Unit 6
Inference for μ when σ is unknown

Inference When σ is Unknown

Until now, we have made the unrealistic assumption that we know the population standard deviation σ of our variable of interest, in order to more easily explain the reasoning behind our methods.

Now that the framework is in place, we are ready to make the transition to the more realistic situation of unknown population standard deviation.

Inference When σ is Unknown

Recall that Z is a standard normal random variable.

If $X \sim N(\mu, \sigma)$, then

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

Inference When σ is Unknown

What do we mean when we say "the distribution of Z"?

If we were to take every possible sample of size n from the population of X, where $X \sim N(\mu, \sigma)$, and plot

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

for each sample (this is theoretical, as there are infinitely many possible samples), then we would get the standard normal curve.

Inference When σ is Unknown

If σ is unknown, then we estimate it by the sample standard deviation s, and so

$$Z \approx \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

t Distribution

This standardized variable is very important, and we give it its own name. If $X \sim \mathrm{N}(\mu, \sigma)$, then the variable

$$T = \frac{\overline{X} - \mu}{s / \sqrt{n}}$$

follows a ***t* distribution**.

t Distribution

That is, if we were to take every possible sample of size n, and plot

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

for each sample, then we would get a probability density curve that we call the t distribution.

The quantity in the denominator of t is called the **standard error** of the sample mean. The standard error of a random variable is its estimated standard deviation.

t Distribution

The ***t* statistic**

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

has a t distribution with ***n*-1 degrees of freedom**.

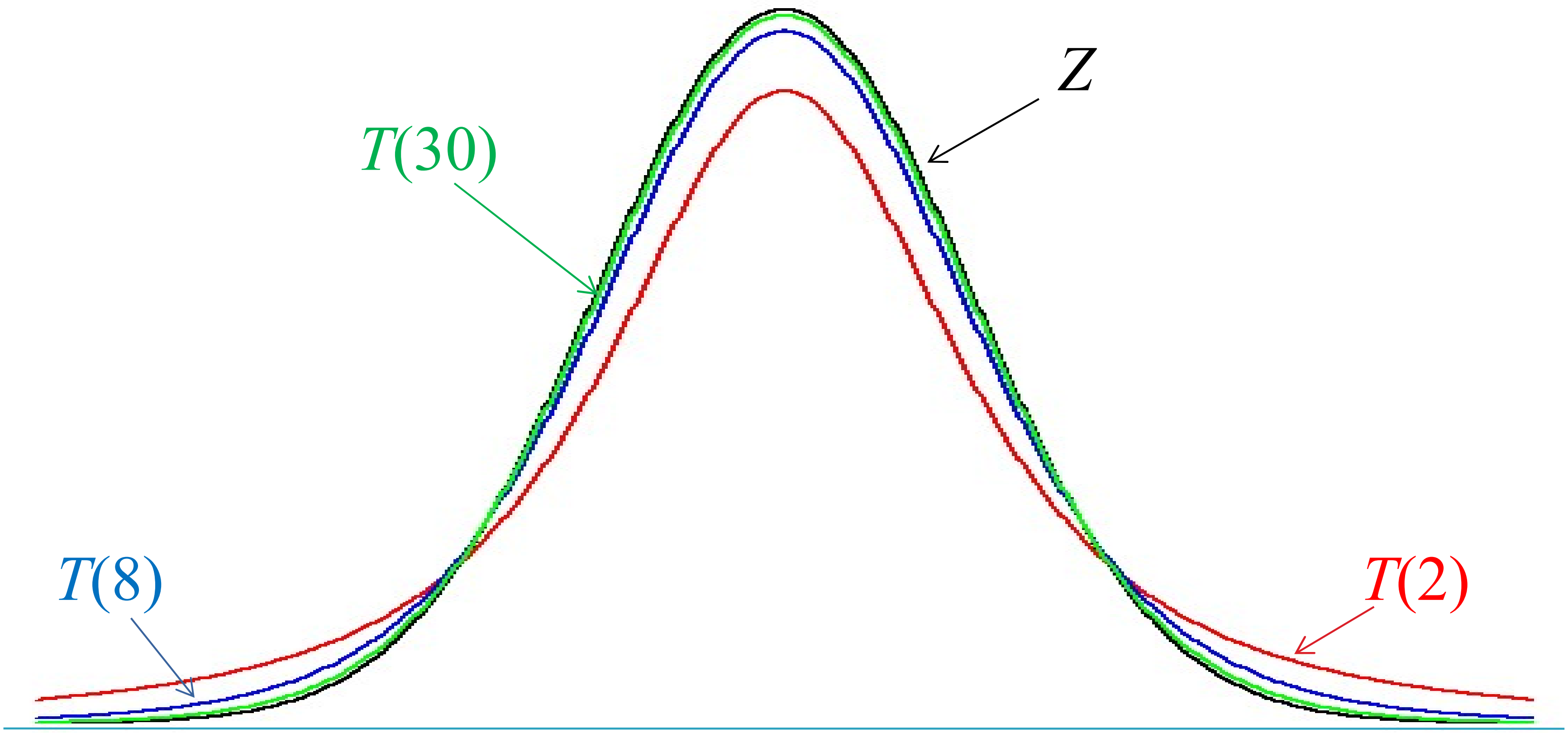
Because the t distributions depend on the degrees of freedom $n - 1$, there is a different t distribution associated with every sample size n. We will write the t distribution with $n - 1$ degrees of freedom as $t(n - 1)$ for short.

T vs. *Z*

Since the form of t and z are quite similar, we expect the shape of the t distribution to be similar to that of the standard normal curve.

In fact, this is the case. The t distribution is symmetric about zero, but the spread for the distribution is **greater** than for the standard normal curve. This is the case because estimating σ by s introduces more variation.

T vs. *Z*



T vs. *Z*

The t distributions have less area near the center and more in the tails than the standard normal distribution.

As the degrees of freedom increase, the t distribution approaches the standard normal distribution.

The critical values for selected t distributions are given in Table 2, for various confidence levels C (located at the bottom) and upper tail probabilities p (located at the top). Instead of looking at the z^* row near the bottom of the table, we go along the row for $n - 1$ degrees of freedom to find our critical values.

Confidence Intervals

We take an SRS of size n from a population having unknown mean μ and unknown standard deviation σ. A level C confidence interval for μ is:

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

where t^* is the upper α/2 critical value of the $t(n\text{-}1)$ distribution. This confidence interval is exact when the population is normal and approximate in other cases when the sample size is large.

Example

The chief of a local police department would like to estimate the true mean response time for all emergency calls in the city. A random sample of seven emergency calls is selected, and the police response times (in minutes) are shown below:

7 4 11 8 7 12 9

We would like to construct a 95% confidence interval for the true mean response time μ of all emergency response calls in the city. We will assume that response times follow a normal distribution.

Example

Some preliminary calculations:

$$n = 7 \qquad \bar{x} = 8.29 \qquad s = 2.69$$

From Table 2, we find the upper 0.025 critical value of the t distribution with $n - 1 = 6$ degrees of freedom to be $t^* = 2.447$. Our 95% confidence interval is

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}} = 8.29 \pm 2.447 \left(\frac{2.69}{\sqrt{7}} \right) = (5.80, 10.78)$$

Z or T?

Do I use Z or T?

Do I know σ?

Yes → Z

No → T

R Code

```
> time <- c(7, 4, 11, 8, 7, 12, 9)
> t.test(time, conf.level = 0.95)
```

```
data:  time
t = 8.1483, df = 6, p-value = 0.0001837
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
  5.797536 10.773892
sample estimates:
mean of x
 8.285714
```

Example

The interpretation of the interval is the same as before. That is, if we repeatedly measured samples of seven response times in this city and constructed an interval in a similar manner, then 95% of such intervals would contain the true mean response time μ.

The convergence of the *t* distribution to the standard normal distribution as the sample size gets high can be seen by comparing the critical values for z^* and t^* for 1000 degrees of freedom.

Practice Question

The *t* distribution with 100 degrees of freedom is closest to:

(A) a normal distribution with mean 0 and s.d. 1.
(B) a normal distribution with mean 1 and s.d. 1.
(C) a normal distribution with mean 0 and s.d. 99.
(D) a normal distribution with mean 0 and s.d. 100.
(E) a *t* distribution with 2 degrees of freedom.

Practice Question

Weights of apples grown in a large orchard are known to follow a normal distribution. The average weight of a random sample of ten apples is calculated to be 140 grams, and the standard deviation is calculated to be 17 grams. Which of the following is a 99% confidence interval for the true mean weight of all apples grown in the orchard?

(A) (122.53, 157.47)
(B) (126.15, 153.85)
(C) (124.44, 155.56)
(D) (123.82, 156.18)
(E) (125.60, 154.40)

σ or s?

If the question is giving you σ:

“The population standard deviation is 17 grams.”

“The standard deviation of weights of all apples in the orchard is 17 grams.”

“Weights follow a normal distribution with standard deviation 17 grams.”

σ or s?

If the question is giving you *s*:

"The sample standard deviation is 17 grams."

"The standard deviation of weights for a sample of ten apples is 17 grams."

"We take a sample of ten apples. The standard deviation of weights for these apples is calculated to be 17 grams."

Practice Question

We would like to estimate the true mean amount of money μ a family spends on groceries per week. For a random sample of nine weeks, the sample mean and standard deviation of the amount spent on groceries are \$287 and \$75, respectively. Weekly grocery expenses for the family are known to follow a normal distribution. What is the margin of error for a 90% confidence interval for μ?

(A) 41.125 (B) 42.375 (C) 44.250

(D) 45.750 (E) 46.500

Hypothesis Tests

We can also use the *t* distributions to conduct tests of significance:

We take a simple random sample of size *n* from a normal distribution with unknown mean μ and unknown standard deviation σ. The null hypothesis H_0: $\mu = \mu_0$ is rejected in favour of the alternative hypothesis H_a when the P-value is less than or equal to the level of significance α.

Example

Consider the previous example. The police chief would like to know if new measures need to be adopted in order to improve response times. He would like to conduct a hypothesis test to determine if the mean response time has increased since the previous year, when the true mean response time was known to be 6.5 minutes.

Example

Step 1

Let $\alpha = 0.10$.

Step 2

We are testing the hypotheses

H_0: The mean response time is the same as last year.
H_a: The mean response time has increased since last year.

Equivalently:

$$H_0\!: \mu = 6.5 \quad \text{vs.} \quad H_a\!: \mu > 6.5$$

Example

<u>Step 3</u>

Reject H_0 if P-value $\leq \alpha = 0.10$.

<u>Step 4</u>

The test statistic is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{8.29 - 6.50}{2.69/\sqrt{7}} = 1.76$$

Example

<u>Step 5</u>

The P-value is $P(T(6) \geq 1.76)$. From Table 2, we see that

$$P(T(6) \geq 1.440) = 0.10 \text{ and } P(T(6) \geq 1.943) = 0.05$$

Since $1.440 < t = 1.76 < 1.943$, it follows that our P-value is between 0.05 and 0.10. For any P-value between 0.05 and 0.10 we would reject the null hypothesis, so our conclusion is as follows…

Example

Step 6

Since the P-value < α = 0.10, we reject the null hypothesis. At a 10% level of significance, we have sufficient evidence to conclude that the department's true mean response has increased since last year. As such, action will be taken to improve response times.

The interpretation of the P-value is the same for *t* tests as it is for *z* tests. If the true mean response time was 6.5 minutes, the probability of observing a sample mean at least as high as 8.29 minutes would be between 0.05 and 0.10.

R Code

```
> time <- c(7, 4, 11, 8, 7, 12, 9)
> t.test(time, mu = 6.50, alternative = "greater")
```

```
data:  time
t = 1.7561, df = 6, p-value = 0.0648
alternative hypothesis: true mean is greater than 6.5
95 percent confidence interval:
 6.309763      Inf
sample estimates:
mean of x
 8.285714
```

Example

We repeat the test using the critical value approach:

Step 1

Let $\alpha = 0.10$.

Step 2

We are testing the hypotheses

H_0: The mean response time is the same as last year.
H_a: The mean response time has increased since last year.

Equivalently:

$$H_0: \mu = 6.5 \quad \text{vs.} \quad H_a: \mu > 6.5$$

Example

Step 3

Reject H_0 if $t \geq t^* = 1.440$.

where $t^* = 1.440$ is the upper 0.10 critical value of the t distribution with $n - 1 = 6$ d.f.

Step 4

The test statistic is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{8.29 - 6.50}{2.69/\sqrt{7}} = 1.76$$

Example

Step 5

Since $t = 1.76 > t^* = 1.440$, we reject the null hypothesis. At a 10% level of significance, we have sufficient evidence to conclude that the department's true mean response time has increased since last year. As such, action will be taken to improve response times.

Example

A fast food restaurant claims that the average waiting time in their drive through is less than a minute. We record the drive through waiting times for a random sample of 31 customers. The sample average time is 56.8 seconds and the sample standard deviation is 17.6 seconds. Suppose waiting times are known to follow a normal distribution.

We will conduct a hypothesis test to examine the significance of the restaurant's claim.

Example

Step 1

Let $\alpha = 0.05$.

Step 2

We are testing the hypotheses

H_0: The true mean waiting time is one minute.
H_a: The true mean waiting time is less than one minute.

Equivalently,

$$H_0: \ \mu = 60 \quad \text{vs.} \quad H_a: \ \mu < 60$$

Example

Notice that we phrase the hypotheses in terms of seconds, since our data were measured in seconds.

Step 3

Reject H_0 if the P-value $\leq \alpha = 0.05$.

Step 4

The test statistic is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{56.8 - 60}{17.6/\sqrt{31}} = -1.01$$

Example

Step 5

The P-value is $P(T(30) \le -1.01) = P(T(30) \ge 1.01)$ by the symmetry of the t distribution. We see from Table 2 that

$$P(T(30) \ge 0.854) = 0.20 \text{ and } P(T(30) \ge 1.055) = 0.15.$$

Since $0.854 < 1.01 < 1.055$, it follows that the P-value is between 0.15 and 0.20. For any P-value between 0.15 and 0.20, we fail to reject the null hypothesis, so our conclusion is as follows…

Example

Step 6

Since the P-value > $\alpha = 0.05$, we fail to reject H_0. At the 5% level of significance, we have insufficient evidence that the true mean waiting time is less than 60 seconds.

Interpretation of P-value:

If the true mean waiting time was 60 seconds, the probability of observing a sample mean at least as low as 56.8 seconds would be between 0.15 and 0.20.

Example

Note that we assumed that wait times in the drive through follow a normal distribution. In reality, wait times are likely skewed to the right.

Nevertheless, we had a high sample size, so we know the sample mean is approximately normally distributed, and so our use of the *t* distribution is justified.

R Code

With our tables, we can only find bounds for the P-value, but we can use R to find the exact P-value:

```
> pt(-1.01, 30)
[1] 0.1757386
```

Example

We conduct the test again using the critical value approach:

Step 1

Let $\alpha = 0.05$.

Step 2

We are testing the hypotheses

H_0: The average waiting time in the drive thru is one minute.
H_a: The average waiting time in the drive thru is less than one minute.

Equivalently,

$$H_0:\ \mu = 60 \ \text{ vs. } \ H_a:\ \mu < 60$$

Example

Step 3

Reject H_0 if $t \leq -t^* = -1.697$.

where $-t^* = -1.697$ is the lower 0.05 critical value of the t distribution with $n - 1 = 30$ d.f.

Step 4

The test statistic is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{56.8 - 60}{17.6/\sqrt{31}} = -1.01$$

Example

Step 5

Since $t = -1.01 > -t^* = -1.697$, we fail to reject the null hypothesis. At a 5% level of significance, we have insufficient evidence to conclude that the true mean waiting time in the drive thru is less than one minute.

Practice Question

Systolic blood pressures of healthy adults follow a normal distribution. We would like to conduct a hypothesis test to determine if the true mean systolic blood pressure of healthy adults is greater than 120. We take a random sample of 25 healthy adults. The sample mean systolic blood pressure is calculated to be 122 and the sample standard deviation is calculated to be 7.8. Conduct an appropriate test at the 5% level of significance.

Practice Question

In testing the hypotheses H_0: $\mu = 2$ vs. H_a: $\mu > 2$, a researcher took a random sample of size $n = 25$, for which it was calculated that $\bar{x} = 3$ and $s = 3$. It is known that the population is normally distributed with standard deviation $\sigma = 2$. The value of the test statistic for the appropriate hypothesis test is:

(A) $t = 2.50$

(B) $z = 3.33$

(C) $t = 7.50$

(D) $z = 2.50$

(E) $t = 3.33$

Practice Question

A scientist is concerned about radiation levels in her laboratory. A room is only considered safe if the true mean radiation level is 425 or less. A random sample of 16 radiation measurements is taken at different locations within the laboratory. These 16 measurements have a mean of 437 and a standard deviation of 20. Radiation levels in the laboratory are known to follow a normal distribution. We conduct a hypothesis test at the 5% level of significance to determine whether there is evidence that the laboratory is unsafe.

Practice Question

What are the hypotheses for the appropriate test of significance?

(A) H_0: $\mu = 425$ vs. H_a: $\mu > 425$
(B) H_0: $\mu = 437$ vs. H_a: $\mu > 437$
(C) H_0: $\mu = 425$ vs. H_a: $\mu \neq 425$
(D) H_0: $\mu = 437$ vs. H_a: $\mu < 437$
(E) H_0: $\mu = 425$ vs. H_a: $\mu < 425$

Practice Question

What is the P-value for the appropriate test of significance?

(A) between 0.005 and 0.01
(B) between 0.01 and 0.02
(C) between 0.02 and 0.025
(D) between 0.025 and 0.05
(E) between 0.05 and 0.10

Practice Question

We conclude that:

(A) there is proof the lab is unsafe.
(B) there is sufficient evidence the lab is safe.
(C) there is sufficient evidence the lab is unsafe.
(D) there is insufficient evidence the lab is safe.
(E) there is insufficient evidence the lab is unsafe.

Example

A machine is designed to fill automobile tires to a mean air pressure of 30 pounds per square inch (psi). The manufacturer tests the machine on a random sample of 25 tires. The sample has a mean air pressure of 30.9 psi and a standard deviation of 1.8 psi. It is known that air pressure follows a normal distribution.

Example

A 95% confidence interval for the true mean fill pressure for this machine is:

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}} = 30.9 \pm 2.064 \left(\frac{1.8}{\sqrt{25}} \right) = (30.16, 31.64)$$

where t^* = 2.064 is the upper 0.025 critical value of the t distribution with 25 – 1 = 24 degrees of freedom.

Example

We interpret the confidence interval as follows:

If we took repeated samples of 25 fill pressures and calculated an interval in a similar manner, then 95% of such intervals would contain the true mean fill pressure for this machine.

We will now conduct a hypothesis test to determine whether the true mean fill pressure for this machine differs from 30 psi.

Example

Step 1

Let $\alpha = 0.05$.

Step 2

H_0: The true mean fill pressure for this machine is 30 psi.

H_a: The true mean fill pressure for this machine differs from 30 psi.

Equivalently,

$$H_0: \mu = 30 \quad \text{vs.} \quad H_a: \mu \neq 30$$

Example

Step 3

Reject H_0 if the P-value $\leq \alpha = 0.05$.

Step 4

The test statistic is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{30.9 - 30}{1.8/\sqrt{25}} = 2.50$$

Example

Step 5

The P-value is $2P(T(24) \geq 2.50)$. We see from Table 2 that

$P(T(24) \geq 2.492) = 0.01$ and $P(T(24) \geq 2.797) = 0.005$.

Since $2.492 < t = 2.50 < 2.797$, it follows that $0.005 < P(T(24) \geq 2.50) < 0.01$, and so the P-value is between $2(0.005) = 0.01$ and $2(0.01) = 0.02$.

For any P-value between 0.01 and 0.02, we reject H_0, and so our conclusion is as follows…

Example

Step 6

Since the P-value < $\alpha = 0.05$, we reject the null hypothesis. At the 5% level of significance, we have sufficient evidence to conclude that the true mean fill pressure for this machine differs from 30 psi.

Note that, since this is a two sided-test using a 5% level of significance, and since we constructed a 95% confidence interval for μ, we could have used that interval to conduct the test. Since $\mu_0 = 30$ is not in the confidence interval, we reject H_0.

R Code

We use R to find the exact P-value:

```
> 2*pt(2.50, 24, lower.tail = FALSE)
[1] 0.01965418
```

Example

We conduct the test again using the critical value approach.

Step 1

Let $\alpha = 0.05$.

Step 2

H_0: The true mean fill pressure for this machine is 30 psi.

H_a: The true fill pressure for this machine differs from 30 psi

Equivalently,

$$H_0: \mu = 30 \quad \text{vs.} \quad H_a: \mu \neq 30$$

Example

Step 3

Reject H_0 if $|t| \geq t^* = 2.064$ (i.e., if $t \leq -2.064$ or $t \geq 2.064$)

where $t^* = 2.064$ is the upper 0.025 critical value of the t distribution with $n - 1 = 24$ d.f.

Step 4

The test statistic is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{30.9 - 30}{1.8/\sqrt{25}} = 2.50$$

Example

Step 5

Since $t = 2.50 > t^* = 2.064$, we reject the null hypothesis. At the 5% level of significance, we have sufficient evidence to conclude that the true mean fill pressure for this machine differs from 30 psi.

Practice Question

The label on bottles of dish soap claims that the bottles contain 350 ml of soap. The amount of soap per bottle is known to follow a normal distribution. We take a random sample of 15 bottles and the sample mean and standard deviation are calculated to be 352.6 ml and 6.5 ml, respectively. Conduct a hypothesis test at the 5% level of significance to determine whether the true mean amount of soap per bottle is different from what the label indicates.

Practice Question

We would like to conduct a hypothesis test to determine if the true mean length of songs on the radio differs from 200 seconds. Based on a random sample of 50 songs on the radio, we calculate a sample mean length of 210 seconds, and a 96% confidence interval for μ is calculated to be (201.1, 218.9). Which of the following statements is true?

(A) At $\alpha = 0.02$, we will fail to reject H_0, since 210 falls inside the 96% confidence interval.

(B) At $\alpha = 0.02$, we will reject H_0, since 200 falls outside the 96% confidence interval.

(C) At $\alpha = 0.04$, we will fail to reject H_0, since 200 falls outside the 96% confidence interval.

(D) At $\alpha = 0.04$, we will fail to reject H_0, since 210 falls inside the 96% confidence interval.

(E) At $\alpha = 0.04$, we will reject H_0, since 200 falls outside the 96% confidence interval.

Summary: P-value for *t* tests

Upper-tailed tests: Go along the row for $n - 1$ d.f. and find the two values that bracket the test statistic t in between. The P-value is between the two probabilities at the top of those rows.

Lower-tailed tests: First, remove the negative from the test statistic. Go along the row for $n - 1$ d.f. and find the two values that bracket the test statistic t in between. The P-value is between the two probabilities at the top of those rows.

Two-sided tests: Follow the same instructions for an upper-tailed test, but double the two probabilities at the top of the appropriate rows.

Summary: Critical Values for *t* tests

Upper-tailed tests: Find the level of significance α in the top row of the table. Go down that column to the row for $n - 1$ d.f. to find t^*. Reject H_0 if $t \geq t^*$.

Lower-tailed tests: Find the level of significance α in the top row of the table. Go down that column to the row for $n - 1$ d.f. to find t^*, and put a negative in front of it. Reject H_0 if $t \leq -t^*$.

Two-sided tests: Find the value of $\alpha/2$ in the top row of the table (or the corresponding confidence level at the bottom). Go down (or up) that column to the row for $n - 1$ d.f. to find t^*. Reject H_0 if $t \leq -t^*$ or $t \geq t^*$.

P-value

Suppose we are conducting a test of

$$H_0 : \mu = \mu_0 \text{ vs. } H_a : \mu > \mu_0$$

We select a sample of size 22 and we get a test statistic of $t = 4.58$. The P-value is $P(T(21) \geq 4.58)$.

We see from Table 2 that $P(T(21) \geq 3.819) = 0.0005$, and that 3.819 is the highest value in this row. As such, since our value is even higher, the P-value must be less than 0.0005.

P-value

Suppose we are conducting a test of

$$H_0 : \mu = \mu_0 \text{ vs. } H_a : \mu > \mu_0$$

We select a sample of size 16 and we get a test statistic of $t = 0.53$. The P-value is $P(T(15) \geq 0.53)$.

We see from Table 2 that $P(T(15) \geq 0.691) = 0.25$, and that 0.691 is the lowest value in this row. As such, since our value is even lower, the P-value must be greater than 0.25.

Cautions About *t* Procedures

Because t is a function of the sample mean and of the sample standard deviation, which are both strongly influenced by outliers, t itself is strongly influenced by outliers.

Although we have an assumption that the population is normally distributed, the t procedures are quite robust against non-normality, especially for large sample sizes.

The most important assumption (except for small samples) is that the data are from an SRS from the population.