Unit 19
Formulating Hypotheses and Making Decisions

Objectives:

- To formulate a null hypothesis and an alternative hypothesis, and to choose a significance level
- To identify the Type I error and the Type II error in a hypothesis test
- To understand the role that sampling error plays in hypothesis testing
- To understand several definitions and concepts associated with hypothesis testing

Our primary reason for studying the properties of the sampling distributions of $\bar{x}$ and $\bar{p}$ is to begin exploring the kinds of conclusions that we may make about a population parameter, such as μ or λ, from a simple random sample. Drawing a conclusion about a parameter, such as μ or λ, from observing a statistic, such as $\bar{x}$ and $\bar{p}$ is what inferential statistics is all about. The type of inferential statistics that we shall consider first is known as *hypothesis testing*. Presently, we shall consider several of the new terms and new concepts which are important in hypothesis testing. Afterwards, we shall define a hypothesis test more formally.

To begin our discussion of hypothesis testing, we consider the claim of the manufacturer of a particular lighter that the lighter will ignite on the first try 75% of the time. To check this claim, we could obtain a simple random sample of observations by trying the lighter several times. Based on the results we observe from the simple random sample, we draw our conclusion concerning whether or not we believe the manufacturer's claim. In this illustration, the sample consists of the random observations that we actually make when we ignite the lighter, and the population from which the sample is selected consists of all the observations that we could ever possibly make from igniting the lighter. The parameter of interest is the long term proportion of times the lighter will ignite on the first try out of all possible attempts to ignite the lighter, and it seems reasonable that the statistic on which to base our conclusions about this population parameter should be the sample proportion of times the lighter ignites on the first try.

It would not be a sound procedure to conclude that the population proportion is different from 0.75 (which is what the manufacturer claims) simply because the proportion in a simple random sample is not exactly equal to 0.75. Even if the population proportion is exactly 0.75 (as the manufacturer claims), the proportion in a simple random sample will most likely not be exactly equal to 0.75 because of the random variation that exists in a sampling distribution. This random variation in a sampling distribution is called *sampling error*. Sampling error does not refer to a mistake or error that someone has made; it refers only to the natural variation that occurs with random selection. The whole purpose of hypothesis testing is to decide whether or not the difference between a hypothesized value for a population parameter and the value of a statistic observed in a simple random sample can be reasonably attributed to sampling error.

Performing a hypothesis test is analogous to conducting a court trial. In a court trial we must decide whether the defendant is guilty or innocent. Evidence is collected and presented. The evidence must then be evaluated. Based on the evaluation of the evidence, a verdict is then returned. Let us now see how performing a hypothesis test is analogous to the process of reaching a verdict in a court trial.

The purpose of a hypothesis test is to decide which of two competing hypotheses should be believed; however, we do not treat these two hypotheses the same. Consider the court trial where the two hypotheses are "The defendant is innocent" and the "The defendant is guilty." Since the court trial proceeds under the assumption that the defendant is innocent until there is sufficient evidence that the defendant is guilty, the hypothesis stating "The defendant is innocent" is assumed to be true at the outset, unless and until there is sufficient evidence suggesting that we should believe the hypothesis "The defendant is guilty." This is precisely how a hypothesis test is performed. One of the two hypotheses is assumed to be true at the outset until there is sufficient evidence indicating that the other hypothesis is true.

In a hypothesis test, the hypothesis which is assumed to be true at the outset is called the *null hypothesis*. The hypothesis for which sufficient evidence is required before it will be believed is called the *alternative hypothesis* (or sometimes also called the *research hypothesis*). In the illustration concerning the lighter, we are trying to decide whether or not to believe the manufacturer's claim that the lighter will ignite on the first try 75% of the time. If you imagine that the manufacturer's claim is on trial, you will see that we assume the claim to be true at the outset, unless and until we find sufficient evidence to suggest that the claim is not true. In other words, "The lighter will ignite on the first try 75% of the time" is the null hypothesis and "The lighter will not ignite on the first try 75% of the time" is the alternative hypothesis.

In either a hypothesis test or a court trial, we can identify two types of errors which can occur. One type of error is to believe the alternative hypothesis when in reality the null hypothesis is true; this is called a *Type I error*. The other type of error is to believe the null hypothesis when in reality the alternative hypothesis is true; this is called a *Type II error*. In a court trial, the null hypothesis is "The defendant is innocent," and the alternative hypothesis is "The defendant is guilty"; consequently, in a court trial, a Type I error is concluding that the defendant is guilty when in reality the defendant is innocent, and a Type II error is concluding that the defendant is innocent when in reality the defendant is guilty. In the illustration concerning the lighter, a Type I error is concluding that the lighter will not ignite on the first try 75% of the time when in reality the lighter will ignite on the first try 75% of the time, and a Type II error is concluding that the lighter will ignite on the first try 75% of the time when in reality the lighter will not ignite on the first try 75% of the time.

The Type I error is generally treated as the more serious error. For instance, in a court trial, declaring an innocent defendant to be guilty is considered a more serious error than declaring a guilty defendant to be innocent. Of course, we would prefer never making either type of error, but since we have to live with the fact that making errors is unavoidable, we are faced with the choice of which error to guard against more closely. The fact that we choose to guard more closely against making a Type I error than against making a Type II error is the reason why court trials and hypothesis tests are structured the way they are.

A *significance level* is defined to be the highest probability of making a Type I error that we are willing to tolerate. In a court trial, the criteria for believing that the defendant is guilty is typically that there be "no reasonable doubt," but this is practically impossible to quantify in terms of a probability. In a hypothesis test, however, we can choose a specific probability for our significance level. In the illustration concerning the lighter, ask yourself what probability could you live with for concluding that the lighter will not ignite on the first try 75% of the time when in reality the lighter will ignite on the first try 75% of the time (or, in other words, what probability could you live with for concluding that the manufacturer’s claim is false when in reality the claim is true) . Ideally, of course, we would have this probability be zero, but this is not practical, since, as we have already stated, the chance of making an error is going to be unavoidable. Commonly chosen significance levels are 0.10, 0.05, and 0.01, and the Greek letter α (alpha) is used to denote the significance level.

With a given sample size n, the probability of making a Type II error increases as we decrease α (our chosen probability of making a Type I error). For this reason we may want to avoid choosing α to be too small. Let us suppose that in the illustration concerning the lighter, we choose a significance level of $\alpha = 0.10$. This implies that we can tolerate a 0.10 probability for concluding that the lighter will not ignite on the first try 75% of the time when in reality the lighter will ignite on the first try 75% of the time. In other words, we are willing to live with a 0.10 probability of believing that the manufacturer's claim is false when in reality the claim is true.

As a general rule in a hypothesis test about a population parameter such as λ , the null hypothesis is a statement of equality, i.e., a statement that the parameter is equal to a specific hypothesized value; the alternative hypothesis is a statement of inequality, such as a statement that the parameter is not equal to the hypothesized value. In the illustration concerning the lighter, we make a decision about which hypothesis to believe by looking at the difference between the sample proportion $\overline{p}$ which is a statistic, and the hypothesized proportion 0.75. If this difference falls within the bounds we would expect from sampling error (i.e., random variation), then we have no reason to doubt that the hypothesized proportion 0.75 is correct; but if this difference cannot be reasonably attributed to sampling error, then we are inclined to believe that the hypothesized proportion 0.75 is not correct. In other words, the difference between the sample proportion $\overline{p}$ and the hypothesized proportion 0.75 plays the same role in our hypothesis test that the evidence against the defendant plays in a court trial.

How can we tell if the difference between the sample proportion $\overline{p}$ and the hypothesized proportion 0.75 falls inside or outside the bounds we would expect from sampling error? We use our knowledge that with a sufficiently large sample size n, the sampling distribution $\bar{p}$ can be treated as a normal distribution. Since we know that practically all of the area under a normal density curve is within two or three standard deviations of the mean, a difference between $\bar{p}$ and 0.75 of more than two or three standard deviations would make it very difficult for us to believe that this difference occurred as a result of random variation; we would be more inclined to believe that this unusually large difference occurred because the hypothesized 0.75 is not correct.

Our suggestion that a difference of more than three standard deviations between the sample proportion $\bar{p}$ and the hypothesized proportion 0.75 inclines us to believe that the hypothesized 0.75 is not correct is not a very precise description of exactly how much evidence against the null hypothesis is required to believe the alternative hypothesis. However, the choice of significance level α tells us precisely how strong our evidence against the null hypothesis must be in order for us to believe the alternative hypothesis. Once we have obtained the value of $\bar{p}$ in our sample, we can calculate the probability of obtaining a sample proportion farther away from (below or above) 0.75 under the assumption that the hypothesized 0.75 is correct. If this probability is smaller than the significance level α, then the fact that our sample is so unlikely under the assumption that the hypothesized 0.75 is correct leads us to believe that 0.75 is not correct; if this probability is larger than the significance level α, then the fact that our sample is not unlikely under the assumption that the hypothesized 0.75 is correct does not provide us with sufficiently strong evidence to doubt that 0.75 is correct.

To illustrate, imagine that you observe a simple random sample of n = 300 attempts to ignite the lighter, and you find that the lighter ignites on the first try in 214 of these attempts. The sample proportion of successes is $\bar{p}$ = 214/300 = 0.7133, which is a distance of 0.75 – 0.7133 = 0.0367 away from the hypothesized 0.75. To decide whether this difference falls inside or outside the bounds we would expect from sampling error, we shall calculate the probability of obtaining a sample proportion farther away from (below or above) 0.75, under the assumption that λ = 0.75 is the population proportion of times the lighter will ignite on the first try (i.e., under the assumption that the null hypothesis is true).

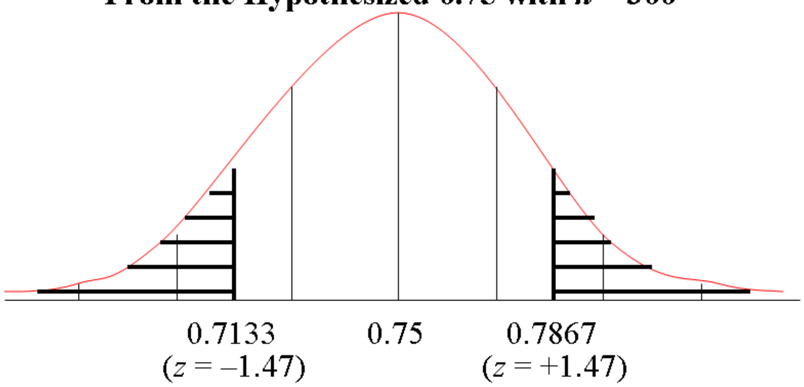
We shall obtain this probability by using the fact that the sampling distribution of $\bar{p}$ can be treated as a normal distribution. Under the assumption that the null hypothesis is true, this sampling distribution has mean $\mu_{\bar{p}} = \lambda = 0.75$ and standard deviation $\sigma_{\bar{p}} = \sqrt{(0.75)(1-0.75)/300} = 0.025$. To find the probability that $\bar{p}$ from a simple random sample of n = 300 attempts to ignite the lighter will be more than 0.0367 away from (below or above) the hypothesized λ = 0.75, we first use $\mu_{\bar{p}}$ = 0.75 and $\sigma_{\bar{p}}$ = 0.025 to find the z-score of 0.7133 to be

$$\frac{0.7133-0.75}{0.025} = -1.47 \ ,$$

and the z-score of 0.7867 to be

$$\frac{0.7867-0.75}{0.025} = +1.47 \ .$$

Figure 19-1
Sample Proportions More Than 0.0367 Away From the Hypothesized 0.75 with n = 300



This desired probability is the proportion of shaded area in Figure 19-1. The shaded area below 0.7133 in Figure 19-1 is a mirror image of the shaded area in the figure at the top of Table A.2; also, the shaded area above 0.7867 in Figure 19-1 corresponds exactly to the shaded area in the figure at the top of Table A.2. Since a normal distribution is symmetric, we can find the total shaded area by doubling the entry of Table A.2 in the row labeled 1.4 and the column labeled 0.07. The probability that $\bar{p}$ from a simple random sample of n = 300 attempts to ignite the lighter will not be between 0.7133 and 0.7867 is 0.0708 + 0.0708 = 0.1416 (or 14.16%).

Earlier, we chose a significance level of $\alpha = 0.10$. In order to be convinced that the hypothesized $\lambda = 0.75$ is not correct, the probability of obtaining a sample proportion $\bar{p}$ farther away from the hypothesized $\lambda = 0.75$ than the observed sample proportion, $\bar{p}$ = 214/300 = 0.7133, must smaller than our significance level $\alpha = 0.10$. Since we found this probability to be 0.1416, which is not smaller than $\alpha = 0.10$, the simple random sample we observed with $\bar{p}$ = 214/300 = 0.7133 does not provide us with sufficient evidence to convince us that the hypothesized $\lambda = 0.75$ is not correct.

Let us now instead imagine that when you observe a simple random sample of $n = 300$ attempts to ignite the lighter, you find that the lighter ignites on the first try in 210 of these attempts. The sample proportion of successes is $\bar{p}$ = 210/300 = 0.70, which is a distance of 0.75 – 0.70 = 0.05 away from the hypothesized 0.75. To decide whether this difference falls inside or outside the bounds we would expect from sampling error, we shall calculate the probability of obtaining a sample proportion farther away from (below or above) 0.75, under the assumption that $\lambda = 0.75$ is the population proportion of times the lighter will ignite on the first try (i.e., under the assumption that the null hypothesis is true).

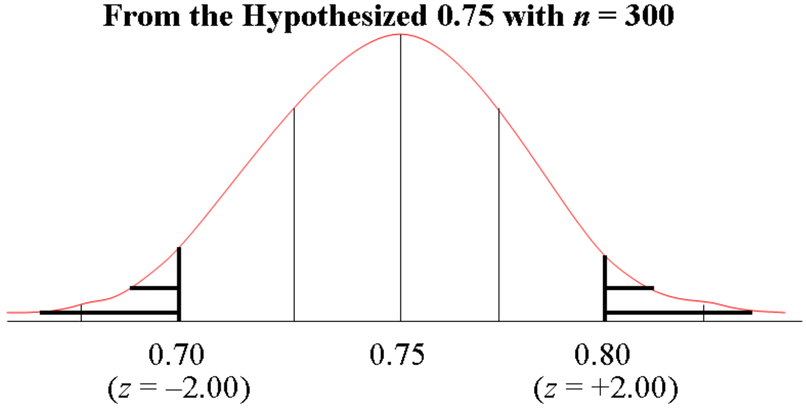
We shall obtain this probability by using the fact that the sampling distribution of $\bar{p}$ can be treated as a normal distribution; as before, under the assumption that the null hypothesis is true, this sampling distribution has mean $\mu_{\bar{p}} = \lambda = 0.75$ and standard deviation $\sigma_{\bar{p}} = \sqrt{(0.75)(1-0.75)/300} = 0.025$. To find the probability that $\bar{p}$ from a simple random sample of $n = 300$ attempts to ignite the lighter will be more than 0.05 away from (below or above) the hypothesized $\lambda = 0.75$, we first use $\mu_{\bar{p}} = 0.75$ and $\sigma_{\bar{p}} = 0.025$ to find the *z*-score of 0.70 to be

$$\frac{0.70-0.75}{0.025} = -2.00 \ ,$$

and the *z*-score of 0.80 to be

$$\frac{0.80-0.75}{0.025} = +2.00 \ .$$

Figure 19-2
Sample Proportions More Than 0.05 Away From the Hypothesized 0.75 with *n* = 300



This desired probability is the proportion of shaded area in Figure 19-2. The shaded area below 0.70 in Figure 19-2 is a mirror image of the shaded area in the figure at the top of Table A.2; also, the shaded area above 0.80 in Figure 19-2 corresponds exactly to the shaded area in the figure at the top of Table A.2. Since a normal distribution is symmetric, we can find the total shaded area by doubling the entry of Table A.2 in the row labeled 2.0 and the column labeled 0.00. The probability that $\bar{p}$ from a simple random sample of $n = 300$ attempts to ignite the lighter will not be between 0.70 and 0.80 is 0.0228 + 0.0228 = 0.0454 (or 4.54%).

In order to be convinced that the hypothesized $\lambda = 0.75$ is not correct, the probability of obtaining a sample proportion $\bar{p}$ farther away from the hypothesized $\lambda = 0.75$ than the observed sample proportion, $\bar{p}$ = 210/300 = 0.70, must smaller than our significance level $\alpha = 0.10$. Since we found this probability to be 0.0454, which is not smaller than $\alpha = 0.10$, the simple random sample we observed with $\bar{p}$ = 210/300 = 0.70 provides us with sufficient evidence to convince us that the hypothesized $\lambda = 0.75$ is not correct.

We have now seen that with a significance level of $\alpha = 0.10$, the sample proportion $\bar{p}$ = 214/300 = 0.7133 does not offer sufficiently strong evidence against the hypothesized $\lambda = 0.75$, whereas the sample proportion $\bar{p}$ = 210/300 = 0.70 does give us sufficiently strong evidence against the hypothesized λ = 0.75. Note that our conclusion in each situation would not change if we chose a significance level of α = 0.05 instead; however, if we chose a significance level of α = 0.01, neither situation would provide us with sufficiently strong evidence against the hypothesized λ = 0.75.

In a court trial, deciding whether or not the evidence is sufficiently strong to believe that the defendant is guilty (i.e. sufficiently strong against the null hypothesis of innocence) often involves subjective judgment. In our hypothesis test, deciding whether or not the evidence is sufficiently strong to believe the alternative hypothesis (i.e. sufficiently strong against the null hypothesis) was simply a matter of deciding whether or not the probability of obtaining a larger difference between the observed sample proportion $\bar{p}$ and the hypothesized λ = 0.75 is smaller than the chosen significance level α; however, the choice of the significance level α is a matter of subjective judgment.

We see then that except for the probability calculations that we employ, the manner in which we make a decision in a hypothesis test and the manner in which we make a decision in a court trial are very similar. Thus far, we have only introduced the basic general concepts in hypothesis testing. Next, we shall discuss a formal four-step process for performing any hypothesis test.

Self-Test Problem 19-1. City officials wish to perform a hypothesis test to see if there is any evidence that the proportion of senior citizens among all the passengers who ride the city bus is different from 0.3.

(a) State the null and alternative hypotheses.

(b) Identify which of the following is a Type I error and which of the following is a Type II error:

(i) concluding that the proportion of passengers who are senior citizens is 0.3, when in reality the proportion of senior citizens is 0.3;

(ii) concluding that the proportion of passengers who are senior citizens is 0.3, when in reality the proportion of senior citizens is different from 0.3;

(iii) concluding that the proportion of passengers who are senior citizens is different from 0.3, when in reality the proportion of senior citizens is 0.3;

(iv) concluding that the proportion of passengers who are senior citizens is different from 0.3, when in reality the proportion of senior citizens is different from 0.3.

(c) Suppose that a simple random sample of 350 passengers is selected, and the sample proportion $\bar{p}$ of senior citizens is obtained. Why can we treat the sampling distribution of $\bar{p}$ as a normal distribution, under the assumption that the hypothesized λ = 0.3 is correct?

(d) Do each of (i) through (iv) which follow under the assumption that the hypothesized λ = 0.3 is correct.

(i) Find the mean and standard deviation for the sampling distribution of $\bar{p}$ with simple random samples of size 350.

(ii) Suppose that in a simple random sample of 350 passengers, 126 are found to be senior citizens. Find the sample proportion $\bar{p}$ of senior citizens, and find the *z*-score for this sample proportion.

(iii) Find the probability of obtaining a sample proportion farther away from (below or above) the hypothesized λ = 0.3 than the one calculated in part (ii).

(iv) Decide whether the random sample of 350 passengers with 126 senior citizens provides sufficient evidence against the hypothesized λ = 0.3 for us to believe that the population proportion is different from 0.3, if a

significance level of α = 0.10 is chosen,
significance level of α = 0.05 is chosen,
significance level of α = 0.01 is chosen.

(*continued next page*)

Self-Test Problem 19-1 (*continued from previous page*)

(e) Do each of (i) through (iv) which follow under the assumption that the hypothesized $\lambda = 0.3$ is correct.

(i) Find the mean and standard deviation for the sampling distribution of $\overline{p}$ with simple random samples of size 700.

(ii) Suppose that in a simple random sample of 700 passengers, 252 are found to be senior citizens. Find the sample proportion $\overline{p}$ of senior citizens, and find the *z*-score for this sample proportion.

(iii) Find the probability of obtaining a sample proportion farther away from (below or above) the hypothesized $\lambda = 0.3$ than the one calculated in part (ii).

(iv) Decide whether the random sample of 700 passengers with 252 senior citizens provides sufficient evidence against the hypothesized $\lambda = 0.3$ for us to believe that the population proportion is different from 0.3, if a

significance level of $\alpha = 0.10$ is chosen,
significance level of $\alpha = 0.05$ is chosen,
significance level of $\alpha = 0.01$ is chosen.

Answers to Self-Test Problems

19-1 (a) The null hypothesis is "The proportion of senior citizens among all passengers is 0.3"; the alternative hypothesis is "The proportion of senior citizens among all passengers is different from 0.3".

(b) Statement (iii) is the Type I error, and statement (ii) is the Type II error.

(c) We may treat the sampling distribution of $\overline{p}$ with $n = 350$ as a normal distribution because $350(0.3) = 105$ and $350(1 – 0.3) = 245$ are both greater than 5.

(d) (i) $\mu_{\overline{p}} = 0.3$ and $\sigma_{\overline{p}} = 0.02449$.

(ii) The sample proportion $\overline{p}$ is 126/350 = 0.36, and the *z*-score is +2.45.

(iii) 0.0142

(iv) Since 0.0142 is smaller than the significance level 0.10, the sample provides sufficient evidence against the hypothesized $\lambda = 0.3$ with $\alpha = 0.10$.
Since 0.0142 is smaller than the significance level 0.05, the sample provides sufficient evidence against the hypothesized $\lambda = 0.3$ with $\alpha = 0.05$.
Since 0.0142 is larger than the significance level 0.01, the sample does not provide sufficient evidence against the hypothesized $\lambda = 0.3$ with $\alpha = 0.01$.

(e) (i) $\mu_{\overline{p}} = 0.3$ and $\sigma_{\overline{p}} = 0.01732$.

(ii) The sample proportion $\overline{p}$ is 252/700 = 0.36, and the *z*-score is +3.46.

(iii) practically zero

(iv) Since the probability in (iii) is smaller than the significance level 0.10, the sample provides sufficient evidence against the hypothesized $\lambda = 0.3$ with $\alpha = 0.10$.
Since the probability in (iii) is smaller than the significance level 0.05, the sample provides sufficient evidence against the hypothesized $\lambda = 0.3$ with $\alpha = 0.05$.
Since the probability in (iii) is smaller than the significance level 0.01, the sample provides sufficient evidence against the hypothesized $\lambda = 0.3$ with $\alpha = 0.01$

Summary

One type of inferential statistics (drawing a conclusion about a parameter, such as μ or λ from observing a statistic, such as $\bar{x}$ or $\bar{p}$) is *hypothesis testing*. The purpose of a hypothesis test is to decide which of two competing hypotheses should be believed; performing a hypothesis test is analogous to the process of reaching a verdict in a court trial. Just as a court trial proceeds under the assumption that the defendant is innocent unless and until there is sufficient evidence to indicate the defendant is guilty, a hypothesis test is performed under the assumption that one hypothesis is true unless and until there is sufficient evidence to indicate the other hypothesis is true. The hypothesis which is assumed to be true at the outset is called the *null hypothesis*. The hypothesis for which sufficient evidence is required before it will be believed is called the *alternative hypothesis* (or sometimes also called the research hypothesis). In general, the null hypothesis is a statement of equality, and the alternative hypothesis is a statement of inequality.

Two types of errors are possible in a hypothesis test. A *Type I error* is believing the alternative hypothesis when in reality the null hypothesis is true; a *Type II error* is believing the null hypothesis when in reality the alternative hypothesis is true. The Type I error is generally treated as the more serious error. A *significance level* is defined to be the highest probability of making a Type I error that we are willing to tolerate. Commonly chosen significance levels are 0.10, 0.05, and 0.01, and the Greek letter α (alpha) is used to denote the significance level. With a given sample size *n*, the probability of making a Type II error increases as we decrease α (our chosen probability of making a Type I error).

In a hypothesis test about a population parameter such as λ, the null hypothesis is usually a statement of equality, i.e., a statement that the parameter is equal to a specific hypothesized value; the alternative hypothesis is a statement of inequality, such as a statement that the parameter is not equal to the hypothesized value. We make a decision about which hypothesis to believe by looking at the difference between a statistic from the sample and the hypothesized value, such as the difference between the sample proportion $\bar{p}$ and a hypothesized value for the population proportion λ . If this difference falls within the bounds we would expect from the random variation in a sampling distribution, which is called *sampling error*, then we have no reason to doubt that the hypothesized value is correct; but if this difference cannot be reasonably attributed to sampling error, then we are inclined to believe that the hypothesized value is not correct.

To decide if the difference between a statistic and a hypothesized value for a parameter, such as the difference between sample proportion $\bar{p}$ and a hypothesized value for the population proportion λ , falls inside or outside the bounds we would expect from sampling error, we calculate the probability of obtaining a larger distance away from (below or above) the hypothesized value under the assumption that the hypothesized value is correct. If this probability is smaller than the significance level α , then the fact that our sample is so unlikely under the assumption that the hypothesized value is correct leads us to believe that the hypothesized value is not correct; if this probability is larger than the significance level α, then the fact that our sample is not unlikely under the assumption that the hypothesized value is correct does not provide us with sufficiently strong evidence to doubt the hypothesized value.