Chapter 9

One- and Two-Sample Estimation Problems

Single Sample: Estimating the Mean

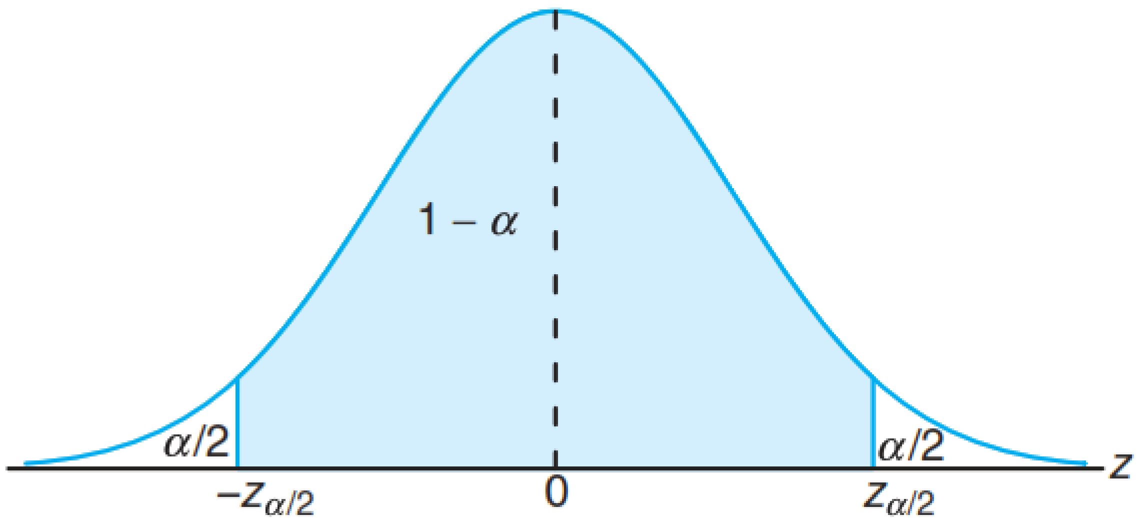

Figure 9.2: $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$.

Confidence Interval on μ, σ^2 Known

Confidence interval for the mean; σ is known

$$\boxed{\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

If $\bar{x}$ is the mean of a random sample of size n from a population with known variance σ^2, a $100(1-\alpha)\%$ confidence interval for μ is given by

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

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

$$\hat{\theta}_L = \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{and} \quad \hat{\theta}_U = \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

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.

Solution: The point estimate of μ is $\bar{x} = 2.6$. The z-value leaving an area of 0.025 to the right, and therefore an area of 0.975 to the left, is $z_{0.025} = 1.96$ (Table A.3). Hence, the 95% confidence interval is

$$2.6 - (1.96)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (1.96)\left(\frac{0.3}{\sqrt{36}}\right),$$

which reduces to $2.50 < \mu < 2.70$. To find a 99% confidence interval, we find the z-value leaving an area of 0.005 to the right and 0.995 to the left. From Table A.3 again, $z_{0.005} = 2.575$, and the 99% confidence interval is

$$2.6 - (2.575)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (2.575)\left(\frac{0.3}{\sqrt{36}}\right),$$

or simply

$$2.47 < \mu < 2.73.$$

Example 9.13. Construct a 95% confidence interval for the population mean based on a sample of measurements

$$2.5,\ 7.4,\ 8.0,\ 4.5,\ 7.4,\ 9.2$$

if measurement errors have Normal distribution, and the measurement device guarantees a standard deviation of $\sigma = 2.2$.

Solution. This sample has size $n = 6$ and sample mean $\bar{X} = 6.50$. To attain a confidence level of

$$1 - \alpha = 0.95,$$

$$\alpha = 0.05 \text{ and } \alpha/2 = 0.025.$$

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 6.50 \pm (1.960)\frac{2.2}{\sqrt{6}}$$

$$= \underline{6.50 \pm 1.76} \text{ or } \underline{[4.74,\ 8.26]}.$$

The $100(1-\alpha)\%$ confidence interval provides an estimate of the accuracy of our point estimate. If μ is actually the center value of the interval, then $\bar{x}$ estimates μ without error. Most of the time, however, $\bar{x}$ will not be exactly equal to μ and the point estimate will be in error. The size of this error will be the absolute value of the difference between μ and $\bar{x}$, and we can be $100(1-\alpha)\%$ confident that this difference will not exceed $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. We can readily see this if we draw a diagram of a hypothetical confidence interval, as in Figure 9.4.

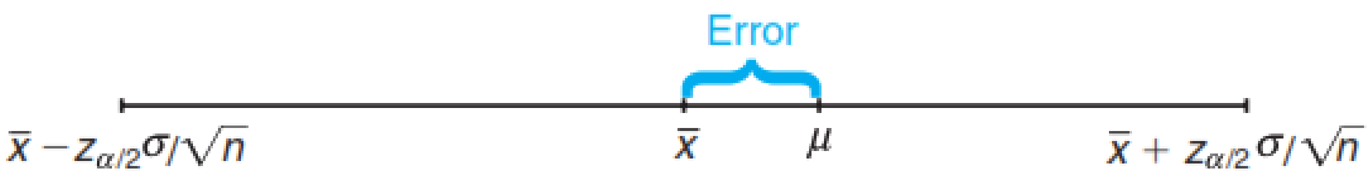

Figure 9.4: Error in estimating μ by $\bar{x}$.

Theorem 9.1: If $\bar{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error will not exceed $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

In Example 9.2, we are 95% confident that the sample mean $\bar{x} = 2.6$ differs from the true mean μ by an amount less than $(1.96)(0.3)/\sqrt{36} = 0.1$ and 99% confident that the difference is less than $(2.575)(0.3)/\sqrt{36} = 0.13$.

Frequently, we wish to know how large a sample is necessary to ensure that the error in estimating μ will be less than a specified amount e. By Theorem 9.1, we must choose n such that $z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = e$. Solving this equation gives the following formula for n.

Theorem 9.2: If $\bar{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error will not exceed a specified amount e when the sample size is

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

How large a sample is required if we want to be 95% confident that our estimate of μ in Example 9.2 is off by less than 0.05?

***Solution*:** The population standard deviation is $\sigma = 0.3$. Then, by Theorem 9.2,

$$n = \left[\frac{(1.96)(0.3)}{0.05}\right]^2 = 138.3.$$

Therefore, we can be 95% confident that a random sample of size 139 will provide an estimate $\bar{x}$ differing from μ by an amount less than 0.05.

One-Sided Confidence Bounds on μ, σ^2 Known

If $\bar{X}$ is the mean of a random sample of size n from a population with variance σ^2, the one-sided $100(1-\alpha)\%$ confidence bounds for μ are given by

$$\text{upper one-sided bound:} \qquad \bar{x} + z_\alpha \sigma/\sqrt{n};$$

$$\text{lower one-sided bound:} \qquad \bar{x} - z_\alpha \sigma/\sqrt{n}.$$

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 average time for the subjects is 6.2 seconds. Give an upper 95% bound for the mean reaction time.

***Solution*:** The upper 95% bound is given by

$$\begin{aligned}\bar{x} + z_\alpha \sigma/\sqrt{n} &= 6.2 + (1.645)\sqrt{4/25} = 6.2 + 0.658 \\ &= 6.858 \text{ seconds.}\end{aligned}$$

Hence, we are 95% confident that the mean reaction time is less than 6.858 seconds.

The Case of σ Unknown

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

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

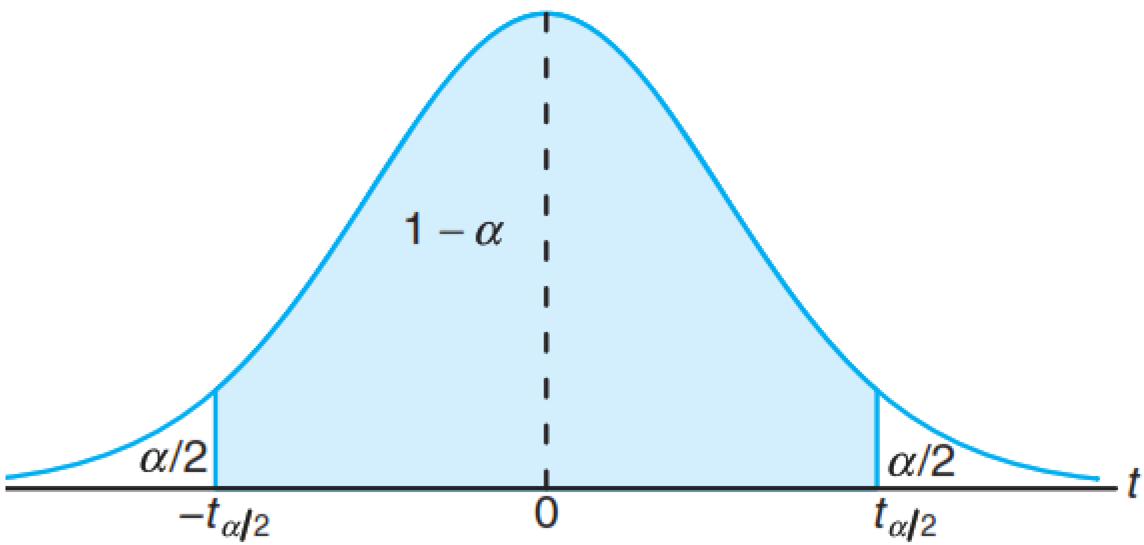

Figure 9.5: $P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha$.

Confidence Interval on μ, σ^2 Unknown

Confidence interval for the mean; σ is unknown

$$\bar{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

where $t_{\alpha/2}$ is a critical value from T-distribution with $n-1$ degrees of freedom

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal population with unknown variance σ^2, a $100(1-\alpha)\%$ confidence interval for μ is

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

where $t_{\alpha/2}$ is the t-value with $v = n-1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

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.

***Solution*:** The sample mean and standard deviation for the given data are

$$\bar{x} = 10.0 \qquad \text{and} \qquad s = 0.283.$$

Using Table A.4, we find $t_{0.025} = 2.447$ for $v = 6$ degrees of freedom.

95% confidence interval for μ is

$$10.0 - (2.447)\left(\frac{0.283}{\sqrt{7}}\right) < \mu < 10.0 + (2.447)\left(\frac{0.283}{\sqrt{7}}\right),$$

which reduces to $9.74 < \mu < 10.26$.

Concept of a Large-Sample Confidence Interval

Often statisticians recommend that even when normality cannot be assumed, σ is unknown, and $n \geq 30$, s can replace σ and the confidence interval

$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

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.

Solution:

we find $z_{0.005} = 2.575$. Hence, a 99% confidence interval for μ is

$$501 \pm (2.575)\left(\frac{112}{\sqrt{500}}\right) = 501 \pm 12.9,$$

$$488.1 < \mu < 513.9.$$

Two Samples: Estimating the Difference between Two Means

Confidence Interval for $\mu_1 - \mu_2$, σ_1^2 and σ_2^2 Known

If $\bar{x}_1$ and $\bar{x}_2$ are means of independent random samples of sizes n_1 and n_2 from populations with known variances σ_1^2 and σ_2^2, respectively, a $100(l - \alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

Confidence interval for the difference between two means

Confidence interval for the difference of means; known standard deviations

$$\boxed{\bar{X} - \bar{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$$

Example 9.16 (Delays at nodes). Internet connections are often slowed by delays at nodes. Let us determine if the delay time increases during heavy-volume times.

Five hundred packets are sent through the same network between 5 pm and 6 pm (sample X), and three hundred packets are sent between 10 pm and 11 pm (sample Y). The early sample has a mean delay time of 0.8 sec with a standard deviation of 0.1 sec whereas the second sample has a mean delay time of 0.5 sec with a standard deviation of 0.08 sec. Construct a 99.5% confidence interval for the difference between the mean delay times.

Solution:

We have $n = 500$, $\bar{X} = 0.8$, $s_X = 0.1$; $m = 300$, $\bar{Y} = 0.5$, $s_Y = 0.08$.

$$\begin{aligned} \bar{X} - \bar{Y} \quad &\pm \quad z_{0.0025}\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}} = (0.8 - 0.5) \pm (2.81)\sqrt{\frac{(0.1)^2}{500} + \frac{(0.08)^2}{300}} \\ &= \quad \underline{0.3 \pm 0.018} \text{ or } \underline{[0.282, 0.318]}. \end{aligned}$$

Variances Unknown but Equal

Consider the case where σ_1^2 and σ_2^2 are unknown. If $\sigma_1^2 = \sigma_2^2 = \sigma^2$, we obtain a standard normal variable of the form

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2[(1/n_1) + (1/n_2)]}}.$$

Confidence Interval for $\mu_1 - \mu_2$, $\sigma_1^2 = \sigma_2^2$ but Both Unknown

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

If $\bar{x}_1$ and $\bar{x}_2$ are the means of independent random samples of sizes n_1 and n_2, respectively, from approximately normal populations with unknown but equal variances, a $100(1 - \alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

where s_p is the pooled estimate of the population standard deviation and $t_{\alpha/2}$ is the t-value with $v = n_1 + n_2 - 2$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

9.35 A random sample of size $n_1 = 25$, taken from a normal population with a standard deviation $\sigma_1 = 5$, has a mean $\bar{x}_1 = 80$. A second random sample of size $n_2 = 36$, taken from a different normal population with a standard deviation $\sigma_2 = 3$, has a mean $\bar{x}_2 = 75$. Find a 94% confidence interval for $\mu_1 - \mu_2$.

$$n_1 = 25, n_2 = 36, \bar{x}_1 = 80, \bar{x}_2 = 75, \sigma_1 = 5, \sigma_2 = 3,$$

$$(80 - 75) - (1.88)\sqrt{25/25 + 9/36} < \mu_1 - \mu_2 < (80 - 75) + (1.88)\sqrt{25/25 + 9/36},$$

$$2.9 < \mu_1 - \mu_2 < 7.1.$$

9.36 Two kinds of thread are being compared for strength. Fifty pieces of each type of thread are tested under similar conditions. Brand A has an average tensile strength of 78.3 kilograms with a standard deviation of 5.6 kilograms, while brand B has an average tensile strength of 87.2 kilograms with a standard deviation of 6.3 kilograms. Construct a 95% confidence interval for the difference of the population means.

$n_A = 50, n_B = 50, \bar{x}_A = 78.3, \bar{x}_B = 87.2, \sigma_A = 5.6$, and $\sigma_B = 6.3$.

$$(87.2 - 78.3) \pm 1.96\sqrt{5.6^2/50 + 6.3^2/50} = 8.9 \pm 2.34,$$

$$\text{or } 6.56 < \mu_A - \mu_B < 11.24.$$

9.38 Two catalysts in a batch chemical process, are being compared for their effect on the output of the process reaction. A sample of 12 batches was prepared using catalyst 1, and a sample of 10 batches was prepared using catalyst 2. The 12 batches for which catalyst 1 was used in the reaction gave an average yield of 85 with a sample standard deviation of 4, and the 10 batches for which catalyst 2 was used gave an average yield of 81 and a sample standard deviation of 5. Find a 90% confidence interval for the difference between the population means, assuming that the populations are approximately normally distributed with equal variances.

$$n_1 = 12, n_2 = 10, \bar{x}_1 = 85, \bar{x}_2 = 81, s_1 = 4, s_2 = 5,$$

$$s_p = 4.478 \qquad t_{0.05} = 1.725$$

$$(85 - 81) \pm (1.725)(4.478)\sqrt{1/12 + 1/10} = 4 \pm 3.31,$$

$$0.69 < \mu_1 - \mu_2 < 7.31.$$

9.50 Two levels (low and high) of insulin doses are given to two groups of diabetic rats to check the insulin-binding capacity, yielding the following data:

Low dose:	$n_1 = 8$	$\bar{x}_1 = 1.98$	$s_1 = 0.51$
High dose:	$n_2 = 13$	$\bar{x}_2 = 1.30$	$s_2 = 0.35$

Assume that the variances are equal. Give a 95% confidence interval for the difference in the true average insulin-binding capacity between the two samples.

$$n_1 = 8, n_2 = 13, \bar{x}_1 = 1.98, \bar{x}_2 = 1.30, s_1 = 0.51, s_2 = 0.35,$$

$$s_p = 0.416.$$

$$(1.98 - 1.30) \pm (2.093)(0.416)\sqrt{1/8 + 1/13} = 0.68 \pm 0.39,$$

$$0.29 < \mu_1 - \mu_2 < 1.07.$$