Chapter 8

Statistical Intervals for a Single Sample

Chapter 8: Interval Estimation of Parameters

LEARNING OBJECTIVES

1. Confidence Interval on μ of a $N(\mu, \sigma^2)$: σ^2 known
2. Confidence Interval on μ of a $N(\mu, \sigma^2)$: σ^2 unknown
3. Confidence Interval on μ of any distribution: large-sample
4. Large-Sample Confidence Interval for a Population Proportion

1. Confidence Interval on μ of a $N(\mu, \sigma^2)$: σ^2 known

Introduction

- A **confidence interval** estimate for μ is an interval of the form $l \leq \mu \leq u$, where the endpoints l and u are computed from the sample data.

- Because different samples will produce different values of l and u, these end-points are values of random variables L and U, respectively.

- Suppose that we can determine values of L and U such that the following probability statement is true:

$$P(L \leq \mu \leq U) = 1 - \alpha$$

There is a probability of 1- α of selecting a sample for which the CI will contain the true value of μ

Introduction

- After we have selected the sample: $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, and computed l and u, the resulting **confidence interval** for μ is

$$l \le \mu \le u$$

Lower-confidence limit

Upper-confidence limit

How to find the random variables L and U ?

CI on μ of a N(μ, σ²): σ² Known

Since $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$, we can find a real number $z_{\alpha/2}$ such that

$$P\left(-z_{\alpha/2} \le \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \le z_{\alpha/2}\right) = 1 - \alpha$$

or

$$P\left(\overline{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$$

L U

CI on μ of a N(μ, σ²): σ² Known

Theorem

If $\bar{x}$ is the sample mean of a random sample of size *n* from a normal population with known variance, (1 – α)-CI on μ is given by

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Solution

Step 1: Find $\bar{x}, \sigma, \mathrm{n}, \mathrm{z}_{\alpha/2}$

Step 2: Find $E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Step 3: CI: $\bar{x} - E \le \mu \le \bar{x} + E$

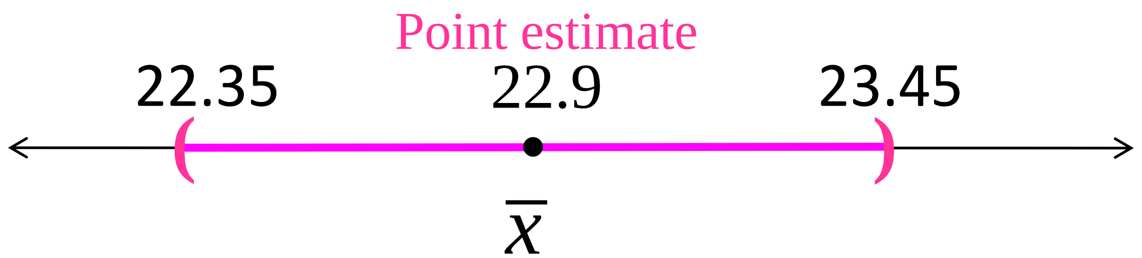
CI on μ of a N(μ, σ²): σ² Known

Example 1

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.

CI on μ of a N(μ, σ²): σ² Known

Confidence interval:



CI on μ of a N(μ, σ^2): σ^2 Known

Example 2

ASTM Standard E23 defines standard test methods for notched bar impact testing of metallic materials. The Charpy V-notch (CVN) technique measures impact energy and is often used to determine whether or not a material experiences a ductile-to-brittle transition with decreasing temperature. Ten measurements of impact energy (*J*) on specimens of A238 steel cut at 60∘C are as follows: 64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, and 64.3. Assume that impact energy is normally distributed with $\sigma = 1$ *J*. We want to find a 95% CI for μ, the mean impact energy.

CI on μ of a $N(\mu, \sigma^2)$: σ^2 Known

Example 3

The average zinc concentration recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 grams per milliliter. Find the 95% and 99% confidence intervals for the mean zinc concentration in the river. Assume that the population standard deviation is 0.3 gram per milliliter.

$\bar{x} = 2,6$

$\sigma^2 = 0,3$

$1-\alpha_1 = 0,95 \rightarrow \alpha_1 = 0,05$

$1-\alpha_2 = 0,99 \rightarrow \alpha_2 = 0,01$

$z_{\alpha_1/2} = 1,96$

$z_{\alpha_1/2} \frac{\sigma}{\sqrt{n}} = 0,1$

95% CI

$\mu \in \{2,42 ; 2,77\}$

$z_{\alpha_2/2} = 2,57$

$z_{\alpha_2/2} \frac{\sigma}{\sqrt{n}} = 0,23$

99% CI

$\mu \in \{2,36 ; 2,834\}$

CI on μ of a N(μ, σ²): σ² Known

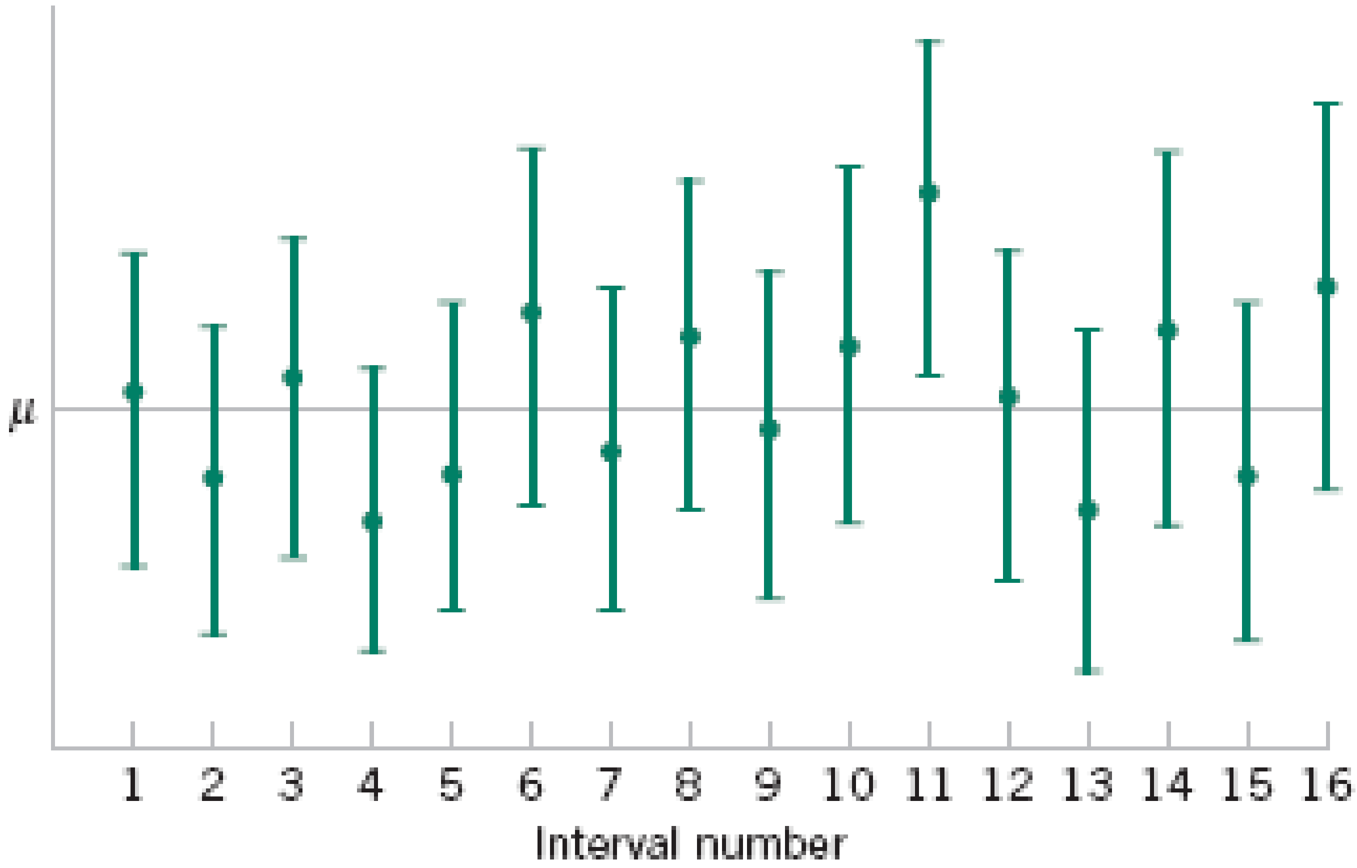


Figure 8-1 Repeated construction of a confidence interval for μ.

CI on μ of a N(μ, σ²): σ² Known

Confidence Level and Precision of Estimation

The length of a confidence interval is a measure of the precision of estimation.

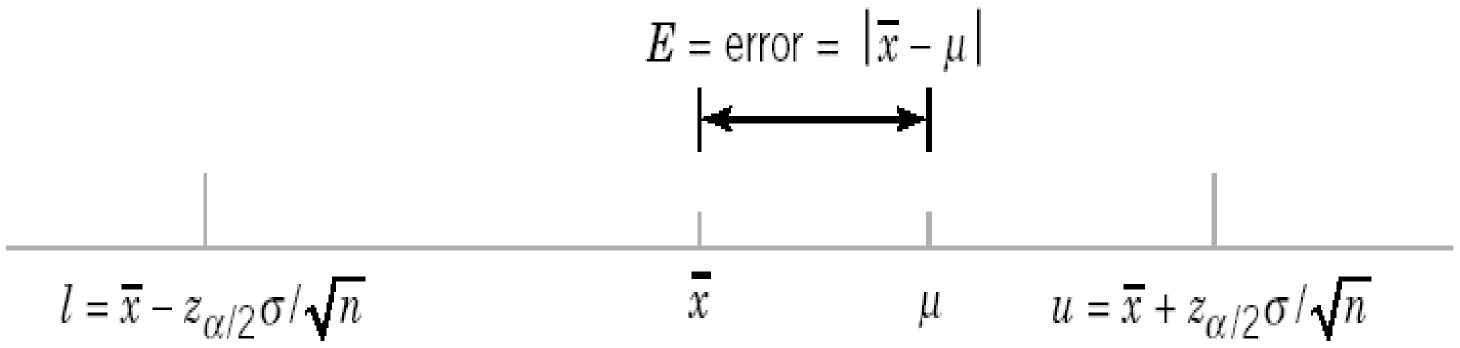


Figure 8-2 Error in estimating μ with $\bar{x}$.

CI on μ of a N(μ, σ^2): σ^2 Known

Theorem — **Choice of Sample Size**

If $\bar{x}$ is used as an estimate of μ , we can be $(1 - \alpha)$-confident that the error $|\,\bar{x} - \mu\,|$ will not exceed a specified amount E when the sample size is

$$n = \left(\frac{z_{\alpha/2}\sigma}{E} \right)^2 \qquad (8.6)$$

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

$$\Rightarrow n = \left[\frac{z_{\alpha/2} \cdot \sigma}{E} \right]^2$$

CI on μ of a N(μ, σ²): σ² Known

Example 4

To illustrate the use of this procedure, consider the CVN test described in Example 2 and suppose that we want to determine how many specimens must be tested to ensure that the 95% CI on μ for A238 steel cut at 60∘C has a length of at most 1.0 *J*. Because the bound on error in estimation *E* is one-half of the length of the CI, to determine *n*, we use Equation 8.6 with $E = 0.5$, $\sigma = 1$, and $z_{\alpha/2} = 1.96$.

$$n_x = \left[\frac{z \cdot \sigma}{E}\right]^2 = \left[\frac{1,96}{0,5}\right]^2 = 15,36$$

$$n = \lceil n_x \rceil = 16$$

Example 9.5: The average zinc concentration recovered from a sample is found to be 2.6 grams per milliliter. Assume that the population standard deviation is 0.3 gram per milliliter. How large a sample is required if we want to be 95% confident that our estimate of μ is off by less than 0.05?

$\mu = 2,6$

$\sigma^2 = 0,3$

$E = 0,05$

$CI = 0,95$

$\alpha = 0,05$

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2 = \left(\frac{1,96 \cdot \sqrt{0,3}}{0,05} \right)^2 = 461$$

$$z_{\alpha/2} = 1,96$$

CI on μ of a N(μ, σ²): σ² Known

Theorem One-Sided Confidence Bounds

A $(1 - \alpha)$ upper-confidence bound for μ is

$$\mu \leq \overline{x} + z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

A $(1 - \alpha)$ lower-confidence bound for μ is

$$\overline{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}} \leq \mu$$

CI on μ of a N(μ, σ²): σ² Known

Example 4

The same data for impact testing from Example 2 are used to construct a lower, one-sided 95% confidence interval for the mean impact energy. Recall that $x = 64.46$, $\sigma = 1J$, and $n = 10$. The interval is

Example

In a psychological testing experiment, 25 subjects are selected randomly and their reaction time, in seconds, to a particular stimulus is measured. Past experience suggests that the variance in reaction times to these types of stimuli is 4 sec^2 and that the distribution of reaction times is approximately normal. The s^2 average time for the subjects is 6.2 seconds. Give an upper 95% bound for the mean reaction time. $\bar{x}$ α

CI 95% UI

$1-\alpha = 0,95 \to \alpha = 0,05$

$z_\alpha = 1,644$

$\sigma^2 = 4$; $\bar{x} = 6,2$

$E = z_\alpha \cdot \frac{\sigma}{\sqrt{n}} = 0,65$

$x \in \{5,54; 6,856\}$

Exercises 1

Past experience has indicated that the breaking strength of yarn used in manufacturing drapery material is normally distributed and that σ = 2 psi. A random sample of nine specimens is tested, and the average breaking strength is found to be 98 psi. Find a 95% two-sided confidence interval on the true mean breaking strength.

$n = 9$

$\sigma = 2$

$\bar{x} = 98$

$1 - \alpha = 0{,}95$

$\alpha = 0{,}05$

$z_{\alpha/2} = 1{,}96$

$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1{,}3$

$x \in \{96{,}7;\ 99{,}3\}$

2. Confidence Interval on μ of a $N(\mu, \sigma^2)$: σ^2 unknown

CI on μ of a N(μ, σ²): σ² Unknown

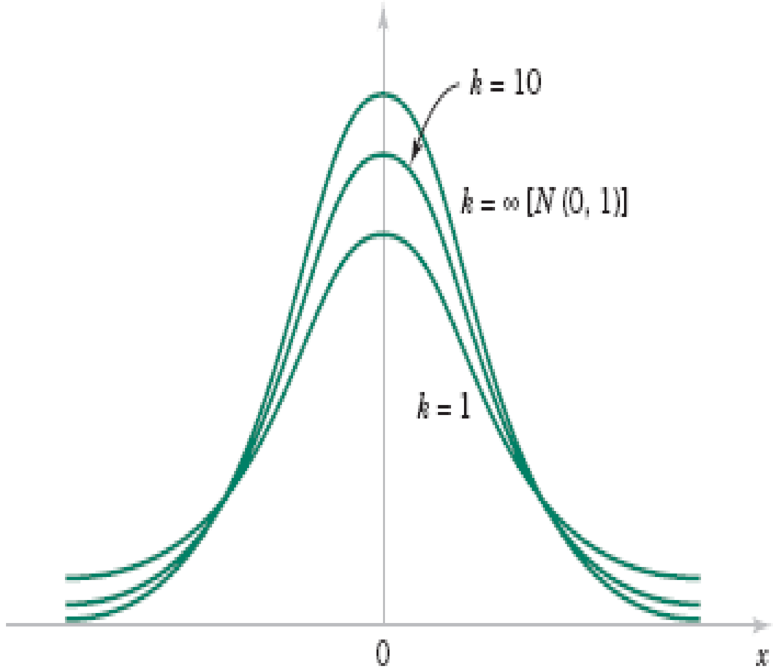
Definition **Student Distribution**

The variable random X with probability density function

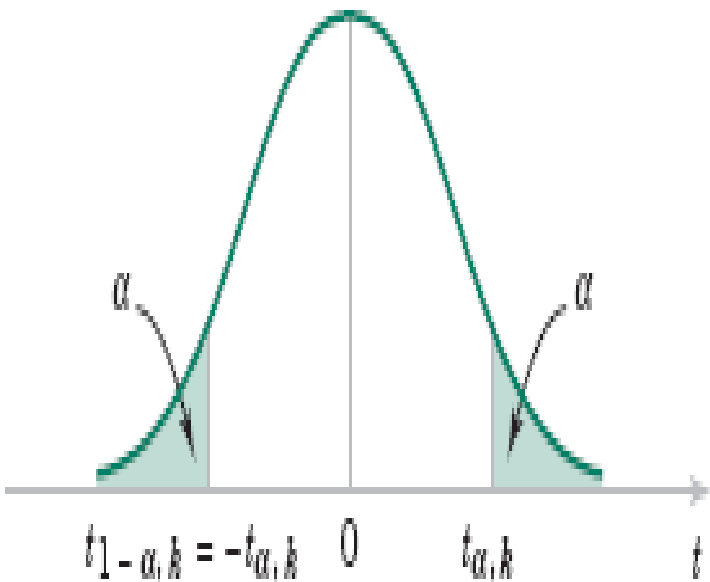
$$f(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k}\,\Gamma(k/2)} \frac{1}{\left(\frac{x^2}{k}+1\right)^{\frac{k+1}{2}}}, \quad -\infty < x < +\infty$$

is called a Student variable random or *t* distribution with *k* degrees of freedom.

CI on μ of a N(μ, σ²): σ² Unknown



Probability density functions of several *t* distributions.



Percentage points of the *t* distribution.

CI ON μ OF A N(μ, σ^2): σ^2 UNKNOWN

Theorem ***t* distribution**

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. The random variable

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

has a *t* distribution with $n - 1$ degrees of freedom.

Theorem

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ^2,

- A $(1 - \alpha)$**-percent confidence interval** on μ is given by

$$\bar{x} - t_{\alpha/2,n-1} \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

- A $(1 - \alpha)$ **upper-confidence bound** for μ is

$$\mu \le \bar{x} + t_{\alpha,n-1} \frac{s}{\sqrt{n}}$$

- A $(1 - \alpha)$ **lower-confidence bound** for μ is

$$\bar{x} - t_{\alpha,n-1} \frac{s}{\sqrt{n}} \le \mu$$

CI ON μ OF A N(μ, σ^2): σ^2 UNKNOWN

Example

An article in the journal *Materials Engineering* describes the results of tensile adhesion tests on 22 U-700 alloy specimens. The load at specimen failure is as follows: 19.8 10.1 14.9 7.5 15.4 15.4 15.4 18.5 7.9 12.7 11.9 11.4 11.4 14.1 17.6 16.7 15.8 19.5 8.8 13.6 11.9 11.4. Find a 95% CI on μ.

$s^2 = 12,62$

$\bar{x} = 13,71$

$t_{\alpha/2;\, n-1} = 2,08$

$n = 22$

$E = t_{\alpha/2;\, n-1} \frac{s}{\sqrt{n}} = 1,57$

$\mu \in \{12,14 \,;\, 15,28\}$

CI ON μ OF A $N(\mu, \sigma^2)$: σ^2 UNKNOWN

Example

The contents of seven similar containers of sulfuric acid are 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, and 9.6 liters. Find a 95% confidence interval for the mean contents of all such containers, assuming an approximately normal distribution.

$\bar{x} = 10,06$

$s = 0,24$

$n = 7$

$\alpha = 0,05$

$t_{\alpha/2;n-1} = 2,447$

$E = t_{\alpha/2;n-1} \dfrac{s}{\sqrt{n}} = 0,23$

$x \in \{9,83 ; 10,28\}$

Exercises for Section 8.2

8.2.1 A random sample has been taken from a normal distribution. Output from a software package follows:

Variable	N	Mean	SE Mean	StDev	Variance	Sum
x	10	?	0.507	1.605	?	251.848

$\mu = \frac{S}{n} = 25,1848$

$\sigma^2 = (SD)^2 = 2,576025$

a. Fill in the missing quantities.

b. Find a 95% CI on the population mean.

$E = z_{\alpha/2} \cdot SE = 0,99372$

$\mu \pm E = 25,1848 \pm 0,99372 = (24,19108 ; 26,17852)$

3. Confidence Interval on μ of any distribution: large-sample

ARBITRARY DISTRIBUTION: A LARGE-SAMPLE

Theorem

When n is large, the quantity

$n \geq 40$

$$\frac{\overline{X} - \mu}{S / \sqrt{n}}$$

has an approximate N (0, 1). Consequently,

$$\overline{x} - z_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2} \frac{s}{\sqrt{n}}$$

is a **large sample confidence interval** for μ, with confidence level of approximately (1 – α).

ARBITRARY DISTRIBUTION: A LARGE-SAMPLE

Example

A sample of fish was selected from 53 Florida lakes and mercury concentration in the muscle tissue was measured (ppm). The mercury concentration values are

1.230	0.490	0.490	1.080	0.590	0.280	0.180	0.100	0.940
1.330	0.190	1.160	0.980	0.340	0.340	0.190	0.210	0.400
0.040	0.830	0.050	0.630	0.340	0.750	0.040	0.860	0.430
0.044	0.810	0.150	0.560	0.840	0.870	0.490	0.520	0.250
1.200	0.710	0.190	0.410	0.500	0.560	1.100	0.650	0.270
0.270	0.500	0.770	0.730	0.340	0.170	0.160	0.270	

Find an approximate 95% CI on μ.

ARBITRARY DISTRIBUTION: A LARGE-SAMPLE

Example

Scholastic Aptitude Test (SAT) mathematics scores of a random sample of 500 high school seniors in the state of Texas are collected, and the sample mean and standard deviation are found to be 501 and 112, respectively. Find a 99% confidence interval on the mean SAT mathematics score for seniors in the state of Texas.

$\mu = 501$

$\sigma = 112$

$n = 500$

$E = z_{\alpha/2} \cdot SE = 2{,}576 \cdot 5 = 12{,}902$

$SE = \frac{s}{\sqrt{n}} = \frac{112}{\sqrt{500}} = 5$

$\mu \pm SE = 501 \pm 12{,}9 = (488{,}1 ; 513{,}9)$

Exercises

Many cardiac patients wear an implanted pacemaker to control their heartbeat. A plastic connector module mounts on the top of the pacemaker. Assuming a standard deviation of 0.0015 inch and an approximately normal distribution, find a 95% confidence interval for the mean of the depths of all connector modules made by a certain manufacturing company. A random sample of 75 modules has an average depth of 0.310 inch.

$s = 0{,}0015$

$\mu = 0{,}31$

$\alpha = 0{,}05$

$n = 75$

$E = z_{\alpha/2} \cdot SE = 0{,}0003908$

$SE = \frac{s}{\sqrt{n}} = 0{,}000173$

$\mu \pm SE = 0{,}31 \pm 0{,}0003908$

$= (0{,}3097 ; 0{,}3097)$

Exercise

The heights of a random sample of 50 college students showed a mean of 174.5 centimeters and a standard deviation of 6.9 centimeters.

(a) Construct a 98% confidence interval for the mean height of all college students.

(b) What can we assert with 98% confidence about the possible size of our error if we estimate the mean height of all college students to be 174.5 centimeters?

$\alpha = 0{,}02$
$n = 50$
$\mu = 174{,}5$
$s = 6{,}9$

$SE = \frac{s}{\sqrt{n}} = 0{,}975$

$E = z_{\alpha/2} \cdot SE = 2{,}326 \cdot 0{,}975 = 2{,}2679$

$\mu \pm E = 174{,}5 \pm 2{,}2679 = (172{,}2321;\ 176{,}7679)$

4. Confidence Interval for the proportion p: large-sample

CI FOR *p*: A LARGE-SAMPLE

Theorem Normal approximation

If n is large, the distribution of

$n \geq 40$
or n is not mentioned as population

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

$\sigma^2 \sim s^2 = p(1-p)$

is approximately $N(0, 1)$.

We can find $z_{\alpha/2}$ such that

$$P\left(-z_{\alpha/2} \leq \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{\alpha/2}\right) \simeq 1 - \alpha$$

or

$$P\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{P} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) \simeq 1 - \alpha$$

CI FOR *p*: A LARGE-SAMPLE

Theorem — CI for p

Let $\hat{p}$ is a point estimation for the proportion p of the population based on a random sample of size n, an approximate $(1-\alpha)$-confidence interval on p is

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Remark: $n\hat{p} > 5$ and $n(1-\hat{p}) > 5$

CI FOR *p*: A LARGE-SAMPLE

Example 1

In a survey of 1219 U.S. adults, 354 said that their favorite sport to watch is football. Construct a 95% confidence interval for the proportion of adults in the United States who say that their favorite sport to watch is football.

$\hat{p} = 0,29$

$\alpha = 0,05$

$E = z_{\alpha/2} \cdot SE = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0,025$

$\bar{p} \pm E = 0,29 \pm 0,025$

$= (0,265 ; 0,315)$

CI FOR *p*: A LARGE-SAMPLE

Example 2

In a random sample of n = 500 families owning television sets in the city of Hamilton, Canada, it is found that x = 340 subscribe to HBO. Find a 95% confidence interval for the actual proportion of families with television sets in this city that subscribe to HBO.

CI FOR *p*: A LARGE-SAMPLE

Theorem Choice of Sample Size

If $\hat{p}$ is used as an estimate of p, we can be $(1 - \alpha)$-confident that the error $|\hat{p} - p|$ will not exceed a specified amount E when the sample size is

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p)$$

$\leq 1/4$

An upper bound on n is given by

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

CI FOR *p*: A LARGE-SAMPLE

Theorem One-Sided Confidence Bounds

A $(1-\alpha)$ **upper-confidence bound** for p is

$$p \le \hat{p} + z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

A $(1-\alpha)$ **lower-confidence bound** for p is

$$\hat{p} - z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p$$

CI FOR *p*: A LARGE-SAMPLE

Example 3

How large a sample is required if we want to be 95% confident that our estimate of p in Example 2 is within 0.02 of the true value?

Exercises for Section 8.4

The 2004 presidential election exit polls from the critical state of Ohio provided the following results. The exit polls had 2020 respondents, 768 of whom were college graduates. Of the college graduates, 412 voted for George Bush.

a. Calculate a 95% confidence interval for the proportion of college graduates in Ohio who voted for George Bush.

b. Calculate a 95% lower confidence bound for the proportion of college graduates in Ohio who voted for George Bush.

Exercises for Section 8.4

The U.S. Postal Service (USPS) has used optical character recognition (OCR) since the mid-1960s. In 1983, USPS began deploying the technology to major post offices throughout the country (*www.britannica.com*). Suppose that in a random sample of 500 handwritten zip code digits, 466 were read correctly.

a. Construct a 95% confidence interval for the true proportion of correct digits that can be automatically read.

b. What sample size is needed to reduce the margin of error to 1%?

c. How would the answer to part (b) change if you had to assume that the machine read only one-half of the digits correctly?