Mathematics for Data Science

Lecture 5: Tests of Hypotheses for a Single Sample

Mohamad GHASSANY

EFREI Paris

INTRODUCTION

In the previous two chapters, we showed how a parameter of a population can be estimated from sample data, using either a **point estimate** or an interval of likely values called a **confidence interval** . In many situations, a different type of problem is of interest; there are two competing claims about the value of a parameter, and the engineer must determine which claim is correct. For example, suppose that an engineer is designing an air crew escape system that consists of an ejection seat and a rocket motor that powers the seat. The rocket motor contains a propellant, and for the ejection seat to function properly, the propellant should have a mean burning rate of 50 cm/sec. If the burning rate is too low, the ejection seat may not function properly, leading to an unsafe ejection and possible injury of the pilot. Higher burning rates may imply instability in the propellant or an ejection seat that is too powerful, again leading to possible pilot injury. So the practical engineering question that must be answered is: Does the mean burning rate of the propellant equal 50 cm/sec, or is it some other value (either higher or lower)? This type of question can be answered using a statistical technique called **hypothesis testing**. This chapter focuses on the basic principles of hypothesis testing and provides techniques for solving the most common types of hypothesis testing problems involving a single sample of data.

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 described in the introduction. 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 deciding whether or not the mean burning rate is 50 centimeters per second. We may express this formally as

$$H_0: \mu = 50 \text{ centimeters per second} \qquad H_1: \mu \neq 50 \text{ centimeter per seconds} \qquad (9\text{-}1)$$

The statement $H_0: \mu = 50$ centimeters per second in Equation 9-1 is called the **null hypothesis**. This is a claim that is initially assumed to be true. The statement $H_1: \mu \neq 50$ centimeters per second is called the **alternative hypothesis** and it is a statement that condradicts 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

$$\begin{array}{ccc} H_0: \mu = 50 \text{ centimeters per second} & & H_0: \mu = 50 \text{ centimeters per second} \\ & \text{or} & \\ H_1: \mu < 50 \text{ centimeter per seconds} & & H_1: \mu > 50 \text{ centimeters per second} \end{array} \qquad (9\text{-}2)$$

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 $\geq$ and when the alternative hyphothesis is stated with the $>$ sign, the implicit claim in the null hypothesis can be taken as $\leq$.

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.

To illustrate the general concepts, consider the propellant burning rate problem introduced earlier. The null hypothesis is that the mean burning rate is 50 centimeters per second, and the alternate is that it is not equal to 50 centimeters per second. That is, we wish to test

$$H_0\colon \mu = 50 \text{ centimeters per second}$$

$$H_1\colon \mu \neq 50 \text{ centimeters per second}$$

Suppose that a sample of $n = 10$ specimens is tested and that the sample mean burning rate $\bar{x}$ is observed. The sample mean is an estimate of the true population mean μ. A value of the sample mean $\bar{x}$ that falls close to the hypothesized value of $\mu = 50$ centimeters per second does not conflict with the null hypothesis that the true mean μ is really 50 centimeters per second. On the other hand, a sample mean that is considerably different from 50 centimeters per second is evidence in support of the alternative hypothesis H_1. Thus, the sample mean is the test statistic in this case.

The sample mean can take on many different values. Suppose that if $48.5 \leq \bar{x} \leq 51.5$, we will not reject the null hypothesis $H_0\colon \mu = 50$, and if either $\bar{x} < 48.5$ or $\bar{x} > 51.5$, we will reject the null hypothesis in favor of the alternative hypothesis $H_1\colon \mu \neq 50$. This is illustrated in Fig. 9-1. The values of $\bar{x}$ that are less than 48.5 and greater than 51.5 constitute the **critical region** for the test; all values that are in the interval $48.5 \leq \bar{x} \leq 51.5$ form a region for which we will fail to reject the null hypothesis. By convention, this is usually called the **acceptance region**. The boundaries between the critical regions and the acceptance region are called the **critical values**. In our example, the critical values are 48.5 and 51.5. It is customary to state conclusions relative to the null hypothesis H_0. Therefore, we reject H_0 in favor of H_1 if the test statistic falls in the critical region and fail to reject H_0 otherwise.

This decision procedure can lead to either of two wrong conclusions. For example, the true mean burning rate of the propellant could be equal to 50 centimeters per second. However, for the randomly selected propellant specimens that are tested, we could observe a value of the test statistic $\bar{x}$ that falls into the critical region. We would then reject the null hypothesis H_0 in favor of the alternate H_1 when, in fact, H_0 is really true. This type of wrong conclusion is called a **type I error**.

Type I Error

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

Now suppose that the true mean burning rate is different from 50 centimeters per second, yet the sample mean $\bar{x}$ falls in the acceptance region. In this case, we would fail to reject H_0 when it is false. This type of wrong conclusion is called a **type II error**.

Type II Error

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

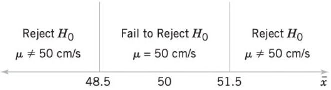

FIGURE 9-1 Decision criteria for testing $H_0: \mu = 50$ centimeters per second versus $H_1: \mu \neq 50$ centimeters per second.

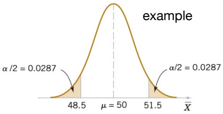

FIGURE 9-2 The critical region for $H_0: \mu = 50$ versus $H_1: \mu \neq 50$ and $n = 10$.

TABLE • 9-1 Decisions in Hypothesis Testing

Decision	H_0 Is True	H_0 Is False
Fail to reject H_0	No error	Type II error
Reject H_0	Type I error	No error

Thus, in testing any statistical hypothesis, four different situations determine whether the final decision is correct or in error. These situations are presented in Table 9-1.

Because our decision is based on random variables, probabilities can be associated with the type I and type II errors in Table 9-1. The probability of making a type I error is denoted by the Greek letter α.

Probability of Type I Error

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true}) \qquad (9\text{-}3)$$

Sometimes the type I error probability is called the **significance level**, the **α-error**, or the **size of the test**.

In evaluating a hypothesis-testing procedure, it is also important to examine the probability of a **type II error**, which we will denote by β. That is,

Probability of Type II Error

$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false}) \quad (9\text{-}4)$$

To calculate β (sometimes called the **β-error**), we must have a specific alternative hypothesis; that is, we must have a particular value of μ. For example, suppose that it is important to reject the null hypothesis H_0: $\mu = 50$ whenever the mean burning rate μ is greater than 52 centimeters per second or less than 48 centimeters per second. We could calculate the probability of a type II error β for the values $\mu = 52$ and $\mu = 48$ and use this result to tell us something about how the test procedure would perform. Specifically, how will the test procedure work if we wish to detect, that is, reject H_0, for a mean value of $\mu = 52$ or $\mu = 48$? Because of symmetry, it is necessary to evaluate only one of the two cases—say, find the probability of accepting the null hypothesis H_0: $\mu = 50$ centimeters per second when the true mean is $\mu = 52$ centimeters per second.

Generally, the analyst controls the type I error probability α when he or she selects the critical values. Thus, it is usually easy for the analyst to set the type I error probability at (or near) any desired value. Because the analyst can directly control the probability of wrongly rejecting H_0, we always think of rejection of the null hypothesis H_0 as a **strong conclusion**.

Because we can control the probability of making a type I error (or significance level), a logical question is what value should be used. The type I error probability is a measure of risk, specifically, the risk of concluding that the null hypothesis is false when it really is not. So, the value of α should be chosen to reflect the consequences (economic, social, etc.) of incorrectly rejecting the null hypothesis. Smaller values of α would reflect more serious consequences and larger values of α would be consistent with less severe consequences. This is often hard to do, so what has evolved in much of scientific and engineering practice is to use the value $\alpha = 0.05$ in most situations unless information is available that this is an inappropriate choice. In the rocket propellant problem with $n = 10$, this would correspond to critical values of 48.45 and 51.55.

> A widely used procedure in hypothesis testing is to use a type 1 error or significance level of $\alpha = 0.05$. This value has evolved through experience and may not be appropriate for all situations.

On the other hand, the probability of type II error β is not a constant but depends on the true value of the parameter. It also depends on the sample size that we have selected. Because the type II error probability β is a function of both the sample size and the extent to which the null hypothesis H_0 is false, it is customary to think of the decision to accept H_0 as a **weak conclusion** unless we know that β is acceptably small. Therefore, rather than saying we "accept H_0," we prefer the terminology "fail to reject H_0." Failing to reject H_0 implies that we have not found sufficient evidence to reject H_0, that is, to make a strong statement. Failing to reject H_0 does not necessarily mean that there is a high probability that H_0 is true. It may simply mean that more data are required to reach a strong conclusion. This can have important implications for the formulation of hypotheses.

A useful analog exists between hypothesis testing and a jury trial. In a trial, the defendant is assumed innocent (this is like assuming the null hypothesis to be true). If strong evidence is found to the contrary, the defendant is declared to be guilty (we reject the null hypothesis). If evidence is insufficient, the defendant is declared to be not guilty. This is not the same as proving the defendant innocent and so, like failing to reject the null hypothesis, it is a weak conclusion.

An important concept that we will use is the power of a statistical test.

Power

> The **power** of a statistical test is the probability of rejecting the null hypothesis H_0 when the alternative hypothesis is true.

The power is computed as $1 - \beta$, and power can be interpreted as **the probability of correctly rejecting a false null hypothesis**. We often compare statistical tests by comparing their power properties. For example, consider the propellant burning rate problem when we are testing $H_0: \mu = 50$ centimeters per second against $H_1: \mu \neq 50$ centimeters per second. Suppose that the true value of the mean is $\mu = 52$. When $n = 10$, we found that $\beta = 0.2643$, so the power of this test is $1 - \beta = 1 - 0.2643 = 0.7357$ when $\mu = 52$.

Power is a very descriptive and concise measure of the **sensitivity** of a statistical test when by sensitivity we mean the ability of the test to detect differences. In this case, the sensitivity of the test for detecting the difference between a mean burning rate of 50 centimeters per second and 52 centimeters per second is 0.7357. That is, if the true mean is really 52 centimeters per second, this test will correctly reject $H_0: \mu = 50$ and "detect" this difference 73.57% of the time. If this value of power is judged to be too low, the analyst can increase either α or the sample size n.

One-Sided and Two-Sided Hypotheses

In formulating one-sided alternative hypotheses, we should remember that rejecting H_0 is always a strong conclusion. Consequently, we should put the statement about which it is important to make a strong conclusion in the alternative hypothesis. In real-world problems, this will often depend on our point of view and experience with the situation.

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.

Operationally, once a *P*-value is computed, we typically compare it to a predefined significance level to make a decision. Often this predefined significance level is 0.05. However, in presenting results and conclusions, it is standard practice to report the observed *P*-value along with the decision that is made regarding the null hypothesis.

CONNECTION BETWEEN HYPOTHESIS TESTS AND CONFIDENCE INTERVALS

A close relationship exists between the test of a hypothesis about any parameter, say θ, and the confidence interval for θ. If $[l, u]$ is a $100(1-\alpha)\%$ confidence interval for the parameter θ, the test of size α of the hypothesis

$$H_0: \theta = \theta_0 \qquad H_1: \theta \neq \theta_0$$

will lead to rejection of H_0 if and only if θ_0 is **not** in the $100(1-\alpha)\%$ CI $[l, u]$. As an illustration, consider the escape system propellant problem with $\bar{x} = 51.3$, $\sigma = 2.5$, and $n = 16$. The null hypothesis H_0: $\mu = 50$ was rejected, using $\alpha = 0.05$. The 95% two-sided CI on μ can be calculated using Equation 8-7. This CI is $51.3 \pm 1.96\left(2.5/\sqrt{16}\right)$ and this is $50.075 \leq \mu \leq 52.525$. Because the value $\mu_0 = 50$ is not included in this interval, the null hypothesis H_0: $\mu = 50$ is rejected.

Although hypothesis tests and CIs are equivalent procedures insofar as decision making or **inference** about μ is concerned, each provides somewhat different insights. For instance, the confidence interval provides a range of likely values for μ at a stated confidence level whereas hypothesis testing is an easy framework for displaying the **risk levels** such as the *P*-value associated with a specific decision.

GENERAL PROCEDURE FOR HYPOTHESIS TESTS

This chapter 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 populatio 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 sampl mean $\bar{X}$ is an **unbiased point estimator** of μ with variance σ^2/n.

HYPOTHESIS TESTS ON THE MEAN

Suppose that we wish to test the hypotheses

$$H_0: \mu = \mu_0 \qquad H_1: \mu \neq \mu_0 \tag{9-7}$$

where μ_0 is a specified constant. We have a random sample $X_1, X_2, \ldots, X_n$ from a normal population. Because $\bar{X}$ has a normal distribution (i.e., the **sampling distribution** of $\bar{X}$ is normal) with mean μ_0 and standard deviation $\sigma/\sqrt{n}$ if the null hypothesis is true, we could calculate a P-value or construct a critical region based on the computed value of the sample mean $\bar{X}$, as in Section 9-1.2.

It is usually more convenient to **standardize** the sample mean and use a test statistic based on the standard normal distribution. That is, the test procedure for $H_0: \mu = \mu_0$ uses the **test statistic**:

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \tag{9-8}$$

If the null hypothesis $H_0: \mu = \mu_0$ is true, $E(\bar{X}) = \mu_0$, and it follows that the distribution of Z_0 is the standard normal distribution [denoted $N(0, 1)$].

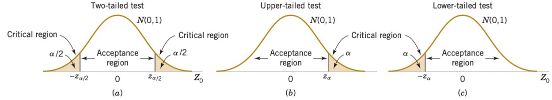

FIGURE 9-11 The distribution of Z_0 when $H_1: \mu = \mu_0$ is true with critical region for (a) The two-sided alternative $H_1: \mu \neq \mu_0$ (b) The one-sided alternative $H_1: \mu > \mu_0$. (c) The one-sided alternative $H_1: \mu < \mu_0$.

The **reference distribution** for this test is the standard normal distribution. The test is usually called a **z-test**.

We can also use the fixed significance level approach with the z-test. The only thing we have to do is determine where to place the critical regions for the two-sided and one-sided alternative hypotheses. First consider the two-sided alternative in Equation 9-10. Now if $H_0: \mu = \mu_0$ is true, the probability is $1 - \alpha$ that the test statistic Z_0 falls between $-z_{\alpha/2}$ and $z_{\alpha/2}$ where $z_{\alpha/2}$ is the $100_{\alpha/2}$ percentage point of the standard normal distribution. The regions associated with $z_{\alpha/2}$ and $-z_{\alpha/2}$ are illustrated in Fig. 9-11(a). Note that the probability is α that the test statistic Z_0 will fall in the region $Z_0 > z_{\alpha/2}$ or $Z_0 < -z_{\alpha/2}$, when $H_0: \mu = \mu_0$ is true. Clearly, a sample producing a value of the test statistic that falls in the tails of the distribution of Z_0 would be unusual if $H_0: \mu = \mu_0$ is true; therefore, it is an indication that H_0 is false. Thus, we should reject H_0 if either

$$z_0 > z_{\alpha/2} \tag{9-14}$$

or

$$z_0 < -z_{\alpha/2} \tag{9-15}$$

and we should fail to reject H_0 if

$$-z_{\alpha/2} \leq z_0 \leq z_{\alpha/2} \tag{9-16}$$

Equations 9-14 and 9-15 define the **critical region** or **rejection region** for the test. The type I error probability for this test procedure is α.

We may develop fixed significance level testing procedures for the one-sided alternatives. Consider the upper-tailed case in Equation 9-10.

In defining the critical region for this test, we observe that a negative value of the test statistic Z_0 would never lead us to conclude that $H_0: \mu = \mu_0$ is false. Therefore, we would place the critical region in the upper tail of the standard normal distribution and reject H_0 if the computed value z_0 is too large. Refer to Fig. 9-11(b). That is, we would reject H_0 if

$$z_0 > z_\alpha \tag{9-17}$$

Similarly, to test the lower-tailed case in Equation 9-12, we would calculate the test statistic Z_0 and reject H_0 if the value of Z_0 is too small. That is, the critical region is in the lower tail of the standard normal distribution as in Fig. 9-11(c), and we reject H_0 if

$$z_0 < -z_\alpha \tag{9-18}$$

Summary of Tests on the Mean, Variance Known

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

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

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

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

The *P*-values and critical regions for these situations are shown in Figs. 9-10 and 9-11.

Example 9-2 **Propellant Burning Rate** 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 $\sigma = 2$ centimeters per second. The experimenter decides to specify a type I error probability or significance level of $\alpha = 0.05$ 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?

We may solve this problem by following the seven-step procedure outlined in Section 9-1.6. This results in

1. **Parameter of interest:** The parameter of interest is μ, the mean burning rate.
2. **Null hypothesis:** H_0: $\mu = 50$ centimeters per second
3. **Alternative hypothesis:** H_1: $\mu \neq 50$ 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:** Reject H_0 if the P-value is less than 0.05. To use a fixed significance level test, the boundaries of the critical region would be $z_{0.025} = 1.96$ and $-z_{0.025} = -1.96$.
6. **Computations:** Because $\bar{x} = 51.3$ and $\sigma = 2$,
$$z_0 = \frac{51.3 - 50}{2 / \sqrt{25}} = 3.25$$
7. **Conclusion:** Because the P-value $= 2\left[1 - \Phi(3.25)\right] = 0.0012$ we reject H_0: $\mu = 50$ 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.

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. The situation is analogous to the one in Section 8-2 where we considered a **confidence interval** on the mean for the same situation. As in that section, the validity of the test procedure we will describe rests on the assumption that the population distribution is at least approximately normal. The important result on which the test procedure relies is that 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. Recall that we used this result in Section 8-2 to devise the t-confidence interval for μ. Now consider testing the hypotheses

$$H_0: \mu = \mu_0 \qquad H_1: \mu \neq \mu_0$$

We will use the **test statistic:**

Test Statistic

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \qquad (9\text{-}26)$$

If the null hypothesis is true, T_0 has a t distribution with $n - 1$ degrees of freedom. When we know the distribution of the test statistic when H_0 is true (this is often called the **reference distribution** or the **null distribution**), we can calculate the P-value from this distribution, or, if we use a fixed significance level approach, we can locate the critical region to control the type I error probability at the desired level.

The single-sample t-test we have just described can also be conducted using the **fixed significance level** approach. Consider the two-sided alternative hypothesis. The null hypothesis would be rejected if the value of the test statistic t_0 falls in the critical region defined by the lower and upper $\alpha/2$ percentage points of the t distribution with $n - 1$ degrees of freedom. That is, reject H_0 if

$$t_0 > t_{\alpha/2,n-1} \quad \text{or} \quad t_0 < -t_{\alpha/2,n-1}$$

For the one-tailed tests, the location of the critical region is determined by the direction to which the inequality in the alternative hypothesis "points." So, if the alternative is $H_1: \mu > \mu_0$, reject H_0 if

$$t_0 > t_{\alpha,n-1}$$

and if the alternative is $H_1: \mu < \mu_0$, reject H_0 if

$$t_0 < -t_{\alpha,n-1}$$

Figure 9-15 provides the locations of these critical regions.

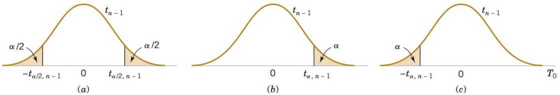

FIGURE 9-15 The distribution of T_0 when $H_0: \mu = \mu_0$ is true with critical region for (a) $H_1: \mu \neq \mu_0$, (b) $H_1: \mu > \mu_0$, and (c) $H_1: \mu < \mu_0$.

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: $T_0 = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$

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. 9-13 and 9-15, respectively.

Example 9-6 **Golf Club Design** The increased availability of light materials with high strength has revolutionized 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 restitution 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 restitution 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 $\alpha = 0.05$) to support a claim that the mean coefficient of restitution 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

The sample mean and sample standard deviation are $\bar{x} = 0.83725$ and $s = 0.02456$. The normal probability plot of the data in Fig. 9-16 supports the assumption that the coefficient of restitution is normally distributed. Because the experiment's objective is to demonstrate that the mean coefficient of restitution exceeds 0.82, a one-sided alternative hypothesis is appropriate.

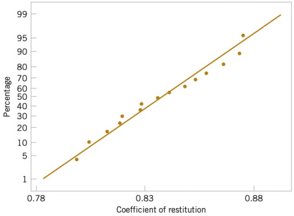

FIGURE 9-16. Normal probability plot of the coefficient of restitution data from Example 9-6.

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 restitution, μ.
2. **Null hypothesis:** $H_0: \mu = 0.82$
3. **Alternative hypothesis:** $H_1: \mu > 0.82$ We want to reject H_0 if the mean coefficient of restitution exceeds 0.82.
4. **Test statistic:** The test statistic is

$$t_0 = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

5. **Reject H_0 if:** Reject H_0 if the *P*-value is less than 0.05.
6. **Computations:** Because $\bar{x} = 0.83725$, $s = 0.02456$, $\mu_0 = 0.82$, and $n = 15$, we have

$$t_0 = \frac{0.83725 - 0.82}{0.02456/\sqrt{15}} = 2.72$$

7. **Conclusions:** From Appendix A Table II we find for a t distribution with 14 degrees of freedom that $t_0 = 2.72$ falls between two values: 2.624, for which $\alpha = 0.01$, and 2.977, for which $\alpha = 0.005$. Because this is a one-tailed test, we know that the *P*-value is between those two values, that is, $0.005 < P < 0.01$. Therefore, because $P < 0.05$, we reject H_0 and conclude that the mean coefficient of restitution exceeds 0.82.

Practical Interpretation: There is strong evidence to conclude that the mean coefficient of restitution exceeds 0.82.

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.

HYPOTHESIS TESTS ON THE VARIANCE

Suppose that we wish to test the hypothesis that the variance of a normal population σ^2 equals a specified value, say σ_0^2, or equivalently, that the standard deviation σ is equal to σ_0. Let $X_1, X_2, \ldots, X_n$ be a random sample of n observations from this population. To test

$$H_0\colon \sigma^2 = \sigma_0^2 \qquad H_1\colon \sigma^2 = \sigma_0^2 \tag{9-34}$$

we will use the test statistic:

Test Statistic

$$X_0^2 = \frac{(n-1)S^2}{\sigma_0^2} \tag{9-35}$$

If the null hypothesis $H_0\colon \sigma^2 = \sigma_0^2$ is true, the test statistic χ_0^2 defined in Equation 9-35 follows the chi-square distribution with $n - 1$ degrees of freedom. This is the reference distribution for this test procedure. To perform a fixed significance level test, we would take a random sample from the population of interest, calculate χ_0^2, the value of the test statistic χ_0^2, and the null hypothesis $H_0\colon \sigma^2 = \sigma_0^2$ would be rejected if

$$\chi_0^2 > \chi^2_{\alpha/2,n-1} \quad \text{or if} \quad \chi_0^2 > \chi^2_{1-\alpha/2,n-1}$$

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. Figure 9-17(a) shows the critical region.

The same test statistic is used for one-sided alternative hypotheses. For the one-sided hypotheses

$$H_0\colon \sigma^2 = \sigma_0^2 \qquad H_1\colon \sigma^2 > \sigma_0^2 \tag{9-36}$$

we would reject H_0 if $\chi_0^2 > \chi^2_{\alpha,n-1}$, whereas for the other one-sided hypotheses

$$H_0\colon \sigma^2 = \sigma_0^2 \qquad H_1\colon \sigma^2 < \sigma_0^2 \tag{9-37}$$

we would reject H_0 if $\chi_0^2 < \chi^2_{1-\alpha,n-1}$. The one-sided critical regions are shown in Fig. 9-17(b) and (c).

Tests on the Variance of a Normal Distribution

Null hypothesis: $H_0\colon \sigma^2 = \sigma^2_0$

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 $\chi_0^2 < \chi^2_{1-\alpha/2,n-1}$
$H_1: \sigma^2 > \sigma_0^2$	$\chi_0^2 > \chi^2_{\alpha,n-1}$
$H_1: \sigma^2 < \sigma_0^2$	$\chi_0^2 < \chi^2_{1-\alpha,n-1}$

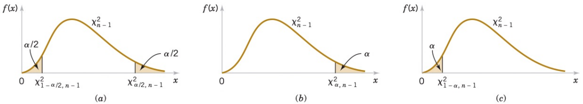

FIGURE 9-17 Reference distribution for the test of $H_0\colon \sigma^2 = \sigma_0^2$ with critical region values for (a) $H_1\colon \sigma^2 \neq \sigma_0^2$. (b) $H_1\colon \sigma^2 > \sigma_0^2$. (c) $H_1\colon \sigma^2 < \sigma_0^2$.

Example 9-8 **Automated Filling** 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 underfilled or overfilled. Is there evidence in the sample data to suggest that the manufacturer has a problem with underfilled or overfilled bottles? Use $\alpha = 0.05$, and assume that fill volume has a normal distribution.

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:** $H_0 : \sigma^2 > 0.01$
4. **Test statistic:** The test statistic is $\chi_0^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$
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:** $\chi_0^2 = \dfrac{19(0.0153)}{0.01} = 29.07$
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.

Tests on a Population Proportion

LARGE-SAMPLE TESTS ON A PROPORTION

Many engineering problems concern a random variable that follows the binomial distribution. For example, consider a production process that manufactures items that are classified as either acceptable or defective. Modelling the occurrence of defectives with the binomial distribution is usually reasonable when the binomial parameter p represents the proportion of defective items produced. Consequently, many engineering decision problems involve hypothesis testing about p.

We will consider testing

$$H_0 : p = p_0 \qquad H_1 : p = p_0 \tag{9-39}$$

An approximate test based on the normal approximation to the binomial will be given. As noted earlier, this approximate procedure will be valid as long as p is not extremely close to 0 or 1, and if the sample size is relatively large. Let X be the number of observations in a random sample of size n that belongs to the class associated with p. Then if the null hypothesis $H_0: p = p_0$ is true, we have $X \sim N[np_0, np_0(1-p_0)]$, approximately. To test $H_0: p = p_0$, calculate the test statistic

Test Statistic

$$Z_0 = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} \tag{9-40}$$

and determine the P-value. Because the test statistic follows a standard normal distribution if H_0 is true, the P-value is calculated exactly like the P-value for the z-tests in Section 9-2. So for the two-sided alternative hypothesis, the P-value is the sum of the probability in the standard normal distribution above $|z_0|$ and the probability below the negative value $-|z_0|$, or

$$P = 2\left[1 - \Phi\left(|z_0|\right)\right]$$

For the one-sided alternative hypothesis $H_0: p > p_0$, the P-value is the probability above z_0, or

$$P = 1 - \Phi(z_0)$$

and for the one-sided alternative hypothesis $H_0: p < p_0$, the P-value is the probability below z_0, or

$$P = \Phi(z_0)$$

We can also perform a **fixed-significance-level** test. For the two-sided alternative hypothesis, we would reject $H_0 : p \neq p_0$ if

$$z_0 > z_{\alpha/2} \quad \text{or} \quad z_0 < -z_{\alpha/2}$$

Critical regions for the one-sided alternative hypotheses would be constructed in the usual manner.

Another form of the test statistic Z_0 in Equation 9-40 is occasionally encountered. Note that if X is the number of observations in a random sample of size n that belongs to a class of interest, then $\hat{P} = X/n$ is the sample proportion that belongs to that class. Now divide both numerator and denominator of Z_0 in Equation 9-40 by n, giving

$$Z_0 = \frac{X/n - p_0}{\sqrt{p_0(1-p_0)/n}} \quad \text{or} \quad Z_0 = \frac{\hat{P} - p_0}{\sqrt{p_0(1-p_0)/n}} \tag{9-41}$$

This presents the test statistic in terms of the sample proportion instead of the number of items X in the sample that belongs to the class of interest.

Summary of Approximate Tests on a Binomial Proportion

Testing Hypotheses on a Binomial Proportion

Null hypotheses: $H_0: p = p_0$

Test statistic: $$Z_0 = \frac{X - np_0}{\sqrt{np_0(1-p_0)}}$$

Alternative Hypotheses	***P*-Value**	**Rejection Criterion for Fixed-Level Tests**
$H_1: p \neq p_0$	Probability above $\lvert z_0 \rvert$ and probability below $-\lvert z_0 \rvert$, $P = 2[1 - \Phi(z_0)]$	$z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$
$H_1: p > p_0$	Probability above z_0, $P = 1 - \Phi(z_0)$	$z_0 > z_\alpha$
$H_1: p < p_0$	Probability below z_0, $P = \Phi(z_0)$	$z_0 < -z_\alpha$

Example 9-10 **Automobile Engine Controller** A semiconductor manufacturer produces controllers used in automobile engine applications. The customer requires that the process fallout or fraction defective at a critical manufacturing step not exceed 0.05 and that the manufacturer demonstrate process capability at this level of quality using $\alpha = 0.05$. The semiconductor manufacturer takes a random sample of 200 devices and finds that four of them are defective. Can the manufacturer demonstrate process capability for the customer?

We may solve this problem using the seven-step hypothesis-testing procedure as follows:

1. **Parameter of interest:** The parameter of interest is the process fraction defective p.
2. **Null hypothesis:** H_0: $p = 0.05$
3. **Alternative hypothesis:** $H_1: p < 0.05$

   This formulation of the problem will allow the manufacturer to make a strong claim about process capability if the null hypothesis H_0: $p = 0.05$ is rejected.
4. **Test statistic:** The test statistic is (from Equation 9-40): $z_0 = \dfrac{x - np_0}{\sqrt{np_0(1-p_0)}}$

   where $x = 4$, $n = 200$, and $p_0 = 0.05$.
5. **Reject H_0 if:** Reject H_0: $p = 0.05$ if the p-value is less than 0.05.
6. **Computation:** The test statistic is

$$z_0 = \frac{4 - 200(0.05)}{\sqrt{200(0.05)(0.95)}} = -1.95$$

7. **Conclusions:** Because $z_0 = -1.95$, the P-value is $\Phi(-1.95) = 0.0256$, so we reject H_0 and conclude that the process fraction defective p is less than 0.05.

   Practical Interpretation: We conclude that the process is capable.

Content is copied from book: Applied Statistics and Probability for Engineers