Chapter 7
Estimates and Sample Sizes

Slide 1

Overview

Slide 2

This chapter presents the beginning of inferential statistics.

The two major applications of inferential statistics involve the use of sample data to

1) estimate the value of a population parameter, and

2) test some claim (or hypothesis) about a population.

Overview

Slide 3

This chapter presents the beginning of inferential statistics.

3) We introduce methods for estimating values of these important population parameters: proportions, means, and variances.

4) We also present methods for determining sample sizes necessary to estimate those parameters.

Chapter index

Assumptions

Slide 4

1. The sample is a simple random sample.
2. The conditions for the binomial distribution are satisfied (See Section 5-3.)
3. The normal distribution can be used to approximate the distribution of sample proportions because $np \geq 5$ and $nq \geq 5$ are both satisfied.

Notation for Proportions

Slide 5

p = population proportion

$\hat{p} = \frac{x}{n}$ sample proportion of x successes in a sample of size n

(pronounced 'p-hat')

$\hat{q} = 1 - \hat{p}$ = sample proportion of failures in a sample size of n

Definition

Slide 6

Point Estimate

- A point estimate is a single value (or point) used to approximate a population parameter.

Definition

Slide 7

Point Estimate

- The sample proportion $\hat{p}$ is the best point estimate of the population proportion p.

Slide 8

Example: In the Chapter Problem, we noted that 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Using these survey results, find the best point estimate of the proportion of all adult Minnesotans opposed to photo-cop use.

Because the sample proportion is the best point estimate of the population proportion, we conclude that the best point estimate of p is 0.51. When using the survey results to estimate the percentage of all adult Minnesotans that are opposed to photo-cop use, our best estimate is 51%.

Definition

Slide 9

Confidence Interval

- A **confidence interval** (or interval estimate) is a range (or an interval) of values used to estimate the true value of a population parameter. A confidence interval is sometimes abbreviated as CI.

Definition

Slide 10

Confidence Interval

- A **confidence level** is the probability $1-\alpha$ (often expressed as the equivalent percentage value) that is the proportion of times that the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times.

$$\text{This is usually} \begin{cases} 90\% \longleftrightarrow \alpha = 10\% \\ 95\% \longleftrightarrow \alpha = 5\% \\ 99\% \longleftrightarrow \alpha = 1\% \end{cases}$$

Definition

Slide 11

Confidence Interval

- The confidence level is also called the **degree of confidence**, or the **confidence coefficient**.

Confidence Interval

Slide 12

- Do not use the *overlapping* of confidence intervals as the basis for making final conclusions about the equality of proportions.

Critical Values

Slide 13

- 1. We know from Section 5-6 that under certain conditions, the sampling distribution of sample proportions can be approximated by a normal distribution, as in Figure 6-2.
- 2. Sample proportions have a relatively small chance (with probability denoted by α) of falling in one of the red tails of Figure 6-2.
- 3. Denoting the area of each shaded tail by $\alpha/2$, we see that there is a total probability of α that a sample proportion will fall in either of the two red tails.

Critical Values

Slide 14

- 4. By the rule of complements (from Chapter 3), there is a probability of 1—α that a sample proportion will fall within the inner region of Figure 6-2.
- 5. The z score separating the right-tail is commonly denoted by $z_{\alpha/2}$, and is referred to as a ***critical value*** because it is on the borderline separating sample proportions that are likely to occur from those that are unlikely to occur.

The Critical Value $z_{\alpha/2}$

Slide 15

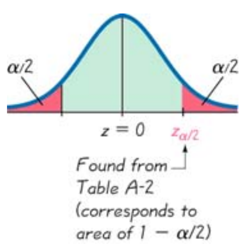


Figure 6-2

Notation for Critical Value

Slide 16

The critical value $z_{\alpha/2}$ is the positive z value that is at the vertical boundary separating an area of $\alpha/2$ in the right tail of the standard normal distribution. (The value of $-z_{\alpha/2}$ is at the vertical boundary for the area of $\alpha/2$ in the left tail). The subscript $\alpha/2$ is simply a reminder that the z score separates an area of $\alpha/2$ in the right tail of the standard normal distribution.

Definition

Slide 17

Critical Value

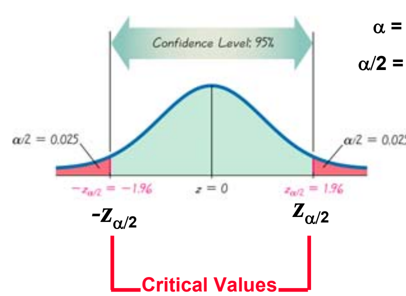
- A **critical value** is the number on the borderline separating sample statistics that are likely to occur from those that are unlikely to occur. The number $z_{\alpha/2}$ is a critical value that is a z score with the property that it separates an area of $\alpha/2$ in the right tail of the standard normal distribution. (See Figure 6-2).

Finding $z_{\alpha/2}$ for 95% Degree of Confidence

Slide 18

α = 5%

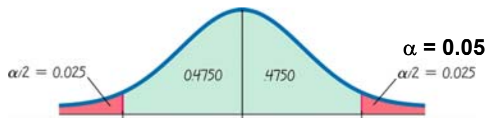
$\alpha/2$ = 2.5% = .025



Finding $z_{\alpha/2}$ for 95% Degree of Confidence

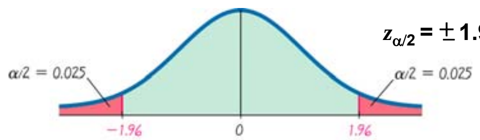
Slide 19

α = 0.05



Use Table A-2
to find a z score of 1.96

$z_{\alpha/2} = \pm 1.96$



Definition

Slide 20

When data from a simple random sample are used to estimate a population proportion *p*, the margin of error, denoted by *E*, is the maximum likely (with probability 1 – α) difference between the observed proportion $\hat{p}$ and the true value of the population proportion *p*.

Margin of Error of the Estimate of *p*

Slide 21

Formula 6-1

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

Confidence Interval for Population Proportion

Slide 22

$$\hat{p} - E < p < \hat{p} + E$$

where

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

Confidence Interval for Population Proportion

Slide 23

$$\hat{p} - E < p < \hat{p} + E$$

$$\hat{p} \pm E$$

$$(\hat{p} - E, \hat{p} + E)$$

Round-Off Rule for Confidence Interval Estimates of *p*

Slide 24

Round the confidence interval limits to

three significant digits.

Procedure for Constructing a Confidence Interval for *p*

Slide 25

- 1. Verify that the required assumptions are satisfied. (The sample is a simple random sample, the conditions for the binomial distribution are satisfied, and the normal distribution can be used to approximate the distribution of sample proportions because $np \geq 5$, and $nq \geq 5$ are both satisfied).
- 2. Refer to Table A-2 and find the critical value $z_{\alpha/2}$ that corresponds to the desired confidence level.
- 3. Evaluate the margin of error $E = \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Procedure for Constructing a Confidence Interval for *p*

Slide 26

- 4. Using the calculated margin of error, E and the value of the sample proportion, $\hat{p}$, find the values of $\hat{p} - E$ and $\hat{p} + E$. Substitute those values in the general format for the confidence interval:

$$\hat{p} - E < p < \hat{p} + E$$

- 5. Round the resulting confidence interval limits to three significant digits.

Slide 27

Example: In the Chapter Problem, we noted that 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

a) Find the margin of error E that corresponds to a 95% confidence level.

b) Find the 95% confidence interval estimate of the population proportion p.

c) Based on the results, can we safely conclude that the majority of adult Minnesotans oppose use the the photo-cop?

Slide 28

Example: In the Chapter Problem, we noted that 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

a) Find the margin of error E that corresponds to a 95% confidence level

First, we check for assumptions. We note that $n\hat{p} = 422.79 \geq 5$, and $n\hat{q} = 406.21 \geq 5$.

Next, we calculate the margin of error. We have found that $\hat{p} = 0.51$, $\hat{q} = 1 - 0.51 = 0.49$, $z_{\alpha/2} = 1.96$, and $n = 829$.

$$E = 1.96\sqrt{\frac{(0.51)(0.49)}{829}}$$

$$E = 0.03403$$

Slide 29

Example: In the Chapter Problem, we noted that 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

b) Find the 95% confidence interval for the population proportion p.

We substitute our values from Part a to obtain:

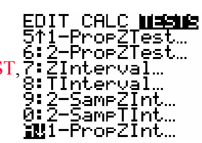
$0.51 - 0.03403 < p < 0.51 + 0.03403$,

$0.476 < p < 0.544$

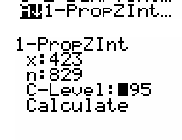
Slide 30

Using TI Calculator: One Population, **Proportion** Confidence Interval

1) Calculate x for the last example, then go to STAT, TEST, then select 1-PropZint.

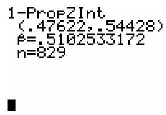


2) Enter the values for x, n and C-level, then select Calculate

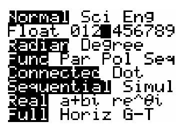


Using TI Calculator:
One Population,
Proportion Confidence Interval

3) Confidence Interval and p-hat is displayed now.



4) You may adjust the number of decimals by selecting mode, float, and then 3 followed by enter.



Example: In the Chapter Problem, we noted that 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

c) Based on the results, can we safely conclude that the majority of adult Minnesotans oppose use of the photo-cop?

Based on the survey results, we are 95% confident that the limits of 47.6% and 54.4% contain the true percentage of adult Minnesotans opposed to the photo-cop. The percentage of opposed adult Minnesotans is likely to be any value between 47.6% and 54.4%. However, a majority requires a percentage greater than 50%, so we *cannot* safely conclude that the majority is opposed (because the *entire* confidence interval is not greater than 50%).

Determining Sample Size

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

(solve for n by algebra)

$$n = \frac{(z_{\alpha/2})^2\,\hat{p}\hat{q}}{E^2}$$

Sample Size for Estimating Proportion *p*

When an estimate of $\hat{p}$ is known:

$$n = \frac{(z_{\alpha/2})^2\,\hat{p}\hat{q}}{E^2}$$ Formula 6-2

When no estimate of p is known:

$$n = \frac{(z_{\alpha/2})^2\,0.25}{E^2}$$ Formula 6-3

Example: Suppose a sociologist wants to determine the current percentage of U.S. households using e-mail. How many households must be surveyed in order to be 95% confident that the sample percentage is in error by no more than four percentage points?

a) Use this result from an earlier study: In 1997, 16.9% of U.S. households used e-mail (based on data from *The World Almanac and Book of Facts*).

$$n = \frac{[z_{\alpha/2}]^2\,\hat{p}\hat{q}}{E^2}$$

$$= \frac{[1.96]^2\,(0.169)(0.831)}{0.04^2}$$

= 337.194

= 338 households

To be 95% confident that our sample percentage is within four percentage points of the true percentage for all households, we should randomly select and survey 338 households.

Example: Suppose a sociologist wants to determine the current percentage of U.S. households using e-mail. How many households must be surveyed in order to be 95% confident that the sample percentage is in error by no more than four percentage points?

b) Assume that we have no prior information suggesting a possible value of $\hat{p}$.

$$n = \frac{[z_{\alpha/2}]^2 \cdot 0.25}{E^2}$$

$$= \frac{(1.96)^2\,(0.25)}{0.04^2}$$

= 600.25

= 601 households

With no prior information, we need a larger sample to achieve the same results with 95% confidence and an error of no more than 4%.

Finding the Point Estimate and E from a Confidence Interval

Slide 37

Point estimate of $\hat{p}$:

$$\hat{p} = \frac{\text{upper confidence limit} + \text{lower confidence limit}}{2}$$

Margin of Error:

$$E = \frac{\text{upper confidence limit} - \text{lower confidence limit}}{2}$$

Chapter index

Assumptions

Slide 38

1. The sample is a simple random sample.
2. The value of the population standard deviation σ is known.
3. Either or both of these conditions is satisfied: The population is normally distributed or $n > 30$.

Definitions

Slide 39

- **Estimator**

  is a formula or process for using sample data to estimate a population parameter.
- **Estimate**

  is a specific value or range of values used to approximate a population parameter.
- **Point Estimate**

  is a single value (or point) used to approximate a population parameter.

The sample mean $\bar{x}$ is the best point estimate of the population mean μ.

Sample Mean

Slide 40

1. For many populations, the distribution of sample means $\bar{x}$ tends to be more consistent (with *less variation*) than the distributions of other sample statistics.
2. For all populations, the sample mean $\bar{x}$ is an unbiased estimator of the population mean μ, meaning that the distribution of sample means tends to center about the value of the population mean μ.

Slide 41

Example: A study found the body temperatures of 106 healthy adults. The sample mean was 98.2 degrees and the sample standard deviation was 0.62 degrees. Find the point estimate of the population mean μ of all body temperatures.

Because the sample mean $\bar{x}$ is the best point estimate of the population mean μ, we conclude that the best point estimate of the population mean μ of all body temperatures is 98.20° F.

Definition
Confidence Interval

Slide 42

As we saw in Section 6-2, a confidence interval is a range (or an interval) of values used to estimate the true value of the population parameter. The confidence level gives us the success rate of the procedure used to construct the confidence interval.

Definition
Level of Confidence

Slide 43

As described in Section 6-2, the confidence level is often expressed as probability 1 - α, where α is the complement of the confidence level.

For a 0.95(95%) confidence level, $\alpha = 0.05$.

For a 0.99(99%) confidence level, $\alpha = 0.01$.

Definition
Margin of Error

Slide 44

is the maximum likely difference observed between sample mean $\bar{x}$ and population mean μ,
and is denoted by *E*.

Definition
Margin of Error

Slide 45

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Formula 6-4

Confidence Interval (or Interval Estimate) for Population Mean μ when σ is known

Slide 46

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

$$\bar{x} \pm E$$

$$(\bar{x} - E, \bar{x} + E)$$

Procedure for Constructing a Confidence Interval for μ when σ is known

Slide 47

1. Verify that the required assumptions are met.
2. Find the critical value $z_{\alpha/2}$ that corresponds to the desired degree of confidence.
3. Evaluate the margin of error $E = z_{\alpha/2} \cdot \sigma / \sqrt{n}$.
4. Find the values of $\bar{x} - E$ and $\bar{x} + E$. Substitute those values in the general format of the confidence interval:

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

5. Round using the confidence intervals roundoff rules.

Round-Off Rule for Confidence Intervals Used to Estimate μ

Slide 48

1. When using the original set of data, round the confidence interval limits to one more decimal place than used in original set of data.

2. When the original set of data is unknown and only the summary statistics $\{n, \bar{x}, s\}$ are used, round the confidence interval limits to the same number of decimal places used for the sample mean.

Slide 49

Example: **A study found the body temperatures of 106 healthy adults. The sample mean was 98.2 degrees and the sample standard deviation was 0.62 degrees. Find the margin of error *E* and the 95% confidence interval for μ.**

$n = 106$
$\bar{x} = 98.20^\circ$
$s = 0.62^\circ$

$\alpha = 0.05$
$\alpha/2 = 0.025$
$z_{\alpha/2} = 1.96$

$$E = z_{\alpha/2} \bullet \frac{\sigma}{\sqrt{n}} = 1.96 \bullet \frac{0.62}{\sqrt{106}} = 0.12$$

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

$$98.20^\circ - 0.12 < \mu < 98.20^\circ + 0.12$$

$$98.08^\circ < \mu < 98.32^\circ$$

Slide 50

Example: **A study found the body temperatures of 106 healthy adults. The sample mean was 98.2 degrees and the sample standard deviation was 0.62 degrees. Find the margin of error *E* and the 95% confidence interval for μ.**

$n = 106$
$\bar{x} = 98.20^\circ$
$s = 0.62^\circ$

$\alpha = 0.05$
$\alpha/2 = 0.025$
$z_{\alpha/2} = 1.96$

$$E = z_{\alpha/2} \bullet \frac{\sigma}{\sqrt{n}} = 1.96 \bullet \frac{0.62}{\sqrt{106}} = 0.12$$

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

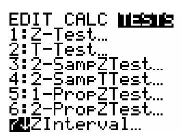
$$98.08^\circ < \mu < 98.32^\circ$$

Based on the sample provided, the confidence interval for the population mean is $98.08^\circ < \mu < 98.32^\circ$. If we were to select many different samples of the same size, 95% of the confidence intervals would actually contain the population mean μ.

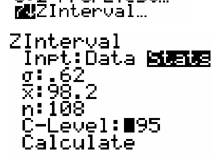
Slide 51

Using TI Calculator:
One Population, Large Sample or σ known
Mean Confidence Interval

1) Select STAT, TESTS, ZInterval.



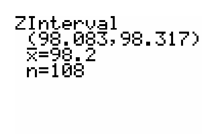
2) Choose Stats, enter values for Population standard deviation, sample mean and sample size, C-Level, then select calculate.



Slide 52

Using TI Calculator:
One Population, Large Sample or σ known
Mean Confidence Interval

3) Confidence Interval is displayed now.



Slide 53

Sample Size for Estimating Mean μ

$$n = \left[\frac{(z_{\alpha/2}) \bullet \sigma}{E}\right]^2$$

Formula 6-5

Slide 54

Round-Off Rule for Sample Size *n*

When finding the sample size *n*, if the use of Formula 6-5 does not result in a whole number, always *increase* the value of *n* to the next *larger* whole number.

Finding the Sample Size n when σ is unknown

Slide 55

1. Use the range rule of thumb (see Section 2-5) to estimate the standard deviation as follows: $\sigma \approx$ range/4.
2. Conduct a pilot study by starting the sampling process. Based on the first collection of at least 31 randomly selected sample values, calculate the sample standard deviation s and use it in place of σ.
3. Estimate the value of σ by using the results of some other study that was done earlier.

Slide 56

Example: Assume that we want to estimate the mean IQ score for the population of statistics professors. How many statistics professors must be randomly selected for IQ tests if we want 95% confidence that the sample mean is within 2 IQ points of the population mean? Assume that σ = 15, as is found in the general population.

$\alpha = 0.05$
$\alpha/2 = 0.025$
$z_{\alpha/2} = 1.96$
$E = 2$
$\sigma = 15$

$$n = \left[\frac{1.96 \bullet 15}{2}\right]^2 = 216.09 = 217$$

With a simple random sample of only 217 statistics professors, we will be 95% confident that the sample mean will be within 2 points of the true population mean μ.

Chapter index

σ Not Known Assumptions

Slide 57

1) The sample is a simple random sample.
2) Either the sample is from a normally distributed population, or $n > 30$.

Use Student t distribution

Student t Distribution

Slide 58

If the distribution of a population is essentially normal, then the distribution of

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

- is essentially a Student t Distribution for all samples of size n, and is used to find critical values denoted by $t_{\alpha/2}$.

Definition

Slide 59

Degrees of Freedom (df)

corresponds to the number of sample values that can vary after certain restrictions have been imposed on all data values

$$df = n - 1$$

in this section.

Margin of Error E for Estimate of μ

Slide 60

Based on an Unknown σ and a Small Simple Random Sample from a Normally Distributed Population

Formula 6-6

$$E = t_{\alpha/2} \frac{s}{\sqrt{n}}$$

where $t_{\alpha/2}$ has n – 1 degrees of freedom.

Confidence Interval for the Estimate of E

Slide 61

Based on an Unknown σ and a Small Simple Random Sample from a Normally Distributed Population

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

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

$t_{\alpha/2}$ found in Table A-3

Procedure for Constructing a Confidence Interval for μ when σ is not known

Slide 62

1. Verify that the required assumptions are met.
2. Using $n - 1$ degrees of freedom, refer to Table A-3 and find the critical value $t_{\alpha/2}$ that corresponds to the desired degree of confidence.
3. Evaluate the margin of error E = $t_{\alpha/2} \cdot s / \sqrt{n}$.
4. Find the values of $\bar{x}$ - E and $\bar{x}$ + E. Substitute those values in the general format for the confidence interval:

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

5. Round the resulting confidence interval limits.

Slide 63

Example: A study found the body temperatures of 106 healthy adults. The sample mean was 98.2 degrees and the sample standard deviation was 0.62 degrees. Find the margin of error E and the 95% confidence interval for µ.

$n = 106$
$\bar{x} = 98.20°$
$s = 0.62°$

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 1.984 \cdot \frac{0.62}{\sqrt{106}} = 0.1195$$

$\alpha = 0.05$
$\alpha/2 = 0.025$
$t_{\alpha/2} = 1.96$

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

$$98.08° < \mu < 98.32°$$

Based on the sample provided, the confidence interval for the population mean is $98.08° < \mu < 98.32°$. The interval is the same here as in Section 6-2, but in some other cases, the difference would be much greater.

Important Properties of the Student *t* Distribution

Slide 64

1. The Student *t* distribution is different for different sample sizes (see Figure 6-5 for the cases $n = 3$ and $n = 12$).
2. The Student *t* distribution has the same general symmetric bell shape as the normal distribution but it reflects the greater variability (with wider distributions) that is expected with small samples.
3. The Student *t* distribution has a mean of $t = 0$ (just as the standard normal distribution has a mean of $z = 0$).
4. The standard deviation of the Student *t* distribution varies with the sample size and is greater than 1 (unlike the standard normal distribution, which has a $\sigma = 1$).
5. As the sample size n gets larger, the Student *t* distribution gets closer to the normal distribution.

Student *t* Distributions for n = 3 and n = 12

Slide 65

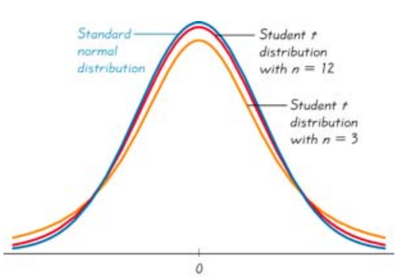


Figure 6-5

Using the Normal and *t* Distribution

Slide 66

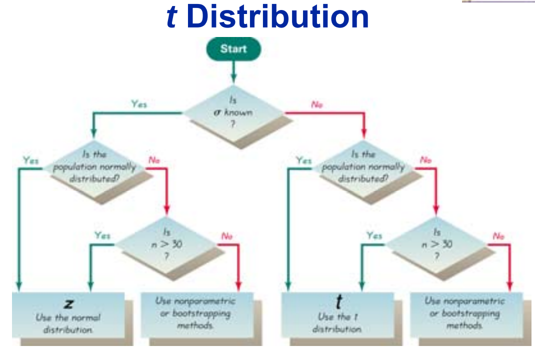


Figure 6-6

Slide 67

Example: Data Set 14 in Appendix B includes the Flesch ease of reading scores for 12 different pages randomly selected from J.K. Rowling's *Harry Potter and the Sorcerer's Stone*. Find the 95% interval estimate of μ, the mean Flesch ease of reading score. (The 12 pages' distribution appears to be bell-shaped.)

$\bar{x} = 80.75$

$s = 4.68$

$$E = t_{\alpha/2} \frac{s}{\sqrt{n}} = \frac{(2.201)(4.68)}{\sqrt{12}} = 2.97355$$

$\alpha = 0.05$

$\alpha/2 = 0.025$

$t_{\alpha/2} = 2.201$

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

$$80.75 - 2.97355 < \mu < 80.75 + 2.97355$$

$$77.77645 < \mu < 83.72355$$

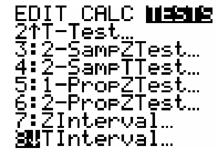
$$77.78 < \mu < 83.72$$

We are 95% confident that this interval contains the mean Flesch ease of reading score for all pages.

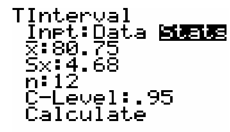
Slide 68

Using TI Calculator: One Population, σ unknown Mean Confidence Interval

1) Select STAT, TESTS, TInterval.



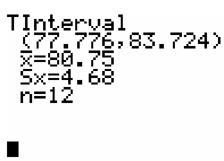
2) Select Stats, enter values for sample mean, standard deviation, size, and C-Level, then select Calculate.



Slide 69

Using TI Calculator: One Population, σ Unknown Mean Confidence Interval

3) Confidence Interval is displayed now.



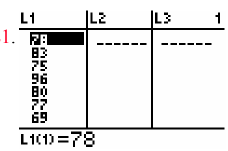
Slide 70

Test scores of 15 students are given:

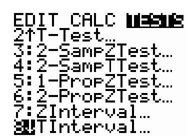
78, 83, 75, 96, 80, 77, 69, 90, 85, 70, 68, 83, 83, 70, 100

Assume test scores are randomly distributed, find a 90% confidence interval for the mean of all test scores.

1) Enter the test scores into L1.



2) Select STAT, TESTS, TInterval,



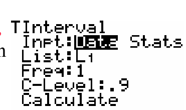
Slide 71

Test scores of 15 students are given:

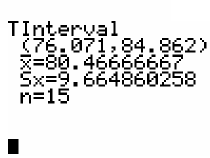
78, 83, 75, 96, 80, 77, 69, 90, 85, 70, 68, 83, 83, 70, 100

Assume test scores are randomly distributed, find a 90% confidence interval for the mean of all test scores.

3) Select Data, enter L1 for List, 1 for Freq:, 0.9 for C-Level, then select Calculate.



4) Confidence Interval, sample mean, and standard deviation are now displayed



Slide 72

Finding the Point Estimate and E from a Confidence Interval

Point estimate of μ:

$$\bar{x} = \frac{\text{(upper confidence limit)} + \text{(lower confidence limit)}}{2}$$

Margin of Error:

$$E = \frac{\text{(upper confidence limit)} - \text{(lower confidence limit)}}{2}$$

Chapter index

Assumptions

Slide 73

1. The sample is a simple random sample.
2. The population must have normally distributed values (even if the sample is large).

Chi-Square Distribution

Slide 74

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$ Formula 6-7

where

n = sample size

s^2 = sample variance

σ^2 = population variance

Properties of the Distribution of the Chi-Square Statistic

Slide 75

1. The chi-square distribution is not symmetric, unlike the normal and Student t distributions.

   As the number of degrees of freedom increases, the distribution becomes more symmetric. (continued)

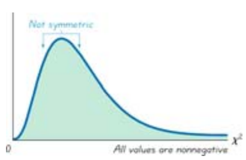


Figure 6-8 Chi-Square Distribution

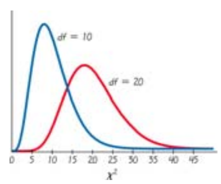


Figure 6-9 Chi-Square Distribution for *df* = 10 and *df* = 20

Properties of the Distribution of the Chi-Square Statistic

Slide 76

(continued)

2. The values of chi-square can be zero or positive, but they cannot be negative.
3. The chi-square distribution is different for each number of degrees of freedom, which is $df = n - 1$ in this section. As the number increases, the chi-square distribution approaches a normal distribution.

In Table A-4, each critical value of χ^2 corresponds to an area given in the top row of the table, and that area represents the total region located to the right of the critical value.

Slide 77

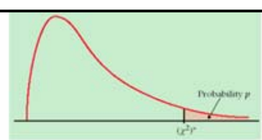


Table entry for p is the critical value $(\chi^2)^*$ with probability p lying to its right.

TABLE F χ^2 distribution critical values

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46	24.10
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26	33.14
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91	34.82
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53	36.48
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12	38.11
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70	39.72
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25	41.31
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79	42.88
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31	44.43
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82	45.97
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31	47.50
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80	49.01

Slide 78

Example: Find the critical values of χ^2 that determine critical regions containing an area of 0.025 in each tail. Assume that the relevant sample size is 10 so that the number of degrees of freedom is 10 – 1, or 9.

α = 0.05

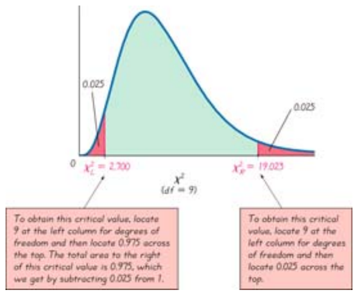
$\alpha/2$ = 0.025

$1 - \alpha/2$ = 0.975

Critical Values: Table A-4

Slide 79

Areas to the right of each tail



Estimators of σ^2

Slide 80

The sample variance s^2 is the best point estimate of the population variance σ^2.

Confidence Interval for the Population Variance σ^2

Slide 81

$$\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$$

Right-tail CV

Confidence Interval for the Population Variance σ^2

Slide 82

$$\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$$

Right-tail CV

Left-tail CV

Confidence Interval for the Population Variance σ^2

Slide 83

$$\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$$

Right-tail CV

Left-tail CV

Confidence Interval for the Population Standard Deviation σ

$$\sqrt{\frac{(n-1)s^2}{\chi^2_R}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_L}}$$

Procedure for Constructing a Confidence Interval for σ or σ^2

Slide 84

1. Verify that the required assumptions are met.
2. Using $n - 1$ degrees of freedom, refer to Table A-4 and find the critical values χ^2_R and χ^2_L that corresponds to the desired confidence level.
3. Evaluate the upper and lower confidence interval limits using this format of the confidence interval:

$$\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$$

continued

Procedure for Constructing a Confidence Interval for σ or σ^2

Slide 85

(continued)

4. If a confidence interval estimate of σ is desired, take the square root of the upper and lower confidence interval limits and change σ^2 to σ.

5. Round the resulting confidence level limits. If using the original set of data to construct a confidence interval, round the confidence interval limits to one more decimal place than is used for the original set of data. If using the sample standard deviation or variance, round the confidence interval limits to the same number of decimals places.

Slide 86

Example: A study found the body temperatures of 106 healthy adults. The sample mean was 98.2 degrees and the sample standard deviation was 0.62 degrees. Find the 95% confidence interval for σ.

$n = 106$

$\bar{x} = 98.2°$

s = 0.62°

$\alpha = 0.05$

$\alpha/2 = 0.025$

$1 - \alpha/2 = 0.975$

$\chi^2_R = 129.561$, $\chi^2_L = 74.222$

$$\frac{(106-1)(0.62)^2}{129.561} < \sigma^2 < \frac{(106-1)(0.62)^2}{74.222}$$

$$0.31 < \sigma^2 < 0.54$$

$$0.56 < \sigma < 0.74$$

We are 95% confident that the limits of 0.56°F and 0.74°F contain the true value of σ. We are 95% confident that the standard deviation of body temperatures of all healthy people is between 0.56°F and 0.74°F.

Determining Sample Size

Slide 87

Table 6-2

Sample Size for σ^2		Sample Size for σ	
To be 95% confident that s^2 is within	of the value of σ^2, the sample size n should be at least	To be 95% confident that s is within	of the value of σ, the sample size n should be at least
1%	77,207	1%	19,204
5%	3,148	5%	767
10%	805	10%	191
20%	210	20%	47
30%	97	30%	20
40%	56	40%	11
50%	37	50%	7
To be 99% confident that s^2 is within	of the value of σ^2, the sample size n should be at least	To be 99% confident that s is within	of the value of σ, the sample size n should be at least
1%	133,448	1%	33,218
5%	5,457	5%	1,335
10%	1,401	10%	335
20%	368	20%	84
30%	171	30%	37
40%	100	40%	21
50%	67	50%	13

Slide 88

Example: We want to estimate σ, the standard deviation off all body temperatures. We want to be 95% confident that our estimate is within 10% of the true value of σ. How large should the sample be? Assume that the population is normally distributed.

From Table 6-2, we can see that 95% confidence and an error of 10% for σ correspond to a sample of size 191.

Chapter index