

# Statistical Intervals for a Single Sample

## 8-1 INTRODUCTION

A **tolerance interval** is another important type of interval estimate. For example, the chemical product viscosity data might be assumed to be normally distributed. We might like to calculate limits that bound 95% of the viscosity values. For a normal distribution, we know that 95% of the distribution is in the interval

$$\mu - 1.96\sigma, \mu + 1.96\sigma \quad (8-1)$$

However, this is not a useful tolerance interval because the parameters  $\mu$  and  $\sigma$  are unknown. Point estimates such as  $\bar{x}$  and  $s$  can be used in Equation 8-1 for  $\mu$  and  $\sigma$ . However, we need to account for the potential error in each point estimate to form a tolerance interval for the distribution. The result is an interval of the form

$$\bar{x} - ks, \bar{x} + ks \quad (8-2)$$

where  $k$  is an appropriate constant (that is larger than 1.96 to account for the estimation error). As for a confidence interval, it is not certain that Equation 8-2 bounds 95% of the distribution, but the interval is constructed so that we have high confidence that it does.

Confidence and tolerance intervals bound unknown elements of a distribution. In this chapter you will learn to appreciate the value of these intervals. A **prediction interval** provides bounds on one (or more) future observations from the population. For example, a

Keep the purpose of the three types of interval estimates clear:

- A confidence interval bounds population or distribution parameters (such as the mean viscosity).
- A tolerance interval bounds a selected proportion of a distribution.
- A prediction interval bounds future observations from the population or 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

Suppose that  $X_1, X_2, \dots, X_n$  is a random sample from a normal distribution with unknown mean  $\mu$  and known variance  $\sigma^2$ . From the results of Chapter 5 we know that the sample mean  $\bar{X}$  is normally distributed with mean  $\mu$  and variance  $\sigma^2/n$ . We may **standardize**  $\bar{X}$  by subtracting the mean and dividing by the standard deviation, which results in the variable

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad (8-3)$$

The random variable  $Z$  has a standard normal distribution.

A **confidence interval** estimate for  $\mu$  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:

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

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  $\mu$ . Once we have selected the sample, so that  $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ , and computed  $l$  and  $u$ , the resulting **confidence interval** for  $\mu$  is

$$l \leq \mu \leq u \quad (8-5)$$

The end-points or bounds  $l$  and  $u$  are called the **lower-** and **upper-confidence limits**, respectively, and  $1 - \alpha$  is called the **confidence coefficient**.

In our problem situation, because  $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$  has a standard normal distribution, we may write

$$P\left\{-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right\} = 1 - \alpha \quad \text{© CourseSmart}$$

$$P\left\{\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha \quad (8-6)$$

**Confidence Interval on the Mean, Variance Known**

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

$$\bar{x} - z_{\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \bar{x} + z_{\alpha/2}\sigma/\sqrt{n} \tag{8-7}$$

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

**EXAMPLE 8-1  
Metallic Material Transition**

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  $60^\circ\text{C}$  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  $\mu$ , 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:

$$\begin{aligned} \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} &\leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \\ 64.46 - 1.96\frac{1}{\sqrt{10}} &\leq \mu \leq 64.46 + 1.96\frac{1}{\sqrt{10}} \\ 63.84 &\leq \mu \leq 65.08 \end{aligned}$$

That is, based on the sample data, a range of highly plausible values for mean impact energy for A238 steel at  $60^\circ\text{C}$  is  $63.84J \leq \mu \leq 65.08J$ .

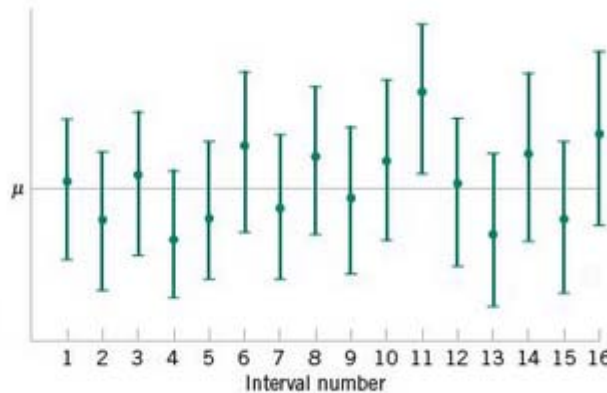


Figure 8-1 Repeated construction of a confidence interval for  $\mu$ .

**Interpreting a Confidence Interval**

How does one interpret a confidence interval? In the impact energy estimation problem in Example 8-1 the 95% CI is  $63.84 \leq \mu \leq 65.08$ , so it is tempting to conclude that  $\mu$  is within this interval with probability 0.95. However, with a little reflection, it's easy to see that this cannot be correct; the true value of  $\mu$  is unknown and the statement  $63.84 \leq \mu \leq 65.08$  is either correct (true with probability 1) or incorrect (false with probability 1). The correct interpretation lies in the realization that a CI is a *random interval* because in the probability statement defining the end-points of the interval (Equation 8-4),  $L$  and  $U$  are random variables. Consequently, the correct interpretation of a  $100(1 - \alpha)\%$  CI depends on the relative frequency view of probability. Specifically, if an infinite number of random samples are collected and a  $100(1 - \alpha)\%$  confidence interval for  $\mu$  is computed from each sample,  $100(1 - \alpha)\%$  of these intervals will contain the true value of  $\mu$ .

### Confidence Level and Precision of Estimation

Notice in Example 8-1 that our choice of the 95% level of confidence was essentially arbitrary. What would have happened if we had chosen a higher level of confidence, say, 99%? In fact, doesn't it seem reasonable that we would want the higher level of confidence? At  $\alpha = 0.01$ , we find  $z_{\alpha/2} = z_{0.01/2} = z_{0.005} = 2.58$ , while for  $\alpha = 0.05$ ,  $z_{0.025} = 1.96$ . Thus, the length of the 95% confidence interval is

$$2(1.96\sigma/\sqrt{n}) = 3.92\sigma/\sqrt{n}$$

whereas the length of the 99% CI is

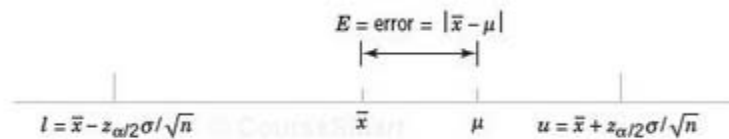
$$2(2.58\sigma/\sqrt{n}) = 5.16\sigma/\sqrt{n}$$

Thus, the 99% CI is longer than the 95% CI. This is why we have a higher level of confidence in the 99% confidence interval. Generally, for a fixed sample size  $n$  and standard deviation  $\sigma$ , the higher the confidence level, the longer the resulting CI.

Thus, the 99% CI is longer than the 95% CI. This is why we have a higher level of confidence in the 99% confidence interval. Generally, for a fixed sample size  $n$  and standard deviation  $\sigma$ , the higher the confidence level, the longer the resulting CI.

The length of a confidence interval is a measure of the **precision** of estimation. From the preceding discussion, we see that precision is inversely related to the confidence level. It is desirable to obtain a confidence interval that is short enough for decision-making purposes and that also has adequate confidence. One way to achieve this is by choosing the sample size  $n$  to be large enough to give a CI of specified length or precision with prescribed confidence.

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



## 8-2.2 Choice of Sample Size

### Sample Size for Specified Error on the Mean, Variance Known

If  $\bar{x}$  is used as an estimate of  $\mu$ , 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

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

### EXAMPLE 8-2 Metallic Material Transition

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  $\mu$  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  $z_{\alpha/2} = 1.96$ . 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  $n = 16$ .

### One-Sided Confidence Bounds on the Mean, Variance Known

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

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

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

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

### EXAMPLE 8-3 One-Sided Confidence Bound

The same data for impact testing from Example 8-1 is used to construct a lower, one-sided 95% confidence interval for the mean impact energy. Recall that  $\bar{x} = 64.46$ ,  $\sigma = 1J$ , and  $n = 10$ . The interval is

$$\begin{aligned} \bar{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}} &\leq \mu \\ 64.46 - 1.64 \frac{1}{\sqrt{10}} &\leq \mu \\ 63.94 &\leq \mu \end{aligned}$$

The lower limit for the two-sided interval in Example 8-1 was 63.84. Because  $z_{\alpha} < z_{\alpha/2}$ , the lower limit of a one-sided interval is always greater than the lower limit of a two-sided interval of equal confidence. The one-sided interval does not bound  $\mu$  from above so that it still achieves 95% confidence with a slightly greater lower limit. If our interest is only in the lower limit for  $\mu$ , then the one-sided interval is preferred because it provides equal confidence with a greater lower limit. Similarly, a one-sided upper limit is always less than a two-sided upper limit of equal confidence.

### 8-3 CONFIDENCE INTERVAL ON THE MEAN OF A NORMAL DISTRIBUTION, VARIANCE UNKNOWN

#### 8-3.1 *t* Distribution

*t* Distribution

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

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

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

The *t* probability density function is

$$f(x) = \frac{\Gamma[(k + 1)/2]}{\sqrt{\pi k} \Gamma(k/2)} \cdot \frac{1}{[(x^2/k) + 1]^{(k+1)/2}} \quad -\infty < x < \infty \tag{8-16}$$

where  $k$  is the number of degrees of freedom. The mean and variance of the *t* distribution are zero and  $k/(k - 2)$  (for  $k > 2$ ), respectively.

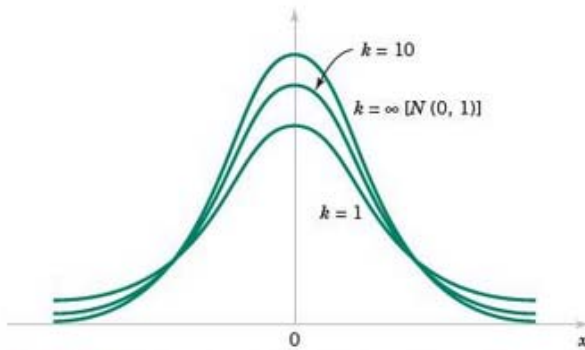


Figure 8-4 Probability density functions of several *t* distributions.

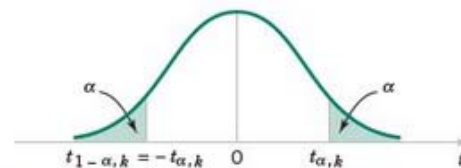


Figure 8-5 Percentage points of the *t* distribution.

$$P(T_{10} > t_{0.05,10}) = P(T_{10} > 1.812) = 0.05$$

### 8-3.2 $t$ Confidence Interval on $\mu$

It is easy to find a  $100(1 - \alpha)$  percent confidence interval on the mean of a normal distribution with unknown variance by proceeding essentially as we did in Section 8-2.1. We know that the distribution of  $T = (\bar{X} - \mu)/(S/\sqrt{n})$  is  $t$  with  $n - 1$  degrees of freedom. Letting  $t_{\alpha/2, n-1}$  be the upper  $100\alpha/2$  percentage point of the  $t$  distribution with  $n - 1$  degrees of freedom, we may write:

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

or

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

© CourseSmart

Rearranging this last equation yields

$$P(\bar{X} - t_{\alpha/2, n-1}S/\sqrt{n} \leq \mu \leq \bar{X} + t_{\alpha/2, n-1}S/\sqrt{n}) = 1 - \alpha \quad (8-17)$$

This leads to the following definition of the  $100(1 - \alpha)$  percent two-sided confidence interval on  $\mu$ .

**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  $\sigma^2$ , a **100(1 -  $\alpha$ ) percent confidence interval on  $\mu$**  is given by

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

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 of a normal distribution are also of interest and are easy to find. Simply use only the appropriate lower or upper confidence limit from Equation 8-18 and replace  $t_{\alpha/2, n-1}$  by  $t_{\alpha, n-1}$ .

**EXAMPLE 8-5**  
**Alloy Adhesion**

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	

The sample mean is  $\bar{x} = 13.71$ , and the sample standard deviation is  $s = 3.55$ . Figures 8-6 and 8-7 show a box plot and a normal probability plot of the tensile adhesion test data, respectively. These displays provide good support for the assumption that the population is normally distributed. We want to find a 95% CI on  $\mu$ . Since  $n = 22$ , we have  $n - 1 = 21$  degrees of freedom for  $t$ , so  $t_{0.025, 21} = 2.080$ . The resulting CI is

$$\begin{aligned} \bar{x} - t_{\alpha/2, n-1}s/\sqrt{n} &\leq \mu \leq \bar{x} + t_{\alpha/2, n-1}s/\sqrt{n} \\ 13.71 - 2.080(3.55)/\sqrt{22} &\leq \mu \leq 13.71 + 2.080(3.55)/\sqrt{22} \\ 13.71 - 1.57 &\leq \mu \leq 13.71 + 1.57 \\ 12.14 &\leq \mu \leq 15.28 \end{aligned}$$

The CI is fairly wide because there is a lot of variability in the tensile adhesion test measurements.



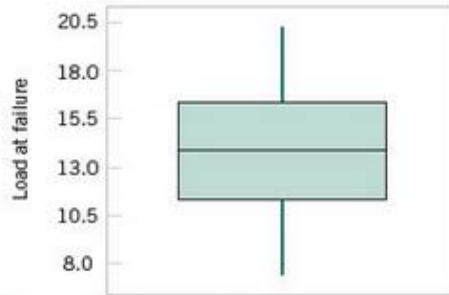


Figure 8-6 Box and whisker plot for the load at failure data in Example 8-5.

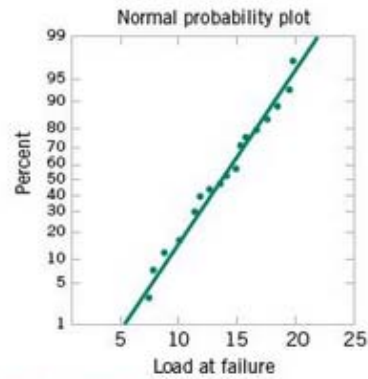


Figure 8-7 Normal probability plot of the load at failure data from Example 8-5.

#### 8-4 CONFIDENCE INTERVAL ON THE VARIANCE AND STANDARD DEVIATION OF A NORMAL DISTRIBUTION

Sometimes confidence intervals on the population variance or standard deviation are needed. When the population is modeled by a normal distribution, the tests and intervals described in this section are applicable. The following result provides the basis of constructing these confidence intervals.

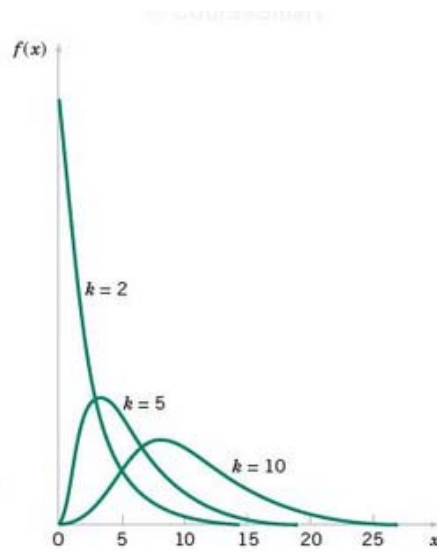
##### $\chi^2$ Distribution

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

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

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

Figure 8-8 Probability density functions of several  $\chi^2$  distributions.



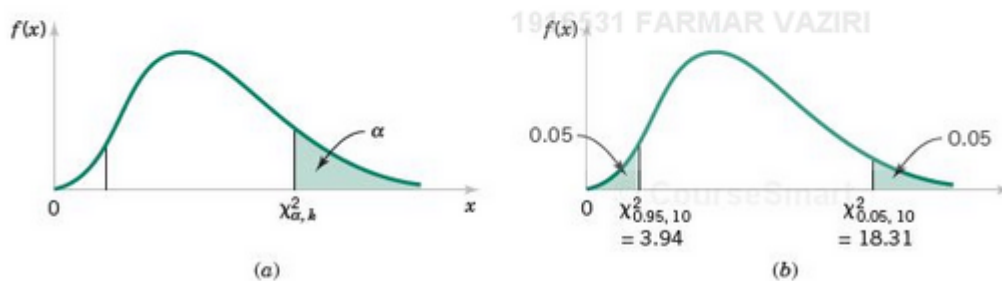
The probability density function of a  $\chi^2$  random variable is

$$f(x) = \frac{1}{2^{k/2} \Gamma(k/2)} x^{(k/2)-1} e^{-x/2} \quad x > 0 \quad (8-20)$$

The **percentage points** of the  $\chi^2$  distribution are given in Table IV of the Appendix. Define  $\chi_{\alpha,k}^2$  as the percentage point or value of the chi-square random variable with  $k$  degrees of freedom such that the probability that  $X^2$  exceeds this value is  $\alpha$ . That is,

$$P(X^2 > \chi_{\alpha,k}^2) = \int_{\chi_{\alpha,k}^2}^{\infty} f(u) du = \alpha$$

This probability is shown as the shaded area in Fig. 8-9(a). To illustrate the use of Table IV, note that the areas  $\alpha$  are the column headings and the degrees of freedom  $k$  are given in the left column. Therefore, the value with 10 degrees of freedom having an area (probability) of 0.05



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

to the right is  $\chi_{0.05,10}^2 = 18.31$ . This value is often called an **upper 5%** point of chi-square with 10 degrees of freedom. We may write this as a probability statement as follows:

$$P(X^2 > \chi_{0.05,10}^2) = P(X^2 > 18.31) = 0.05$$

Conversely, a **lower 5%** point of chi-square with 10 degrees of freedom would be  $\chi_{0.95,10}^2 = 3.94$  (from Appendix A). Both of these percentage points are shown in Figure 8-9(b).

The construction of the  $100(1 - \alpha)\%$  CI for  $\sigma^2$  is straightforward. Because

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

is chi-square with  $n - 1$  degrees of freedom, we may write

$$P(\chi_{1-\alpha/2,n-1}^2 \leq X^2 \leq \chi_{\alpha/2,n-1}^2) = 1 - \alpha$$

so that

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

This last equation can be rearranged as

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

This leads to the following definition of the confidence interval for  $\sigma^2$ .

**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  $\sigma^2$ , then a  $100(1 - \alpha)\%$  confidence interval on  $\sigma^2$  is

$$\frac{(n-1)s^2}{\chi_{\alpha/2, n-1}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{1-\alpha/2, n-1}^2} \quad (8-21)$$

where  $\chi_{\alpha/2, n-1}^2$  and  $\chi_{1-\alpha/2, n-1}^2$  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  $\sigma$  has lower and upper limits that are the square roots of the corresponding limits in Equation 8-21.

It is also possible to find a  $100(1 - \alpha)\%$  lower confidence bound or upper confidence bound on  $\sigma^2$ .

**One-Sided Confidence Bounds on the Variance**

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

$$\frac{(n-1)s^2}{\chi_{\alpha, n-1}^2} \leq \sigma^2 \quad \text{and} \quad \sigma^2 \leq \frac{(n-1)s^2}{\chi_{1-\alpha, n-1}^2} \quad (8-22)$$

respectively.

**EXAMPLE 8-6 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  $s^2 = 0.0153$  (fluid ounces)<sup>2</sup>. 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. A 95% upper-confidence interval is found from Equation 8-22 as follows:

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

or

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

This last expression may be converted into a confidence interval on the standard deviation  $\sigma$  by taking the square root of both sides, resulting in

$$\sigma \leq 0.17$$

Therefore, at the 95% level of confidence, the data indicate that the process standard deviation could be as large as 0.17 fluid ounce.

8-7 TOLERANCE AND PREDICTION INTERVALS

© CourseSmart

8-7.1 Prediction Interval for a Future Observation

In some problem situations, we may be interested in predicting a future observation of a variable. This is a different problem than estimating the mean of that variable, so a confidence interval is not appropriate. In this section we show how to obtain a  $100(1 - \alpha)\%$  **prediction interval** on a future value of a normal random variable.

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	$\mu$	Standard deviation $\sigma$ known	8-2	9-2	
Mean of arbitrary distribution with large sample size	$\mu$	Sample size large enough that central limit theorem applies and $\sigma$ is essentially known	8-2.5	9-2.5	Large sample size is often taken to be $n \geq 40$
Mean of normal distribution	$\mu$	Standard deviation $\sigma$ unknown and estimated	8-3	9-3	
Variance (or standard deviation) of normal distribution	$\sigma^2$	Mean $\mu$ unknown and estimated	8-4	9-4	
Population Proportion	$p$	None	8-5	9-5	

Prediction Interval

A  $100(1 - \alpha)\%$  prediction interval on a single future observation from a normal distribution is given by

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

EXAMPLE 8-9 Alloy Adhesion

Reconsider the tensile adhesion tests on specimens of U-700 alloy described in Example 8-5. 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  $\mu$  was  $12.14 \leq \mu \leq 15.28$ . We plan to test a twenty-third specimen.

A 95% prediction interval on the load at failure for this specimen is

$$\begin{aligned} \bar{x} - t_{\alpha/2, n-1} s \sqrt{1 + \frac{1}{n}} &\leq X_{n+1} \leq \bar{x} + t_{\alpha/2, n-1} s \sqrt{1 + \frac{1}{n}} \\ 13.71 - (2.080)3.55 \sqrt{1 + \frac{1}{22}} &\leq X_{23} \leq 13.71 + (2.080)3.55 \sqrt{1 + \frac{1}{22}} \\ 6.16 &\leq X_{23} \leq 21.26 \end{aligned}$$

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

### 8-7.2 Tolerance Interval for a Normal Distribution

Consider a population of semiconductor processors. Suppose that the speed of these processors has a normal distribution with mean  $\mu = 600$  megahertz and standard deviation  $\sigma = 30$  megahertz. Then the interval from  $600 - 1.96(30) = 541.2$  to  $600 + 1.96(30) = 658.8$  megahertz captures the speed of 95% of the processors in this population because the interval from  $-1.96$  to  $1.96$  captures 95% of the area under the standard normal curve. The interval from  $\mu - z_{\alpha/2}\sigma$  to  $\mu + z_{\alpha/2}\sigma$  is called a **tolerance interval**.

If  $\mu$  and  $\sigma$  are unknown, we can use the data from a random sample of size  $n$  to compute  $\bar{x}$  and  $s$ , and then form the interval  $(\bar{x} - 1.96s, \bar{x} + 1.96s)$ . However, because of sampling variability in  $\bar{x}$  and  $s$ , it is likely that this interval will contain less than 95% of the values in the population. The solution to this problem is to replace 1.96 by some value that will make the proportion of the distribution contained in the interval 95% with some level of confidence. Fortunately, it is easy to do this.

#### Tolerance Interval

A tolerance interval for capturing at least  $\gamma\%$  of the values in a normal distribution with confidence level  $100(1 - \alpha)\%$  is

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

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

#### EXAMPLE 8-10 Alloy Adhesion

Let's reconsider the tensile adhesion tests originally described in Example 8-5. The load at failure for  $n = 22$  specimens was observed, and we found that  $\bar{x} = 13.71$  and  $s = 3.55$ . We want to find a tolerance interval for the load at failure that includes 90% of the values in the population with 95% confidence. From Appendix Table XII the tolerance factor  $k$  for  $n = 22$ ,  $\gamma = 0.90$ , and 95% confidence is  $k = 2.264$ . The desired tolerance interval is

$$(\bar{x} - ks, \bar{x} + ks) \quad \text{or} \quad [13.71 - (2.264)3.55, 13.71 + (2.264)3.55]$$

which reduces to  $(5.67, 21.74)$ . We can be 95% confident that at least 90% of the values of load at failure for this particular alloy lie between 5.67 and 21.74 megapascals.