Section 7.2

Estimating μ When σ Is Unknown

Estimating μ When σ Is Unknown

In order to use the normal distribution to find confidence intervals for a population mean μ we need to know the value of σ, the population standard deviation.

However, much of the time, when μ is unknown, σ is unknown as well. In such cases, we use the sample standard deviation s to approximate σ.

When we use s to approximate σ, the sampling distribution for $\bar{x}$ follows a new distribution called a *Student's t distribution*.

Student's *t* Distributions

Student's *t* distributions were discovered in 1908 by W. S. Gosset. He was employed as a statistician by Guinness brewing company, a company that discouraged publication of research by its employees.

As a result, Gosset published is research under the pseudonym *Student*. Gosset was the first to recognize the importance of developing statistical methods for obtaining reliable information from samples of populations with unknown σ.

Student's *t* Distributions

Assume that x has a normal distribution with mean μ. For samples of size n with sample mean $\overline{x}$ and sample standard deviation s, the ***t* variable**

$$t = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}} \tag{12}$$

has a **Student's *t* distribution** with **degrees of freedom** $d.f. = n - 1$.

If many random samples of size *n* are drawn, then we get many *t* values from Equation (12).

These *t* values can be organized into a frequency table, and a histogram can be drawn, thereby giving us an idea of the shape of the *t* distribution (for a given *n*).

Student's *t* Distributions

Table 6 of Appendix II gives values of the variable *t* corresponding to what we call the number of *degrees of freedom*, abbreviated *d.f*. For the methods used in this section, the number of degrees of freedom is given by the formula

$$d.f. = n - 1 \tag{13}$$

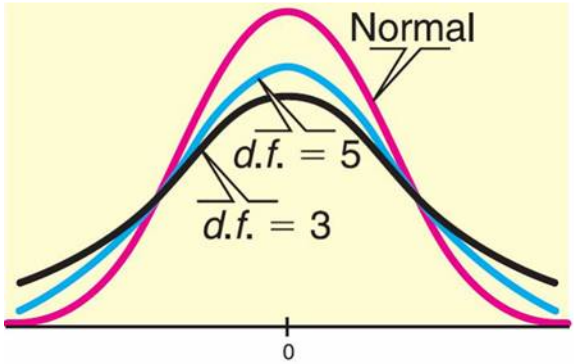
where *d.f*. stands for the degrees of freedom and *n* is the sample size. Each choice for *d.f*. gives a different *t* distribution.

The graph of a *t* distribution is always symmetrical about its mean, which (as for the *z* distribution) is 0.

Student’s *t* Distributions

The main observable difference between a *t* distribution and the standard normal *z* distribution is that a *t* distribution has somewhat thicker tails.

Figure 7-5 shows a standard normal *z* distribution and Student’s *t* distribution with *d.f.* = 3 and *d.f.* = 5.



A Standard Normal Distribution and Student’s *t* Distribution with *d.f.* = 3 and *d.f.* = 5

Figure 7-5

Student’s *t* Distributions

Properties of a Student’s *t* distribution

1. The distribution is *symmetric* about the mean 0.
2. The distribution depends on the *degrees of freedom, d.f.* ($d.f. = n - 1$ for μ confidence intervals).
3. The distribution is *bell-shaped*, but has thicker tails than the standard normal distribution.
4. As the degrees of freedom increase, the *t* distribution *approaches* the standard normal distribution.
5. The area under the entire curve is 1.

Using Table 6 to Find Critical Values for Confidence Intervals

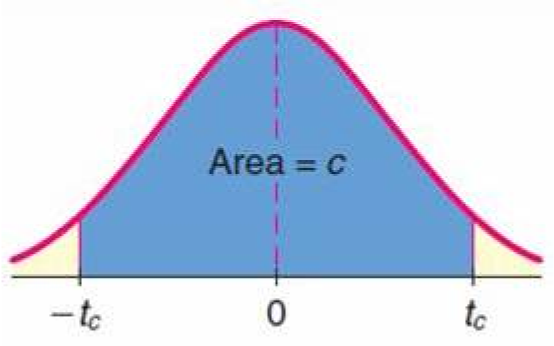
Table 6 of Appendix II gives various t values for different degrees of freedom *d.f*. We will use this table to find *critical values* t_c for a c confidence level.

In other words, we want to find t_c such that an area equal to c under the t distribution for a given number of degrees of freedom falls between $-t_c$ and t_c in the language of probability, we want to find t_c such that

$$P(-t_c \leq t \leq t_c) = c$$

Using Table 6 to Find Critical Values for Confidence Intervals

This probability corresponds to the shaded area in Figure 7-6.



Area Under the t Curve Between $-t_c$ and t_c

Figure 7-6

Table 6 of Appendix II has been arranged so that c is one of the column headings, and the degrees of freedom *d.f*. are the row headings.

Using Table 6 to Find Critical Values for Confidence Intervals

To find t_c for any specific *c*, we find the column headed by that *c* value and read down until we reach the row headed by the appropriate number of degrees of freedom *d.f*.

(You will notice two other column headings: one-tail area and two-tail area. We will use these later, but for the time being, just ignore them.)

Convention for using a Student's *t* distribution table

If the degrees of freedom *d.f.* you need are not in the table, use the closest *d.f.* in the table that is *smaller*. This procedure results in a critical value t_c that is more conservative, in the sense that it is larger. The resulting confidence interval will be longer and have a probability that is slightly higher than *c*.

Example 4 – *Student's t distribution*

Use Table 7-3 (an excerpt from Table 6 of Appendix II) to find the critical value t_c for a 0.99 confidence level for a t distribution with sample size n = 5.

one-tail area	–	–	–	–
two-tail area	–	–	–	–
d.f. \ **c**	**. . . 0.900**	**0.950**	**0.980**	**0.990 . . .**
⋮				
3	. . . 2.353	3.182	4.541	5.841 . . .
4	. . . 2.132	2.776	3.747	4.604 . . .
⋮				
7	. . . 1.895	2.365	2.998	3.449 . . .
8	. . . 1.860	2.306	2.896	3.355 . . .

Student's t Distribution Critical Values (Excerpt from Table 6, Appendix II)

Table 7-3

Example 4 – *Solution*

a. First, we find the column with c heading 0.990.

b. Next, we compute the number of degrees of freedom:
$d.f. = n - 1 = 5 - 1 = 4$

c. We read down the column under the heading $c = 0.99$ until we reach the row headed by 4 (under $d.f.$).
The entry is 4.604. Therefore, $t_{0.99} = 4.604$.

Confidence Intervals for μ When σ Is Unknown

Confidence Intervals for μ When σ Is Unknown

In Section 7.1, we found bounds $\pm E$ on the margin of error for a c confidence level. Using the same basic approach, we arrive at the conclusion that

$$E = t_c \frac{s}{\sqrt{n}}$$

is the maximal margin of error for a c confidence level when σ is unknown (i.e., $|\bar{x} - \mu| < E$ with probability c).

The analogue of Equation (1) is

$$P\left(-t_c \frac{s}{\sqrt{n}} < \bar{x} - \mu < t_c \frac{s}{\sqrt{n}}\right) = c \qquad \textbf{(14)}$$

Confidence Intervals for μ When σ Is Unknown

Procedure:

HOW TO FIND A CONFIDENCE INTERVAL FOR μ WHEN σ IS UNKNOWN

Requirements

Let x be a random variable appropriate to your application. Obtain a simple random sample (of size n) of x values from which you compute the sample mean $\bar{x}$ and the sample standard deviation s.

If you can assume that x has a normal distribution or simply a mound-shaped, symmetric distribution, then any sample size n will work. If you cannot assume this, then use a sample size of $n \geq 30$.

Confidence interval for μ when σ is unknown

$$\bar{x} - E < \mu < \bar{x} + E \qquad (16)$$

where $\bar{x}$ = sample mean of a simple random sample

$$E = t_c \frac{s}{\sqrt{n}}$$

c = confidence level ($0 < c < 1$)

t_c = critical value for confidence level c and degrees of freedom $d.f. = n - 1$ (See Table 6 of Appendix II.)

Example 5 – *Confidence Intervals for μ, σ Unknown*

Suppose an archaeologist discovers seven fossil skeletons from a previously unknown species of miniature horse. Reconstructions of the skeletons of these seven miniature horses show the shoulder heights (in centimeters) to be

45.3 47.1 44.2 46.8 46.5 45.5 47.6

For these sample data, the mean is $\bar{x} \approx 46.14$ and the sample standard deviation $s \approx 1.19$. Let μ be the mean shoulder height (in centimeters) for this entire species of miniature horse, and assume that the population of shoulder heights is approximately normal.

Example 5 – *Confidence Intervals for μ, σ Unknown*

cont'd

Find a 99% confidence interval for μ, the mean shoulder height of the entire population of such horses.

Solution:

Check Requirements We assume that the shoulder heights of the reconstructed skeletons form a random sample of shoulder heights for all the miniature horses of the unknown species.

Example 5 – *Solution*

cont'd

The *x* distribution is assumed to be approximately normal. Since σ is unknown, it is appropriate to use a Student's *t* distribution and sample information to compute a confidence interval for μ.

In this case, $n = 7$, so $d.f. = n - 1 = 7 - 1 = 6$ For $c = 0.999$, Table 6 of Appendix II gives $t_{0.99} = 3.707$ (for *d.f.* 6). The sample standard deviation is $s = 1.19$.

Example 5 – *Solution*

cont'd

$$E = t_c \frac{s}{\sqrt{n}} = (3.707)\,\frac{1.19}{\sqrt{7}} \approx 1.67$$

The 99% confidence interval is

$$\bar{x} - E < \mu < \bar{x} + E$$

$$46.14 - 1.67 < \mu < 46.14 + 1.67$$

$$44.5 < \mu < 47.8$$

Example 5 – *Solution*

cont'd

Interpretation The archaeologist can be 99% confident that the interval from 44.5 cm to 47.8 cm is an interval that contains the population mean μ for shoulder height of this species of miniature horse.

Confidence Intervals for μ When σ Is Unknown

Summary: Confidence intervals for the mean

Assume that you have a random sample of size n from an x distribution and that you have computed $\bar{x}$ and s. A confidence interval for μ is

$$\bar{x} - E < \mu < \bar{x} + E$$

where E is the margin of error. How do you find E? It depends on how much you know about the x distribution.

Confidence Intervals for μ When σ Is Unknown

Situation I (most common)

You don't know the population standard deviation σ. In this situation, you use the t distribution with margin of error

$$E = t_c \frac{s}{\sqrt{n}}$$

where degrees of freedom

$$d.f. = n - 1$$

Although a t distribution can be used in many situations, you need to observe some guidelines. If n is less than 30, x should have a distribution that is mound-shaped and approximately symmetric. It's even better if the x distribution is normal. If n is 30 or more, the central limit theorem implies that these restrictions can be relaxed.

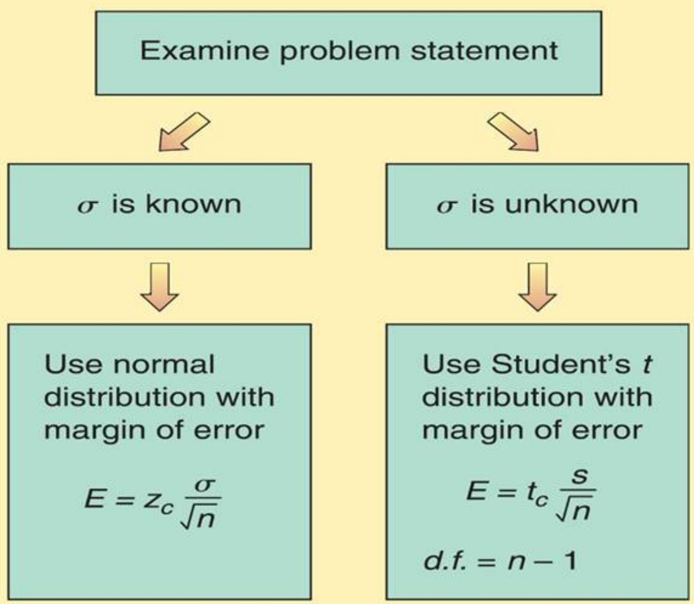
Confidence Intervals for μ When σ Is Unknown

Situation II (almost never happens!)

You actually know the population value of σ. In addition, you know that x has a normal distribution. If you don't know that the x distribution is normal, then your sample size n must be 30 or larger. In this situation, you use the standard normal z distribution with margin of error

$$E = z_c \frac{\sigma}{\sqrt{n}}$$

Which distribution should you use for $\overline{x}$?



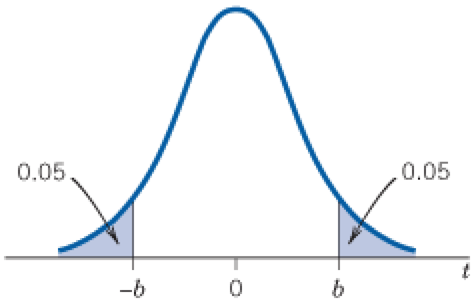
Example:

A random sample of size n=9 from a normal population produced the mean $\bar{X}$= 8.3 and the standard deviation s= 1.2. Obtain 90% confidence interval for μ.

A 90% CI for μ is of the form

$$CI = \left(\bar{x} - ME, \bar{x} + ME\right)$$

$$where,\ ME = t_{\alpha/2} \frac{s}{\sqrt{n}}$$



Here, $t_{0.05} = 1.860$ correspond to 9-1=8 degree of freedom.

$$ME = 1.860 * \frac{1.2}{\sqrt{9}} = 0.744$$

Hence, 95% CI is given by (8.3-0.744, 8.3+0.744)

=(7.556, 9.044)