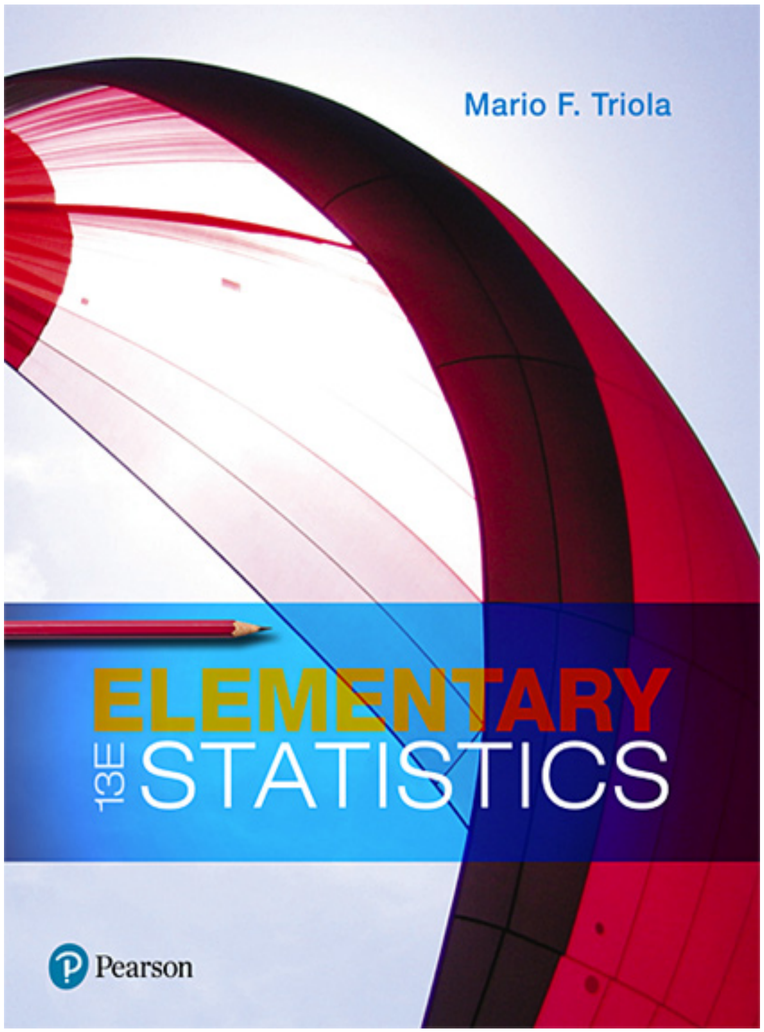
Elementary Statistics

Thirteenth Edition



Chapter 7

Estimating Parameters and Determining Sample Sizes

Estimating Parameters and Determining Sample Sizes

7-1 Estimating a Population Proportion

7-2 Estimating a Population Mean

7-3 Estimating a Population Standard Deviation or Variance

7-4 Bootstrapping: Using Technology for Estimates

Key Concept

This section presents methods for using a sample standard deviation s (or a sample variance s^2) to estimate the value of the corresponding population standard deviation σ (or population variance σ^2). Here are the main concepts:

- **Point Estimate:** The sample variance s^2 is the best **point estimate** of the population variance σ^2.
- **Confidence Interval:** When constructing a **confidence interval** estimate of a population standard deviation, we construct the confidence interval using the χ^2 **distribution.**

Chi-Square Distribution (1 of 6)

Key points about the χ^2 (chi-square or chi-squared) distribution:

- In a normally distributed population with variance σ^2, if we randomly select independent samples of size n and, for each sample, compute the sample variance s^2, the sample statistic $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$ has a sampling distribution called the **chi-square distribution**.

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

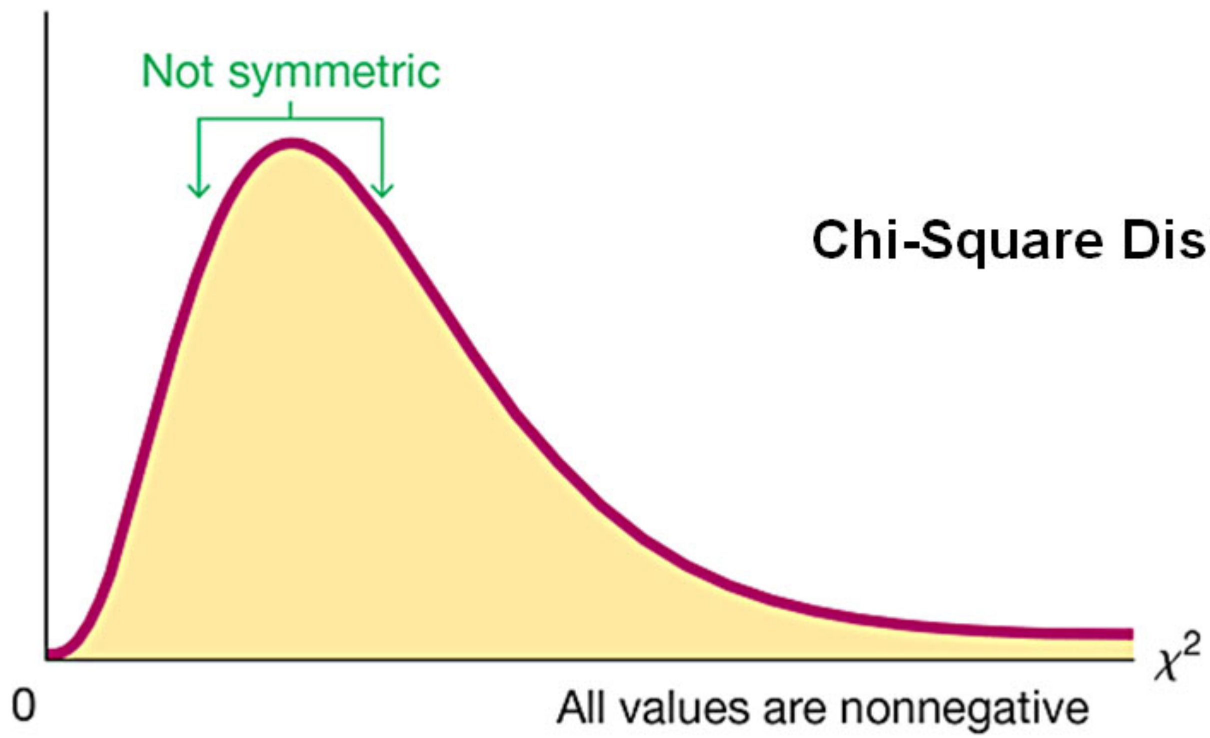
Chi-Square Distribution (2 of 6)

- **Critical Values of χ^2** We denote a right-tailed critical value by χ_R^2 and we denote a left-tailed critical value by χ_L^2. Those critical values can be found by using technology or Table A-4, and they require that we first determine a value for the number of **degrees of freedom**.
- **Degrees of Freedom**
  For the methods of this section, the number of degrees of freedom is the sample size minus 1.

$$\text{Degrees of freedom: df} = n - 1$$

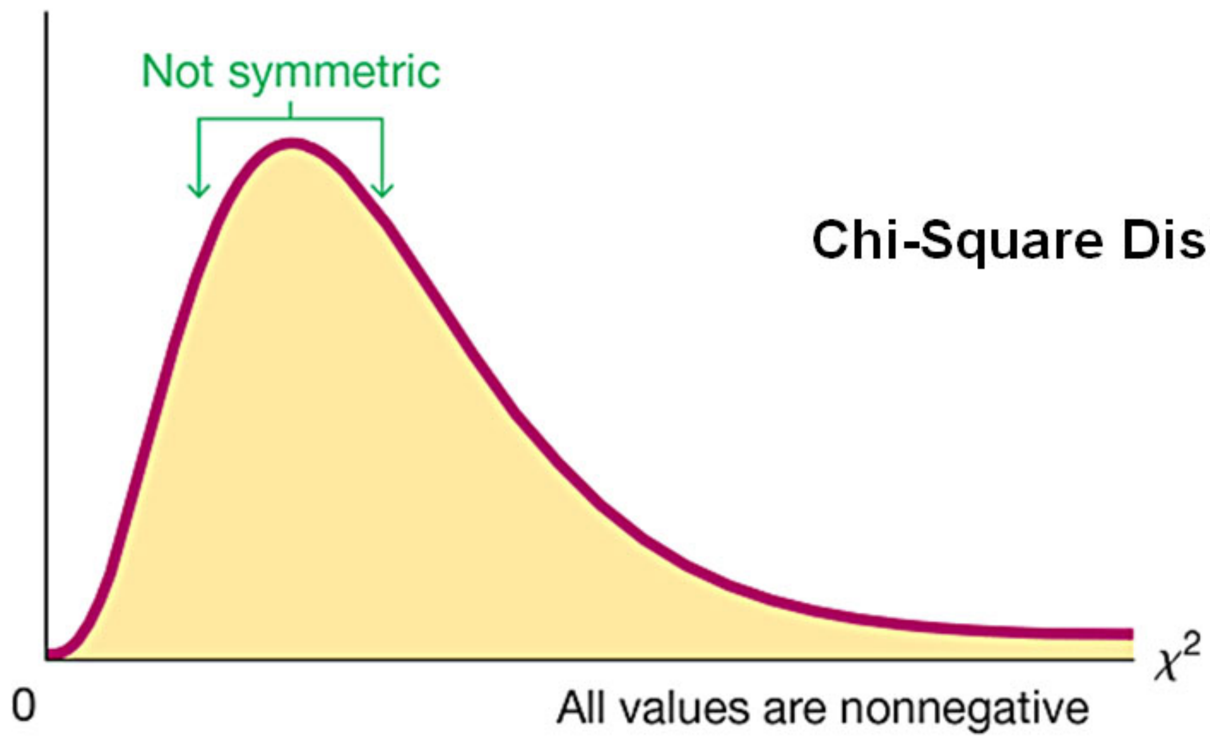
Chi-Square Distribution (3 of 6)

- The chi-square distribution is skewed to the right, unlike the normal and Student *t* distributions.



Chi-Square Distribution (4 of 6)

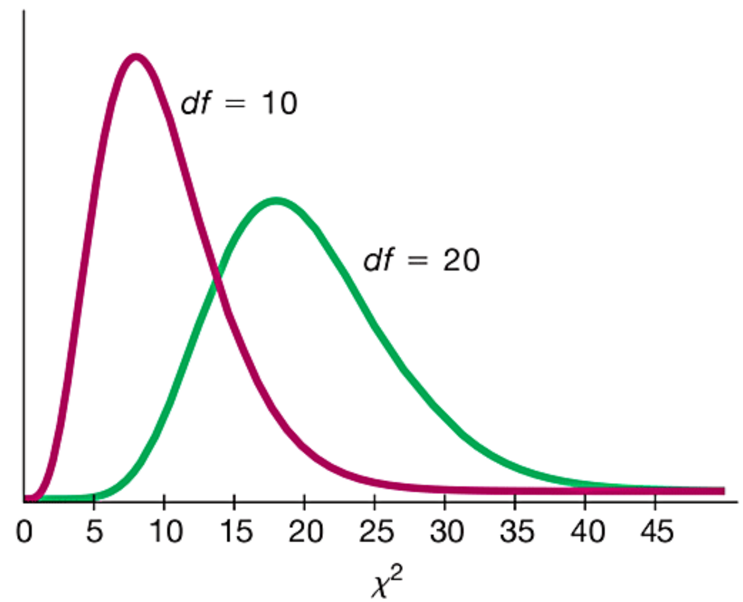
- The values of chi-square can be zero or positive, but they cannot be negative.



Chi-Square Distribution (5 of 6)

- The chi-square distribution is different for each number of degrees of freedom. As the number of degrees of freedom increases, the chi-square distribution approaches a normal distribution.

Chi-Square Distribution for df = 10 and df = 20



Chi-Square Distribution (6 of 6)

Because the chi-square distribution is not symmetric, a confidence interval estimate of σ^2 does not fit a format of $s^2 - E < \sigma^2 < s^2 + E$, so we must do separate calculations for the upper and lower confidence interval limits. If using Table A-4 for finding critical values, note the following design feature of that table:

In Table A-4, each critical value of χ^2 in the body of the table corresponds to an area given in the top row of the table, and each area in that top row is a **cumulative area to the right** of the critical value.

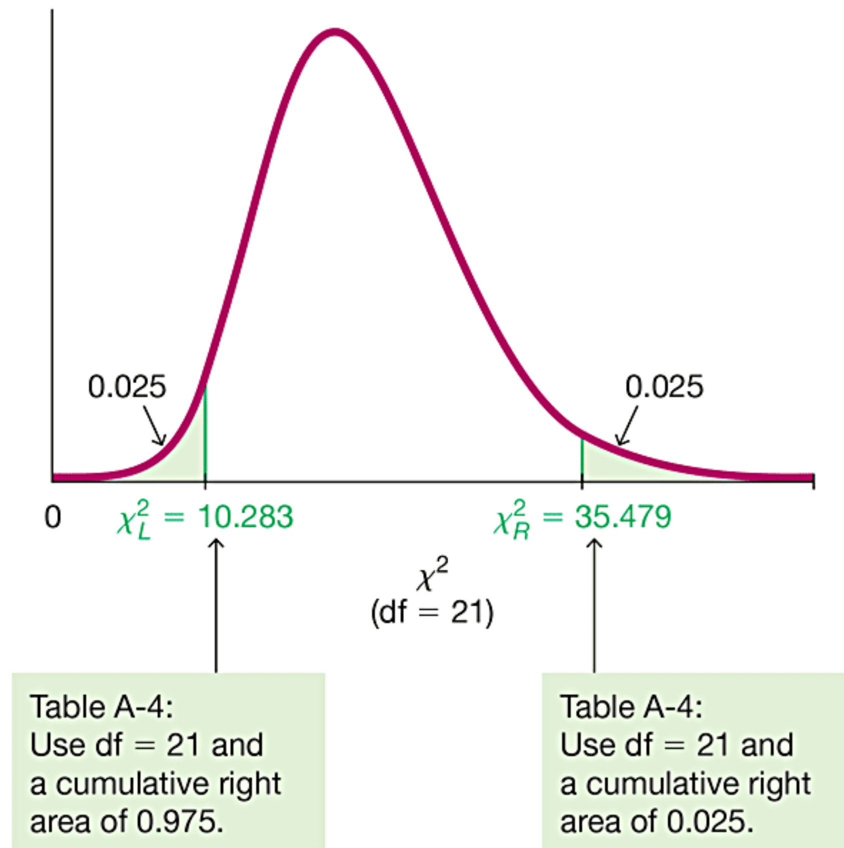
Example: Finding Critical Value of χ^2 (1 of 4)

A simple random sample of 22 IQ scores is obtained. Construction of a confidence interval for the population standard deviation s requires the left and right critical values of χ^2 corresponding to a confidence level of 95% and a sample size of $n = 22$. Find χ^2_L (the critical value of X^2 separating an area of 0.025 in the left tail), and find χ^2_R (the critical value of X^2 separating an area of 0.025 in the right tail).

Example: Finding Critical Value of χ^2 (2 of 4)

Solution

With a sample size of $n = 22$, the number of degrees of freedom is df = $n - 1 = 21$.

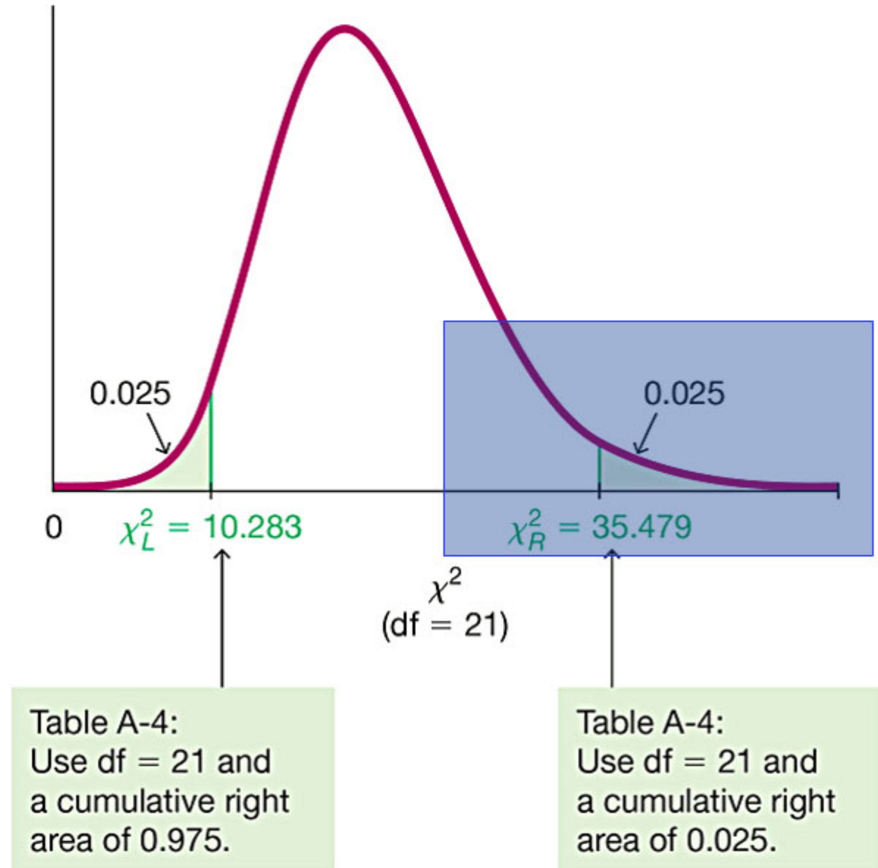


Example: Finding Critical Value of χ^2 (3 of 4)

Solution

$\chi^2_R = 35.479$

The critical value to the right is obtained from Table A-4 in a straightforward manner by locating 21 in the degrees-of-freedom column at the left and 0.025 across the top row.

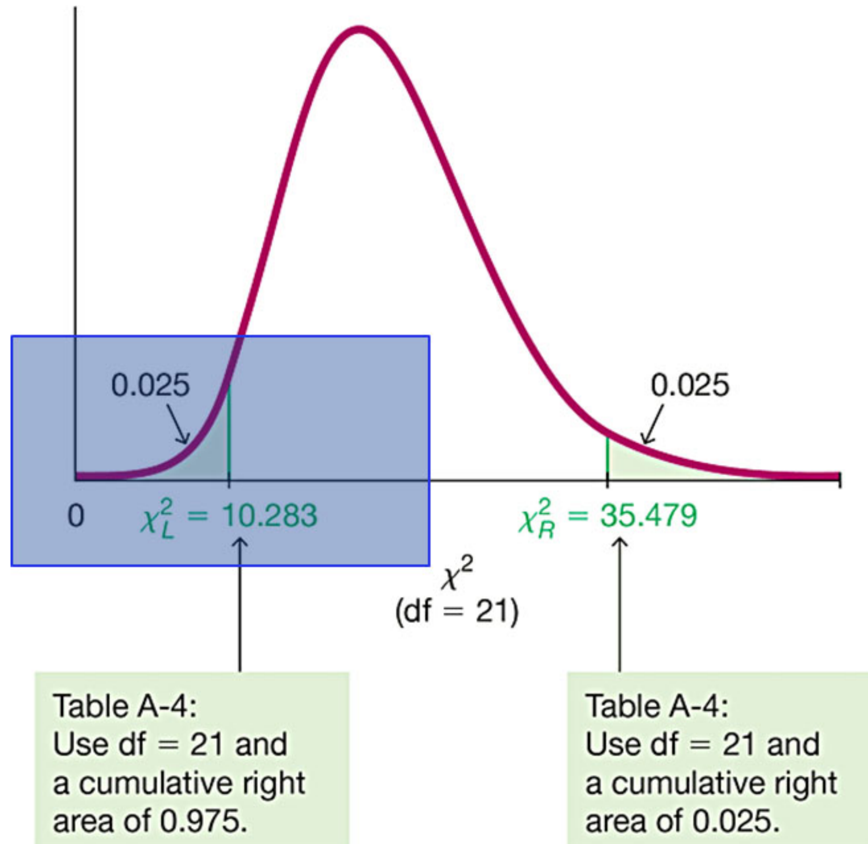


Example: Finding Critical Value of χ^2 (4 of 4)

Solution

$\chi^2_L = 10.283$

But, we must locate 0.975 (or 1 − 0.025) across the top row because the values in the top row are always **areas to the right** of the critical value. The total area to the right of is 0.975.



Confidence Interval for Estimating a Population Standard Deviation or Variance: Objective

Construct a confidence interval estimate of a population standard deviation or variance.

Confidence Interval for Estimating a Population Standard Deviation or Variance: Notation

σ = population standard deviation

σ^2 = population variance

s = sample standard deviation

s^2 = sample variance

n = number of sample values

E = margin of error χ^2

χ^2_L = left-tailed critical value of χ^2

χ^2_R = left-tailed critical value of χ^2

Confidence Interval for Estimating a Population Standard Deviation or Variance: Requirements

1. The sample is a simple random sample.
2. The population must have normally distributed values. The requirement of a normal distribution is much stricter here than in earlier sections, so large departures from normal distributions can result in large errors.

Confidence Interval for Estimating a Population Standard Deviation or Variance: Confidence Interval for the Population Variance σ^2

$$\frac{(n-1)s^2}{\chi_R^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_L^2}$$

Confidence Interval for Estimating a Population Standard Deviation or Variance: Confidence Interval for the Population Standard Deviation σ

$$\sqrt{\frac{(n-1)s^2}{\chi_R^2}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi_L^2}}$$

Confidence Interval for Estimating a Population Standard Deviation or Variance: Round-Off Rule

1. **Original Data:** When using the **original set of data** values, round the confidence interval limits to one more decimal place than is used for the original data.

2. **Summary Statistics:** When using the **summary statistics** (n, s), round the confidence interval limits to the same number of decimal places used for the sample standard deviation.

Procedure for Constructing a Confidence Interval for σ or σ^2 (1 of 2)

1. Verify that the two requirements are satisfied.
2. Using $n - 1$ degrees of freedom, find the critical values χ^2_R and χ^2_L that correspond to the desired confidence level.
3. To get a confidence interval estimate of σ^2, use the following:

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

Procedure for Constructing a Confidence Interval for σ or σ^2 (2 of 2)

4. To get a confidence interval estimate of σ, take the square root of each component of the above confidence interval.
5. Round the confidence interval limits using the round-off rule.

Using Confidence Intervals for Comparisons or Hypothesis Tests

Comparisons Confidence intervals can be used **informally** to compare the variation in different data sets, but **the overlapping of confidence intervals should not be used for making formal and final conclusions about equality of variances or standard deviations.**

Example: Confidence Interval for Estimating σ of IQ Scores (1 of 6)

Data Set 7 “IQ and Lead” in Appendix B lists IQ scores for subjects in three different lead exposure groups. The 22 full IQ scores for the group with medium exposure to lead (Group 2) have a standard deviation of 14.29263. Consider the sample to be a simple random sample and construct a 95% confidence interval estimate of σ, the standard deviation of the population from which the sample was obtained.

Example: Confidence Interval for Estimating σ of IQ Scores (2 of 6)

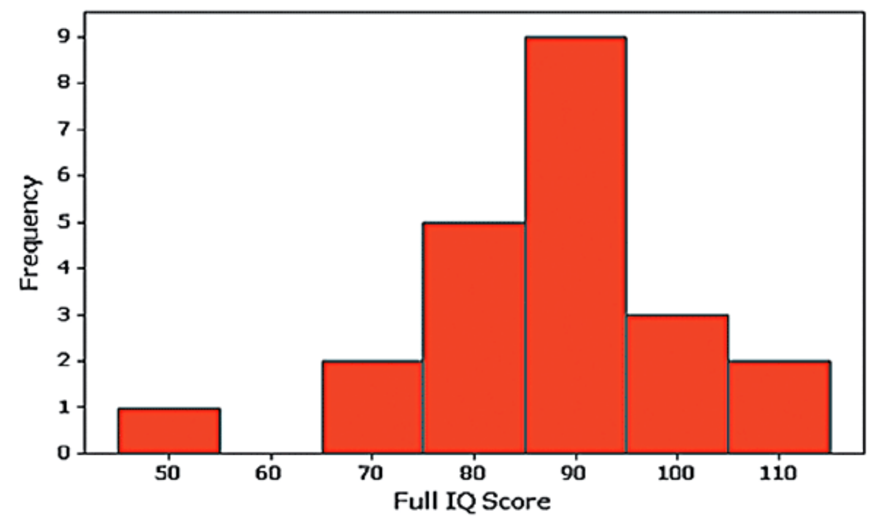
Solution

Requirement Check

Step 1: Check requirements. (1) The sample can be treated as a simple random sample.

(2) The accompanying histogram has a shape very close to the bell shape of a normal distribution.

Minitab



Example: Confidence Interval for Estimating σ of IQ Scores (3 of 6)

Solution

Step 2: If using Table A-4, we use the sample size of $n = 22$ to find degrees of freedom: df = $n - 1 = 21$. Refer to the row corresponding to 21 degrees of freedom, and refer to the columns with areas of 0.975 and 0.025. (For a 95% CI, we divide $\alpha = 0.05$ equally between the two tails of the chi-square distribution, and we refer to the values of 0.975 and 0.025 across the top row.)

The critical values are $\chi^2_L = 10.283$ and $\chi^2_R = 35.479$.

Example: Confidence Interval for Estimating σ of IQ Scores (4 of 6)

Solution

Step 3: Using the critical values of 10.283 and 35.479, the sample standard deviation of *s* = 14.29263 and the sample size of *n* = 22, we construct the 95% confidence interval by evaluating the following:

$$\frac{(n-1)s^2}{\chi_R^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_L^2}$$

$$\frac{(22-1)(14.29263)^2}{35.479} < \sigma^2 < \frac{(22-1)(14.29263)^2}{10.283}$$

Example: Confidence Interval for Estimating σ of IQ Scores (5 of 6)

Solution

Step 4: Evaluating the expression results in

$$120.9 < \sigma^2 < 417.2.$$

Finding the square root of each part (before rounding), then rounding to one decimal place, yields this 95% confidence interval estimate of the population standard deviation: $11.0 < \sigma < 20.4$.

Example: Confidence Interval for Estimating σ of IQ Scores (6 of 6)

Interpretation

Based on this result, we have 95% confidence that the limits of 11.0 and 20.4 contain the true value of s. The confidence interval can also be expressed as (11.0, 20.4), **but it cannot be expressed in a format of** $s \pm E$.

Determining Sample Sizes (1 of 4)

The procedures for finding the sample size necessary to estimate σ are much more complex than the procedures given earlier for means and proportions. For normally distributed populations, the table on the following slide, or the following formula can be used:

$$n = \frac{1}{2}\left(\frac{z_{\frac{\alpha}{2}}}{d}\right)^2.$$

Determining Sample Sizes (2 of 4)

Finding Sample Size

σ

To be 95% confident that *s* is within …	**Of the value of σ, the sample size *n* should be at least**
1%	19,205
5%	768
10%	192
20%	48
30%	21
40%	12
50%	8

Determining Sample Sizes (3 of 4)

Finding Sample Size

σ

To be 99% confident that s is within …	Of the value of σ, the sample size n should be at least
1%	33,218
5%	1,338
10%	336
20%	85
30%	38
40%	22
50%	14

Determining Sample Sizes (4 of 4)

Statdisk also provides sample sizes. With Statdisk, select **Analysis, Sample Size Determination,** and then **Estimate Standard Deviation.** Excel, StatCrunch, and the TI-83/84 Plus calculator do not provide such sample sizes.

Example: Finding Sample Size for Estimating *σ* (1 of 2)

We want to estimate the standard deviation σ of all IQ scores of people with exposure to lead. We want to be 99% confident that our estimate is within 5% of the true value of σ. How large should the sample be? Assume that the population is normally distributed.

Example: Finding Sample Size for Estimating *σ* (2 of 2)

Solution

From the table given for finding sample size, we can see that 99% confidence and an error of 5% for *s* correspond to a sample of size 1336. We should obtain a simple random sample of 1336 IQ scores from the population of subjects exposed to lead.