Lecture-8

Hypothesis testing

Learning Objectives

After careful study of this lecture, you should be able to do the following:

1. Structure engineering decision-making problems as hypothesis tests
2. Test hypotheses on the mean of a normal distribution using either a Z-test or a t-test procedure
3. Test hypotheses on the variance or standard deviation of a normal distribution
4. Test hypotheses on a population proportion
5. Use the p-value approach for making decisions in hypothesis tests
6. Use the chi-square goodness-of-fit test to check distributional assumptions
7. Use contingency table tests

STATISTICAL HYPOTHESES:

An **interval** estimate for a **population parameter** is called a **confidence interval**. However, many problems in engineering require that we decide which of two competing **claims** or **statements** about **some parameter** is true. The statements are called **hypotheses**, and the decision-making procedure is called **hypothesis testing**. This is one of the most useful aspects of statistical inference, because many types of decision making problems, tests, or experiments in the engineering world can be formulated as hypothesis-testing problems. Furthermore, as we will see, a very close connection exists between **hypothesis testing and confidence intervals**.

Statistical hypothesis testing and confidence interval estimation of parameters are the fundamental methods used at the data analysis stage of a **comparative experiment** in which the engineer is interested, for example, in comparing the mean of a population to a specified value. These simple comparative experiments are frequently encountered in practice and provide a good foundation for the more complex experimental design problems. In this lecture, we discuss comparative experiments involving a **single population**, and our focus is on testing hypotheses concerning the parameters of the population.

We now give a formal definition of a statistical hypothesis.

A statistical hypothesis is a statement about the parameters of one or more populations.

Because we use probability distributions to represent **populations**, a statistical hypothesis may also be thought of as a statement about the probability distribution of a random variable. The hypothesis will usually involve one or more parameters of this distribution.

For example, consider the air crew escape system. Suppose that we are interested in the **burning rate of the solid propellant**. Burning rate is a random variable that can be described by a probability distribution. Suppose that our interest focuses on the **mean burning rate** (a parameter of this distribution). Specifically, we are interested in determining whether or not the mean burning rate is **50 centimeters per second**. We may express this formally as

$H_0: \mu = 50$: centimeters per second; $H_1: \mu \neq 50$: centimeter per seconds (8.1)

The statement H_0: μ = 50 centimeters per second in equation (8.1) is called the **null hypothesis**. This is a claim that is initially assumed to be true. The statement H_1: μ ≠ 50: centimeters per second is called the **alternative hypothesis** and it is a statement that **contradicts the null hypothesis**. Because the alternative hypothesis specifies values of μ that could be either greater or less than 50 centimeters per second, it is called a **two-sided alternative hypothesis**. In some situations, we may wish to formulate a one-sided alternative hypothesis, as in

H_0: μ = 50 centimeters per second	H_0: μ = 50: centimeters per second	
or	or	(8.2)
H_1: μ< 50 centimeters per second	H_1: μ > 50: centimeters per second	

We will always state the **null hypothesis as an equality claim**. However when the alternative hypothesis is stated with the < sign, the implicit claim in the null hypothesis can be taken as ≥ and when the alternative hypothesis is stated with the > sign, the implicit claim in the null hypothesis can be taken as ≤.

It is important to remember that hypotheses are always statements about the population or distribution under study, not statements about the sample. The value of the population parameter specified in the null hypothesis (50 centimeters per second in the preceding example) is usually determined in one of three ways.

First, it may result from past experience or knowledge of the process or even from previous tests or experiments. The objective of hypothesis testing, then, is usually to determine whether the parameter value has changed.

Second, this value may be determined from some theory or model regarding the process under study. Here the objective of hypothesis testing is to verify the theory or model.

A third situation arises when the value of the population parameter results from external considerations, such as design or engineering specifications, or from contractual obligations. In this situation, the usual objective of hypothesis testing is conformance testing.

A procedure leading to a decision about the null hypothesis is called a **test of a hypothesis**. Hypothesis-testing procedures rely on using the information in a random sample from the population of interest. If this information is consistent with the null hypothesis, we will not reject it; however, if this information is inconsistent with the null hypothesis, we will conclude that the null hypothesis is false and reject it in favor of the alternative. We emphasize that the truth or falsity of a particular hypothesis can never be known with certainty unless we can examine the entire population. This is usually impossible in most practical situations. Therefore, a hypothesis-testing procedure should be developed with the probability of reaching a wrong conclusion in mind. Testing the hypothesis involves taking a random sample, computing a test statistic from the sample data, and then using the test statistic to make a decision about the null hypothesis.

This decision procedure can lead to either of two wrong conclusions:

Type I Error: Rejecting the null hypothesis H_0 when it is true is defined as a type I error.

Type II Error: Failing to reject the null hypothesis when it is false is defined as a type II error.

P-Values in Hypothesis Tests

The p-value is the smallest level of significance that would lead to rejection of the null hypothesis H_0 with the given data.

It is customary to consider the test statistic (and the data) significant when the null hypothesis H_0 is rejected; therefore, we may think of the p-value as the smallest level α at which the data are significant.

General procedure for hypothesis tests:

This lecture develops hypothesis-testing procedures for many practical problems. Use of the following sequence of steps in applying hypothesis-testing methodology is recommended.

1. **Parameter of interest**: From the problem context, identify the parameter of interest.
2. **Null hypothesis, H_0**: State the null hypothesis, H_0
3. **Alternative hypothesis, H_1:** Specify an appropriate alternative hypothesis, H_1
4. **Test statistic:** Determine an appropriate test statistic.
5. **Reject H_0 if:** State the rejection criteria for the null hypothesis.
6. **Computations:** Compute any necessary sample quantities, substitute these into the equation for the test statistic, and compute that value.
7. **Draw conclusions:** Decide whether or not H_0 should be rejected and report that in the problem context.

Tests on the Mean of a Normal Distribution, Variance Known:

In this section, we consider hypothesis testing about the mean μ of a single normal population where the variance of the population σ^2 is known. We will assume that a random sample $X_1, X_2, \ldots, X_n$ has been taken from the population. Based on our previous discussion, the sample mean $\bar{X}$ is an unbiased point estimator of μ with variance σ^2/n.

Summary of Tests on the Mean, Variance Known

Testing Hypotheses on the Mean, Variance Known (Z-Tests)

Null hypothesis: $H_0: \mu = \mu_0$

Test statistic: $Z_0 = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

Alternative Hypotheses	P-Value	Rejection Criterion for Fixed-Level Tests
$H_1: \mu \neq \mu_0$	Probability above $\lvert z_0 \rvert$ and probability below $-\lvert z_0 \rvert$, $P = 2\left[1 - \Phi(\lvert z_0 \rvert)\right]$	$z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$
$H_1: \mu > \mu_0$	Probability above z_0, $P = 1 - \Phi(z_0)$	$z_0 > z_\alpha$
$H_1: \mu < \mu_0$	Probability below z_0, $P = \Phi(z_0)$	$z_0 < -z_\alpha$

The P-values and critical regions for these situations are shown in Figs. 8.1 and 8.2.

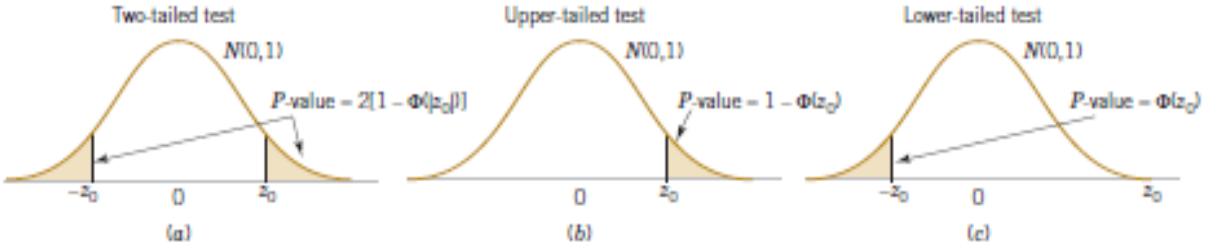


Figure 8.1: The p-value for a z-test. (a) The two-sided alternative H_1: $\mu \neq \mu_0$. (b) The one-sided alternative H_1: $\mu > \mu_0$. (c) The one-sided alternative H_1: $\mu < \mu_0$.

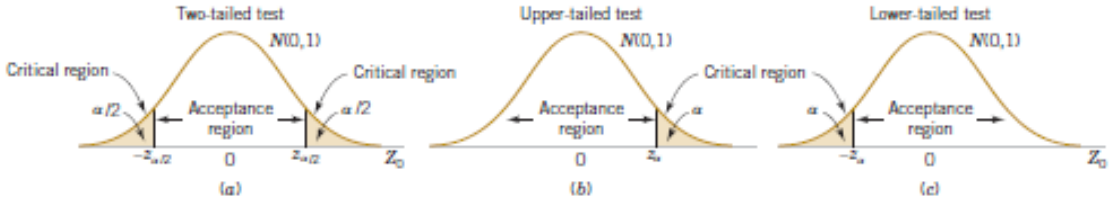


Figure 8.2: The distribution of Z when H_1: $\mu = \mu_0$ is true with critical region for (a) The two-sided alternative H_1: $\mu \neq \mu_0$ (b) The one-sided alternative H_1: $\mu > \mu_0$. (c) The one-sided alternative H_1: $\mu < \mu_0$.

Example8.1: Air crew escape systems are powered by a solid propellant. The burning rate of this propellant is an important product characteristic. Specifications require that the mean burning rate must be **50** centimeters per second. We know that the standard deviation of burning rate is **σ = 2** centimeters per second. The experimenter decides to specify a type I error probability or significance level of **α = 0.05** and selects a random sample of **n = 25** and obtains a sample average burning rate **of $\bar{x}$ = 51.3** centimeters per second. What conclusions should be drawn?

Answer:

1. **Parameter of interest:** The parameter of interest is μ, the mean burning rate.
2. **Null hypothesis:** H_0: μ = 50 centimeters per second
3. **Alternative hypothesis:** H_1: μ ≠ 50 centimeters per second
4. **Test statistic:** The test statistic is $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$
5. **Reject H_0:** If the p-value is less than 0.05 or **the value of test statistic is in the critical region**.
6. **Computations:** Because $\bar{x}$ = 51.3 and σ = 2, $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{51.3-50}{2/\sqrt{25}} = \mathbf{3.25}$
7. **Conclusion:** (i) Here the p-value $= 2[1 - \Phi(3.25)] = 0.0012$, which is less than 0.05. So we reject H_0: μ = 50 at the 0.05 level of significance.

(ii) The tabular value of the critical region: $z_{0.025} > 1.96$ and $-z_{0.025} < -1.96$. Here the value of test statistic is $z_0 = 3.25 > 1.96$. We reject H_0 at the 0.05 level of significance.

Practical Interpretation: We conclude that the mean burning rate differs from 50 centimeters per second, based on a sample of 25 measurements. In fact, there is strong evidence that the mean burning rate exceeds 50 centimeters per second.

Example 8.2:

A random sample of **100** recorded deaths in the United States during the past year showed an average life time of 71.8 years. Assuming a population standard deviation of 8.9 years, does this seem to indicate that the mean life time today is greater than 70 years? Use a 0.05 level of significance.

Answer:

1. **Parameter of interest:** The parameter of interest is μ, the average life time.
2. **Null hypothesis:** H_0: μ = 70 years
3. **Alternative hypothesis:** H_1: μ > 70 years
4. **Test statistic:** The test statistic is $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$
5. **Reject H_0:** If the p-value is less than $\alpha = 0.05$ or **the value of test statistic is in the critical region**.
6. **Computations:** Here $\bar{x}$ = 71.8 and σ = 8.9, $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{71.8-70}{8.9/\sqrt{100}} = 2.02$
7. **Conclusion:** (i) Here the p-value = $1 - \Phi(2.02) = 0.0217$. We reject H_0 at the 0.05 level of significance and conclude that the mean life time today is greater than 70 years. Or

(ii) The tabular value of the critical region: $z_{table} > 1.645$. Here the value of test statistic is $z_0 = 2.02 > 1.96$. We reject H_0 at the 0.05 level of significance and conclude that the mean life time today is greater than 70 years.

Example 8.3:

A manufacturer of sports equipment has developed a new synthetic fishing line that the company claims has a mean breaking strength of 8 kilograms with a standard deviation of 0.5 kilogram. Test the hypothesis that μ = 8 kilograms against the alternative that μ ≠ 8 kilograms if a random sample of 50 lines is tested and found to have a mean breaking strength of 7.8 kilograms. Use a 0.01 level of significance.

Answer:

1. **Parameter of interest:** The parameter of interest is μ, the mean breaking strength.
2. **Null hypothesis:** H_0: μ = 8 kilograms

3. **Alternative hypothesis:** H_1: μ ≠ 8 kilograms

4. **Test statistic:** The test statistic is $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

5. **Reject H₀:** If the p-value is less than $\alpha = 0.01$ or **the value of test statistic is in the critical region**.

6. **Computations:** Here $\bar{x}$ = 7.8 and σ = 0.5, n = 50, then $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{7.8-8}{0.5/\sqrt{50}} = -2.83$

7. **Conclusion:** (i) Here the p-value = $2[1 - \Phi(-2.83)] = 0.0046$. We reject H_0 at the 0.01 level of significance and conclude that the average breaking strength is not equal to 8 but is, in fact, less than 8 kilograms. Or

(ii) The tabular value of the critical region: $z_{0.005} > 2.575$ and $z_{0.005} < -2.575$. Here the value of test statistic is $z_0 = -2.83 < -2.575$. Hence we reject H_0 at the 0.01 level of significance and conclude that the average breaking strength is not equal to 8 but is, in fact, less than 8 kilograms.

Tests on the Mean of a Normal Distribution, Variance Unknown:

We now consider the case of hypothesis testing on the mean of a population with unknown variance σ^2. If $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with mean μ and variance σ^2, the random variable $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ has a t distribution with n – 1 degrees of freedom.

Summary for the One Sample t-test:

Testing Hypotheses on the Mean of a Normal Distribution, Variance Unknown

Null hypothesis: $H_0: \mu = \mu_0$

Test statistic: $T_0 = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$

Alternative Hypotheses	*P*-Value	Rejection Criterion for Fixed-Level Tests
$H_1: \mu \neq \mu_0$	Probability above $\lvert t_0 \rvert$ and probability below $-\lvert t_0 \rvert$	$t_0 > t_{\alpha/2,n-1}$ or $t_0 < -t_{\alpha/2,n-1}$
$H_1: \mu > \mu_0$	Probability above t_0	$t_0 > t_{\alpha,n-1}$
$H_1: \mu < \mu_0$	Probability below t_0	$t_0 < -t_{\alpha,n-1}$

The calculations of the P-values and the locations of the critical regions for these situations are shown in Figs. 8.3 and 8.4, respectively.

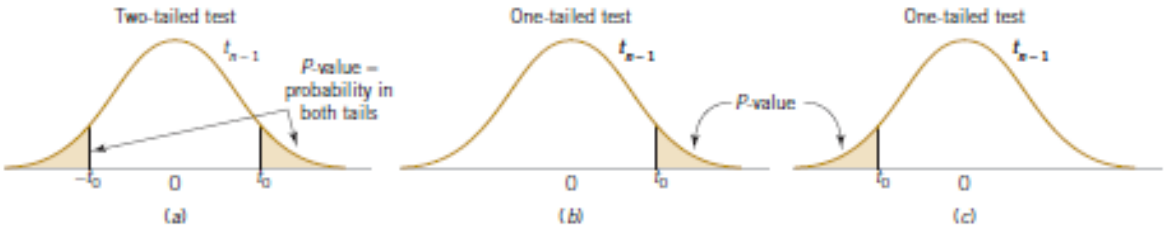


Figure 8.3: Calculating the P-value for a t-test: (a) H_1: $\mu \neq \mu_0$, (b) H_1: $\mu > \mu_0$, (c) H_1: $\mu < \mu_0$.

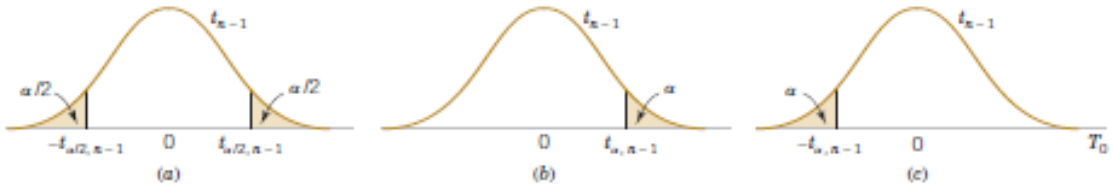


Figure 8.4: The distribution of T_0 when H_0: $\mu = \mu_0$ is true with critical region for: (a) H_1: $\mu \neq \mu_0$, (b) H_1: $\mu > \mu_0$, (c) H_1: $\mu < \mu_0$.

Example 8.4:

The Edison Electric Institute has published figures on the number of kilowatt hours used annually by various home appliances. It is claimed that a vacuum cleaner uses an average of 46 kilowatt hours per year. If a random sample of 12 homes included in a planned study indicates that vacuum cleaners use an average of 42 kilowatt hours per year with a standard deviation of 11.9 kilowatt hours, does this suggest at the 0.05 level of significance that vacuum cleaners use, on average, less than 46 kilowatt hours annually? Assume the population of kilowatt hours to be normal.

Answer: The solution using the seven-step procedure for hypothesis testing is as follows:

1. **Parameter of interest:** The parameter of interest is the average of kilowatt hours used per year, μ.
2. **Null hypothesis:** H_0: μ = 46 kilowatt hours
3. **Alternative hypothesis:** H_1: μ < 46 kilowatt hours
4. **Test statistic:** The test statistic is $t_0 = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$
5. **Reject H_0:** If the value of test statistic is in the critical region.
6. **Computations:** Here $\bar{x}$ = 42 kilowatt hours, s = 11.9 kilowatt hours and n = 12, we have $t_0 = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{42-46}{11.9/\sqrt{12}} = -1.16$, with 11 degrees of freedom.
7. **Conclusions:** The critical region at $\alpha = 0.05$ with υ = 11 degrees of freedom is t < -1.796. Here the value of test statistic is t = -1.16 > -1.796. Hence do not reject H_0 and

conclude that the average number of kilowatt hours used annually by home vacuum cleaners is not significantly less than 46.

Example 8.5

The increased availability of light materials with high strength has reformed the design and manufacture of golf clubs, particularly drivers. Clubs with hollow heads and very thin faces can result in much longer tee shots, especially for players of modest skills. This is due partly to the "spring-like effect" that the thin face imparts to the ball. Firing a golf ball at the head of the club and measuring the ratio of the ball's outgoing velocity to the incoming velocity can quantify this spring-like effect. The ratio of velocities is called the coefficient of return of the club. An experiment was performed in which 15 drivers produced by a particular club maker were selected at random and their coefficients of refund measured. In the experiment, the golf balls were fired from an air cannon so that the incoming velocity and spin rate of the ball could be precisely controlled. It is of interest to determine whether there is evidence (with α = 0.005) to support a claim that the mean coefficient of refund exceeds 0.82. The observations follow:

0.8411	0.8191	0.8182	0.8125	0.8750
0.8580	0.8532	0.8483	0.8276	0.7983
0.8042	0.8730	0.8282	0.8359	0.8660

Answer: The solution using the seven-step procedure for hypothesis testing is as follows:

1. **Parameter of interest:** The parameter of interest is the **mean coefficient of refund**, μ.
2. **Null hypothesis:** H_0: μ = 0.82
3. **Alternative hypothesis:** H_1: μ > 0.82
4. **Test statistic:** The test statistic is $t_0 = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$
5. **Reject H_0:** If the value of test statistic is in the critical region.
6. **Computations:** Because $\bar{x}$ = 0.83725, s = 0.02456, μ_0 = 0.82, and n = 15, we have

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{0.83725 - 0.82}{0.02456/\sqrt{15}} = 2.72$$

7. **Conclusions:** From Table we find for a t distribution with $v = n - 1$ =14 degrees of freedom that t_0 = **2.72** falls between two values: 2.624, for which α = 0.01, and **2.977**, for which **α = 0.005**. Therefore, we reject H_0 and conclude that the mean coefficient of

refund exceeds 0.82 at α = 0.01. On the other hand, we accept H_0 and conclude that the mean coefficient of refund is 0.82 at **α = 0.005.**

Practical Interpretation: There is strong evidence to conclude that the mean coefficient of refund exceeds 0.82 at 1% level of significance. On the other hand, the mean coefficient of refund accept 0.82 at 0.5% level of significance.

Tests on the Variance and Standard Deviation of a Normal Distribution:

Sometimes hypothesis tests 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.

Tests on the Variance of a Normal Distribution:

Null hypothesis: $H_0: \sigma^2 = \sigma_0^2$

Test statistic: $\chi_0^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$

Alternative Hypothesis	Rejection Criteria
$H_1: \sigma^2 \neq \sigma_0^2$	$\chi_0^2 > \chi^2_{\alpha/2,n-1}$ or $\chi_0^2 < \chi^2_{1-\alpha/2,n-1}$
$H_1: \sigma^2 > \sigma_0^2$	$\chi_0^2 > \chi^2_{\alpha,n-1}$
$H_1: \sigma^2 < \sigma_0^2$	$\chi_0^2 < \chi^2_{1-\alpha,n-1}$

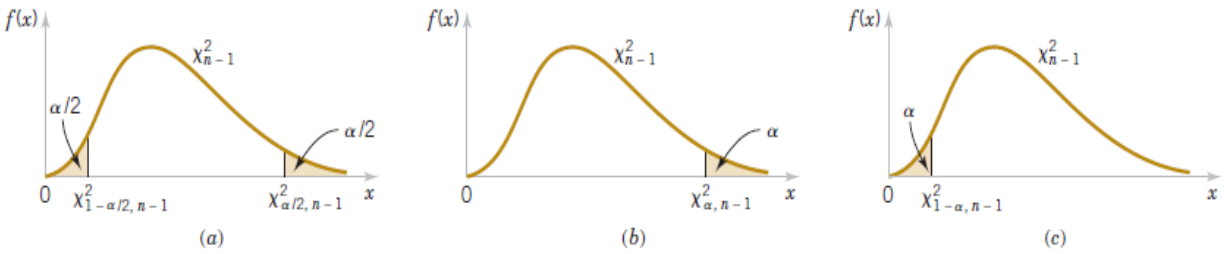


Figure 8.5 Reference distribution for the test of $H_0 : \sigma^2 = \sigma_0^2$ with critical region values for (a) $H_1 : \sigma^2 \neq \sigma_0^2$, (b) $H_0 : \sigma^2 > \sigma_0^2$, (c) $H_0 : \sigma^2 < \sigma_0^2$.

Example 8.6: An automated 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 exceeds 0.01 (fluid ounces)^2, an unacceptable proportion of bottles will be under filled or overfilled. Is there evidence in the sample data to suggest that the manufacturer has a problem with under filled or over filled bottles? Use α = 0.05, and assume that fill volume has a normal distribution.

Answer:

Using the seven-step procedure results in the following:

1. **Parameter of interest:** The parameter of interest is the population variance σ^2.
2. Null hypothesis: H_0: $\sigma^2 = 0.01$
3. Alternative hypothesis: H_a: $\sigma^2 > 0\ 01$
4. Test statistic: The test statistic is $\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$
5. Reject H_0 if: Use $\alpha = 0.05$, and reject H_0 if $\chi_0^2 > \chi^2_{0.05,19} = 30.14$.
6. Computations: $\chi_0^2 = \frac{19(0.0153)}{(0.01)^2} = 29.07$
7. Conclusions: Because $\chi_0^2 = 29.07 < \chi^2_{0.05,19} = 30.14$, we conclude that there is no strong evidence that the variance of fill volume exceeds 0.01 (fluid ounces)^2. So there is no strong evidence of a problem with incorrectly filled bottles.

Example 8.7

The population standard deviation of strengths of steel bars produced by a large manufacturer is 2.95. In order to meet tighter specifications engineers are trying to reduce the variability of the process. A sample of 28 bars gives a sample standard deviation of 2.65. Assume that the strengths of steel bars are normally distributed. Is there evidence at the 5% level of significance that the standard deviation has decreased?

Answer: H_0: $\sigma^2 = (2.95)^2 = 8.70$

Ha: $\sigma^2 < 8.70$ (one-tailed test)

The test statistic will be $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$. If $\chi^2_{calculated}$ is sufficiently small, then H_0 is not likely to be true.

$\chi^2_{calculated} = \frac{(n-1)s^2}{\sigma^2} = \frac{(28-1)(2.65)^2}{(2.95)^2} = 20.98$

From Table A3, for 5% probability in the lower tail (the left hand tail) and therefore 95% probability in the upper tail, the one to the right, and with 28 – 1 = 27 degrees of freedom, we find $\chi^2_{calculated} > \chi^2_{critical}$, so the calculated critical value does not fall in the cross-hatched tail for 5% probability. The population variance is not significantly less than 8.70, so the population standard deviation is not significantly less than 2.95. We do not have evidence at the 5% level of significance that the standard deviation of strengths of the steel bars has decreased.

For Practice:

1. The mean water temperature downstream from a discharge pipe at a power plant cooling tower should be no more than 100°F. Past experience has indicated that the standard deviation of temperature is 2°F. The water temperature is measured on nine randomly chosen days, and the average temperature is found to be 98°F.

(a) Is there evidence that the water temperature is acceptable at $\alpha = 0.05$?

(b) What is the P-value for this test?

(c) What is the probability of accepting the null hypothesis at $\alpha = 0.05$ if the water has a true mean temperature of 104°F?

2. Medical researchers have developed a new artificial heart constructed primarily of titanium and plastic. The heart will last and operate almost indefinitely once it is implanted in the patient's body, but the battery pack needs to be recharged about every four hours. A random sample of 50 battery packs is selected and subjected to a life test. The average life of these batteries is 4.05 hours. Assume that battery life is normally distributed with standard deviation $\sigma = 0.2$ hour.

(a) Is there evidence to support the claim that mean battery life exceeds 4 hours? Use $\alpha = 0.05$.

(b) What is the P-value for the test in part (a)?

3. A 1992 article in the Journal of the American Medical Association ("A Critical Appraisal of 98.6 Degrees F, the Upper Limit of the Normal Body Temperature, and Other Legacies of Carl Reinhold August Wunderlich") reported body temperature, gender, and heart rate for a number of subjects. The body temperatures for 25 female subjects follow: 97.8, 97.2, 97.4, 97.6, 97.8, 97.9, 98.0, 98.0, 98.0, 98.1, 98.2, 98.3, 98.3, 98.4, 98.4, 98.4, 98.5, 98.6, 98.6, 98.7, 98.8, 98.8, 98.9, 98.9, and 99.0.

Test the hypothesis H_0: $\mu = 98.6$ versus H_1: $\mu \neq 98.6$, using $\alpha = 0.05$. Find the P-value.

4. The sodium content of twenty 300-gram boxes of organic cornflakes was determined. The data (in milligrams) are as follows: 131.15, 130.69, 130.91, 129.54, 129.64, 128.77, 130.72, 128.33, 128.24, 129.65, 130.14, 129.29, 128.71, 129.00, 129.39, 130.42, 129.53, 130.12, 129.78, and 130.92.

Can you support a claim that mean sodium content of this brand of cornflakes differs from 130 milligrams? Use $\alpha = 0.05$. Find the P-value.