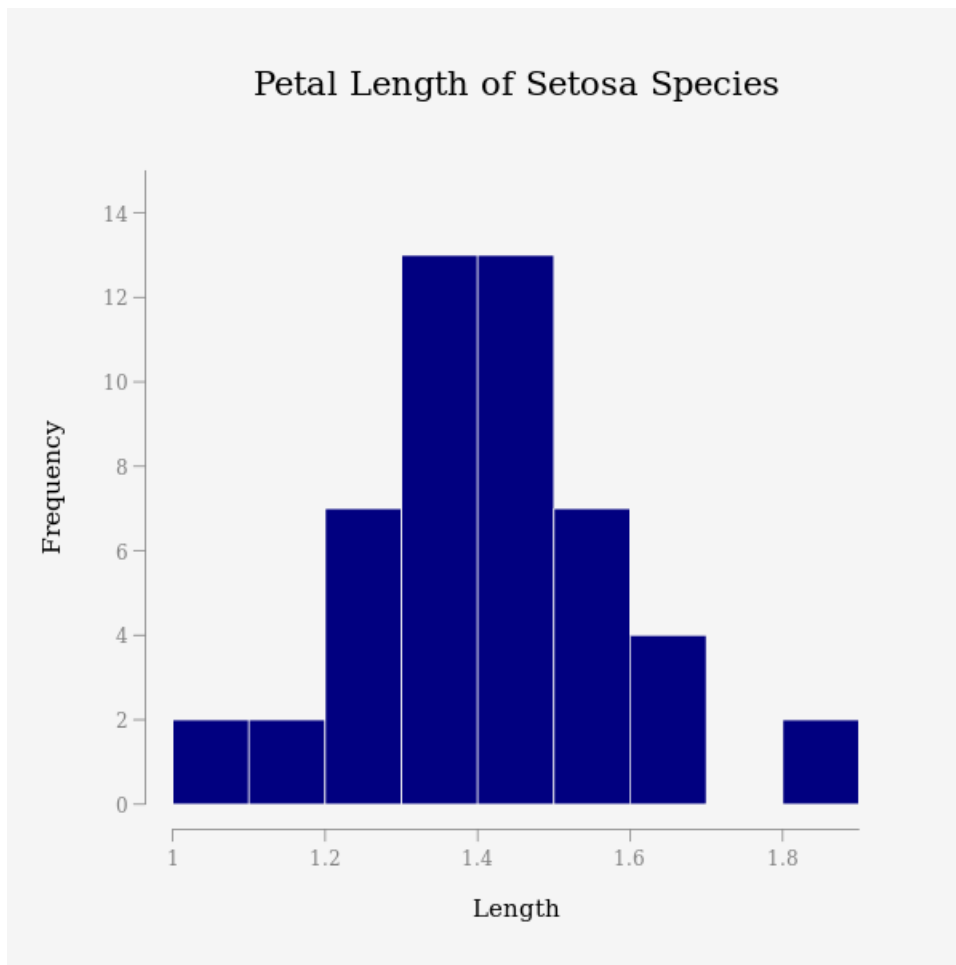
If we believe the Central Limit Theorem is going to be accurate (and assuming the data were collected in a random fashion), we can perform a hypothesis test to test the claim that the population mean petal length is 1.3 centimeters. Assume the population standard deviation of petal lengths is 0.5 cm.

Question #2 Describe the distribution of the sample. Does the shape of the distribution suggest that the Central Limit Theorem will be roughly accurate and so we can model the sampling distribution of all possible sample means as a normal distribution?

The sample is roughly normally distributed, suggesting that the Central Limit Theorem should hold for samples of size 50 and therefore that the sampling distribution of all possible sample means can be modeled as a normal distribution.

Question #3 Write the null hypothesis for this test using an appropriate symbol for the parameter.

H_0: μ = 1.3 cm

Question #4 What is the distribution of sample means under the null hypothesis? Remember that we need to specify the family/shape of the distribution and all relevant parameters of the distribution.

Under H$_0$, the sample mean is distributed N(1.3, 0.5/sqrt(50)) = N(1.3, 0.071) by Central Limit Theorem

First let's work in the Neyman-Pearson framework. Recall that in this framework we need to specify an alternative value of our parameter. Let's suppose that a meaningful difference is an increase of 1.5 cm; that is, under the alternative hypothesis, our parameter value is 1.5 cm.

Question #5 Assuming α = 0.05, find the critical region for this test.

The critical region is $\bar{x} \geq 1.4163$.

Question #6 Is our observed sample mean in the critical region? What should we conclude about the population mean petal length (in the N-P framework)?

Our observed sample mean of 1.462 is in the critical region. Thus, we can accept the alternative hypothesis and conclude that the population mean petal length is 1.5 cm.

Question #7 What is the sampling distribution of our sample means if in fact the alternative hypothesis H$_1$ is correct?

Under H$_0$, the sample mean is distributed N(1.5, 0.5/sqrt(50)) = N(1.5, 0.071) by Central Limit Theorem

Question #8 What is the power of our test to detect the specific alternative value of 1.5 cm?

The power of our test to detect the specific alternative value of 1.5 cm is about 0.882.

Now let's work in the Null Hypothesis Significance Testing framework. Recall that in this framework our alternative hypothesis is an <u>inequality</u>. Let's assume that if the population mean is not 1.3 cm, then it must be larger than 1.3 cm.

Question #9 Write the alternative hypothesis H$_a$ for this test using an appropriate symbol for the parameter.

H$_a$: μ > 1.3 cm

Recall that the p-value is the probability of observing a sample with <u>our observed value</u> of the test statistic, or a value more favorable to the alternative hypothesis. Use the *Probability Calculator* with an appropriate distribution and inequality statement to find the p-value corresponding to our <u>observed</u> sample mean.

Question #10 What is the p-value for our test? Using a 5% significance level, what should we conclude about the population mean petal length (in the NHST framework)?

The p-value is approximately 0.011. At a 5% significance level, we reject H$_0$ and conclude that the population mean petal length is greater than 1.3 cm.

Lab Assignment #17

Let's again work with our *sim_squirrels* dataset from Lab 15. Recall that this dataset has the sample mean and sample standard deviation of 1000 different samples of size 4 from a N(1.7, 0.5) distribution. You may want to start by double-clicking the dataset to *View* it.

Question #1 Compute the standard error of the mean for the *first* sample.

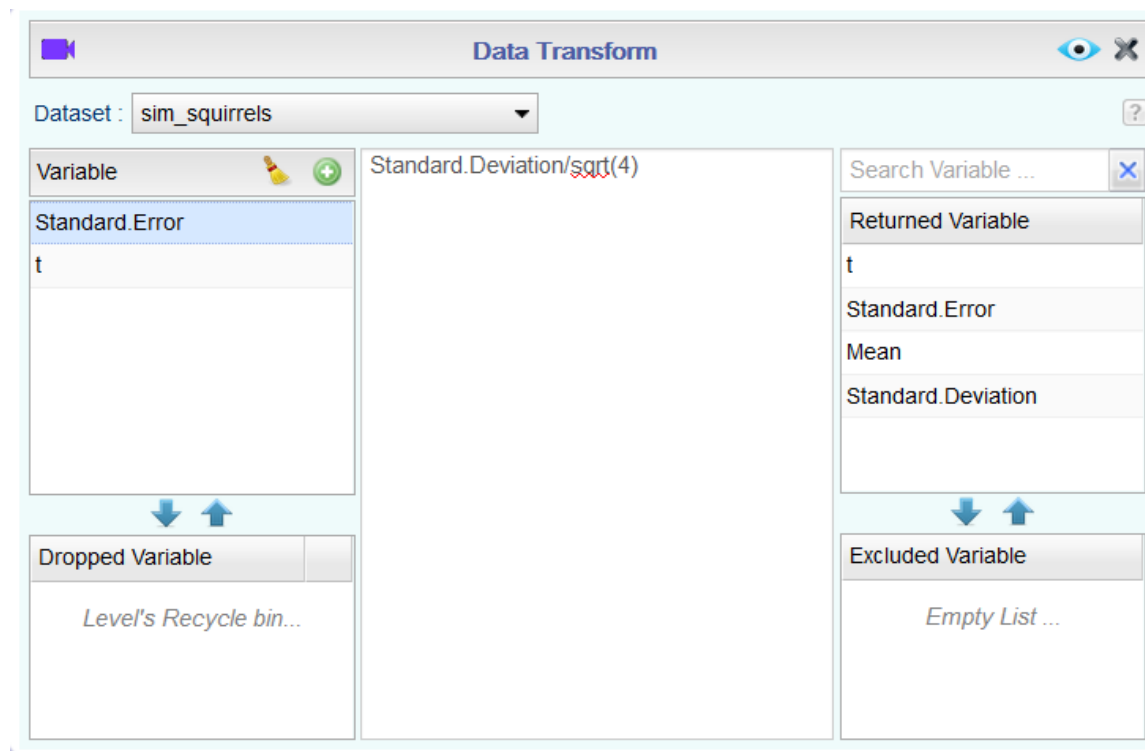**The sample standard deviation of the first sample is 0.6085. Thus the standard error is 0.6085/sqrt(4) = 0.3043.**

Question #2 For the *first* sample, compute the t-statistic corresponding to the sample mean.
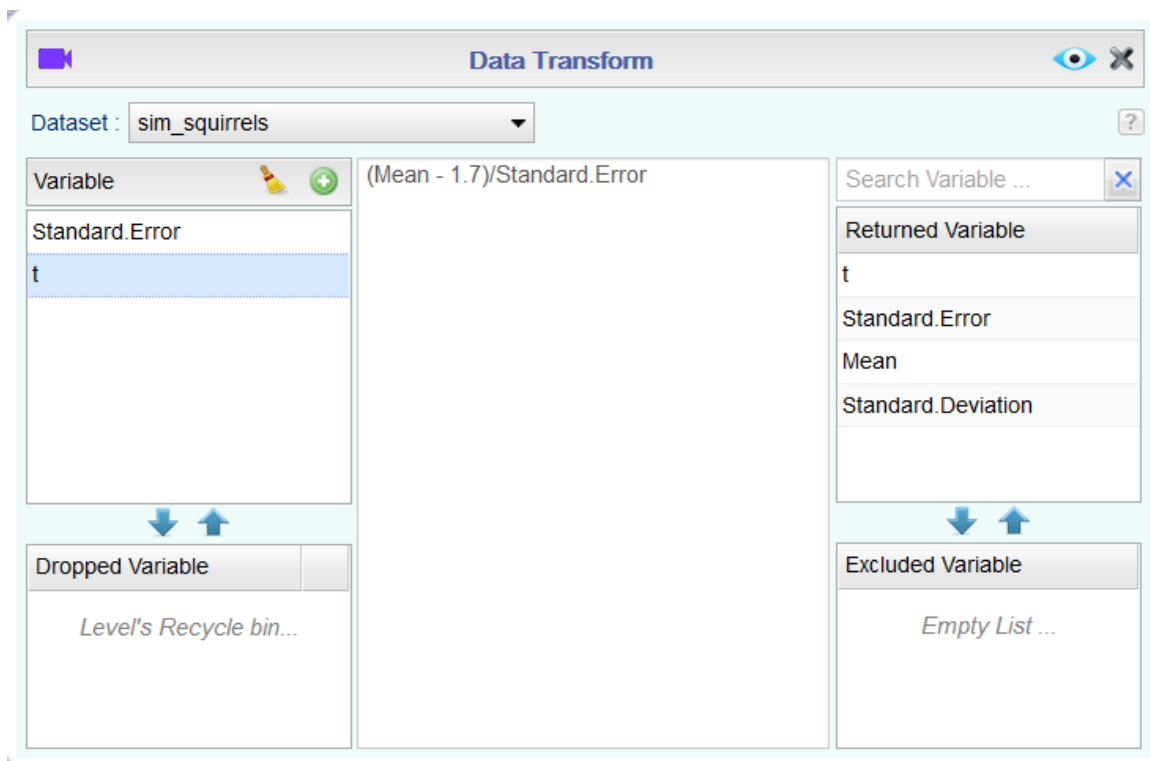
The sample mean is 2.075. Thus the t-statistic is

$$\frac{\dot{x}-\mu}{standard\ error} = \frac{2.075-1.7}{0.3043} = 1.23$$

Next, we will do the process automatically to get the distribution of all 1000 t-statistics.

Right-click the dataset, then click *Functions* ✉ *Transform*. Add two variables, *Standard.Error* and *t*, and fill in the center box as shown below.

Data Transform

Dataset : sim_squirrels

Variable

(Mean - 1.7)/Standard.Error

Search Variable ...

Standard.Error

Returned Variable

t

t

Standard.Error

Mean

Standard.Deviation

Dropped Variable

Excluded Variable

Level's Recycle bin...

Empty List ...

Save the resulting dataset as *sim_squirrels4*. This will remind you that the standard errors and t-statistics are computed based on samples of size 4.

Question #3 Create a histogram of the 1000 t-statistics and paste it below. Overlay both the density curve and the normal curve. How is the distribution of the t-statistics similar to a normal distribution? How is it different?

Simulated t-Statistics for Samples of Size 4
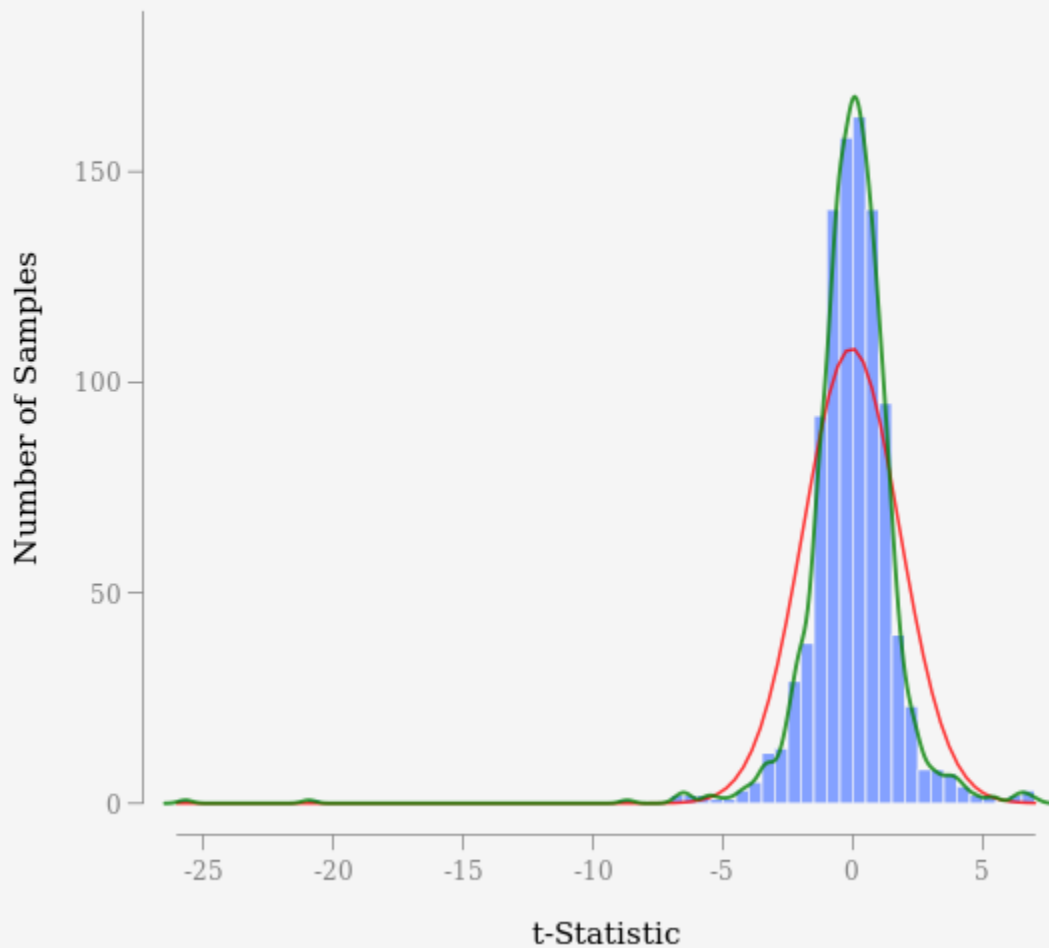
The distribution of t-statistics is also a bell curve. It <u>should</u> look flatter than the normal distribution but in this case there are a couple of bizarre outliers (below -20) that mess up the normal distribution curve.

Question #4 What are the degrees of freedom associated with this t-statistic?

Degrees of freedom = sample size – 1 = 3

Go back to your simulation from Lab 15 and change the sample size from 4 to 100. Simulate 1000 new sample means and sample standard deviations.

Question #5 Compute the standard error of the mean for the *first new* sample.

The sample standard deviation of the first sample is 0.4763. Thus the standard error is 0.4763/sqrt(100) = 0.0476.

**Question #6** For the *first new* sample, compute the t-statistic corresponding to the sample mean.

Just like before, use the *Transform* dialog to automatically create all 1000 t-statistics. Remember to change the sample size in your formula for the t-statistic! Save the resulting dataset as *sim_squirrels100*.

Question #7 Create a histogram of the 1000 *new* t-statistics and paste it below. Overlay both the density curve and the normal curve.
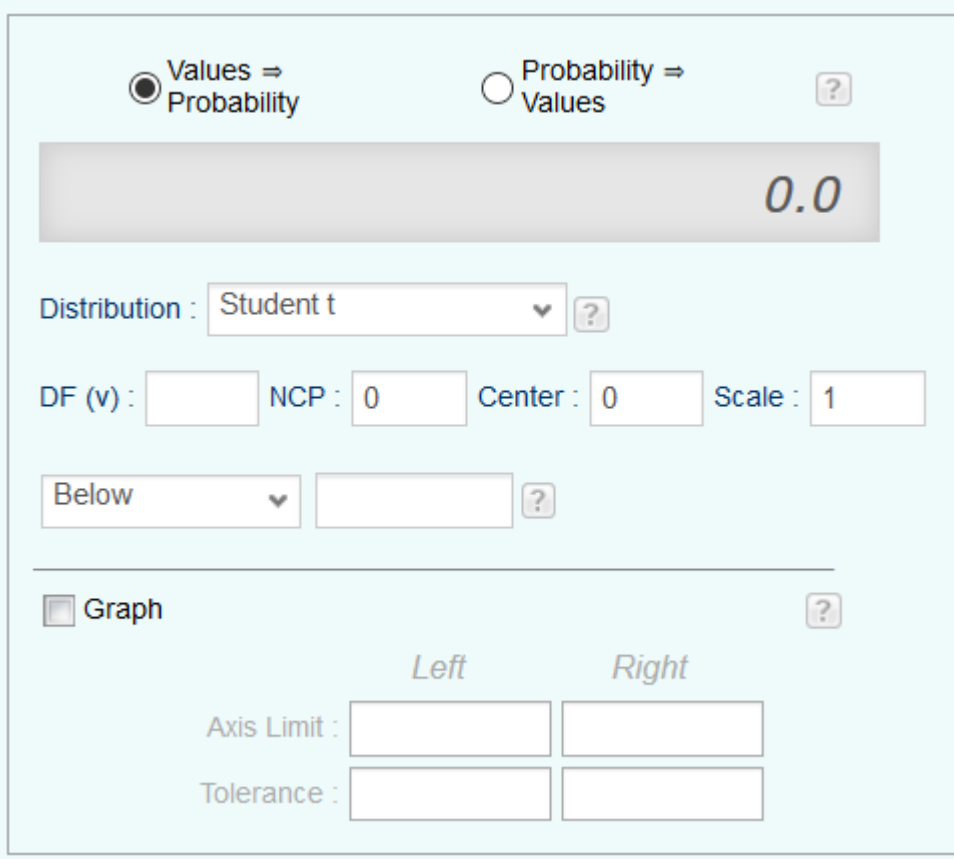

Simulated t-Statistics for Samples of Size 100
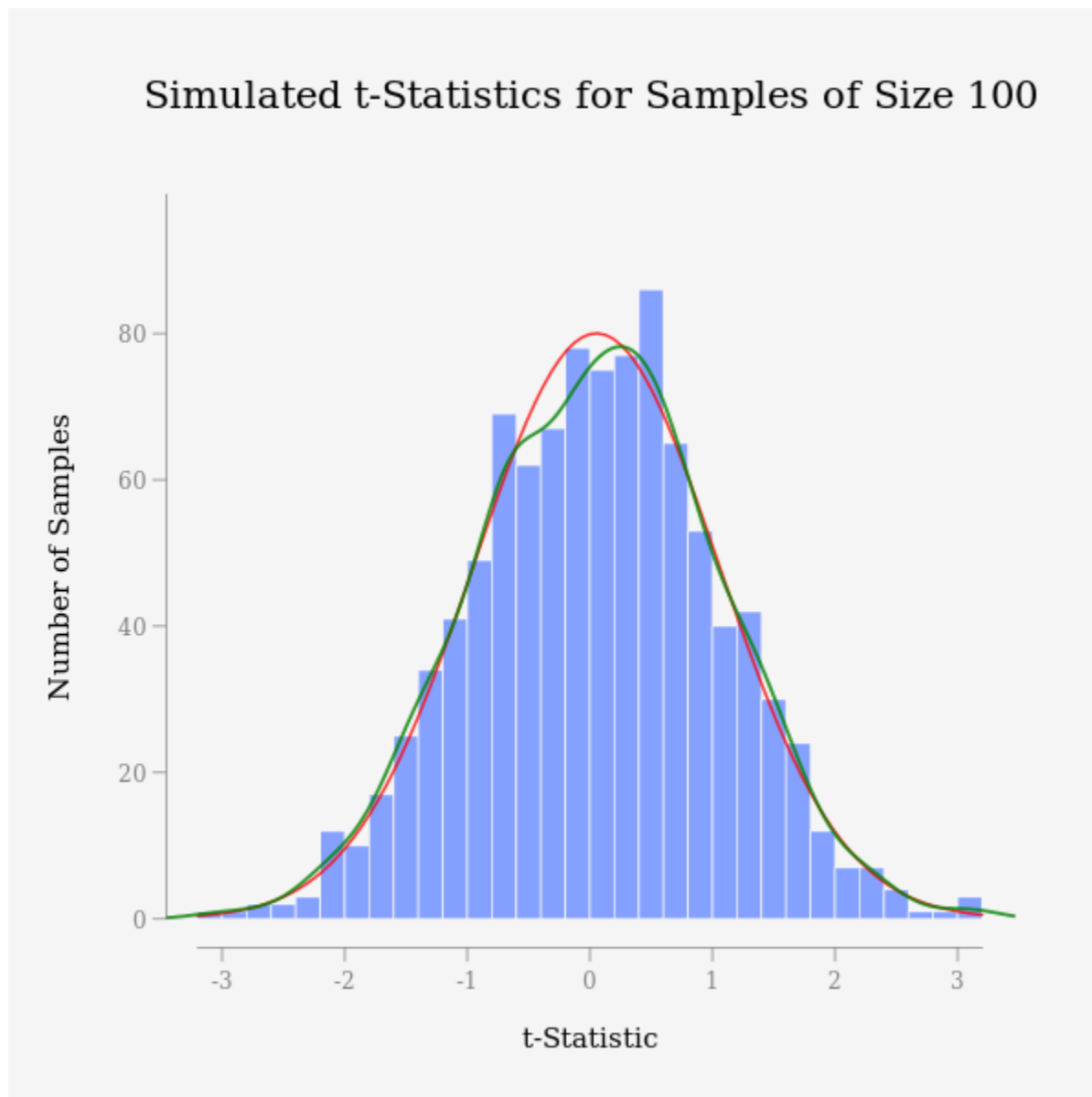
Question #8 What are the degrees of freedom associated with this t-statistic?

Degrees of freedom = sample size – 1 = 99

Finally, we will do some probability computations involving the t-statistic.

Open the *Distribution Calculator* ✉ *Continuous* and select the *Student t* Distribution.

⦿ Values ⇒ Probability	◯ Probability ⇒ Values

0.0

Distribution : Student t ▾ ?

DF (v) : [ ]   NCP : 0   Center : 0   Scale : 1

Below ▾ [ ] ?

☐ Graph

        Left   *Right*

Axis Limit : [ ] [ ]

Tolerance : [ ] [ ]

In the *DF* box, enter the degrees of freedom for the distribution. Ignore all the other boxes on that line. Then, fill in the line below it to answer **Question #9**.

Question #9 For a sample of size 4, what is the probability of obtaining a sample mean at least 2 standard errors away from the population mean? What about for a sample of size 100?

For a sample of size 4, the probability of getting a sample mean more than 2 standard errors away from the population mean is 0.139; however, for a sample of size 100, this probability decreases to 0.048.

Question #10 Briefly describe how the standard error and t-distribution change as the sample size increases.

As the sample size increases, the standard error decreases and the t-distribution changes to look more normal – less probability in the tails.

In Lab 16 we tested the null hypothesis $\mu = 1.3\,cm$ against the alternative hypothesis $H_1 : \mu = 1.5\,cm$ (in the Neyman-Pearson framework) or $H_a : \mu > 1.3\,cm$ (in the NHST framework). We assumed that the data were collected in a random fashion (or at least that the iris sample can be treated as such), that the Central Limit Theorem holds for samples of iris flowers of size 50, and that the population standard deviation of petal length for our particular *setosa* species is 0.5 cm.

Now we will do the same test, except we will <u>not</u> make any assumptions about the population standard deviation.

Question #1 Compute the standard error of the mean for our sample of 50 *setosa* petal lengths.

Standard error = $s / \sqrt{n} = 0.176334 / \sqrt{50} = 0.0246$

Question #2 Compute the t-statistic corresponding to the sample mean under H_0.

$$t = \frac{\acute{x} - \mu}{s / \sqrt{n}} = \frac{1.462 - 1.3}{0.0246} = 6.584$$

(The actual value is about 6.596 but due to rounding this may be a little off)

Question #3 What are the degrees of freedom associated with this t-statistic?

Degrees of freedom = n – 1 = 49

Open the *Distribution Calculator* ✉ *Continuous* and select the *Student t* Distribution. In the *DF* box, enter the degrees of freedom for the distribution. Ignore all the other boxes on that line. Then, select the appropriate type of calculation (*Values* ✉ *Probability* or *Probability* ✉ *Values*) and fill in the rest of the dialog to answer Questions #4 and #5 below.

Question #4 What is the critical region on the *t* scale for α = 0.05? Do you still accept the alternative hypothesis H_1?
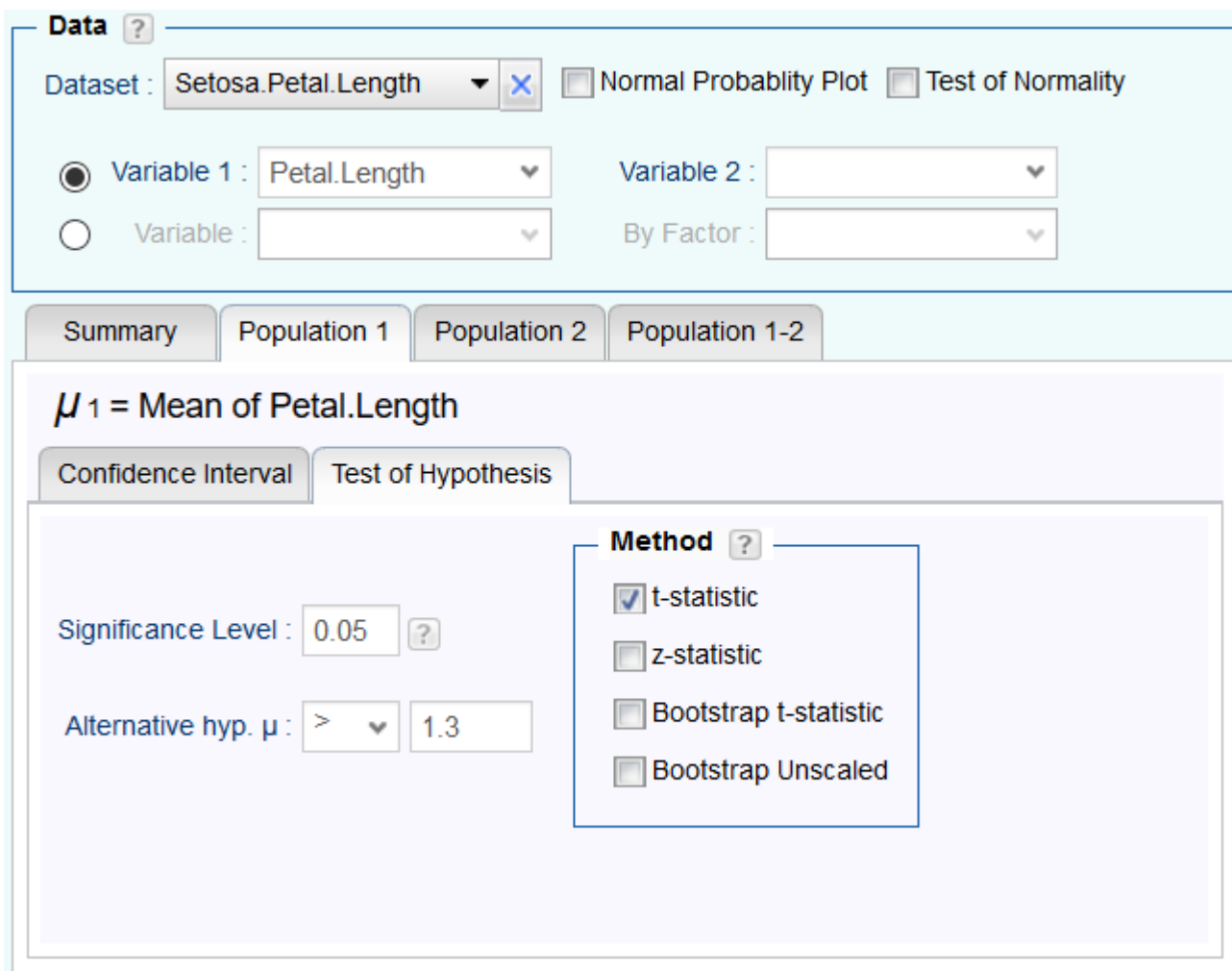
The critical region is $t \geq 1.677$. Since the observed t-statistic value of 6.596 is in the critical region, we still accept H_1.

Question #5 What is the probability of getting a sample with a *t* statistic as large or larger than the value you obtained (in **Question #2**)? At a significance level of 0.05, do you still reject the null hypothesis H_0 and accept the alternative hypothesis H_a?

The probability is approximately 1.4 x 10^{-8}. At a significance level of 0.05, we definitely still reject the null hypothesis.

Now let's do all of this work in the *Mean Inference* section. Click on *Analytics* ✉ *Mean Inference*. Select the Setosa.Petal.Length Dataset and select Petal.Length as *Variable 1*. The boxes under *Population 1* at the bottom of the dialog should now be filled in with values of the sample statistics.

To do a one-sample *t*-test, click the *Population 1* tab. Select the *Test of Hypothesis* tab and fill in the appropriate alternative hypothesis (in the NHST framework). By default the *z-statistic* box is checked, uncheck it and check the *t-statistic* box instead.

Question #6 Paste the output of the t-test (I just need the table with the p-value). Do you get the same t-statistic and p-value as in **Question #5**?

Test of Hypothesis: t-Test
Petal.Length

Research Hypothesis Ha: Mean of 'Petal.Length' is greater than 1.3
5% upper critical value in units of data = 1.34118

Sample Mean	Std Error	Obs t Stat	DF	5% t-Upper Critical	P-value
1.46200	0.0245598	6.59615	49	1.67655	1.40136e-08

Test is significant at 5% level.

t-statistic and p-value are roughly the same (+/- a little bit of rounding error) as in Question #5.

To perform a t power analysis in Rguroo, click the *Details* button and select the *Power Analysis* tab. Everything should already be filled in except the alternative value of the population mean under H_1. Fill that in, then preview.

Question #7 Paste the output of the power analysis (again, just the table is sufficient). Is 50 flowers a large enough sample to detect the difference between $\mu = 1.3\,cm$ and $\mu = 1.5\,cm$?

Power: t-Test for Mean
Petal.Length

Research Hypothesis Ha: Mean of 'Petal.Length' is greater than 1.3
Sample Size = 50
Standard Deviation = 0.173663996480184
Significance Level = 5%

Null	Alternative	Effect Size	Approx. Power	Exact Power
1.30000	1.50000	1.15165	1.00000	1.00000

Approximate Power is computed via normal approximation.

The power is 1, so 50 flowers is more than enough. Perhaps we should use a lower significance level to reduce power so that α ≤ β.

Now let's investigate what would happen if we used a two-sided alternative hypothesis instead. We test the null hypothesis $\mu = 1.3\,cm$ against the alternative hypothesis $H_a : \mu \neq 1.3\,cm$ (in the NHST framework).

In *either* the *Basics* ✉ *Population 1* ✉ *Test of Hypothesis* or *Details* ✉ *Power Analysis* sections, change the > sign to ≠.

Question #8 Paste the output of the t-test (I just need the table with the p-value). Compare the values of the test statistic and p-value with that of the one-sided test (in **Question #6**).

Test of Hypothesis: t-Test
Petal.Length

Alternative Hypothesis Ha: Mean of 'Petal.Length' is not equal to 1.3
2.5% lower critical value in units of data = 1.25065
2.5% upper critical value in units of data = 1.34935

Sample Mean	Std Error	Obs t Stat	DF	2.5% t-Lower Critical	2.5% t-Upper Critical	P-value
1.462	0.02456	6.5961	49	-2.0096	2.0096	2.8027e-08

Test is significant at 5% level.

The t-statistic has the same value but, as expected, the p-value doubled.

Using a two-sided alternative hypothesis in the NHST framework is equivalent to splitting the critical region so that half is below 1.3 cm and the other half is above 1.3 cm. Thus, when we do Neyman-Pearson testing, we need to calculate the probability of being in each part of the critical region separately, then add those probabilities.

Question #9 Paste the output of the power analysis (again, just the table is sufficient). Is 50 flowers still a large enough sample to detect the difference between $\mu=1.3\,cm$ and $\mu=1.5\,cm$?

Power: t-Test for Mean
Petal.Length

Alternative Hypothesis Ha: Mean of 'Petal.Length' is not equal to 1.3
Sample Size = 50
Standard Deviation = 0.173663996480184
Significance Level = 5 %

Null	Alternative	Effect Size	Approx. Power	Exact Power
1.3	1.5	1.1516	1	1

Approximate Power is computed via normal approximation.
Exact Power is computed based on a non-central t distribution.

50 flowers is still more than enough to detect the difference.

In this lab we will explore the difference between paired and two independent samples t-procedures.

Note: this lab is long. To compensate, Lab 20 will be shorter and both labs will be due at the same time.

Researchers were interested in "proving" statistically that the pockets in women's jeans are a sham. For each brand, researchers recorded the measurements of the front and back pocket in 20 pairs of straight or boot-cut women's jeans and the same measurements for an equivalent-style pair of men's jeans from the same brand at the same store.

Assume that these 20 brands constitute a random sample of all brands of jeans that offer both men's and women's jeans (and that there are way more than 20 brands, making this a sample from a very large population).

Question #1 Given the above description, which type of t-test (one sample, matched pairs or two independent samples) is most appropriate? Why?

Matched pairs is most appropriate because each women's jean is matched with a men's jean.

This lab requires two datasets found on Titanium, straight_jeans and straight_jeans2. These two datasets contain the same information but recorded in two different ways. Import both files to Rguroo.

Question #2 What is an observational unit in the straight_jeans dataset (that is, what does one row in the dataset represent)? What is an observational unit in the straight_jeans2 dataset?

In the straight_jeans dataset, an observational unit is a pair of jeans. In straight_jeans2, an observational unit is a brand of jeans.

For this lab, we will focus on the width of the back pocket. The variable maxWidthBack represents the width (in cm) of a back pocket at the opening and the variable minWidthBack represents the width (in cm) of the same pocket at the point before it starts to taper.

First, create a histogram using the straight_jeans dataset. Select the variable *maxWidthBack*. In the *Plot by Group* section, select *menWomen* as the *Factor* variable. Check the *Uniform x-Limit* box. This will ensure that both histograms (one for men and one for women) are plotted on the same x-axis scale, so that we can compare center and variability as well as shape.

Histogram

Data [?]

Dataset : straight_jeans ▼

Variable : maxWidthBack ▼

Label [?]

Title :

X-Axis :

Y-Axis :

Bins & Bars [?]

Method : Freedman-Diaconis ▼

Color : #84A1FF

Border : whitesmoke

Plot by group [?]

Factor : menWomen ▼   ⦿ Multiple ○ Overlay

Rows :   ⬍ Cols :   ⬍ ☑ Uniform x-Limit

Type [?]

⦿ Frequency

○ Relative Frequency

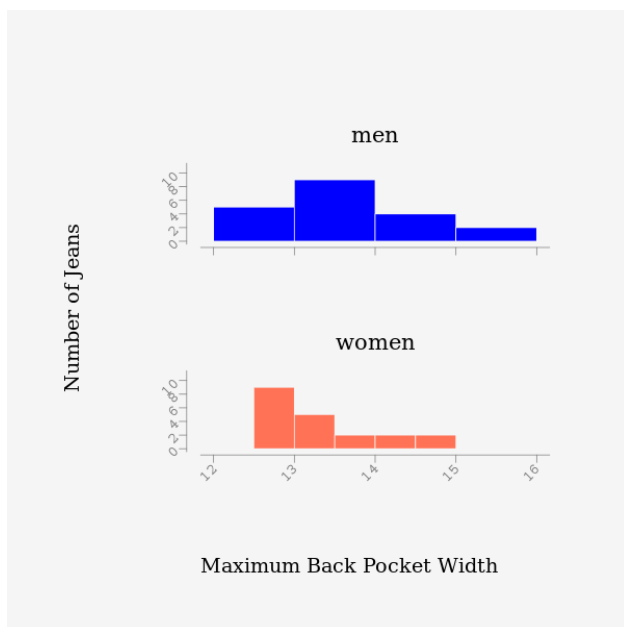○ Density

Value Label [?]

☐ Counts

☐ Percent

Smoothing [?]

☐ Density

☐ Normal

Question #3 Paste below your figure (two total histograms). What differences do you notice between men's and women's jeans based on the variable *maxWidthBack*?

The shapes are both skewed right, but the men's jeans are much more variable.

The men's jeans have a peak at 13-14 but women's jeans have a peak at 12.5-13. The men's jeans look like they are larger on average.

Now, open up a new histogram dialog and do the same thing for the variable *minWidthBack*.

Question #4 Paste below your figure (two total histograms). What differences do you notice between men's and women's jeans based on the variable *minWidthBack*?
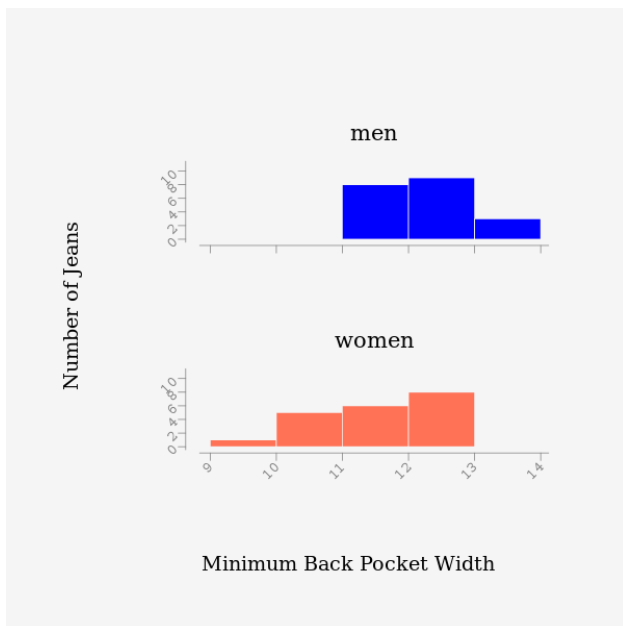

Here men's jeans are clearly larger than women's jeans. Although both have a peak at 12-13, the men's jeans are skewed right and the women's jeans are skewed left.

The women's jeans seem to have more variability.

Now, go to *Analytics* ✉ *Mean Inference* ✉ *One and Two Population*. Select straight_jeans as your dataset. Click the radio button to enable the *Variable/By Factor* variable entry, then select price as your *Variable* and gender as your *Factor*. In the *Summary* tab below, select men as the *Level* for *Population 1* and women as the *Level* for *Population 2*.

Mean Inference

Data ?

Dataset : straight_jeans ▼ ✕   ☐ Normal Probablity Plot ☐ Test of Normality

○ Variable 1 :   Variable 2 :

◉ Variable : maxWidthBack ▼   By Factor : menWomen ▼

| Summary | Population 1 | Population 2 | Population 1-2 |

☐ Paired Data

Population 1 ?

Level : men ▼

Label : men

Sample Mean : 13.845

Sample S.d. : 0.89881

Pop. S.d. :

Sample Size : 20

Population 2 ?

Level : women ▼

Label : women

Sample Mean : 13.485

Sample S.d. : 0.78087

Pop. S.d. :

Sample Size : 20

Click on the *Population 1-2* tab. Let's do a *t* hypothesis test (with significance level 0.05). Although the researchers want to prove that the pockets in women's jeans are a sham, we'll pretend like we don't know beforehand which gender should have narrower pockets and use a two-sided test instead (H$_a$: μ_{women} − μ_{men} ≠ 0 or H$_a$: μ_{men} − μ_{women} ≠ 0).

Question #5 Paste below the Rguroo table for the *Test of Hypothesis (t-test)*. Report and interpret the t-statistic and p-value for this test.

Test of Hypothesis: t-test
maxWidthBack (men) - maxWidthBack (women)

Alternative Hypothesis Ha: Mean of 'maxWidthBack (men) - maxWidthBack (women)' is not equal to 0
2.5% lower critical value in units of data = -0.539312
2.5% upper critical value in units of data = 0.539312
Unequal population variances was assumed.

Diff of Means	Std Error	Obs t Stat	DF	2.5% t-Lower Critical	2.5% t-Upper Critical	P-value
0.36	0.26624	1.3522	37.272	-2.0257	2.0257	0.18446

Test is not significant at 5% level.

The t-statistic is 1.352 and the p-value is 0.184. Our difference in sample means was 1.352 standard errors above expected under the null hypothesis, and the probability of getting a t-statistic larger in magnitude if H_0 is correct is 0.184.

Follow the same instructions to perform a two-sample t-test for *minWidthBack*.

Question #6 Paste below the Rguroo table for the *Test of Hypothesis (t-test)*. Report and interpret the t-statistic p-value for this test.

Test of Hypothesis: t-test
minWidthBack (men) - minWidthBack (women)

Alternative Hypothesis Ha: Mean of 'minWidthBack (men) - minWidthBack (women)' is not equal to 0
2.5% lower critical value in units of data = -0.578572
2.5% upper critical value in units of data = 0.578572
Unequal population variances was assumed.

Diff of Means	Std Error	Obs t Stat	DF	2.5% t-Lower Critical	2.5% t-Upper Critical	P-value
0.55	0.2858	1.9244	37.999	-2.0244	2.0244	0.061814

Test is not significant at 5% level.

The t-statistic is 1.924 and the p-value is 0.062. Our difference in sample means was 1.924 standard errors above expected under the null hypothesis, and the probability of getting a t-statistic larger in magnitude if H_0 is correct is 0.062.

Now let's look at what happens if we use matched pairs t-procedures instead. The straight_jeans2 dataset provides data in a format that will be easier to work with in the matched pairs framework.

First, create a histogram using the straight_jeans2 dataset, but this time, instead of selecting a variable, type *maxWidthBackMens – maxWidthBackWomens* in the *Variable* box. This histogram represents the distribution of <u>paired differences</u> once the jeans have been matched by brand. Thus, these paired differences account for the fact that some brands have wider pockets than others.

Question #7 Paste below your histogram. After matching for brand, does one gender appear to have wider pockets based on the maximum width?

Based on the histogram, it looks like for most brands, the men's jeans have wider pockets than women's jeans.

Paired Differences (men - women)
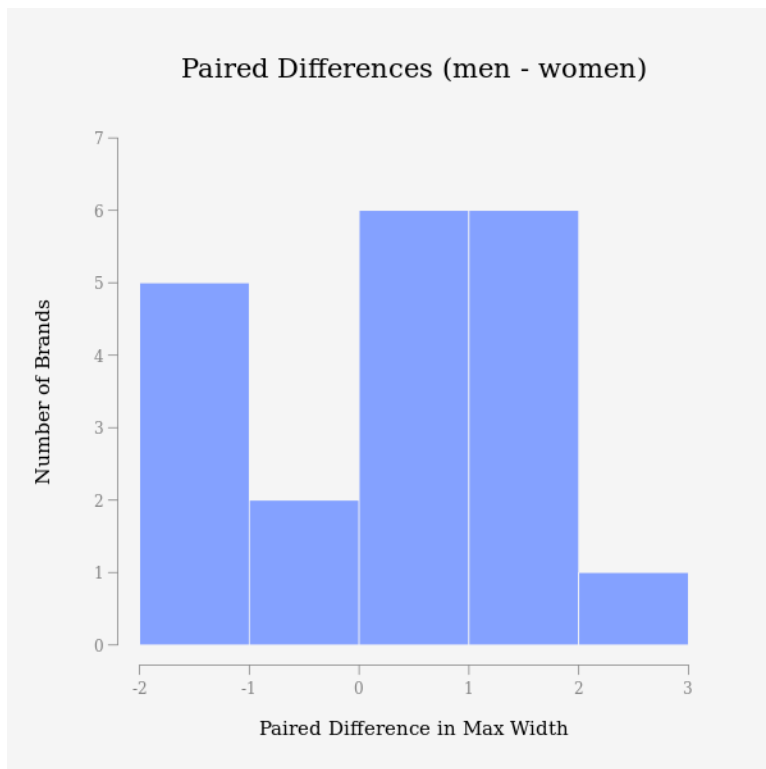
Now do the same for *minWidthBackMens – minWidthBackWomens*.

Question #8 Paste below your histogram. After matching for brand, does one gender appear to have wider pockets based on the minimum width?
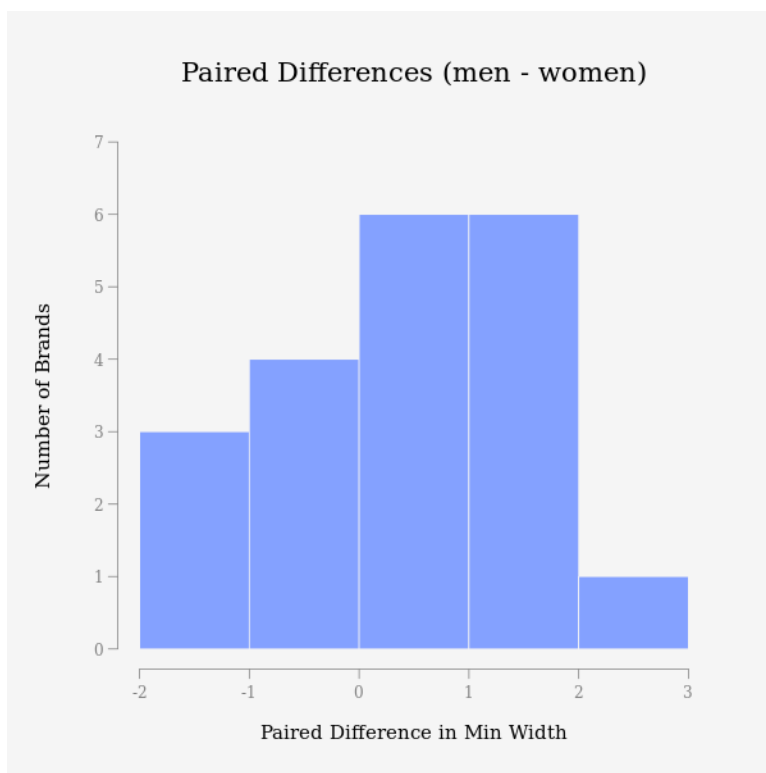
Paired Differences (men - women)

Based on the histogram, it looks like for most brands, the men's jeans have wider pockets than women's jeans.

Now let's do the actual hypothesis test in the matched pairs framework. Open up a new *Analytics ✉ Mean Inference ✉ One and Two Population* with the straight_jeans2 dataset. Click the radio button to enable *Variable 1/Variable 2* as the variable entry, then choose *maxWidthBackMens* as *Variable 1* and *maxWidthBackWomens* as *Variable 2* (or the other way around; since we're doing a two-sided test it

doesn't matter). Check the *Paired Data* box in the *Summary*, then do a *Population 1-2* ✉ *Test of Hypothesis* ✉ *t-statistic*.

Mean Inference

Data ?

Dataset : straight_jeans2 ▾ ✕   ☐ Normal Probablity Plot   ☐ Test of Normality

◉ Variable 1 : maxWidthBackMens ▾   Variable 2 : WidthBackWomens ▾
○ Variable :   ▾   By Factor :   ▾

| Summary | Population 1 | Population 2 | Population 1-2 |

☑ Paired Data

Paired Difference ?

Level :   ▾
Label : xWidthBackWomens
Sample Mean : 0.36
Sample S.d. : 1.2445
Pop. S.d. :
Sample Size : 20

Population 2 ?

Level :   ▾
Label :
Sample Mean :
Sample S.d. :
Pop. S.d. :
Sample Size :

Question #9 Paste below the Rguroo table for the *Test of Hypothesis (t-test)*. Report and interpret the t-statistic and p-value for this test.

Test of Hypothesis: Paired t-test
[maxWidthBackMens - maxWidthBackWomens]

Alternative Hypothesis Ha: Mean of '[maxWidthBackMens - maxWidthBackWomens]' is not equal to 0
2.5% lower critical value in units of data = -0.582456
2.5% upper critical value in units of data = 0.582456
Differences of pairs were used.

Mean of Paired Diff	Std Error	Obs t Stat	DF	2.5% t-Lower Critical	2.5% t-Upper Critical	P-value
0.36	0.27828	1.2936	19	-2.093	2.093	0.2113

Test is not significant at 5% level.

The t-statistic is 1.294 and the p-value is 0.211. Our difference in sample means was 1.294 standard errors above expected under the null hypothesis, and the probability of getting a t-statistic larger in magnitude if H_0 is correct is 0.211.

Open a new *Mean Inference* dialog and follow the same instructions for the minimum back pocket width (*minWidthBackMens* and *minWidthBackWomens*).

Question #10 Paste below the Rguroo table for the *Test of Hypothesis (t-test)*. Report and interpret the t-statistic and p-value for this test.

Test of Hypothesis: Paired t-test
[minWidthBackMens - minWidthBackWomens]

Alternative Hypothesis Ha: Mean of '[minWidthBackMens - minWidthBackWomens]' is not equal to 0
2.5% lower critical value in units of data = -0.476679
2.5% upper critical value in units of data = 0.476679
Differences of pairs were used.

Mean of Paired Diff	Std Error	Obs t Stat	DF	2.5% t-Lower Critical	2.5% t-Upper Critical	P-value
0.55	0.22775	2.415	19	-2.093	2.093	0.025983

Test is significant at 5% level.

The t-statistic is 2.415 and the p-value is 0.026. Our difference in sample means was 2.415 standard errors above expected under the null hypothesis, and the probability of getting a t-statistic larger in magnitude if H_0 is correct is 0.026.

Question #11 Compare your answer to **Question #9** with your answer to **Question #5** and your answer to **Question #10** with your answer to **Question #6**. Did the values of the t-statistics change when we switched frameworks? What about the p-values?

When we switched frameworks, the t-statistics values changed and the p-values changed. For maxWidthBack, the t-statistic was larger (in magnitude) in the two-sample framework. For minWidthBack the t-statistic was larger (in magnitude) in the matched pairs framework.

In this lab, we will test whether there is a difference in the mean sepal length among all three iris species.

To start, let's do some exploratory data analysis. Using the iris dataset, create side-by-side *Boxplot*s of the sepal length for each of the three species by selecting *Sepal.Length* as the *Numerical* variable and *Species* as the *Factor* variable. Add an appropriate title and axis labels. As usual, you can mess around in the *Factor Level Editor* to change colors and other things.
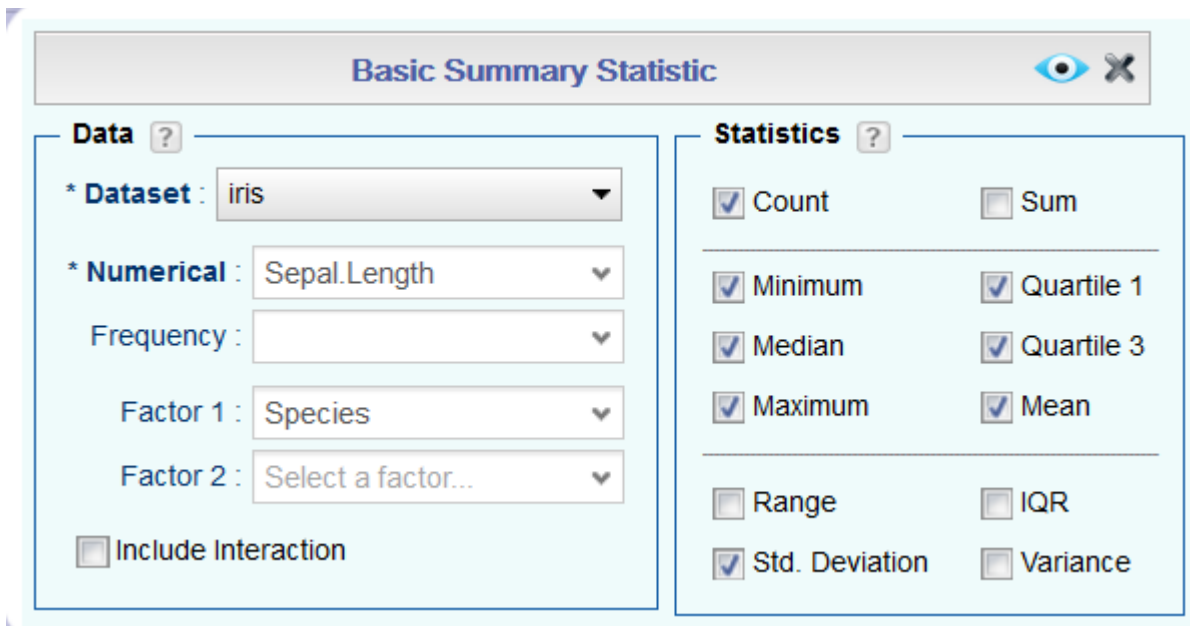
Question #1 Insert the final set of boxplots below.
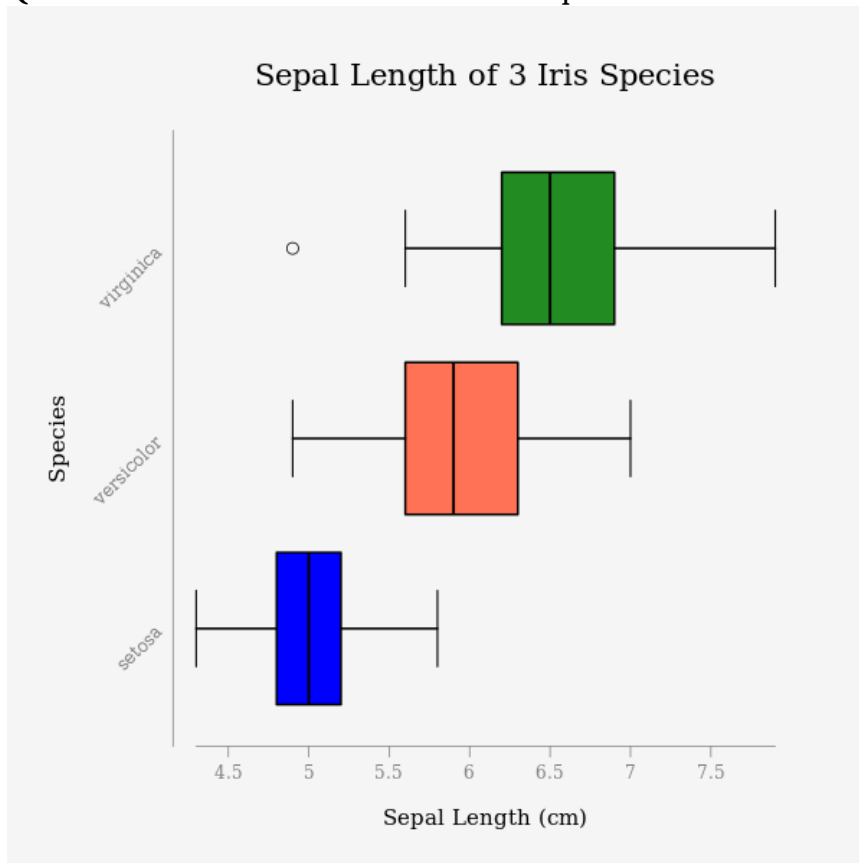
Question #2 Write the (null) hypothesis for a one-way ANOVA test of the claim that the three species have the same mean sepal length. Recall that this is a Fisher-type test and so only a null hypothesis is specified.

$H_0: \mu_{setosa} = \mu_{versicolor} = \mu_{virginica}$

Now let's check our variability. Right-click the dataset and select *Functions ✉ Summary Statistic*. Select *Sepal.Length* as the *Numerical* variable and *Species* as the *Factor 1* as shown below:
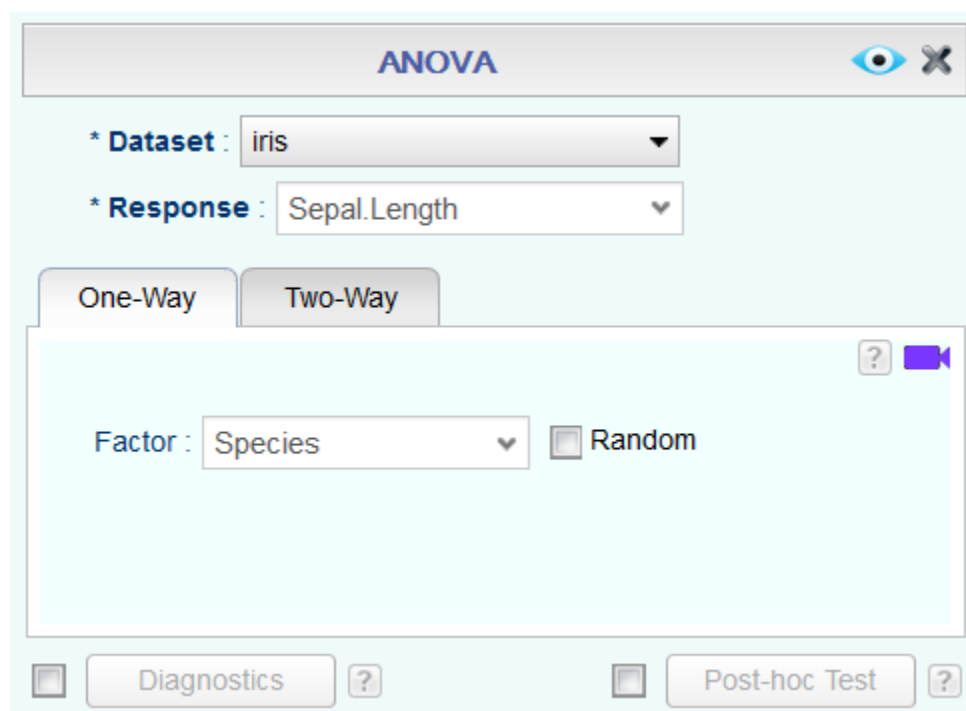
Next, click the + next to *Species* to expand to see the summary by group.

Question #3 What are the three sample standard deviations? Is our rule of thumb about the population standard deviation satisfied?

For setosa the standard deviation is 0.352, for versicolor it is 0.516, and for virginica it is 0.636. Since 0.636 < 2*(0.352), the largest standard deviation is less than twice the smallest and our ANOVA test will be robust to failure of the assumption "population standard deviations are equal."

We will now do the one-way ANOVA test by clicking *Analytics* ☑ *Analysis* ☑ *Anova*. Once you have selected the dataset, select *Sepal.Length* as the *Response* and *Species* as the *Factor* as shown below:

Question #4 Copy and paste the *Anova Table* from the output below. Note that this table does <u>not</u> include the "Total" row shown in lecture (we can derive all numbers in that row from the values in the table).

ANOVA Table
Model: Sepal.Length ~ Species

H0: The means for all levels are equal

Source	DF	Sum of Squares	Mean Square	F Value	Pr>F	BFB
Species	2	63.212	31.606	119.26	0	
Residual	147	38.956	0.26501			

Question #5 From the output in **Question #4**, identify the test statistic, the sampling distribution it comes from (don't forget to include <u>all</u> relevant degrees of freedom parameters), and its observed value in <u>this</u> sample of 150 flowers.

The test statistic is an F statistic and it comes from an F(2, 147) distribution. Its observed value in this sample of flowers is 119.26.

Question #6 What is the p-value for this one-way ANOVA F test? At the 5% significance level, is the model assumption "all three populations have the same mean sepal length" reasonable?

The p-value is essentially 0. We reject the hypothesis and claim that the model assumption "all three populations have the same mean sepal length" is not reasonable.

Your answer to **Question #6** should be that the model assumption is not reasonable. In this case, we want to do *post hoc* tests to determine *which* means are different. Click on *Basics* and this time check the box (in the bottom right) to do a *Post-hoc Test*.

Question #7 Copy and paste below the table labeled *Tukey's HSD: Multiple Comparison of Means*. Which pairs of means appear to be different?

Tukey's HSD: Multiple Comparison of Means

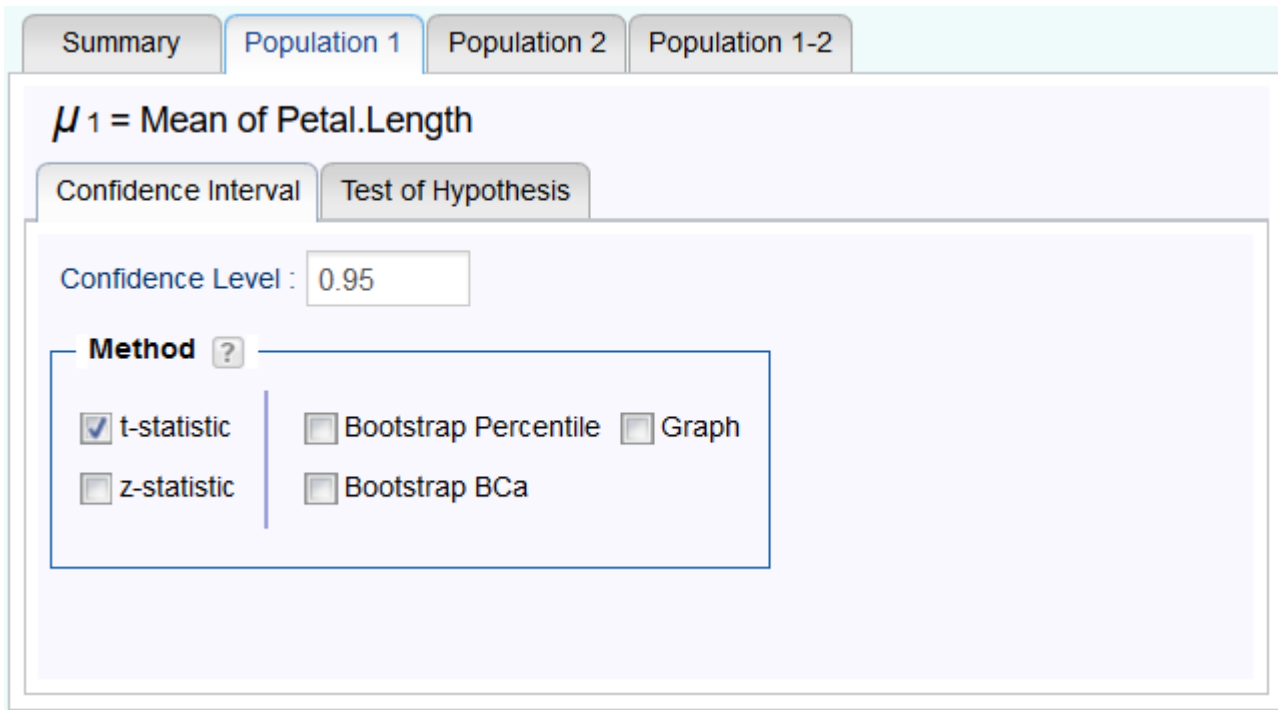Table of 95% family-wise confidence level

Difference Levels	Mean	Lower Limit	Upper Limit	Adjusted p-value
versicolor-setosa	0.93	0.68623	1.1738	3.3862e-14
virginica-setosa	1.582	1.3382	1.8258	2.9976e-15
virginica-versicolor	0.652	0.40823	0.89577	8.2876e-09

Based on the "Adjusted p-value" column, it appears that versicolor is different from setosa and virginica is different from both setosa and versicolor. So all three means are different from each other.

Recall from lecture that confidence intervals come out of Neyman-Pearson ideas. However, instead of making a choice between <u>two</u> possible parameter values, we estimate an entire <u>interval</u> of values within which we suspect the true parameter value lies. Thus, we can use our confidence interval estimates to do two-sided hypothesis tests in the NHST framework.

First, let's open up our Setosa.Petal.Length dataset. Recall that we used this dataset to test the null hypothesis that the population mean petal length was 1.3 cm. Now we will test this null hypothesis using a confidence interval instead.

To obtain a confidence interval for one mean, open the *Mean Inference* ✉ *One and Two Population* dialog. Select the appropriate dataset and variable (remember that we have already subset to include just the *setosa* irises). Then, in the *Population 1* tab, check the box for a *t-statistic* confidence interval:

Question #1 Paste the output table labeled *t-Based Confidence Interval*. Write a sentence interpreting the 95% confidence interval for the population mean petal length of *setosa* irises.

t-Based Confidence Interval

95% Confidence interval

Variable	Mean	Std Error	DF	Lower CL	Upper CL	Margin of Error
Petal.Length	1.46200	0.0245598	49	1.41265	1.51135	0.0493548

We are 95% confident in our estimate that the population mean petal length of *setosa* irises is between 1.413 and 1.511 cm.

Question #2 Based on our confidence interval, does the population mean petal length of *setosa* irises appear to be greater than 1.3 cm or less than 1.3 cm? Or can you not tell? Explain your answer.
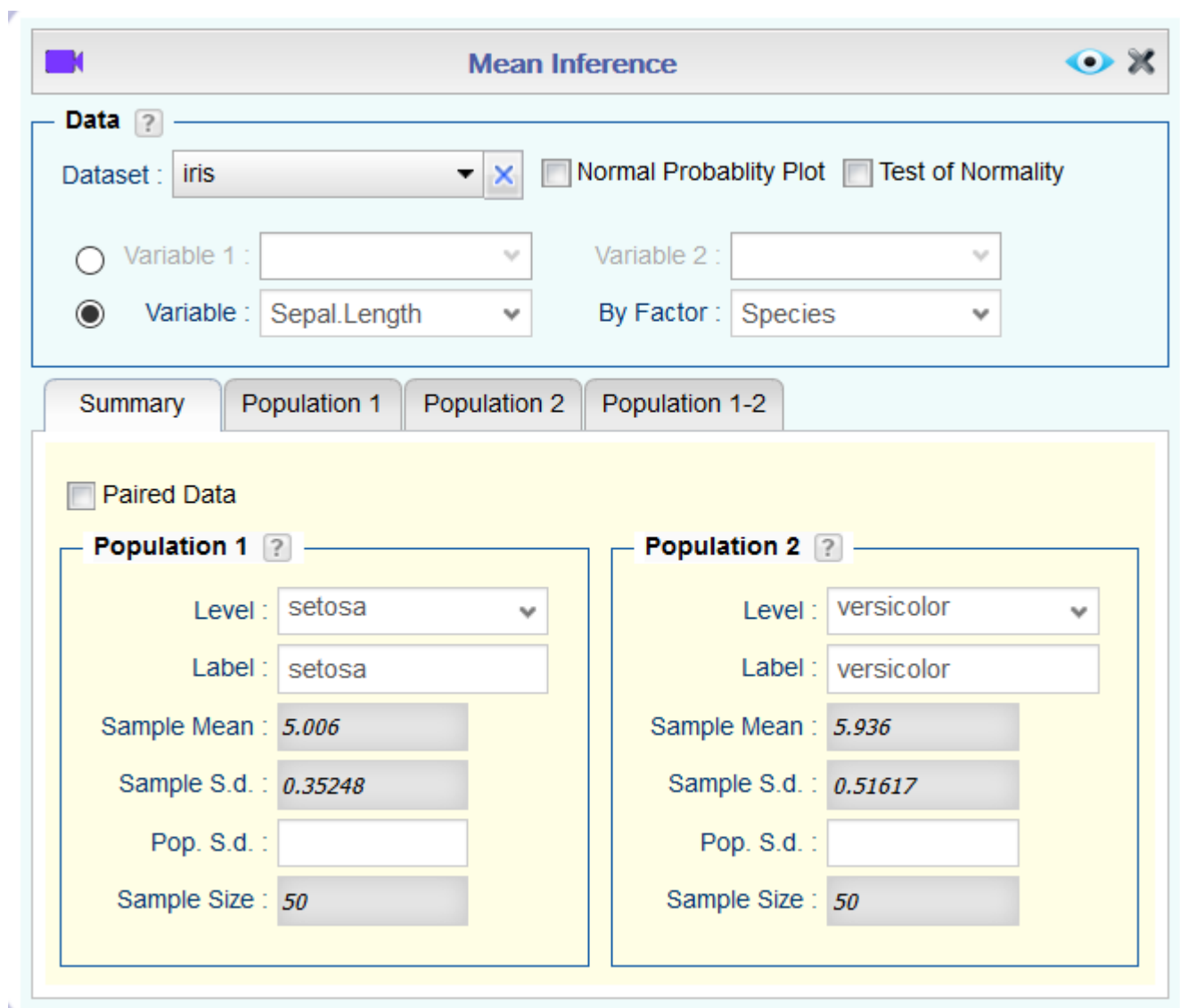
Since the entire confidence interval is greater than 1.3 cm, the population mean petal length of *setosa* irises appears to be greater than 1.3 cm.

Question #3 Based on our confidence interval, can we reject H_0: μ = 1.3 in favor of the two-sided NHST alternative H_a: $\mu \neq$ 1.3 at the 5% significance level? Why or why not?

We can reject the null hypothesis H_0: μ = 1.3 in favor of H_a: $\mu \neq$ 1.3 at the 5% significance level because 1.3 is <u>not</u> in the 95% confidence interval.

Now, let's look at the difference of population mean sepal lengths between *setosa* and *versicolor*. We'll need the *iris* dataset for this one. Since our *setosa* and *versicolor* samples are unrelated, this is a two-sample *t* confidence interval.

In *Mean Inference* ✉ *One and Two Population* dialog, we'll select *Variable/By Factor* and select Sepal.Length as the *Variable* and Species as the *Factor*. Then, we'll select *setosa* as Population 1 and *versicolor* as *Population 2*:

Go to the *Population 1-2* tab and under *Confidence Interval*, check the *t-statistic* box.

Question #4 Paste the output table labeled *t-Based Confidence Interval*. Write a sentence interpreting the 95% confidence interval for the difference of population mean sepal lengths between setosa and versicolor.

t-Based Confidence Interval

95% Confidence interval
Assumed unequal variances for variables

Variable	Midpoint	Std Error	DF	Lower CL	Upper CL	Margin of Error
Sepal.Length (setosa) - Sepal.Length (versicolor)	-0.93	0.088395	86.538	-1.1057	-0.75429	0.17571

We are 95% confident in our estimate that the population mean sepal length for *setosa* irises is between 0.754 cm and 1.106 cm LESS than the population mean sepal length for *versicolor* irises.

Question #5 Based on our confidence interval, which species has longer petals? Or can you not tell? Explain your answer.

Since the entire confidence interval is negative, this suggests that the *versicolor* irises have longer petals, on average.

Question #6 Based on our confidence interval, can we reject H$_0$: $\mu_1 - \mu_2 = 0$ in favor of the two-sided NHST alternative H$_a$: $\mu_1 - \mu_2 \neq 0$ at the 5% significance level? Why or why not?

We can reject the null hypothesis H$_0$: $\mu_1 - \mu_2 = 0$ in favor of H$_a$: $\mu_1 - \mu_2 \neq 0$ at the 5% significance level because 0 is <u>not</u> in the 95% confidence interval.

Finally, we will do a *t* confidence interval for the population mean of paired differences (matched pairs confidence interval). We used the straight_jeans2 dataset for this one and investigated the mean of the paired differences in maximum back pocket width (*maxWidthBackMens* and *maxWidthBackWomens*).

In *Mean Inference* ✉ *One and Two Population* dialog, we'll select *maxWidthBackMens* as *Variable 1* and *maxWidthBackWomens* as *Variable 1*. Then, we'll check the *Paired Data* box and, in the *Population 1-2* tab, check the box for a *t-statistic* based *Confidence Interval*. Just to get the practice, we'll change our *Confidence Level* to 0.99 (99%).

Question #7 Paste the output table labeled *t-Based Confidence Interval*. Write a sentence interpreting the 99% confidence interval for the population mean of the paired differences in maximum back pocket width.

t-Based Confidence Interval

99% Confidence interval
Paired differences are used

Variable	Midpoint	Std Error	DF	Lower CL	Upper CL	Margin of Error
[maxWidthBackMens - maxWidthBackWomens]	0.36	0.27828	19	-0.43615	1.1562	0.79615

We are 95% confident that on average, men's jeans back pockets are between 0.436 cm narrower and 1.156 cm wider compared to women's jeans of the same brand.

Question #8 Based on our confidence interval, after matching by brand, whose jeans have wider back pockets – men or women? Or can you not tell? Explain your answer.

Because the confidence interval includes 0, we cannot tell whose jeans have wider back pockets. There are some values of μ_d in the interval for which men's pockets are wider and some values of μ_d for which women's pockets are wider.

Question #9 Based on our confidence interval, can we reject H_0: $\mu_d = 0$ in favor of the two-sided NHST alternative H_a: $\mu_d \neq 0$ at the 1% significance level? Why or why not?

We cannot reject the null hypothesis H_0: $\mu_d = 0$ in favor of H_a: $\mu_d \neq 0$ at the 1% significance level because 0 <u>is</u> in the 99% confidence interval.

Question #10 Based on your answers in this lab, what information can you get from a <u>confidence interval</u> that you cannot get from a hypothesis test?

The confidence interval gives us an entire range of values that are reasonable estimates for the parameter we are interested in. A hypothesis test will only tell us whether we can accept a proposed value for the parameter/whether we must reject that proposed value.