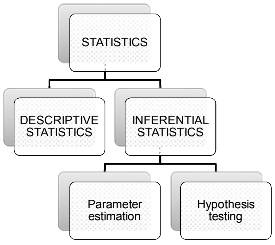
STATISTICS

Confidence Intervals

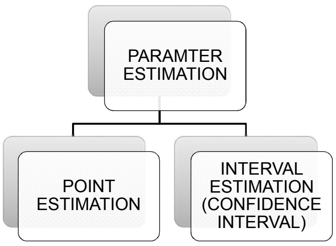
Review

Statistics is the science of collecting, organizing, analyzing, and interpreting data in order to make decisions.



•

Review



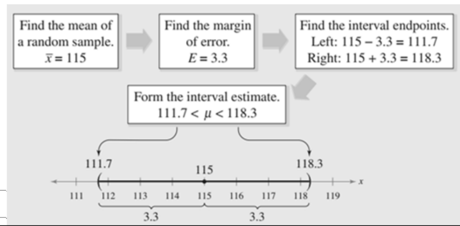
•

Motivating example

For instance, a chess club wants to estimate the mean IQ of its members. The mean of a random sample of members is 115. Because this estimate consists of a single number represented by a point on a number line, **it is called a point estimate.**

The problem with using a point estimate is that it is rarely equal to the exact parameter (mean, standard deviation, or proportion) of the population.

Suppose the club wants to be 90% confident of its estimate for the mean IQ of its members. **Interval Estimate**



•

Section 6.1

Confidence Intervals for the Mean (σ Known)

Point Estimate for Population μ

Point Estimate

- A single value estimate for a population parameter
- Most unbiased point estimate of the population mean μ is the sample mean $\bar{x}$

Estimate Population Parameter…	with Sample Statistic
Mean: μ	$\bar{x}$

Example: Point Estimate for Population μ (1 of 2)

A social networking website allows its users to add friends, send messages, and update their personal profiles. The following represents a random sample of the number of friends for 40 users of the website. Find a point estimate of the population mean μ. *(Adapted from Facebook)*

140	105	130	97	80	165	232	110	214	201	122
98	65	88	154	133	121	82	130	211	153	114
58	77	51	247	236	109	126	132	125	149	122
74	59	218	192	90	117	105				

Example: Point Estimate for Population μ (2 of 2)

The sample mean of the data is

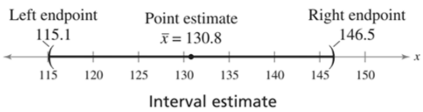
$$\bar{x} = \frac{\sum x}{n} = \frac{5232}{40} = 130.8.$$

the point estimate for the mean number of friends for all users of the website is 130.8 friends

Interval Estimate

Interval estimate

- An interval, or range of values, used to estimate a population parameter.

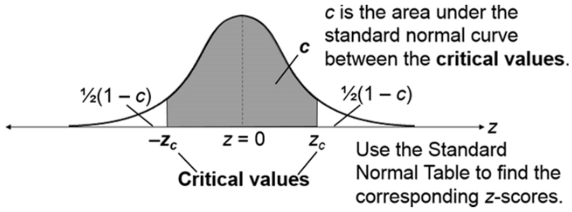


How confident do we want to be that the interval estimate contains the population mean μ?

Level of Confidence (1 of 2)

Level of confidence *c*

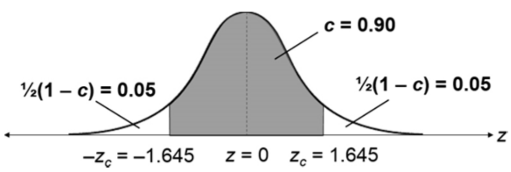
- The probability that the interval estimate contains the population parameter.



The remaining area in the tails is 1 – *c*.

Level of Confidence (2 of 2)

If the level of confidence is 90%, this means that we are 90% confident that the interval contains the population mean μ.



The corresponding *z*-scores are ± 1.645.

Level of confidence

Level of Confidence	z_c
90%	1.645
95%	1.96
99%	2.575

Sampling Error

Sampling error

- The difference between the point estimate and the actual population parameter value.
- For μ:
  - the sampling error is the difference $\bar{x} - \mu$
  - μ is generally unknown
  - $\bar{x}$ varies from sample to sample

Margin of Error

Margin of error

- The greatest possible distance between the point estimate and the value of the parameter it is estimating for a given level of confidence, *c*.
- Denoted by *E*.

$$E = z_c \sigma_{\bar{x}} = z_c \frac{\sigma}{\sqrt{n}}$$

1. The sample is random.
2. Population is normally distributed or $n \geq 30$. Sometimes called the maximum error of estimate or error tolerance.

Example: Finding the Margin of Error (1 of 3)

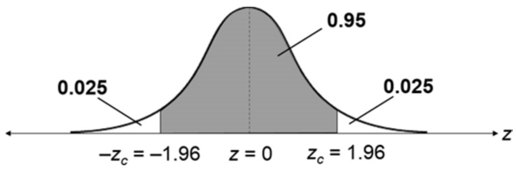
Use the data in example and a 95% confidence level to find the margin of error for the mean number of friends for all users of the website.

Assume that the sample standard deviation is about 53.0

Example: Finding the Margin of Error (2 of 3)

Solution

- First find the critical values



95% of the area under the standard normal curve falls within 1.96 standard deviations of the mean. (You can approximate the distribution of the sample means with a normal curve by the Central Limit Theorem, because $n = 40 \geq 30$.)

Example: Finding the Margin of Error (3 of 3)

Using the values $z_c = 1.96$, $\sigma \approx s \approx 53.0$, and $n = 40$,

$$\begin{aligned} E &= z_c \frac{\sigma}{\sqrt{n}} \\ &\approx 1.96 \cdot \frac{53.0}{\sqrt{40}} \\ &\approx 16.4. \end{aligned}$$

You are 95% confident that the margin of error for the population mean is about 16.4 friends.

Confidence Intervals for the Population Mean

A *c*-confidence interval for the population mean μ

- $\bar{x} - E < \mu < \bar{x} + E$ where $E = z_c \frac{\sigma}{\sqrt{n}}$
- The probability that the confidence interval contains μ is *c*, assuming that the estimation process is repeated a large number of times.

Constructing Confidence Intervals for *μ* (1 of 2)

Finding a Confidence Interval for a Population Mean (σ Known)

In Words	In Symbols
1. Verify that σ known, sample is random, and either the population is normally distributed or $n \geq 30$.	
2. Find the sample statistics n and $\bar{x}$.	$\bar{x} = \frac{\Sigma x}{n}$

Constructing Confidence Intervals for *μ* (2 of 2)

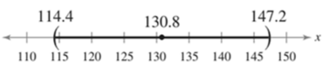
In Words	In Symbols
3. Find the critical value z_c that corresponds to the given level of confidence.	Use Table 4, Appendix B.
4. Find the margin of error E.	$E = z_c \frac{\sigma}{\sqrt{n}}$
5. Find the left and right endpoints and form the confidence interval.	Left endpoint: $\bar{x} - E$ Right endpoint: $\bar{x} + E$ Interval: $\bar{x} - E < \mu < \bar{x} + E$

Example 1: Constructing a Confidence Interval (1 of 2)

Construct a 95% confidence interval for the mean number of friends for all users of the website.

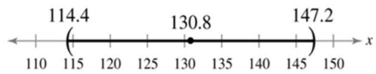
Left Endpoint	Right Endpoint
$\bar{x} - E \approx 130.8 - 16.4 = 114.4$	$\bar{x} + E \approx 130.8 + 16.4 = 147.2$

$$114.4 < \mu < 147.2$$



Example 1: Constructing a Confidence Interval (2 of 2)

14.4 < μ < 147.2



With 95% confidence, you can say that the population mean number of friends is between 114.4 and 147.2.

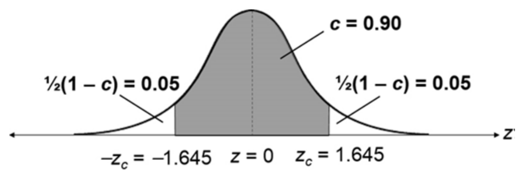
Example 2: Constructing a Confidence Interval (1 of 4)

A college admissions director wishes to estimate the mean age of all students currently enrolled. In a random sample of 20 students, the mean age is found to be 22.9 years. From past studies, the standard deviation is known to be 1.5 years, and the population is normally distributed. Construct a 90% confidence interval of the population mean age.

Example 2: Constructing a Confidence Interval (2 of 4)

Solution

- First find the critical values



z_c = 1.645

Example 2: Constructing a Confidence Interval (3 of 4)

- Margin of error:

$$E = z_c \frac{\sigma}{\sqrt{n}} = 1.645 \cdot \frac{1.5}{\sqrt{20}} \approx 0.6$$

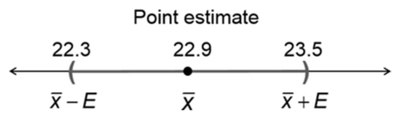
- Confidence interval:

Left Endpoint:	Right Endpoint:
$\bar{x} - E$	$\bar{x} + E$
$\approx 22.9 - 0.6$	$\approx 22.9 + 0.6$
$= 22.3$	$= 23.5$

22.3 < *μ* < 23.5

Example 2: Constructing a Confidence Interval (4 of 4)

22.3 < *μ* < 23.5



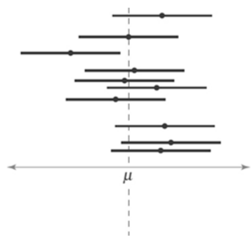
With 90% confidence, you can say that the mean age of all the students is between 22.3 and 23.5 years.

Interpreting the Results (1 of 2)

- *μ* is a fixed number. It is either in the confidence interval or not.
- **Incorrect:** "There is a 90% probability that the actual mean is in the interval (22.3, 23.5)."
- **Correct:** "If a large number of samples is collected and a confidence interval is created for each sample, approximately 90% of these intervals will contain *μ*.

Interpreting the Results (2 of 2)

The horizontal segments represent 90% confidence intervals for different samples of the same size. In the long run, 9 of every 10 such intervals will contain *μ*.



Sample Size

- Given a *c*-confidence level and a margin of error *E*, the minimum sample size *n* needed to estimate the population mean μ is

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

- If σ is unknown, you can estimate it using *s* provided you have a preliminary sample with at least 30 members.

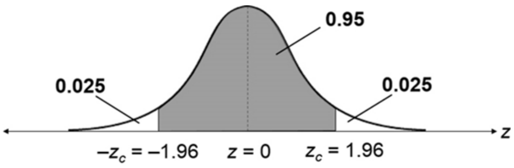
Example: Determining a Minimum Sample Size (1 of 3)

You want to estimate the mean number of friends for all users of the website. How many users must be included in the sample if you want to be 95% confident that the sample mean is within seven friends of the population mean? Assume the sample standard deviation is about 53.0.

Example: Determining a Minimum Sample Size (2 of 3)

Solution

- First find the critical values



$z_c = 1.96$

Example: Determining a Minimum Sample Size (3 of 3)

$z_c = 1.96 \qquad \sigma \approx s = 53.0 \qquad E = 7$

$$n = \left(\frac{z_c \sigma}{E}\right)^2 \approx \left(\frac{1.96 \cdot 53.0}{7}\right)^2 \approx 220.23$$

When necessary, **round up** to obtain a whole number.

You should include **at least 221** users in your sample.

Section 6.2

Confidence Intervals for the Mean $(\sigma \text{ Unknown})$

Slide 35

The *t*-Distribution

- When the population standard deviation is unknown, the sample size is less than 30, and the random variable *x* is approximately normally distributed, it follows a ***t*-distribution**.

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

- Critical values of *t* are denoted by t_c.

Properties of the *t*-Distribution (1 of 3)

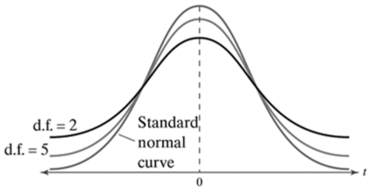
1. The mean, median, and mode of the *t*-distribution are equal to 0.
2. The *t*-distribution is bell shaped and symmetric about the mean.
3. The total area under a *t*-curve is 1.
4. The tails in the *t*-distribution are “thicker” than those in the standard normal distribution.
5. The standard deviation of the *t*-distribution varies with the sample size, but it is greater than 1.

Properties of the *t*-Distribution (2 of 3)

6. The *t*-distribution is a family of curves, each determined by a parameter called the degrees of freedom. The **degrees of freedom** are the number of free choices left after a sample statistic such as $\bar{x}$ is calculated. When you use a *t*-distribution to estimate a population mean, the degrees of freedom are equal to one less than the sample size.
   - $\text{d.f.} = n - 1$ Degrees of freedom

Properties of the *t*-Distribution (3 of 3)

7. As the degrees of freedom increase, the *t*-distribution approaches the normal distribution. For 30 or more degrees of freedom, the *t*-distribution is close to the standard normal distribution.



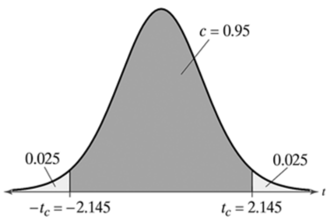
Example: Finding Critical Values of *t*

Find the critical value t_c for a 95% confidence when the sample size is 15.

	Level of confidence, *c*	0.80	0.90	0.95	0.98	0.99
	One tail, α	0.10	0.05	0.025	0.01	0.005
d.f.	Two tails, α	0.20	0.10	0.05	0.02	0.01
1		3.078	6.314	12.706	31.821	63.657
2		1.886	2.920	4.303	6.965	9.925
3		1.638	2.353	3.182	4.541	5.841
12		1.356	1.782	2.179	2.681	3.055
13		1.350	1.771	2.160	2.650	3.012
14		1.345	1.761	2.145	2.624	2.977
15		1.341	1.753	2.131	2.602	2.947
16		1.337	1.746	2.120	2.583	2.921

Solution: Critical Values of *t* (1 of 2)

From the table, you can see that $t_c = 2.145$. The figure shows the *t*-distribution for 14 degrees of freedom, $c = 0.95$, and $t_c = 2.145$.



Constructing a Confidence Interval for a Population Mean (*Sigma* Unknown) (1 of 2)

In Words	*In Symbols*
1. Verify that σ is not known, the sample is random, and the population is normally distributed or $n \geq 30$.	
2. Find the sample statistics *n*, $\bar{x}$, and *s*.	$\bar{x} = \frac{\sum x}{n}$ $s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$

Constructing a Confidence Interval for a Population Mean (*Sigma* Unknown) (2 of 2)

In Words	*In Symbols*
3. Identify the degrees of freedom, the level of confidence *c*, and the critical value t_c.	$\text{d.f.} = n-1$; Use Table 5 in Appendix B.
4. Find the margin of error *E*.	$E = t_c \frac{s}{\sqrt{n}}$
5. Find the left and right endpoints and form the confidence interval.	Left endpoint: $\bar{x} - E$ Right endpoint: $\bar{x} + E$ Interval: $\bar{x} - E < \mu < \bar{x} + E$

Example: Constructing a Confidence Interval (1 of 2)

You randomly select 16 coffee shops and measure the temperature of the coffee sold at each. The sample mean temperature is $162.0°\text{F}$ with a sample standard deviation of $10.0°\text{F}$. Construct a 95% confidence interval for the population mean temperature of coffee sold. Assume the temperatures are approximately normally distributed.

Solution: Constructing a Confidence Interval (1 of 6)

Because σ is unknown, the sample is random, and the temperatures are approximately normally distributed, use the *t*-distribution. Using $n = 16, \bar{x} = 162.0, s = 10.0$, $c = 0.95$, and $\text{d.f.} = 15$, you can use Table 5 to find that $t_c = 2.131$.

The margin of error at the 95% confidence level is

$$E = t_c \frac{s}{\sqrt{n}} = 2.131 \cdot \frac{10}{\sqrt{16}} \approx 5.3$$

Solution: Constructing a Confidence Interval (2 of 6)

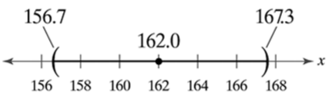
- Confidence interval:

Left Endpoint:	Right Endpoint:
$\bar{x} - E$	$\bar{x} + E$
$\approx 162 - 5.3$	$\approx 162 + 5.3$
$= 156.7$	$= 167.3$

$$\mathbf{156.7 < \mu < 167.3}$$

Solution: Constructing a Confidence Interval (3 of 6)

- $\mathbf{156.7 < \mu < 167.3}$



With 95% confidence, you can say that the mean temperature of coffee sold is between $156.7°\text{F}$ and $167.3°\text{F}$.

Example: Constructing a Confidence Interval (2 of 2)

You randomly select 36 cars of the same model that were sold at a car dealership and determine the number of days each car sat on the dealership's lot before it was sold. The sample mean is 9.75 days, with a sample standard deviation of 2.39 days. Construct a 99% confidence interval for the population mean number of days the car model sits on the dealership's lot.

Solution: Constructing a Confidence Interval (4 of 6)

Solution

Because σ is unknown, the sample is random, and $n = 36 \geq 30$, use the *t*-distribution. Using $n = 36$, $\bar{x} = 9.75, s = 2.39, c = 0.99$, and $\text{d.f.} = 35$, you can use Table 5 to find that $t_c = 2.724$. The margin of error at the 99% confidence level is

$$E = t_c \frac{s}{\sqrt{n}} = 2.724 \cdot \frac{2.39}{\sqrt{36}} \approx 1.09.$$

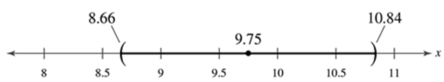
Solution: Constructing a Confidence Interval (5 of 6)

Solution

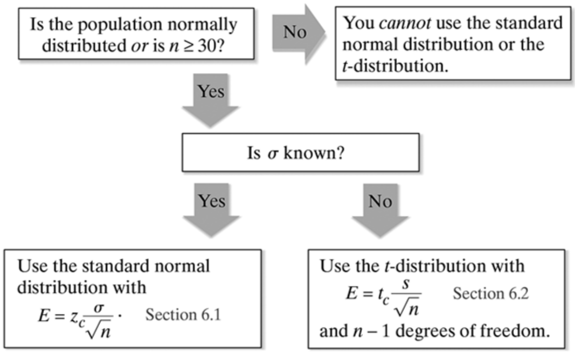
The confidence interval is constructed as shown.

Left Endpoint	Right Endpoint
$\bar{x} - E \approx 9.75 - 1.09$	$\bar{x} + E \approx 9.75 + 1.09$
$= 8.66$	$= 10.84$

$$8.66 < \mu < 10.84$$



Normal or *t*-Distribution?



Example: Normal or *t*-Distribution?

You randomly select 25 newly constructed houses. The sample mean construction cost is $\$181,000$ and the population standard deviation is $\$28,000$. Assuming construction costs are normally distributed, should you use the normal distribution, the *t*-distribution, or neither to construct a 95% confidence interval for the population mean construction cost? Explain your reasoning.

Solution: Normal or *t*-Distribution?

Solution

Is the population normally distributed or is $n \geq 30$?

Yes, the population is normally distributed. Note that even though $n = 25 < 30$ you can still use either the standard normal distribution or the *t*-distribution because the population is normally distributed.

Is σ known?

Yes.

Decision:

Use the standard normal distribution.

Next Lecture

- HYPOTHESIS TESTING WITH ONE SAMPLE