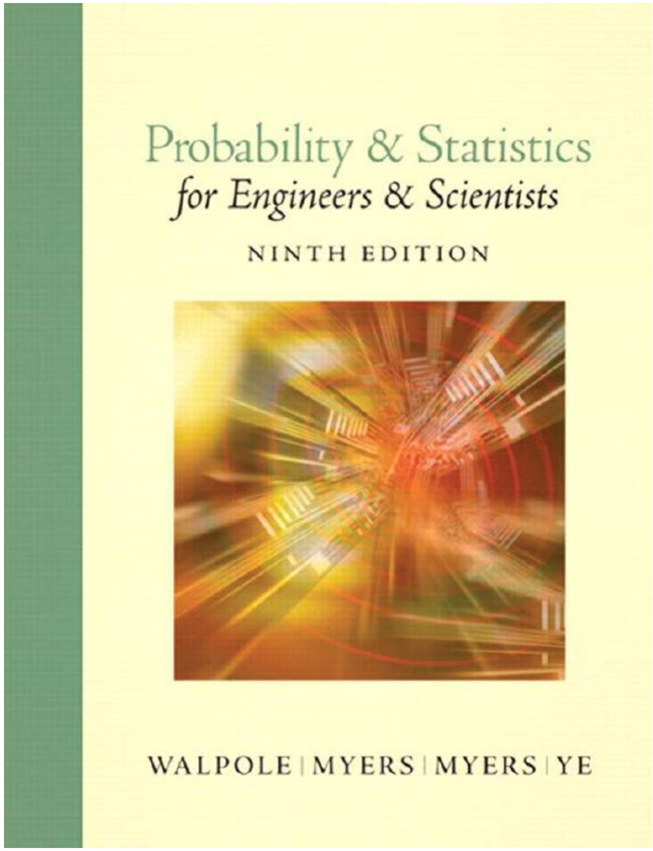
Statistical Analysis

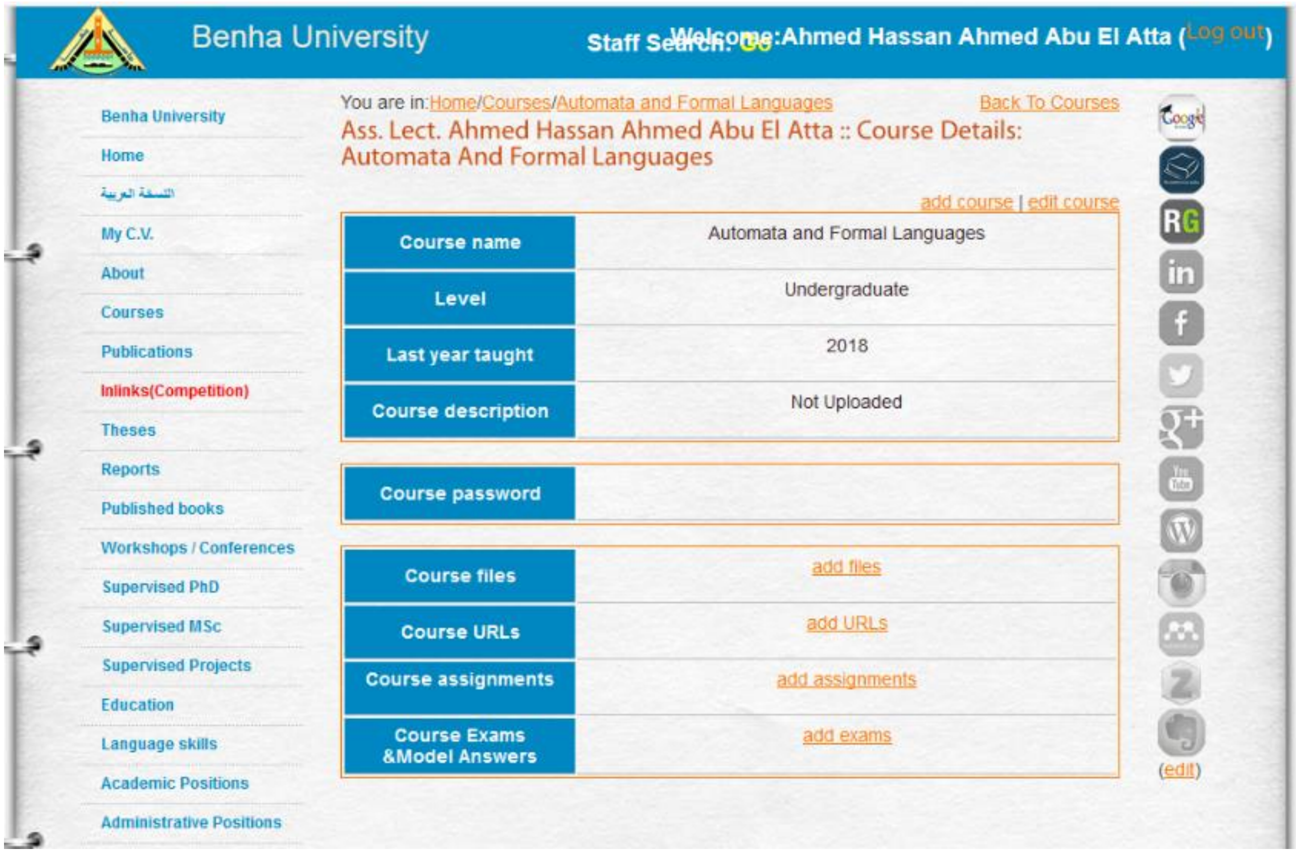
Lecture 04

Books


Probability & Statistics
for Engineers & Scientists
NINTH EDITION
WALPOLE | MYERS | MYERS | YE

PowerPoint

http://www.bu.edu.eg/staff/ahmedaboalatah14-courses/14767



Ex

8.59 If S_1^2 and S_2^2 represent the variances of independent random samples of size $n_1 = 8$ and $n_2 = 12$, taken from normal populations with equal variances, find $P(S_1^2/S_2^2 < 4.89)$.

$P\left(\frac{S_1^2}{S_2^2} < 4.89\right) = P\left(\frac{S_1^2/\sigma^2}{S_2^2/\sigma^2} < 4.89\right) = P(F < 4.89) = 0.99$, where F has 7 and 11 degrees of freedom.

One- and Two-Sample Estimation Problems

CHAPTER 9

Estimation Problems

Estimation of population parameters.

One and two samples

Estimation Types

A **point estimate** of some population parameter θ is a single value $\hat{\theta}$ of a statistic $\hat{\Theta}$. For example, the value $\bar{x}$ of the statistic $\bar{X}$, computed from a sample of size n, is a point estimate of the population parameter μ.

An interval estimate of a population parameter θ is an interval of the form $\hat{\theta}_L < \theta < \hat{\theta}_U$, where $\hat{\theta}_L$ and $\hat{\theta}_U$ depend on the value of the statistic $\hat{\Theta}$ for a particular sample and also on the sampling distribution of $\hat{\Theta}$.

Interpretation of Interval Estimates

If, for instance, we find $\hat{\Theta}_L$ and $\hat{\Theta}_U$ such that

$$P(\hat{\Theta}_L < \theta < \hat{\Theta}_U) = 1 - \alpha,$$

for $0 < \alpha < 1$, then we have a probability of $1-\alpha$ of selecting a random sample that will produce an interval containing θ. The interval $\hat{\theta}_L < \theta < \hat{\theta}_U$, computed from the selected sample, is called a $100(1-\alpha)\%$ **confidence interval**, the fraction $1-\alpha$ is called the **confidence coefficient** or the **degree of confidence**, and the endpoints, $\hat{\theta}_L$ and $\hat{\theta}_U$, are called the lower and upper **confidence limits**.

Thus, when $\alpha = 0.05$, we have a 95% confidence interval, and when $\alpha = 0.01$, we obtain a wider 99% confidence interval. The wider the confidence interval is, the more confident we can be that the interval contains the unknown parameter. Of course, it is better to be 95% confident that the average life of a certain television transistor is between 6 and 7 years than to be 99% confident that it is between 3 and 10 years.

9.4 Single Sample: Estimating the Mean

Point Estimation for the Mean

The sampling distribution of $\bar{X}$ is centered at μ, and in most applications the variance is smaller than that of any other estimators of μ. Thus, the sample mean $\bar{x}$ will be used as a point estimate for the population mean μ. Recall that $\sigma^2_{\bar{X}} = \sigma^2/n$, so a large sample will yield a value of $\bar{X}$ that comes from a sampling distribution with a small variance. Hence, $\bar{x}$ is likely to be a very accurate estimate of μ when n is large.

The Interval Estimation for the Mean

If our sample is selected from a normal population or if "n" is sufficiently large, we can establish a confidence interval for "μ" by considering the sampling distribution of $\bar{X}$.

The Interval Estimation for the Mean

According to the Central Limit Theorem, we can expect the sampling distribution of $\bar{X}$ to be approximately normally distributed with mean $\mu_{\bar{X}} = \mu$ and

standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n}$. Writing $z_{\alpha/2}$ for the z-value above which we find an area of $\alpha/2$ under the normal curve, we can see from Figure 9.2 that

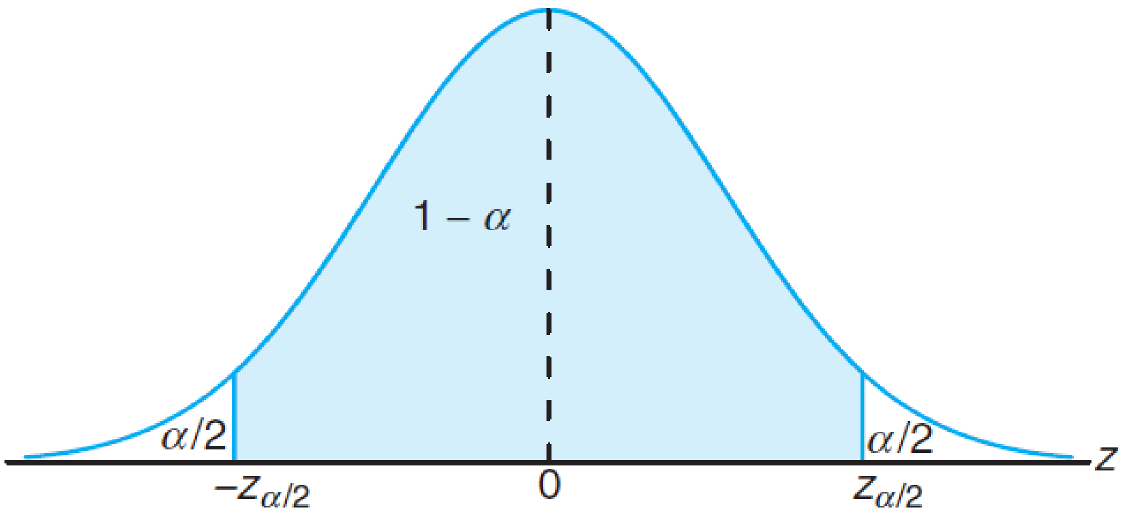
$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha,$$

where

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

The Interval Estimation for the Mean

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



The Interval Estimation for the Mean

Multiplying each term in the inequality by $\sigma/\sqrt{n}$ and then subtracting $\bar{X}$ from each term and multiplying by -1 (reversing the sense of the inequalities), we obtain

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Confidence Interval on μ, σ^2 Known

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.

Example 9.2:

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

$$2.47 < \mu < 2.73.$$

We now see that a longer interval is required to estimate μ with a higher degree of confidence.

The Point Estimation Error

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.

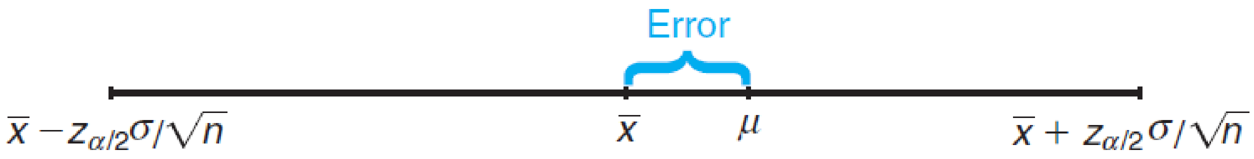


Figure 9.4: Error in estimating μ by $\bar{x}$.

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 $(2.575)(0.3)/\sqrt{36} = 0.13$.

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

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

Example 9.3:

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?

Example 9.3:

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

One-sided confidence bounds are developed in the same fashion as two-sided intervals. However, the source is a one-sided probability statement that makes use of the Central Limit Theorem:

$$P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_\alpha\right) = 1-\alpha.$$

One can then manipulate the probability statement much as before and obtain

$$P(\mu > \bar{X} - z_\alpha\sigma/\sqrt{n}) = 1-\alpha.$$

Similar manipulation of $P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > -z_\alpha\right) = 1-\alpha$ gives

$$P(\mu < \bar{X} + z_\alpha\sigma/\sqrt{n}) = 1-\alpha.$$

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

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

Example 9.4:

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.

Example 9.4:

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

Single Sample: Estimating the Mean σ Unknown

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

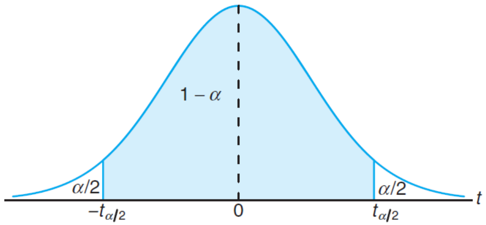


Figure 9.5: $P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha$.

$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha,$$

where $t_{\alpha/2}$ is the t-value with $n-1$ degrees of freedom, above which we find an area of $\alpha/2$. Because of symmetry, an equal area of $\alpha/2$ will fall to the left of $-t_{\alpha/2}$. Substituting for T, we write

$$P\left(-t_{\alpha/2} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}\right) = 1 - \alpha.$$

Multiplying each term in the inequality by $S/\sqrt{n}$, and then subtracting $\bar{X}$ from each term and multiplying by -1, we obtain

$$P\left(\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

For a particular random sample of size n, the mean $\bar{x}$ and standard deviation s are computed and the following $100(1-\alpha)\%$ confidence interval for μ is obtained.

Confidence Interval on μ, σ^2 Unknown

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.

One-Sided Confidence Bounds

Computed one-sided confidence bounds for μ with σ unknown are as the reader would expect, namely

$$\bar{x} + t_\alpha \frac{s}{\sqrt{n}} \qquad \text{and} \qquad \bar{x} - t_\alpha \frac{s}{\sqrt{n}}.$$

They are the upper and lower $100(1-\alpha)\%$ bounds, respectively. Here t_α is the t-value having an area of α to the right.

Example 9.5:

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.

Example 9.5:

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. Hence, the

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 $9.74 < \mu < 10.26$.