Section 7.1

Estimating μ When σ Is Known

Estimating μ When σ Is Known

Because of time and money constraints, difficulty in finding population members, and so forth, we usually do not have access to *all* measurements of an *entire* population. Instead we rely on information from a sample.

In this section, we develop techniques for estimating the population mean μ using sample data. We assume the population standard deviation σ is known.

Let's begin by listing some basic assumptions used in the development of our formulas for estimating μ when σ is known.

Estimating μ When σ Is Known

Assumptions about the random variable *x*

1. We have a *simple random sample* of size n drawn from a population of x values.
2. The value of σ, the population standard deviation of x, *is known*.
3. If the x *distribution is normal*, then our methods work for *any sample size* n.
4. If x has an unknown distribution, then we require a *sample size* $n \geq 30$. However, if the x distribution is distinctly skewed and definitely not mound-shaped, a sample of size 50 or even 100 or higher may be necessary.

Estimating μ When σ Is Known

An estimate of a population parameter given by a single number is called a *point estimate* for that parameter. It will come as no great surprise that we use $\bar{x}$ (the sample mean) as the point estimate for μ (the population mean).

A **point estimate** of a population parameter is an estimate of the parameter using a single number.

$\bar{x}$ is the **point estimate** for μ.

Even with a large random sample, the value of $\bar{x}$ usually is not *exactly* equal to the population mean μ. The *margin of error* is the magnitude of the difference between the sample point estimate and the true population parameter value.

Estimating μ When σ Is Known

When using $\overline{x}$ as a point estimate for μ, the **margin of error** is the magnitude of $\overline{x} - \mu$ or $|\overline{x} - \mu|$.

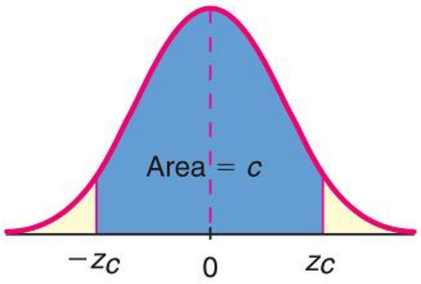
We cannot say exactly how close $\overline{x}$ is to μ when μ is unknown. Therefore, the exact margin of error is unknown when the population parameter is unknown. Of course, μ is usually not known, or there would be no need to estimate it.

In this section, we will use the language of probability to give us an idea of the size of the margin of error when we use $\overline{x}$ as a point estimate for μ.

Estimating μ When σ Is Known

First, we need to learn about *confidence levels*. The reliability of an estimate will be measured by the confidence level.

Suppose we want a confidence level of c (see Figure 7-1). Theoretically, we can choose c to be any value between 0 and 1, but usually c is equal to a number such as 0.90, 0.95, or 0.99.



Confidence Level c and Corresponding Critical Value z_c Shown on the Standard Normal Curve

Figure 7.1

Estimating μ When σ Is Known

In each case, the value z_c is the number such that the area under the standard normal curve falling between $-z_c$ and z_c is equal to c. The value z_c is called the *critical value* for a confidence level of c.

For a confidence level c, the **critical value** z_c is the number such that the area under the standard normal curve between $-z_c$ and z_c equals c.

The area under the normal curve from to $-z_c$ is the probability that the standardized normal variable z lies in that interval. This means that

$$P(-z_c < z < z_c) = c$$

Example 1 – *Find a critical value*

Let us use Table 5 of Appendix II to find a number $z_{0.99}$ such that 99% of the area under the standard normal curve lies between $-z_{0.99}$ and $z_{0.99}$. That is, we will find $z_{0.99}$ such that

$$P(-z_{0.99} < z < z_{0.99}) = 0.99$$

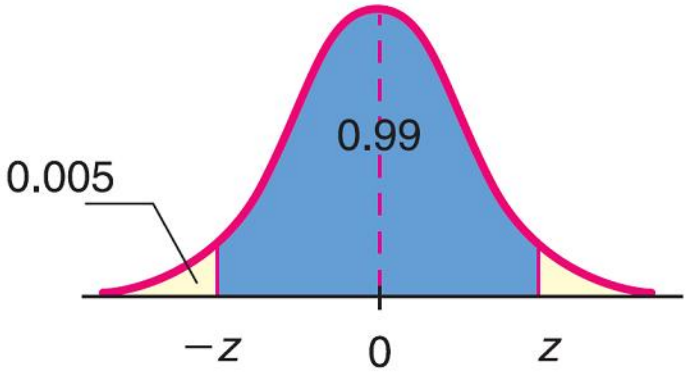
Solution:

We saw how to find the z value when we were given an area between $-z$ and z. The first thing we did was to find the corresponding area to the left of $-z$.

Example 1 – *Solution*

cont'd

If A is the area between $-z$ and z, then $(1 - A)/2$ is the area to the left of z. In our case, the area between $-z$ and z is 0.99.The corresponding area in the left tail is $(1 - 0.99)/2 = 0.005$ (see Figure 7-2).



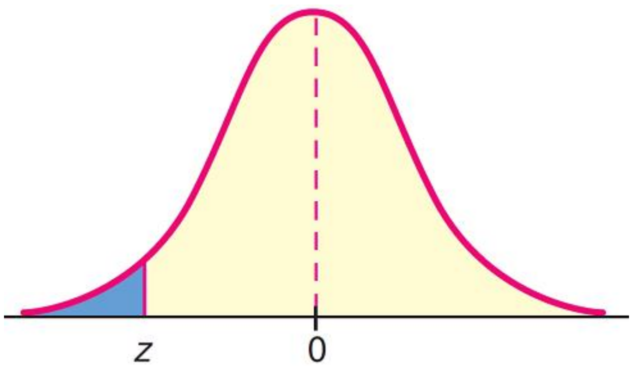
Area Between $-z$ and z Is 0.99

Figure 7-2

Example 1 – *Solution*

cont'd

Next, we use Table 5 of Appendix II to find the z value corresponding to a left-tail area of 0.0050. Table 7-1 shows an excerpt from Table 5 of Appendix II.



z	.00	07	.08	.09
−3.4	.0003	.0003	.0003	.0002
⋮				
−2.5	.0062	.0051	.0049	.0048
		↑ .0050		

Excerpt from Table 5 of Appendix II

Table 7-1

Example 1 – *Solution*

cont'd

From Table 7-1, we see that the desired area, 0.0050, is exactly halfway between the areas corresponding to $z = –2.58$ and $z = –2.57$. Because the two area values are so close together, we use the more conservative z value –2.58 rather than interpolate.

In fact, $z_{0.99} \approx 2.576$. However, to two decimal places, we use $z_{0.99} = 2.58$ as the critical value for a confidence level of $c = 0.99$. We have

$$P(–2.58 < z < 2.58) \approx 0.99$$

Estimating μ When σ Is Known

The *margin of error* (or absolute error) using $\bar{x}$ as a point estimate for μ is $|\bar{x} - \mu|$. In most practical problems, μ is unknown, so the margin of error is also unknown.

However, Equation (1) allows us to compute an *error tolerance E* that serves as a bound on the margin of error. Using a *c*% level of confidence, we can say that the point estimate $\bar{x}$ differs from the population mean μ by a *maximal margin of error*

$$E = z_c \frac{\sigma}{\sqrt{n}} \tag{5}$$

Estimating μ When σ Is Known

Note:

Formula (5) for *E* is based on the fact that the sampling distribution for $\bar{x}$ is exactly normal, with mean μ and standard deviation $\sigma/\sqrt{n}$.

This occurs whenever the *x* distribution is normal with mean μ and standard deviation σ.

If the *x* distribution is not normal, then according to the central limit theorem, large samples ($n \geq 30$) produce an $\bar{x}$ distribution that is approximately normal, with mean μ and standard deviation $\sigma/\sqrt{n}$.

Estimating μ When σ Is Known

Using Equations (1) and (5), we conclude that

$$P(-E < \overline{x} - \mu < E) = c \qquad (6)$$

Equation (6) states that the probability is c that the difference between $\overline{x}$ and μ is no more than the maximal error tolerance E. If we use a little algebra on the inequality

$$-E < \overline{x} - \mu < E \qquad (7)$$

Estimating μ When σ Is Known

for μ, we can rewrite it in the following mathematically equivalent way:

$$\bar{x} - E < \mu < \bar{x} + E \qquad (8)$$

Since formulas (7) and (8) are mathematically equivalent, their probabilities are the same. Therefore, from (6), (7), and (8), we obtain

$$P(\bar{x} - E < \mu < \bar{x} + E) = c \qquad (9)$$

Estimating μ When σ Is Known

Equation (9) states that there is a chance *c* that the interval from $\bar{x} - E$ to $\bar{x} + E$ contains the population mean μ. We call this interval a *c confidence interval for* μ.

A ***c* confidence interval for** μ is an interval computed from sample data in such a way that c is the probability of generating an interval containing the actual value of μ. In other words, c is the proportion of confidence intervals, based on random samples of size n, that actually contain μ.

Estimating μ When σ Is Known

Procedure:

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

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 $\overline{x}$. The value of σ is already known (perhaps from a previous study).

If you can assume that x has a normal 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 known

$$\overline{x} - E < \mu < \overline{x} + E \qquad (10)$$

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

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

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

z_c = critical value for confidence level c based on the standard normal distribution (see Table 5(b) of Appendix II for frequently used values).

Confidence Interval for Mean when population standard deviation is known

1. Population Standard Deviation (σ)-known

If the data has normal distribution with mean and standard deviation σ *given, with*   the mean of the observed values of random variables from that distribution, then 100*(1-)% Confidence interval for mean is given by

$$CI = \left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

18

Standard Error and Margin of Error

A credit company randomly selected 50 contested items and recorded the dollar amount being contested. These contested items had sample mean $\overline{x} = 95.74$ dollars and $s = 24.63$ dollars. Construct a point estimate for the population mean contested amount, μ, and give its 90% error margin.

ANSWER ⊕

SOLUTION ⊖

Estimated mean weakly amount contested is $\overline{x} = \$95.74$.

Estimated standard error is $\frac{s}{\sqrt{n}} = \frac{24.63}{\sqrt{50}} = \$3.483$.

90% error margin is $z_{0.05}\frac{s}{\sqrt{n}} = 1.645\,(3.483) = \$5.730$.

Standard Error and Margin of Error

When estimating μ from a large sample, suppose that one has found the 95% error margin of $\overline{X}$ to be 4.2. From this information, determine:

(a) The estimated S.E. of $\overline{X}$.

ANSWER ⊕

SOLUTION ⊖

Given that the 95% error margin is $1.96\frac{s}{\sqrt{n}} = 4.2$, we know that the estimatedstandard error is $\frac{s}{\sqrt{n}} = \frac{4.2}{1.96} = 2.143$.

(b) The 90% error margin.

ANSWER ⊕

SOLUTION ⊖

90% error margin is $1.645\frac{s}{\sqrt{n}} = 1.645\,(2.143) = 3.525$.

20

Example 2 – *Confidence interval for μ with σ known*

Julia enjoys jogging. She has been jogging over a period of several years, during which time her physical condition has remained constantly good. Usually, she jogs 2 miles per day. The standard deviation of her times is $\sigma = 1.80$ minutes. During the past year, Julia has recorded her times to run 2 miles. She has a random sample of 90 of these times.

For these 90 times, the mean was $\bar{x} = 15.60$ minutes. Let μ be the mean jogging time for the entire distribution of Julia's 2-mile running times (taken over the past year). Find a 0.95 confidence interval for μ.

Example 2 – *Solution*

Check Requirements We have a simple random sample of running times, and the sample size n = 90 is large enough for the $\overline{x}$ distribution to be approximately normal.

We also know σ. It is appropriate to use the normal distribution to compute a confidence interval for μ.

Example 2 – *Solution*

cont'd

To compute E for the 95% confidence interval $\bar{x} - E$ to $\bar{x} + E$, we use the fact that $z_c = 1.96$ (see Table 7-2), together with the values $n = 90$ and $\sigma = 1.80$.

Level of Confidence c	Critical Value z_c
0.70, or 70%	1.04
0.75, or 75%	1.15
0.80, or 80%	1.28
0.85, or 85%	1.44
0.90, or 90%	1.645
0.95, or 95%	1.96
0.98, or 98%	2.33
0.99, or 99%	2.58

Some Levels of Confidence and Their Corresponding Critical Values

Table 7-2

Example 2 – *Solution*

cont' d

Therefore,

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

$$E = 1.96\left(\frac{1.80}{\sqrt{90}}\right)$$

$$E \approx 0.37$$

Using Equation (10), the given value of $\bar{x}$, and our computed value for E, we get the 95% confidence interval for μ.

Example 2 – *Solution*

cont'd

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

$$15.60 - 0.37 < \mu < 15.60 + 0.37$$

$$15.23 < \mu < 15.97$$

Interpretation We conclude with 95% confidence that the interval from 15.23 minutes to 15.97 minutes is one that contains the population mean μ of jogging times for Julia.

Example: Zoology; Hummingbirds

A small group of 15 hummingbirds has been under study. The average weight for these birds is found to be 3.15 g. Based on previous studies, we can assume that the weights have a normal probability distribution with a population standard deviation of 0.33 g.

a)Find an 80% confidence interval for the average weight of these hummingbirds in the study region.

$$CI = \left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

Hence, the 80% confidence interval:

(3.15- 1.28*(0.33/15^0.5), 3.15+1.28*(0.33/15^0.5))

➔(3.04, 3.26)

With 80% confidence, the true average weight of the hummingbirds in the study region is between 3.04g and 3.26g

b) Find a 95% confidence interval for the average weight of these hummingbirds in the study region.

26

Example : SAT Scores

Question:

A random sample of students who have taken the Student Aptitude Test (SAT) is taken from each of the 50 states. The sample mean of the SAT math scores for these 50 states in 2004 is 537.98. Given the population standard deviation is 34 points, construct a 90% confidence interval for the true mean score and interpret its meaning.

Solution:

For a 90% confidence interval, since the population standard deviation is known, we

$$CI = \left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

Hence, the 90% confidence interval:

(537.98-1.645*(34/50^0.5), 537.98+1.645*(34/50^0.5))

➔(530.07, 545.89)

Conclusion: With 90% confidence, the true SAT math scores for the 50 states in 2004 is between 530.07points and 545.89points

Sample Size for Estimating the Mean μ

Sample Size for Estimating the Mean μ

In the design stages of statistical research projects, it is a good idea to decide in advance on the confidence level you wish to use and to select the *maximal* margin of error E you want for your project.

How you choose to make these decisions depends on the requirements of the project and the practical nature of the problem.

Whatever specifications you make, the next step is to determine the sample size. Solving the formula that gives the maximal margin of error E for n enables us to determine the minimal sample size.

Sample Size for Estimating the Mean μ

Procedure:

HOW TO FIND THE SAMPLE SIZE n FOR ESTIMATING μ WHEN σ IS KNOWN

Requirements

The distribution of sample means $\overline{x}$ is approximately normal.

Formula for sample size

$$n = \left(\frac{z_c \sigma}{E}\right)^2 \tag{11}$$

where E = specified maximal margin of error

σ = population standard deviation

z_c = critical value from the normal distribution for the desired confidence level c. Commonly used values of z_c can be found in Table 5(b) of Appendix II.

If n is not a whole number, increase n to the next higher whole number. Note that n is the minimal sample size for a specified confidence level and maximal error of estimate E.

Example 3 – *Sample size for estimating μ*

A wildlife study is designed to find the mean weight of salmon caught by an Alaskan fishing company. A preliminary study of a random sample of 50 salmon showed $s \approx 2.15$ pounds.

How large a sample should be taken to be 99% confident that the sample mean $\bar{x}$ is within 0.20 pound of the true mean weight μ?

Solution:

In this problem, $z_{0.99} = 2.58$ and $E = 0.20$. The preliminary study of 50 fish is large enough to permit a good approximation of σ of by $s = 2.15$.

Example 3 – *Solution*

cont' d

Therefore, Equation (6) becomes

$$n = \left(\frac{z_c \sigma}{E}\right)^2 \approx \left(\frac{(2.58)(2.15)}{0.20}\right)^2 = 769.2$$

Note:

In determining sample size, any fractional value of *n* is always rounded to the *next higher whole number*.

We conclude that a sample size of 770 will be large enough to satisfy the specifications. Of course, a sample size larger than 770 also works.

Example

Determining a Sample Size for Collecting Water Samples

A limnologist wishes to estimate the mean phosphate content per unit volume of lake water. It is known from studies in previous years that the standard deviation has a fairly stable value of $\sigma = 4$. How many water samples must the limnologist analyze to be 90% certain that the error of estimation does not exceed 0.8 milligrams?

SOLUTION

Here $\sigma = 4$ and $1 - \alpha = .90$, so $\alpha / 2 = .05$. The upper .05 point of the $N(0, 1)$ distribution is $z_{.05} = 1.645$. The tolerable error is $d = .8$. Computing

$$n = \left[\frac{1.645 \times 4}{.8} \right]^2 = 67.65$$

we round up to determine that the required sample size is $n = 68$.

33