Part I

Syllabus and Software

Part II

Summaries and displays, and one-, two-, and many-way tests of means

Chapter 2

Estimation in One-Sample Problems

Contents

Learning objectives

After completing this topic, you should be able to:

select graphical displays that meaningfully communicate properties of a sample.

assess the assumptions of the one-sample t-test visually.

decide whether the mean of a population is different from a hypothesized value.

recommend action based on a hypothesis test.

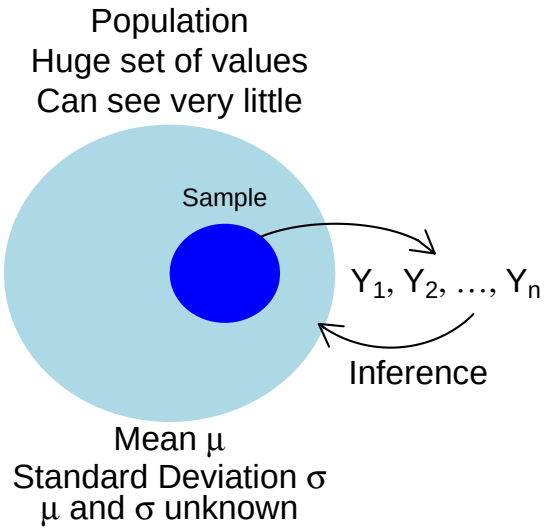
Achieving these goals contributes to mastery in these course learning outcomes:

1. organize knowledge.
5. define parameters of interest and hypotheses in words and notation.
6. summarize data visually, numerically, and descriptively.
8. use statistical software.
12. make evidence-based decisions.

2.1 Inference for a population mean

Suppose that you have identified a population of interest where individuals are measured on a single quantitative characteristic, say, weight, height or IQ. You select a random or representative sample from the population with the goal of estimating the (unknown) **population mean** value, identified by μ. You cannot see much of the population, but you would like to know what is typical in the population (μ). The only information you can see is that in the sample.

This is a standard problem in statistical inference, and the first inferential problem that we will tackle. For notational convenience, identify the measurements on the sample as $Y_1, Y_2, \ldots, Y_n$, where n is the sample size. Given the data, our best guess, or estimate, of μ is the sample mean: $\bar{Y} = \frac{\sum_i Y_i}{n} = \frac{Y_1+Y_2+\cdots+Y_n}{n}$.



There are two main methods that are used for inferences on μ: **confidence intervals** (CI) and **hypothesis tests**. The standard CI and test procedures are based on the sample mean and the sample standard deviation, denoted by s.

■ CLICKERQs — Inference for a population mean, 2 ■

2.1.1 Standard error, LLN, and CLT

The **standard error (SE)** is the standard deviation of the **sampling distribution** of a statistic.

The **sampling distribution** of a statistic is the distribution of that statistic, considered as a random variable, when derived from a random sample of size n.

The **standard error of the mean (SEM)** is the standard deviation of the sample-mean's estimate of a population mean. (It can also be viewed as the standard deviation of the error in the sample mean relative to the true mean, since the sample mean is an unbiased estimator.) SEM is usually estimated by the sample estimate of the population standard deviation (sample standard

deviation) divided by the square root of the sample size (assuming statistical independence of the values in the sample):

$$SE_{\bar{Y}} = s/\sqrt{n}$$

where s is the sample standard deviation (i.e., the sample-based estimate of the standard deviation of the population), and n is the size (number of observations) of the sample.

In probability theory, the **law of large numbers (LLN)** is a theorem that describes the result of performing the same experiment a large number of times. According to the law, the average of the results obtained from a large number of trials (the sample mean, $\bar{Y}$) should be close to the expected value (the population mean, μ), and will tend to become closer as more trials are performed.

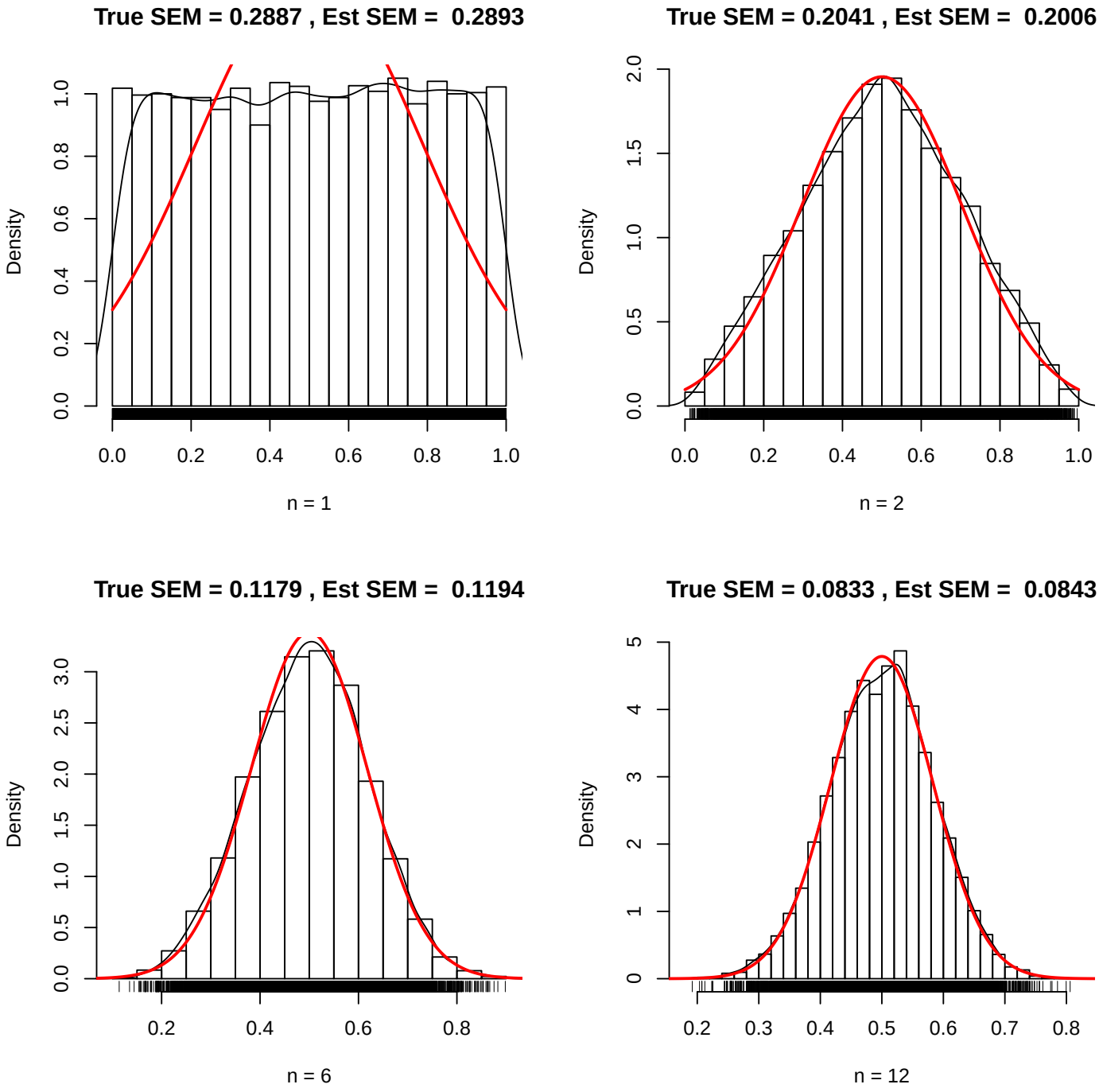
In probability theory, the **central limit theorem (CLT)** states that, given certain conditions, the mean of a sufficiently large number of independent random variables, each with finite mean and variance, will be approximately normally distributed[1].

As a joint illustration of these concepts, consider drawing random variables following a Uniform(0,1) distribution, that is, any value in the interval $[0, 1]$ is equally likely. By definition, the mean of this distribution is $\mu = 1/2$ and the variance is $\sigma^2 = 1/12$ (so the standard deviation is $\sigma = \sqrt{1/12} = 0.289$). Therefore, if we draw a sample of size n, then the standard error of the mean will be $\sigma/\sqrt{n}$, and as n gets larger the distribution of the mean will increasingly follow a normal distribution. We illustrate this by drawing $N = 10000$ samples of size n and plot those N means, computing the expected and observed SEM and how well the histogram of sampled means follows a normal distribution, Notice, indeed, that even with samples as small as 2 and 6 that the properties of the SEM and the distribution are as predicted.

[1]The central limit theorem has a number of variants. In its common form, the random variables must be identically distributed. In variants, convergence of the mean to the normal distribution also occurs for non-identical distributions, given that they comply with certain conditions.

```r
#### Illustration of Central Limit Theorem, Uniform distribution
# demo.clt.unif(N, n)
#   draws N samples of size n from Uniform(0,1)
#   and plots the N means with a normal distribution overlay
demo.clt.unif <- function(N, n) {
  # draw sample in a matrix with N columns and n rows
  sam <- matrix(runif(N*n, 0, 1), ncol=N);
  # calculate the mean of each column
  sam.mean <- colMeans(sam)
  # the sd of the mean is the SEM
  sam.se <- sd(sam.mean)
  # calculate the true SEM given the sample size n
  true.se <- sqrt((1/12)/n)
  # draw a histogram of the means
  hist(sam.mean, freq = FALSE, breaks = 25
       , main = paste("True SEM =", round(true.se, 4)
                      , ", Est SEM = ", round( sam.se, 4))
       , xlab = paste("n =", n))
  # overlay a density curve for the sample means
  points(density(sam.mean), type = "l")
  # overlay a normal distribution, bold and red
  x <- seq(0, 1, length = 1000)
  points(x, dnorm(x, mean = 0.5, sd = true.se), type = "l", lwd = 2, col = "red")
  # place a rug of points under the plot
  rug(sam.mean)
}

par(mfrow=c(2,2));
demo.clt.unif(10000,  1);
demo.clt.unif(10000,  2);
demo.clt.unif(10000,  6);
demo.clt.unif(10000, 12);
```

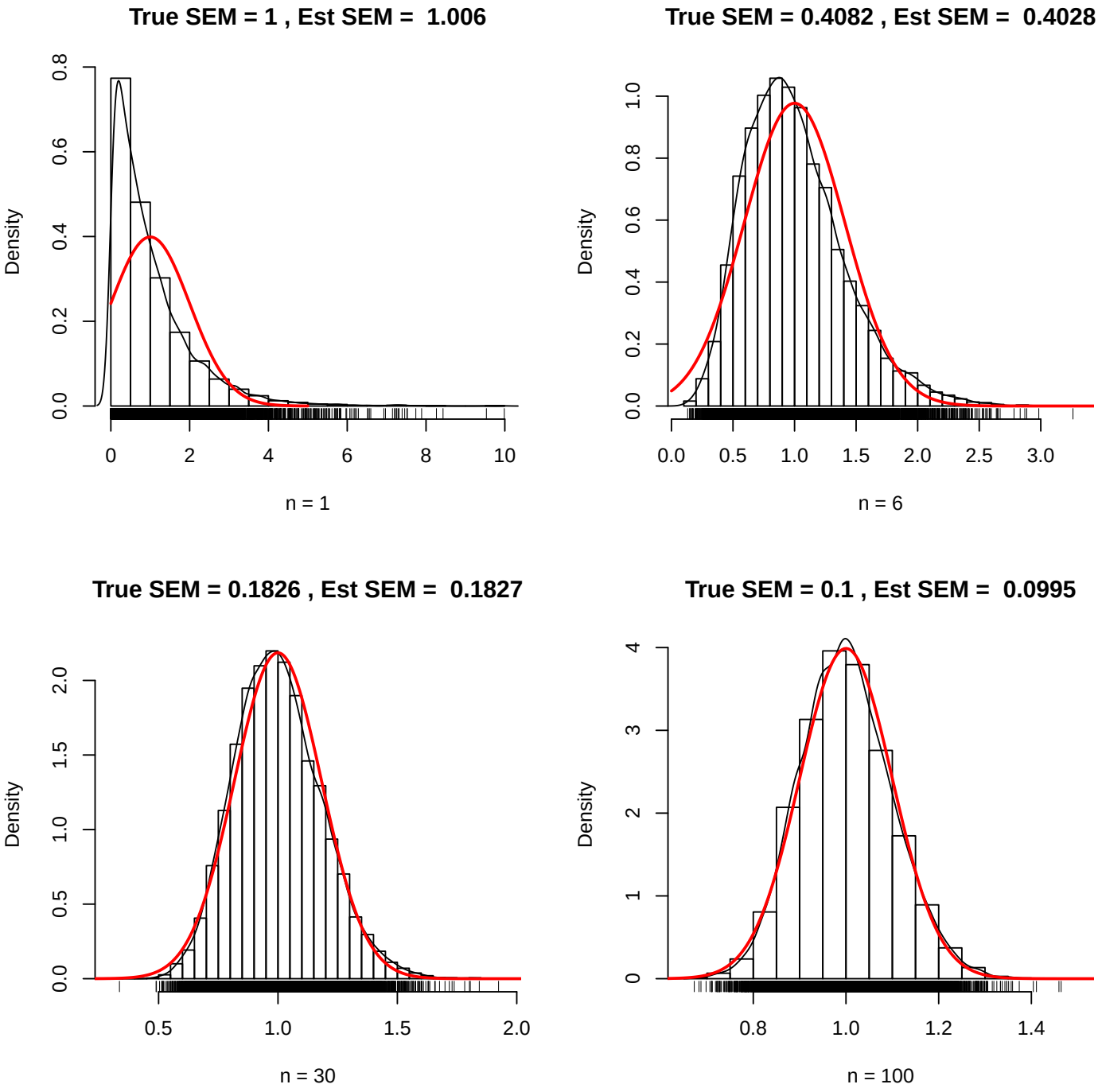


In a more extreme example, we draw samples from an Exponential(1) distribution ($\mu = 1$ and $\sigma = 1$), which is strongly skewed to the right. Notice that the normality promised by the CLT requires larger samples sizes, about $n \geq 30$, than for the previous Uniform(0,1) example, which required about $n \geq 6$.

```
#### Illustration of Central Limit Theorem, Exponential distribution
# demo.clt.exp(N, n) draws N samples of size n from Exponential(1)
# and plots the N means with a normal distribution overlay
demo.clt.exp <- function(N, n) {
```

```
  # draw sample in a matrix with N columns and n rows
  sam <- matrix(rexp(N*n, 1), ncol=N);
  # calculate the mean of each column
  sam.mean <- colMeans(sam)
  # the sd of the mean is the SEM
  sam.se <- sd(sam.mean)
  # calculate the true SEM given the sample size n
  true.se <- sqrt(1/n)
  # draw a histogram of the means
  hist(sam.mean, freq = FALSE, breaks = 25
       , main = paste("True SEM =", round(true.se, 4), ", Est SEM = ", round(sam.se, 4))
       , xlab = paste("n =", n))
  # overlay a density curve for the sample means
  points(density(sam.mean), type = "l")
  # overlay a normal distribution, bold and red
  x <- seq(0, 5, length = 1000)
  points(x, dnorm(x, mean = 1, sd = true.se), type = "l", lwd = 2, col = "red")
  # place a rug of points under the plot
  rug(sam.mean)
}

par(mfrow=c(2,2));
demo.clt.exp(10000,   1);
demo.clt.exp(10000,   6);
demo.clt.exp(10000,  30);
demo.clt.exp(10000, 100);
```



Note well that the further the population distribution is from being normal, the larger the sample size is required to be for the sampling distribution of the sample mean to be normal. If the population distribution is normal, what's the minimum sample size for the sampling distribution of the mean to be normal?

For more examples, try:

```
#### More examples for Central Limit Theorem can be illustrated with this code
# install.packages("TeachingDemos")
library(TeachingDemos)
# look at examples at bottom of the help page
?clt.examp
```

2.1.2 z-score

Given a distribution with mean $\bar{x}$ and standard deviation s, a location-scale transformation known as a z-score will shift the distribution to have mean 0 and scale the spread to have standard deviation 1:

$$z = \frac{x - \bar{x}}{s}.$$

Below, the original variable x has a normal distribution with mean 100 and standard deviation 15, Normal$(100, 15^2)$, and z has a Normal$(0, 1)$ distribution.

```
# sample from normal distribution
df <- data.frame(x = rnorm(100, mean = 100, sd = 15))
df$z <- scale(df$x)  # by default, this performs a z-score transformation

summary(df)
```

```
##        x                 z.V1
##  Min.   : 39.64   Min.   :-3.446123
##  1st Qu.: 90.99   1st Qu.:-0.485300
##  Median :100.00   Median : 0.033925
##  Mean   : 99.41   Mean   : 0.000000
##  3rd Qu.:110.72   3rd Qu.: 0.652006
##  Max.   :132.70   Max.   : 1.919736
```

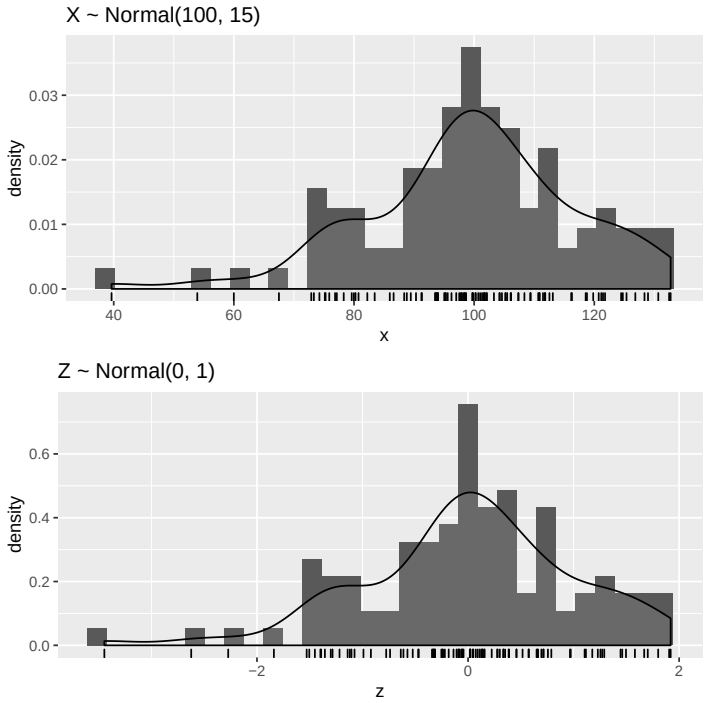
```
## ggplot
library(ggplot2)
p1 <- ggplot(df, aes(x = x))
  # Histogram with density instead of count on y-axis
p1 <- p1 + geom_histogram(aes(y=..density..))
p1 <- p1 + geom_density(alpha=0.1, fill="white")
p1 <- p1 + geom_rug()
p1 <- p1 + labs(title = "X ~ Normal(100, 15)")

p2 <- ggplot(df, aes(x = z))
  # Histogram with density instead of count on y-axis
p2 <- p2 + geom_histogram(aes(y=..density..))
p2 <- p2 + geom_density(alpha=0.1, fill="white")
```

```
p2 <- p2 + geom_rug()
p2 <- p2 + labs(title = "Z ~ Normal(0, 1)")

library(gridExtra)
grid.arrange(grobs = list(p1, p2), ncol=1)
```

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```

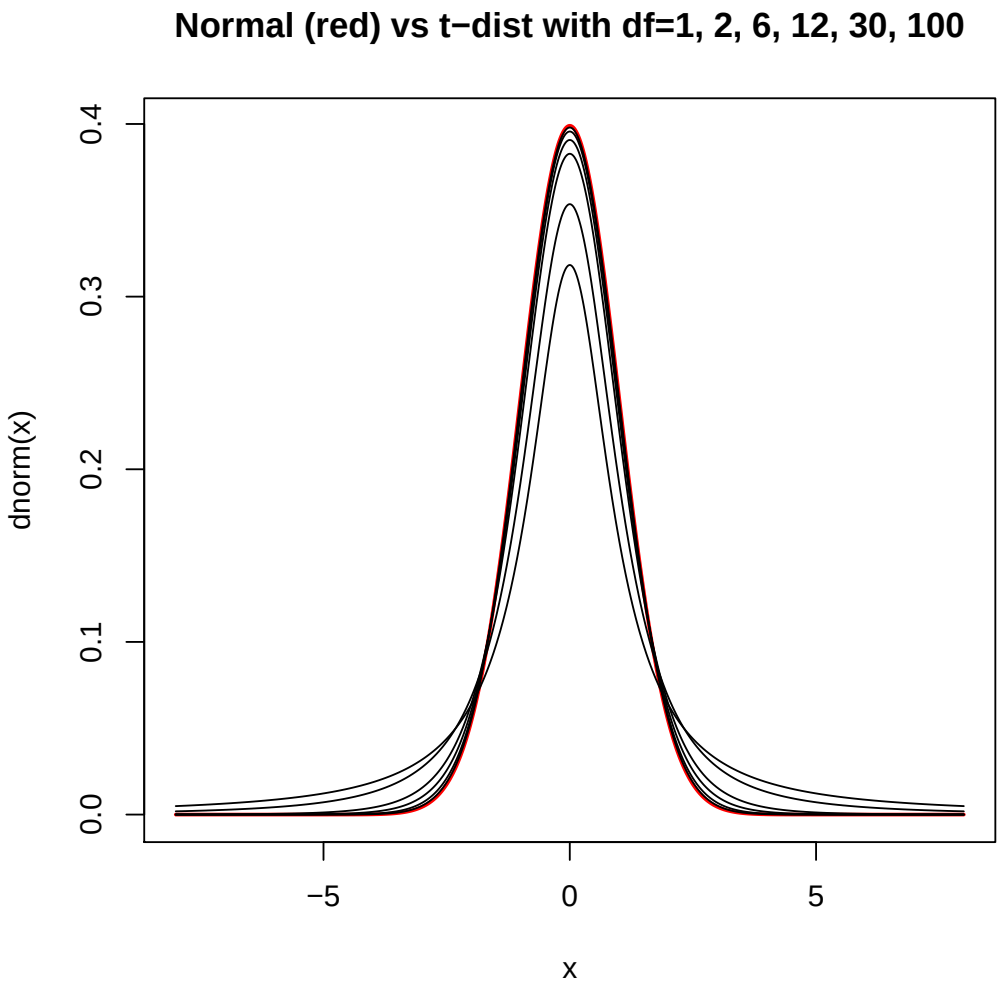


2.1.3 t-distribution

The Student's **t-distribution** is a family of continuous probability distributions that arises when estimating the mean of a normally distributed population in situations where the *sample size is small* and population *standard deviation is unknown*. The t-distribution is symmetric and bell-shaped, like the normal distribution, but has heavier tails, meaning that it is more prone to producing values that fall far from its mean. Effectively, the t-distribution is wider than the normal distribution because in addition to estimating the mean μ with $\bar{Y}$, we *also* have to estimate σ^2 with s^2, so there's some additional uncertainty. The degrees-of-freedom (df) parameter of the t-distribution is the sample size n minus the number of variance parameters estimated. Thus, $df = n - 1$ when we have one sample and $df = n - 2$ when we have two samples. As n increases,

the t-distribution becomes close to the normal distribution, and when $n = \infty$ the distributions are equivalent.

```
#### Normal vs t-distributions with a range of degrees-of-freedom
x <- seq(-8, 8, length = 1000)
par(mfrow=c(1,1))
plot(x, dnorm(x), type = "l", lwd = 2, col = "red"
    , main = "Normal (red) vs t-dist with df=1, 2, 6, 12, 30, 100")
points(x, dt(x,  1), type = "l")
points(x, dt(x,  2), type = "l")
points(x, dt(x,  6), type = "l")
points(x, dt(x, 12), type = "l")
points(x, dt(x, 30), type = "l")
points(x, dt(x,100), type = "l")
```



2.2 CI for μ

Statistical inference provides methods for drawing conclusions about a population from sample data. In this chapter, we want to make a claim about population mean μ given sample statistics $\bar{Y}$ and s.

A CI for μ is a range of plausible values for the unknown population mean μ, based on the observed data, of the form "Best Guess $\pm$ Reasonable Error of the Guess". To compute a CI for μ:

1. Define the **population parameter**, "Let μ = mean [characteristic] for population of interest".
2. Specify the **confidence coefficient**, which is a number between 0 and 100%, in the form $100(1-\alpha)\%$. Solve for α. (For example, 95% has $\alpha = 0.05$.)
3. Compute the t-critical value: $t_{\text{crit}} = t_{0.5\alpha}$ such that the area under the t-curve ($df = n-1$) to the right of t_{crit} is 0.5α. See appendix or internet for a t-table.
4. Report the CI in the form $\bar{Y} \pm t_{\text{crit}} SE_{\bar{Y}}$ or as an interval (L, U). The desired CI has lower and upper endpoints given by $L = \bar{Y} - t_{\text{crit}} SE_{\bar{Y}}$ and $U = \bar{Y} + t_{\text{crit}} SE_{\bar{Y}}$, respectively, where $SE_{\bar{Y}} = s/\sqrt{n}$ is the standard error of the sample mean.
5. Assess method assumptions (see below).

In practice, the confidence coefficient is large, say 95% or 99%, which correspond to $\alpha = 0.05$ and 0.01, respectively. The value of α expressed as a percent is known as the **error rate** of the CI.

The CI is determined once the confidence coefficient is specified and the data are collected. Prior to collecting the data, the interval is unknown and is viewed as random because it will depend on the actual sample selected. Different samples give different CIs. The **"confidence"** in, say, the 95% CI (which has a 5% error rate) can be interpreted as follows. *If you repeatedly sample the population and construct 95% CIs for μ, then 95% of the intervals will contain μ, whereas 5% will not.* The interval you construct from your data will either cover μ, or it will not.

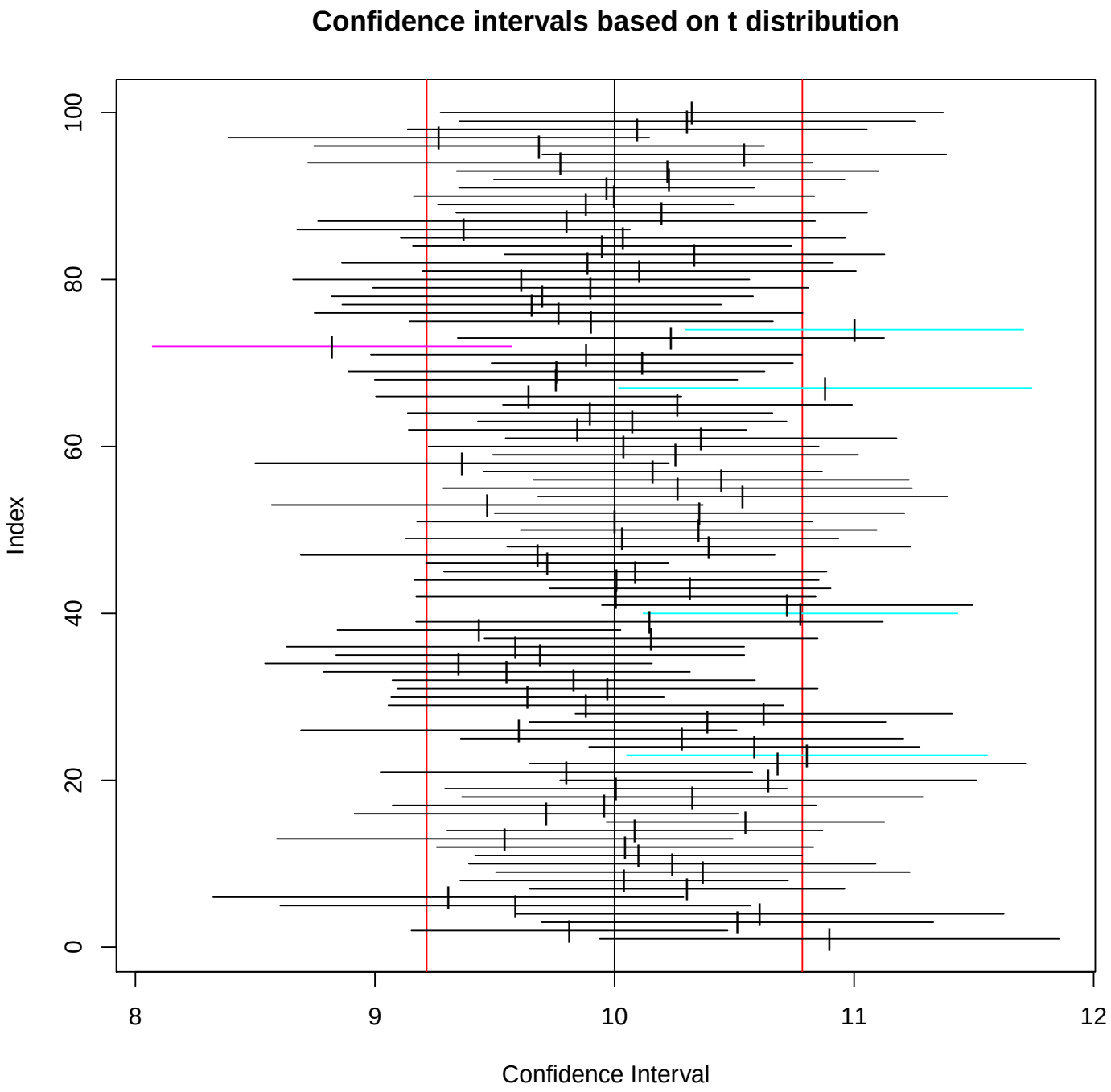
The length of the CI

$$U - L = 2t_{\text{crit}} SE_{\bar{Y}}$$

depends on the accuracy of our estimate $\bar{Y}$ of μ, as measured by the standard error of $\bar{Y}$, $SE_{\bar{Y}} = s/\sqrt{n}$. Less precise estimates of μ lead to wider intervals for a given level of confidence.

An example with 100 CIs Consider drawing a sample of 25 observations from a normally distributed population with mean 10 and sd 2. Calculate the 95% t-CI. Now do that 100 times. The plot belows reflects the variability of that process. We expect 95 of the 100 CIs to contain the true population mean of 10, that is, on average 5 times out of 100 we draw the incorrect inference that the population mean is in an interval when it does not contain the true value of 10.

```
#### Illustration of Confidence Intervals (consistent with their interpretation)
library(TeachingDemos)
ci.examp(mean.sim = 10, sd = 2, n = 25
       , reps = 100, conf.level = 0.95, method = "t")
```



■ ClickerQs — CI for μ, 2 ■

2.2.1 Assumptions for procedures

I described the classical CI. The procedure is based on the assumptions that the data are a **random sample** from the population of interest, and that the

population frequency curve is normal. The population frequency curve can be viewed as a "smoothed histogram" created from the population data.

The normality assumption can never be completely verified without having the entire population data. You can assess the reasonableness of this assumption using a stem-and-leaf display or a boxplot of the sample data. The stem-and-leaf display from the data should resemble a normal curve.

In fact, the assumptions are slightly looser than this, the population frequency curve can be anything provided the sample size is large enough that it's reasonable to assume that the **sampling distribution of the mean is normal**.

Assessing assumptions using the bootstrap

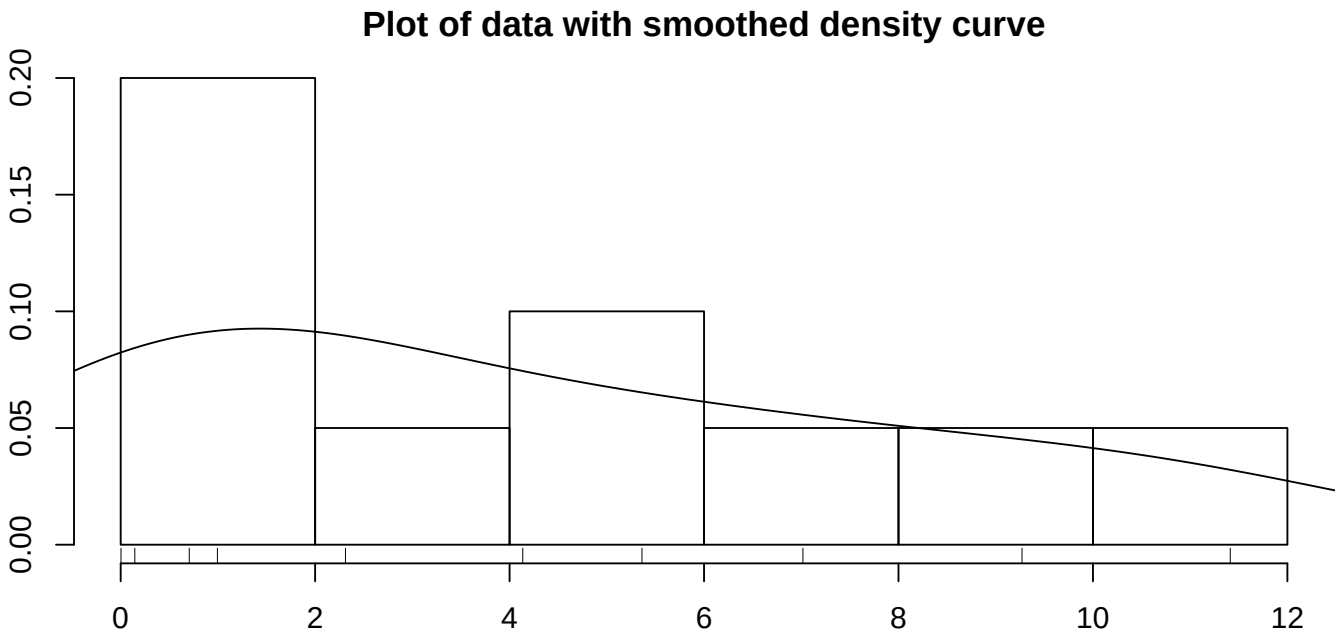
We will cover the bootstrap at the end of this course, but a brief introduction here can help us check model assumptions. Recall in Section 2.1.1 the sampling distribution examples. Assume the sample is representative of the population. Let's use our sample as a proxy for the population and repeatedly draw samples (with replacement) of size n and calculate the mean, then plot the bootstrap sampling distribution of means. If this bootstrap sampling distribution strongly deviates from normal, then that's evidence from the data that inference using the t-distribution is not appropriate. Otherwise, if roughly normal, then the t-distribution may be sensible.

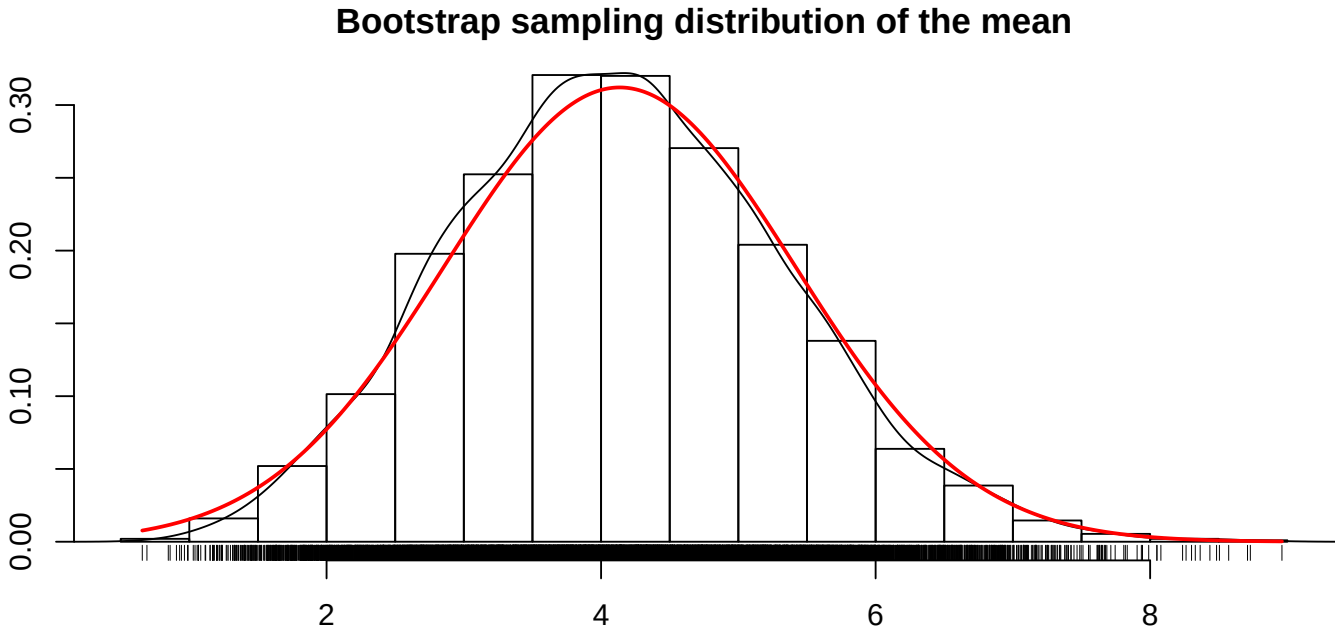
```
#### Visual comparison of whether sampling distribution is close to Normal via Bootstrap
# a function to compare the bootstrap sampling distribution with
#   a normal distribution with mean and SEM estimated from the data
bs.one.samp.dist <- function(dat, N = 1e4) {
  n <- length(dat);
  # resample from data
  sam <- matrix(sample(dat, size = N * n, replace = TRUE), ncol=N);
  # draw a histogram of the means
  sam.mean <- colMeans(sam);
  # save par() settings
  old.par <- par(no.readonly = TRUE)
  # make smaller margins
  par(mfrow=c(2,1), mar=c(3,2,2,1), oma=c(1,1,1,1))
  # Histogram overlaid with kernel density curve
  hist(dat, freq = FALSE, breaks = 6
```

```
      , main = "Plot of data with smoothed density curve")
  points(density(dat), type = "l")
  rug(dat)

  hist(sam.mean, freq = FALSE, breaks = 25
      , main = "Bootstrap sampling distribution of the mean"
      , xlab = paste("Data: n =", n
                     , ", mean =", signif(mean(dat), digits = 5)
                     , ", se =", signif(sd(dat)/sqrt(n)), digits = 5))
  # overlay a density curve for the sample means
  points(density(sam.mean), type = "l")
  # overlay a normal distribution, bold and red
  x <- seq(min(sam.mean), max(sam.mean), length = 1000)
  points(x, dnorm(x, mean = mean(dat), sd = sd(dat)/sqrt(n))
       , type = "l", lwd = 2, col = "red")
  # place a rug of points under the plot
  rug(sam.mean)
  # restore par() settings
  par(old.par)
}

# example data, skewed --- try others datasets to develop your intuition
x <- rgamma(10, shape = .5, scale = 20)
bs.one.samp.dist(x)
```





Example: Age at First Heart Transplant Let us go through a hand-calculation of a CI, using R to generate summary data. I'll show you later how to generate the CI in R.

We are interested in the mean age at first heart transplant for a population of patients.

1. Define the population parameter

Let μ = mean age at the time of first heart transplant for population of patients.

2. Calculate summary statistics from sample

The ages (in years) at first transplant for a sample of 11 heart transplant patients are as follows:

54, 42, 51, 54, 49, 56, 33, 58, 54, 64, 49.

Summaries for the data are: $n = 11$, $\bar{Y} = 51.27$, and $s = 8.26$ so that $SE_{\bar{Y}} = 8.26/\sqrt{11} = 2.4904$. The degrees of freedom are $df = 11 - 1 = 10$.

3. Specify confidence level, find critical value, calculate limits

Let us calculate a 95% CI for μ. For a 95% CI $\alpha = 0.05$, so we need to find $t_{\text{crit}} = t_{0.025}$, which is 2.228. Now $t_{\text{crit}} SE_{\bar{Y}} = 2.228 \times 2.4904 = 5.55$. The lower limit on the CI is $L = 51.27 - 5.55 = 45.72$. The upper limit is $U = 51.27 + 5.55 = 56.82$.

4. Summarize in words For example, I am 95% confident that the population mean age at first transplant is 51.3 ± 5.55, that is, between 45.7 and 56.8 years (rounding off to 1 decimal place).

5. Check assumptions We will see this in several pages, sampling distribution is reasonably normal.

2.2.2 The effect of α on a two-sided CI

A two-sided $100(1-\alpha)\%$ CI for μ is given by $\bar{Y} \pm t_{\text{crit}} SE_{\bar{Y}}$. The CI is centered at $\bar{Y}$ and has length $2t_{\text{crit}} SE_{\bar{Y}}$. The confidence coefficient $100(1-\alpha)\%$ is **increased** by **decreasing** α, which increases t_{crit}. That is, increasing the confidence coefficient makes the CI wider. This is sensible: to increase your confidence that the interval captures μ you must pinpoint μ with less precision by making the CI wider. For example, a 95% CI is wider than a 90% CI.

■ ClickerQs — CI, quickie ■

2.3 Hypothesis Testing for μ

A **hypothesis test** is used to make a **decision** about a population parameter.

Suppose you are interested in checking whether the population mean μ is equal to some prespecified value, say μ_0. This question can be formulated as a two-sided hypothesis test, where you are trying to decide which of two contradictory claims or hypotheses about μ is more reasonable given the observed data. The **null hypothesis**, or the hypothesis under test, is $H_0 : \mu = \mu_0$, whereas the **alternative hypothesis** is $H_A : \mu \neq \mu_0$.

I will explore the ideas behind hypothesis testing later. At this point, I focus on the mechanics behind the test. The steps in carrying out the test are:

1. Set up the **null and alternative hypotheses** in words and notation.
   In words: "The population mean for [what is being studied] is different from [value of μ_0]." (Note that the statement in words is in terms of the alternative hypothesis.)
   In notation: $H_0 : \mu = \mu_0$ versus $H_A : \mu \neq \mu_0$ (where μ_0 is specified by the context of the problem).
2. Choose the **size** or **significance level** of the test, denoted by α. In practice, α is set to a small value, say, 0.01 or 0.05, but theoretically can be any value between 0 and 1.
3. Compute the **test statistic**

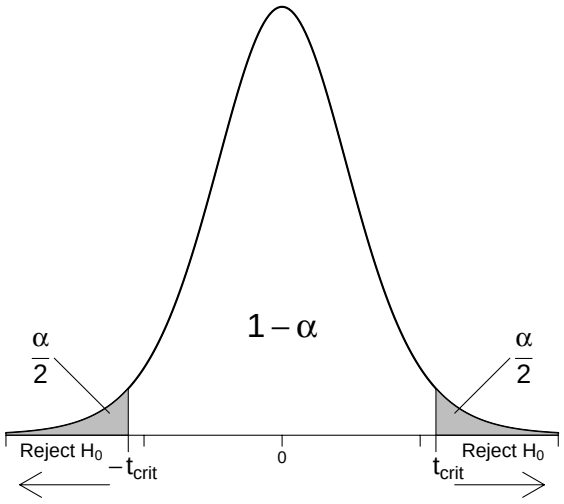
$$t_s = \frac{\bar{Y} - \mu_0}{SE_{\bar{Y}}},$$

   where $SE_{\bar{Y}} = s/\sqrt{n}$ is the standard error.
   Note: I sometimes call the test statistic t_{obs} to emphasize that the computed value depends on the observed data.
4. Compute the **critical value** $t_{\text{crit}} = t_{0.5\alpha}$ (or p-value from the test statistic) in the direction of the alternative hypothesis from the t-distribution table with degrees of freedom $df = n - 1$.
5. State the **conclusion** in terms of the problem.
   Reject H_0 in favor of H_A (i.e., decide that H_0 is false, based on the data)

if $|t_s| > t_{\text{crit}}$ or p-value $< \alpha$, that is, reject if $t_s < -t_{\text{crit}}$ or if $t_s > t_{\text{crit}}$. Otherwise, **Fail to reject** H_0.
(Note: We DO NOT *accept* H_0 — more on this later.)

6. **Check assumptions** of the test, when possible (could do earlier to save yourself some effort if they are not met).

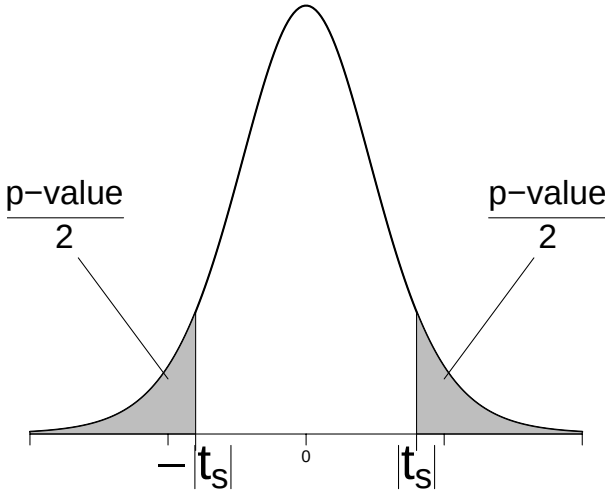
The process is represented graphically below. The area under the t-probability curve outside $\pm t_{\text{crit}}$ is the size of the test, α. One-half α is the area in each tail. You reject H_0 in favor of H_A only if the test statistic is outside $\pm t_{\text{crit}}$.



2.3.1 P-values

The **p-value**, or **observed significance level** for the test, provides a measure of plausibility for H_0. Smaller values of the p-value imply that H_0 is less plausible. To compute the p-value for a two-sided test, you

1. Compute the test statistic t_s as above.
2. Evaluate the area under the t-probability curve (with $df = n-1$) outside $\pm|t_s|$.



The p-value is the total shaded area, or twice the area in either tail. A useful **interpretation** of the p-value is that *it is the chance of obtaining data favoring H_A by this much or more* if H_0 *actually is true.* Another interpretation is that

> the **p-value** is *the probability of observing a sample mean at least as extreme as the one observed assuming μ_0 from H_0 is the true population mean.*

If the p-value is small then the sample we obtained is pretty unusual to have obtained if H_0 is true — but we actually got the sample, so probably it is not very unusual, so we would conclude H_0 is false (it would not be unusual if H_A is true).

Most, if not all, statistical packages summarize hypothesis tests with a p-value, rather than a decision (i.e., reject or not reject at a given α level). You can make a decision to reject or not reject H_0 for a size α test based on the p-value as follows — *reject H_0 if the p-value is less than α.* This decision is identical to that obtained following the formal rejection procedure given earlier. The reason for this is that the p-value can be interpreted as the smallest value you can set the size of the test and still reject H_0 given the observed data.

There are a lot of terms to keep straight here. α and t_{crit} are constants we choose (actually, one determines the other so we really only choose one, usually α) to set how rigorous evidence against H_0 needs to be. t_s and the p-value (again, one determines the other) are random variables because they are calculated from the random sample. They are the evidence against H_0.

2.3.2 Assumptions for procedures

I described the classical t-test, which assumes that the data are a random sample from the population and that the population frequency curve is normal. These are the same assumptions as for the CI.

Example: Age at First Transplant (Revisited) The ages (in years) at first transplant for a sample of 11 heart transplant patients are as follows: 54, 42, 51, 54, 49, 56, 33, 58, 54, 64, 49. Summaries for these data are: $n = 11$, $\bar{Y} = 51.27$, $s = 8.26$ and $SE_{\bar{Y}} = 2.4904$. Test the hypothesis that the mean age at first transplant is 50. Use $\alpha = 0.05$.

As in the earlier analysis, define

μ = mean age at time of first transplant for population of patients.

We are interested in testing $H_0 : \mu = 50$ against $H_A : \mu \neq 50$, so $\mu_0 = 50$.

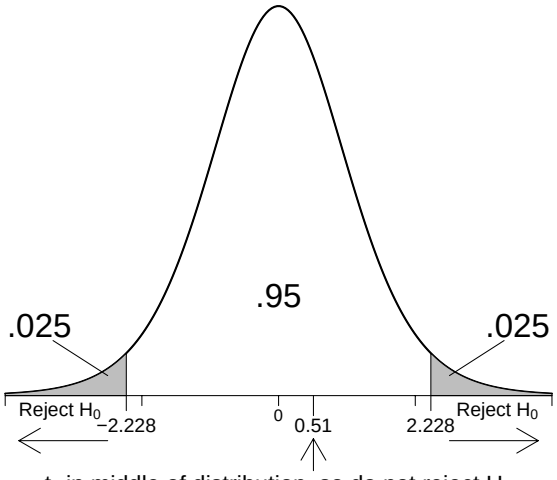
The degrees of freedom are $df = 11 - 1 = 10$. The critical value for a 5% test is $t_{\text{crit}} = t_{0.025} = 2.228$. (Note $\alpha/2 = 0.05/2 = 0.025$). The same critical value was used with the 95% CI.

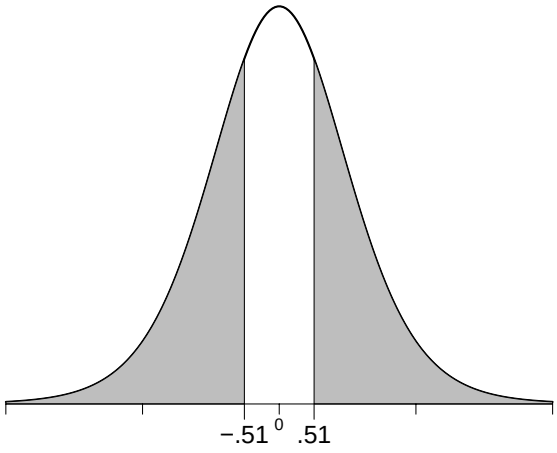
For the test,

$$t_s = \frac{\bar{Y} - \mu_0}{SE_{\bar{Y}}} = \frac{51.27 - 50}{2.4904} = 0.51.$$

Since $t_{\text{crit}} = 2.228$, we do not reject H_0 using a 5% test. Notice the placement of t_s relative to t_{crit} in the picture below. Equivalently, the p-value for the test is 0.62, thus we fail to reject H_0 because $0.62 > 0.05 = \alpha$. The results of the hypothesis test should not be surprising, since the CI tells you that 50 is a plausible value for the population mean age at transplant. Note: All you can

say is that the data *could have* come from a distribution with a mean of 50 — this is not convincing evidence that μ actually *is* 50.





Example: Age at First Transplant R output for the heart transplant problem is given below. Let us look at the output and find all of the summaries we computed. Also, look at the graphical summaries to assess whether the t-test and CI are reasonable here.

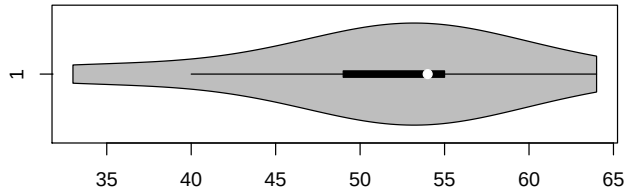
```
#### Example: Age at First Transplant
# enter data as a vector
age <- c(54, 42, 51, 54, 49, 56, 33, 58, 54, 64, 49)
```

The age data is unimodal, skewed left, no extreme outliers.

```
par(mfrow=c(2,1))
# Histogram overlaid with kernel density curve
hist(age, freq = FALSE, breaks = 6)
points(density(age), type = "l")
rug(age)

# violin plot
library(vioplot)
vioplot(age, horizontal=TRUE, col="gray")
```

```
## [1] 33 64
```




```
# stem-and-leaf plot
stem(age, scale=2)
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   3 | 3
##   3 |
##   4 | 2
##   4 | 99
##   5 | 1444
##   5 | 68
##   6 | 4
```

```
# t.crit
qt(1 - 0.05/2, df = length(age) - 1)
```

```
## [1] 2.228139
```

```
# look at help for t.test
?t.test
  # defaults include: alternative = "two.sided", conf.level = 0.95
t.summary <- t.test(age, mu = 50)
t.summary
```

```
##
##  One Sample t-test
##
## data:  age
## t = 0.51107, df = 10, p-value = 0.6204
## alternative hypothesis: true mean is not equal to 50
## 95 percent confidence interval:
##  45.72397 56.82149
```

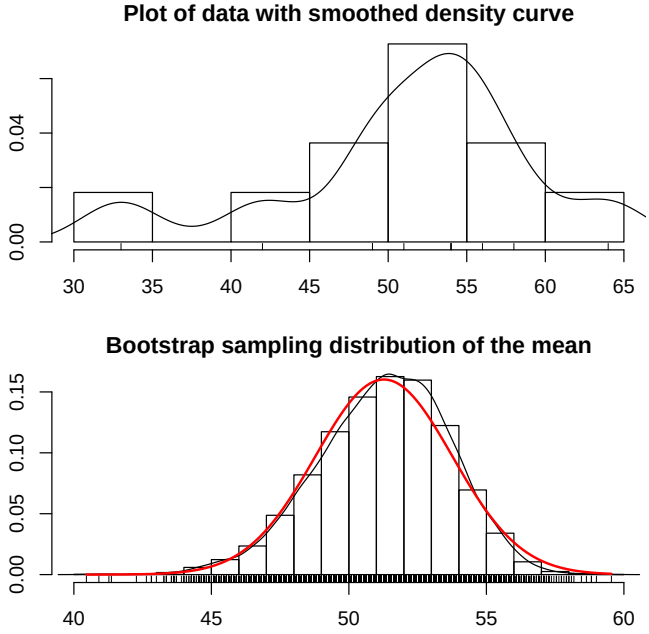
```
## sample estimates:
## mean of x
##  51.27273
```

```
summary(age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   33.00   49.00   54.00   51.27   55.00   64.00
```

The assumption of normality of the sampling distribution appears reasonablly close, using the bootstrap discussed earlier. Therefore, the results for the t-test above can be trusted.

```
bs.one.samp.dist(age)
```

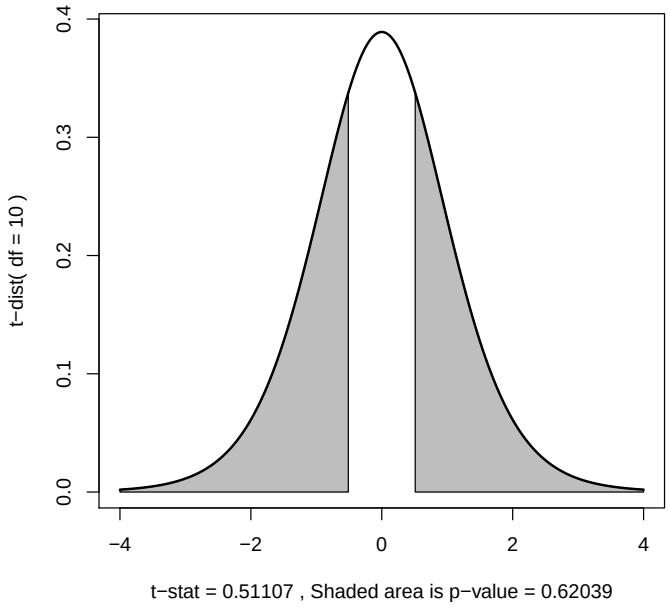


Aside: To print the shaded region for the p-value, you can use the result of `t.test()` with the function `t.dist.pval()` defined here.

```
# Function to plot t-distribution with shaded p-value
t.dist.pval <- function(t.summary) {
  par(mfrow=c(1,1))
  lim.extreme <- max(4, abs(t.summary$statistic) + 0.5)
  lim.lower <- -lim.extreme;
  lim.upper <-  lim.extreme;
  x.curve <- seq(lim.lower, lim.upper, length=200)
  y.curve <- dt(x.curve, df = t.summary$parameter)
  plot(x.curve, y.curve, type = "n"
       , ylab = paste("t-dist( df =", signif(t.summary$parameter, 3), ")")
       , xlab = paste("t-stat =", signif(t.summary$statistic, 5)
                      , ", Shaded area is p-value =", signif(t.summary$p.value, 5)))
  if ((t.summary$alternative == "less")
      | (t.summary$alternative == "two.sided")) {
    x.pval.l <- seq(lim.lower, -abs(t.summary$statistic), length=200)
```

```
    y.pval.l <- dt(x.pval.l, df = t.summary$parameter)
    polygon(c(lim.lower, x.pval.l, -abs(t.summary$statistic))
          , c(0, y.pval.l, 0), col="gray")
  }
  if ((t.summary$alternative == "greater")
      | (t.summary$alternative == "two.sided")) {
    x.pval.u <- seq(abs(t.summary$statistic), lim.upper, length=200)
    y.pval.u <- dt(x.pval.u, df = t.summary$parameter)
    polygon(c(abs(t.summary$statistic), x.pval.u, lim.upper)
          , c(0, y.pval.u, 0), col="gray")
  }
  points(x.curve, y.curve, type = "l", lwd = 2, col = "black")
}

# for the age example
t.dist.pval(t.summary)
```



Aside: Note that the `t.summary` object returned from `t.test()` includes a number of quantities that might be useful for additional calculations.

```
names(t.summary)
```

```
##  [1] "statistic"   "parameter"   "p.value"     "conf.int"
##  [5] "estimate"    "null.value"  "stderr"      "alternative"
##  [9] "method"      "data.name"
```

```
t.summary$statistic
```

```
##         t
## 0.5110715
```

```
t.summary$parameter
```

```
## df
## 10
```

```
t.summary$p.value
```

```
## [1] 0.6203942
```

```
t.summary$conf.int
```

```
## [1] 45.72397 56.82149
## attr(,"conf.level")
## [1] 0.95
```

```
t.summary$estimate
```

```
## mean of x
##  51.27273
```

```
t.summary$null.value
```

```
## mean
##   50
```

```
t.summary$alternative
```

```
## [1] "two.sided"
```

```
t.summary$method
```

```
## [1] "One Sample t-test"
```

```
t.summary$data.name
```

```
## [1] "age"
```

Example: Meteorites One theory of the formation of the solar system states that all solar system meteorites have the same evolutionary history and thus have the same cooling rates. By a delicate analysis based on measurements of phosphide crystal widths and phosphide-nickel content, the cooling rates, in degrees Celsius per million years, were determined for samples taken from meteorites named in the accompanying table after the places they were found. The Walker[2] County (Alabama, US), Uwet[3] (Cross River, Nigeria), and Tocopilla[4] (Antofagasta, Chile) meteorite cooling rate data are below.

Suppose that a hypothesis of solar evolution predicted a mean cooling rate of $\mu = 0.54$ degrees per million years for the Tocopilla meteorite. Do the observed cooling rates support this hypothesis? Test at the 5% level. The boxplot and stem-and-leaf display (given below) show good symmetry. The assumption of a normal distribution of observations basic to the t-test appears to be realistic.

Meteorite	Cooling rates
Walker County	0.69 0.23 0.10 0.03 0.56 0.10 0.01 0.02 0.04 0.22
Uwet	0.21 0.25 0.16 0.23 0.47 1.20 0.29 1.10 0.16
Tocopilla	5.60 2.70 6.20 2.90 1.50 4.00 4.30 3.00 3.60 2.40 6.70 3.80

Let

[2] http://www.lpi.usra.edu/meteor/metbull.php?code=24204

[3] http://www.lpi.usra.edu/meteor/metbull.php?code=24138

[4] http://www.lpi.usra.edu/meteor/metbull.php?code=17001

μ = mean cooling rate over all pieces of the Tocopilla meteorite.

To answer the question of interest, we consider the test of $H_0 : \mu = 0.54$ against $H_A : \mu \neq 0.54$. Let us go carry out the test, compute the p-value, and calculate a 95% CI for μ. The sample summaries are $n = 12$, $\bar{Y} = 3.892$, $s = 1.583$. The standard error is $SE_{\bar{Y}} = s/\sqrt{n} = 0.457$.

R output for this problem is given below. For a 5% test (i.e., $\alpha = 0.05$), you would reject H_0 in favor of H_A because the p-value ≤ 0.05. The data strongly suggest that $\mu \neq 0.54$. The 95% CI says that you are 95% confident that the population mean cooling rate for the Tocopilla meteorite is between 2.89 and 4.90 degrees per million years. Note that the CI gives us a means to assess how different μ is from the hypothesized value of 0.54.

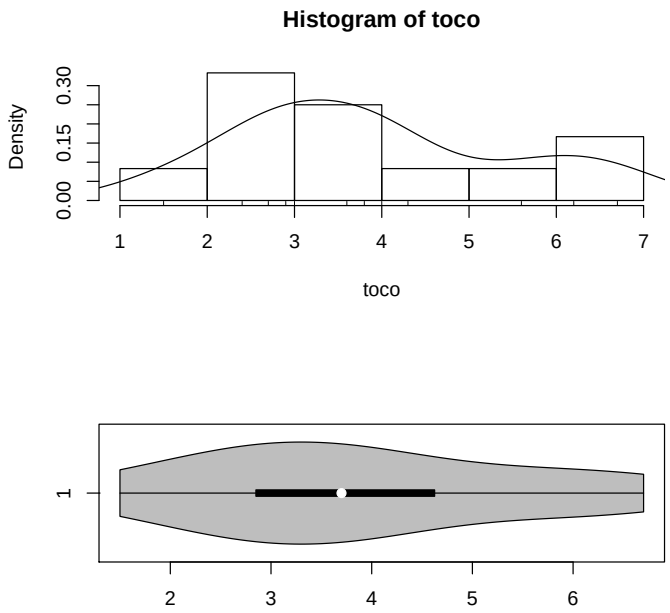
```
#### Example: Meteorites
# enter data as a vector
toco <- c(5.6, 2.7, 6.2, 2.9, 1.5, 4.0, 4.3, 3.0, 3.6, 2.4, 6.7, 3.8)
```

The Tocopilla data is unimodal, skewed right, no extreme outliers.

```
par(mfrow=c(2,1))
# Histogram overlaid with kernel density curve
hist(toco, freq = FALSE, breaks = 6)
points(density(toco), type = "l")
rug(toco)

# violin plot
library(vioplot)
vioplot(toco, horizontal=TRUE, col="gray")
```

```
## [1] 1.5 6.7
```




```
# stem-and-leaf plot
stem(toco, scale=2)
```

```
##
##   The decimal point is at the |
##
##   1 | 5
##   2 | 479
##   3 | 068
##   4 | 03
##   5 | 6
##   6 | 27
```

```
# t.crit
qt(1 - 0.05/2, df = length(toco) - 1)
```

```
## [1] 2.200985
```

```
t.summary <- t.test(toco, mu = 0.54)
t.summary
```

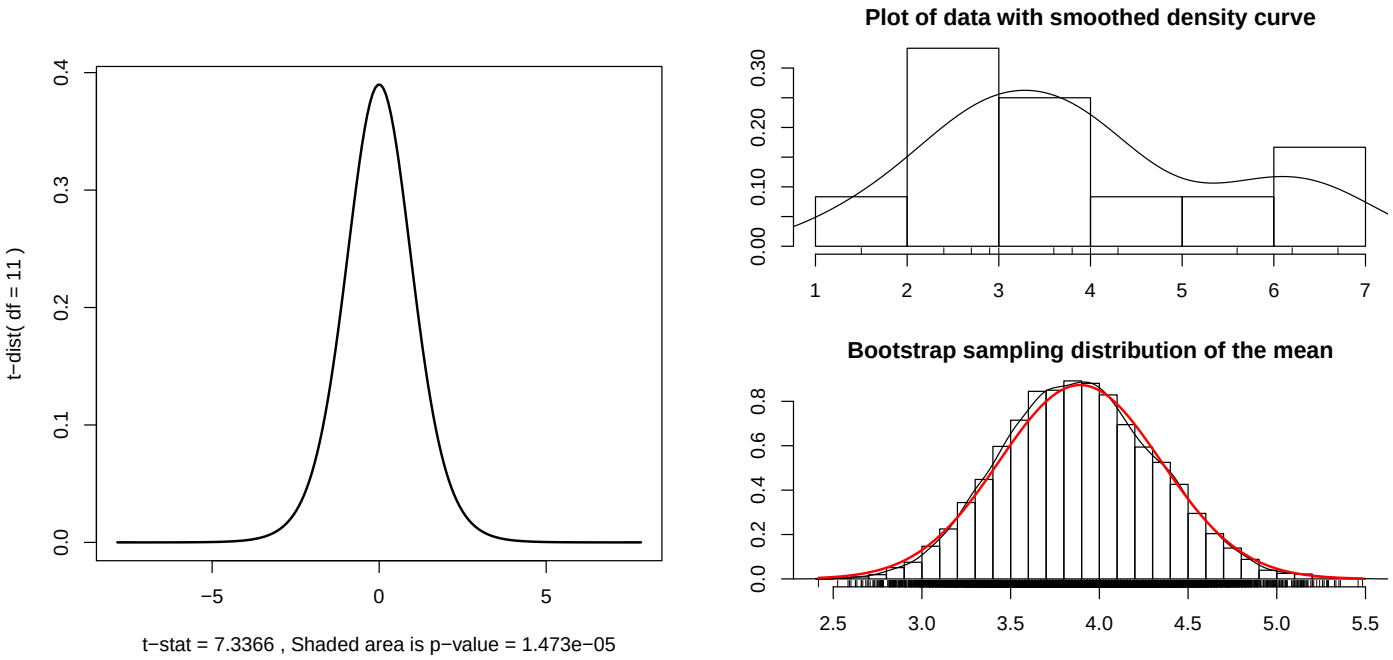
```
##
##  One Sample t-test
##
## data:  toco
## t = 7.3366, df = 11, p-value = 1.473e-05
## alternative hypothesis: true mean is not equal to 0.54
## 95 percent confidence interval:
##  2.886161 4.897172
## sample estimates:
## mean of x
##  3.891667
```

```
summary(toco)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.500   2.850   3.700   3.892   4.625   6.700
```

The assumption of normality of the sampling distribution appears reasonable. Therefore, the results for the t-test above can be trusted.

```
t.dist.pval(t.summary)
bs.one.samp.dist(toco)
```



2.3.3 The mechanics of setting up hypothesis tests

When setting up a test you should imagine you are the researcher conducting the experiment. In many studies, the researcher wishes to establish that there has been a change from the **status quo**, or that they have developed a method that produces a **change** (possibly in a specified direction) in the typical response. The researcher sets H_0 to be the **status quo** and H_A to be the **research hypothesis** — the claim the researcher wishes to make. In some studies you define the hypotheses so that H_A is the **take action** hypothesis — rejecting H_0 in favor of H_A leads one to take a radical action.

Some perspective on testing is gained by understanding the mechanics behind the tests. A hypothesis test is a decision process in the face of uncertainty. You are given data and asked which of two contradictory claims about a pop-

ulation parameter, say μ, is more reasonable. Two decisions are possible, but whether you make the correct decision depends on the true state of nature which is unknown to you.

	State of nature	
Decision	H_0 true	H_A true
Fail to reject [accept] H_0	correct decision	*Type-II error*
Reject H_0 in favor of H_A	*Type-I error*	correct decision

For a given problem, only one of these errors is possible. For example, if H_0 is true you can make a Type-I error but not a Type-II error. Any reasonable decision rule based on the data that tells us when to reject H_0 and when to not reject H_0 will have a certain probability of making a Type-I error if H_0 is true, and a corresponding probability of making a Type-II error if H_0 is false and H_A is true. For a given decision rule, define

$$\alpha = \text{Prob(Reject } H_0 \text{ given } H_0 \text{ is true)} = \text{Prob(Type-I error)}$$

and

$$\beta = \text{Prob(Fail to reject } H_0 \text{ when } H_A \text{ true)} = \text{Prob(Type-II error).}$$

The mathematics behind hypothesis tests allows you to prespecify or control α. For a given α, the tests we use (typically) have the smallest possible value of β. Given the researcher can control α, you set up the hypotheses so that committing a Type-I error is more serious than committing a Type-II error. The magnitude of α, also called the **size** or **level** of the test, should depend on the seriousness of a Type-I error in the given problem. The more serious the consequences of a Type-I error, the smaller α should be. In practice α is often set to 0.10, 0.05, or 0.01, with $\alpha = 0.05$ being the scientific standard. By setting α to be a small value, you reject H_0 in favor of H_A only if the data **convincingly indicate** that H_0 is false.

Let us piece together these ideas for the meteorite problem. Evolutionary history predicts $\mu = 0.54$. A scientist examining the validity of the theory is trying to decide whether $\mu = 0.54$ or $\mu \neq 0.54$. Good scientific practice dictates that rejecting another's claim when it is true is more serious than not being able to reject it when it is false. This is consistent with defining $H_0 : \mu = 0.54$ (the

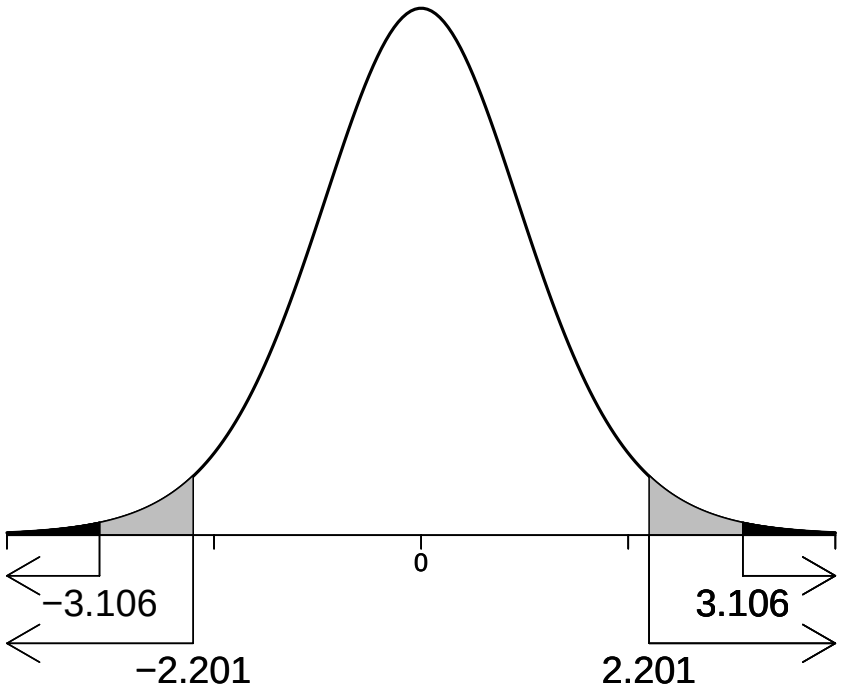
status quo) and $H_A : \mu \neq 0.54$. To convince yourself, note that the implications of a Type-I error would be to claim the evolutionary theory is false when it is true, whereas a Type-II error would correspond to not being able to refute the evolutionary theory when it is false. With this setup, the scientist will refute the theory only if the data overwhelmingly suggest that it is false.

2.3.4 The effect of α on the rejection region of a two-sided test

For a size α test, you reject $H_0 : \mu = \mu_0$ if

$$t_s = \frac{\bar{Y} - \mu_0}{SE_{\bar{Y}}}$$

satisfies $|t_s| > t_{\text{crit}}$.



Rejection Regions for .05 and .01 level tests

The critical value is computed so that the area under the t-probability curve (with $df = n - 1$) outside $\pm t_{\text{crit}}$ is α, with 0.5α in each tail. Reducing α makes t_{crit} larger. That is, reducing the size of the test makes rejecting H_0 harder because the rejection region is smaller. A pictorial representation is given above for the Tocopilla data, where $\mu_0 = 0.54$, $n = 12$, and $df = 11$. Note that $t_{\text{crit}} = 2.201$ and 3.106 for $\alpha = 0.05$ and 0.01, respectively.

The mathematics behind the test presumes that H_0 is true. Given the data, you use

$$t_s = \frac{\bar{Y} - \mu_0}{SE_{\bar{Y}}}$$

to measure how far $\bar{Y}$ is from μ_0, relative to the spread in the data given by $SE_{\bar{Y}}$. For t_s to be in the rejection region, $\bar{Y}$ must be significantly above or below μ_0, relative to the spread in the data. To see this, note that rejection

rule can be expressed as: **Reject** H_0 if

$$\bar{Y} < \mu_0 - t_{\text{crit}} SE_{\bar{Y}} \quad \text{or} \quad \bar{Y} > \mu_0 + t_{\text{crit}} SE_{\bar{Y}}.$$

The rejection rule is sensible because $\bar{Y}$ is our best guess for μ. You would reject $H_0 : \mu = \mu_0$ only if $\bar{Y}$ is so far from μ_0 that you would question the reasonableness of assuming $\mu = \mu_0$. How far $\bar{Y}$ must be from μ_0 before you reject H_0 depends on α (i.e., how willing you are to reject H_0 if it is true), and on the value of $SE_{\bar{Y}}$. For a given sample, reducing α forces $\bar{Y}$ to be further from μ_0 before you reject H_0. For a given value of α and s, increasing n allows smaller differences between $\bar{Y}$ and μ_0 to be **statistically significant** (i.e., lead to rejecting H_0). In problems where small differences between $\bar{Y}$ and μ_0 lead to rejecting H_0, you need to consider whether the observed differences are important.

In essence, the t-distribution provides an objective way to calibrate whether the observed $\bar{Y}$ is typical of what sample means look like when sampling from a normal population where H_0 is true. If all other assumptions are satisfied, and $\bar{Y}$ is inordinately far from μ_0, then our only recourse is to conclude that H_0 must be incorrect.

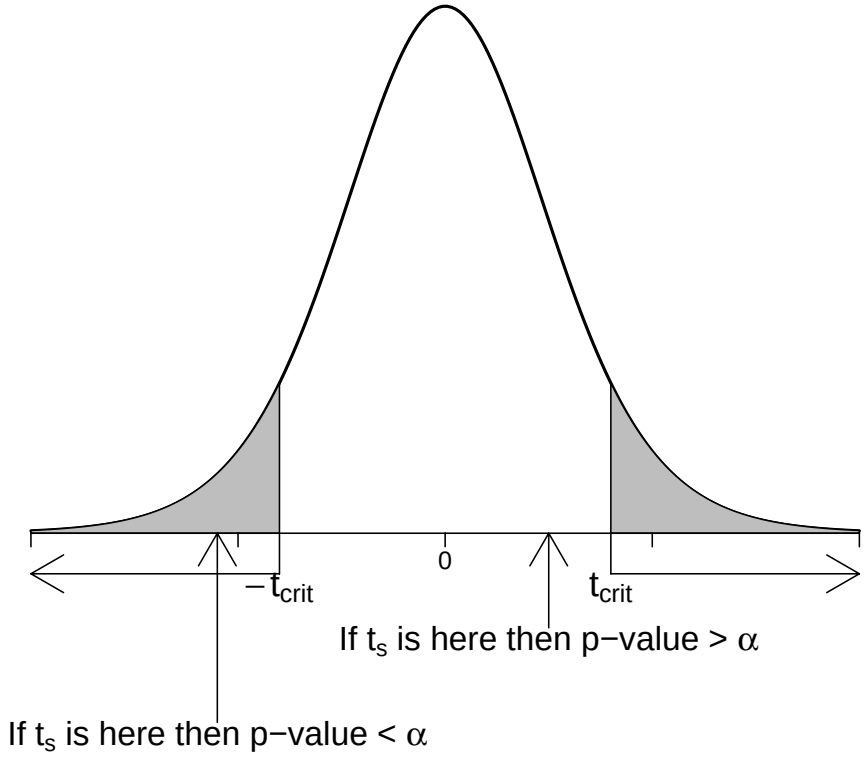
2.4 Two-sided tests, CI and p-values

An important relationship among two-sided tests of $H_0 : \mu = \mu_0$, CI, and p-values is that

$$\text{size } \alpha \text{ test rejects } H_0 \Leftrightarrow 100(1-\alpha)\% \text{ CI does not contain } \mu_0 \Leftrightarrow \text{p-value} \leq \alpha, \text{ and}$$

$$\text{size } \alpha \text{ test fails to reject } H_0 \Leftrightarrow 100(1-\alpha)\% \text{ CI contains } \mu_0 \Leftrightarrow \text{p-value} > \alpha.$$

For example, an $\alpha = 0.05$ test rejects $H_0 \Leftrightarrow$ 95% CI does not contain $\mu_0 \Leftrightarrow$ p-value ≤ 0.05. The picture below illustrates the connection between p-values and rejection regions.



Either a CI or a test can be used to decide the plausibility of the claim that $\mu = \mu_0$. Typically, you use the test to answer the question **is there a difference?** If so, you use the CI to assess **how much of a difference exists**. I believe that scientists place too much emphasis on hypothesis testing.

2.5 Statistical versus practical significance

Suppose in the Tocopilla meteorite example, you rejected $H_0 : \mu = 0.54$ at the 5% level and found a 95% two-sided CI for μ to be 0.55 to 0.58. Although you have sufficient evidence to conclude that the population mean cooling rate μ differs from that suggested by evolutionary theory, the range of plausible values for μ is small and contains only values close to 0.54. Although you have

shown statistical significance here, you need to ask yourself whether the actual difference between μ and 0.54 is large enough to be important. The answer to such questions is always problem specific.

2.6 Design issues and power

An experiment may not be sensitive enough to pick up true differences. For example, in the Tocopilla meteorite example, suppose the true mean cooling rate is $\mu = 1.00$. To have a 50% chance of correctly rejecting $H_0 : \mu = 0.54$, you would need about $n = 48$ observations. If the true mean is $\mu = 0.75$, you would need about 221 observations to have a 50% chance of correctly rejecting H_0. In general, the smaller the difference between the true and hypothesized mean (relative to the spread in the population), the more data that is needed to reject H_0. If you have prior information on the expected difference between the true and hypothesized mean, you can design an experiment appropriately by choosing the sample size required to likely reject H_0.

The **power** of a test is the probability of correctly rejecting H_0 when it is false. Equivalently,

power $= 1 -$ Prob(failing to reject H_0 when it is false) $= 1 -$ Prob(Type-II error).

For a given sample size, the tests I have discussed have maximum power (or smallest probability of a Type-II error) among all tests with fixed size α. However, the actual power may be small, so sample size calculations, as briefly highlighted above, are important prior to collecting data. See your local statistician.

```
#### Power calculations (that you can do on your own)
# install.packages("pwr")
library(pwr)
?power.t.test
power.t.test(n = NULL, delta = 1.00 - 0.54, sd = sd(toco),
             , sig.level = 0.05, power = 0.50
             , type = "one.sample", alternative = "two.sided")
power.t.test(n = NULL, delta = 0.75 - 0.54, sd = sd(toco),
             , sig.level = 0.05, power = 0.50
```

```
            , type = "one.sample", alternative = "two.sided")
```

For more examples, try:

```
# install.packages("TeachingDemos")
library(TeachingDemos)
# Demonstrate concepts of Power.
?power.examp
```

2.7 One-sided tests on μ

There are many studies where a one-sided test is appropriate. The two common scenarios are the **lower one-sided test** $H_0 : \mu = \mu_0$ (or $\mu \geq \mu_0$) versus $H_A : \mu < \mu_0$ and the **upper one-sided test** $H_0 : \mu = \mu_0$ (or $\mu \leq \mu_0$) versus $H_A : \mu > \mu_0$. Regardless of the alternative hypothesis, the tests are based on the t-statistic:

$$t_s = \frac{\bar{Y} - \mu_0}{SE_{\bar{Y}}}.$$

For the **upper one-sided test**

1. Compute the critical value t_{crit} such that the area under the t-curve to the **right** of t_{crit} is the desired size α, that is $t_{\text{crit}} = t_\alpha$.
2. Reject H_0 if and only if $t_s \geq t_{\text{crit}}$.
3. The p-value for the test is the area under the t-curve to the **right** of the test statistic t_s.

The **upper one-sided test** uses the **upper tail** of the t-distribution for a rejection region. The p-value calculation reflects the form of the rejection region. You will reject H_0 only for large positive values of t_s which require $\bar{Y}$ to be significantly greater than μ_0. Does this make sense?

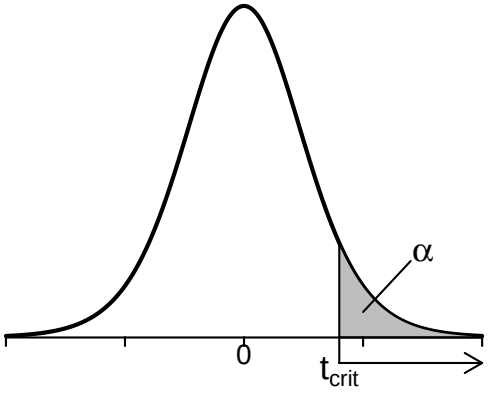
For the **lower one-sided test**

1. Compute the critical value t_{crit} such that the area under the t-curve to the **right** of t_{crit} is the desired size α, that is $t_{\text{crit}} = t_\alpha$.
2. Reject H_0 if and only if $t_s \leq -t_{\text{crit}}$.
3. The p-value for the test is the area under the t-curve to the **left** of the test statistic t_s.

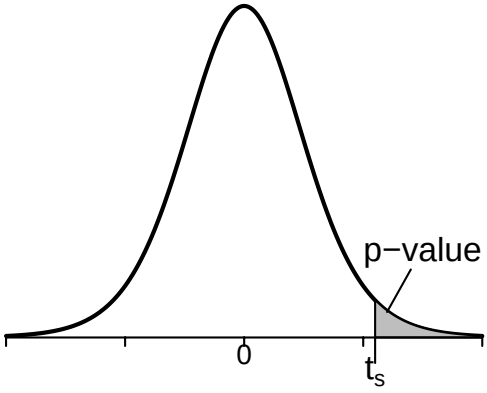
The **lower one-sided test** uses the **lower tail** of the t-distribution for a rejection region. The calculation of the rejection region in terms of $-t_{\text{crit}}$ is awkward but is necessary for hand calculations because many statistical tables only give upper tail percentiles. Note that here you will reject H_0 only for large negative values of t_s which require $\bar{Y}$ to be significantly less than μ_0.

As with two-sided tests, the p-value can be used to decide between rejecting or not rejecting H_0 for a test with a given size α. A picture of the rejection region and the p-value evaluation for one-sided tests is given below.

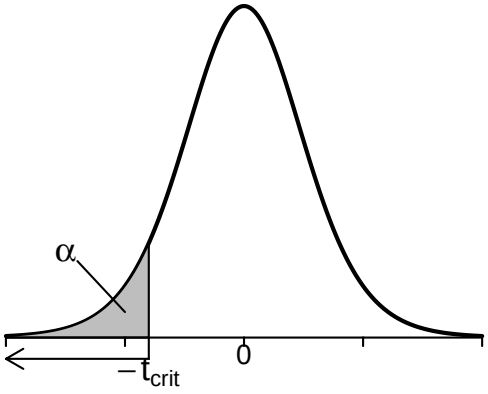
Upper One−Sided Rejection Region



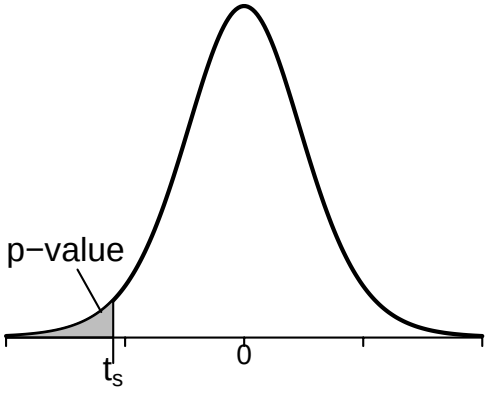
Upper One−Sided p−value



Lower One−Sided Rejection Region



Lower One−Sided p−value



■ CLICKER Qs — One-sided tests on μ ■

Example: Weights of canned tomatoes A consumer group suspects that the average weight of canned tomatoes being produced by a large cannery is less than the advertised weight of 20 ounces. To check their conjecture, the group purchases 14 cans of the canner's tomatoes from various grocery stores. The weights of the contents of the cans to the nearest half ounce were as follows: 20.5, 18.5, 20.0, 19.5, 19.5, 21.0, 17.5, 22.5, 20.0, 19.5, 18.5, 20.0, 18.0, 20.5. Do the data confirm the group's suspicions? Test at the 5% level.

Let μ = the population mean weight for advertised 20 ounce cans of tomatoes produced by the cannery. The company claims that $\mu = 20$, but the consumer group believes that $\mu < 20$. Hence the consumer group wishes to test $H_0 : \mu = 20$ (or $\mu \geq 20$) against $H_A : \mu < 20$. The consumer group will reject H_0 only if the data overwhelmingly suggest that H_0 is false.

You should assess the normality assumption prior to performing the t-test. The stem-and-leaf display and the boxplot suggest that the distribution might be slightly skewed to the left. However, the skewness is not severe and no outliers are present, so the normality assumption is not unreasonable.

R output for the problem is given below. Let us do a hand calculation using the summarized data. The sample size, mean, and standard deviation are 14, 19.679, and 1.295, respectively. The standard error is $SE_{\bar{Y}} = s/\sqrt{n} = 0.346$. We see that the sample mean is less than 20. But is it sufficiently less than 20 for us to be willing to publicly refute the canner's claim? Let us carry out the test, first using the rejection region approach, and then by evaluating a p-value.

The test statistic is

$$t_s = \frac{\bar{Y} - \mu_0}{SE_{\bar{Y}}} = \frac{19.679 - 20}{0.346} = -0.93.$$

The critical value for a 5% one-sided test is $t_{0.05} = 1.771$, so we reject H_0 if $t_s < -1.771$ (you can get that value from r or from the table). The test statistic is not in the rejection region. Using the t-table, the p-value is between

0.15 and 0.20. I will draw a picture to illustrate the critical region and p-value calculation. The exact p-value from R is 0.185, which exceeds 0.05.

Both approaches lead to the conclusion that we do not have sufficient evidence to reject H_0. That is, we do not have sufficient evidence to question the accuracy of the canner's claim. If you did reject H_0, is there something about how the data were recorded that might make you uncomfortable about your conclusions?

```
#### Example: Weights of canned tomatoes
tomato <- c(20.5, 18.5, 20.0, 19.5, 19.5, 21.0, 17.5
          , 22.5, 20.0, 19.5, 18.5, 20.0, 18.0, 20.5)

par(mfrow=c(2,1))
# Histogram overlaid with kernel density curve
hist(tomato, freq = FALSE, breaks = 6)
points(density(tomato), type = "l")
rug(tomato)

# violin plot
library(vioplot)
vioplot(tomato, horizontal=TRUE, col="gray")
```

```
## [1] 17.5 22.5
```

```
# t.crit
qt(1 - 0.05/2, df = length(tomato) - 1)
```

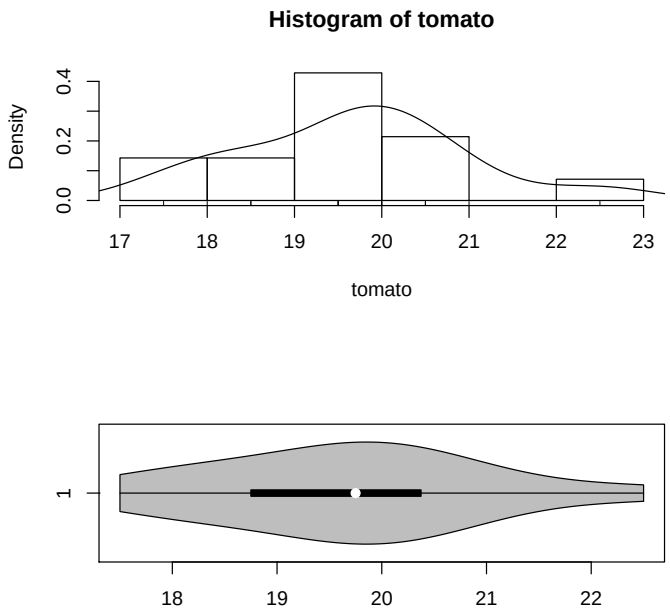
```
## [1] 2.160369
```

```
t.summary <- t.test(tomato, mu = 20, alternative = "less")
t.summary
```

```
##
##  One Sample t-test
##
## data:  tomato
## t = -0.92866, df = 13, p-value = 0.185
## alternative hypothesis: true mean is less than 20
## 95 percent confidence interval:
##      -Inf 20.29153
## sample estimates:
## mean of x
##  19.67857
```

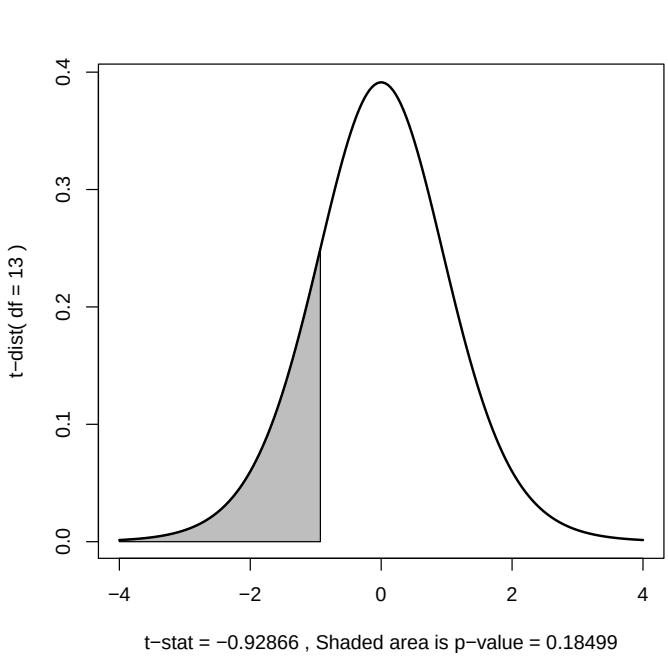
```
summary(tomato)
```

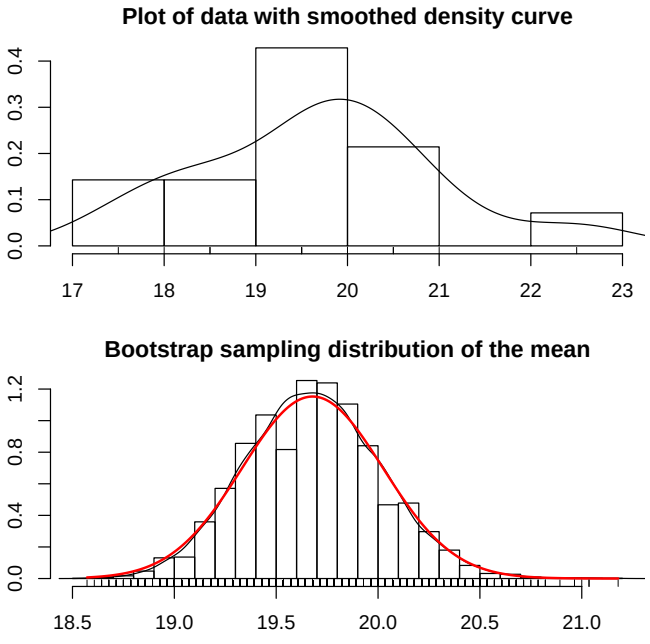
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   17.50   18.75   19.75   19.68   20.38   22.50
```



The assumption of normality of the sampling distribution appears reasonable. Therefore, the results for the t-test above can be trusted.

```
t.dist.pval(t.summary)
bs.one.samp.dist(tomato)
```





2.7.1 One-sided CIs

How should you couple a one-sided test with a CI procedure? For a **lower one-sided test**, you are interested only in an **upper bound** on μ. Similarly, with an **upper one-sided test** you are interested in a **lower bound** on μ. Computing these type of bounds maintains the consistency between tests and CI procedures. The general formulas for lower and upper $100(1-\alpha)\%$ confidence bounds on μ are given by

$$\bar{Y} - t_{\text{crit}} SE_{\bar{Y}} \qquad \text{and} \qquad \bar{Y} + t_{\text{crit}} SE_{\bar{Y}}$$

respectively, where $t_{\text{crit}} = t_\alpha$.

In the cannery problem, to get an upper 95% bound on μ, the critical value is the same as we used for the one-sided 5% test: $t_{0.05} = 1.771$. The upper bound on μ is

$$\bar{Y} + t_{0.05} SE_{\bar{Y}} = 19.679 + 1.771 \times 0.346 = 19.679 + 0.613 = 20.292.$$

Thus, you are 95% confident that the population mean weight of the canner's 20oz cans of tomatoes is less than or equal to 20.29. As expected, this interval covers 20.

If you are doing a one-sided test in R, it will generate the correct one-sided bound. That is, a lower one-sided test will generate an upper bound, whereas an upper one-sided test generates a lower bound. If you only wish to compute a one-sided bound without doing a test, you need to specify the direction of the alternative which gives the type of bound you need. An upper bound was generated by R as part of the test we performed earlier. The result agrees with the hand calculation.

Quite a few packages, do not directly compute one-sided bounds so you have to fudge a bit. In the cannery problem, to get an upper 95% bound on μ, you take the upper limit from a 90% two-sided confidence limit on μ. The rational for this is that with the 90% two-sided CI, μ will fall above the upper limit 5% of the time and fall below the lower limit 5% of the time. Thus, you are 95% confident that μ falls below the upper limit of this interval, which gives us

our one-sided bound. Here, you are 95% confident that the population mean weight of the canner's 20 oz cans of tomatoes is less than or equal to 20.29, which agrees with our hand calculation.

```
One-Sample T: Cans
Variable   N     Mean   StDev  SE Mean         90% CI
Cans      14  19.6786  1.2951   0.3461  (19.0656, 20.2915)
```

The same logic applies if you want to generalize the one-sided confidence bounds to arbitrary confidence levels and to lower one-sided bounds — always double the error rate of the desired one-sided bound to get the error rate of the required two-sided interval! For example, if you want a lower 99% bound on μ (with a 1% error rate), use the lower limit on the 98% two-sided CI for μ (which has a 2% error rate).

■ CLICKER Qs — P-value ■

Part III

Nonparametric, categorical, and regression methods

Part IV

Additional topics