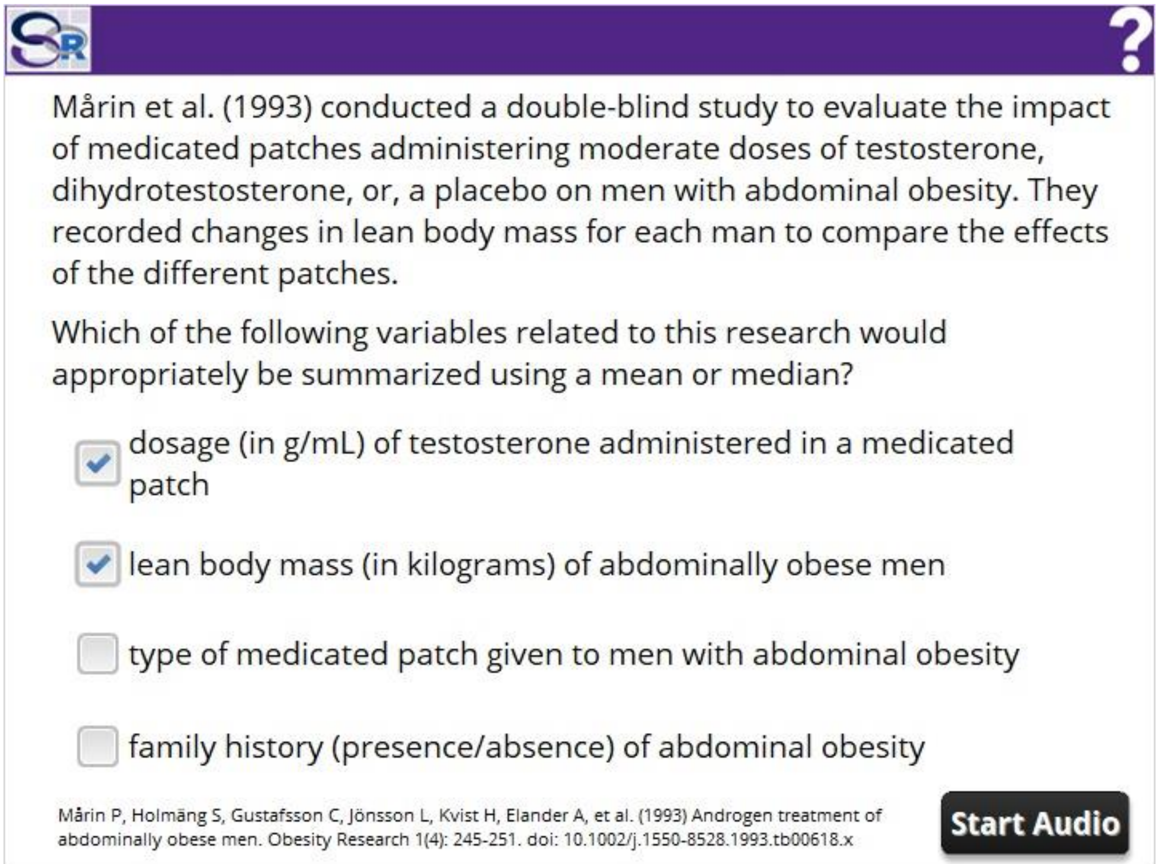
Lab 5 - t procedures for mu

1. Are you ready?

1.1 Warm-up Question 1

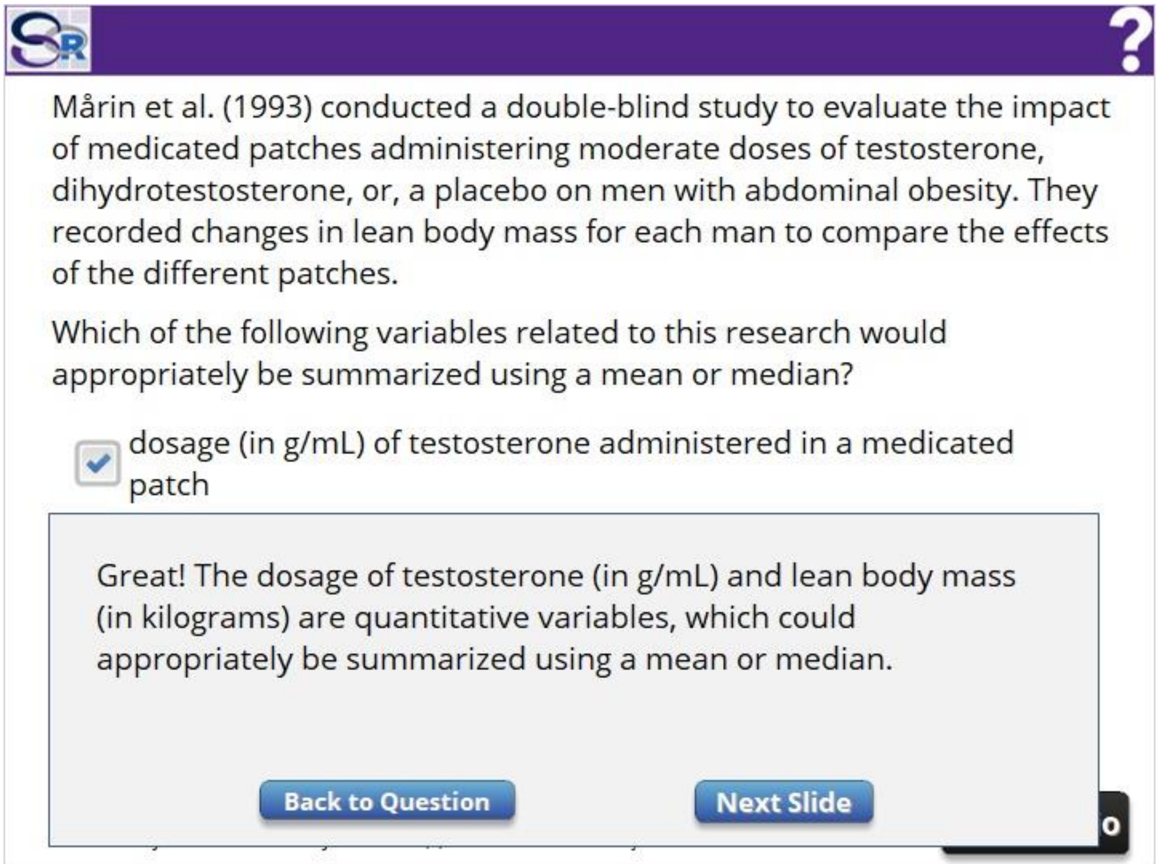


Notes:

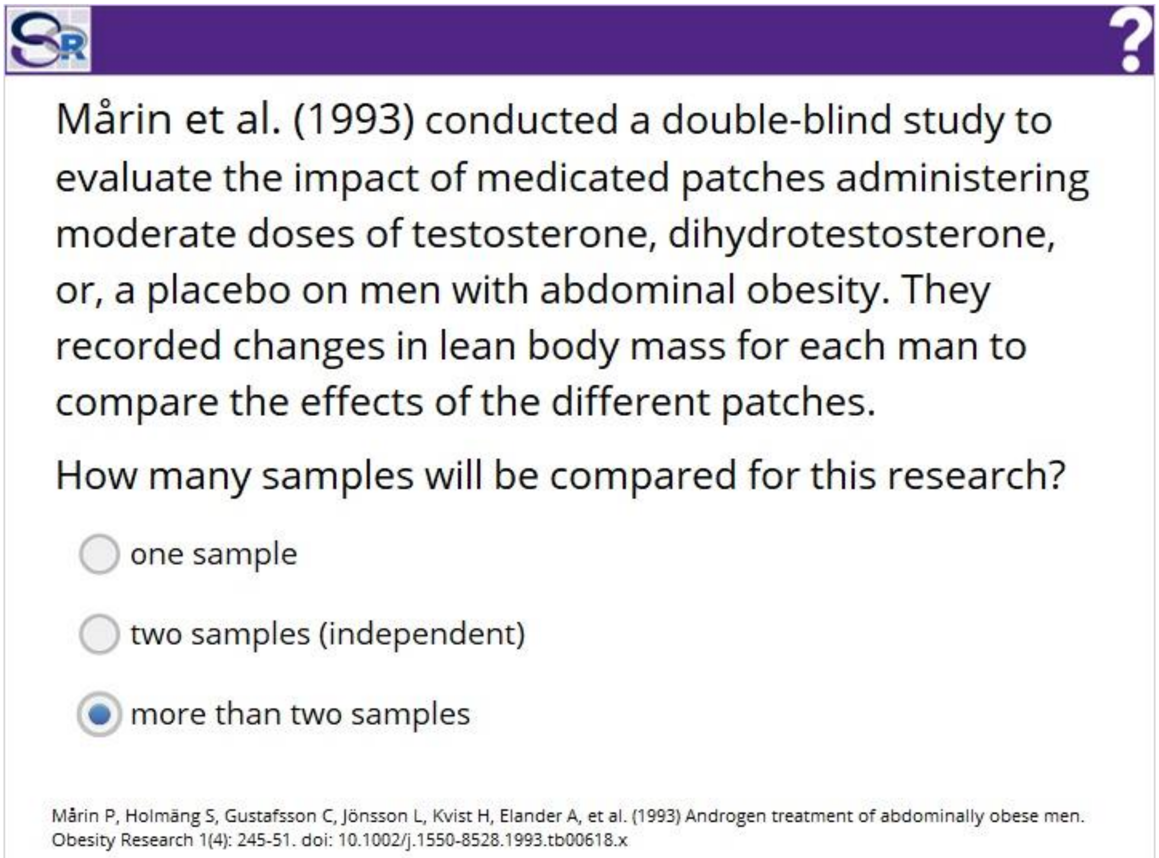
Let's think about a couple of questions to get warmed up for the methods in this module. Question 1 - Mårin and colleagues conducted a double-blind study to evaluate the impact of medicated patches administering moderate doses of testosterone, dihydrotestosterone, or, a placebo on men with abdominal obesity. They recorded changes in lean body mass for each man to compare the effects of the different patches.
Which of the following variables related to this research would appropriately be summarized using a mean or median?

More than one option may be appropriate, so select all that apply.

Correct (Slide Layer)



1.2 Warm-up Question 2



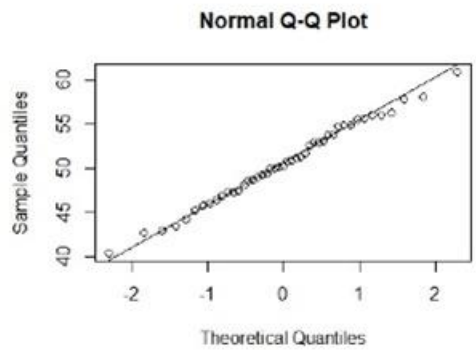
Notes:

Question 2: Let's consider the same study by Marin and co-authors. How many samples will be compared for this research?

Correct	Choice	Feedback
	one sample	One sample is incorrect. The researchers are comparing the effects of three different treatments: testosterone, dihydrotestosterone, and a placebo. Consequently, this study compared more than two samples.
	two samples (independent)	Two samples (independent) is incorrect. The researchers are comparing the effects of three different treatments: testosterone, dihydrotestosterone, and a placebo. Consequently, this study compared more than two samples.
X	more than two samples	Great! The researchers are comparing the change in lean body mass for mean in three different treatments: testosterone, dihydrotestosterone, and a placebo. Note that this module deals with one sample situations, so this study would use methods described elsewhere to conduct the comparison.

1.3 Warm-up Question 3

The plot provided is a Normal quantile plot (QQ plot), like one that might be generated for sample data of lean body mass (in kilograms) of abdominally obese men.



What information can be obtained from a Normal quantile plot?

- whether a variable fits a binomial distribution
- whether a variable fits a Normal distribution
- whether a variable fits a Poisson distribution

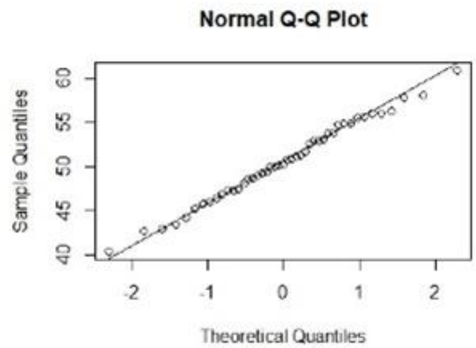
Notes:

Now let's consider how we might investigate and describe a distribution during an exploration of our data.

Question 3: The plot provided here is a Normal quantile plot (also known as a QQ plot), like one that might be generated for sample data of lean body mass (in kilograms) of abdominally obese men.

What information can be obtained from a Normal quantile plot?

The plot provided is a Normal quantile plot (QQ plot), like one that might be generated for sample data of lean body mass (in kilograms) of abdominally obese men.



What information can be obtained from a Normal quantile plot?

Great! Normal QQ plots are useful for visually evaluating whether a variable fits a Normal distribution. This module will describe how to produce QQ plots in R to evaluate the Normality assumption of one-sample t-tests.

For more information about Normal distributions and other probability models, see Chapter 3 Diez, Barr, and Cetinkaya-Rundel (2015) OpenIntro Statistics. 3rd Edition. Available at https://www.openintro.org/stat/textbook.php

Back to Question

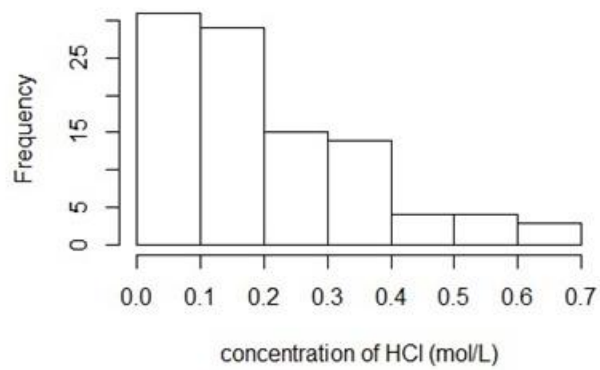
Next Slide

1.4 Warm-up Question 4

This histogram summarizes the concentration of hydrochloric acid (mol/L) in the stomach acid of a sample of 100 patients.

What is the best description of the shape of this distribution?

- ○ approximately symmetric
- ○ left (negative) skewed
- ◉ right (positive) skewed



Notes:

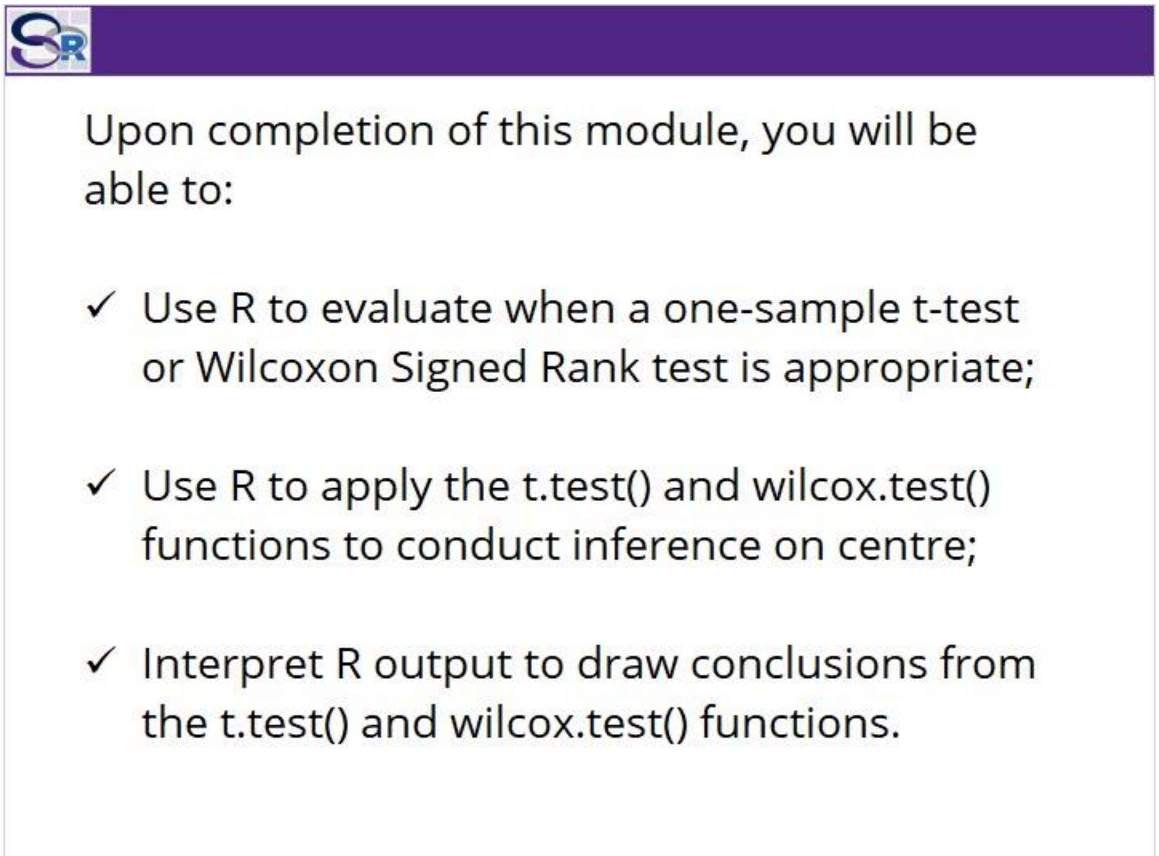
Question 4: This histogram summarizes the concentration of hydrochloric acid (in mol/L) in the stomach acid of a sample of 100 patients.

What is the best description of the shape of this distribution?

Correct	Choice	Feedback
	approximately symmetric	Unfortunately, approximately symmetric is not a correct description of the shape of this distribution. If we ‘cut’ a symmetric distribution in half, the left and right sides of the distribution would be (approximately) mirror images of each other. For a refresher on descriptors of shape, see http://stats.onlinelearning.utoronto.ca/modules/summarizing-data/
	left (negative) skewed	Unfortunately, left (negative) skewed is not a correct description of the shape of this distribution. A left skewed distribution would have a long ’tail’ on the left side of the distribution. For a refresher on descriptors of shape, see http://stats.onlinelearning.utoronto.ca/modules/summarizing-data/
X	right (positive) skewed	Great! This is an example of a right (positive) skewed distribution.

2. Module goals

2.1 Outcomes



Notes:

In this module, you will learn how to conduct inference procedures on the location of centre for one population. In particular, you'll learn about using R to evaluate when one sample t procedures or Wilcoxon Signed Rank test procedures are appropriate, and how to conduct and interpret these two procedures in R.

NOTE: Ignore reference to the Wilcoxon Signed Rank test—this content is not testable and has (more or less) been removed from this lesson.

3. One-sample t-test

3.1 Research Setting: Objective

Image by Arne von Bril is licensed under CC BY 2.0

B. Hare, M. Brown, C. Williamson, and M. Tomasello. 2002. The domestication of social cognition in dogs. Science 298(5598): 1634-1636. doi: 10.1126/science.1072702

"We test among the three hypotheses for the origin of dogs' ability to use human social cues...in an object choice task"

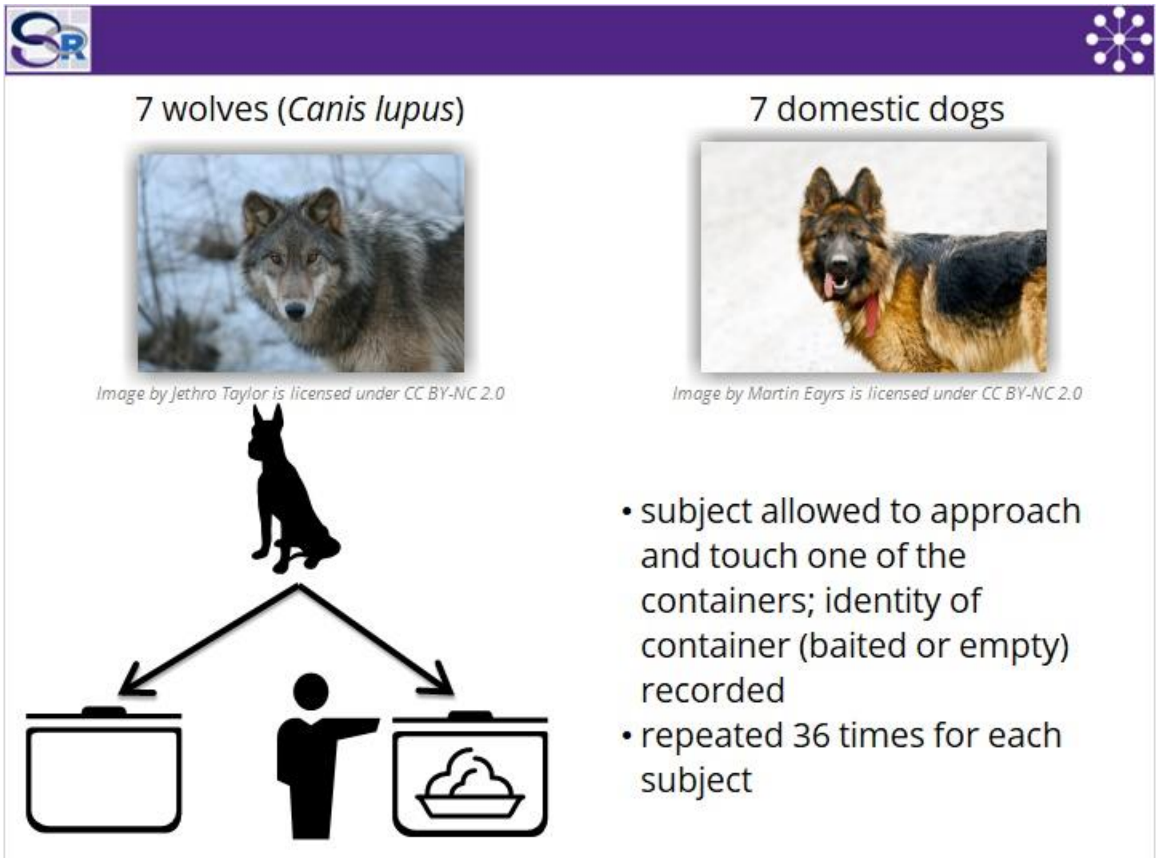
The domestication hypothesis predicts that dogs should be more skillful than wolves at using human social skills to find an object because of selection pressure for social cognition during domestication.

Notes:

Domestic dogs show an interesting ability to use many different visual cues given by humans when presented with an 'object choice task'; that is, they can follow a human's gaze (or other behavior like pointing) to select a particular object when presented with the choice of two objects. This skill is noteworthy because even human's close relatives-chimpanzees and other primates-are not very good at interpreting such social cues.
This begs the question of how such a skill has arisen in dogs. There are several hypotheses explaining this skill. Hare and colleagues set out to--in their own words--"test among the three hypotheses for the origin of dogs' ability to use human social cues".
One of the three hypotheses of interest to us for this module, is referred to as the domestication hypothesis, which suggests that dogs have evolved the ability to interpret human social cues as a result of selection pressure during domestication. This particular hypothesis, therefore, predicts that dogs

should be more skillful than wolves-their close relative-at using human social skills because of the absence of domestication process and therefore, the selection pressure, in the evolutionary history of wolves.

3.2 Research Setting: Methods

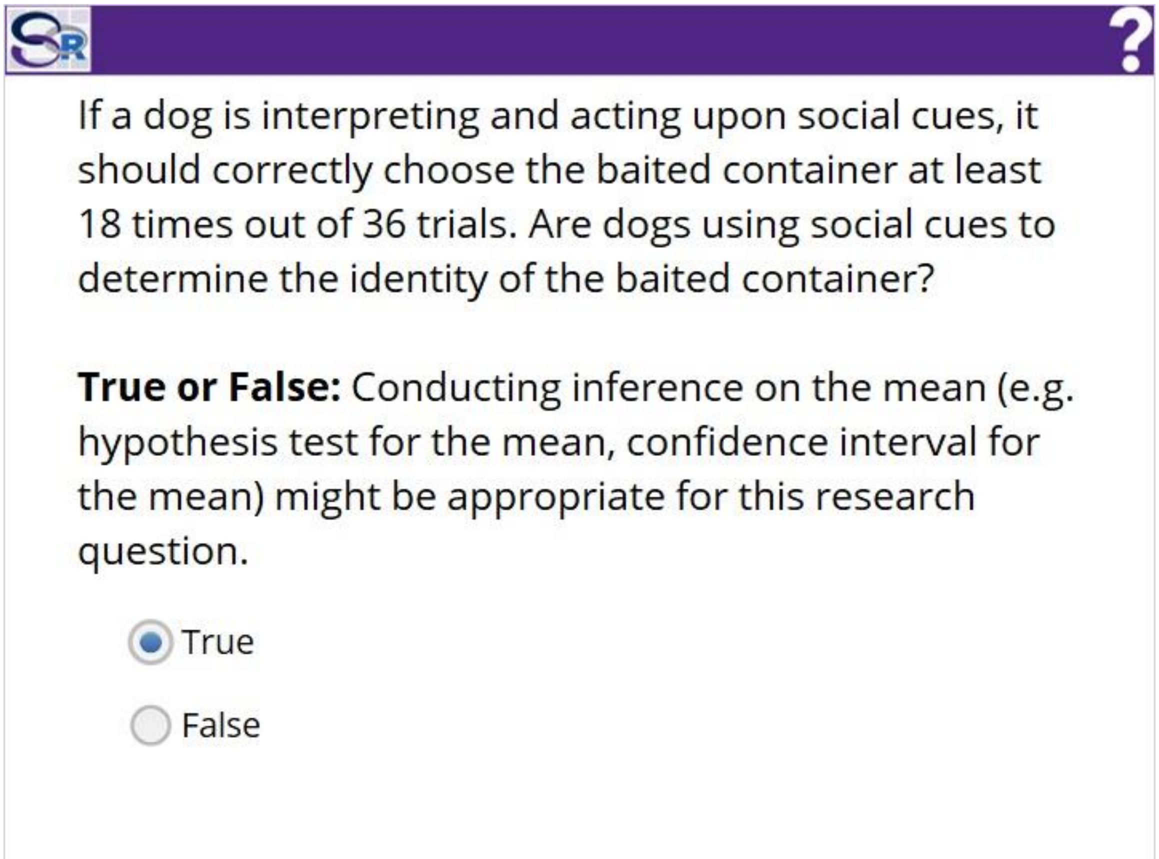


Notes:

To investigate the domestication hypothesis, the researchers set up an experiment following a typical object choice trial. In their experiment, a researcher presented the test subject-either a wolf or a dog-with two containers. One of the containers had been previously baited with food, and the other with no food. At the start of the trial, the test subject was unaware of which container had been baited. The researcher then used a combination of social cues that included gazing at, pointing at, and tapping the baited container to communicate its identity to the test subject. The subject was then allowed to approach and touch one of the containers. The researcher recorded whether the test subject correctly identified the baited container or not.

This procedure was repeated for a sample of 7 wolves and of 7 domestic dogs, giving each individual 36 trials at identifying the baited container.

3.3 Thinking ahead - appropriate inference



Notes:

In the plan stage of the scientific inquiry cycle, it's a good idea to think ahead toward the analyses that might be appropriate for the expected data. One of the first requirements in the object choice experiment by Hare and colleagues was to identify whether the subjects were responding to social cues at all. If a dog is interpreting and acting upon social cues, it should correctly choose the baited container at least 18 times out of 36 trials. A suitable research question might then be, are dogs using social cues to determine the identity of the baited container?

True or False: Conducting inference on the mean number of correctly chosen containers (that is, a hypothesis test for the mean, or a confidence interval for the mean) might be an appropriate procedure for this research.

If a dog is interpreting and acting upon social cues, it should correctly choose the baited container at least 18 times out of 36 trials. Are dogs using social cues to determine the identity of the baited container?

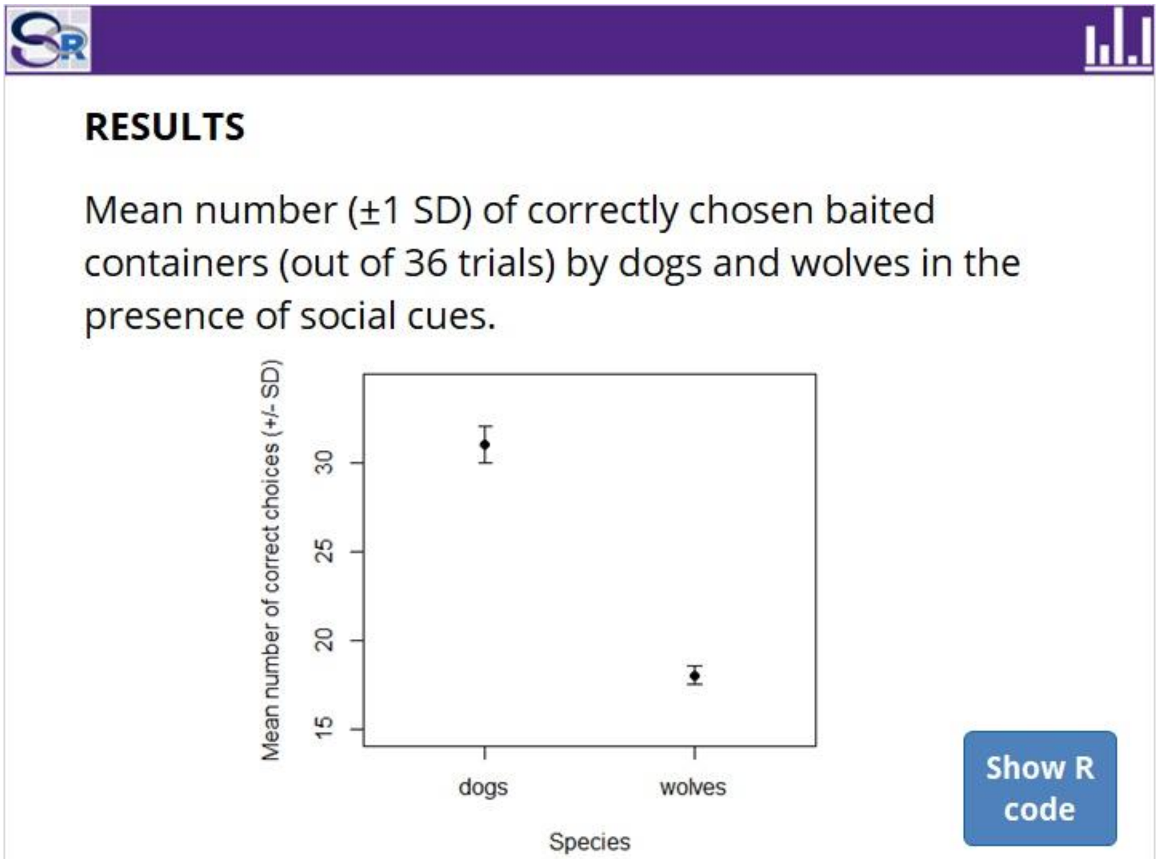
True or False: Conducting inference on the mean (e.g. hypothesis test for the mean, confidence interval for the mean) might be appropriate for this research

That's right! The data collected in this experiment are quantitative (number of times the baited container was correctly chosen). These data can be summarized by a mean number of times the baited container was correctly chosen by dogs. We could then test the null hypothesis that the mean number is 18, or estimate the mean number using a confidence interval.

Back to Question

Next Slide

3.4 Data



Notes:

Hare and colleagues collected data on the number of correctly chosen baited containers that the dogs and wolves identified in the object choice test. In their 2002 publication, the researchers summarized the count data using means for each species, wolves or dogs. Estimates of these statistics based on their publication are summarized in the mean plot illustrated below, which was created in R, with error bars representing plus and minus one standard deviation from the mean number of correctly chosen baited containers (out of 36 trials) for each species. Take a moment to inspect these data; do they suggest that dogs, and separately, wolves are interpreting human social cues to identify the baited containers? Before we move on, a word of caution is necessary when investigating data with a mean plot; there is no universal standard for what the error bars on a mean plot represent. In this mean plot, the error bars represent plus and minus one standard deviation from the mean. Some authors choose to use plus/minus 2 (or more) standard deviations, and others still might use standard error rather than standard deviation.Always look carefully for a description of what the error bars represent in a mean plot so you can accurately interpret the spread of the data.

means plot R script (Slide Layer)

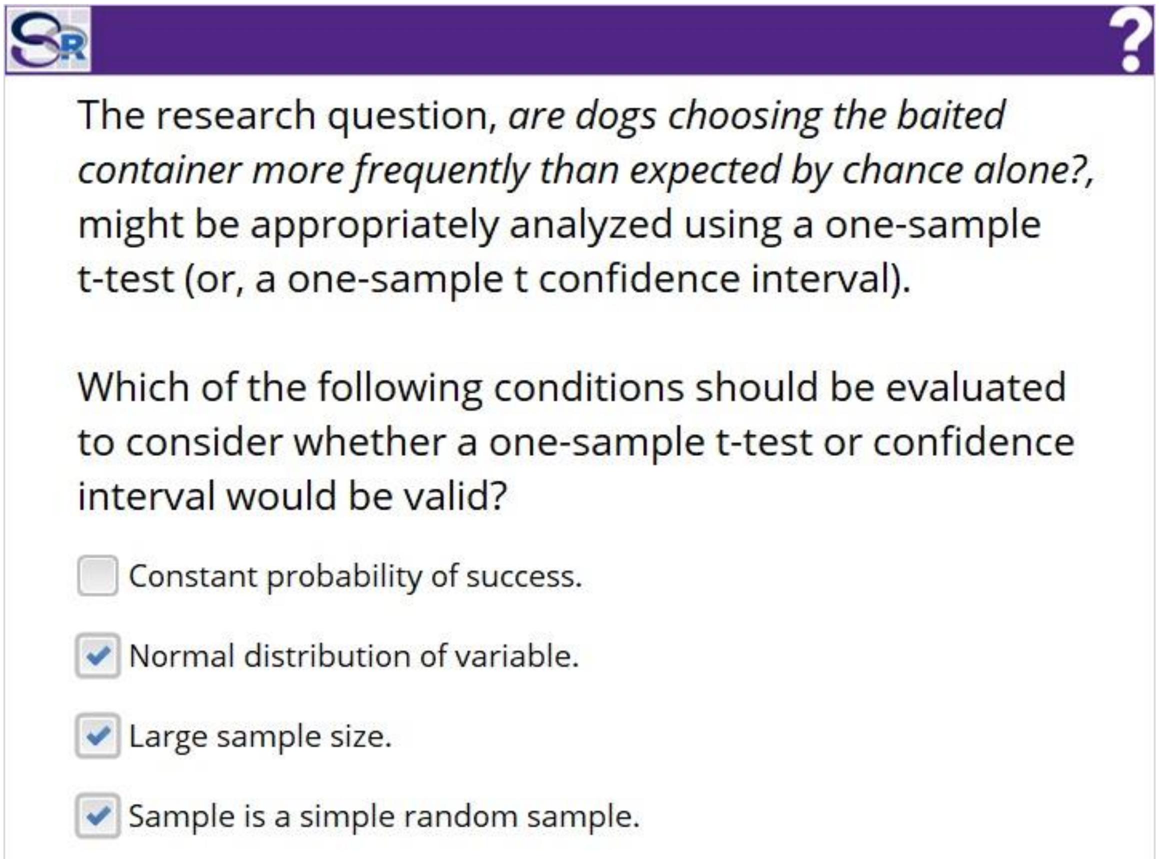
The data used to generate the means plot were estimated from Figure 2 in Hare et al. (2002). The means were estimated and the standard deviations back-calculated from the standard error of the means.

```
install.packages("psych")
library(psych)
species<-data.frame(mean=c(31,18),se=c(1.06, 0.53))
rownames(species)<-c("dogs","wolves")
error.bars(stats=species,pch=19, xlab="Species",ylab="Mean number of correct choices (+/-SD)",eyes=FALSE,main=NULL,cex.main=0.9)
```

B. Hare, M. Brown, C. Williamson, and M. Tomasello. 2002. The domestication of social cognition in dogs. Science 298(5598): 1634-1636. doi: 10.1126/science.1072702

Back to graph

3.5 Thinking ahead: Conditions for t-test



Notes:

The research question, *are dogs choosing the baited container more frequently than expected by chance alone?,* might be appropriately analyzed using a one-sample t-test.

Which of the following conditions should be evaluated to consider whether a one-sample t-test would be valid? More than one option may be appropriate, so select all that would apply.

The research question, *are dogs choosing the baited container more frequently than expected by chance alone?*, might be appropriately analyzed using a one-sample t-test (or, a one-sample t confidence interval).

Which of the following conditions should be evaluated to consider whether a one-sample t-test or confidence interval would be valid?

Great! To conduct a valid one-sample t-test,
1. the sample should be a simple random sample
2. the sample data are drawn from a Normally distributed population distribution. However, the t-procedures are robust to deviations from Normality when sample size is large.

Back to Question

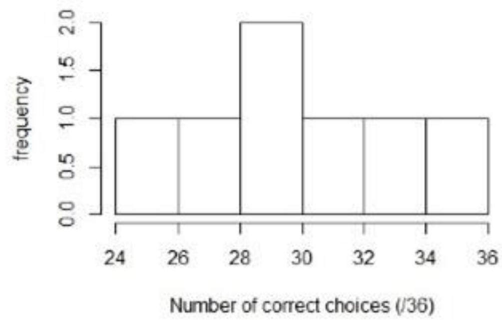
Next Slide

3.6 Analysis: Conditions for t-test

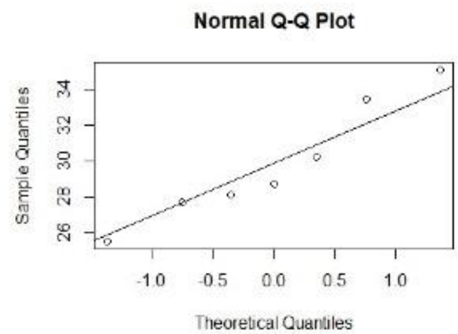
EVALUATING NORMALITY CONDITION

For a vector of data values, x

Histogram to evaluate shape:
hist(x)



Sample size
length(x)



Quantile plot (and fit line) to evaluate fit of Normal model:
qqnorm(x)
qqline(x)

Show R code

Notes:

It is always important to check the validity of the conditions for an inference procedure. The one sample t procedures for inference on a mean requires that the sample be drawn from a Normally distributed population distribution.

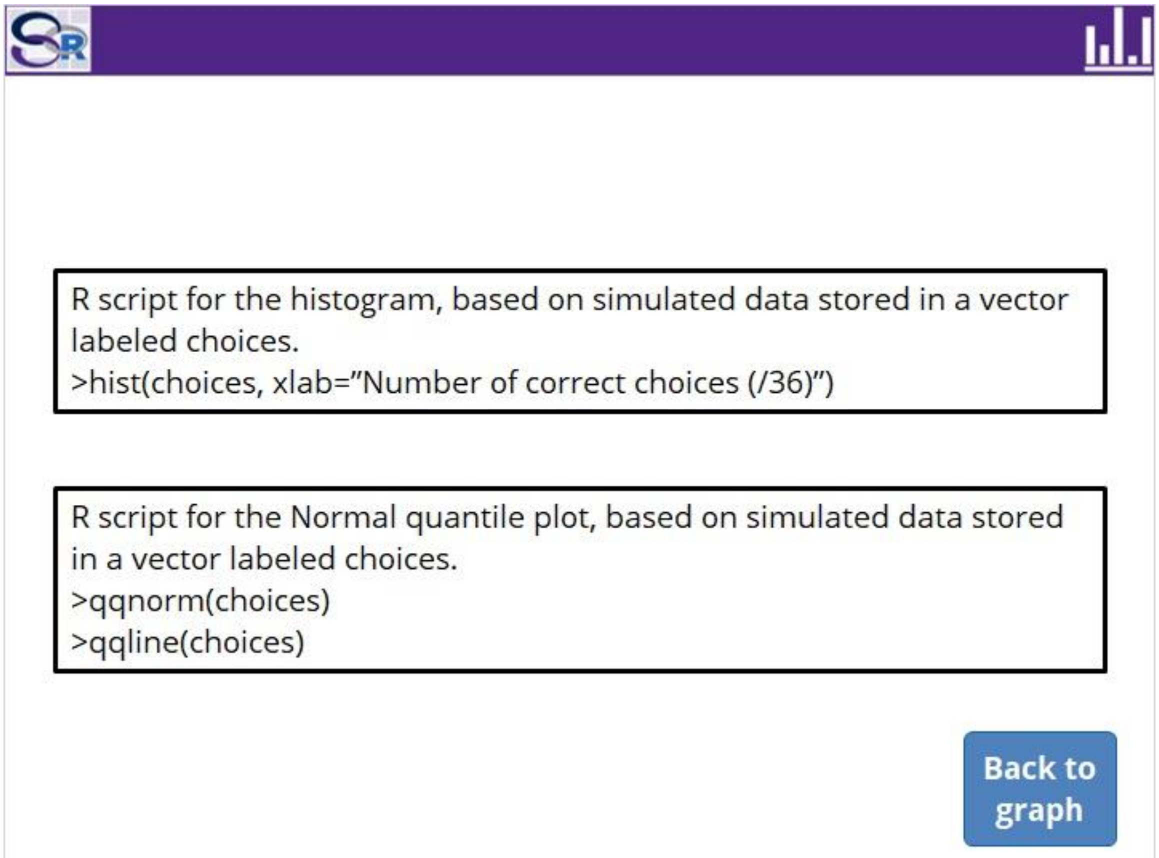
There are various methods to evaluate the validity of this condition. In R, we could generate a histogram for our sample data using the hist() function to assess the general shape of the distribution of data in our sample as an indicator of the distribution of the population.

Or, we could assess the fit of a Normal probability model with a Normal quantile plot (also known as a QQ plot) for our sample data using the qqnorm() function; again, sample data that are fit by a Normal model may suggest that the population distribution is approximately Normal.

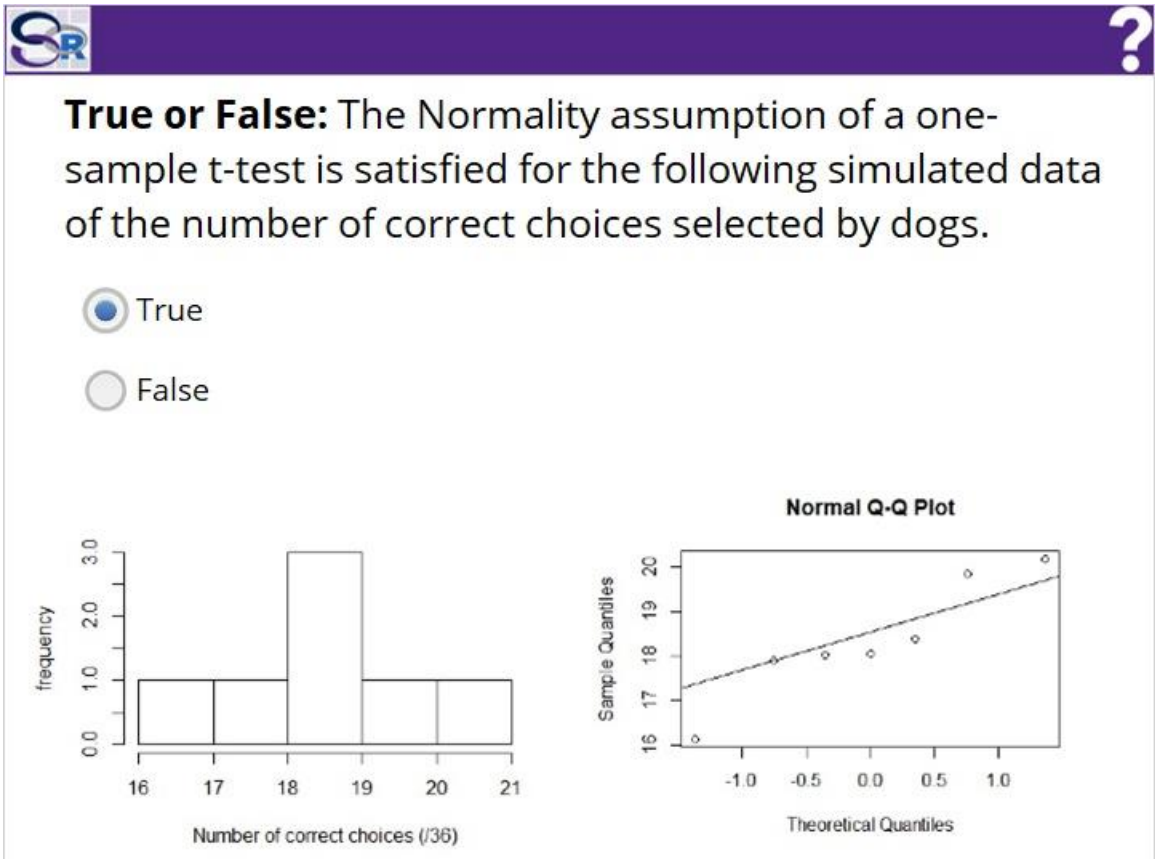
If we wish to identify the sample size we are working with to consider whether we can rely on the robustness of the t procedures, we could use the length() function.

Once we have confirmed that the Normality condition is reasonably met, we can move on to using R to conduct the t test.

R script for histogram (Slide Layer)



3.7 Application: conditions for t-test



Notes:

Consider the histogram and Normal Quantile plot given here.

True or False: The Normality assumption of a one sample t-test is satisfied for the following simulated data on the number of correctly chosen baited containers selected by dogs.

True or False: The Normality assumption of a one-sample t-test is satisfied for the following simulated data of the number of correct choices selected by dogs.

- (•) True
- ( ) False

Great! The Normal quantile (QQ plot) shows that the sample data reasonably fit the theoretical distribution for a Normal distribution; this histogram also confirms that the sample data are reasonably unimodal, bell-shaped, and symmetric.

Back to Question

Next Slide

3.8 t.test() syntax

One-sample t test and confidence interval for μ in R

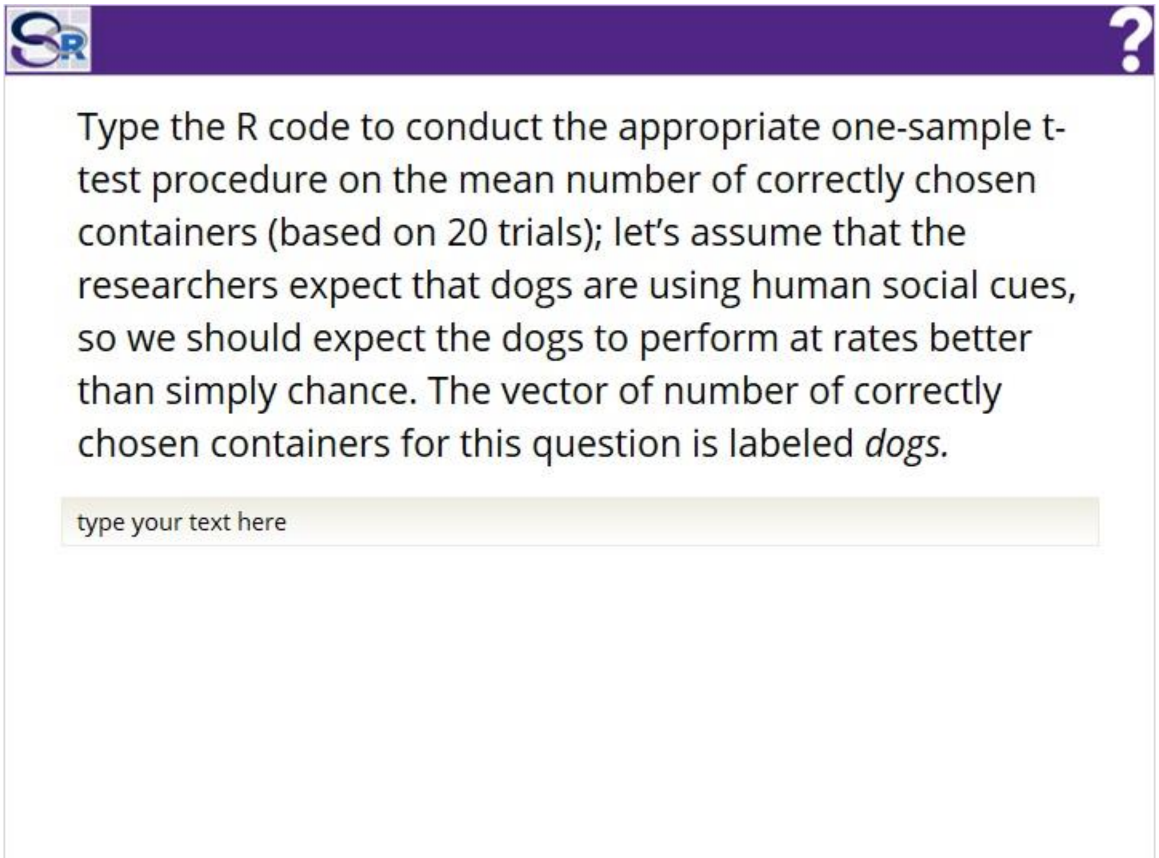
t.test()

Argument	Description	Value	Default
x	vector of data (quantitative variable)	real numbers	***required***
alternative	direction of test (i.e., sign in alternative hypothesis)	"two.sided" "less" "greater"	"two.sided"
mu	hypothesized value of population mean	real number	0
conf.level	level of confidence interval	number between 0 and 1	0.95

Notes:

The t.test function runs the t procedures in R; this includes both the one sample AND two sample t procedures, depending on the arguments included. The minimum argument that the t.test function requires is x, the vector of data; including only x, will run the one-sample procedure--which is the topic of this module.
You can customize the procedure by varying the other arguments. For instance, if you wanted to set a specific value of the mean to test as a null hypothesis, then use the "mu=" argument; the default is set to a mean of zero. You can conduct a one-tailed test and corresponding confidence interval using the "alternative=" argument and can construct a confidence interval for the mean using a confidence level other than 95% (which is the default) by using the "conf.level=" argument.
There are other possible arguments for the t.test() procedures; however, they are not relevant to one sample t procedures, and therefore, are not summarized in this table.

3.9 Application: what's the R code?



Notes:

Here’s an opportunity for you to try typing the R code that will run the appropriate procedure to address the preliminary research question of whether dogs are using human social cues to select the baited containers.

Type the R code to conduct the appropriate one-sample t-test procedure on the mean number of correctly chosen containers (based on 20 trials); let’s assume that the researchers expect that dogs are using human social cues, so we should expect the dogs to perform at rates better than simply chance. The vector of number of correctly chosen containers for this question is labeled *dogs.*

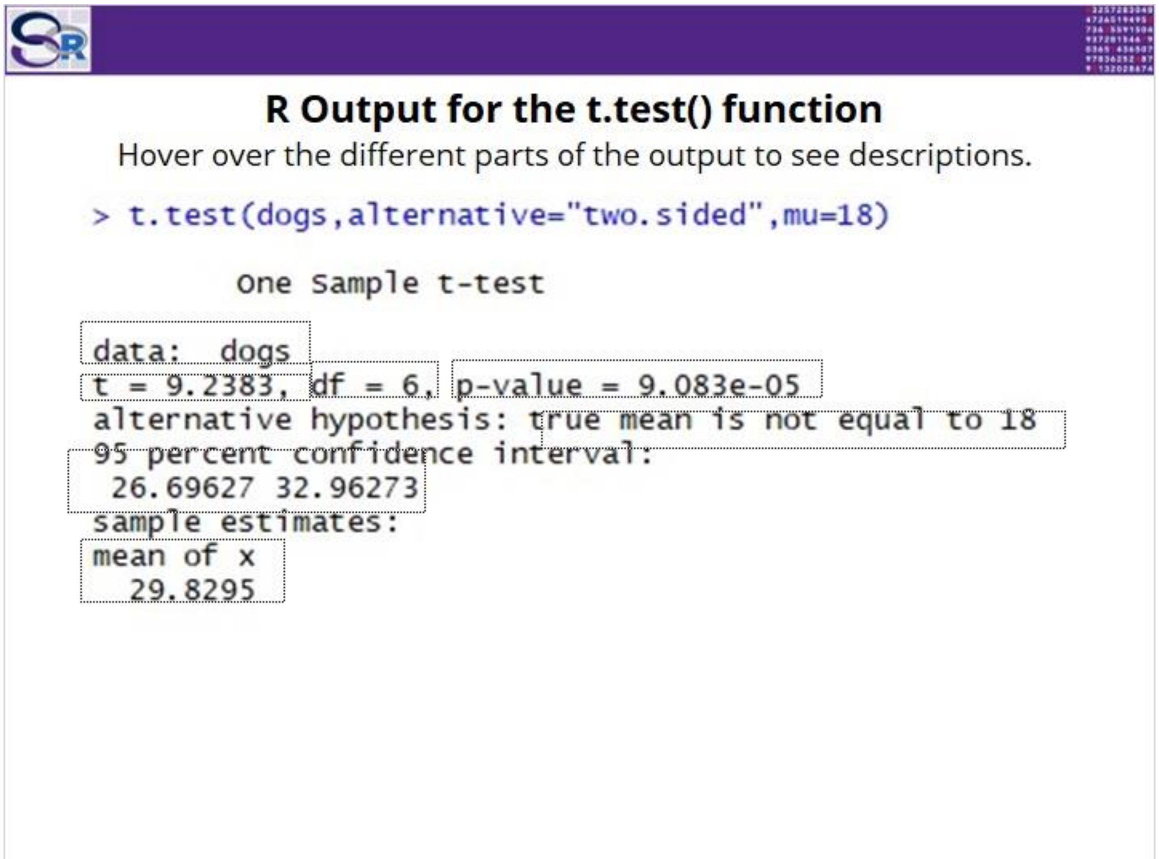
Hmmm, your answer didn’t quite match ours, but that doesn’t mean it is incorrect. There are lots of ways to obtain the same outcome. Compare your answer to the following code (or try it in R to see if it runs!):
t.test(dogs,”greater”,10)
Remember, if a dog was selecting based on chance, we would expect 10 correctly chosen containers, on average (i.e. 50% of 20). If we are expecting dogs to perform better than chance, then we need the right-tailed alternative hypothesis, mu>10.
The default value of conf.level (95%) is ok here.

If you wanted to explicitly include all the arguments, you could type their values in order in the brackets:
t.test(dogs,10,”greater”,0.95) or you could set each argument within the brackets (order need not be the same here):
t.test(dogs,mu=10,alternative=”greater”,conf.level=0.95)

Remember, R code is case-sensitive. Extra spaces within the brackets are often ok, but stray punctuation marks are not.

Try Again   Next Slide

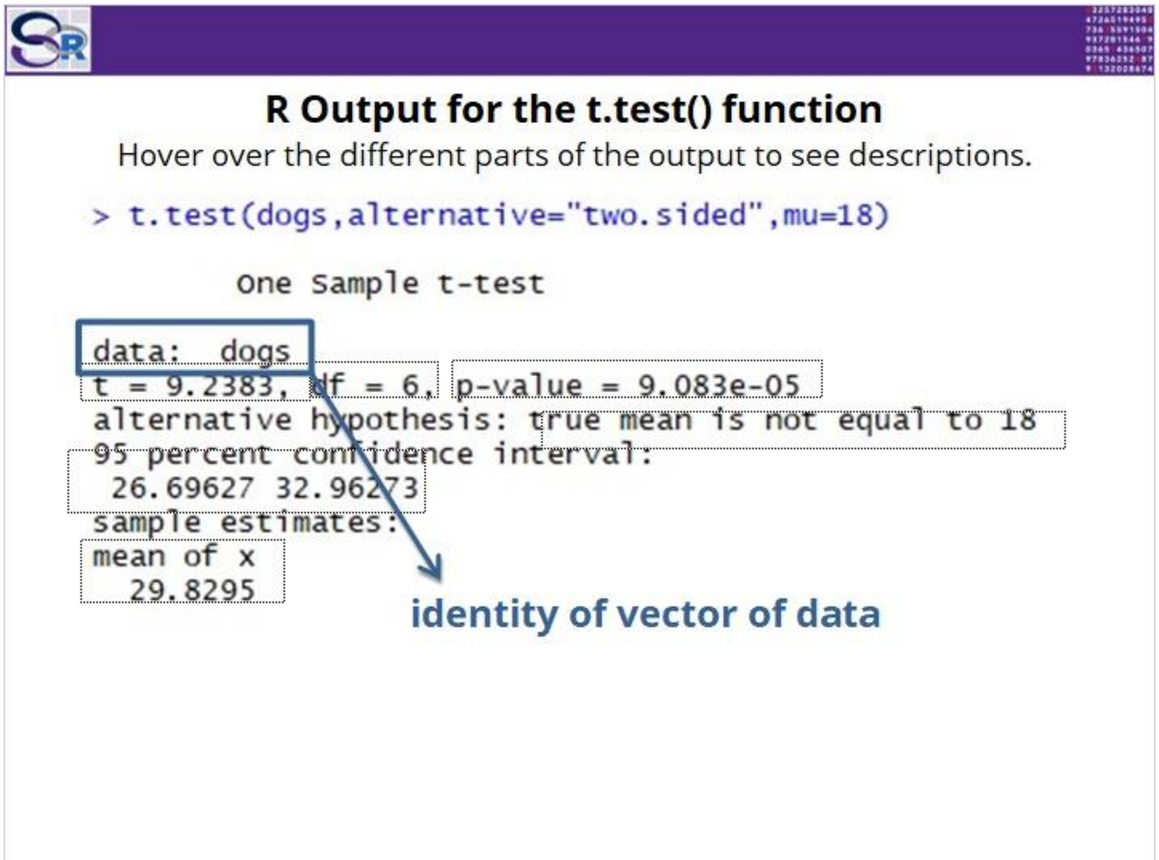
3.10 Analysis: R output for t.test()



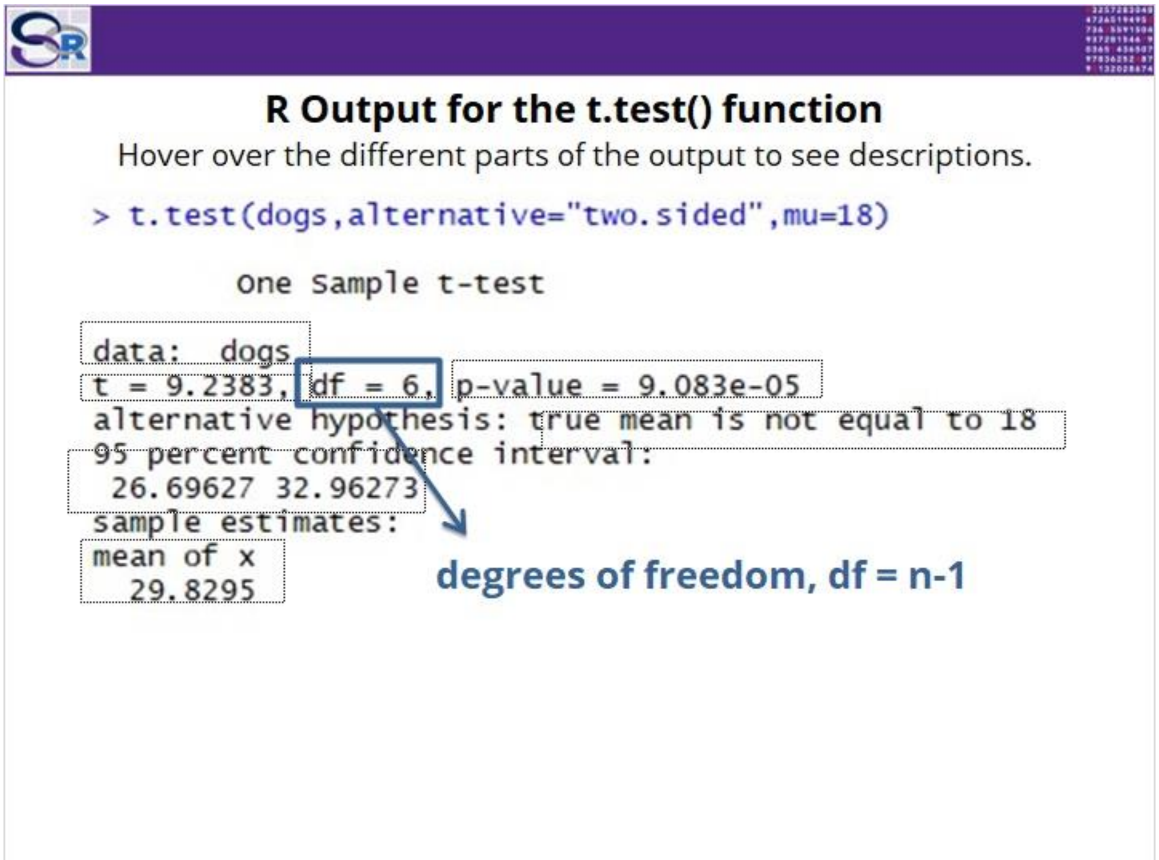
Notes:

R produces this output for the t.test function when used for a one sample procedure. The data, test statistic, degrees of freedom, a confidence interval and hypothesis test for the mean are all summarized here. Hover over the different pieces of the output with your mouse for more information.

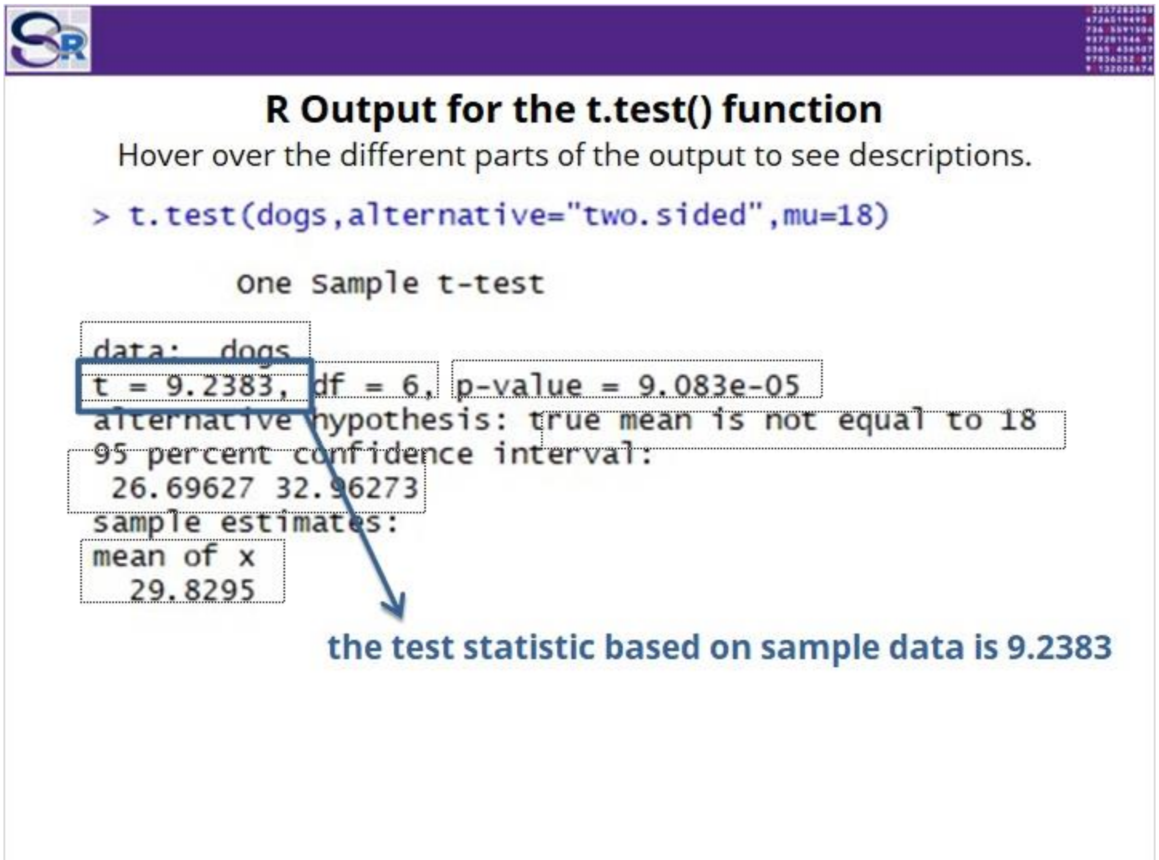
data vector (Slide Layer)



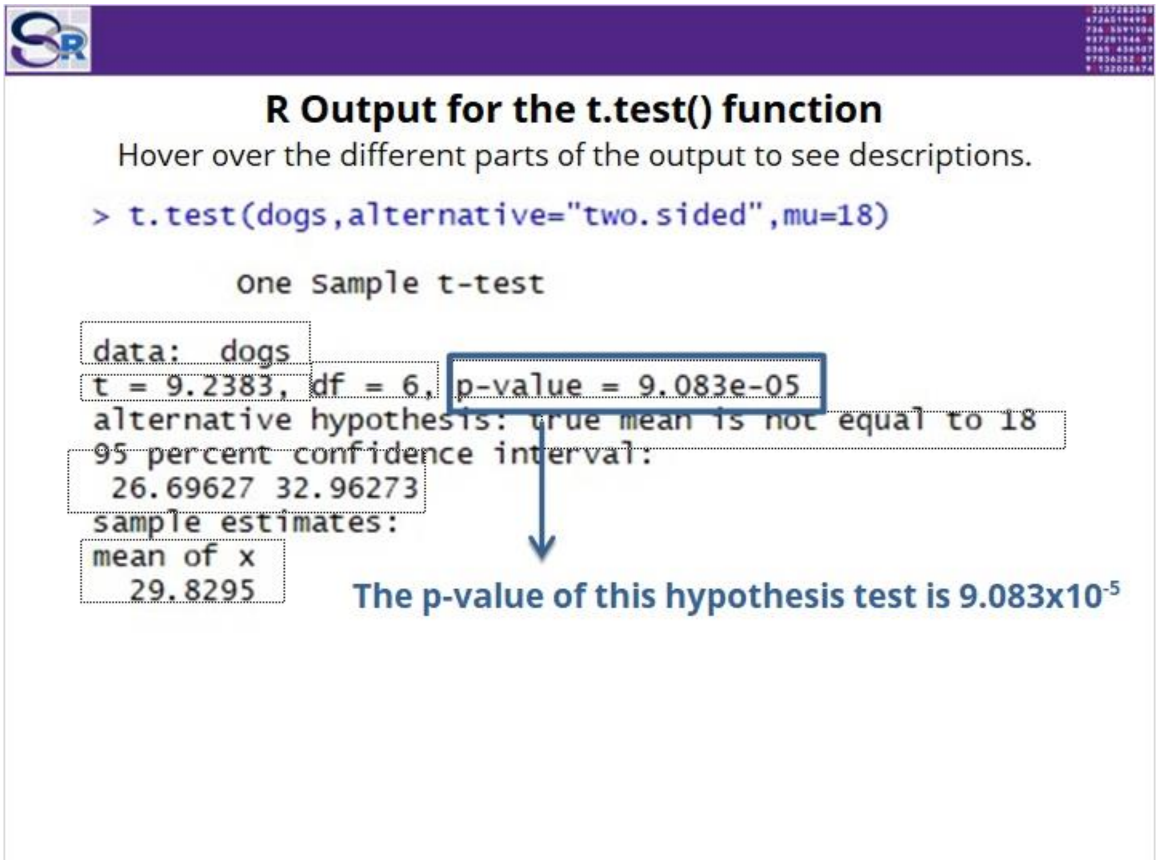
df (Slide Layer)



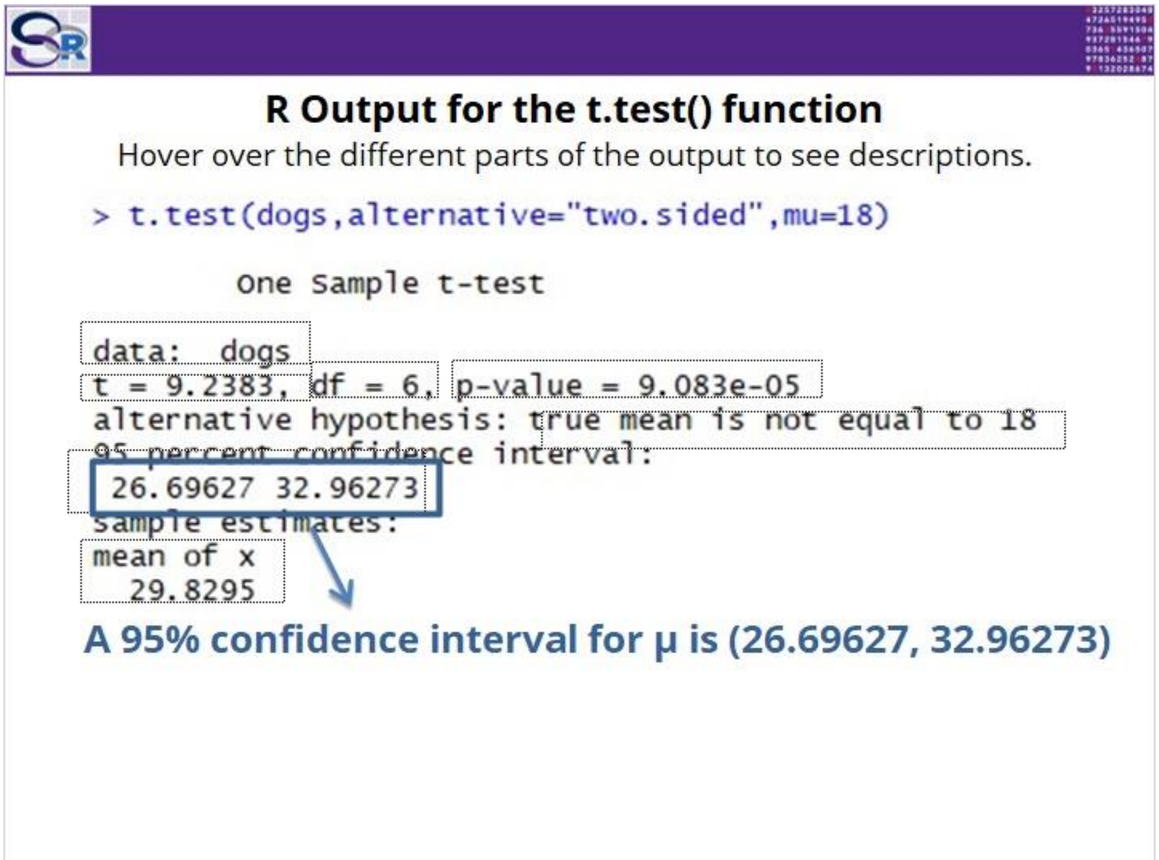
test statistic (Slide Layer)



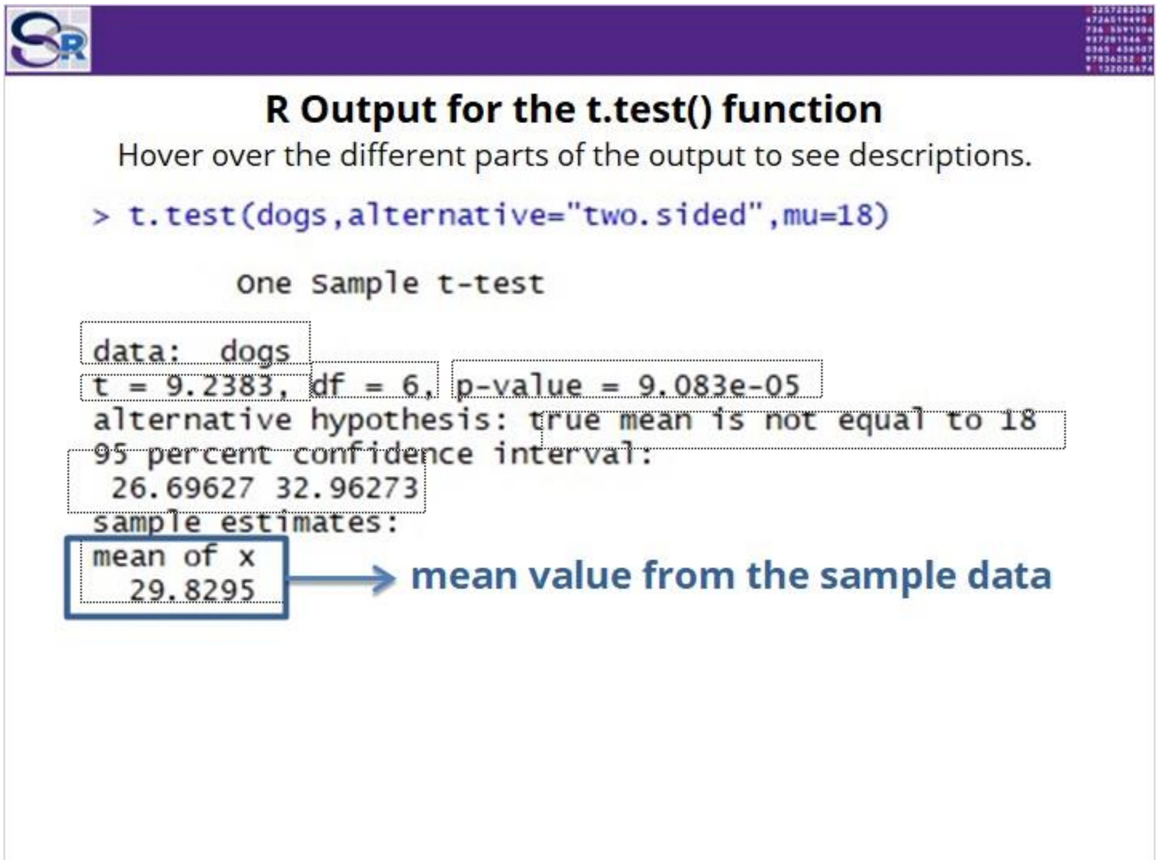
P-value (Slide Layer)



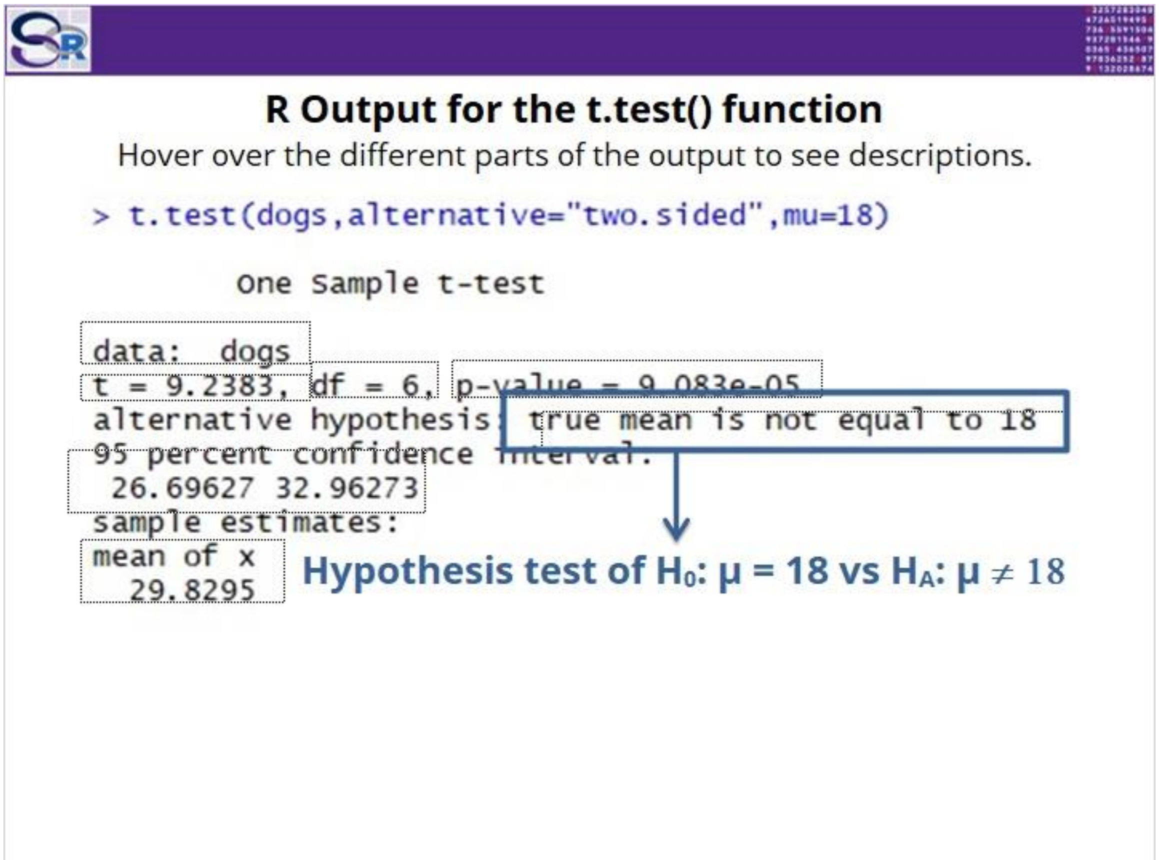
Confidence interval (Slide Layer)



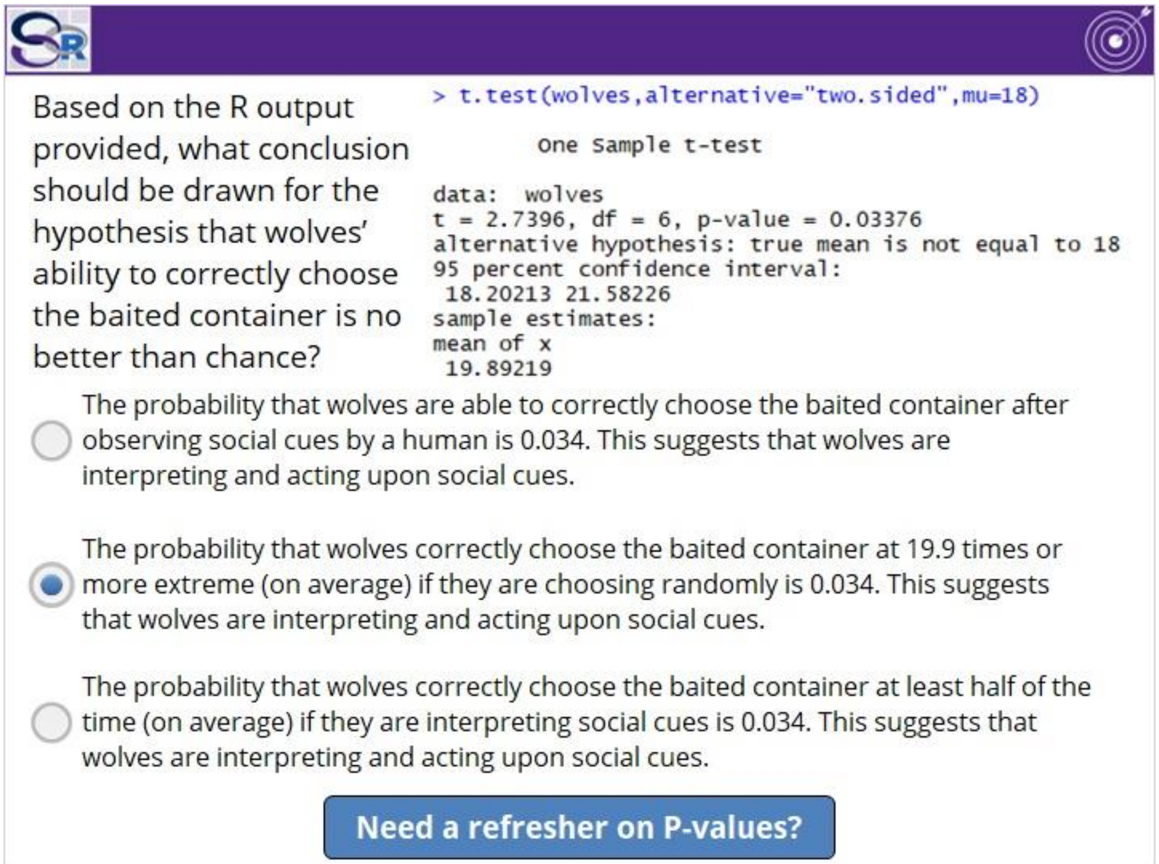
Sample mean (Slide Layer)



hypothesis (Slide Layer)



3.11 Conclusion



Notes:

Now that we can get R to produce output for an appropriate inference procedure, we need to think about what these statistical findings mean so we can address a research question; that is, we are at the Conclusion stage of the scientific inquiry cycle.

As part of their research project, Hare and colleagues tested wolves' ability to identify the baited container. The sample output provided here is based on simulating sample data having a sample mean and standard deviation estimated from the graphs provided in the Hare and colleagues publication, as they relate to putting wolves through the object choice test that provided social cues for the identity of the baited container. Based on the R output provided, what conclusion should be drawn for the hypothesis that wolves' ability to correctly choose the baited container is no better than chance?

Based on the R output provided, what conclusion should be drawn for the hypothesis that wolves' ability to correctly choose the baited container is no better than chance?

```
> t.test(wolves,alternative="two.sided",mu=18)

        One Sample t-test

data:  wolves
t = 2.7396, df = 6, p-value = 0.03376
alternative hypothesis: true mean is not equal to 18
95 percent confidence interval:
 18.20213 21.58226
sample estimates:
mean of x
 19.89219
```

- The probability that wolves are able to correctly choose the baited container after observing social cues by a human is 0.034. This suggests that wolves are interpreting and acting upon social cues.

That's right! A P-value of 0.03 suggests that it is unlikely for wolves to choose the correct container 19.9 times (or more extreme) if they are simply choosing the containers randomly. This might be considered good evidence that wolves are acting upon social cues.

Back to Question | Next Slide

P-value refresher (Slide Layer)

These types of hypotheses tests assume the null hypothesis is true and assess evidence against that assumption.

A **small p-value** means the results look unusual or are unlikely under that assumption, so we reject the null hypothesis (H_0) in favour of the alternative hypothesis (H_A).

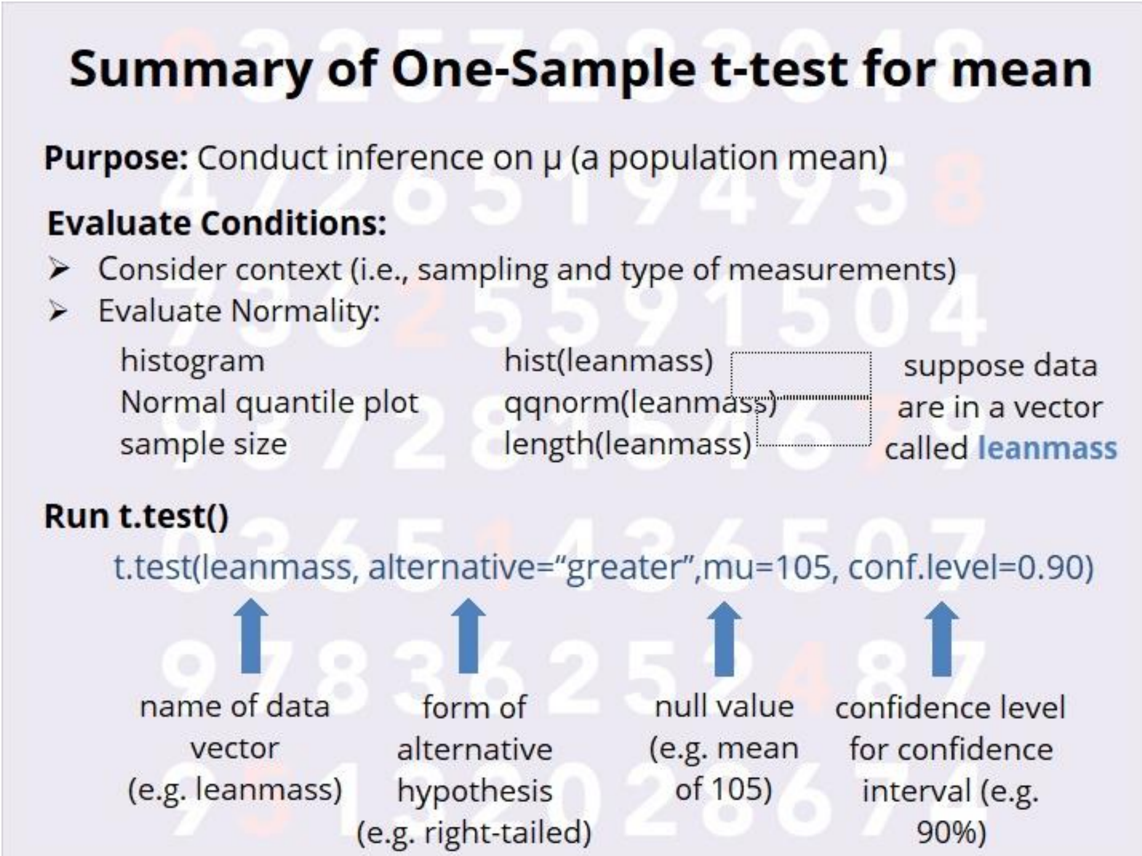
More resources:

• See http://stats.onlinelearning.utoronto.ca/modules/statistical-tests-i/

• Open Intro – ***Section 4.3 (Hypothesis Testing)*** in:
Diez, Barr, and Cetinkaya-Rundel (2015) *OpenIntro Statistics*. 3rd Edition. Available at https://www.openintro.org/stat/textbook.php

Back

3.12 Summary: one sample t procedures

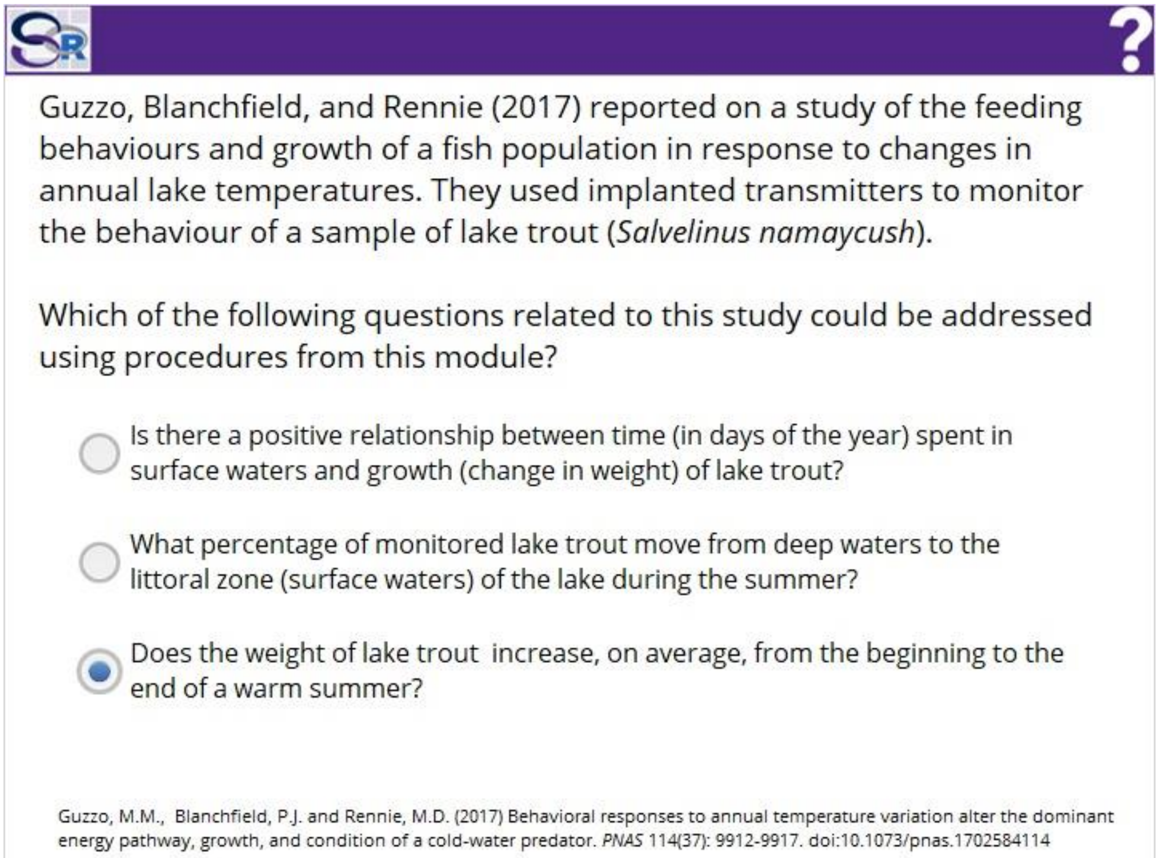


Notes:

Here is a summary of the one sample t procedures for a population mean (including when they may be useful and what conditions are necessary), along with instructions on how to run them in R.

4. Check your understanding

4.1 Review Question 1



Notes:

Let’s work through a series of review questions to recap the procedures discussed in this module. An important consideration in the scientific inquiry cycle is recognizing when certain statistical methods are called for and when they are not.

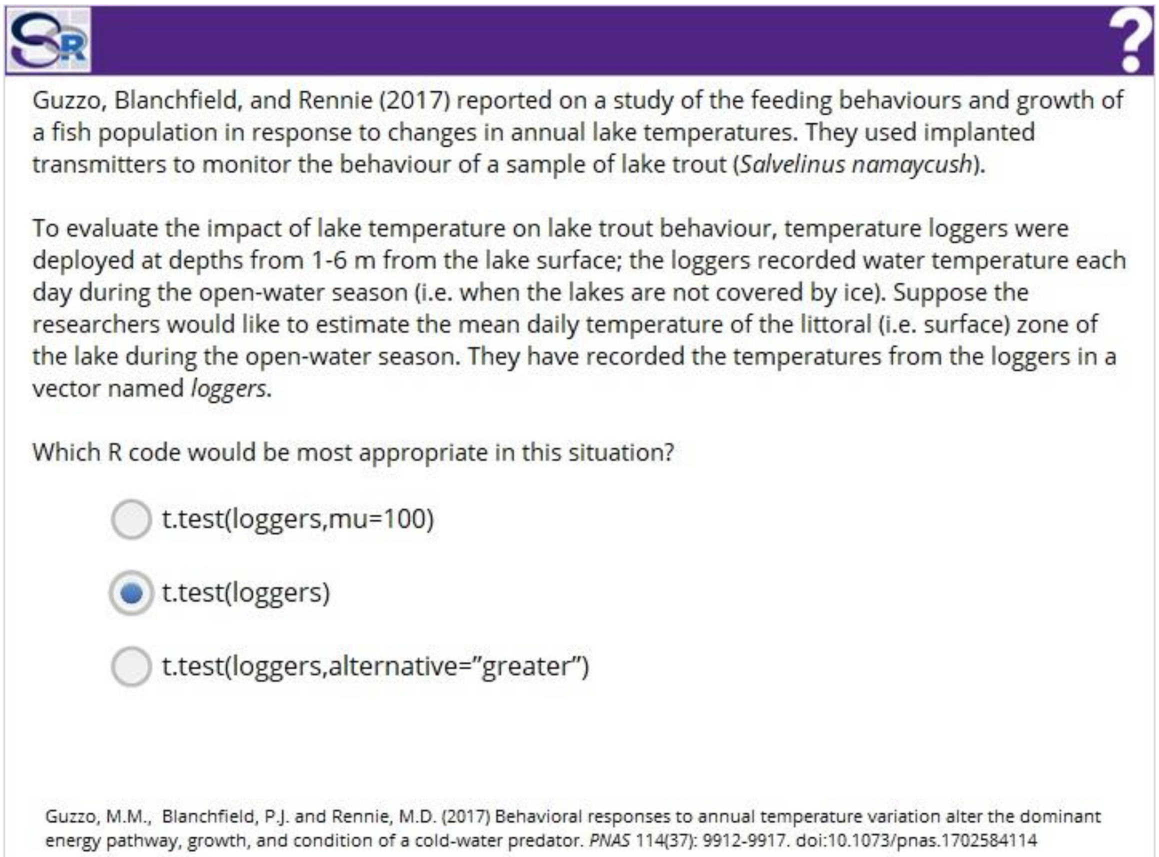
Consider some research reported in 2017 by Guzzo, Blanchfield and Rennie, who conducted a study of the feeding behaviours and growth of a fish population in response to changes in annual lake temperatures. They used implanted transmitters to monitor the behaviour of a sample of lake trout.

Which of the following questions related to this study could be addressed using procedures from this module? That is, either one sample t procedures, or one sample Wilcoxon Signed rank test procedures for inference on a population centre.

Correct	Choice	Feedback
	Is there a positive relationship between time (in days of the year) spent in surface waters and growth (change in weight) of lake trout?	That is incorrect. While the variables involved are quantitative (weight and time), this question is looking at the relationship between two quantitative variables. Some form of regression (e.g. linear regression) analysis would be appropriate to determine the nature of a relationship (if one exists). This module focused on inference on one measure of centre (e.g. mean or median).
	What percentage of monitored lake trout move from deep waters to the littoral zone (surface waters) of the lake during the summer?	That is incorrect. The data collected for this research question would be qualitative (categorical): whether the trout are in the littoral zone or not. This research question would be best addressed using a confidence interval to estimate the proportion (i.e. percentage) rather than inference on a location of centre (i.e. mean or median).
X	Does the weight of lake trout increase, on average, from the beginning to the end of a warm summer?	Yes! The data collected are quantitative (lake trout weight). If the data are collected as matched pairs (i.e. repeated measures on the same fish, to calculate a difference in weight from beginning to end of the summer), it

	would be appropriate to use one sample inference on the mean or median of the differences.

4.2 Review Question 2



Notes:

Let's continue to think about the research by Guzzo, Blanchfield, and Rennie.

To evaluate the impact of lake temperature on lake trout behaviour, temperature loggers were deployed at depths from 1 to 6 metres from the lake surface; the loggers recorded water temperature each day during the open-water season (that is, when the lakes are not covered by ice). Suppose the researchers would like to estimate the mean daily temperature of the littoral (i.e. surface) zone of the lake during the open-water season. They have recorded the temperatures from the loggers in a vector named *loggers*.

Which R code would be most appropriate in this situation?

Correct	Choice	Feedback
	t.test(loggers,mu=100)	Not quite. Although the one sample t procedures might be appropriate here because temperature is a quantitative variable (especially if the condition of a Normal distribution is met), setting the null value to mu=100 will produce a confidence interval that estimates the mean temperature plus 100 (i.e. μ + 100).
X	t.test(loggers)	Yes! The one sample t inference procedures (t test and confidence interval) might be appropriate here because temperature is a quantitative variable. Leaving the alternative hypothesis as the default (i.e., two-sided) will produce a confidence interval estimate for μ, the true mean daily temperature of the littoral zone during the open-water season. Note that, if we fail to meet the necessary conditions for the t procedures, we might consider the Wilcoxon Signed Rank procedures, with R code wilcox.test(loggers, conf.int=TRUE). Remember, by default, the wilcox.text() will not produce a confidence interval;

	we have to specify the conf.int=TRUE argument to obtain one.
t.test(loggers,alternative=”greater”)	Not quite. Although one sample t procedures might be appropriate here because temperature is a quantitative variable, setting the alternative to “greater” (or right tailed) will produce a one-sided confidence interval rather than a two-sided confidence interval.