CH.8 Statistical Intervals for a Single Sample

- Introduction
- Confidence interval on the mean of a normal distribution, variance known
- Confidence interval on the mean of a normal distribution, variance unknown
- Confidence interval on the variance and standard deviation of a normal distribution
- Large-sample confidence interval for a population proportion
- Tolerance and prediction intervals

8-1 Introduction

- In the previous chapter we illustrated how a parameter can be estimated from sample data. However, it is important to understand how good is the estimate obtained.

- Bounds that represent an interval of plausible values for a parameter are an example of an interval estimate.

- Three types of intervals will be presented:
    - Confidence intervals
    - Prediction intervals
    - Tolerance intervals

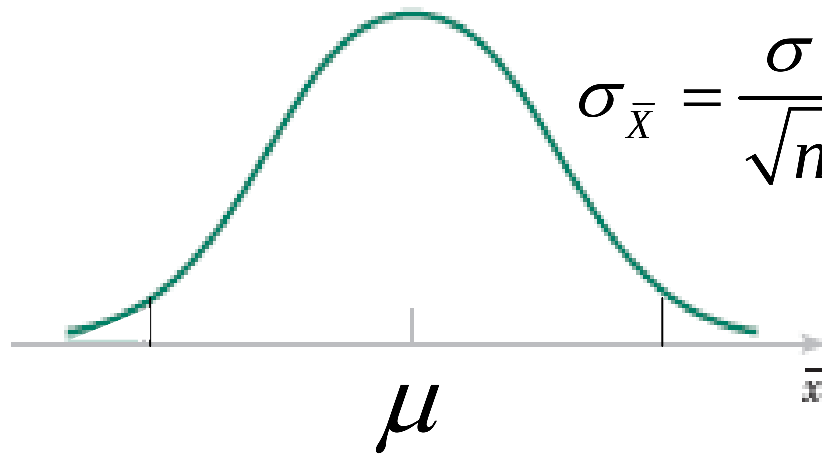
8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

8-2.1 Development of the Confidence Interval and its Basic Properties

POPULATION
$N(\mu,\sigma^2)$
σ^2: known
μ : unknown

Sample $X_1, X_2, \ldots, X_n$

$$\bar{X} \sim N\left(\mu, \sigma^2/n\right)$$



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

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

has a standard normal distribution

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

8-2.1 Development of the Confidence Interval and its Basic Properties

A **confidence interval** estimate for μ is an interval of the form $l \leq \mu \leq u$, where the end-points 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:

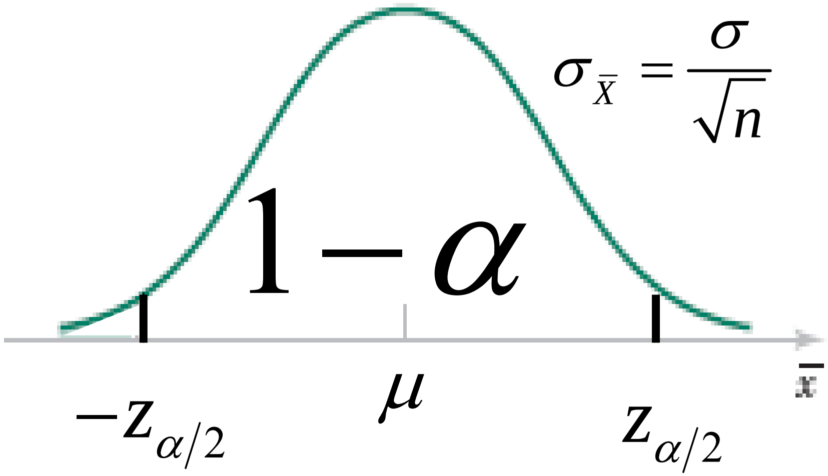
$$P\{L \leq \mu \leq U\} = 1 - \alpha \qquad (8\text{-}4)$$

where $0 \leq \alpha \leq 1$. There is a probability of $1 - \alpha$ of selecting a sample for which the CI will contain the true value of μ. Once we have selected the sample, so that $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, and computed l and u, the resulting **confidence interval** for μ is

$$l \leq \mu \leq u \qquad (8\text{-}5)$$

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

8-2.1 Development of the Confidence Interval and its Basic Properties



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

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

$$P\left\{\bar{X} - z_{\alpha/2}\,\sigma/\sqrt{n} \le \mu \le \bar{X} + z_{\alpha/2}\,\sigma/\sqrt{n}\right\} = 1-\alpha$$

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

8-2.1 Development of the Confidence Interval and its Basic Properties

Definition

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

$$\overline{x} - z_{\alpha/2}\sigma/\sqrt{n} \le \mu \le \overline{x} + z_{\alpha/2}\sigma/\sqrt{n} \qquad (8\text{-}7)$$

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

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-1

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 $\sigma = 1J$. We want to find a 95% CI for μ, the mean impact energy.

$n = 10$

$\sigma = 1$

$\bar{x} = 64.46$

$\alpha = 0.05$

$z_{\alpha/2} = z_{0.025} = 1.96$

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-1

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 $\sigma = 1J$. We want to 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 8-7 as follows:

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

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

$$63.84 \le \mu \le 65.08$$

That is, based on the sample data, a range of highly plausible vaules for mean impact energy for A238 steel at 60°C is $63.84J \le \mu \le 65.08J$.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

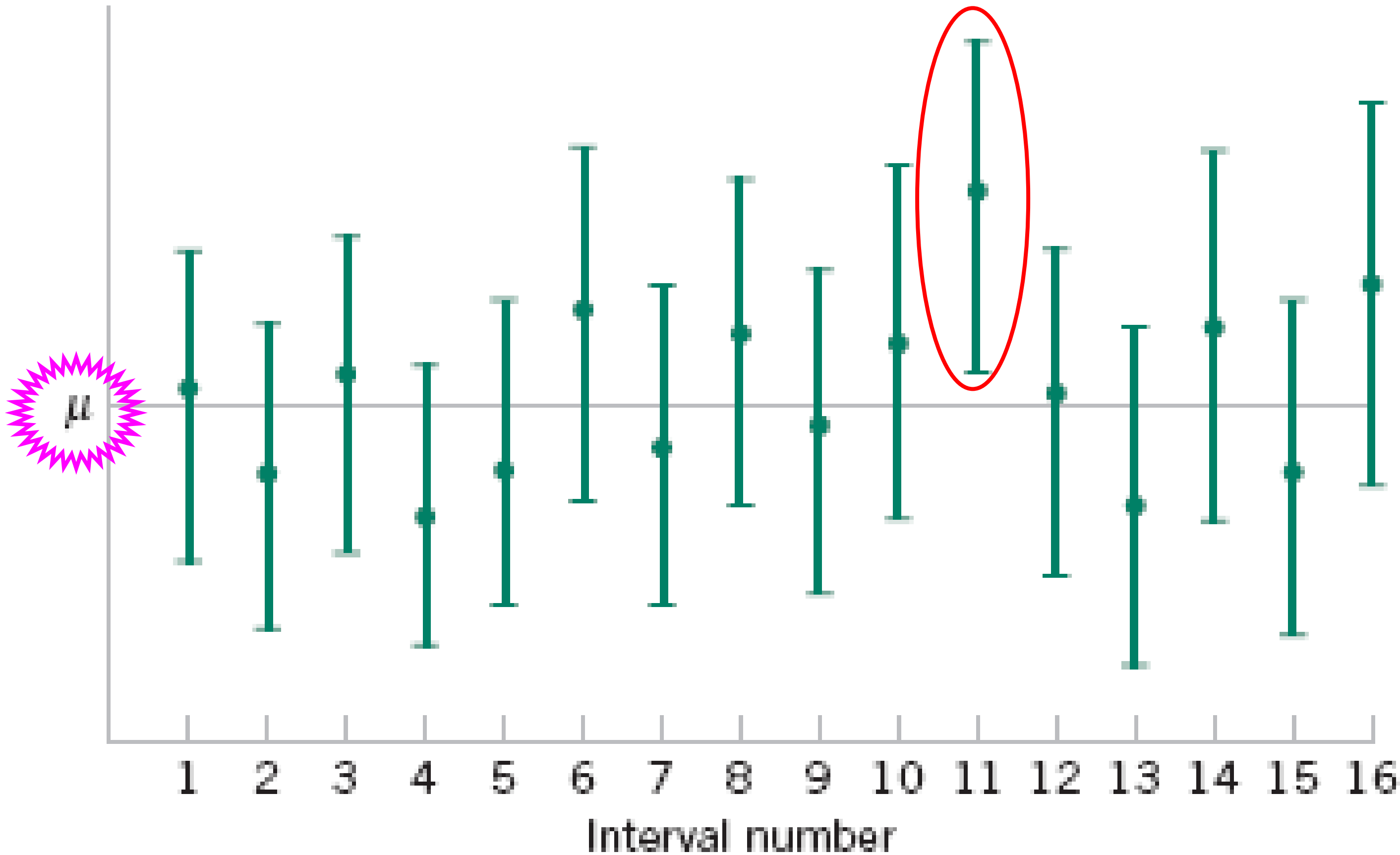
Interpreting a Confidence Interval

- The confidence interval is a random interval
- The appropriate interpretation of a confidence interval (for example on μ) is:

The observed interval [l, u] brackets the true value of μ, with confidence $100(1-\alpha)\%$.

- Examine the figure on the next slide.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known



Repeated construction of a confidence interval for μ.

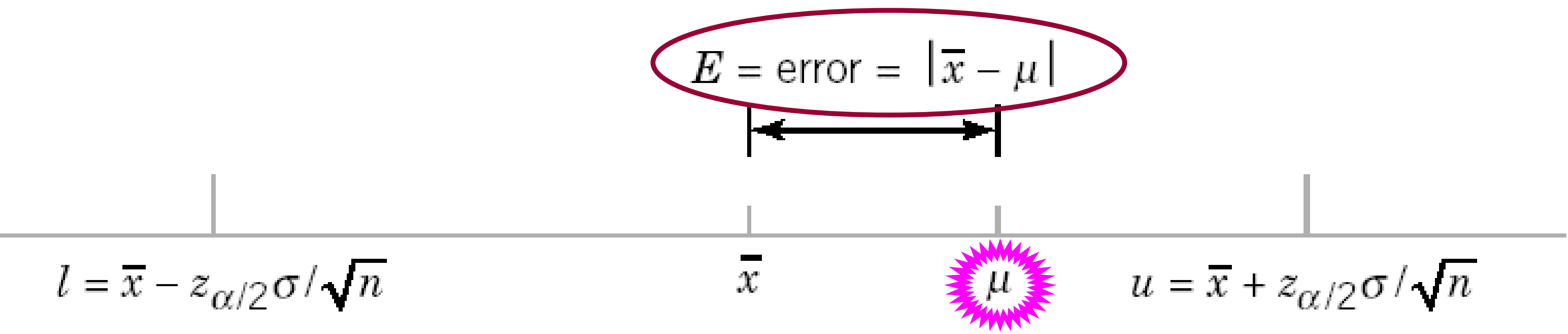
A 95% CI means in the long run only 5% of the intervals would fail to contain μ.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Confidence Level and Precision of Estimation

For fixed n and σ, the higher the confidence level, the longer the resulting confidence interval.

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



By using $\bar{x}$ to estimate μ, the error E is less than or equal to $z_{\alpha/2}\,\sigma/\sqrt{n}$ with confidence 100(1-α)%.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

8-2.2 Choice of Sample Size

Bound on error is $\left[E \geq z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right]$

Choose n to be large enough to assure a predetermined CI and precision !

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

$$n = \left(\frac{z_{\alpha/2}\sigma}{E} \right)^2 \qquad (8\text{-}8)$$

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-2

To illustrate the use of this procedure, consider the CVN test described in Example 8-1, and suppose that we wanted 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*.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-2

To illustrate the use of this procedure, consider the CVN test described in Example 8-1, and suppose that we wanted 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 $1.0J$. Since the bound on error in estimation E is one-half of the length of the CI, to determine n we use Equation 8-8 with $E = 0.5$, $\sigma = 1$, and $z_{\alpha/2} = 0.025$. The required sample size is 16

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

and because n must be an integer, the required sample size is $n = 16$.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

8-2.3 One-Sided Confidence Bounds

Definition

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

$$\mu \le u = \bar{x} + z_{\alpha}\sigma/\sqrt{n} \qquad (8\text{-}9)$$

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

$$\bar{x} - z_{\alpha}\sigma/\sqrt{n} = l \le \mu \qquad (8\text{-}10)$$

Attention!

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-3

The same data for impact testing from example 8.1.

Construct a lower, one sided 95% CI for the mean impact energy

$$z_\alpha = z_{0.05} = 1.64$$

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

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

$$63.94 \le \mu$$

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

8-2.5 A Large-Sample Confidence Interval for μ

-Holds regardless of the population distribution
-Additional variability because of replacing σ by S
-So **n>=40**

Definition

When n is large, the quantity

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

has an approximate standard normal distribution. Consequently,

$$\overline{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \overline{x} + z_{\alpha/2}\frac{s}{\sqrt{n}} \qquad (8\text{-}13)$$

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

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-4

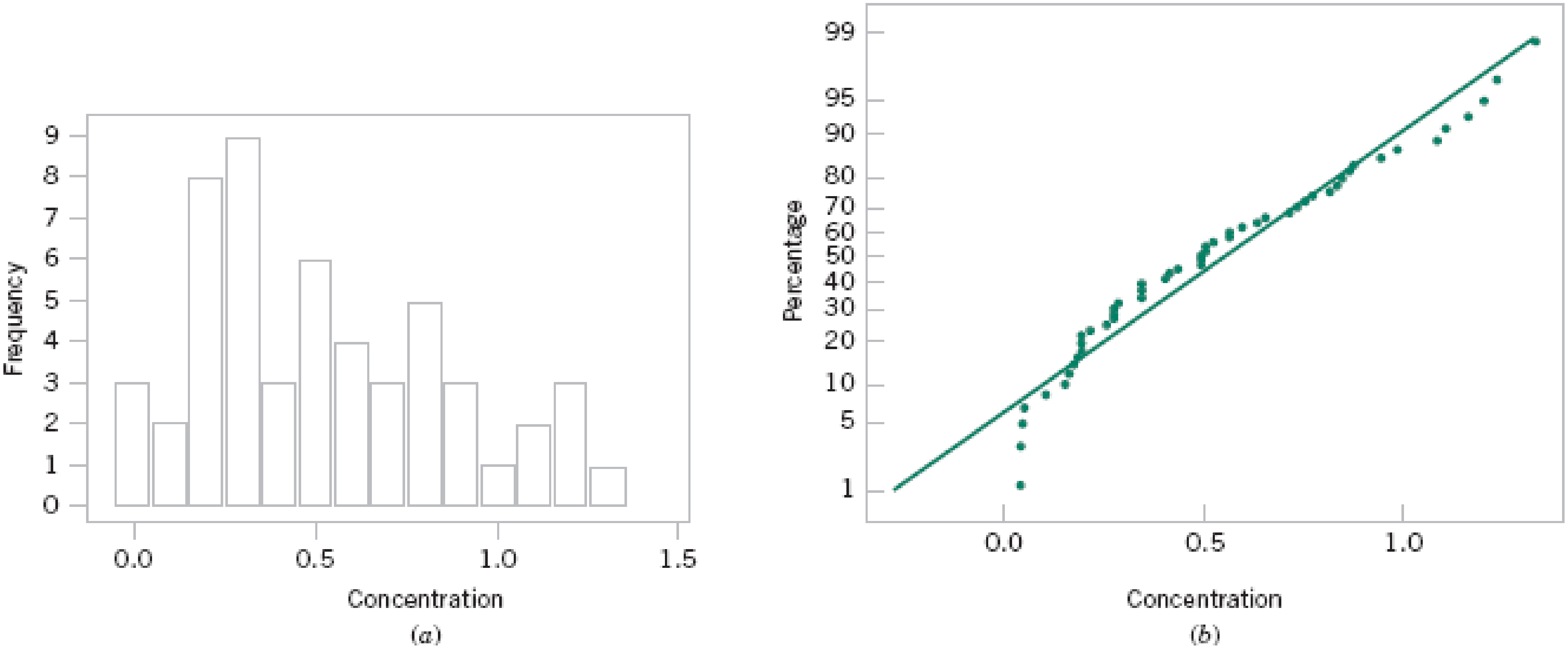
An article in the 1993 volume of the *Transactions of the American Fisheries Society* reports the results of a study to investigate the mercury contamination in largemouth bass. A sample of fish was selected from 53 Florida lakes and mercury concentration in the muscle tissue was measured (ppm). The 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.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-4 (continued)



Mercury concentration in largemouth bass

(a) Histogram. (b) Normal probability plot

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-4 (continued)

The summary statistics from Minitab are displayed below:

Descriptive Statistics: Concentration

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Concentration	53	0.5250	0.4900	0.5094	0.3486	0.0479

Variable	Minimum	Maximum	Q1	Q3
Concentration	0.0400	1.3300	0.2300	0.7900

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-4 (continued)

Figure 8-3(a) and (b) presents the histogram and normal probability plot of the mercury concentration data. Both plots indicate that the distribution of mercury concentration is not normal and is positively skewed. We want to find an approximate 95% CI on μ. Because $n > 40$, the assumption of normality is not necessary to use Equation 8-13. The required quantities are $n = 53$, $\bar{x} = 0.5250$, $s = 0.3486$, and $z_{0.025} = 1.96$. The approximate 95% CI on μ is

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

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

$$0.4311 \leq \mu \leq 0.6189$$

This interval is fairly wide because there is a lot of variability in the mercury concentration measurements.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

A General Large Sample Confidence Interval

If the estimator of a population parameter Θ
- has an approximate normal distribution
- is approximately unbiased for Θ
- has standard deviation that can be estimated from the sample data

Then the estimator has an approximate normal distribution and a large-sample approximate CI for Θ is as follows

$$\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\Theta}} \leq \theta \leq \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\Theta}} \qquad (8\text{-}14)$$

8-3 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

8-3.1 The *t* distribution

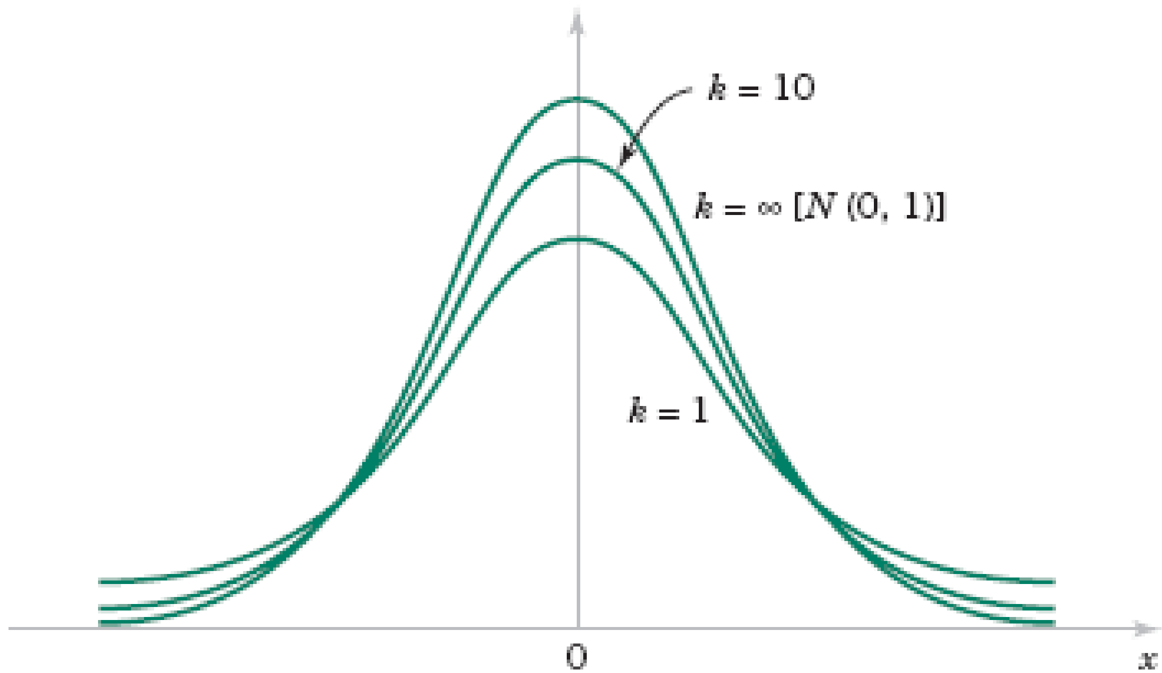
Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. The random variable

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \tag{8-15}$$

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

8-3 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

8-3.1 The *t* distribution

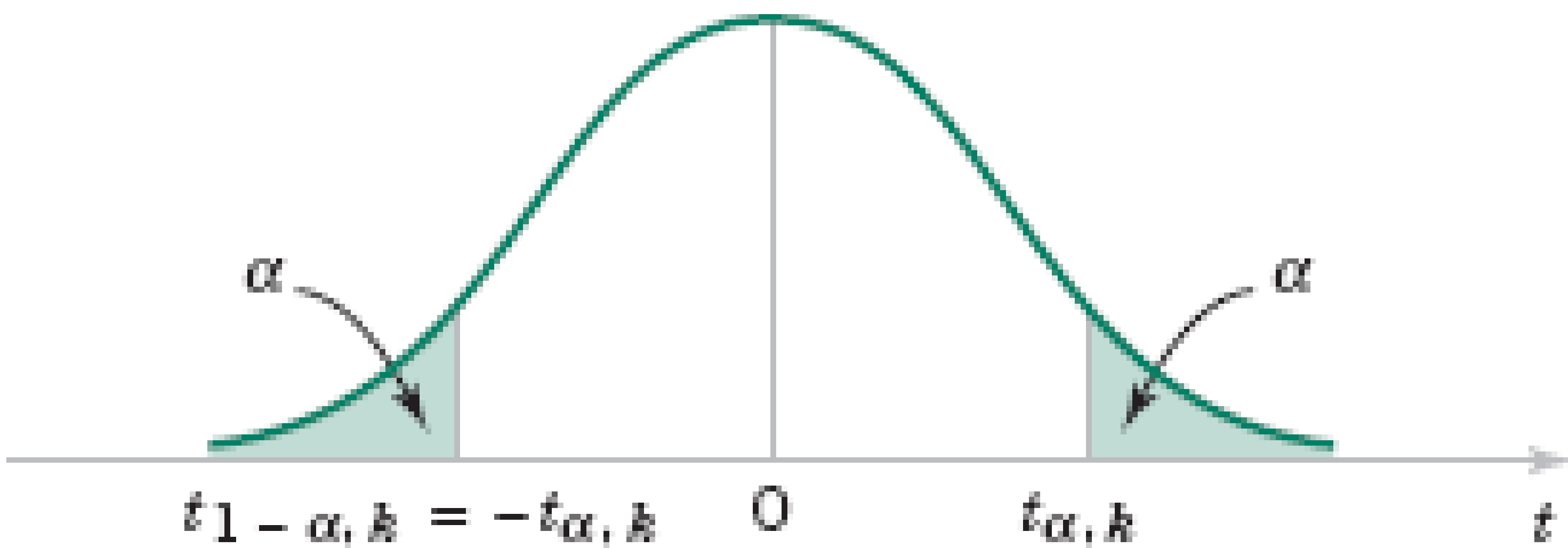


Probability density functions of several *t* distributions: symmetric, unimodal.

t distribution has more probability in the tails than the normal distribution.

8-3 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

8-3.1 The *t* distribution



Percentage points of the *t* distribution.

$t_{1-\alpha,k} = -t_{\alpha,k}$

8-3 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

8-3.2 The *t* Confidence Interval on μ

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 − α) percent confidence interval on μ** is given by

$$\bar{x} - t_{\alpha/2,n-1}s/\sqrt{n} \le \mu \le \bar{x} + t_{\alpha/2,n-1}s/\sqrt{n} \qquad (8\text{-}18)$$

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

One-sided confidence bounds on the mean are found by replacing $t_{\alpha/2,n-1}$ in the equation with $t_{\alpha,n-1}$.

8-3 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

Example 8-5

An article in the journal *Materials Engineering* (1989, Vol. II, No. 4, pp. 275–281) describes the results of tensile adhesion tests on 22 U-700 alloy specimens. 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	

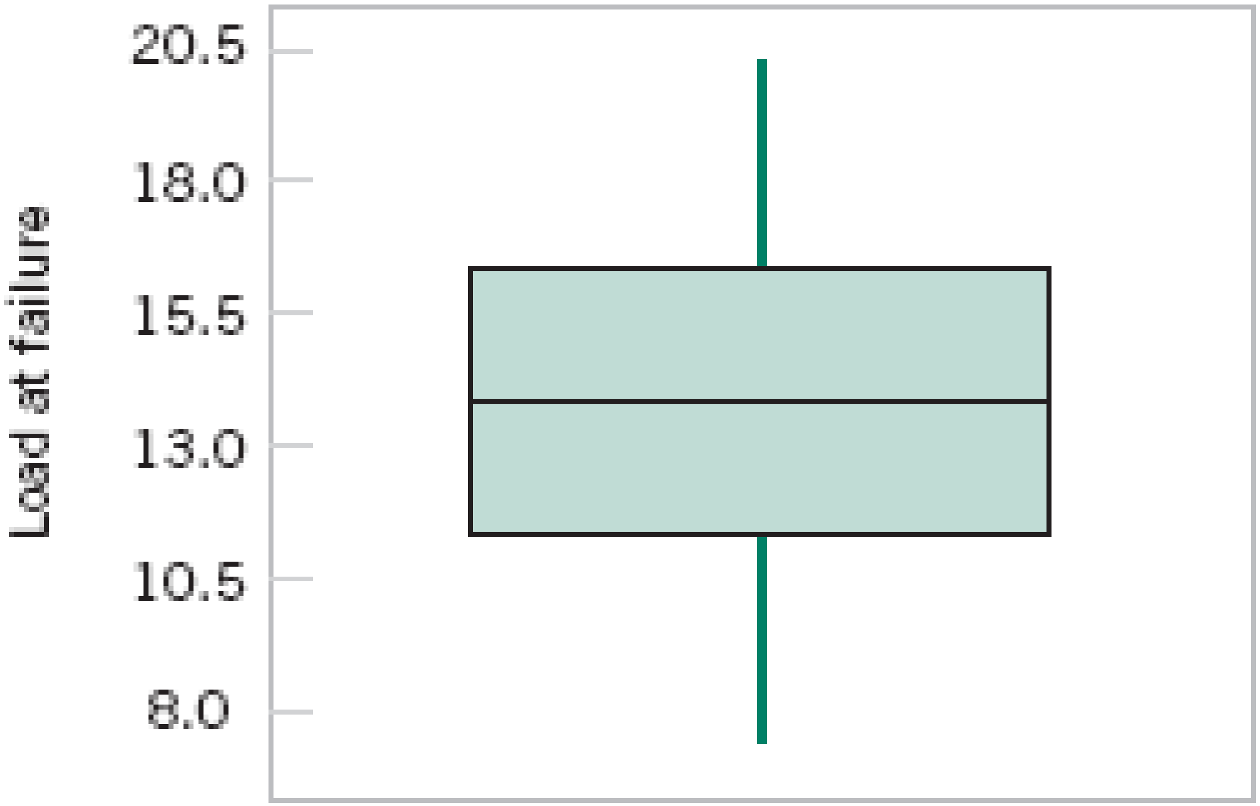
Find a 95% CI on μ.

$$\bar{x} = 13.71 \qquad s = 3.55$$

Let's check whether we can assume that the population is normally distributed by

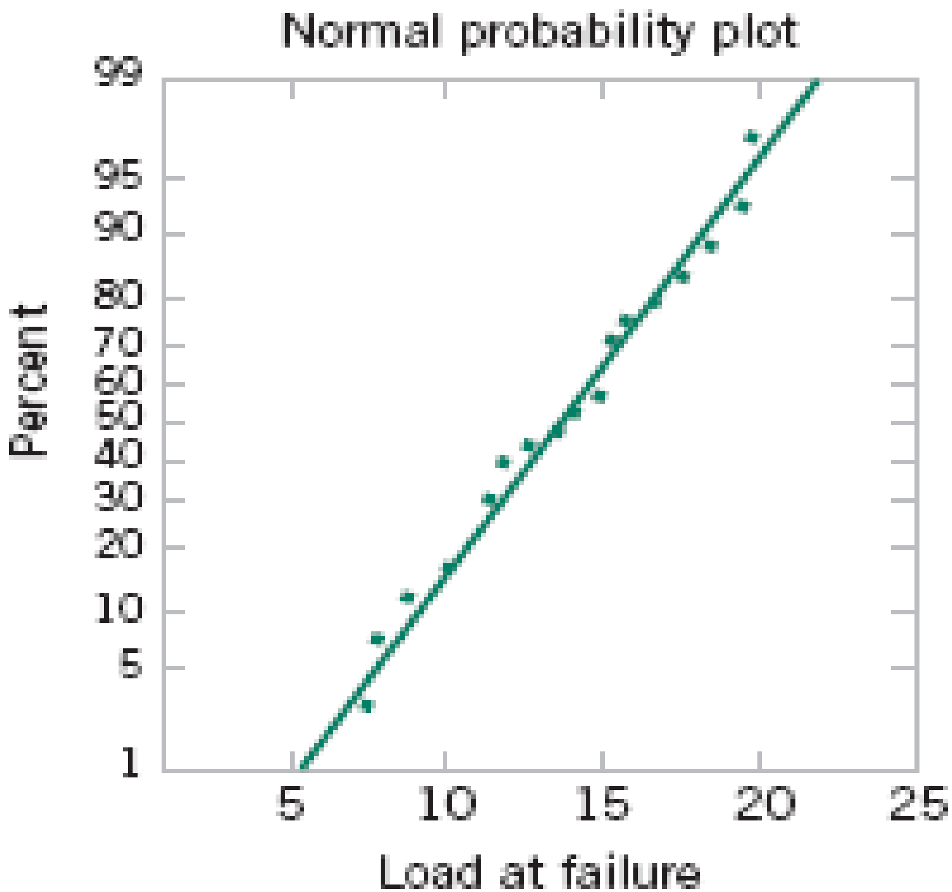
- box plot
- normal probability plot

8-3 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown



Box and Whisker plot for the load at failure data in Example 8-5.

8-3 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown



Normal probability plot of the load at failure data in Example 8-5.

8-3 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

Example 8-5 (continues)

- Box plot and normal probability plot provide good support for the assumption that the population is normally distributed.
- Since $n = 22$, we have n-1 = 21 degrees of freedom for t.
- $t_{0.025,21} = 2.080$
- The resulting 95% CI is

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

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

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

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

8-4 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

Definition

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

$$X^2 = \frac{(n-1)\,S^2}{\sigma^2} \tag{8-19}$$

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

8-4 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

Remember

The gamma distribution function is

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)} \quad for\ x > 0. \quad \lambda > 0, r > 0$$

The chi-squared distribution is a special case of the gamma distribution where **λ =1/2** and **r=1/2, 1, 3/2, 2,…**

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

where **k is the number of degrees of freedom, k=1,2,3….**

8-4 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

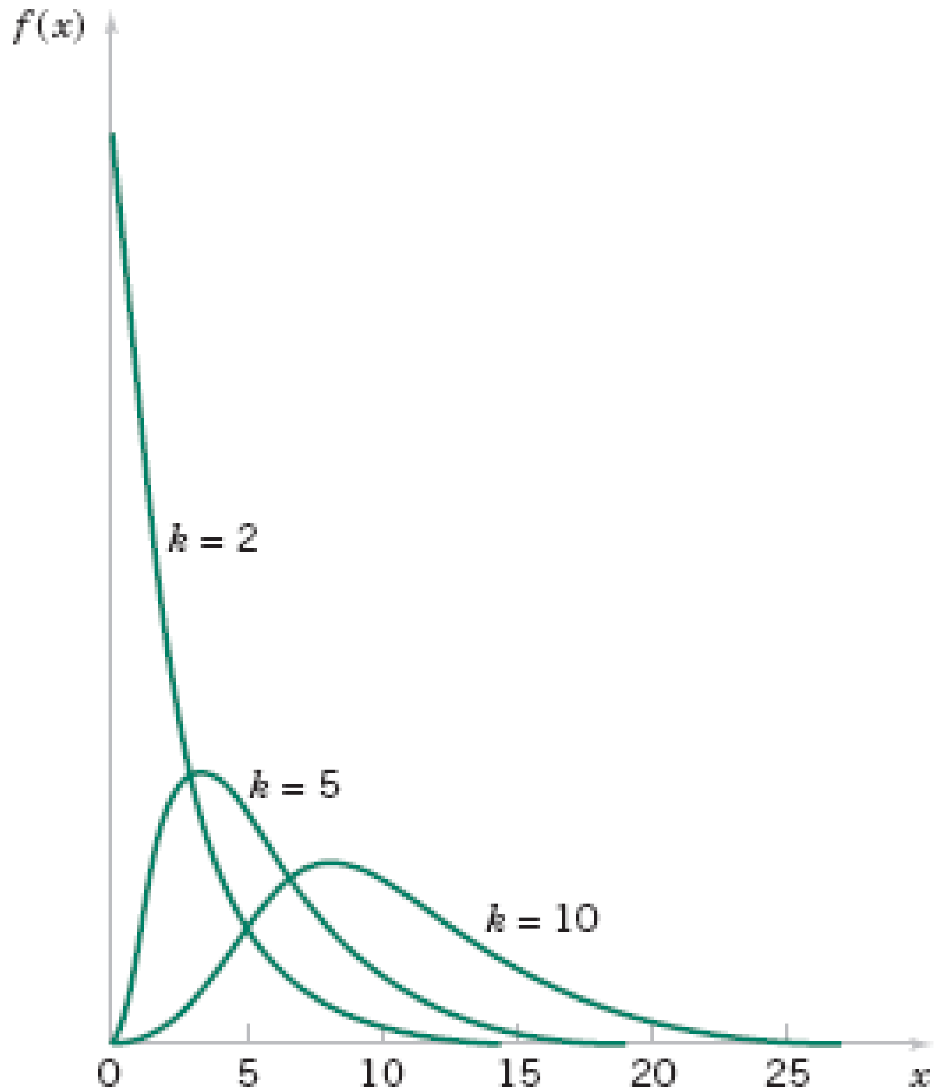
Probability density functions of several χ^2 distributions.

$$As \quad k \rightarrow \infty$$

The limiting form of χ^2 distribution is the normal distribution.

Note χ^2 distribution is not symmetric !

It is right skewed !



8-4 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

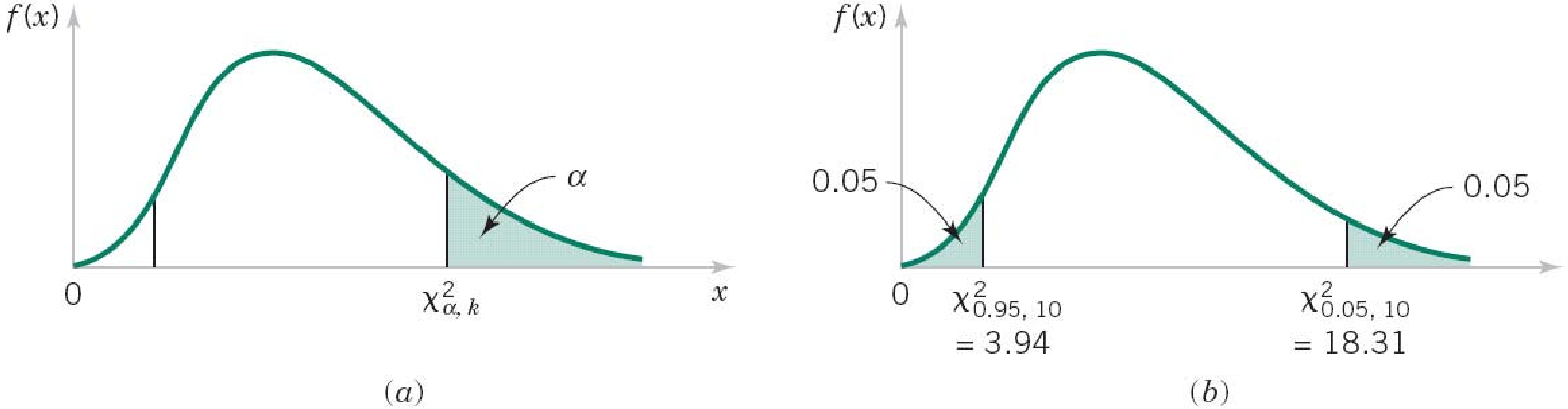


Figure 8-9 Percentage point of the χ^2 distribution. (a) The percentage point $\chi^2_{\alpha,k}$. (b) The upper percentage point $\chi^2_{0.05,10} = 18.31$ and the lower percentage point $\chi^2_{0.95,10} = 3.94$.

$$P\left(\chi^2_{1-\alpha/2,n-1} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\alpha/2,n-1}\right) = 1-\alpha$$

8-4 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

Definition

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 − α)% confidence interval on** σ^2 **is**

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} \qquad (8\text{-}21)$$

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 8-21.

8-4 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

One-Sided Confidence Bounds

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

$$\frac{(n-1)s^2}{\chi^2_{\alpha,n-1}} \le \sigma^2 \quad \text{and} \quad \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha,n-1}} \qquad (8\text{-}22)$$

respectively.

8-4 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

Example 8-6

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 (fluid ounces)^2. If the variance of fill volume is too large, an unacceptable proportion of bottles will be under- or overfilled. We will assume that the fill volume is approximately normally distributed. Compute a 95% upper-confidence interval for the variance.

$$\sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{0.95,19}}$$

$$\sigma^2 \leq \frac{(19)0.0153}{10.117} = 0.0287 \text{ (fluid ounce)}^2$$

$$\sigma \leq 0.17$$

8-5 A Large-Sample Confidence Interval For a Population Proportion

Normal Approximation for 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.

The quantity $\sqrt{p(1-p)/n}$ is called the standard error of the point estimator $\hat{P}$ which is $\sigma_{\hat{P}}$.

8-5 A Large-Sample Confidence Interval For a Population 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

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad (8\text{-}25)$$

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

NOTE:

$$\left.\begin{array}{l} np >= 5 \\ n(1-p) >= 5 \end{array}\right\} \text{ is required !}$$

8-5 A Large-Sample Confidence Interval For a Population Proportion

Example 8-7

In a random sample of 85 automobile engine crankshaft bearings, 10 have a surface finish that is rougher than the specifications allow. Compute a 95% two-sided confidence interval for the proportion of bearings in the population that exceeds the roughness specification.

$$\hat{p} = x/n = 10/85 = 0.12$$

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

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

$$0.05 \le p \le 0.19$$

8-5 A Large-Sample Confidence Interval For a Population Proportion

Choice of Sample Size

The sample size for a specified value E is given by

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p) \qquad (8\text{-}26)$$

Note that p(1-p) will be maximum at p=0.5. So **p(1-p)≤0.25**

An upper bound on n is given by

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 (0.25) \qquad (8\text{-}27)$$

8-5 A Large-Sample Confidence Interval For a Population Proportion

Example 8-8

Consider the situation in Example 8-7.
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?

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

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, sample size is found as:

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

Notice that 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.

8-5 A Large-Sample Confidence Interval For a Population Proportion

One-Sided Confidence Bounds

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

$$\hat{p} - z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \qquad \text{and} \qquad p \le \hat{p} + z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{8-28}$$

respectively.

8-7 Tolerance and Prediction Intervals

8-7.1 Prediction Interval for Future Observation

•Predicting the next future observation with a 100(1-α)% prediction interval

•A random sample of X_1, X_2, …, X_n from a normal population

•What will be X_{n+1} ?

•A point prediction of X_{n+1} is the sample mean X-bar.

•The prediction error is (X_{n+1} - X-bar). Since X_{n+1} and X-bar are independent, prediction error is normally distributed with

$$\begin{cases} E(X_{n+1} - \bar{X}) = \mu - \mu = 0 \\ V(X_{n+1} - \bar{X}) = \sigma^2 + \dfrac{\sigma^2}{n} = \sigma^2\left(1 + \dfrac{1}{n}\right) \end{cases}$$

$$Z = \frac{X_{n+1} - \bar{X}}{\sigma\sqrt{1 + \frac{1}{n}}} \qquad \textit{replace } \sigma \textit{ with } S \qquad T = \frac{X_{n+1} - \bar{X}}{S\sqrt{1 + \frac{1}{n}}}$$

8-7 Tolerance and Prediction Intervals

8-7.1 Prediction Interval for Future Observation

A 100(1 − α)% prediction interval on a single future observation from a normal distribution is given by

$$\bar{x} - t_{\alpha/2,n-1}\, s\sqrt{1+\frac{1}{n}} \le X_{n+1} \le \bar{x} + t_{\alpha/2,n-1}\, s\sqrt{1+\frac{1}{n}} \qquad (8\text{-}29)$$

The prediction interval for X_{n+1} will always be longer than the confidence interval for μ.

8-7 Tolerance and Prediction Intervals

Example 8-9

Reconsider the tensile adhesion tests on specimens of U-700 alloy described in Example 8-4. The load at failure for $n = 22$ specimens was observed, and we found that $\bar{x} = 13.71$ and $s = 3.55$. The 95% confidence interval on μ was $12.14 \le \mu \le 15.28$. We plan to test a twenty-third specimen. A 95% prediction interval on the load at failure for this specimen is

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

$$13.71 - (2.080)3.55\sqrt{1 + \frac{1}{22}} \le X_{23} \le 13.71 + (2.080)3.55\sqrt{1 + \frac{1}{22}}$$

$$6.16 \le X_{23} \le 21.26$$

Notice that the prediction interval is considerably longer than the CI.

8-7 Tolerance and Prediction Intervals

8-7.2 Tolerance Interval for a Normal Distribution

Definition

If **μ and σ are unkown**, capturing a specific percentage of values of a population will contain less than this percentage (probably) because of sampling variability in x-bar and s

(covering)

A **tolerance interval** for capturing at least γ% of the values in a normal distribution with confidence level 100(1 − α)% is

$$\bar{x} - ks, \qquad \bar{x} + ks$$

where k is a tolerance interval factor found in Appendix Table XII. Values are given for γ = 90%, 95%, and 99% and for 90%, 95%, and 99% confidence.

8-7 Tolerance and Prediction Intervals

Example 8-10

Consider the tensile adhesion tests originally described in Example 8-4. The load at failure for n =22 specimens was observed, and we found that $\overline{x}$ = 13.71 and s =3.55.Find a tolerance interval for the load at failure that includes 90% of the values in the population with 95% confidence.

$n = 22$, $\gamma = 0.90$, and 95% confidence ➡ k=2.264

$$(\overline{x} - ks, \overline{x} + ks) \Rightarrow [13.71 - (2.264)3.55\,,\,13.71 + (2.264)3.55]$$

$$(5.67\,,\,21.74)$$