MA026IU
PROBABILITY, STATISTIC AND RANDOM PROCESS

Part 2 B

Point Estimation of Parameters and Statistical Intervals for a Single Sample

I. Point Estimation

- A **point estimate** is a reasonable value of a population parameter.
- $X_1, X_2, \ldots, X_n$ are **random variables**.
- Functions of these random variables, $\bar{x}$ and s^2, are also random variables called **statistics**.
- Statistics have their unique distributions which are called **sampling distributions**.

Point Estimator

Point Estimator

A **point estimate** of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic $\hat{\Theta}$. The statistic $\hat{\Theta}$ is called the **point estimator**.

As an example, suppose that the random variable X is normally distributed with an unknown mean μ. The sample mean is a point estimator of the unknown population mean μ. That is, $\hat{\mu} = \overline{X}$. After the sample has been selected, the numerical value $\overline{x}$ is the point estimate of μ. Thus, if $x_1 = 25$, $x_2 = 30$, $x_3 = 29$, and $x_4 = 31$, the point estimate of μ is

$$\overline{x} = \frac{25 + 30 + 29 + 31}{4} = 28.75$$

Some Parameters & Their Statistics

Estimation problems occur frequently in engineering. We often need to estimate

- The mean μ of a single population
- The variance σ^2 (or standard deviation σ) of a single population
- The proportion p of items in a population that belong to a class of interest
- The difference in means of two populations, $\mu_1 - \mu_2$
- The difference in two population proportions, $p_1 - p_2$

Reasonable point estimates of these parameters are as follows:

- For μ, the estimate is $\hat{\mu} = \overline{x}$, the sample mean.
- For σ^2, the estimate is $\hat{\sigma}^2 = s^2$, the sample variance.
- For p, the estimate is $\hat{p} = x/n$, the sample proportion, where x is the number of items in a random sample of size n that belong to the class of interest.
- For $\mu_1 - \mu_2$, the estimate is $\hat{\mu}_1 - \hat{\mu}_2 = \overline{x}_1 - \overline{x}_2$, the difference between the sample means of two independent random samples.
- For $p_1 - p_2$, the estimate is $\hat{p}_1 - \hat{p}_2$, the difference between two sample proportions computed from two independent random samples.

- There could be choices for the point estimator of a parameter.
- To estimate the mean of a population, we could choose the:
  - Sample mean.
  - Sample median.
  - Average of the largest & smallest observations in the sample.

Some Definitions

- The random variables $X_1, X_2, \ldots, X_n$ are a **random sample** of size n if:
  a) The $X_i's$ are independent random variables.
  b) Every X_i has the same probability distribution.
- A **statistic** is any function of the observations in a random sample.
- The probability distribution of a statistic is called a **sampling distribution**.

Central Limit Theorem

Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2 and if $\overline{X}$ is the sample mean, the limiting form of the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{7.1}$$

as $n \to \infty$, is the standard normal distribution.

Example | Central Limit Theorem

Suppose that a random variable X has a continuous uniform distribution

$$f(x) = \begin{cases} 1/2, & 4 \le x \le 6 \\ 0, & \text{otherwise} \end{cases}$$

Find the distribution of the sample mean of a random sample of size $n = 40$.

The mean and variance of X are $\mu = 5$ and $\sigma^2 = (6-4)^2/12 = 1/3$. The central limit theorem indicates that the distribution of $\overline{X}$ is approximately normal with mean $\mu_{\overline{X}} = 5$ and variance $\sigma^2_{\overline{X}} = \sigma^2/n = 1/[3(40)] = 1/120$. See the distributions of X and $\overline{X}$ in Figure 7.5.

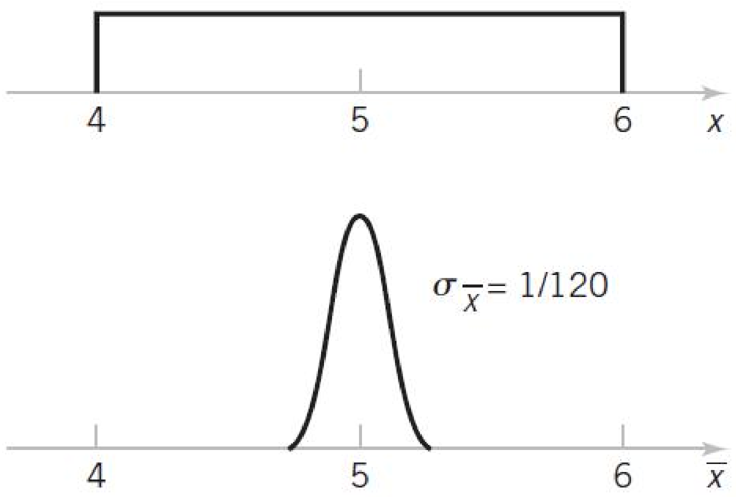

FIGURE 7.5

The distribution of X and $\overline{X}$ for Example 7.2.

Sampling Distribution of a Difference in Sample Means

Approximate Sampling Distribution of a Difference in Sample Means

If we have two independent populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 and if $\overline{X}_1$ and $\overline{X}_2$ are the sample means of two independent random samples of sizes n_1 and n_2 from these populations, then the sampling distribution of

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \tag{7.4}$$

is approximately standard normal if the conditions of the central limit theorem apply. If the two populations are normal, the sampling distribution of Z is exactly standard normal.

Unbiased Estimators

Bias of an Estimator

The point estimator $\hat{\Theta}$ is an **unbiased estimator** for the parameter θ if

$$E(\hat{\Theta}) = \theta \tag{7.5}$$

If the estimator is not unbiased, then the difference

$$E(\hat{\Theta}) - \theta \tag{7.6}$$

is called the **bias** of the estimator $\hat{\Theta}$.

When an estimator is unbiased, the bias is zero; that is, $E(\hat{\Theta}) - \theta = 0$.

Example | Sample Mean and Variance are Unbiased

- Suppose X is a random variable with mean μ and variance σ^2. Let $X_1, X_2, \dots, X_n$ be a random sample of size n from the population represented by X
- Show that the sample mean $\bar{X}$ and sample variance S^2 are unbiased estimators of μ and σ^2 respectively.

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right)$$

$$= \frac{1}{n}\left[E(X_1) + E(X_2) + \dots + E(X_n)\right]$$

$$= \frac{1}{n}[\mu + \mu + \dots + \mu] = \frac{n\mu}{n} = \mu$$

Now consider the sample variance. We have

$$E(S^2) = E\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}\right] = \frac{1}{n-1} E \sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$= \frac{1}{n-1} E \sum_{i=1}^{n}(X_i^2 + \bar{X}^2 - 2\bar{X}X_i)$$

$$= \frac{1}{n-1} E\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i^2) - nE(\bar{X}^2)\right]$$

The last equality follows the equation for the mean of a linear function in Chapter 5. However, because $E(X_i^2) = \mu^2 + \sigma^2$ and $E(\bar{X}^2) = \mu^2 + \sigma^2/n$, we have

$$E(S^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n}(\mu^2 + \sigma^2) - n(\mu^2 + \sigma^2/n)\right]$$

$$= \frac{1}{n-1}(n\mu^2 + n\sigma^2 - n\mu^2 - \sigma^2) = \sigma^2$$

Therefore, the sample variance S^2 is an unbiased estimator of the population variance σ^2.

Variance of a Point Estimator

Minimum Variance Unbiased Estimator

If we consider all unbiased estimators of θ, the one with the smallest variance is called the **minimum variance unbiased estimator** (MVUE).

If $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normal distribution with mean μ and variance σ^2, the sample mean $\overline{X}$ is the MVUE for μ.

Standard Error: Reporting a Point Estimate

Standard Error of an Estimator

The **standard error** of an estimator $\hat{\Theta}$ is its standard deviation given by $\sigma_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})}$. If the standard error involves unknown parameters that can be estimated, substitution of those values into $\sigma_{\hat{\Theta}}$ produces an **estimated standard error**, denoted by $\hat{\sigma}_{\hat{\Theta}}$.

Suppose that we are sampling from a normal distribution with mean μ and variance σ^2. Now the distribution of $\overline{X}$ is normal with mean μ and variance σ^2/n, so the **standard error** of $\overline{X}$ is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

If we did not know σ but substituted the sample standard deviation S into the preceding equation, the **estimated standard error** of $\overline{X}$ would be

$$SE(\overline{X}) = \hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n}}$$

Example | Thermal Conductivity

- The following 10 measurements of thermal conductivity of Armco iron were obtained:

$$41.60, 41.48, 42.34, 41.95, 41.86,$$

$$42.18, 41.72, 42.26, 41.81, 42.04$$

- A point estimate of the mean thermal conductivity at 100 °F and 550

$$\bar{x} = 41.924 \text{ Btu/hr-ft-}^\circ\text{F}$$

watts is the sample mean or

$$SE(\overline{X}) = \hat{\sigma}_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{0.284}{\sqrt{10}} = 0.0898$$

- Because σ^2 is unknown, we may replace it with the standard deviation s = 0.284 to obtain the estimated standard error of $\bar{x}$ as:

II. Confidence Interval & Properties

A **confidence interval** estimate for μ is an interval of the form

$$P\{L \leq \mu \leq U\} = 1 - \alpha \qquad (8.2)$$

where the end-points L and U are computed from the sample data.

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

- The endpoints or bounds L and U are called lower- and upper-confidence limits and is called the confidence coefficient.

Confidence Interval on the Mean, Variance Known

Confidence Interval on the Mean, Variance Known

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

$$\overline{x} - z_{\alpha/2}\sigma/\sqrt{n} \le \mu \le \overline{x} + z_{\alpha/2}\sigma/\sqrt{n} \tag{8.5}$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.

Example 1 | Metallic Material Transition

- Ten measurements of impact energy (*J*) on specimens of A238 steel cut at 60$\circ$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. The impact energy is normally distributed with $\sigma = 1J$. Find a 95% CI for μ, the mean impact energy.
- The required quantities are $z_{\alpha/2} = z_{0.025} = 1.96$, $n = 10$, $\sigma = 1$, and $\bar{x} = 64.46$. The resulting 95% CI is found from Equation 8-1 as follows:

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

$$64.46 - 1.96\frac{1}{\sqrt{10}} \leq \mu \leq 64.46 + 1.96\frac{1}{\sqrt{10}}$$

$$63.84 \leq \mu \leq 65.08$$

- **<u>Interpretation:</u>** Based on the sample data, a range of highly plausible values for mean impact energy for A238 steel at 60$\circ$C is $63.84J \leq \mu \leq 65.08J$

Choice of Sample Size

Sample Size for Specified Error on the Mean, Variance Known

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

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

Example 2 | Metallic Material Transition

- Consider the CVN test described in Example 1.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*.
- The bound on error in estimation *E* is one-half of the length of the CI.
- Use Equation 8.6 to determine n with $E = 0.5$, $\sigma = 1$, and $z_{\alpha/2} = 1.96$.

$$n = \left(\frac{z_{\alpha/2}\,\sigma}{E}\right)^2 = \left[\frac{(1.96)1}{0.5}\right]^2 = 15.37$$

- Since, n must be an integer, the required sample size is $n = 16$.

One-sided Confidence Bounds

One-Sided Confidence Bounds on the Mean, Variance Known

A $100(1-\alpha)\%$ **upper-confidence bound** for μ is

$$\mu \le \bar{x} + z_{\alpha}\sigma/\sqrt{n} \tag{8.7}$$

and a $100(1-\alpha)\%$ **lower-confidence bound** for μ is

$$\bar{x} - z_{\alpha}\sigma/\sqrt{n} \le \mu \tag{8.8}$$

Example 3 | One-sided Confidence Bound

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

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

$$64.46 - 1.64\ \frac{1}{\sqrt{10}} \leq \mu$$

$$63.94 \leq \mu$$

Large-sample Confidence Interval for μ

Large-Sample Confidence Interval on the Mean

When n is large, the quantity

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

has an approximate standard normal distribution. Consequently,

$$\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}} \qquad (8.11)$$

is a large-sample confidence interval for μ, with confidence level of approximately $100(1 - \alpha)\%$.

Example 4 | Mercury Contamination

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

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

- Find an approximate 95% CI on μ.

Example 4 | Mercury Contamination (ctd.)

- The summary statistics for these data are as follows:

Variable	N	Mean	Median	StDev	Minimum	Maximum	$Q1$	$Q3$
Concentration	53	0.5250	0.4900	0.3486	0.0400	1.3300	0.2300	0.7900

- Because $n > 40$, the assumption of normality is not necessary to use in Equation 8.11. The required values are $n = 53$, $\bar{x} = 0.5250$, $s = 0.3586$, and $z_{0.025} = 1.96$.

$$\bar{x} - z_{0.025}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + z_{0.025}\frac{s}{\sqrt{n}}$$

$$0.5250 - 1.96\,\frac{0.3486}{\sqrt{53}} \le \mu \le 0.5250 + 1.96\,\frac{0.3486}{\sqrt{53}}$$

$$0.4311 \le \mu \le 0.6189$$

Interpretation: This interval is fairly wide because there is variability in the mercury concentration measurements. A larger sample size would have produced a shorter interval.

Large-sample Approximate Confidence Interval

Suppose that θ is a parameter of a probability distribution, and let $\hat{\theta}$ be an estimator of θ. Then a large-sample approximate CI for θ is given by

Large-Sample Approximate Confidence Interval

$$\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\Theta}} \le \theta \le \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\Theta}} \tag{8.12}$$

The t distribution

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

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \tag{8.13}$$

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

The Confidence Interval on Mean, Variance Unknown

Confidence Interval on the Mean, Variance Unknown

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 $100(1 - \alpha)\%$ confidence interval on μ is given by

$$\bar{x} - t_{\alpha/2,n-1}s/\sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2,n-1}s/\sqrt{n} \tag{8.16}$$

where $t_{\alpha/2,n-1}$ is the upper $100\alpha/2$ percentage point of the t distribution with $n - 1$ degrees of freedom.

Example 5 | Alloy Adhesion

- The load at specimen failure is as follows (in megapascals):

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	

- Construct a 95% CI on μ.
- The sample mean is $\bar{x} = 13.71$ and sample standard deviation is $s = 3.55$. Since $n = 22$, we have $n \quad 1 = 21$ degrees of freedom for t, so $t_{0.025,21} = 2.080$.

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

$$13.71 - 2.080(3.55)/\sqrt{22} \le \mu \le 13.71 + 2.080(3.55)/\sqrt{22}$$

$$13.71 - 1.57 \le \mu \le 13.71 + 1.57$$

$$12.14 \le \mu \le 15.28$$

Interpretation: The CI is fairly wide because there is a lot of variability in the measurements. A larger sample size would have led to a shorter interval.

Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

χ^2 Distribution

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

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \tag{8.17}$$

has a chi-square (χ^2) distribution with $n - 1$ degrees of freedom.

The probability density function of a χ^2 random variable is

$$f(x) = \frac{1}{2^{k/2}\,\Gamma(k/2)}\, x^{(k/2)-1} e^{-x/2} \qquad x > 0$$

Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

Confidence Interval on the Variance

If s^2 is the sample variance from a random sample of n observations from a normal distribution with unknown variance σ^2, then a $100(1-\alpha)\%$ confidence interval on σ^2 is

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} \tag{8.19}$$

where $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-\alpha/2,n-1}$ are the upper and lower $100\alpha/2$ percentage points of the chi-square distribution with $n-1$ degrees of freedom, respectively. A confidence interval for σ has lower and upper limits that are the square roots of the corresponding limits in Equation 8.19.

One-Sided Confidence Bounds on the Variance

The $100(1-\alpha)\%$ lower and upper confidence bounds on σ^2 are

$$\frac{(n-1)s^2}{\chi^2_{\alpha,n-1}} \leq \sigma^2 \quad \text{and} \quad \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha,n-1}} \tag{8.20}$$

respectively.

Example 6 | Detergent Filling

- An automatic filling machine is used to fill bottles with liquid detergent. A random sample of 20 bottles results in a sample variance of fill volume of $s^2 = 0.0153^2$. Assume that the fill volume is approximately normal. Compute a 95% upper confidence bound.
- A 95% upper confidence bound is f ows

$$\sigma^2 \le \frac{(n-1)s^2}{\chi^2_{0.95,19}} \qquad \sigma^2 \le \frac{(19)0.0153}{10.117} = 0.0287 \text{ (fluid ounce)}^2$$

$$\sigma \le 0.17$$

- A confidence interval on the standard deviation σ can be obtained by taking the square root on both sides, resulting in
- ***Practical Interpretation:*** Therefore, at the 95% level of confidence, the data indicates that the process standard deviation could be as large as 0.17 fluid ounce. The process engineer or manager now needs to determine whether a standard deviation this large could lead to an operational problem with under-or-over-filled bottles

Large-Sample Confidence Interval for a Population Proportion

Normal Approximation for a Binomial Proportion

If n is large, the distribution of

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

is approximately standard normal.

To construct the confidence interval on p, note that

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) \simeq 1 - \alpha$$

so

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

This may be rearranged as

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

Large-Sample Confidence Interval for a Population Proportion

Approximate Confidence Interval on a Binomial Proportion

If $\hat{p}$ is the proportion of observations in a random sample of size n that belongs to a class of interest, an approximate $100(1-\alpha)\%$ confidence interval on the proportion p of the population that belongs to this class is

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad (8.23)$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution.

Example 7 | Crankshaft Bearings

- In a random sample of 85 automobile engine crankshaft bearings, 10 have a surface finish that is rougher than the specifications allow. Construct a 95% two-sided confidence interval for *p.*
- A 95% two-sided confidence interval for *p* is computed from Equation 8-11 as

$$\hat{p} - z_{0.025}\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} \le p \le \hat{p} + z_{0.025}\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}$$

$$0.12 - 1.96\sqrt{\frac{0.12\left(0.88\right)}{85}} \le p \le 0.12 + 1.96\sqrt{\frac{0.12\left(0.88\right)}{85}}$$

$$0.0509 \le p \le 0.2243$$

- *Practical Interpretation:* This is a wide CI. Although the sample size does not appear to be small (*n* = 85), the value of $\hat{p}$ is fairly small, which leads to a large standard error for contributing to the wide CI.

Choice of Sample Size

Sample Size for a Specified Error on a Binomial Proportion

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p) \tag{8.24}$$

The sample size from Equation 8.24 will always be a maximum for *p = 0.5* [that is, *p(1 − p)* ≤ *0.25* with equality for *p = 0.5*], and can be used to obtain an upper bound on *n.*

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 (0.25) \tag{8.25}$$

Example 8 | Crankshaft Bearings

- Consider the situation in Example 7. How large a sample is required if we want to be 95% confident that the error in using $\hat{p}$ to estimate *p* is less than 0.05?

- Using $\hat{p} = 0.12$ as an initial estimate of *p*, we find from Equation 8-12 that the required sample size is

$$n = \left(\frac{z_{0.025}}{E} \right)^2 \hat{p}(1 - \hat{p}) = \left(\frac{1.96}{0.05} \right)^2 0.12(0.88) \cong 163$$

- If we wanted to be *at least* 95% confident that our estimate $\hat{p}$ of the true proportion *p* was within 0.05 regardless of the value of *p*, we would use Equation 8-13 to find the sample size

$$n = \left(\frac{z_{0.025}}{E} \right)^2 (0.25) = \left(\frac{1.96}{0.05} \right)^2 (0.25) \cong 385$$

- ***Practical Interpretation***: If we have information concerning the value of *p*, either from a preliminary sample or from past experience, we could use a smaller sample while maintaining both the desired precision of estimation and the level of confidence.

Approximate One-Sided Confidence Bounds on a Binomial Proportion

Approximate One-Sided Confidence Bounds on a Binomial Proportion

The approximate $100(1-\alpha)\%$ lower and upper confidence bounds are

$$\hat{p} - z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \qquad \text{and} \qquad p \leq \hat{p} + z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{8.26}$$

respectively.

Guidelines for Constructing Confidence Intervals

- Most difficult step in constructing a confidence interval is often the match of the appropriate calculation to the objective of the study

- Two primary comments can help identify the analysis
  1. Determine the parameter (and the distribution of the data) that will be bounded by the confidence interval or tested by the hypothesis
  2. Check if other parameters are known or need to be estimated

Guidelines for Constructing Confidence Intervals

TABLE 8.1 The Roadmap for Constructing Confidence Intervals and Performing Hypothesis Tests, One-Sample Case

Parameter to Be Bounded by the Confidence Interval or Tested with a Hypothesis?	Symbol	Other Parameters?	Confidence Interval Section	Hypothesis Test Section	Comments
Mean of normal distribution	μ	Standard deviation σ known	8.1	9.2	Large sample size is often taken to be $n \geq 40$
Mean of arbitrary distribution with large sample size	μ	Sample size large enough that central limit theorem applies and σ is essentially known	8.1.5	9.2.3	
Mean of normal distribution	μ	Standard deviation σ unknown and estimated	8.2	9.3	
Variance (or standard deviation) of normal distribution	σ^2	Mean μ unknown and estimated	8.3	9.4	
Population proportion	p	None	8.4	9.5	