Chapter ten

The Normal Distribution and Its Application to Tests of Statistical Significance

Parametric tests for a normal population distribution

What are the Characteristics of the Normal Frequency Distribution?

What is the *z*-Score?

When can We Use the Normal Sampling Distribution?

What are the Assumptions of the One-Sample *z*-Test for Means?

Using the normal sampling distribution when population parameters are unknown

How Can We Make Assumptions About an Unknown Population?

What is the Central Limit Theorem?

When Can It Be Used?

Two examples

How Can We Define a Sampling Distribution When the Parameters Are Unknown?

What is the *z*-Test for Proportions?

What is the *t*-Test for Means?

D. Weisburd and C. Britt, *Statistics in Criminal Justice*, DOI 10.1007/978-1-4614-9170-5_10,

In chapters 8 and 9, tests of statistical significance were presented that did not make assumptions about the population distribution of the characteristics studied. We now turn to a different type of test of statistical significance in which the researcher must make certain assumptions about the population distribution. These tests, called parametric tests, are widely used in criminal justice and criminology because they allow the researcher to test hypotheses in reference to interval-level scales.

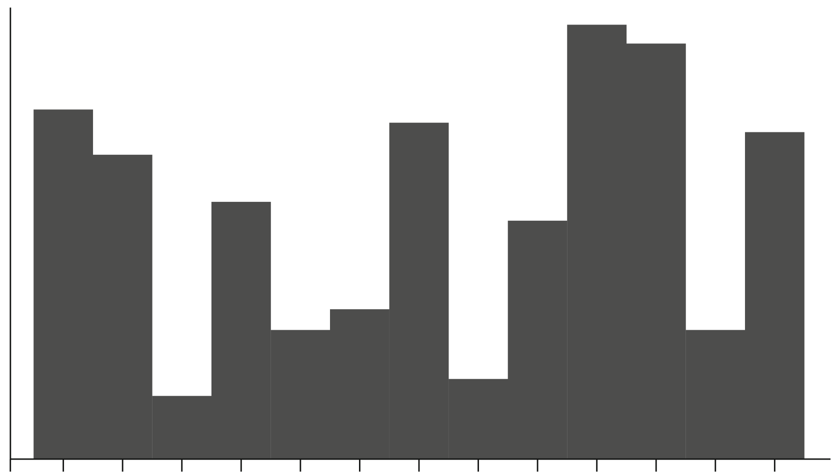
We begin by introducing the normal sampling distribution and its application to tests of significance for measures that are normally distributed in the population. We then turn to a basic dilemma faced by researchers in the use of parametric tests. The purpose of statistical inference is to make statements about populations from what is known about samples. However, parametric tests require that we make assumptions about the population at the outset. If population parameters are generally unknown, how can we make assumptions about them? In this chapter, we examine this dilemma in the context of two types of parametric tests that are based on the normal distribution.

The Normal Frequency Distribution, or Normal Curve

In Chapter 3, we noted that frequency distributions may take many different forms. Sometimes there is no pattern to a distribution of scores. This is the case for the example in Figure 10.1, in which the frequency of scores goes up and down without consistency. But often a distribution begins to take a specific shape. For example, Floyd Allport suggested more than half a century ago that the distribution of deviant behavior is shaped like a J.[1] His J curve, represented in Figure 10.2, fits many types

[1]F. H. Allport, "The J-Curve Hypothesis of Conforming Behavior," *Journal of Social Psychology* 5 (1934): 141–183.

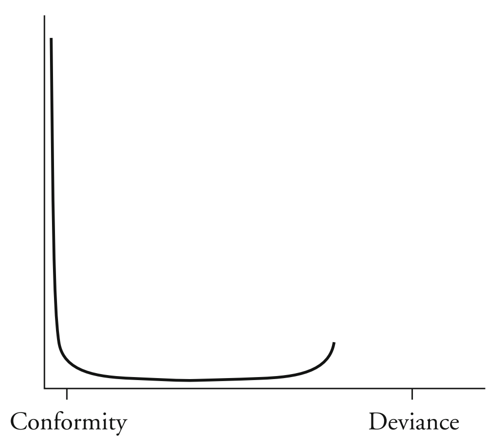
Figure 10.1 *Random Frequency Distribution*



of rule-breaking behavior and suggests a theory of deviance in which social control leads most people to conform more or less to societal rules. Allport fit a J curve to behaviors as diverse as parking violations, conformity to religious rituals in church, and stopping at a stop sign.

The most widely utilized distributional form in statistics is what is defined as the **normal frequency distribution** or **normal curve.** The normal distribution is the basis for a number of parametric statistical

Figure 10.2 *The J Curve*



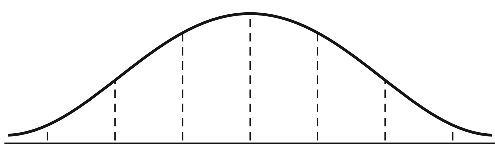
tests. This is the case in good part because of a set of special characteristics associated with the normal curve.

Characteristics of the Normal Frequency Distribution

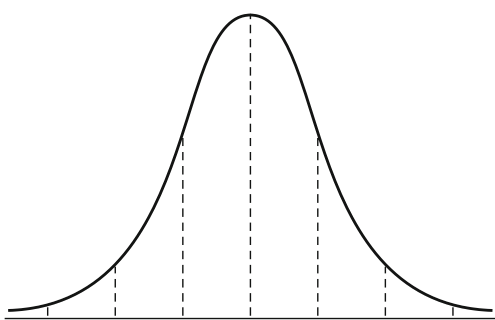
A normal distribution is always symmetrical and bell shaped. By that we mean that it is shaped exactly the same on both sides of its mean. If you represent a normal distribution as a curve, you can fold it over at its mean and gain two half-curves that are exactly alike. Of course, there are many different potential bell-shaped curves that are symmetrical, as illustrated in Figure 10.3. The curve in part a of Figure 10.3, for example, is fairly flat. What this means is that the scores are fairly widely spread around the mean. The curve in part b, in contrast, is very peaked. Here, the scores are tightly clustered around the mean. In the statistical language developed in earlier chapters, we can say that the standard deviation of the first distribution is much larger than that of the second.

In a true normal distribution, the mean, mode, and median are always the same. This can be seen in the normal curves in Figure 10.3. If the distribution is completely symmetrical, then the 50th percentile score, or the median, must be right in the middle of the distribution. In turn, since

Figure 10.3 *Two Examples of Normal Curves*



(a) *Normal Curve with a Large Standard Deviation*

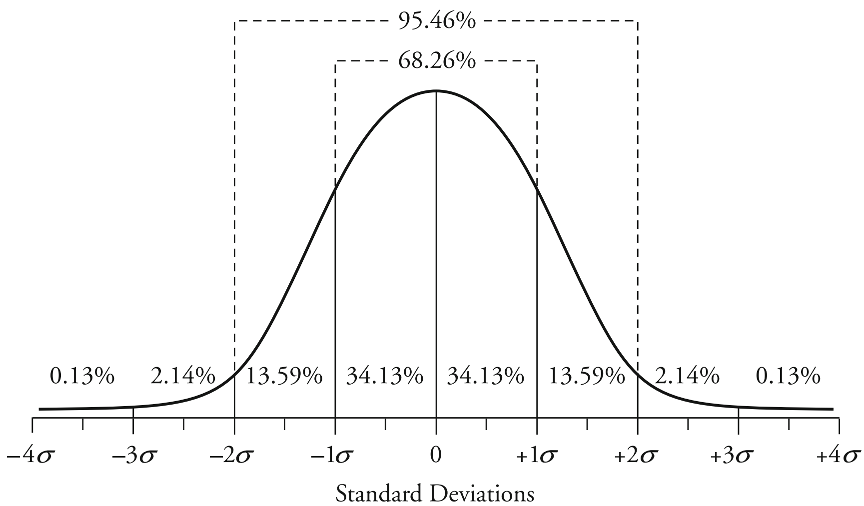


(b) *Normal Curve with a Small Standard Deviation* the middle of the distribution represents its highest peak, and thus the largest frequency of scores, it is also the location of the mode for the normal distribution. Finally, given that there is an exactly equal distribution of scores below and above that peak, the same value must also be the mean for a normal distribution.

All of these traits help to define a normal distribution. However, the most useful characteristic of a normal distribution develops from the fact that the percentage of cases between its mean and points at a measured distance from the mean is always fixed. The measure in this case is the **standard deviation unit.** A standard deviation unit is simply the standard deviation for the particular distribution being examined. For example, let's say that you were examining the results of a standardized test for assessing adjustment of prisoners and that the distribution obtained was a normal distribution. You obtained a mean score of 90 and a standard deviation of 10 for your sample. The standard deviation unit of this distribution would be 10. That is, if you measured one standard deviation unit from the mean in either direction, you would move 10 points from the mean, to 100 and 80. If you measured two standard deviation units from the mean, you would move 20 points, to 110 and 70.

In a normal distribution, 68.26% of the cases in the distribution are found within one standard deviation unit above and below the mean (see Figure 10.4). Because the normal distribution is symmetrical, this means that 34.13% of the cases lie within one standard deviation unit to either the right (positive side) or the left (negative side) of the mean.

Figure 10.4 *Percentage of Cases Under Portions of the Normal Curve*



Fully 95.46% of the cases are found within two standard deviation units above and below the mean. Virtually all of the cases in a distribution with a normal form are within three standard deviation units of the mean, although in theory the tails (or extremes) of this distribution go on forever. For the sample of inmates discussed above, we thus know that slightly more than two-thirds have adjustment scores of between 80 and 100 (one standard deviation unit above and below the mean). Very few members of the sample have adjustment scores above 120, which represents a score that is three standard deviation units from the mean.

z-Scores

Using a simple equation, we can convert all normal distributions, irrespective of their particular mean or standard deviation, to a single **standard normal distribution.** This distribution can then be used to identify the exact location of any score. We do this by converting the actual scores in our sample or population to ***z*-scores,** which represent standard deviation units for the standard normal distribution. This distribution has a mean of 0 and a standard deviation unit of 1. The formula for converting a specific score to a *z*-score is represented by Equation 10.1.

$$z = \frac{X_i - \mu}{\sigma} \qquad \textbf{Equation 10.1}$$

For this equation, we take the score of interest and subtract from it the mean score for the population distribution (represented by μ). We then divide that result by the standard deviation of the population distribution we are examining (represented by σ). In practice, what this formula does is allow us to convert any specific score in any normal distribution to a *z*-score in a standard normal distribution. We can then use a standardized table to identify the location of that score. A concrete example will make this conversion easier to understand.

Intelligence quotient (IQ) scores are normally distributed in the U.S. population, with a mean of 100 and a standard deviation of about 15. Suppose a probation officer is writing a report on a young offender. She finds that the young man has an IQ of 124. She wants to give the sentencing judge a good sense of what this means in terms of how this young man compares to others. She can do this by transforming the mean IQ of the offender (124) to a *z*-score and then identifying where this *z*-score fits in the standard normal distribution. We use Equation 10.1 for this purpose.

As shown in the numerator of the equation, we subtract the population mean (μ) of IQ scores, which we already noted was 100, from the score of 124. By doing this we shift the position of our score. We now have its location if the mean of our distribution were 0—the mean of a standard normal distribution.

If the mean were 0, then the score for this offender would be 24 (and not 124). As a second step, we divide this result by 15, the standard deviation (σ) of IQ scores in the U.S. population. This is equivalent to converting our sample standard deviation unit to 1, the standard deviation of the standard normal distribution, since each score of 15 is equivalent to one z standard deviation unit. The result is 1.60.

Working It Out

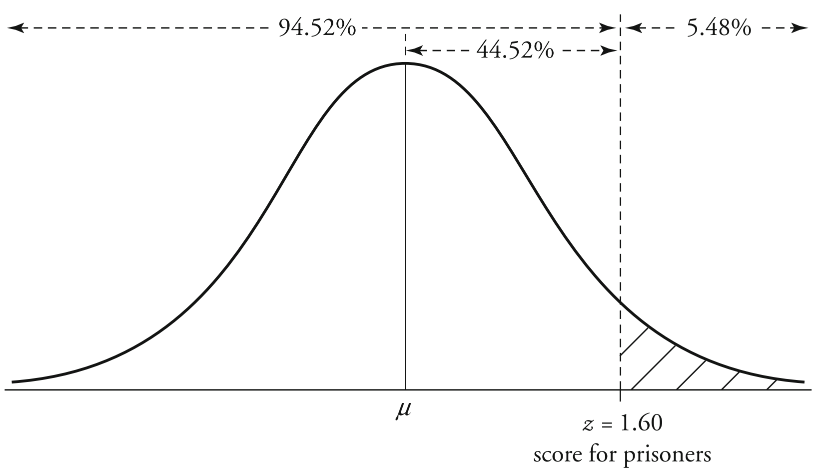
$$z = \frac{X_i - \mu}{\sigma}$$

$$= \frac{124 - 100}{15}$$

$$= 1.60$$

Our final step is to compare this z-score to an already prepared table of the standard normal distribution, provided in Appendix 3. You will notice that the z table goes up to only 0.50. This is because it provides us with only half of the normal curve. On this half of the normal curve, our z-score of 1.60 is equivalent to 0.4452, meaning that 44.52% of the scores lie between 0 and +1.60 standard deviations from 0. In Figure 10.5, our result is illustrated in the context of the normal curve. Because our result is a positive score, we place the value on the right-hand side of the normal distribution.

Figure 10.5 *IQ Score of Young Prisoner Compared to Average IQ Score of the General Population*



To identify the percentage of people in the general population with IQ scores higher than that of the young offender, we subtract our result of 0.4452 from 0.50 (the proportion of cases in this half of the curve). Our result of 0.0548 means that only a bit more than 5% of the general population has higher IQ scores than this offender. Conversely, almost 95% of the population has lower IQ scores than this offender. By converting the offender's score to a score on the standard normal distribution, we are able to place his intelligence in context. From our finding, we can see that he is indeed a highly intelligent young man, based on his IQ score.

Developing Tests of Statistical Significance Based on the Standard Normal Distribution: The Single-Sample *z*-Test for Known Populations

The normal distribution can also be used as a sampling distribution. When a population of scores is distributed normally, the sampling distribution of sample means will also be distributed normally. This means in practice that we can use the normal distribution as our sampling distribution for a test of statistical significance if we know at the outset that the variable of interest is normally distributed in the population. This is the case for the IQ test, so we will continue to use it as an example. Let's say that you were interested in whether American prisoners differ from Americans generally in terms of average IQ scores. The population characteristics for all Americans are, as discussed above, known. The mean score for the population is 100, and the standard deviation of the population mean is 15. You conduct a study of 125 prisoners selected through an independent random sampling procedure from the population of American prisoners. You find that the mean IQ in your sample is 90.[2] This mean is different from the mean of the American population. But we know that samples vary, and thus you might get a mean of 90 even if the mean for American prisoners were the same as that for the general population. What we want to know is how likely we are to get such an outcome in our sample if the distribution of American prisoners is the same as that of the general American population.[3] Be-

---

[2]Our hypothesized results mirror those found in prior studies; see R. J. Hernstein, "Some Criminogenic Traits of Offenders," in J. Q. Wilson (ed.), *Crime and Public Policy* (San Francisco: Institute for Contemporary Studies, 1983). Whether these differences mean that offenders are, on average, less intelligent than nonoffenders is an issue of some controversy in criminology, in part because of the relationship of IQ to other factors, such as education and social status.

[3]By implication, we are asking whether it is reasonable to believe that our sample of prisoners was drawn from the general population. For this reason, the *z*-test can also be used to test for random sampling. If you have reason to doubt the sampling methods of a study, you can conduct this test, comparing the observed characteristics of your sample with the known parameters of the population from which your sample was drawn.

cause the population parameters of the American population are known, a ***single-sample z-test*** for known populations is appropriate.

We set up our test of statistical significance the same way we did other tests in previous chapters.

Assumptions:

Level of Measurement: Interval scale.

Population Distribution: Normal distribution.

Sampling Method: Independent random sampling.

Sampling Frame: The American prison population.

Hypotheses:

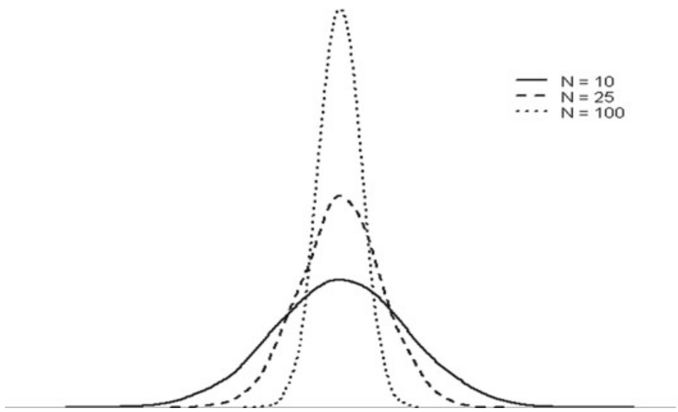
H_0: The mean IQ of the population from which our sample of prisoners was drawn is the same as the mean IQ of the general population of Americans ($\mu = 100$).

H_1: The mean IQ of the population from which our sample of prisoners was drawn is not the same as the mean IQ of the general population of Americans ($\mu \neq 100$).

As required by the single-sample *z*-test for known populations, IQ scores are measured at an interval level. As already noted, IQ is also normally distributed in the general population, meaning that it is appropriate to use a normal sampling distribution to conduct our test of statistical significance. Our sample, as required by our test, is drawn randomly with replacement from the American prison population. Our null hypothesis is that the mean IQ of American prisoners is the same as the mean IQ of the general American population ($\mu = 100$). Our research hypothesis is that the mean IQ of prisoners is different from that of the average American ($\mu \neq 100$).

The Sampling Distribution The mean of the sampling distribution we use for our test of statistical inference is defined, as in other tests, by our null hypothesis. In statistical tests using the normal distribution we can assume that the mean of the sampling distribution is the same as the mean of the population distribution. In this case, it is 100, or the mean IQ for the American population. The standard deviation is drawn from our knowledge of the standard deviation of the population distribution of scores. However, we cannot simply take the standard deviation of scores for the population distribution and apply it to the sampling distribution for our test, because the standard deviation of the sampling distribution is influenced by the number of observations in a sample. This is illustrated in Figure 10.6, which presents three different sampling distributions for the same population distribution of scores. In the first, there are only 10 cases in the samples from which the sampling distribution is developed. In the second, there are 25 cases in each sample. Finally, in the third

Figure 10.6 *Normal Distribution of Scores from Samples of Varying Sizes: N = 10, N = 25, and N = 100*



distribution, there are 100 cases in each sample. What is clear here is that the spread of scores is reduced as the size of samples in the distribution increases. This fact is a very important one in statistics and follows what our common sense tells us: Our sampling distribution becomes more tightly clustered around the mean as N increases. This implies, in practice, that we are less likely to draw deviant samples (those far from the mean of the sampling distribution) as the *N* of cases in our samples increases. Put in lay terms, larger samples are more trustworthy or more likely to reflect the true population score, all else being equal, than are smaller samples.

In order to differentiate between the standard deviation of a population distribution of scores and that of a sampling distribution, statisticians call the standard deviation of a sampling distribution the **standard error**. Using Equation 10.2, we adjust our standard error for the fact that the dispersion of sample means decreases as sample size increases. In order to distinguish the standard deviation (σ) from the standard error in this text, we will use the subscripts sd (for sampling distribution) whenever we refer to the standard error of a sampling distribution. Accordingly, the standard error of a sampling distribution is represented as σ_{sd} in Equation 10.2.

$$\text{Standard error} = \sigma_{\text{sd}} = \frac{\sigma}{\sqrt{N}}$$ **Equation 10.2**

For our example, we find the standard error of the sampling distribution by dividing the population standard deviation of IQ, 15, by the square root of our sample *N*. The result is 1.342.

Working It Out

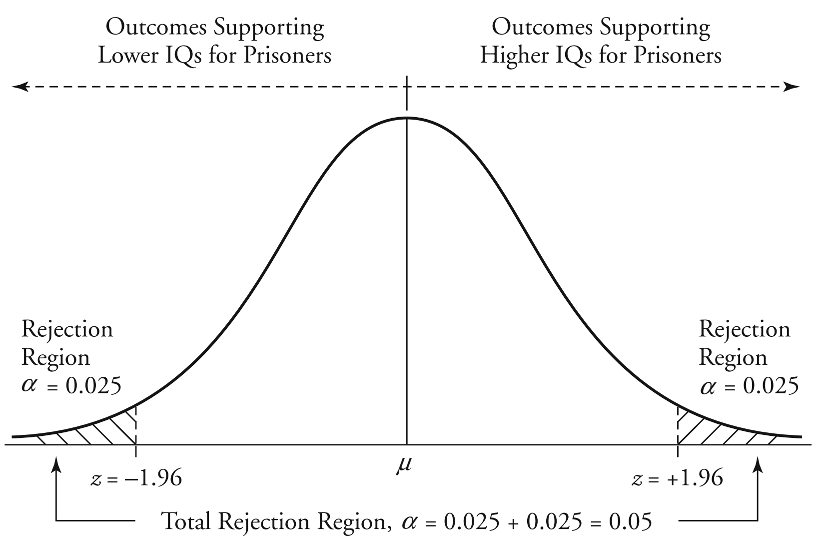
$$\text{Standard error} = \frac{\sigma}{\sqrt{N}}$$

$$= \frac{15}{\sqrt{125}}$$

$$= 1.342$$

Significance Level and Rejection Region Given that no special concerns have been stated in regard to the risk of either a Type I or a Type II error, we use a conventional 0.05 significance threshold. As our research hypothesis is nondirectional, we use a two-tailed test. What this means for our rejection region is illustrated in Figure 10.7. On the right-hand side of the distribution are outcomes greater than the average American IQ of 100. On the left-hand side of the distribution are outcomes less than the average. Because our research hypothesis is not directional, we split our total rejection region of 5% between both tails of the distribution. This is represented by the shaded area. Each shaded area represents half the total rejection region, or 0.025.

Figure 10.7 *Rejection Region on a Normal Frequency Distribution for a 0.05 Two-Tailed Significance Test*

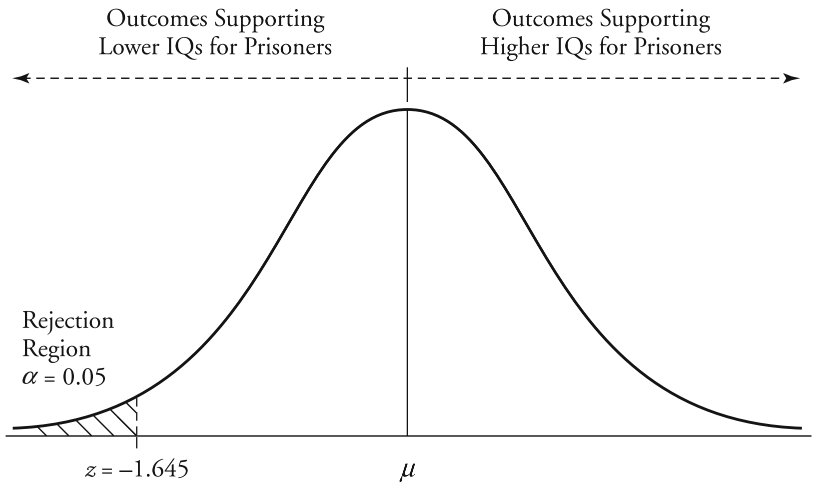


To define the z-score that corresponds with our rejection region, we must turn to the table of probability values associated with the z distribution in Appendix 3. As discussed earlier in the chapter, the z table represents only half of the normal curve. We look at the value associated with 0.4750 (0.5000 − 0.0250) in the table, which is 1.96. If we observe a test statistic either greater than 1.96 or less than −1.96, we will reject the null hypothesis of our test (see Figure 10.7). In this case, our observed significance level would be less than the 0.05 criterion for our test.

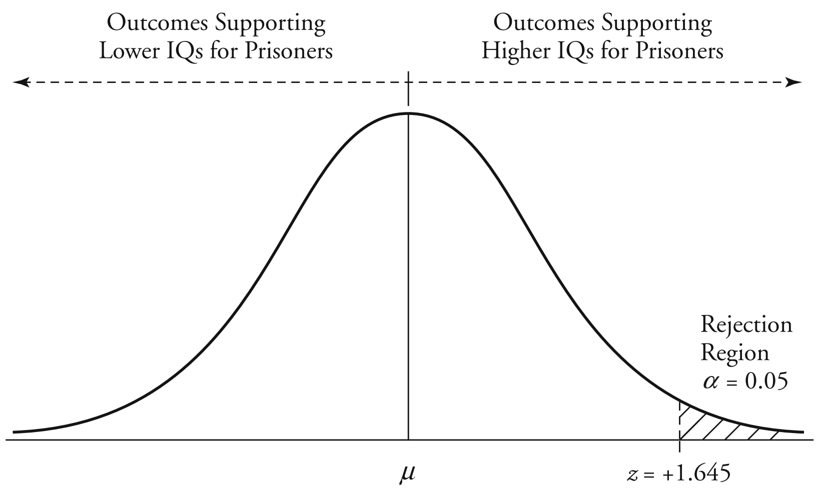
If we had stated a directional research hypothesis, we would place the entire rejection region ($\alpha = 0.05$) in one of the two tails of the normal distribution. In this case, we would conduct a one-tailed statistical test. Parts a and b of Figure 10.8 represent the rejection regions for two different one-tailed tests of statistical significance. If our research hypothesis stated that average IQs for prisoners were less than those for the U.S. population, we would place the entire rejection region of 0.0500 in the left tail of the distribution (see Figure 10.8a). We again consult the z table in Appendix 3 to identify the z-score associated with a value of 0.4500 (0.5000 − 0.0500). We observe that 0.4500 falls exactly halfway between two values in the table —0.4495 and 0.4505—corresponding to z-scores of −1.64 and −1.65, respectively. How do we determine the value of z in such a case? The most accurate value for z would be found by interpolating between −1.64 and −1.65, which would give −1.645, since the value we are looking for is halfway between the two proportions reported in the z table. In this case, if our test statistic is less than −1.645, then we reject our null hypothesis and conclude that the average IQ for prisoners is less than the U.S. average. If our research hypothesis stated that the average IQ of prisoners was greater than the U.S. average, we would place the rejection region on the right side of the distribution (see Figure 10.8b). In such a case, our critical value would be +1.645, meaning that if our test statistic was greater than 1.645, we would reject the null hypothesis and conclude that the average IQ for prisoners was greater than the U.S. average.

The Test Statistic To calculate our test statistic, we can use the same formula we did in examining the relative position of a score in the standard normal distribution, with two important differences. In this case, we have to take into account the fact that sampling distributions become more tightly spread around their mean as the N of sample cases becomes larger. As discussed in defining the sampling distribution above, we need to adjust the standard deviation of the population distribution by dividing it by the square root of the N of our sample. This provides us with the standard error (σ_{sd}) for our distribution. We also need to subtract the

Figure 10.8 *Rejection Region on a Normal Frequency Distribution for a 0.05 One-Tailed Significance Test*



(a) *Lower IQs for Prisoners*



(b) *Higher IQs for Prisoners*

mean (μ) of the population score from $\overline{X}$, rather than X_i. These adjustments are made in Equation 10.3.

$$z = \frac{\overline{X} - \mu}{\sigma_{\text{sd}}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{N}}$$ **Equation 10.3**

Inserting into our equation the mean value of our sample and its N of cases and the mean and standard deviation for the population of scores, we obtain a z-test statistic of -7.453.

Working It Out

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{N}}$$

$$= \frac{90 - 100}{15/\sqrt{125}}$$

$$= -7.453$$

The Decision Because our test statistic is less than our negative critical value ($-7.453 < -1.96$) and falls in the rejection region, we reject the null hypothesis. We conclude on the basis of our study that the mean IQ of the population from which our sample was drawn is different from that of the general American population.

Applying Normal Sampling Distributions to Nonnormal Populations

The example of IQ presents a case where the single-sample z-test can be used to test hypotheses involving interval-scale measures. However, it requires that the population distribution for the measure be normal. In some fields in the social sciences, measures are constructed in such a way that they are normally distributed in practice.[4] But in criminology, there

[4]In principle, any distribution may be arranged in such a way that it conforms to a normal shape. This can be done simply by ranking scores and then placing the appropriate number within standard deviation units appropriate for constructing a standard normal distribution.

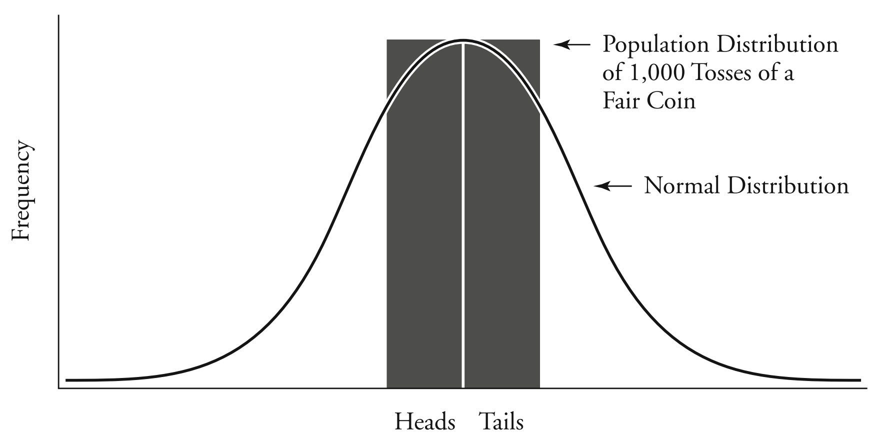
has been much less use of distributions that are standardized in normal form, in part because the distributions of the behaviors and populations that we confront do not often conform to the shape of a normal curve. Even measures that do begin to approximate the shape of the normal distribution seldom meet all the requirements of a true normal distribution.

How, then, can parametric tests based on the normal distribution be widely used to make statistical inferences? Not only do they demand that we make an assumption about a population we usually know little about, but the assumption we are being asked to make does not make very much sense for criminal justice measures. The answer may be found in an important distinction between population distributions on the one hand and sampling distributions on the other. While we have every reason to be hesitant in assuming that the population distribution of scores is normal for criminal justice measures, we can assume with a good deal of confidence that the sampling distributions for such measures are approximately normal. Using the toss of a fair coin as an example, we can provide a simple illustration of this fact.

In Figure 10.9, we overlay the distribution of scores for a population of 1,000 tosses of a fair coin over the normal distribution. As is apparent, outcomes in a coin toss are not distributed normally. This makes good sense, since there are only two possible scores for the coin toss: heads and tails. No matter what the outcome, it is impossible for a coin toss to approximate the form of the normal distribution.

But let's now turn to a sampling distribution for the coin toss. In this case, we want to know the likelihood of gaining a specific number of heads in a set number of coin tosses. This is the logic we used

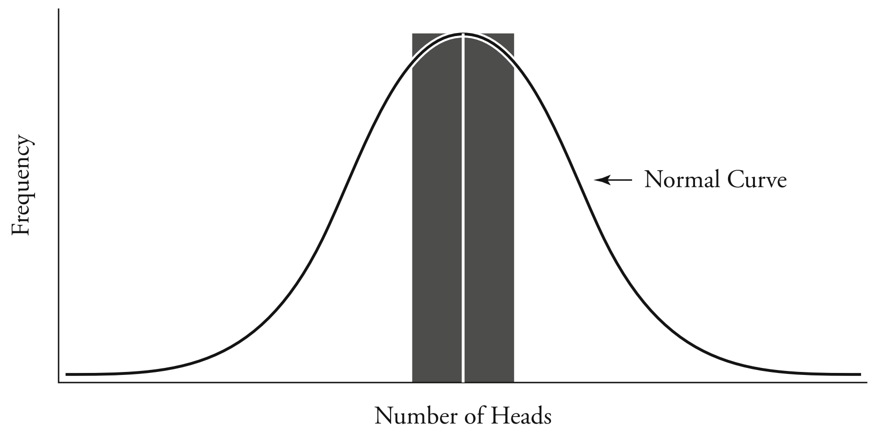
Figure 10.9 *Distribution of 1,000 Tosses of a Fair Coin Contrasted to the Normal Distribution*



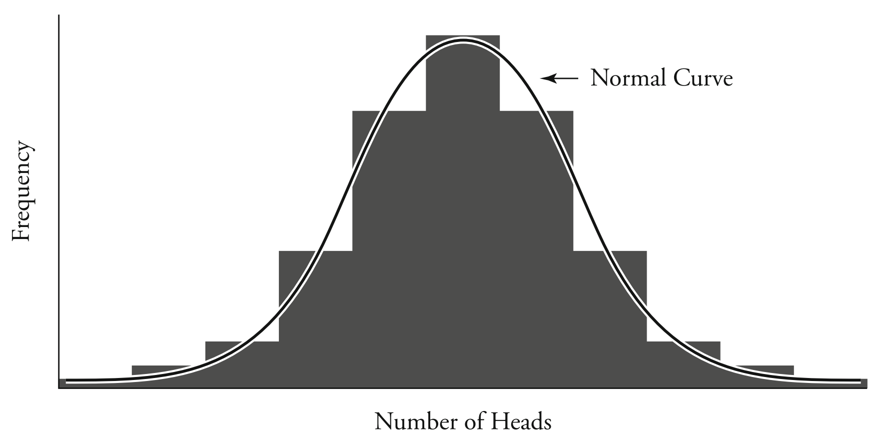
in developing the binomial probability distribution in Chapter 7. Figure 10.10 presents the binomial distribution for different-size samples of the coin toss under the null hypothesis that the coin is fair.

For a sample size of 1 (Figure 10.10a), the shape of the sampling distribution is the same as the shape of the population distribution of scores. However, notice what happens as the size of the samples used to construct the sampling distributions grows. For a sample of 10 (Figure 10.10b), the histogram for the distribution of scores is still jagged, but it has begun to take a shape similar to the normal distribution. Importantly, for a sample of 10, we do not have two potential outcomes, which

Figure 10.10 *Sampling Distribution of Coin Tosses*

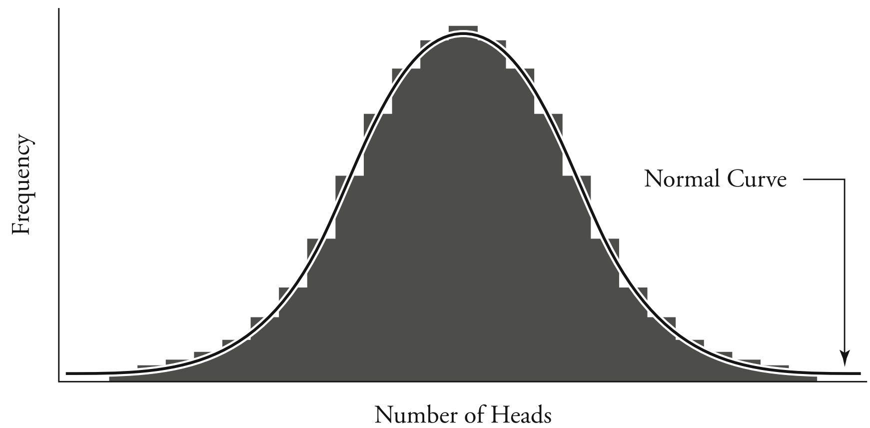


(a) *1 Toss of a Fair Coin*

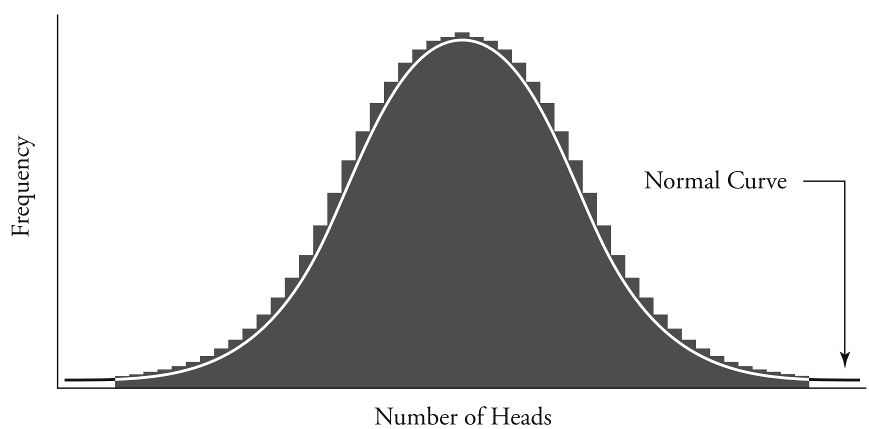


(b) *10 Tosses of a Fair Coin*

Figure 10.10 *Sampling Distribution of Coin Tosses (cont.)*



(c) *100 Tosses of a Fair Coin*



(d) *400 Tosses of a Fair Coin*

would make a normal shape impossible, but 11 potential outcomes (no heads, one head, two heads, three heads, four heads, . . . to ten heads). This is the case because we are flipping the coin ten times for each sample. The sampling distribution is telling us the number of times we would expect to gain a specific number of heads in ten tosses of a fair coin in a very large number of trials. For a sample of 100 flips of a fair coin (Figure 10.10c), the sampling distribution even more closely approximates the normal curve. By the time we get to a sample of 400 flips of the coin (Figure 10.10d), the sampling distribution of a fair coin is almost indistinguishable from the normal curve.

The population distribution of scores for a fair coin is very far from a normal form. Yet sampling distributions for the same population begin to approximate the normal distribution as the size of the sample of coin tosses grows. This remarkable fact is summarized in a very important theorem, or statement, about sampling distributions called the **central limit theorem.** The central limit theorem allows us to overcome our initial dilemma because it says that under many circumstances we can use a normal sampling distribution for making inferences about a population that is not normal in shape.

Central Limit Theorem

If repeated independent random samples of size N are drawn from a population, then as N grows large, the sampling distribution of sample means will be approximately normal.

The central limit theorem tells us that when the number of cases in a sample is large, we can assume that the sampling distribution of sample means is approximately normal even if the population distribution itself is not normal. This is what is meant by the statement "then as N grows large, the sampling distribution of sample means will be approximately normal." However, the theorem does not provide us with a clear statement about how large the number of cases in a sample must be before we can make this assumption.

One reason for this ambiguity is that the number of cases needed before the sampling distribution begins to approximate normality depends in part on the actual distribution of the measure examined in the population. As can be seen from the example of the coin toss, even when the population distribution departs markedly from the normal distribution, the sampling distribution fits fairly closely to the normal curve with a sample size of 100. For this reason, you will find wide agreement that a normal sampling distribution can be assumed for samples of 100 or more, irrespective of the distribution of scores in a population.

There is much less agreement about what to do when a sample is smaller than 100 cases. Some statisticians argue that with 50 cases you can be fairly confident that the central limit theorem applies in most circumstances. Others apply this yardstick to 25 or 30 cases, and still others argue that under certain circumstances—for example, when prior studies suggest a population distribution fairly close to normality—only 15 cases is enough. In conducting research in criminal justice, you should recognize that there is no hard and fast rule regarding sample size and the central limit theorem. In practice, in criminal justice, researchers generally assume that 30 cases is enough for applying the central limit theorem. However, when a distribution strongly departs from normality, as is the case with a proportion, it is safer to require more than 100 cases.

While the central limit theorem solves a major problem in applying normal distribution tests to criminological questions, we are still faced with a barrier in actually carrying out such tests. As we saw earlier (see Equation 10.3), the standard error of the z sampling distribution is gained from knowledge about the standard deviation of the population distribution. How can we identify the standard error of a sampling distribution if we do not know the standard deviation of the population distribution? In the following sections, we illustrate two methods for defining σ for an unknown population. In the first, we take advantage of a special relationship between the mean and the standard deviation of a proportion. In the second, we estimate the unknown parameter based on information gained in our sample.

Comparing a Sample to an Unknown Population: The Single-Sample *z*-Test for Proportions

One implication of the central limit theorem is that we can use a normal sampling distribution to test hypotheses involving proportions. This might seem strange at first, since we estimate the shape of a normal distribution through knowledge of its mean and standard deviation. As discussed in Chapter 4, the mean and standard deviation are not appropriate statistics to use with a nominal-level measure such as a proportion.

Nonetheless, as illustrated in the previous section, the sampling distribution of a proportion—in our example, the coin toss (see Figure 10.10)—begins to approximate a normal distribution when the number of cases for the sample becomes large. The central tendency of this distribution and its dispersion are measured by the mean and standard error, just as in distributions that develop from interval-level data. Accordingly, although it would be inappropriate to use the mean and standard deviation to describe a sample or population distribution of a proportion, the mean and standard error are appropriate statistics for describing the normal sampling distribution that is associated with the same proportion.

Computing the Mean and Standard Deviation for the Sampling Distribution of a Proportion

How do we compute the mean and standard deviation of a proportion? One way to do this would be to simply apply the formula for the mean and the standard deviation to the scores associated with a proportion. However, there is a simpler way to arrive at the same result. It turns out that the mean of a proportion is equal to the proportion itself. This is illustrated in Table 10.1, which shows an example in which the mean and proportion are calculated for five heads in ten tosses of a coin.

Table 10.1 Calculating the Mean and Proportion of 5 Heads in 10 Tosses of a Coin

CALCULATING THE MEAN FOR FIVE HEADS	CALCULATING THE PROPORTION FOR FIVE HEADS
$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{1+1+1+1+1+0+0+0+0+0}{10} = 0.5$	$\text{Proportion} = \frac{N_{\text{successes}}}{N_{\text{total}}} = \frac{5}{10} = 0.5$

Note: head = 1 : tail = 0

For the numerator of the mean, we sum the scores on the ten trials (five ones and five zeros) and get 5. The numerator of a proportion is the N of cases in the category of interest. If the category is heads, then we also get a result of 5. The denominators for both equations are the same (10), and thus the outcomes are also the same. As a general rule, we state that for a proportion $\mu = P$.

What about the standard deviation of a proportion? It turns out that we can calculate the standard deviation with knowledge of only the proportion itself. This is illustrated in Table 10.2.

Taking the sum of the squared deviations from the mean and dividing it by N, we get a result of 0.25. But we can get this same result by multiplying the proportion of heads (P) by the proportion of tails (Q)—in our case, multiplying 0.5 by 0.5. Accordingly, we can substitute $P \bullet Q$ for

$$\frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}$$

Table 10.2 Calculating the Standard Deviation of 5 Heads in 10 Tosses of a Coin

CALCULATING THE STANDARD DEVIATION FROM THE RAW SCORES	CALCULATING THE STANDARD DEVIATION FROM *P* AND *Q*
$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}}$ $= \sqrt{\frac{0.25+0.25+0.25+0.25+0.25+0.25+0.25+0.25+0.25+0.25}{10}}$ $= \sqrt{0.25} = 0.5$	$\sigma = \sqrt{PQ} = \sqrt{(0.5)(0.5)} = \sqrt{0.25} = 0.5$

in the equation for the standard deviation for the mean:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \overline{X})^2}{N}} = \sqrt{PQ} \qquad \textbf{Equation 10.4}$$

Because of this relationship between the mean and the standard deviation of a proportion, when we state the proportion of successes expected under the null hypothesis, we also state by implication the mean and the standard deviation for the population distribution of scores. So if we state in the null hypothesis that the proportion of successes in the population is 0.50, we know that the mean of the population distribution of scores for our test of the null hypothesis is 0.50 and its standard deviation is 0.25 ($\sigma = \sqrt{PQ} = \sqrt{(0.50)(0.50)} = \sqrt{0.25} = 0.50$).

What this means in practice is that we need not have any a priori knowledge of the shape of the population distribution to construct a sampling distribution for our test of proportions. With a large *N*, we can assume a normal sampling distribution, irrespective of the actual form of the population distribution. Through our null hypothesis, we can define both the mean and the standard deviation of the population distribution for our test. We are now ready to use the normal distribution to test hypotheses about unknown population parameters.

Testing Hypotheses with the Normal Distribution: The Case of a New Prison Program

Suppose that you were asked to evaluate a new prison education program. The foundation sponsoring the effort sought to achieve a program success rate of 75% among the 100,000 prisoners enrolled in the program. Success was defined as completion of a six-month course supported by the foundation. Managers of the program claim that the success rate is actually much greater than the criteria set by the foundation. However, a recent newspaper exposé claims that the success rate of the program is actually much below 75%. You are able to collect information on 150 prisoners, selected using independent random sampling. You find that 85% of your sample successfully completed the course. What conclusions can you make, based on your sample results, about the claims of managers and the newspaper exposé?

Assumptions:

Level of Measurement: Interval scale (program success is measured as a proportion).

Population Distribution: Normal distribution (relaxed because *N* is large).

Sampling Method: Independent random sampling.

Sampling Frame: 100,000 prisoners in the program.

Hypotheses:

H_0: The success rate of the program is 0.75 ($P = 0.75$).

H_1: The success rate of the program is not 0.75 ($P \neq 0.75$).

Because the number of cases in our sample is greater than the threshold of 100 suggested for invoking the central limit theorem in the case of a proportion, we can ignore—or, in statistical terms, **relax—assumptions** regarding the shape of the population distribution. In the special case of a proportion, we can also relax the assumption of an interval scale of measurement.[5] Our sample, as assumed by our test, is drawn randomly with replacement from the sampling frame of 100,000 prisoners in the program.

Our research hypothesis is nondirectional. Managers of the program claim that the program has a success rate of greater than 0.75 ($P > 0.75$). The newspaper exposé claims that the success rate is much below 75% ($P < 0.75$). Accordingly, we want to be able to examine both of these potential outcomes in our test. The null hypothesis is that the rate of success for the program is 0.75 ($P = 0.75$).

The Sampling Distribution In calculating the mean and standard deviation or standard error for our sampling distribution, we rely on our null hypothesis. Our null hypothesis states that the proportion of successes in the population is 75%. This means that the mean of the sampling distribution is also 0.75. We can calculate the standard error of the sampling

[5] It would not make sense, however, to use a normal distribution test for nominal-scale measures with more than two categories. The normal distribution assumes scores above and below a mean. The sampling distribution of a proportion follows this pattern because it includes only two potential outcomes, which then are associated with each tail of the distribution. In a multicategory nominal-scale measure, we have more than two outcomes and thus cannot fit each outcome to a tail of the normal curve. Because the order of these outcomes is not defined, we also cannot place them on a continuum within the normal distribution. This latter possibility would suggest that the normal distribution could be applied to ordinal-level measures. However, because we do not assume a constant unit of measurement between ordinal categories, the normal distribution is often considered inappropriate for hypothesis testing with ordinal scales. In the case of a proportion, there is a constant unit of measurement between scores simply because there are only two possible outcomes (e.g., success and failure).

distribution by adjusting Equation 10.2 to the case of a proportion, as illustrated in Equation 10.5:

$$\sigma_{\text{sd}} = \frac{\sigma}{\sqrt{N}} = \frac{\sqrt{PQ}}{\sqrt{N}} = \sqrt{\frac{PQ}{N}} \qquad \textbf{Equation 10.5}$$

Applying this equation to our problem, we obtain a standard error of 0.035 for the normal sampling distribution associated with our null hypothesis:

Working It Out

$$\begin{aligned}\sigma_{\text{sd}} &= \sqrt{\frac{PQ}{N}} \\ &= \sqrt{\frac{(0.75)(0.25)}{150}} \\ &= \frac{\sqrt{0.1875}}{\sqrt{150}} \\ &= \frac{0.433}{12.25} \\ &= 0.0353\end{aligned}$$

In order to test our hypothesis, we will convert this sampling distribution, with mean 0.75 and standard error 0.035, to the standard normal distribution (or z), which has a mean of 0 and a standard deviation or standard error of 1. This calculation is done when we calculate the test statistic below.

Significance Level and Rejection Region Given that no special concerns have been stated in regard to the risk of either a Type I or a Type II error, we use a conventional 0.05 significance threshold. As our research hypothesis is nondirectional, we use a two-tailed test. As our level of significance is the same as in our previous problem, we follow the same procedure and arrive at a critical value of 1.96. If we observe a test statistic either greater than 1.96 or less than −1.96, we will reject the null hypothesis of our test.

The Test Statistic We can rely on the same formula used in the single-sample z-test for known populations, presented earlier in Equation 10.3. However, in Equation 10.6, we express the formula with proportions rather than means:

$$z = \frac{\bar{X} - \mu}{\sigma_{sd}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}} = \frac{p - P}{\sqrt{PQ/N}}$$ **Equation 10.6**

The mean of the sample (p) is 0.85, since this is the outcome of the study. The mean of the sampling distribution P (0.75) is taken from our null hypothesis. The standard error of the sampling distribution (σ_{sd}) was calculated earlier based on our null hypothesis that the proportion of successes was 0.75. Our result is a z-score of 2.833.

Working It Out

$$z = \frac{p - P}{\sqrt{PQ/N}}$$

$$= \frac{0.85 - 0.75}{\sqrt{(0.75)(0.25)/150}}$$

$$= \frac{0.10}{0.0353}$$

$$= 2.8329$$

The Decision Our test statistic is well within the rejection region of our test (which includes scores greater than 1.96 or less than −1.96), meaning that our observed significance level is less than the significance level we set for our test at the outset ($p < 0.05$). We therefore reject the null hypothesis at a 0.05 significance level. We come out on the side of the managers of the program. Our sample results support their position that the overall program has exceeded the criterion for success of the foundation.

Comparing a Sample to an Unknown Population: The Single-Sample *t*-Test for Means

The proportion provides us with a special case in which we can calculate the standard error of our sampling distribution based on our null hypothesis. But this is not possible when our null hypothesis relates to a mean of an interval-level measure. In this case, there is not one specific variance or standard deviation associated with a mean but an infinite number of potential variances or standard deviations. How, then, can we test hypotheses about unknown parameters in the case of the mean?

One obvious method is to simply use the variance of our sample as a "guesstimate" of the variance of the population distribution. The problem with this solution is that the variance of a sample is a somewhat biased estimate of the variance of the population. By this we mean that the average of repeated observations of the variance (s^2) tends in the long run not to be equivalent to the value of σ^2. We can transform s^2 to a better estimate of σ^2 through a very small correction to the equation for the variance. This new statistic (expressed as $\hat{\sigma}^2$ since it is an estimate of σ^2) is represented in Equation 10.7.[6] An estimate of the standard deviation ($\hat{\sigma}$) can be gained by taking the square root of this value.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} (X_i - \overline{X})^2}{N - 1}$$ **Equation 10.7**

In order to use this new statistic to test hypotheses, we must also use a slightly different sampling distribution, called the t distribution. It is sometimes called Student's t because its inventor, W. S. Gossett (1876–1936), first described the distribution under the pen name Student in 1908.

The t distribution (Appendix 4) is very similar to the z distribution (Appendix 3). However, as with the chi-square test, the shape of the t distribution is dependent on the number of degrees of freedom. The number of degrees of freedom for a single-sample t-test is defined as $N - 1$. When the number of cases in a sample is greater than 500, then the t and z distributions are virtually identical. However, as the number of cases in a sample gets smaller and smaller, and so accordingly does the number of degrees of freedom, the t distribution becomes flatter and a larger and larger test statistic is needed to reject the null hypothesis.

This fact can be illustrated by looking at the t table in Appendix 4. As you can see, the t table lists the critical values associated with six significance thresholds for both one- and two-tailed tests. Let's focus on the fourth column, which is the critical value associated with a two-tailed, 5% significance level. When the number of degrees of freedom is 500, the critical value for the t-statistic is the same as for the z distribution: 1.960. At 120, the t value needed to reject the null hypothesis is still almost the same: 1.980. At 100, the value is 1.982: at 50, it is 2.008; and at 25, it is 2.060. The largest differences come for even smaller degrees of freedom.

[6]As noted on page 105 (footnote 1), computerized statistical analysis packages, such as SPSS, use this corrected estimate in calculating the variance and standard deviation for sample estimates.

The t distribution presents a new problem as well in making inferences to unknown populations. Relaxing the assumption of normality is generally considered more risky in a t-test than in a z-test. This makes good sense because we are now using an estimate of σ rather than the actual population parameter. As the number of cases increases, our confidence in this estimate grows.[7] How large should N be before you are willing to use a t-test? With samples of more than 30 cases, your statistical conclusions are not likely to be challenged. However, the t distribution is particularly sensitive to outliers. Conclusions based on smaller samples should be checked carefully to make sure that one or two observations are not the cause of a very large statistical outcome.

Testing Hypotheses with the *t* Distribution

We are now ready to turn to a practical example. Suppose that the study described earlier also examined the average test scores for those prisoners who had completed the program. The foundation set a standard of success of 65 on the test. Program managers say that prisoners who have completed the program achieve average scores much higher than this. The newspaper exposé again claims that the average scores are considerably lower than those expected by the foundation. In this case, you are able to take an independent random sample of 51 prisoners who have completed the test. You find that the test mean for the sample is 60, and the standard deviation is 15. What conclusions about the larger population of prisoners can you come to based on your sample results?

Assumptions:

Level of Measurement: Interval scale.

Population Distribution: Normal distribution (relaxed because N is large).

Sampling Method: Independent random sampling.

Sampling Frame: Prisoners who have completed the program.

Hypotheses:

H_0: The mean test score for prisoners who have completed the program is 65 ($\mu = 65$).

H_1: The mean test score for prisoners who have completed the program is not 65 ($\mu \neq 65$).

---

[7]Our statistical problem is that we assume that μ and σ are independent in developing the t distribution. When a distribution is normal, this is indeed the case. However, for other types of distributions, we cannot make this assumption, and when N is small, a violation of this assumption is likely to lead to misleading approximations of the observed significance level of a test.

Following the assumptions of our test, we use an interval scale (the mean of test scores) and an independent random sampling method. We relax the assumption of normality because N is larger than the minimum threshold of 30 recommended for interval-level measures. Our research hypothesis is once again nondirectional so that we can examine the positions of both the managers of the program and the newspaper exposé. The null hypothesis is that the mean test score for the population of prisoners completing the program is 65 (the foundation standard), or that $\mu = 65$.

The Sampling Distribution Because σ is unknown and cannot be deduced from our null hypothesis, we will use the t distribution. The number of degrees of freedom for our example is defined as $N - 1$, or $51 - 1 = 50$.

Significance Level and Rejection Region Again, we have no reason in this example to depart from the 0.05 significance threshold. Because our research hypothesis is not directional, we use a two-tailed test. Turning to the t table, we find that a t-score of 2.008 is associated with a two-tailed, 5% significance threshold (at 50 degrees of freedom). This means that we will reject our null hypothesis if we obtain a test statistic greater than 2.008 or less than −2.008. For these observed values of our test statistic, the observed significance level of our test is less than the criterion of 0.05 that we have selected.

The Test Statistic The test statistic for the t distribution is similar to that for the z distribution. The only difference is that we now use an estimate of the standard deviation ($\hat{\sigma}$) rather than σ itself.

$$t = \frac{\overline{X} - \mu}{\sigma_{\text{sd}}} = \frac{\overline{X} - \mu}{\hat{\sigma}/\sqrt{N}}$$

Equation 10.8

Although we can get an estimate of σ by adjusting the calculation for s, the formula for t may also be written in a way that allows us to calculate t from the unadjusted sample standard deviation.

$$t = \frac{\overline{X} - \mu}{\hat{\sigma}/\sqrt{N}} = \frac{\overline{X} - \mu}{\sqrt{\dfrac{\sum_{i=1}^{N} (X_i - \overline{X})^2}{N - 1}} \bigg/ \sqrt{N}}$$

$$= \frac{\overline{X} - \mu}{\sqrt{\dfrac{\sum_{i=1}^{N} (X_i - \overline{X})^2}{N}} \bigg/ \sqrt{N - 1}}$$

This means that we can simplify the equation for the *t*-test as follows:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{N - 1}}$$ **Equation 10.9**

Applying the *t* formula to our example, we use the mean of the sample, 60, as $\bar{X}$; μ is defined by the null hypothesis as 65; *s* is our sample standard deviation of 15; and *N* is the number of cases for our sample (51).

Working It Out

$$\begin{aligned} t &= \frac{\bar{X} - \mu}{s/\sqrt{N - 1}} \\ &= \frac{60 - 65}{15/\sqrt{51 - 1}} \\ &= \frac{-5}{15/\sqrt{50}} \\ &= \frac{-5}{2.1213} \\ &= -2.3570 \end{aligned}$$

The Decision Because the test statistic of −2.3570 is less than −2.008, we reject the null hypothesis and conclude that the result is significantly different from the goal set by the foundation. In this case, our decision is on the side of the newspaper exposé. We can conclude from our sample (with a 5% level of risk of falsely rejecting the null hypothesis) that the test scores in the population of prisoners who have completed the program are below the foundation goal of 65.

Chapter Summary

Parametric tests of statistical significance allow us to make inferences about a population from samples using interval-level data. In a parametric test, we make certain assumptions about the shape of the population distribution at the outset.

The **normal distribution,** or **normal curve,** is widely used in statistics. It is symmetrical and bell shaped. Its mean, mode, and median are always the same. There will always be a set number of cases between the mean and points a measured distance from the mean. The measure of this distance is the **standard deviation unit.** All normal distributions, irrespective of their mean or standard deviations, can be converted to a single standard normal distribution by converting the actual scores in the sample or population to *z*-scores. To use a normal sampling distribution for a test of statistical significance, we must assume that the characteristic studied is normally distributed in the population.

An important dilemma in statistical inference is created by this assumption. How can we make assumptions about the population distribution when its characteristics are generally unknown? The **central limit theorem** describes an important fact that allows us to solve this problem. As stated in the theorem, when the number of cases in a sample is large, the sampling distribution will be approximately normal in shape, even if the population distribution itself is not. In the field of criminal justice, it is generally assumed that the central limit theorem can be applied where the sample size is 30 or greater. When dealing with proportions, though, it is safer to require a sample size of at least 100. In such circumstances, we may **relax the assumption** of normality. We can now make inferences using a normal sampling distribution, even though the shape of the population distribution is unknown.

In order to define the sampling distribution, we need information about the population parameters—information that is not usually available. In the case of a test involving proportions, the null hypothesis can be used to define both the mean and the standard error of the population distribution. Once the population parameters have been defined by the null hypothesis, we can apply the formula for the *z*-test of statistical significance. In the case of a test of means, the null hypothesis cannot be used directly to define the standard error. We may, however, use the *t* sampling distribution, which relies on an estimate of the **standard error.**

Key Terms

central limit theorem A theorem that states: "If repeated independent random samples of size *N* are drawn from a population, as *N* grows large, the sampling distribution of sample means will be approximately normal." The central limit theorem enables the researcher to make inferences about an unknown population using a normal sampling distribution.

normal curve A normal frequency distribution represented on a graph by a continuous line.

normal frequency distribution A bell-shaped frequency distribution, symmetrical in form. Its mean, mode, and median are always the same. The percentage of cases between the mean and points at a measured distance from the mean is fixed.

relaxing an assumption Deciding that we need not be concerned with that assumption. For example, the assumption that a population is normal may be relaxed if the sample size is sufficiently large to invoke the central limit theorem.

single-sample *t*-test A test of statistical significance that is used to examine whether a sample is drawn from a specific population with a known or hypothesized mean. In a *t*-test, the standard deviation of the population to which the sample is being compared is unknown.

single-sample *z*-test A test of statistical significance that is used to examine whether a sample is drawn from a specific population with a known or hypothesized mean. In a *z*-test, the standard deviation of the population to which the sample is being compared either is known or—as in the case of a proportion— is defined by the null hypothesis.

standard deviation unit A unit of measurement used to describe the deviation of a specific score or value from the mean in a *z* distribution.

standard error The standard deviation of a sampling distribution.

standard normal distribution A normal frequency distribution with a mean of 0 and a standard deviation of 1. Any normal frequency distribution can be transformed into the standard normal distribution by using the *z* formula.

***z*-score** Score that represents standard deviation units for a standard normal distribution.

Symbols and Formulas

p Proportion of successes (sample)

P Proportion of successes (population)

Q Proportion of failures (population)

σ_{sd} Standard error of a normal distribution

t *t*-score

$\hat{\sigma}$ Estimate of σ

To determine the *z*-score for a single observation:

$$z = \frac{X_i - \mu}{\sigma}$$

To determine the standard error of a sampling distribution:

$$\sigma_{sd} = \frac{\sigma}{\sqrt{N}}$$

To determine the z-score for a sample mean:

$$z = \frac{\bar{X} - \mu}{\sigma_{\text{sd}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}$$

To determine the standard deviation of a proportion:

$$\sigma = \sqrt{PQ}$$

To determine the z-score for a sample proportion:

$$z = \frac{\bar{X} - \mu}{\sigma_{\text{sd}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}} = \frac{p - P}{\sqrt{PQ/N}}$$

To estimate the value of σ from data in a sample:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N - 1}}$$

To determine the value of t:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{N - 1}}$$

Exercises

10.1 In which of the following circumstances would a researcher be justified in using a normal sampling distribution? Explain how or why for each case.

a. A sample of 10 subjects is drawn to study a variable known to be normally distributed in the population.

b. A sample of 50 subjects is drawn to study a variable known to be normally distributed in the population.

c. A sample of 10 subjects is drawn to study a variable. The shape of the distribution of this variable in the population is unknown.

d. A sample of 50 subjects is drawn to study a variable. The shape of the distribution of this variable in the population is unknown.

e. A sample of 50 subjects is drawn to study a proportion. The shape of the distribution of this proportion in the population is unknown.

10.2 A team of psychologists has created an index they claim measures an individual's "ability to control anger." The index is calculated from the answers to a detailed questionnaire and is normally distributed among U.S. adult males, with a mean of 100 and a standard deviation of 30. Researchers assess a group of ten prisoners, all of whom have been convicted for violent rapes. They discover that the mean score for the group is 50.8.

a. What percentage of U.S. adult males would be expected to obtain a score equal to or less than that of the rapists?

b. The psychologists who constructed the index consider the bottom 10% of U.S. adult males on their distribution to be "strongly inclined to use violence to solve social problems." Albert is a respectable businessman who scores 60.6 on the scale. Is Albert included in this category? Explain why.

c. What percentage of U.S. adult males would be expected to score between 110 and 120 on the "anger index"?

10.3 A teacher gives the following assignment to 200 students: Check the local newspaper every morning for a week and count how many times the word "gun" is mentioned on the "local news" pages. At the end of the week, the students report their totals. The mean result is 85, with a standard deviation of 8. The distribution of scores is normal.

a. How many students would be expected to count fewer than 70 cases?

b. How many students would be expected to count between 80 and 90 cases?

c. Karen is a notoriously lazy student. She reports a total of 110 cases at the end of the week. The professor tells her that he is convinced she has not done the assignment, but has simply made up the number. Are his suspicions justified?

10.4 The professors who teach the Introduction to Psychology course at State University pride themselves on the normal distributions of exam scores. After the first exam, the current professor reports to the class that the mean for the exam was 73, with a standard deviation of 7.

a. What proportion of student would be expected to score above 80?

b What proportion of students would be expected to score between 55 and 75?

c. What proportion of students would be expected to score less than 65?

d. If the top 10% of the class receive an A for the exam, what score would be required for a student to receive an A?

e. If the bottom 10% of the class fail the exam, what score would earn a student a failing grade?

10.5 A noted criminologist, Leslie Wilkins, has suggested that the distribution of deviance in the population follows a normal bell-shaped curve, with "sinners" at one extreme, "saints" at the other, and most of us falling somewhere in between the two. Working on the basis of this theory, a researcher constructs a detailed self-report survey whereby individuals are given a score based on the offenses they have committed in the past year, with the score weighted according to the relative triviality or seriousness of each offense. The lower the score, the nearer the individual approximates "sinner" status, and the higher the score, the closer he or she is to being a "saint." From his initial sample of 100 adults in a specific state, the researcher computes a mean score of 30, with a standard deviation of 5.

a. If the researcher's model is correct, below which score should he expect to find the 5% of U.S. society with the greatest propensity to deviance?

b. In his sample of 100, the researcher is surprised to discover that 50 subjects score greater than 35 on the deviance test. How many cases would be expected under the assumption of a normal distribution of saints and sinners? What does this suggest about the original theory?

10.6 An established test measuring "respect for authority" has a mean among U.S. adults of 73 and a standard error of 13.8. Brenda gives the test to 36 prison inmates and finds the mean score to be 69.

a. Is this enough evidence to suggest that the prisoners belong to a population that has significantly less respect for authority than the general U.S. adult population?

b. Assuming there is enough information, test whether this sample differs significantly from the population. Use a significance level of 5% and outline each of the stages of a test of statistical significance.

10.7 The governor of Stretford Prison has a biographical record of all the inmates. The mean age of all the inmates is 22, with a standard deviation of 7.5. A recent survey by a hostile researcher makes damaging criticisms of the educational standards in the prison. The prison governor suspects that the 50 prisoners interviewed for the study were not chosen at random. The mean age of the prisoners chosen is 20. Show how a test for statistical significance can be used by the governor to cast doubt on the sampling method of the survey. Use a significance level of 5% and outline each of the stages of a test of statistical significance.

10.8 A hundred years ago, an anonymous scientist wrote a famous indictment of a notoriously cruel prison somewhere in the United States. Without ever referring to the prison by name, the scientist checked the records of all those who were imprisoned over its 50-year history and found that 15% of those who entered died within. Henry, a historian, is intrigued by the old report and publishes an article in a historical journal in which he states his conviction that the report was referring to Grimsville Prison, which existed about that time. In a subsequent issue of the journal, a rival historian claims that Henry has shown no evidence to support his theory.

Henry finds the records from Grimsville, and from a sample of 80 prisoner records he discovers that 11% of the prisoners died inside. Can he use this information to substantiate his claim that the object of the report is indeed Grimsville? Use a significance level of 5% and outline each of the stages of a test of statistical significance.

10.9 Every pupil at Foggy Lane College was asked a series of questions, which led to an overall score grading "satisfaction" with the college's discipline procedures. The overall mean score was 65. Roger suspects that the black students at the college feel differently. He takes a random sample of 25 black students from the college and finds that their mean satisfaction score is 61, with a standard deviation of 8.

Are the black students' views on discipline significantly different from those of the general student population? Use a significance level of 1% and outline each of the stages of a test of statistical significance.

10.10 A special police unit has spent several years tracking all the members of a large child-abuse ring. In an interview with a daily newspaper, a junior detective on the unit claims that the ringleaders have been tracked down and will shortly be arrested. In response to questions from the interviewer about the makeup of the child-abuse ring, the detective replies, "We have gathered details on every last member of this criminal group—they come from very varied backgrounds and their average age is 36."

X is the chairperson of a charitable club, which is in fact a front for a substantial child-abuse circle. He reads the newspaper article and fears that it might refer to him and his group. He looks through the club's membership files and draws a sample of 50 members, finding an average age of 40 with a standard deviation of 9.

Can X be confident that the detective interviewed in the newspaper was not referring to *his* criminal group?

10.11 A civil rights group is concerned that Hispanic drug offenders are being treated more severely than all drug offenders in Border State. A state government web site reports that all drug offenders were sentenced to an average of 67 months in prison. The group conducts a small study by taking a random sample of public court records. For the 13 Hispanic drug offenders in the sample, the average sentence was 72 months ($s = 8.4$). Use a 5% significance level and test whether Hispanic drug offenders in Border State are sentenced more severely. Be sure to outline the steps in a test of statistical significance.

10.12 A researcher believes that offenders who are arrested for committing homicides in her city are younger than the national average. A review of FBI arrest statistics for recent years indicates that the mean age of homicide offenders is 18.7. The researcher collects information on a random sample of 25 persons arrested for homicide in her city and finds the mean age to be 16.8, with a standard deviation of 4.1. Can the researcher conclude that homicide offenders in her city are younger than the national average? Use a significance level of 0.05. Be sure to outline the steps in a test of statistical significance.

10.13 Following a revolution, the new leadership of the nation of Kippax decides to hold a national referendum on whether the practice of capital punishment should be introduced. In the buildup to the referendum, a leading army general wishes to gauge how the people are likely to vote so that he can make a public statement in line with popular feeling on the issue. He commissions Greg, a statistician, to carry out a secret poll of how people expect to vote. The results of Greg's poll are as follows: The sample proportion in favor of introducing capital punishment is 52%.

Do the results indicate that the majority of the population favors introducing capital punishment? Use a significance level of 5% and outline each of the stages of a test of statistical significance.

10.14 The Silver Star Treatment Center claims to be effective at reducing drug addiction among the persons who go through its treatment regimen. As evidence of the effectiveness of the Silver Star treatment, the director claims that 63% of all drug users nationally have a relapse within 12 months of treatment, but in a random sample of 91 cases treated by Silver Star, only 52% had a relapse within 12 months of completing the treatment. Use a 1% level of significance to test whether Silver Star's treatment is effective at reducing drug use. Be sure to outline the steps in a test of statistical significance.

10.15 A federal judge issues an opinion claiming that nonviolent drug offenders should make up no more than 20% of the local jail population. If a jail is found to have more than 20% nonviolent drug offenders, the jail will fall under court order and be required to release inmates until the composition of the jail population conforms to the judge's standard. The local sheriff draws a random sample of 33 inmates and finds that 23% have been convicted of nonviolent drug offenses. Should the sheriff be concerned about the jail coming under court supervision? Use a significance level of 0.05. Be sure to outline the steps in a test of statistical significance.