UNIT III INFERENTIAL STATISTICS

Populations – samples – random sampling – Sampling distribution- standard error of the mean - Hypothesis testing – z-test – z-test procedure –decision rule – calculations – decisions – interpretations - one-tailed and two-tailed tests – Estimation – point estimate – confidence interval – level of confidence – effect of sample size.

3.1 POPULATIONS

*Any complete set of observations (or potential observations) may be characterized as a **Population.*** Accurate descriptions of populations specify the nature of the observations to be taken. For example, a population might be described as "attitudes toward abortion of currently enrolled students at Bucknell University" or as "SAT critical reading scores of currently enrolled students at Rutgers University".

1. **Real Populations**

Pollsters, such as the Gallup Organization, deal with real populations. A *real* population is one in which all potential observations are accessible at the time of sampling. Examples of real populations, the ages of all visitors to Disneyland on a given day, the ethnic backgrounds of all current employees of the U.S. Postal Department, and presidential preferences of all currently registered voters in the United States. Incidentally, federal law requires that a complete survey be taken every 10 years of the real population of all U.S. house- holds at considerable expense, involving thousands of data collectors as a means of revising election districts for the House of Representatives. (An estimated undercount of millions of people, particularly minorities, in both the 2000 and 2010 censuses has revived a suggestion, long endorsed by statisticians, that the entire U.S. population could be estimated more accurately if a highly trained group of data collectors focused only on a random sample of households.).

2. **Hypothetical Populations**

A *hypothetical* population is one in which all potential observations are not accessible at the time of sampling. In most experiments, subjects are selected from very small, uninspiring real populations: the lab rats housed in the local animal colony or student volunteers from general psychology classes. Experimental subjects often are viewed, nevertheless, as a sample from a much larger hypothetical population, loosely described as "the scores of all similar animal subjects (or student volunteers) who could conceivably undergo the present experiment." According to the rules of inferential statistics, generalizations should be made only to real populations that, in fact, have been sampled. Generalizations to hypothetical populations should be viewed, therefore, as provisional conclusions based on the wisdom of the researcher rather than on any logical or statistical necessity. In effect, it's an open question often answered only by additional experimentation whether or not a given experimental finding merits the generality assigned to it by the researcher.

3.1.2 SAMPLES

Any subset of observations from a population may be characterized as a **sample**. In typical applications of inferential statistics, the sample size is small relative to the population size. For example, less than 1 percent of all U.S. worksites are included in the Bureau of Labor Statistics' monthly survey to estimate the rate of unemployment. And although, only 1475 likely voters had been sampled in the final poll for the 2012 presidential election by the NBC News/*Wall Street Journal*, it correctly predicted that Obama would be the slim winner of the popular vote.

Optimal Sample Size

There is no simple rule of thumb for determining the best or optimal sample size for any particular situation. Often sample sizes are in the hundreds or even the thousands for surveys, but they are less than 100 for most experiments. Optimal sample size depends on the answers to a number of questions, including "What is the estimated variability among observations?" and "What is an acceptable amount of error in our conclusion?" Once these types of questions have been answered that is the result, the specific procedures can be followed to determine the optimal sample size for any situation.

3.1.3 RANDOM SAMPLING

The valid use of techniques from inferential statistics requires that samples be random.

***Random sampling* occurs if, at each stage of sampling, the selection process guarantees that all potential observations in the population have an equal chance of being included in the sample.**

It's important to note that randomness describes the *selection process* that is, the conditions under which the sample is taken and not the particular pattern of observations in the sample. Having established that sampling is random, you still can't predict anything about the unique pattern of observations in that sample. The observations in the sample should be representative of those in the population, but there is no guarantee that they actually will be.

Casual or Haphazard, Not Random

A casual or haphazard sample doesn't qualify as a random sample. Not every student at UC San Diego has an equal chance of being sampled if, for instance, a pollster casually selects only students who enter the student union. Obviously excluded from this sample are all those students (few, we hope) who never enter the student union. Even the final selection of students from among those who do enter the student union might reflect the pollster's various biases, such as an unconscious preference for attractive students who are walking alone.

3.2 Sampling distribution

WHAT IS A SAMPLING DISTRIBUTION?

Random samples rarely represent the underlying population exactly. Even a mean math score of 533 could originate, just by chance, from a population of freshmen whose mean equals the national average of 500. Accordingly, generalizations from a single sample to a population are much more tentative. Indeed, generalizations are based not merely on the single sample mean of 533 but also on its distribution a distribution of sample means for all possible random samples. Representing the statistician's model of random outcomes,

The *sampling distribution of the mean* refers to the probability distribution of means for all possible random samples of a given size from some population.

In effect, this distribution describes the variability among sample means that could occur just by chance and thereby serves as a frame of reference for generalizing from a single sample mean to a population mean.

The sampling distribution of the mean allows us to determine whether, given the variability among all possible sample means, the one observed sample mean can be viewed as a *common* outcome or as a *rare* outcome (from a distribution centered, in this case, about a value of 500). If the sample mean of 533 qualifies as a *common* outcome in this sampling distribution, then the difference between 533 and 500 isn't large enough, relative to the variability of all possible sample means, to signify that anything special is happening in the underlying population. Therefore, we can conclude that the mean math score for the entire freshman class could be the same as the national average of 500. On the other hand, if the sample mean of 533 qualifies as a *rare* outcome in this sampling distribution, then the difference between 533 and 500 is large enough, relative to the variability of all possible sample means, to signify that something special probably is happening in the underlying population. Therefore, we can conclude that the mean math score for the entire freshman class probably exceeds the national average of 500.

All Possible Random Samples

When attempting to generalize from a single sample mean to a population mean, must consult the sampling distribution of the mean. In the present case, this distribution is based on *all possible* random samples, each of size 100 that can be taken from the local population of freshmen. *All possible random samples* refers not to the number of samples of size 100 required to *survey completely* the local population of freshmen but to the number of different ways in which a *single* sample of size 100 can be selected from this population.

"All possible random samples" tends to be a huge number. For instance, if the local population contained at least 1,000 freshmen, the total number of possible random samples, each of size 100, would be astronomical in size. The 301 digits in this number would dwarf even the national debt. Even with the aid of a computer, it would be a horrendous task to construct this sampling distribution from scratch, itemizing each mean for all possible random samples.

Fortunately, statistical theory supplies us with considerable information about the sampling distribution of the mean, as will be discussed in the remainder of this chapter. Armed with this information about sampling distributions, we'll return to the current example in the next chapter and test the claim that the mean math score for the local population of freshmen equals the national average of 500. Only at that point and not at the end of this chapter should you expect to understand completely the role of sampling distributions in practical applications.

3.2.1 CREATING A SAMPLING DISTRIBUTION FROM SCRATCH

Let's establish precisely what constitutes a sampling distribution by creating one from scratch under highly simplified conditions. Imagine some ridiculously small population of four observations with values of 2, 3, 4, and 5, as shown in Figure 9.1**.** Next, itemize all possible random samples, each of size two, that could be taken from this population. There are four possibilities on the first draw from the population and also four possibilities on the second draw from the population, as indicated in Table 9.1**.*** The two sets of possibilities combine to yield a total of 16 possible samples. At this point, remember, we're clarifying the notion of a sampling distribution of the mean. In practice, only a single random sample, not 16 possible samples, would be taken from the population; the sample size would be very small relative to a much larger population size, and, of course, not all observations in the population would be known.

For each of the 16 possible samples, Table 9.1 also lists a sample mean (found by adding the two observations and dividing by 2) and its probability of occurrence (expressed as $^1/16$, since each of the 16 possible samples is equally likely). When cast into a relative frequency or probability distribution, as in Table 9.2**,** the 16 sample means constitute the sampling distribution of the mean, previously defined as the probability distribution of means for all possible random samples of a given size from some population. Not all values of the sample mean occur with equal probabilities in Table 9.2 since some values occur more than once among the 16 possible samples. For instance, a sample mean value of 3.5 appears among 4 of 16 possibilities and has a probability of $^4/16$.

1. **Probability of a Particular Sample Mean**

The distribution in Table 9.2 can be consulted to determine the probability of obtaining a particular sample mean or set of sample means. For example, the probability of a randomly selected sample mean of 5.0 equals $^1/16$ or .0625. According to the addition rule for mutually exclusive outcomes, the probability of a ran�domly selected sample mean of either 5.0 or 2.0 equals 1 /16 + 1 /16 = 2 /16 = .1250.

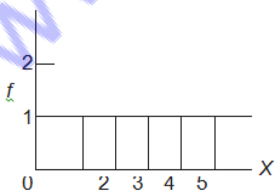


FIGURE 9.1
Graph of a miniature population.

Table 9.1
ALL POSSIBLE SAMPLES OF SIZE TWO FROM A MINIATURE POPULATION

	ALL POSSIBLE SAMPLES	MEAN ($\overline{X}$)	PROBABILITY
(1)	2,2	2.0	$\frac{1}{16}$
(2)	2,3	2.5	$\frac{1}{16}$
(3)	2,4	3.0	$\frac{1}{16}$
(4)	2,5	3.5	$\frac{1}{16}$
(5)	3,2	2.5	$\frac{1}{16}$
(6)	3,3	3.0	$\frac{1}{16}$
(7)	3,4	3.5	$\frac{1}{16}$
(8)	3,5	4.0	$\frac{1}{16}$
(9)	4,2	3.0	$\frac{1}{16}$
(10)	4,3	3.5	$\frac{1}{16}$
(11)	4,4	4.0	$\frac{1}{16}$
(12)	4,5	4.5	$\frac{1}{16}$
(13)	5,2	3.5	$\frac{1}{16}$
(14)	5,3	4.0	$\frac{1}{16}$
(15)	5,4	4.5	$\frac{1}{16}$
(16)	5,5	5.0	$\frac{1}{16}$

Table 9.2
SAMPLING DISTRIBUTION OF THE MEAN (SAMPLES OF SIZE TWO FROM A MINIATURE POPULATION)

SAMPLE MEAN ($\overline{X}$)	PROBABILITY
5.0	$\frac{1}{16}$
4.5	$\frac{2}{16}$
4.0	$\frac{3}{16}$
3.5	$\frac{4}{16}$
3.0	$\frac{3}{16}$
2.5	$\frac{2}{16}$
2.0	$\frac{1}{16}$

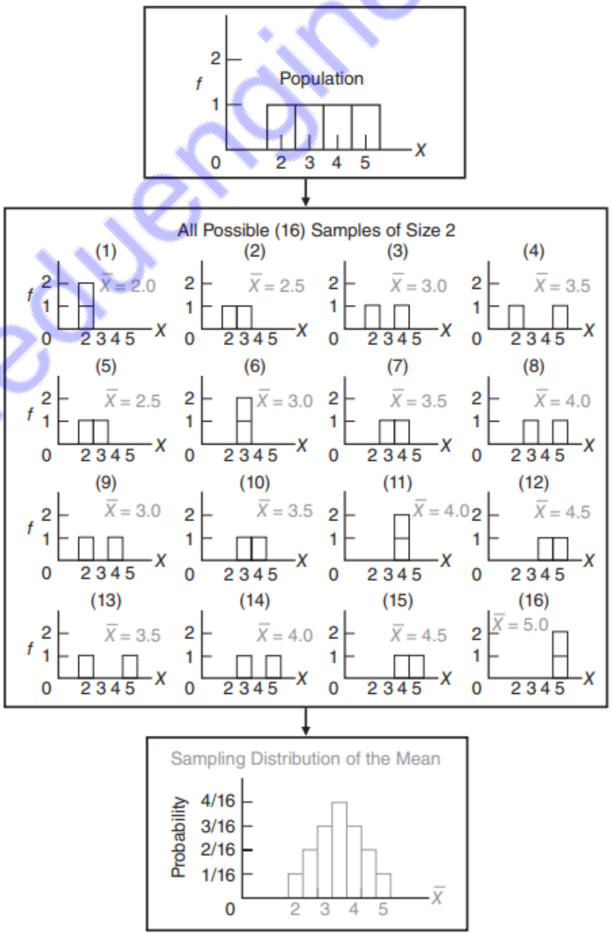


FIGURE 9.2
Emergence of the sampling distribution of the mean from all possible samples.

Table 9.3 SYMBOLS FOR THE MEAN AND STANDARD DEVIATION OF THREE TYPES OF DISTRIBUTIONS		
TYPE OF DISTRIBUTION	**MEAN**	**STANDARD DEVIATION**
Sample	$\overline{X}$	s
Population	μ	σ
Sampling distribution of the mean	$\mu_{\overline{X}}$	$\sigma_{\overline{X}}$ (standard error of the mean)

3.2.3 MEAN OF ALL SAMPLE MEANS ($\mu_{\overline{X}}$)

The distribution of sample means itself has a mean. The mean of the sampling distribution of the mean always equals the mean of the population.

Expressed in symbols,

MEAN OF THE SAMPLING DISTRIBUTION

$$\mu_{\overline{X}} = \mu \tag{9.1}$$

Where $\mu_{\overline{X}}$ represents the mean of the sampling distribution and μ represents the mean of the population.

1. **Interchangeable Means**

The mean of all sample means ($\mu_{\overline{X}}$) always equals the mean of the population (μ), these two terms are interchangeable in inferential statistics. Any claims about thepopulation mean can be transferred directly to the mean of the sampling distribution, and vice versa. If, as claimed, the mean math score for the local population of freshmen equals the national average of 500, then the mean of the sampling distribution also automatically will equal 500. For the same reason, it's permissible to view the one observed sample mean of 533 as a deviation either from the mean of the sampling distribution or from the mean of the population. It should be apparent, therefore, that whether an expression involves μ_{X} or μ, it reflects, at most, a difference in emphasis on either the sampling distribution or the population, respectively, rather than any difference in numerical value.

Explanation

Although important, it's not particularly startling that the mean of all sample means equals the population mean. As can be seen in Figure 9.2, samples are not exact replicas of the population, and most sample means are either larger or smaller than the population mean (equal to 3.5 in Figure 9.2). By taking the mean of all sample means, however, you effectively neutralize chance differences between sample means and retain a value equal to the population mean.

3.2.4 STANDARD ERROR OF THE MEAN ($\sigma_{\overline{X}}$)

The distribution of sample means also has a standard deviation, referred to as the standard error of the mean.

The standard error of the mean equals the standard deviation of the population divided by the square root of the sample size.

1. **STANDARD ERROR OF THE MEAN ($\sigma_{\overline{X}}$)**

Expressed in symbols,

STANDARD ERROR OF THE MEAN

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \tag{9.2}$$

where $\sigma_{\overline{X}}$ represents the standard error of the mean; σ represents the standard deviation of the population; and n represents the sample size.

2. **Special Type of Standard Deviation**

The standard error of the mean serves as a special type of standard deviation that measures variability in the sampling distribution. It supplies us with a standard, much like a yardstick, that describes the amount by which sample means deviate from the mean of the sampling distribution or from the population mean. The error in standard error refers not to computational errors, but to errors in generalizations attributable to the fact that, just by chance, most random samples aren't exact replicas of the population.

The standard error of the mean as a rough measure of the average amount by which sample means deviate from the mean of the sampling distribution or from the population mean.

Insofar as the shape of the distribution sample means approximates a normal curve, as described in the next section, about 68 percent of all sample means deviate less than one standard error from the mean of the sampling distribution, whereas only about 5 percent of all sample means deviate more than two standard errors from the mean of this distribution.

3. **Effect of Sample Size**

A most important implication of Formula 9.2 is that whenever the sample size equals two or more, the variability of the sampling distribution is less than that in the population. A modest demonstration of this effect appears in Figure 9.2, where the means of all possible samples cluster closer to the population mean (equal to 3.5) than do the four original observations in the population. A more dramatic demonstration occurs with larger sample sizes. Earlier in this chapter, for instance, 110 was given as the value of σ, the population standard deviation for SAT scores. Much smaller is the variability in the sampling distribution of mean SAT scores, each based on samples of 100 freshmen. According to Formula 9.2, in the present example,

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{110}{\sqrt{100}} = \frac{110}{10} = 11$$

there is a tenfold reduction in variability, from 110 to 11, when our focus shifts from the population to the sampling distribution.

According to Formula 9.2, any increase in sample size translates into a smaller standard error and, therefore, into a new sampling distribution with less variability. With a larger sample size, sample means cluster more closely about the mean of the sampling distribution and about the mean of the population and, therefore, allow more precise generalizations from samples to populations.

3.2.5 SHAPE OF THE SAMPLING DISTRIBUTION

A product of statistical theory, expressed in its simplest form, **the central limit theorem states that, regardless of the shape of the population, the shape of the sampling distribution of the mean approximates a normal curve if the sample size is sufficiently large.**

According to this theorem, it doesn't matter whether the shape of the parent population is normal, positively skewed, negatively skewed, or some nameless, bizarre shape, as long as the sample size is sufficiently large. What constitutes "sufficiently large" depends on the shape of the parent population. If the shape of the parent population is normal, then any sample size (even a sample size of one) will be sufficiently large. Otherwise, depending on the degree of non-normality in the parent population, a sample size between 25 and 100 is sufficiently large.

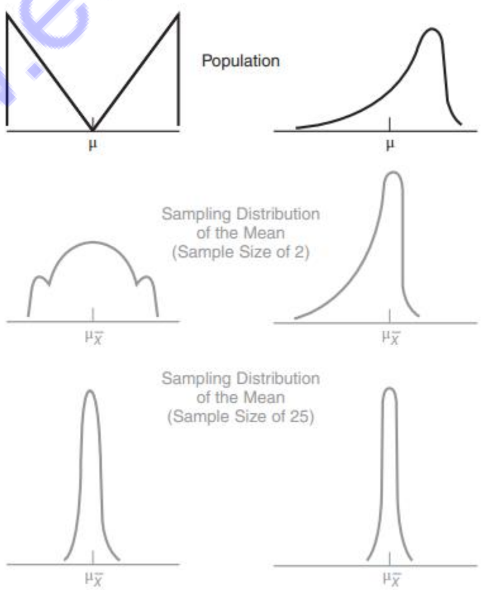


FIGURE 9.3
Effect of the central limit theorem.

1. **Why the Central Limit Theorem Works**

In a normal curve, you will recall, intermediate values are the most prevalent, and extreme values, either larger or smaller, occupy the tapered flanks. Why, when the sample size is large, does the sampling distribution approximate a normal curve, even though the parent population might be non-normal?

2. **Many Sample Means with Intermediate Values**

When the sample size is large, it is most likely that any single sample will contain the full spectrum of small, intermediate, and large scores from the parent population, whatever its shape. The calculation of a mean for this type of sample tends to neutralize or dilute the effects of any extreme scores, and the sample mean emerges with some intermediate value.

Accordingly, intermediate values prevail in the sampling distribution, and they cluster around a peak frequency representing the most common or modal value of the sample mean, as suggested at the bottom of Figure 9.3.

3. **Few Sample Means with Extreme Values**

To account for the rarer sample mean values in the tails of the sampling distribution, focus on those relatively infrequent samples that, just by chance, contain less than the full spectrum of scores from the parent population. Sometimes, because of the relatively large number of extreme scores in a particular direction, the calculation of a mean only slightly dilutes their effect, and the sample mean emerges with some more extreme value. The likelihood of obtaining extreme sample mean values declines with the extremity of the value, producing the smoothly tapered, slender tails that characterize a normal curve.

3.3 HYPOTHESIS TESTING

3.3.1 TESTING A HYPOTHESIS ABOUT SAT SCORES

Test the hypothesis that, with respect to the national average, nothing special is happening in the local population. Insofar as an investigator usually suspects just the opposite namely, that something special is happening in the local population he or she hopes to reject the hypothesis that nothing special is happening, henceforth referred to as the null hypothesis and defined more formally in a later section.

1. **Hypothesized Sampling Distribution**

If the null hypothesis is true, then the distribution of sample means that is, the sampling distribution of the mean for all possible random samples, each of size 100, from the local population of freshmen will be centered about the national average of 500. (Remember, the mean of the sampling distribution always equals the population mean).

In Figure 10.1, this sampling distribution is referred to as the hypothesized sampling distribution, since its mean equals 500, the hypothesized mean reading score for the local population of freshmen.

Anticipating the key role of the hypothesized sampling distribution in our hypothesis test, let's focus on two more properties of this distribution:

i. In Figure 10.1, vertical lines appear, at intervals of size 11, on either side of the hypothesized population mean of 500. These intervals reflect the size of the standard error of the mean, $\overline{X}$. To verify this fact, originally demonstrated in Chapter 9, substitute 110 for the population standard deviation, σ, and 100 for the sample size, n, in Formula 9.2 to obtain

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{110}{\sqrt{100}} = \frac{110}{10} = 11$$

ii. Notice that the shape of the hypothesized sampling distribution in Figure 10.1 approximates a normal curve, since the sample size of 100 is large enough to satisfy the requirements of the central limit theorem. Eventually, with the aid of normal curve tables, we will be able to construct boundaries for common and rare outcomes under the null hypothesis.

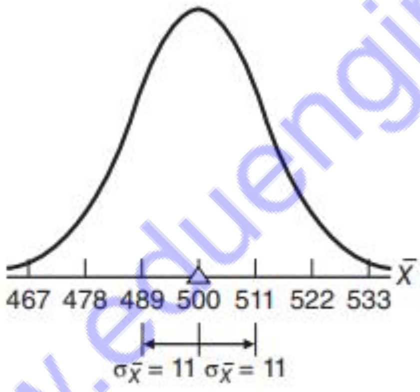


FIGURE 10.1
Hypothesized sampling distribution of the mean centered about a hypothesized population mean of 500.

The null hypothesis that the population mean for the freshman class equals 500 is tentatively assumed to be true. It is tested by determining whether the one observed sample mean qualifies as a common outcome or a rare outcome in the hypothesized sampling distribution of Figure 10.1.

2. Common Outcomes

An observed sample mean qualifies as a common outcome if the difference between its value and that of the hypothesized population mean is small enough to be viewed as a probable outcome under the null hypothesis.

That is, a sample mean qualifies as a common outcome if it doesn't deviate too far from the hypothesized population mean but appears to emerge from the dense concentration of possible sample means in the middle of the sampling distribution. A common outcome signifies a lack of evidence that, with respect to the null hypothesis, something special is happening in the underlying population. Because now there is no compelling reason for rejecting the null hypothesis, it is retained.

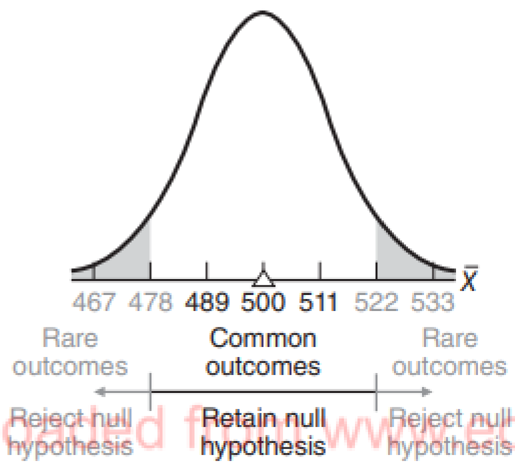
3. Rare Outcomes

An observed sample mean qualifies as a rare outcome if the difference between its value and the hypothesized population mean is too large to be reasonably viewed as a probable outcome under the null hypothesis.

That is, a sample mean qualifies as a rare outcome if it deviates too far from the hypothesized mean and appears to emerge from the sparse concentration of possible sample means in either tail of the sampling distribution. A rare outcome signifies that, with respect to the null hypothesis, something special probably is happening in the underlying population. Because now there are grounds for suspecting the null hypothesis, it is rejected.

4. Boundaries for Common and Rare Outcomes

Superimposed on the hypothesized sampling distribution in Figure 10.2 is one possible set of boundaries for common and rare outcomes, expressed in values of X. If the one observed sample mean is located between 478 and 522, it will qualify as a common outcome (readily attributed to variability) under the null hypothesis, and the null hypothesis will be retained. If, however, the one observed sample mean is greater than 522 or less than 478, it will qualify as a rare outcome (not readily attributed to variability) under the null hypothesis, and the null hypothesis will be rejected. Because the observed sample mean of 533 does exceed 522, the null hypothesis is rejected. On the basis of the present test, it is unlikely that the sample of 100 freshmen, with a mean math score of 533, originates from a population whose mean equals the national average of 500, and, therefore, the investigator can conclude that the mean math score for the local population of freshmen probably differs from (exceeds) the national average.



3.3.2 z TEST FOR A POPULATION MEAN

For the hypothesis test with SAT math scores, it is customary to base the test not on the hypothesized sampling distribution of X shown in Figure 10.2, but on its standardized counterpart, the hypothesized sampling distribution of z shown in Figure 10.3. Now z represents a variation on the familiar standard score, and it displays all of the properties of standard scores.

Furthermore, like the sampling distribution of $\overline{X}$, the **sampling distribution of z** represents the distribution of z values that would be obtained if a value of z were calculated for each sample mean for all possible random samples of a given size from some population.

The conversion from X to z yields a distribution that approximates the standard normal curve in Table A of Appendix C, since, as indicated in Figure 10.3, the original hypothesized population mean (500) emerges as a z score of 0 and the original standard error of the mean (11) emerges as a z score of 1. The shift from $\overline{X}$ to z eliminates the original units of measurement and standardizes the hypothesis test across all situations without, however, affecting the test results.

1. **Converting a Raw Score to z**

To convert a raw score into a standard score, express the raw score as a distance from its mean (by subtracting the mean from the raw score), and then split this distance into standard deviation units (by dividing with the standard deviation). Expressing this definition as a word formula, we have in which, of course, the standard score indicates the deviation of the raw score in standard deviation units, above or below the mean.

$$\text{Standard score} = \frac{\text{raw score} - \text{mean}}{\text{standard deviation}}$$

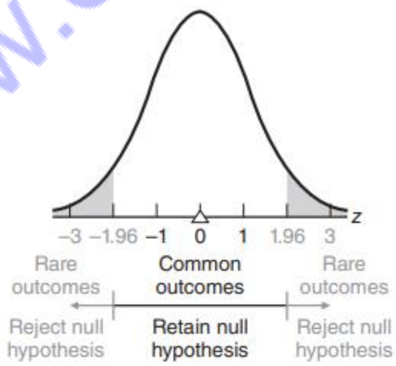


FIGURE 10.3
Common and rare outcomes (values of z).

2. Converting a Sample Mean to z

The z for the present situation emerges as a slight variation of this word formula: Replace the raw score with the one observed sample mean X; replace the mean with the mean of the sampling distribution, that is, the hypothesized population mean μ_{hyp};and replace the standard deviation with the standard error of the mean $\sigma_{\bar{x}}$. Now

z RATIO FOR A SINGLE POPULATION MEAN

$$z = \frac{\bar{X} - \mu_{\text{hyp}}}{\sigma_{\bar{x}}} \qquad (10.1)$$

where z indicates the deviation of the observed sample mean in standard error units, above or below the hypothesized population mean.

To test the hypothesis for SAT scores, we must determine the value of z from Formula 10.1. Given a sample mean of 533, a hypothesized population mean of 500, and a standard error of 11, we find

$$z = \frac{533 - 500}{11} = \frac{33}{11} = 3$$

The observed z of 3 exceeds the value of 1.96 specified in the hypothesized sampling distribution in Figure 10.3. Thus, the observed z qualifies as a rare outcome under the null hypothesis, and the null hypothesis is rejected. The results of this test with z are the same as those for the original hypothesis test with $\bar{x}$.

Progress Check *10.1 Calculate the value of the z test for each of the following situations:

(a) $\bar{X} = 566;\ \sigma = 30;\ n = 36;\ \mu_{\text{hyp}} = 560$

(b) $\bar{X} = 24;\ \sigma = 4;\ n = 64;\ \mu_{\text{hyp}} = 25$

(c) $\bar{X} = 82;\ \sigma = 14;\ n = 49;\ \mu_{\text{hyp}} = 75$

(d) $\bar{X} = 136;\ \sigma = 15;\ n = 25;\ \mu_{\text{hyp}} = 146$

3.3.3 STEP-BY-STEP PROCEDURE

The more important features of hypothesis testing, let's take a detailed look at the test for SAT scores. The test procedure lends itself to a step-by-step description, beginning with a brief statement of the problem that inspired the test and ending with an interpretation of the test results. The following box summarizes the step-by-step procedure for the current hypothesis test.

3.3.4 STATEMENT OF THE RESEARCH PROBLEM

The formulation of a research problem often represents the most crucial and exciting phase of an investigation. Indeed, the mark of a skillful investigator is to focus on an important research problem that can be answered. Do children from broken families score lower on tests of personal adjustment? Do aggressive TV cartoons incite more disruptive behavior in preschool children? Does profit sharing increase the productivity of employees? Because of our emphasis on hypothesis testing, research problems appear in this book as finished products, usually in the first one or two sentences of a new example.

HYPOTHESIS TEST SUMMARY: *z* TEST FOR A POPULATION MEAN (SAT SCORES)

Research Problem

Does the mean SAT math score for all local freshmen differ from the national average of 500?

Statistical Hypotheses

$$H_0 : \mu = 500$$
$$H_1 : \mu \neq 500$$

Decision Rule

Reject H_0 at the .05 level of significance if $z \geq 1.96$ or if $z \leq -1.96$.

Calculations

Given

$$\overline{X} = 533;\ \mu_{\text{hyp}} = 500;\ \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{110}{\sqrt{100}} = 11$$

$$z = \frac{533 - 500}{11} = 3$$

Decision

Reject H_0 at the .05 level of significance because $z = 3$ exceeds 1.96.

Interpretation

The mean SAT math score for all local freshmen does not equal—it exceeds—the national average of 500.

3.3.5 NULL HYPOTHESIS (H_0)

Once the problem has been described, it must be translated into a statistical hypoth- esis regarding some population characteristic. Abbreviated as H_0 , the null hypothesis becomes the focal point for the entire test procedure (even though we usually hope to reject it). In the test with SAT scores, the null hypothesis asserts that, with respect to the national average of 500, nothing special is happening to the mean score for the local population of freshmen. An equivalent statement, in symbols, reads:

$$H_0 : \mu = 500$$

where H_0 represents the null hypothesis and μ is the population mean for the local freshman class.

Generally speaking, the **null hypothesis (H_0)** is a statistical hypothesis that usually asserts that nothing special is happening with respect to some characteristic of the underlying population. Because the hypothesis testing procedure requires that the hypothesized sampling distribution of the mean be centered about a single number (500), the null hypothesis equals a single number (H : μ=500). Furthermore, the null hypothesis always makes a precise statement about a characteristic of the population, never about a sample. Remember, the purpose of a hypothesis test is to determine whether a particular outcome, such as an observed sample mean, could have reason- ably originated from a population with the hypothesized characteristic.

Finding the Single Number for H_0

The single number actually used in H_0 varies from problem to problem. Even for a given problem, this number could originate from any of several sources. For instance, it could be based on available information about some relevant population other than the target population, as in the present example in which 500 reflects the mean SAT math scores for all college-bound students during a recent year. It also could be based on some existing standard or theory for example, that the mean math score for the current population of local freshmen should equal 540 because that happens to be the mean score achieved by all local freshmen during recent years.

If, as sometimes happens, it's impossible to identify a meaningful null hypothesis, don't try to salvage the situation with arbitrary numbers. Instead, use another entirely different technique, known as estimation, which is described in Chapter 12.

3.3.6 ALTERNATIVE HYPOTHESIS (H_1)

In the present example, the alternative hypothesis asserts that, with respect to the national average of 500, something special is happening to the mean math score for the local population of freshmen (because the mean for the local population doesn't equal the national average of 500). An equivalent statement, in symbols, reads:

$$H_1 : \mu \neq 500$$

represents the alternative hypothesis, μ is the population mean for the local freshman class, and signifies, "is not equal to." The **alternative hypothesis (H_1)** asserts the opposite of the null hypothesis. A decision to retain the null hypothesis implies a lack of support for the alternative hypothesis, and a decision to reject the null hypothesis implies support for the alternative hypothesis.

3.3.7 DECISION RULE

A **decision rule** specifies precisely when H0 should be rejected (because the observed z qualifies as a rare outcome). There are many possible decision rules, as will be seen in Section 11.3. A very common one, already introduced in Figure 10.3, specifies that H0 should be rejected if the observed z equals or is more positive than 1.96 or if the observed z equals or is more negative than –1.96. Conversely, H0 should be retained if the observed z falls between ± 1.96.

1. **Critical z Scores**

Figure 10.4 indicates that z scores of ± 1.96 define the boundaries for the middle .95 of the total area (1.00) under the hypothesized sampling distribution for z. Derived from the normal curve table, as you can verify by checking Table A in Appendix C, these two z scores separate common from rare outcomes and hence dictate whether H0 should be retained or rejected. Because of their vital role in the decision about H0 , these scores are referred to as critical z scores.

2. **Level of Significance (α)**

Figure 10.4 also indicates the proportion (.025 .025 .05) of the total area that is identified with rare outcomes. Often referred to as the level of significance of the statistical test, this proportion is symbolized by the Greek letter α (alpha). In the present example, the level of significance, α, equals .05.

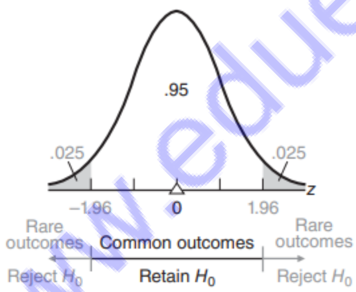


FIGURE 10.4
Proportions of area associated with common and rare outcomes (α = 05).

The level of significance (α) indicates the degree of rarity required of an observed outcome in order to reject the null hypothesis (H_0). For instance, the .05 level of significance indicates that H0 should be rejected if the observed z could have occurred just by chance with a probability of only .05 (one chance out of twenty) or less.

3.3.8 CALCULATIONS

Use information from the sample to calculate a value for z. As has been noted previously, use Formula 10.1 to convert the observed sample mean of 533 into a z of 3.

3.3.9 DECISION

Either retain or reject H_0 , depending on the location, of the observed z value relative to the critical z values specified in the decision rule. According to the present rule, H_0 should be rejected at the .05 level of significance because the observed z of 3 exceeds the critical z of 1.96 and, therefore, qualifies as a rare outcome, that is, an unlikely outcome from a population centered about the null hypothesis.

1. Retain or Reject H_0?

If you are ever confused about whether to retain or reject H_0, recall the logic behind the hypothesis test. You want to reject H only if the observed value of z qualifies as a rare outcome because it deviates too far into the tails of the sampling distribution. Therefore, you want to reject H_0.

Only if the observed value of z equals or is more positive than the upper critical z (1.96) or if it equals or is more negative than the lower critical z (–1.96). Before deciding, you might find it helpful to sketch the hypothesized sampling distribution, along with its critical z values and shaded rejection regions, and then use some mark, such as an arrow (↑), to designate the location of the observed value of z (3) along the z scale. If this mark is located in the shaded rejection region or farther out than this region, as in Figure 10.4—then H_0 should be rejected.

3.3.10 INTERPRETATION

Finally, interpret the decision in terms of the original research problem. In the present example, it can be concluded that, since the null hypothesis was rejected, the mean SAT math score for the local freshman class probably differs from the national average of 500. Although not a strict consequence of the present test, a more specific conclusion is possible. Since the sample mean of 533 (or its equivalent z of 3) falls in the upper rejection region of the hypothesized sampling distribution, it can be concluded that the population mean SAT math score for all local freshmen probably exceeds the national average of 500. By the same token, if the observed sample mean or its equivalent z had fallen in the lower rejection region of the hypothesized sampling distribution, it could have been concluded that the population mean for all local freshmen probably is below the national average. If the observed sample mean or its equivalent z had fallen in the retention region of the hypothesized sampling distribution, it would have been concluded (somewhat weakly, as discussed in Section 11.2) that there is no evidence that the population mean for all local freshmen differs from the national average of 500.

3.4.1 WHY HYPOTHESIS TESTS?

There is a crucial link between hypothesis tests and the need of investigators, whether pollsters or researchers, to generalize beyond existing data. If the 100 freshmen in the SAT example of the previous chapter had been not a sample but a census of the entire freshman class, there wouldn't have been any need to generalize beyond existing data, and it would have been inappropriate to conduct a hypothesis test. Now, the observed difference between the newly observed population mean of 533 and the national average of 500, by itself, would have been sufficient grounds for concluding that the mean SAT math score for all local freshmen exceeds the national average. Indeed, any observed difference in favor of the local freshmen, regardless of the size of the difference, would have supported this conclusion.

If we must generalize beyond the 100 freshmen to a larger local population, as was actually the case, the observed difference between 533 and 500 cannot be interpreted at face value. The basic problem is that the sample mean for a second random sample of 100 freshmen probably would differ, just by chance, from the sample mean of 533 for the first sample. Accordingly, the variability among sample means must be considered when we attempt to decide whether the observed difference between 533 and 500 is real or merely transitory.

1. Importance of the Standard Error

To evaluate the effect of chance, we use the concept of a sampling distribution, that is, the concept of the sample means for all possible random outcomes. A key element in this concept is the standard error of the mean, a measure of the average amount by which sample means differ, just by chance, from the population mean. Dividing the observed difference (533−500) by the standard error (11) to obtain a value of z (3) locates the original observed difference along a z scale of either common outcomes (reasonably attributable to chance) or rare outcomes (not reasonably attributable to chance). If, when expressed as z, the ratio of the observed difference to the standard error is small enough to be reasonably attributed to chance, we retain H0. Otherwise, if the ratio of the observed difference to the standard error is too large to be reasonably attributed to chance, as in the SAT example, we reject H0.

Before generalizing beyond the existing data, we must always measure the effect of chance; that is, we must obtain a value for the standard error. To appreciate the vital role of the standard error in the SAT example, increase its value from 11 to 33 and note that even though the observed difference remains the same (533−500), we would retain, not reject, H0 because now z would equal 1 (rather than 3) and be less than the critical z of 1.96.

2. Possibility of Incorrect Decisions

Having made a decision about the null hypothesis, we never know absolutely whether that decision is correct or incorrect, unless, of course, we survey the entire population. Even if H0 is true (and, therefore, the hypothesized distribution of z about H0 also is true), there is a slight possibility that, just by chance, the one observed z actually originates from one of the shaded rejection regions of the hypothesized distribution of z, thus causing the true H0 to be rejected. This type of incorrect decision—rejecting a true H0—is referred to as a type I error or a false alarm.

On first impulse, it might seem desirable to abolish the shaded rejection regions in the hypothesized sampling distribution to ensure that a true H0 never is rejected. A most unfortunate consequence of this strategy, however, is that no H0, not even a radically false H0, ever would be rejected. This second type of incorrect decision—retaining a false H0—is referred to as a type II error or a miss. Both type I and type II errors are described in more detail later in this chapter.

11.2 STRONG OR WEAK DECISIONS

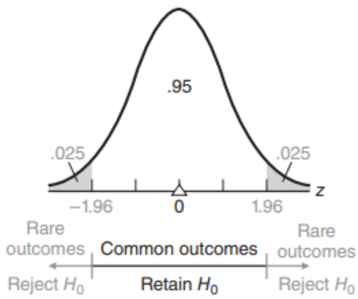


FIGURE 11.1
Proportions of area associated with common and rare outcomes (α = .05).

3. Minimizing Incorrect Decisions

Traditional hypothesis-testing procedures, such as the one illustrated in Figure 11.1, tend to minimize both types of incorrect decisions. If H_0 is true, there is a high probability that the observed z will qualify as a common outcome under the hypothesized sampling distribution and that the true H0 will be retained. (In Figure 11.1, this probability equals the proportion of white area (.95) in the hypothesized sampling distribution.)

On the other hand, if H_0 is seriously false, because the hypothesized population mean differs considerably from the true population mean, there is also a high probability that the observed z will qualify as a rare outcome under the hypothesized distribution and that the false H_0 will be rejected. (In Figure 11.1, this probability can't be determined since; in this case, the hypothesized sampling distribution does not actually reflect the true sampling distribution.

Even though we never really know whether a particular decision is correct or incorrect, it is reassuring that in the long run, most decisions will be correct— assuming the null hypotheses are either true or seriously false

3.4.2 STRONG OR WEAK DECISIONS

1. Retaining H0 Is a Weak Decision

There are subtle but important differences in the interpretation of decisions to retain H0 and to reject H_0. H_0 is retained whenever the observed z qualifies as a common outcome on the assumption that H_0 is true. Therefore, H_0 could be true. However, the same observed result also would qualify as a common outcome when the original value in H_0 (500) is replaced with a slightly different value. Thus, the retention of H_0 must be viewed as a relatively weak decision. Because of this weakness, many statisticians prefer to describe this decision as simply a failure to reject H_0 rather than as the retention of H_0. In any event, the retention of H_0 can't be interpreted as proving H_0 to be true. If H_0 had been retained in the present example, it would have been appropriate to conclude not that the mean SAT math score for all local fresh⍰men equals the national average, but that the mean SAT math score could equal the national average, as well as many other possible values in the general vicinity of the national average.

2. Rejecting H0 Is a Strong Decision

On the other hand, H0 is rejected whenever the observed z qualifies as a rare outcome one that could have occurred just by chance with a probability of .05 or less on the assumption that H_0 is true. This suspiciously rare outcome implies that H_0 is probably false (and conversely, that H_1 is probably true). Therefore, the rejection of H_0 can be viewed as a strong decision. When H_0 was rejected in the present example, it was appropriate to report a definitive conclusion that the mean SAT math score for all local freshmen probably exceeds the national average.

To summarize,

The decision to retain H_0 implies not that H_0 is probably true, but only that H0 could be true, whereas the decision to reject H_0 implies that H_0 is probably false (and that H_1 is probably true).

Since most investigators hope to reject H_0 in favor of H_1, the relative weakness of the decision to retain H_0 usually does not pose a serious problem.

3. Why the Research Hypothesis Isn't Tested Directly

Even though H0, the null hypothesis, is the focus of a statistical test, it is usually of secondary concern to the investigator. Nevertheless, there are several reasons why, although of primary concern, the research hypothesis is identified with H1 and tested indirectly.

4. Lacks Necessary Precision

The research hypothesis, but not the null hypothesis, lacks the necessary precision to be tested directly.

To be tested, a hypothesis must specify a single number about which the hypothesized sampling distribution can be constructed. Because it specifies a single number, the null hypothesis, rather than the research hypothesis, is tested directly. In the SAT example, the null hypothesis specifies that a precise value (the national average of 500) describes the mean for the current population of interest (all local freshmen). Typically, the research hypothesis lacks the required precision. It merely specifies that some inequality exists between the hypothesized value (500) and the mean for the current population of interest (all local freshmen).

5. Supported by a Strong Decision to Reject

Logical considerations also argue for the indirect testing of the research hypothesis and the direct testing of the null hypothesis.

Because the research hypothesis is identified with the alternative hypothesis, the decision to reject the null hypothesis, should it be made, will provide strong support for the research hypothesis, while the decision to retain the null hypothesis, should it be made, will provide, at most, weak support for the null hypothesis.

As mentioned, the decision to reject the null hypothesis is stronger than the decision to retain it. Logically, a statement such as "All cows have four legs" can never be proven in spite of a steady stream of positive instances. It only takes one negative instance—one cow with three legs—to disprove the statement. By the same token, one positive instance (common outcome) doesn't prove the null hypothesis, but one negative instance (rare outcome) disproves the null hypothesis. (Strictly speaking, however, since a rare outcome implies that the null hypothesis is probably but not definitely false, remember that there always is a very small possibility that the rare outcome reflects a true null hypothesis).

Logically, therefore, it makes sense to identify the research hypothesis with the alternative hypothesis. If, as hoped, the data favor the research hypothesis, the test will generate strong support for your hunch: It's probably true. If the data do not favor the research hypothesis, the hypothesis test will generate, at most, weak support for the null hypothesis: It could be true. Weak support for the null hypothesis is of little consequence, as this hypothesis that nothing special is happening in the population usually serves only as a convenient testing device.

3.4.3 ONE-TAILED AND TWO-TAILED TESTS

Two-Tailed Test Generally, the alternative hypothesis, H1, is the complement of the null hypothesis, H0. Under typical conditions, the form of H1 resembles that shown for the SAT example, namely,

$$H_1 : \mu \neq 500$$

This alternative hypothesis says that the null hypothesis should be rejected if the mean reading score for the population of local freshmen differs in either direction from the national average of 500. An observed z will qualify as a rare outcome if it deviates too far either below or above the national average. Panel A of Figure 11.2 shows rejection regions that are associated with both tails of the hypothesized sampling distribution. The corresponding decision rule, with its pair of critical z scores of ±1.96, is referred to as a two-tailed or nondirectional test.

1. One-Tailed Test (Lower Tail Critical)

Now let's assume that the research hypothesis for the investigation of SAT math scores was based on complaints from instructors about the poor preparation of local freshmen. Assume also that if the investigation supports these complaints, a remedial program will be instituted. Under these circumstances, the investigator might prefer a hypothesis test that is specially designed to detect only whether the population mean math score for all local freshmen is less than the national average.

This alternative hypothesis reads:

$$H_1 : \mu \leq 500$$

It reflects a concern that the null hypothesis should be rejected only if the population mean math score for all local freshmen is less than the national average of 500. Accord□ingly, an observed z triggers the decision to reject H0 only if z deviates too far below the national average. Panel B of Figure 11.2 illustrates a rejection region that is associated with only the lower tail of the hypothesized sampling distribution. The corresponding decision rule, with its critical z of –1.65, is referred to as a one-tailed or directional test with the lower tail critical. Use Table A in Appendix C to verify that if the critical z equals –1.65; then .05 of the total area under the distribution of z has been allocated to the lower rejection region. Notice that the level of significance, α, equals .05 for this one-tailed test and also for the original two-tailed test.

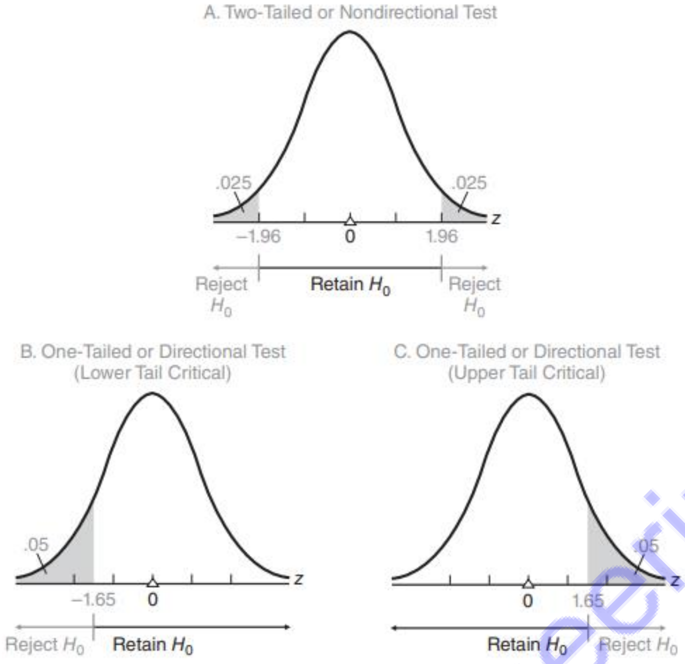


FIGURE 11.2
Three different types of tests ($\alpha = .05$).

2. Extra Sensitivity of One-Tailed Tests

This new one-tailed test is extra sensitive to any drop in the population mean for the local freshmen below the national average. If H0 is false because a drop has occurred, then the observed z will be more likely to deviate below the national average. As can be seen in panels A and B of Figure 11.2, an observed deviation in the direction of concern below the national average is more likely to penetrate the broader rejection region for the one-tailed test than that for the two-tailed test. Therefore, the decision to reject a false H0 (in favor of the research hypothesis) is more likely to occur in the one-tailed test than in the two-tailed test.

3. One-Tailed Test (Upper Tail Critical)

Panel C of Figure 11.2 illustrates a one-tailed or directional test with the upper tail critical. This one-tailed test is the mirror image of the previous test. Now the alternative hypothesis reads:

$$H_1 : \mu > 500$$

and its critical z equals 1.65. This test is specially designed to detect only whether the population mean math score for all local freshmen exceeds the national average. For example, the research hypothesis for this investigation might have been inspired by the possibility of eliminating an existing remedial math program if it can be demonstrated that, on the average, the SAT math scores of all local freshmen exceed the national average.

4. One or Two Tails?

Before a hypothesis test, if there is a concern that the true population mean differs from the hypothesized population mean only in a particular direction, use the appropriate one-tailed or directional test for extra sensitivity. Otherwise, use the more customary two-tailed or nondirectional test. Having committed yourself to a one-tailed test with its single rejection region, you must retain H_0, regardless of how far the observed z deviates from the hypothesized population mean in the direction of "no concern." For instance, if a one-tailed test with the lower tail critical had been used with the data for 100 freshmen from the SAT example, H_0 would have been retained because, even though the observed z equals an impressive value of 3, it deviates in the direction of no concern in this case, above the national average. Clearly, a one-tailed test should be adopted only when there is absolutely no concern about deviations, even very large deviations, in one direction. If there is the slightest concern about these deviations, use a two-tailed test. The selection of a one- or two-tailed test should be made before the data are collected. Never "peek" at the value of the observed z to determine whether to locate the rejection region for a one-tailed test in the upper or the lower tail of the distribution of z. To qualify as a one-tailed test, the location of the rejection region must reflect the investigator's concern only about deviations in a particular direction before any inspection of the data. Indeed, the investigator should be able to muster a compelling reason, based on an understanding of the research hypothesis, to support the direction of the one-tailed test.

5. New Null Hypothesis for One-Tailed Tests

When tests are one-tailed, a complete statement of the null hypothesis also should include all possible values of the population mean in the direction of no concern. For example, given a one-tailed test with the lower tail critical, such as H1: $\mu < 500$, the complete null hypothesis should be stated as H0: $\mu \geq 500$ instead of H0: $\mu = 500$. By the same token, given a one-tailed test with the upper tail critical, such as H1: $\mu > 500$, the complete null hypothesis should be stated as H0: $\mu \leq 500$. If you think about it, the complete H0 describes all of the population means that could be true if a one-tailed test results in the retention of the null hypothesis. For instance, if a one-tailed test with the lower tail critical results in the retention of H0: $\mu \geq 500$, the complete H0 accurately reflects the fact that not only $\mu = 500$ could be true, but also that any other value of the population mean in the direction of no concern, that is, $\mu > 500$, could be true. (Remember, when the test is one-tailed, even a very deviant result in the direction of no concern possibly reflecting a mean much larger than 500 still would trigger the decision to retain H0.) Henceforth, whenever a one-tailed test is employed, write H0 to include values of the population mean in the direction of no concern even though the single number in the complete H0 identified by the equality sign is the one value about which the hypothesized sampling distribution is centered and, therefore, the one value actually used in the hypothesis test.

3.5 ESTIMATION

3.5.1 POINT ESTIMATE FOR μ

A point estimate for μ uses a single value to represent the unknown population mean. This is the most straightforward type of estimate. If a random sample of 100 local freshmen reveals a sample mean SAT score of 533, then 533 will be the point estimate of the unknown population mean for all local freshmen. The best single point estimate for the unknown population mean is simply the observed value of the sample mean.

A Basic Deficiency

Although straightforward, simple, and precise, point estimates suffer from a basic deficiency. They tend to be inaccurate. Because of sampling variability, it's unlikely that a single sample mean, such as 533, will coincide with the population mean. Since point estimates convey no information about the degree of inaccuracy due to sampling variability, statisticians supplement point estimates with another, more realistic type of estimate, known as interval estimates or confidence intervals.

3.5.2 CONFIDENCE INTERVAL (CI) FOR μ

A confidence interval for μ uses a range of values that, with a known degree of certainty, includes the unknown population mean. For instance, the SAT investigator might use a confidence interval to claim, with 95 percent confidence, that the interval between 511.44 and 554.56 includes the population mean math score for all local freshmen. To be 95 percent confident signifies that if many of these intervals were constructed for a long series of samples, approximately 95 percent would include the population mean for all local freshmen. In the long run, 95 percent of these confidence intervals are true because they include the unknown population mean. The remaining 5 percent are false because they fail to include the unknown population mean.

1. Why Confidence Intervals Work

To understand confidence intervals, you must view them in the context of three important properties of the sampling distribution of the mean.

For the sampling distribution from which the sample mean of 533 originates, as shown in Figure 12.1, the three important properties are as follows:

- ■ The mean of the sampling distribution equals the unknown population mean for all local freshmen, whatever its value, because the mean of this sampling distribution always equals the population mean.
- ■ The standard error of the sampling distribution equals the value (11) obtained from dividing the population standard deviation (110) by the square root of the sample size $(\sqrt{100})$
- ■ The shape of the sampling distribution approximates a normal distribution because the sample size of 100 satisfies the requirements of the central limit theorem.

2. A Series of Confidence Intervals

In practice, only one sample mean is actually taken from this sampling distribution and used to construct a single 95 percent confidence interval. However, imagine taking not just one but a series of randomly selected sample means from this sampling distribution. Because of sampling variability, these sample means tend to differ among themselves. For each sample mean, construct a 95 percent confidence interval by adding 1.96 standard errors to the sample mean and subtracting 1.96 standard errors from the sample mean; that is, use the expression.

$$\overline{X} \pm 1.96\sigma_{\overline{X}},$$

to obtain a 95 percent confidence interval for each sample mean.

3. True Confidence Intervals

According to statistical theory, do 95 percent of these confidence intervals include the unknown population mean? As indicated in Figure 12.2, because the sampling distribution is normal, 95 percent of all sample means are within 1.96 standard errors of the unknown population mean, that is, 95 percent of all sample means deviate less than 1.96 standard errors from the unknown population mean. Therefore, and this is the key point, when sample means are expanded into confidence intervals by adding and subtracting 1.96 standard errors 95 percent of all possible confidence intervals are true because they include the unknown population mean. To illustrate

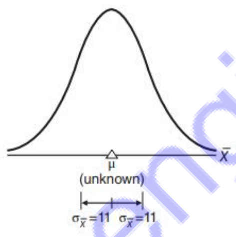


FIGURE 12.1
Sampling distribution of the mean (SAT scores).

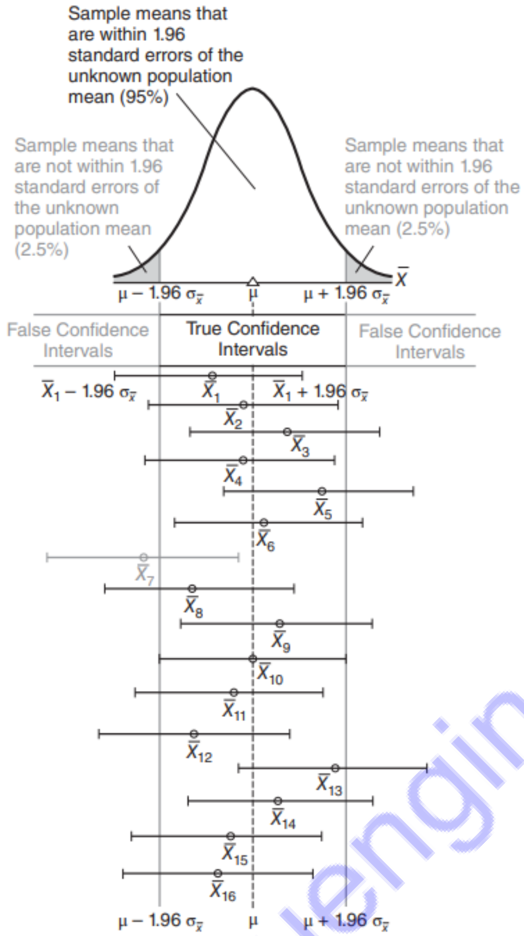


FIGURE 12.2

A series of 95 percent confidence intervals (emerging from a sampling distribution).

this point, 15 of the 16 sample means shown in Figure 12.2 are within 1.96 standard errors of the unknown population mean. The corresponding 15 confidence intervals have ranges that span the broken line for the population mean, thereby qualifying as true intervals because they include the value of the unknown population mean.

4. False Confidence Intervals

Five percent of all confidence intervals fail to include the unknown population mean. As indicated in Figure 12.2, 5 percent of all sample means (2.5 percent in each tail) deviate more than 1.96 standard errors from the unknown population mean. Therefore, when sample means are expanded into confidence intervals—by adding and subtracting 1.96 standard errors—5 percent of all possible confidence intervals are false because they fail to include the unknown population mean. To illustrate this point, only 1 of the 16 sample means shown in Figure 12.2 is not within 1.96 standard errors of the unknown population mean.

The resulting confidence interval, shown as shaded, has a range that does not span the broken line for the population mean, thereby being designated as a false interval because it fails to include the value of the unknown population mean.

5. **Confidence Interval for μ Based on z**

To determine the previously reported confidence interval of 511.44 to 554.56 for the unknown mean math score of all local freshmen, use the following general expression:

CONFIDENCE INTERVAL FOR μ (BASED ON z)	
$\bar{X} \pm (z_{conf})(\sigma_{\bar{X}})$	(12.1)

where $\bar{X}$ represents the sample mean; z_{conf} represents a number from the standard normal table that satisfies the confidence specifications for the confidence interval; and σx represents the standard error of the mean. Given that $\bar{X}$, the sample mean SAT math score, equals 533, that z_{conf} equals 1.96 (from the standard normal tables, where z scores of ±1.96 define the middle 95 percent of the area under the normal curve), and that the standard error, σ_x, equals 11, Formula 12.1 becomes where $\bar{X}$ represents the sample mean; z_{conf} represents a number from the standard normal table that satisfies the confidence specifications for the confidence interval; and σx represents the standard error of the mean. Given that $\bar{X}$, the sample mean SAT math score, equals 533, that z_{conf} equals 1.96 (from the standard normal tables, where z scores of ±1.96 define the middle 95 percent of the area under the normal curve), and that the standard error, σx, equals 11, Formula 12.1 becomes

$$533 \pm (1.96)(11) = 533 \pm 21.56 = \begin{cases} 554.56 \\ 511.44 \end{cases}$$

where 554.56 and 511.44 represent the upper and lower limits of the confidence inter�val. Now it can be claimed, with 95 percent confidence, that the interval between 511.44 and 554.56 includes the value of the unknown mean math score for all local freshmen.

3.5.3 INTERPRETATION OF A CONFIDENCE INTERVAL

A 95 percent confidence claim reflects a long-term performance rating for an extended series of confidence intervals. If a series of confidence intervals is constructed to estimate the same population mean, as in Figure 12.2, approximately 95 percent of these intervals should include the population mean. In practice, only one confidence interval, not a series of intervals, is constructed, and that one interval is either true or false, because it either includes the population mean or fails to include the population mean. Of course, we never really know whether a particular confidence interval is true or false unless the entire population is surveyed. However, when the level of confidence equals 95 percent or more, we can be reasonably confident that the one observed confidence interval includes the true population mean.

For instance, we can be reasonably confident that the true population mean math score for all local freshmen is neither less than 511.44 nor more than 554.56. That's the same as being reasonably confident that the true population mean for all local freshmen is between 511.44 and 554.56.

3.5.4 LEVEL OF CONFIDENCE

The level of confidence indicates the percent of time that a series of confidence intervals includes the unknown population characteristic, such as the population mean. Any level of confidence may be assigned to a confidence interval merely by substituting an appropriate value for z_{conf} in Formula 12.1. For instance, to construct a 99 percent confidence interval from the data for SAT math scores, first consult Table A in Appendix C to verify that z_{conf} values of ±2.58 define the middle 99 percent of the total area under the normal curve. Then substitute numbers for symbols in Formula 12.1 to obtain

$$533 \pm (2.58)(11) = 533 \pm 28.38 = \begin{cases} 561.38 \\ 504.62 \end{cases}$$

It can be claimed, with 99 percent confidence, that the interval between 504.62 and 561.38 includes the value of the unknown mean math score for all local freshmen. This implies that, in the long run, 99 percent of these confidence intervals will include the unknown population mean.

1. Effect on Width of Interval

Notice that the 99 percent confidence interval of 504.62 to 561.38 is wider and, therefore, less precise than the corresponding 95 percent confidence interval of 511.44 to 554.56. The shift from a 95 percent to a 99 percent level of confidence requires an increase in the value of zconf from 1.96 to 2.58. This increase, in turn, causes a wider, less precise confidence interval. Any shift to a higher level of confidence always produces a wider, less precise confidence interval unless offset by an increase in sample size.

2. Choosing a Level of Confidence

Although many different levels of confidence have been used, 95 percent and 99 percent are the most prevalent. Generally, a larger level of confidence, such as 99 per�cent, should be reserved for situations in which a false interval might have particularly serious consequences, such as the failure of a national opinion pollster to predict the winner of a presidential election.

3.5.5 EFFECT OF SAMPLE SIZE

The larger the sample size, the smaller the standard error and, hence, the more precise (narrower) the confidence interval will be. Indeed, as the sample size grows larger, the standard error will approach zero and the confidence interval will shrink to a point estimate. Given this perspective, the sample size for a confidence interval, unlike that for a hypothesis test, never can be too large.

Selection of Sample Size

The hypothesis tests, sample size can be selected according to specifications established before the investigation. To generate a confidence interval that possesses the desired precision (width). Valid use of these formulas requires that before the investigation, the population standard deviation be either known or estimated.

UNIT III

INFERENTIAL STATISTICS

Populations – samples – random sampling – Sampling distribution- standard error of the mean - Hypothesis testing – z-test – z-test procedure –decision rule – calculations – decisions – interpretations - one-tailed and two-tailed tests – Estimation – point estimate – confidence interval – level of confidence – effect of sample size.

PART - A

1) What is population?

In statistics, population is the entire set of items from which you draw data for a statistical study. It can be a group of individuals, a set of items, etc. It makes up the data pool for a study.

2) What is a sample?

A sample represents the group of interest from the population, which you will use to represent the data. The sample is an unbiased subset of the population that best represents the whole data.

3) When are samples used?

- The population is too large to collect data.
- The data collected is not reliable.
- The population is hypothetical and is unlimited in size. Take the example of a study that documents the results of a new medical procedure. It is unknown how the procedure will affect people across the globe, so a test group is used to find out how people react to it.

4) Difference Between Population and Sample?

Population	Samples
All residents of a country would constitute the Population set	All residents who live above the poverty line would be the Sample
All residents above the poverty line in a	All residents who are millionaires would make

country would be the Population	up the Sample
All employees in an office would be the Population	Out of all the employees, all managers in the office would be the Sample

5) Define Hypothetical Population

A population containing a finite number of individuals, members or units is a class. ... All the 400 students of 10th class of particular school is an example of existent type of population and the population of heads and tails obtained by tossing a coin on infinite number of times is an example of hypothetical population.

6) What Is Random Samplings

Random sampling occurs if, at each stage of sampling, the selection process guarantees that all potential observations in the population have an equal chance of being included in the sample

8) What is Sampling Distribution ?

The sampling distribution of the mean refers to the probability distribution of means for all possible random samples of a given size from some population.

9) What are the types of Sampling Distribution?

- Sampling distribution of mean
- Sampling distribution of proportion
- T-distribution

10) Define Sampling distribution of mean

The most common type of sampling distribution is of the mean. It focuses on calculating the mean of every sample group chosen from the population and plotting the data points. The graph shows a normal distribution where the center is the mean of the sampling distribution, which represents the mean of the entire population.

11) What is mean by Sampling distribution of proportion

This sampling distribution focuses on proportions in a population. Samples are selected and their proportions are calculated. The mean of the sample proportions from each group represent the proportion of the entire population,

12) Define T-distribution

A T-distribution is a sampling distribution that involves a small population or one where not much is known about it. It is used to estimate the mean of the population and other statistics such as confidence intervals, statistical differences and linear regression. The T-distribution uses a t- score to evaluate data that wouldn't be appropriate for a normal distribution.

The formula for t-score is: $\boxed{\mathbf{t = [\ x - \mu\ ] / [\ s / sqrt(\ n\ )\ ]}}$

In the formula, "x" is the sample mean and "μ" is the population mean and signifies standard deviation.

13) Define MEAN OF ALL THE SAMPLE MEAN

The mean of the sampling distribution of the mean always equals the mean of the population.

14) Standard Error Of The Mean

The standard error of the mean equals the standard deviation of the population divided by the square root of the sample size

15) What is the Special Type Of Standard Deviation

You might find it helpful to think of the standard error of the mean as a rough measure of the average amount by which sample means deviate from the mean of the sampling distribution or from the population mean.

16) What Is The Hypothesis Testing

Hypothesis testing is a form of statistical inference that uses data from a sample to draw conclusions about a population parameter or a population probability distribution. First, a tentative assumption is made about the parameter or distribution. This assumption is called the null hypothesis and is denoted by H_0.

17) Hypothesized Sampling Distribution

When you perform a hypothesis test of a single population mean μ using a normal distribution (often called a z-test), you take a simple random sample from the population. ... Then the binomial distribution of a sample (estimated) proportion can be approximated by the normal distribution with μ = p and σ=√pqn σ = p q n .

18) Define Decision Rule

A decision rule specifies precisely when H0 should be rejected (because the observed z qualifies as a rare outcome). There are many possible decision rules, as will be seen in Section 11.3. A very common one, already introduced in Figure 10.3, specifies that H0 should be rejected if the observed z equals or is more positive than 1.96 or if the observed z equals or is more negative than –1.96. Conversely, H0 should be retained if the observed z falls between ± 1.96.

19) Define null hypothesis?

The null hypothesis is a typical statistical theory which suggests that no statistical relationship and significance exists in a set of given single observed variable, between two sets of observed data and measured phenomena.

20) What is Level of Significance

Total area that is identified with rare outcomes. Often referred to as the level of significance of the statistical test, this proportion is symbolized by the Greek letter α (alpha) and discussed more thoroughly in Section 11.4. In the present example, the level of significance, α, equals 05.

21) Define One-Tailed And Two-Tailed Tests

Before a hypothesis test, if there is a concern that the true population mean differs from the hypothesized population mean only in a particular direction, use the appropriate one-tailed or directional test for extra sensitivity. Otherwise, use the more customary two-tailed or non directional test

22) What is Two-Tailed Test with example

Generally, the alternative hypothesis, H1, is the complement of the null hypothesis, H0. Under typical conditions, the form of H1 resembles that shown for the SAT example, namely,

$$H1: \mu \neq 500$$

This alternative hypothesis says that the null hypothesis should be rejected if the mean reading score for the population of local freshmen differs in either direction from the national average of 500. An observed z will qualify as a rare outcome if it deviates too far either below or above the national average. Panel A of Figure 11.2 shows rejection regions that are associated with both tails of the hypothesized sampling distribution. The corresponding decision rule, with its pair of critical z scores of ±1.96, is referred to as a two-tailed or non directional test.

23) what is One-Tailed Test (Lower Tail Critical)

Now let's assume that the research hypothesis for the investigation of SAT math scores was based on complaints from instructors about the poor preparation of local freshmen. Assume also that if the investigation supports these complaints, a remedial program will be instituted. Under these circumstances, the investigator might prefer a hypothesis test that is specially designed to detect only whether the population mean math score for all local freshmen is less than the national average. This alternative hypothesis reads:

$$H1: \mu \leq 500$$

24) What is One-Tailed Test (Upper Tail Critical)

Panel C of Figure 11.2 illustrates a one-tailed or directional test with the upper tail critical. This one-tailed test is the mirror image of the previous test. Now the alternative hypothesis reads:

$$H1: \mu > 500$$

and its critical z equals 1.65. This test is specially designed to detect only whether the population mean math score for all local freshmen exceeds the national average. For example, the research hypothesis for this investigation might have been inspired by the possibility of eliminating an existing remedial math program if it can be demonstrated that, on the average, the SAT math scores of all local freshmen exceed the national average

25) Define Consequences of Reducing Standard Error

As can be seen by comparing Figure 11.5 and Figure 11.6, the reduction of the standard error from 2.5 to 1.5 has two important consequences:

1. It shrinks the upper retention region back toward the hypothesized population mean of 100.
2. It shrinks the entire true sampling distribution toward the true population mean of 103.

26) Define Power curve

A graph showing power as a function of some other variable; specifically a graph of the power output of a vehicle or aircraft against engine speed. 2 figurative Chiefly Business. The current thinking or trend. 3Statistics. A graphical representation of the power function of a statistical test.

27) For a one-tailed or directional test with the lower tail critical

$$H0: \mu \geq \text{SOME NUMBERS}$$

$$H1: \mu < \text{SOME NUMBERS}$$

28) For a one-tailed or directional test with the upper tail critical,

$$H0: \mu \leq \text{SOME NUMBERS}$$

$$H1: \mu > \text{SOME NUMBERS}$$

29) What are four possible outcomes for any hypothesis test:

- If H0 really is true, it is a correct decision to retain the true H0.
- If H0 really is true, it is a type I error to reject the true H0.
- If H0 really is false, it is a type II error to retain the false H0.
- If H0 really is false, it is a correct decision to reject the false H0.

30) Define Point Estimate

A point estimate for μ uses a single value to represent the unknown population mean.

31) What is mean by confidence interval (ci) for μ

A confidence interval for μ uses a range of values that, with a known degree of certainty, includes the unknown population mean.

32) Define Effect Of Sample Size

The larger the sample size, the smaller the standard error and, hence, the more precise (narrower) the confidence interval will be. Indeed, as the sample size grows larger, the standard error will approach zero and the confidence interval will shrink to a point estimate. Given this perspective, the sample size for a confidence interval, unlike that for a hypothesis test, never can be too large.

PART B

1) Explain population and samples. And difference?
2) Describe random sampling?
3) Explain sampling distribution and types?
4) Describe null hypothesis test in detail?
5) Explain in detail hypothesis testing and examples?
6) Does the mean of SAT math score for all local freshman differ for all local average of 500? (ztest for population mean)
7) Explain one tailed and two tailed test.
8) Define estimation .Explain in detail about point estimation.