Chapter 3

Frequency Distributions

1. Introduction

The most fundamental type of data collection problem that lends itself to statistical analysis involves the measurements of each of a set of objects on a single variable; for example, achievement test scores of a set of students. While there are many different types of analysis that such data could be subjected to, in this chapter we will introduce the most basic type of statistical analysis, the *frequency distribution*; a simple tally or count of how frequently each value of the variable occurs among the set of measured objects.

2. Creation of a Frequency Distribution

To better understand the procedure and rationale for creating a frequency distribution, we can consider a typical example of its application. Table 1 presents the test performance of 36 students. The data could just as well represent the number of accounts opened by a set of sales reps, the number of arrests among a set of criminals, the number of absences among a set of employees, the number of operations performed among a set of surgeons, the heights of a set of children; or, if the objects were other than people, the maze-running errors of a set of rats, the crop yields of various plots of land, the prices of stocks, the recall levels for different TV commercials, etc.

While an inspection of the raw data presented in Table 1 provides information about the individual performance of the students, it is difficult to get a concise picture of their overall collective performance. The data would be more comprehensible if it could be summarized into a more compact and interpretable form. This is where the frequency distribution comes into play. In examining the data we notice that many values of the variable occur more than once, so it is reasonable that we can summarize the data by making a *tally* or count of how frequently each value occurs. In constructing such a frequency

Table 1 Test scores of 36 individuals.

Individuals	Test scores	Individuals	Test scores
Anderson, B.	12	Kornfield, L.	11
Andrews, T.	9	Lee, R.	10
Barclay, S.	8	Logan, B.	14
Bishop, C.	10	Marsh, N.	8
Brody, R.	15	Melrose, G.	10
Carlton, M.	11	Moran, C.	9
Clark, D.	7	Noble, V.	10
Cox, S.	14	Parker, L.	12
Dewey, D.	10	Potter, D.	13
Edelman, P.	13	Rhodes, F.	8
Farrell, J.	11	Rubin, B.	10
Frank, R.	7	Schultz, R.	11
Gibbs, J.	9	Silver, W.	9
Gray, W.	11	Stack, E.	13
Harmon, G.	12	Thomas, J.	12
Hodge, N.	14	Vargas, R.	11
Irving, T.	6	Weiss, C.	10
Kent, N.	16	Wheeler, E.	9

distribution, the most logical arrangement of the values of the variable is either from lowest to highest, or from highest to lowest, whichever we prefer.

In the present example, the test scores range from a low of 6 to a high of 16, and are listed in the first column of Table 2. In the second column of Table 2 is a tally of how frequently each value of the variable occurs. It was obtained by working down the list of 36 values appearing in Table 1, and making a tally mark for each one beside its corresponding value in Table 2. Finally, in column (3), the tally counts are shown in their numerical form. Columns (1) and (3), then, comprise a frequency distribution table. The column (2) tallies are typically dropped from the table since they represent an intermediate work step, although they do serve as a rough graphical portrayal of the distribution.

A comparison of Table 1 with columns (1) and (3) of Table 2 reveals the data reduction function of a frequency distribution. We find that our original set of data has been condensed and summarized into a more compact and interpretable form. At a glance we know the range of scores as well as where they cluster most heavily. These features of the data were not readily apparent from a mere inspection of Table 1. With larger collections of data, the benefits of the frequency distribution are even more apparent.

Relative frequencies. It is often helpful to convert the frequencies of a frequency distribution into *relative frequencies*, which are nothing more than

Table 2 Frequency distributions of 36 test scores.

(1) Test scores	(2) Tally of individuals	(3) Frequencies	(4) Relative frequencies	(5) Cumulative frequencies	(6) Cumulative relative frequencies
6	|	1	2.8%	1	2.8%
7	||	2	5.6	3	8.3
8	|||	3	8.3	6	16.7
9	𝍸	5	13.9	11	30.6
10	𝍸 ||	7	19.4	18	50.0
11	𝍸 |	6	16.7	24	66.7
12	||||	4	11.1	28	77.8
13	|||	3	8.3	31	86.1
14	|||	3	8.3	34	94.4
15	|	1	2.8	35	97.2
16	|	1	2.8	36	100.0
		36	100.0%		

the observed frequencies converted into percentages based on the total number of observations.

Such relative frequencies appear in column (4) of Table 2. They were obtained simply by dividing each frequency in column (3) by the total number of observations, 36. The relative frequencies tell us at a glance what percentage of the 36 students had a score of a given value. This information is helpful since it is often easier to interpret a percentage figure than a raw frequency figure. For example, it is somewhat more informative to be told that 19.4% of the students had a score of 10 on the test, than to say that 7 of the 36 students had such a score.

Cumulative frequencies. There are many times when not only do we want to know how many observations in a data collection are of a particular value, but also how many are *above* or *below* a given value. For instance, with respect to the student test scores we might want to know how many students scored, say, 10 or lower. To answer such questions we can create a *cumulative frequency distribution*, a distribution that indicates how many of the observations in the data collection occur *up to and including each particular value*.

Such cumulative frequencies are shown in column (5) of Table 2. They will be seen to represent a running total or "accumulation" of the frequencies appearing in column (3). At a glance, then, we see that 18 of the 36 students had a score of 10 or lower; or that 34 had a score of 14 or lower. Necessarily, the highest score in the distribution, in this case 16, must be associated with a

cumulative frequency corresponding to the total number of observations, in this instance 36.

Cumulative relative frequencies. Once we have formed a cumulative frequency distribution, it is a simple matter to convert it into a *cumulative relative frequency distribution*. All we need to do is divide each cumulative frequency by the total frequency of observations. An alternative approach is to accumulate the relative frequencies. Except for differences due to rounding the two methods will yield the same results.

The cumulative relative frequency distribution tells us what percentage of the total observations in our data collection are of a particular value or lower. Column (6) of Table 2 shows the cumulative relative frequencies for the student test score example. We can see at a glance that 66.7% of the students scored 11 or lower, or that 94.4% scored 14 or lower. If we wanted to know what percentage scored above a given value, we merely have to subtract the cumulative relative frequency from 100%. For example, it is easily verifiable that 83.3% of the students scored above 8.

3. Graphical Presentation

It is usually very informative to make a graphical representation of a frequency distribution. Such pictorial presentations can show aspects of the distribution not readily apparent from a tabular presentation. While tally marks, such as those in Table 1, give us a good picture of the frequency distribution, we can create a more formal and aesthetic presentation.

The two most common methods of graphically portraying a frequency distribution are shown in Figure 1. Part *a* of the figure shows a *histogram*, or bar chart, in which the horizontal axis shows the values of the variable in question, and the heights of the bars above each value represent their respective frequencies of occurrence. The data is taken from Table 2, the distribution of student test scores. Notice that Figure 1*a* is simply an alternative representation of the tally marks in Table 2, where the values of the variable are arranged horizontally rather than vertically.

Notice in Figure 1*a* that we have chosen to portray both the raw frequency distribution and the relative frequency distribution with the same histogram. This was accomplished by using the left vertical axis to represent the raw frequencies and the right vertical axis for the relative frequencies; the conversion of the observed frequencies to percentages having no effect on the shape of the distribution.

Figure 1*b* shows an alternative method of graphically portraying a frequency distribution. It is referred to as a *frequency polygon*, and is constructed by connecting the points which have heights corresponding to the

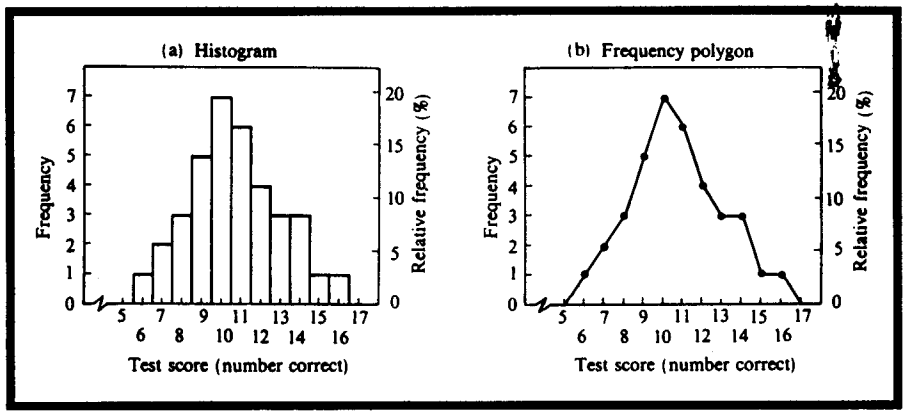


Figure 1 Alternative methods for portraying the frequency distribution of the test scores of 36 individuals as given in Table 2.

frequencies associated with each value along the horizontal axis. Alternatively, it can be thought of as the connection of the midpoints of the tops of the bars forming the histogram in part *a* of the figure. Again, the left vertical axis has been used to scale the observed frequencies, while the right axis shows the relative frequencies.

4. Grouping of Data

In the student test scores example we were fortunate that there were only eleven different values of the variable that occurred, ranging from 6 to 16 in integral values. At other times, we must contend with variables which assume a large number of values. For example, if instead of test scores we measured the students with respect to the variable of weight, we might expect to get values ranging from under 100 to over 200, and if the weight was measured to the nearest tenth of a pound, the number of possible values would be very large. In such situations, it is unlikely that any given value will occur with a frequency greater than one. Consequently, little is to be accomplished by making a frequency distribution of each value. The solution to such a data reduction problem is to create *intervals of values* of the variable in question, and then make a frequency tally of the number of observations falling within each interval.

To demonstrate the grouping procedure we can reanalyze the student test data using score intervals two units wide. The resulting frequency distribution

or less equal to each other. In such a nation, the number of five-year-olds is approximately the same as the number of ten-year-olds or the number of twenty-year-olds or the number of any given age; until, or course, we reach the upper age brackets where death due to aging begins to take its toll. But to that point the distribution of ages is more or less rectangular in form.

The *U-shaped* distribution shown in Figure 4*b* reveals a polarization of observed values for a given variable; either they tend to be very high or very low, with relatively few intermediate values occurring among the set of measured objects. An example of this type of data configuration is the distribution of consumer purchase-interest ratings for a particular product. Most people are either very favorably disposed to it, or they are turned off by it, with relatively few having an intermediate degree of interest in the product.

The distribution shown in Figure 4*c*, looking much like half of a *U*-shaped distribution, is often referred to as a *J-shaped* distribution. The one shown actually looks like a backward *J*, but it could also occur in the reverse direction with most of the observations piled up at the right instead of the left. The distribution shown could be characteristic of a variable such as the number of defects found in various batches of a quality controlled product. Most batches would have zero defects, fewer would have one defect, fewer yet would have two defects, even fewer three defects, and so on until the frequency of batches with numerous defects approaches zero. Another example of this type of reverse *J*-shaped distribution is the age distribution in a developing country in which the number of births increases each year and the expected longevity is short, resulting in more one-year-olds than two-year-olds, and more two-year-olds than three-year-olds, etc.

The *bell-shaped* distribution shown in Figure 4*d* is perhaps the most common of all distributions. The examples presented earlier in the chapter were of this type, situations in which observed values of the variable become increasingly more frequent at the intermediate values. Different bell-shaped distributions may vary in their specific profile, but there is one such distribution that has a specific shape, with a precise mathematical definition, that will be of special interest to us, for it is so pervasive in nature, and could well be said to be the most important distribution in the field of statistical analysis, and it is called the *normal distribution*. It is that distribution that we have already seen three times—in Figures 2*d*, 3, and 4*d*—and before the end of our study we should expect to see it or refer to it nearly a hundred times more, and its shape and characteristics will be as familiar to us as the multiplication tables are to a grade schooler, so basic is it to statistical analysis.

Distributions that are not symmetrical in form, those that tail off either to the right or the left, such as the *J*-shaped distribution, are referred to as *skewed* distributions. The distribution shown in Figure 4*e* is an example of a distribution skewed to the *right*, the direction of the skew being the direction of the

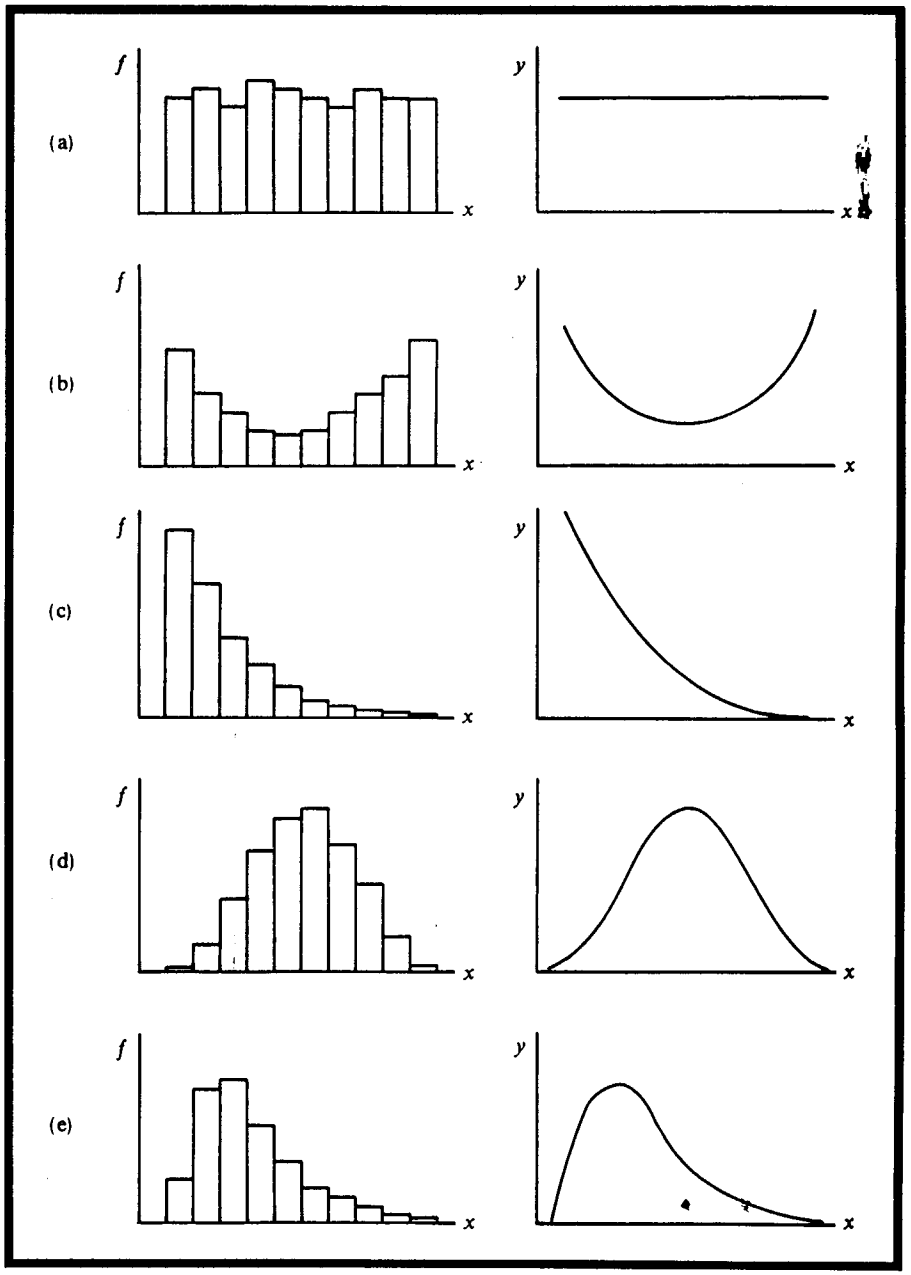


Figure 4 Common shapes of frequency distributions. Histograms and continuous distributions shown side by side: (a) uniform, (b) *U*-shaped, (c) reverse *J*-shaped, (d) bell-shaped, and (e) skewed right.

8. Concluding Comments

The creation of a frequency distribution for a set of data is usually the very first analysis that we perform. It gives us a concise overview of the data that is not possible from an examination of the unorganized observations. The shape of the frequency distribution, which we found could vary widely, provides the initial descriptive and interpretive summary of the data. In the following chapters we will discuss further data reduction techniques and interpretations based on the frequency distribution.

Chapter 4

Central Tendency

1. Introduction

Although a frequency distribution reduces a large collection of data into a relatively compact form, it would be highly desirable to further summarize the frequency distribution itself. While we can use verbal expressions such as "bell-shaped" or "skewed to the right" to describe the overall distribution of scores, such descriptions say nothing about the specific numerical values of the variable in question.

What would be beneficial would be a single summary value that would suggest a typical or representative observation, a measure of *central tendency*, or *location* as it is also called; that is, at which value do the observations "tend to center," or equivalently, where along the values of the variable in question are the observations clustered or "located." While such a measure of central tendency, by itself, would necessarily sacrifice much of the information inherent in the frequency distribution, it would nonetheless serve as a very concise description of a body of data; and when used in conjunction with a frequency distribution would provide a richer summary of the observations than either data reduction technique itself.

In the following sections we will discuss the most useful measures of central tendency, including their respective advantages and limitations.

2. The Mode

Since, as we have seen in the preceding sections, the observations in most data distributions tend to cluster heavily about certain values, one logical measure of central tendency would be that value which occurs most frequently; and that value is referred to as the *modal value*, or simply the *mode*.

For example, in the following data collection consisting of nine observations arranged in value from lowest to highest

9 12 15 15 15 16 16 20 26

the modal value is 15, occurring three times, more frequently than any of the other values. Also, it is important to remember that the mode is the value of the variable occurring most often, and not the frequency associated with that value.

When dealing with a distribution based on data grouped into intervals, the mode is often taken as the *midpoint* or *class mark* of that interval containing the highest frequency of observations. Alternatively, the interval itself could be cited as the modal value. For example, if a grouped frequency distribution of the heights of a set of corn stalks revealed that the 60 to 62 inch interval contained the highest frequency of observations, the mode could be reported as either 61 inches, or as the 60–62 inch interval, whichever we prefer.

There are instances in which a data distribution will have two modal values, characterized by two humps like the back of a camel. Such distributions are referred to as *bimodal*, even though one of the two modal values may be associated with a somewhat higher frequency than the other. Whenever we confront a bimodal distribution we should immediately question the composition of the collection of measured objects. Chances are that two distinct populations of objects, differing on the variable in question, are intermingled; e.g., the distribution of heights of a mixture of males and females.

The chief advantage of using the mode as a measure of central tendency is the ease with which it can be obtained, and its common sense interpretation. However, its limited mathematical properties make it less than the ideal measure of central tendency for more advanced analyses.

3. The Median

If we arrange a set of observations from lowest to highest in value, and then single out the middle value, we have identified what is called the *median value*, or simply the *median*, the value above and below which 50% of the observations fall.

For example, in the following simple collection of nine observations

22 24 24 25 27 30 31 35 40

the median value is 27. If we are dealing with an even number of observations instead of an odd number, as in the following collection of ten scores

22 24 24 25 27 30 31 35 40 47

then the median lies between the fifth and sixth values, midway between the values of 27 and 30—namely, 28.5.

In the case of a frequency distribution based on grouped data, the median can be reported roughly as that interval in which the cumulative relative frequency reaches 50%, or as the midpoint of that interval, or, for a more precise measurement, the median value can be determined by interpolation within the said interval.

Aside from its common sense interpretation as a truly central value, the most attractive feature of the median is its insensitivity to the values of the very extreme scores in a distribution, which are atypical and sometimes flukes. For example, if the highest value in the set of ten scores listed in the preceding paragraph were 85 instead of 47, the median would not be affected in the least. The distribution of individual incomes is a good example of how the median, because of its independence from the values of the extreme observations, is a good indicator of central tendency. While most individuals have incomes within a relatively narrow band, there is a minority who have exceptionally high incomes, many times the typical value. Since the definition of the median does not take into account the actual values of the extreme scores, it will not be affected by such deviant data points. For this reason, the median is especially appropriate as a measure of central tendency of a skewed distribution of data. Its main limitation, like that of the mode, is that it does not have mathematical properties that lend itself to more advanced analyses.

4. The Mean

The most important measure of central tendency, and one of the basic building blocks of all statistical analysis, is the *arithmetic mean*, or simply the *mean*. It is nothing more than the sum of a set of values, divided by the number of values involved.

Consider, for example, the following set of seven observations

3 4 4 5 6 8 10

Summing the values and dividing by the number of values, we have

$$\text{Mean} = \frac{3+4+4+5+6+8+10}{7} = \frac{40}{7} = 5.7$$

As in the above example, we typically calculate a mean to one decimal point beyond that occurring in the data itself, for it would be highly misleading as to the precision of our original observations if we reported the mean as, say, 5.714286.

The mean will be recognized as equivalent to the popular concept of an "average" of a set of numbers; more specifically, the arithmetic average. In other uses, the term "average" is used loosely to mean typical or representative, or central tendency in a most general sense.

Since the mean is such a fundamental concept to statistical analysis, it is useful to designate it in a shorthand form. The letter M is sometimes used to denote the mean, or it may be subscripted to identify the variable in question; for example, M_x would indicate the mean of variable x, while M_y would signify the mean of variable y.

An alternative and more common notation for the mean of a variable x, is $\bar{x}$, which is read "x bar" or "the mean value of variable x." By the same token, $\bar{y}$ signifies the mean value of a variable y. With this notation we can define the mean more concisely as

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

where n is understood to be the number of observations in our data collection, and $x_1, x_2, \ldots, x_n$ are the various observations.

Historically, the capital letter S was used to stand for "sum of" which allowed the even more compact definition

$$\bar{x} = \frac{S(x_i)}{n}$$

where $\bar{x}$, the mean of variable x, is shown to be the sum of the individual observations—$S(x_i)$—divided by the number of observations, n. However, since the letter S signifies another important statistical concept, it has been replaced with the capital Greek letter *Sigma*, designated Σ. Thus, the standard mathematical definition of the mean becomes

$$\bar{x} = \frac{\Sigma x_i}{n} \tag{1}$$

where Σx_i stands for the sum of the individual values of the variable x, n is the number of values in question, and $\bar{x}$ is the resultant mean value of the observations, as before.

The above formula can be made more explicit in the form

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}} \tag{2}$$

in which the range of the subscripts have been added to the summation sign. We read it, "the mean of variable x is equal to the sum of the values x_i, where i ranges from 1 to n, that sum divided by n." Typically, however, we will just use the summation sign Σ and understand implicitly the range of values being summed. And yet other times, for the sake of simplicity, we will also drop the subscript from x_i and simply write $\bar{x} = \Sigma x/n$.

The $\bar{x}$ notation is reserved for the mean of a *sample* of data. To distinguish the mean of a *population* of observations we use the Greek letter *mu*, designated μ, and pronounced *mew*. Thus, both $\bar{x}$ and μ signify a mean, but $\bar{x}$ refers to a sample, while μ refers to a population of observations. The distinction will take on greater significance when we undertake topics of inference, the drawing of conclusions about populations based on sample observations.

Whether one prefers the verbal or the symbolic definition of the mean makes no difference, so long as the concept is understood. In either case the mean should be made an integral part of one's intellectual hardware; for hardly a topic will go by from now on that does not depend either directly or indirectly on the mean, so important is it to the field of statistical analysis.

The major advantage of the mean over the mode and the median as a measure of central tendency is that it takes into account the *numerical value of every single observation* in the data distribution. It represents a balance point, or center of gravity, in that the sum of the distances to the observations below it, is equal to the sum of the distances to the observations above it. This mathematical characteristic of the mean, as we will see, makes it a cornerstone of statistical analysis.

Ironically, this feature of the mean—its sensitivity to every numerical value—becomes its chief drawback as a measure of central tendency in situations where the data distribution is highly skewed, and when there are one or two freaky "outliers" in the data. For example, the mean of the values 5, 6, 7, 8, and 9 is 7.0; while the mean for the set of values 5, 6, 7, 8, and 30 is 11.2, the dramatic increase in the value of the mean being due to the single exceptional score of 30. The median, on the other hand, uninfluenced by the deviant value, remains at 7.

When a data distribution is basically symmetrical in form, the mode, median, and mean will have very nearly the same value. In a skewed distribution the mean tends to get dragged *toward the tail* of the distribution, toward those few exceptional values, as in the example of the preceding paragraph. The median in such distributions will typically fall between the mode and the mean. This relationship between the values of the mode, median, and mean for symmetrical and skewed distributions is summarized in Figure 1.

Knowing only the values of the mean and the median of a data distribution, we can generally guess the shape of the distribution. If these measures of

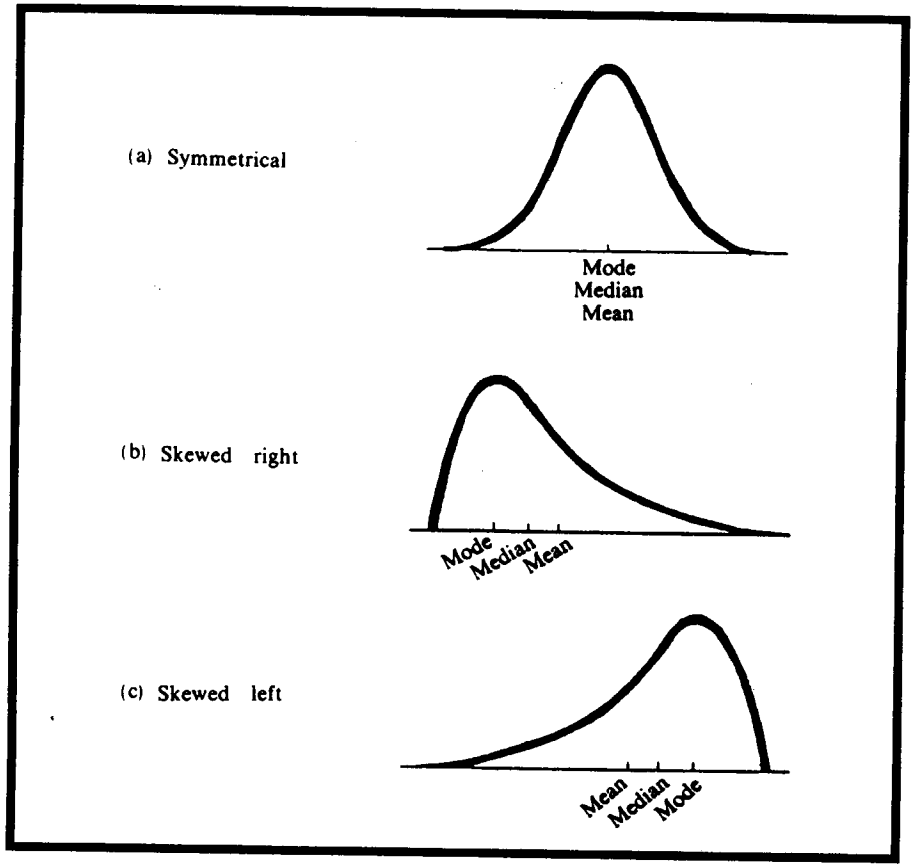


Figure 1 A comparison of the mode, median, and mean for distributions differing in shape.

central tendency are approximately equal, we know that the distribution is probably symmetrical in form. If the mean is *less* than the median, we know that the distribution is skewed to the *left*; i.e., a few low scores have disproportionately affected the mean. When the mean is *higher* than the median, the distribution is skewed to the *right*, the mean being dragged in that direction by a relatively few high scores.

For descriptive purposes, then, we would do well to report the mode, median, and mean when trying to characterize a distribution of data, for each in their own way is a measure of central tendency, though reflecting different features of the data. However, when we are dealing with roughly symmetrical distributions, in which case the mode, median, and mean are more or less

equal, and we are interested in making an inference about the central tendency of the population from which our sample observations have been drawn, we should rely on the mean, for the mean is known to be a more *efficient* estimate, in that for repeated samples of a given size the mean will show less fluctuation in value than either the mode or the median. This should not be surprising since the mean takes into account the information inherent in every single observation, whereas the mode and median do not take advantage of the numerical values of every single data point, wasting information as it were.

5. The Weighted Mean

Typically, every observation in a data collection has equal importance to us. There are situations, however, in which a data collection is based on a set of objects that differ in their *importance* for one reason or another. For example, a student's test scores are not equally important—quiz scores are less important than a midterm score, which in turn is less important than a final exam score. Similarly, a marketer's brand share in a large sales territory is more important than its share in a small territory.

In such situations a straight-forward mean might be misleading. To compensate for the varying importance of the observations, we can calculate a *weighted mean*—more precisely, a *mean based on weighted observations*—in which each observation is multiplied by its importance weight, and then dividing the sum of these weighted observations by the sum of the weights. In symbols, we have

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i} \tag{3}$$

where $\bar{x}_w$ is the weighted mean, w_i is the importance weight associated with an individual observation x_i, $\sum w_i x_i$ is the sum of the products of the observations multiplied by their respective weights, $\sum w_i$ is the sum of the weights, and the subscript i is implicitly understood to vary from 1 to n, the number of observations in the data collection.

Consider, for example, the results of a poll that finds a particular referendum is favored by 44%, 56%, 64%, and 51% of the prospective voters in a city's North, South, East, and West voting districts, respectively. The ordinary mean of these values is 53.8%. If, however, we take into account the unequal importance of the four voting districts based on the sizes of their voter bases—2,500, 1,000, 400, and 2,000 in the North, South, East, and West

Chapter 5

Variation

1. Introduction

A measure of central tendency is not enough to summarize a set of observations. To more fully describe a data distribution we also need a summary measure of the *variation* or *dispersion* of the observed values; i.e., the extent to which the observations differ among themselves in value.

The need for such a measure of variation is evident from the distributions shown in Figure 1. The distributions shown in part *a* of the figure have the same central tendency, or location, but differ markedly in their variation; those in part *b* differ in central tendency, but have the same variation; while those shown in part *c* differ both in variation and central tendency. So just as we have the mean as a measure of central tendency, we need a summary measure of the variation of a distribution of data.

2. The Range

Perhaps the simplest and at the same time the crudest measure of variation is the *range*, the difference between the highest and lowest observed value in a collection of data.

Consider, for example, the following set of ten scores:

7 10 12 15 16 16 19 20 25 29

The range for this set of data is 22, the difference between the extreme values of 29 and 7. We could summarize this collection of observations by stating that it has a mean of 16.9 and a range of 22, extending from 7 to 29; a description which is much more comprehensive than reporting the mean alone.

While the range is an easy measure to determine, easily understood, and a seemingly satisfactory measure of the variation of a set of scores, it suffers from the same weakness as the median; namely, it does not take into account

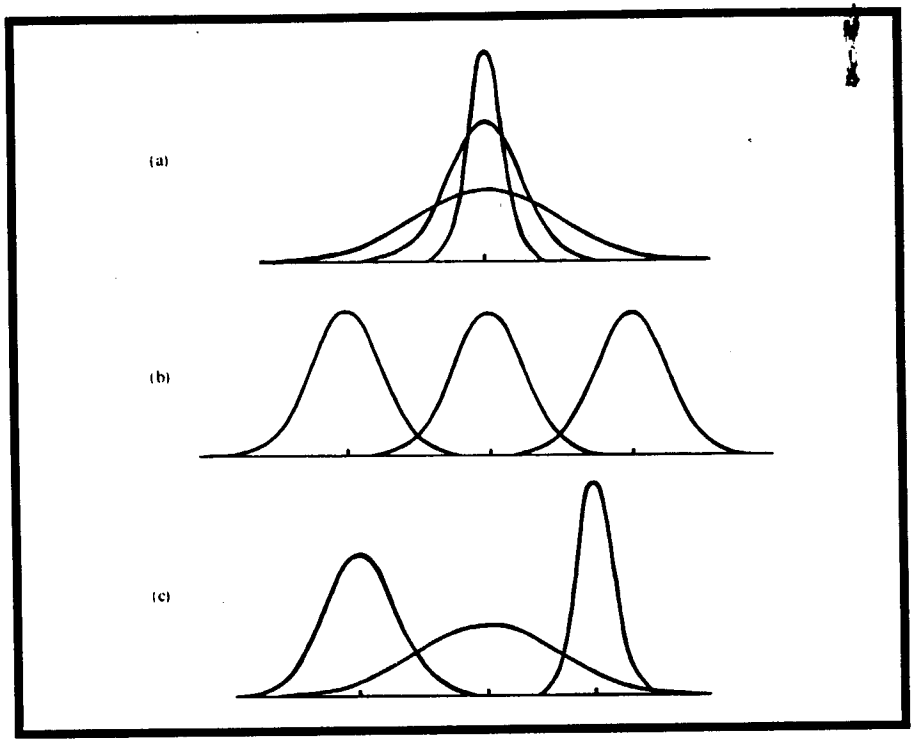


Figure 1 Distributions which have (a) the same central tendency but different variation, (b) different central tendency but the same variation, and (c) different central tendency and variation.

the numerical value of each and every observation. It is based entirely on two values, the highest and the lowest. The shortcoming of the range as a comprehensive measure of variation will be evident from a comparison of the above distribution with the following one:

7 14 15 15 16 16 17 18 22 29

This distribution has exactly the same range as that in the above paragraph, but clearly this set of values, taken as a whole, varies less about the mean than the previous example. Yet, the range, 22 in both instances, does not reflect this facet of the data.

Because of this insensitivity of the range to the "internal" variation of a distribution of data, it should be used only as the roughest measure of variability, primarily to provide information about the values beyond which no observations fall. This, in itself, is very useful information, whatever other limitations the range may have.

3. Mean Absolute Deviation

The greater the variation of a set of scores, the greater will be the deviations of the values from the mean. Consequently, a logical measure of variation would be the average value of those deviations. However, it will be recalled that the mean is the center of gravity, or balance point, of a set of observations, in that the sum of the distances or deviations from the mean of those values falling above it equals the sum of the deviations from the mean of those values falling below it. Since the deviations from the mean, $x_i - \bar{x}$, always sum to zero, their average value will also always equal zero, leaving us no information about the sizes of the individual deviations.

One way to get around the fact that the deviations from the mean sum to zero, is to strip the deviations of their positive and negative signs, and then take the arithmetic mean of these *absolute values* of the deviations. Such a measure of variation is known as the average absolute deviation or *mean absolute deviation* (*M.A.D.*). In symbols, we have

$$M.A.D. = \frac{\sum|x_i - \bar{x}|}{n} \tag{1}$$

where $|x_i - \bar{x}|$ signifies the absolute value of a deviation from the mean, and the summation is across the $i = 1, 2, \ldots, n$ observations in the data set.

As a simple illustration of the calculations of the mean absolute deviation, consider the following set of five scores that have a mean value of 9:

5 6 9 11 14

Their respective deviations from the mean are:

$-4 \quad -3 \quad 0 \quad +2 \quad +5$

As expected, the sum of the positive deviations from the mean ($+7$) balances the sum of the negative deviations from the mean (-7). However, in calculating the mean absolute deviation we disregard the positive and negative signs of the deviations, averaging instead their absolute values. Summing 4, 3, 0, 2, and 5 yields a total of 14, which when divided by $n = 5$, the number of observations in the data set, results in a mean absolute deviation of 2.8.

The logic of the mean absolute deviation as a measure of variation cannot be faulted, but how do we interpret it? While we can easily visualize the meaning of the range of a given distribution of data, how do we visualize the meaning of an average deviation. There is the further complication that, due to its reliance on absolute values, it cannot be used in certain mathematical operations necessary for the development of more advanced statistical techniques.

Nonetheless, the rationale behind the average deviation as a measure of variation is sound enough, and in the following two sections we will see how a minor variation on its theme will provide us with two much more useful measures of dispersion, measures that we will use throughout our subsequent study of statistical methods.

4. The Variance

An alternative to dealing with the absolute values of the deviations from the mean is to *square* each deviation, thereby yielding all positive quantities. The *mean* of these squared deviations could then be determined. Rather than having an average deviation we would have an *average squared deviation*. This very important measure of variation is called the *variance*, and for a *population* of observations we can define it concisely as follows:

$$\textit{Population Variance} = \frac{\textit{Sum of the squared deviations from the mean}}{\textit{Number of observations}}$$

In symbolic form, a population variation is denoted with the square of the lower case Greek letter *sigma*, σ^2, and can be read "sigma squared" or "the variance of variable x." In notational form, the above definition of a population variance becomes

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N} \tag{2}$$

where σ^2 is the variance, $x_i - \mu$ is the deviation of a given observation from the population mean, N is the number of observations in the data set, and the summation Σ is across the $i = 1, 2, \ldots, N$ observations. We could subscript the variance notation as follows, σ_x^2, to signify that we are referring to the variance of variable x, but that is understood in the absence of a subscript. If the variance refers to other than a variable x then it will be subscripted.

The above formula is appropriate for defining the variance of a population of observations. However, if we apply it to a *sample* of observations in an attempt to estimate the variance of the parent population from which the sample was drawn, we will discover that it is a *biased* estimate—it tends to *underestimate* the population variance. That is, if repeated samples were drawn from the population, and the variance calculated for each according to the above formula using the sample mean $\bar{x}$ instead of the population mean μ, the average of these variances would be somewhat lower than the true value of the population variance, were we able to measure every single member of it.

An adequate adjustment to the formula to avoid this bias is to divide the sum of squared deviations not by the number of observations in the sample, but by *one less* than the number of observations. We then have an *unbiased estimate of a population variance based on a sample of observations*, and we define it as follows:

$$\textit{Sample Variance} = \frac{\textit{Sum of the squared deviations from the sample mean}}{\textit{Number of observations less one}}$$

The sample estimate of a population variance is denoted s^2, and is defined symbolically as

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} \tag{3}$$

where the summation is across all n observations in the sample. Notice that we have used n to designate the size of a sample, and N for the size of a population. For formula (3) to yield an unbiased estimate of a population variance, we must assume that the population is infinite in size, or for practical purposes that N is much larger than n—say, at least fifty times as large.

So, when we are interested in estimating the variance of a population of values based on a random sample from that population, we will use formula (3). If we are merely interested in describing a body of data that we define as our population of interest, with no intent of generalizing to a larger parent universe, then we will use formula (2).

It should be apparent that as the size of our sample of data increases, the effect of dividing by the number of observations as opposed to one less than the number of observations becomes slight. Still, for theoretical purposes, the distinction between the two variances should be kept in mind.

5. The Standard Deviation

Since the variance is in units of measurement that are squared, as well as for other reasons that will gradually become apparent, it is convenient to take the *square root* of the variance and define a quantity known as the *standard deviation*:

$$\textit{Standard Deviation} = \sqrt{\textit{Variance}}$$

If we are interested in the *standard deviation of a population* of observations, we

will take the square root of the *population variance*. If, on the other hand, we are interested in *estimating* the standard deviation of a population we will take the square root of the *sample variance*, the unbiased estimate of that popula tion parameter.

In symbols, the *standard deviation of a population of values* is given by

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}} \tag{4}$$

In turn, the *unbiased sample estimate of the population standard deviation* i given by

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} \tag{5}$$

The standard deviation formulas (4) and (5) will be recognized as the squar roots of the variance formulas (2) and (3), respectively, as they should b definition.

Since the standard deviation equals the "square root" of the "mean" c the "squared deviations" it is also known as the *root mean square* value of data collection, or simply as the *RMS* value. This terminology is especiall common in the engineering fields and physical sciences.

As an example of the calculation of the standard deviation, consider th observations

12.1 15.4 14.1 14.4

which represent the widths in centimeters of four skulls found in an archeolog cal dig. Subtracting their mean value of 14.0 from each, squaring the resultar deviations, summing them, dividing by $n-1=3$, and then taking the squa root, we have

$$s = \sqrt{\frac{(12.1-14.0)^2+(15.4-14.0)^2+(14.1-14.0)^2+(14.4-14.0)^2}{4-1}}$$

$$= \sqrt{\frac{5.74}{3}} = 1.38 \text{ cm}$$

It should be noted that the standard deviation is in units corresponding t those of the variable we are measuring, centimeters in this instance.

Assuming the four studied skulls were a random sample of a larger population (which is probably not tenable), then the standard deviation of $s = 1.38$ would be an unbiased estimate of the variation of the larger population of skulls. If we had wanted to treat the four skulls as a population in itself, then we would divide by $N = 4$ in the above expression, instead of by $n - 1 = 3$. So doing, we obtain a standard deviation of $\sigma = 1.20$ cm for the set of four skulls, when treated as a population.

When data is arranged in a frequency distribution consisting of intervals of values of the variable in question, the variance and standard deviation of the original ungrouped data can be *approximated* by taking the midpoint of each interval as representative of the values falling within them, and then weighting their deviations from the mean by their corresponding frequency of occurrence. However, as pointed out in the discussion of the calculation of an approximate mean from grouped data, the procedure is not recommended, since modern computing facilities make it easy enough to obtain exact values from the original ungrouped data.

6. Variation of the Normal Distribution

While both the variance and the standard deviation have wide applications in statistical analysis, we will concentrate initially on the standard deviation. The applications of the variance will be encountered in later chapters. Though we now understand the simple manner in which the standard deviation is calculated, so far we know little of how to interpret it. What does a standard deviation of 1.2 mean, or one of 23, or 955, or whatever else we might calculate from a set of data. At the very least we can surmise that the greater the value of the standard deviation the greater the variation of scores about the mean. This follows from the fact that the standard deviation is based on the *deviations* of the observations from the mean. For example, if two distributions each have a mean of 100, but one has a standard deviation of 5 while the other has a standard deviation of 12, we can safely conclude that the distribution with the larger standard deviation has a wider dispersion of scores.

We can be more specific in our interpretation of a standard deviation provided our distribution of data is of a well-defined form. For our purposes we will be primarily interested in the *normal distribution*, the characteristic bell-shaped curve that we have been alerted to from the very start. Although the curve representing the normal distribution is defined by a precise mathematical equation, we can do without it for our purposes. Rather, we will recognize the normal distribution by its graphical representation, as shown in Figure 2, and by its special characteristics vis-a-vis its mean and standard deviation.

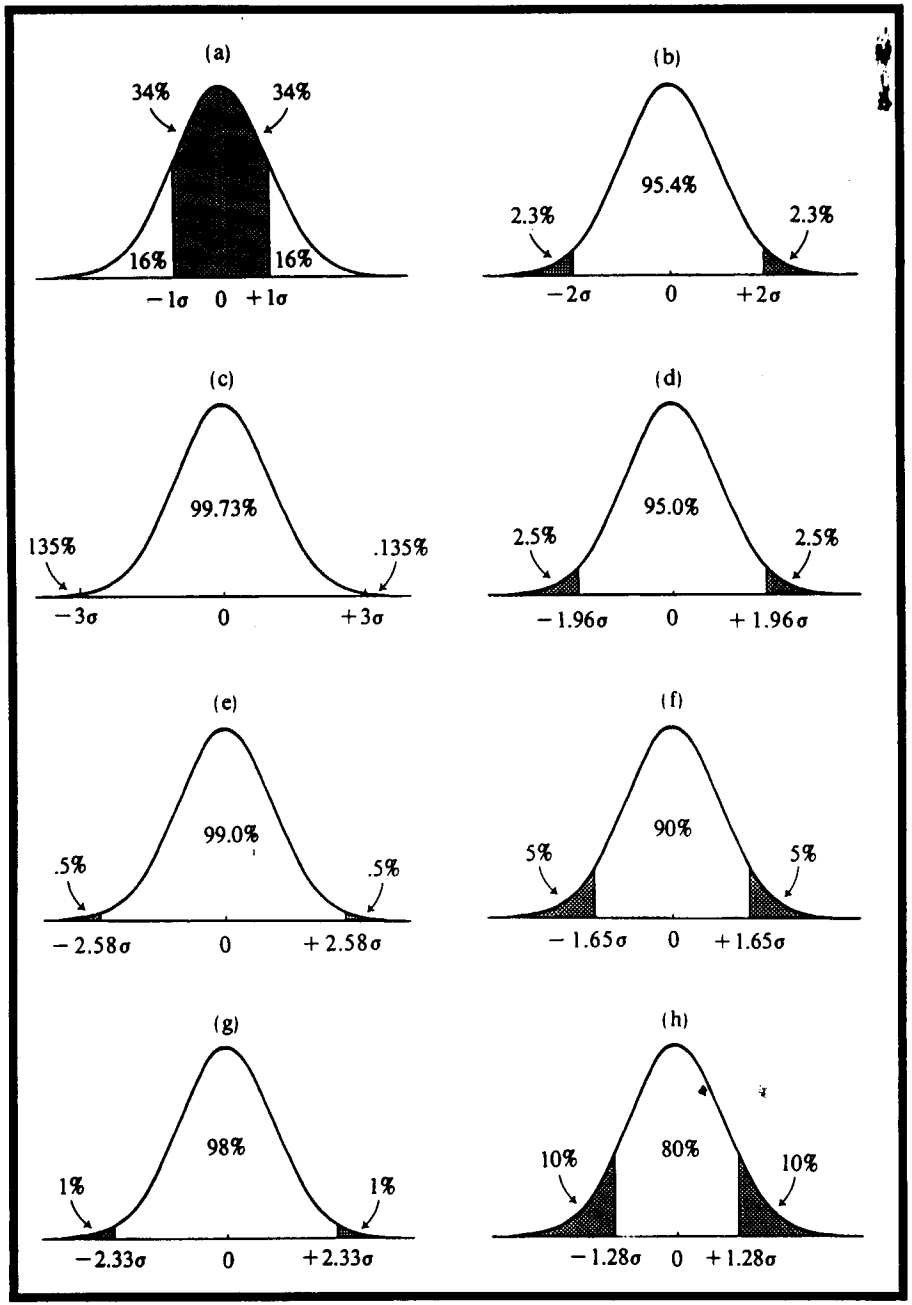


Figure 2 Areas under the normal curve for various standard deviations from the mean.

Chapter 9

Hypothesis Testing

1. Introduction

We have seen that one approach to the problem of statistical inference is the *direct* estimation of population parameters from sample observations—either as point estimates or interval estimates. In this section we will consider an alternative approach, one that is *indirect* in nature; i.e., *hypothesis testing*. Rather than using our sample observations to derive statistics which are approximations of the population parameter in question, we will use our sample statistics to support or discredit *a priori* hypotheses, or speculations, about the true value of the population parameter.

The hypotheses about the population parameters that we wish to test can be based either on prior observations or on theoretical grounds. Whatever the basis for the hypothesis, our sample observations will be used to test the likelihood or tenability of its being true. If it is found to be untenable, from a probability point of view, then we are forced to believe in an alternative hypothesis. But just as we cannot be 100 percent sure with regard to our interval estimates of a parameter, neither can we be absolutely certain of our conclusions with regard to the truth or falsity of our tested hypotheses. We can be 90 percent sure, 95 percent sure, 99 percent sure, and so on, but never 100 percent sure. With this in mind let us begin our discussion of hypothesis testing by considering the various types of hypotheses in which we might be interested.

2. Types of Hypotheses Sets

The most common approach to hypothesis testing is to establish a set of two mutually exclusive and exhaustive hypotheses about the true value of the parameter in question. Then, our sample statistics will be used to support one or the other of the hypothesized alternatives.

Exhaustive hypotheses sets. A researcher might be interested in testing the hypothesis that the mean weight loss from a particular three-week diet is 10 pounds. We designate this *working or "null" hypothesis as* H_0, and write

$$H_0: \mu = 10 \text{ pounds}$$

The alternative to this hypothesis, which we designate H_1, is that the mean weight loss is *not* equal to 10 pounds, and we write

$$H_1: \mu \neq 10 \text{ pounds}$$

where $\neq$ is the symbol for "is not equal to." These two hypotheses, then cover all possibilities, for either the true weight loss due to the diet *is* equal to 10 pounds (H_0) or it *is not* (H_1), and therefore can be referred to as a *exhaustive* hypotheses set.

To be perfectly general, we can again let θ stand for any particular parameter we wish—whether it be a mean, a difference between two means, a mean difference of paired measures, a proportion, a difference between two proportions, etc.—and then our hypotheses set can be expressed

$$H_0: \theta = a$$

$$H_1: \theta \neq a$$

where a is some specified value.

Truncated hypotheses sets. The hypotheses set outlined above is not the only kind in which we might be interested. There are some situations in which the hypotheses set

$$H_0: \theta = a$$

$$H_1: \theta < a$$

is more meaningful, where $<$ is the symbol for "is less than." That is, if the hypothesis $\theta = a$ is not true then the only alternative is that $\theta < a$. But here one might object that while these two alternatives are mutually exclusive, they are not exhaustive, since we have not admitted the possibility $\theta > a$, that θ is greater than a. And this is a crucial point, for when using such a *truncated curtailed* or *one-sided* hypotheses set we must be absolutely certain, usually on logical grounds, that the third and omitted possibility has a zero probability of occurrence. We cannot use an exhaustive hypotheses set such as

$$H_0: \theta \geq a$$

$$H_1: \theta < a$$

since *we have no idea what the sampling distribution for* H_0: $\theta \geq a$ *looks like if we admit the possibility that* $\theta > a$. And it is the H_0 sampling distribution upon which the theory of our test rests.

We might use a truncated or one-sided hypotheses set, for example, when evaluating the breakage level of a product shipped with additional packing material. If the breakage rate is not the same as that which occurred with the old packing technique (H_0: $\mu = a$), then it must be lower (H_1: $\mu < a$). On purely logical grounds we can rule out the possibility that the breakage rate could be higher.

Of course there is another truncated hypotheses set possible, which is simply a variation of the preceding one, namely

$$H_0: \theta = a$$

$$H_1: \theta > a$$

In this type of situation, the alternative to $\theta = a$ is $\theta > a$, and we rule out the possibility that $\theta < a$. For example, if we revise a billboard advertisement by increasing the size of the print with which the brand name appears, we can safely assume that the percentage of consumers recalling the brand name will either be unchanged (H_0: $\theta = a$) or it will be higher (H_1: $\theta > a$), where a is a level of recall for the original version of the billboard.

Whether we use an exhaustive or truncated hypotheses set will be dictated by the nature of the particular problem; that is, which alternatives are possible. The importance of the distinction between these types of hypotheses sets will become more apparent in the following sections where we will consider the techniques for choosing among the alternative hypotheses. Which *type of hypotheses set* we choose is a logical decision, which *hypothesis in the set* we eventually believe is a statistical question.

3. Test of a Mean

Suppose a municipality wants to test the hypothesis that a particular model school bus averages 20 miles per gallon of gasoline. The hypotheses set is

$$H_0: \mu = 20 \ mpg$$

$$H_1: \mu \neq 20 \ mpg$$

Our task is to test the credibility of H_0 vs. H_1.

Since we cannot realistically test every single bus that came off the assembly line, we must test a sample of buses of the particular model of

interest. The mean gas mileage $\bar{x}$ of this sample would then be used to decide the credibility of the hypothesis that $\mu = 20$. What if we obtained a mean value of $\bar{x} = 21$, would the hypothesis of $\mu = 20$ be credible? What if we obtained a sample mean of $\bar{x} = 22$? Or $\bar{x} = 16$?

By now we realize that to answer these questions we need additional information—namely, characteristics of the sampling distribution of the mean $\bar{x}$; which, as we have learned, is dependent on the size of the sample n, and the shape and standard deviation σ of the population of values which was sampled. Knowing the population standard deviation σ and the size of the sample n, we can easily determine the standard error of the sampling distribution of the mean using

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Now if we further know that the sampled population is normally distributed, we will also know that the sampling distribution of $\bar{x}$ is normally distributed. Therefore, we can express our observed sample mean $\bar{x}$ as a deviation from the hypothesized mean μ in standard error units; namely

$$\boxed{z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}} \qquad (1)$$

However, if we do not know the value of the population standard deviation σ, which we need to calculate the standard error of the mean $\sigma_{\bar{x}}$, then we must use the sample standard deviation s as a substitute estimate. Our observed sample mean $\bar{x}$ can then be expressed as a deviation from the hypothesized population mean μ in estimated standard error units; namely

$$z \doteq \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

where $\doteq$ is the symbol for "is approximately equal to." So long as our sample size n is sufficiently large, say greater than thirty, we can be confident that the above z will be very nearly normally distributed, and approximately equal to the value which would be obtained if we knew the value of $\sigma_{\bar{x}}$.

Returning, then, to our example, let us assume that we test $n = 100$ buses and obtain a mean mileage figure of $\bar{x} = 19.1$ *mpg*. Let us also assume, for the time being, that we know the population standard deviation to be $\sigma = 3$ *mpg*.

We can then calculate the standard error of the mean as

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{100}} = .3 \text{ mpg}$$

Next, we can express the observed mean $\bar{x} = 19.1$ as a deviation from the hypothesized value $\mu = 20.0$, expressing it as a normal deviate

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{19.1 - 20.0}{.3} = -3.00$$

In other words, our sample mean of $\bar{x} = 19.1$ deviates from the hypothesized value of $\mu = 20.0$ by three standard errors of the mean. This deviation is shown graphically in Figure 1.

The question we must ask is *whether it is likely that we would get such a deviant sample mean* $\bar{x}$ *if in fact the true population value were* $\mu = 20.0$. If it is too unlikely, then we will tend to disbelieve the hypothesis. But what is our

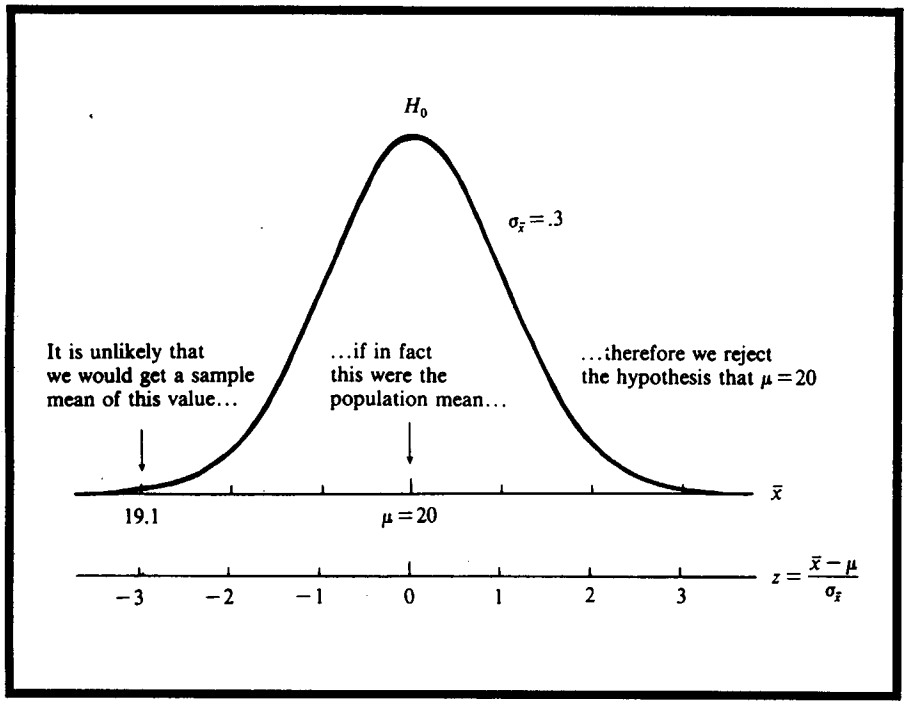


Figure 1 The deviation of a sample mean $\bar{x} = 19.1$ from the hypothesized mean $\mu = 20$.

standard for "unlikely." By convention, a probability less than .05 (1 chance in 20), or less than .01 (1 chance in 100), is usually considered an unlikely event. In other situations we may want to establish either a more stringent or a less stringent criterion for what we consider an unlikely event under the condition that H_0 is true, but in most instances .05 and .01 are accepted standards.

Significance levels. The particular probability that we use as our criterion for an unlikely outcome, supposing that our working hypothesis H_0 is true, is referred to as the *significance level* of our statistical test and is designated with the Greek letter *alpha*, α. It is essential that we stipulate α *prior to conducting our statistical test*, for to choose one after the data has been analyzed would be less than objective. In many research studies "p values" are reported, signifying the after-the-fact probability of obtaining the test statistic in question, given that the null hypothesis were true. While such information is useful, it is not a substitute for prespecifying α, our criterion for rejecting the null hypothesis. For if we do not specify beforehand our standard for an unlikely event, then we will be tempted to decide after the fact, which is not a "well-defined" procedure. It would be like bragging about the number of fish we caught at a particular fishing hole, without first having specified our criterion for what would be a good catch.

Three common significance levels ($\alpha = .01$, $\alpha = .05$, and $\alpha = .10$) are shown in Figure 2. The extreme values of the normal distribution corresponding to these probabilities are designated by the shaded areas, and are referred to as the *critical regions*, or *rejection regions*, for our test of the hypothesis H_0.

It will be further noted in Figure 2 that the probability corresponding to the various significance levels can either be split between the two tails of the distribution, or placed all in one tail. If the probability of an unlikely outcome is divided between the two tails of the distribution, as in the left parts of Figure 2, we refer to our statistical test as a *two-tailed* or *two-sided* test. If, however, we are dealing with a truncated hypotheses set in which we assume that any departure from the H_0 hypothesis must be in a particular direction, then we will choose to put the entire probability of an unlikely outcome in one tail or the other of the distribution, as shown in parts *b*, *d*, and *f* of Figure 2, and we call this type of test a *one-tailed* or *one-sided* test.

Compare, for example, the value of z needed to reach the critical region for a significance level of $\alpha = .05$ when we are dealing with a two-tailed vs. a one-tailed test. For a two-tailed test, a deviation from the mean beyond $z = 1.96$ standard errors will occur by chance 5% of the time. For a one-tailed test, a deviation of only $z = 1.65$ standard errors from the mean, in a given direction, will occur by chance 5% of the time. For a significance level $\alpha = .01$, a value of z greater than 2.58 vs. 2.33 is needed to reach the rejection region for a two-tailed vs. one-tailed test, respectively, as shown in parts *e* and *f* of Figure 2. Of course, the z's reported above are their absolute values.

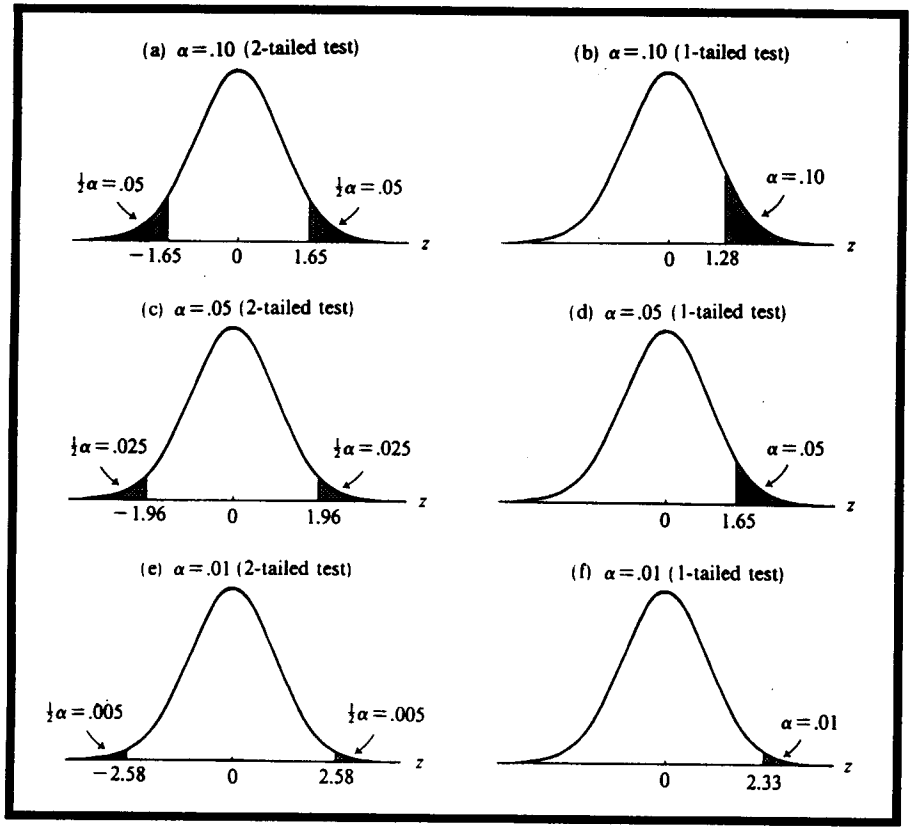


Figure 2 The most commonly used significance levels α, and their corresponding critical regions (when σ is known or sample size is large).

In our bus gas mileage example, assuming we had specified beforehand an $\alpha = .05$ significance level for judging the hypotheses H_0: $\mu = 20$, our observed sample mean $\bar{x} = 19.1$, which deviates from $\mu = 20$ by 3 standard errors, can be considered an unlikely outcome if indeed the hypothesis $\mu = 20$ were true. Rather than believe that we have simply witnessed a *chance* deviation from $\mu = 20$ due to sampling error, we prefer to disbelieve the hypothesis that $\mu = 20$. In "rejecting" the hypothesis H_0: $\mu = 20.0$, we are "accepting" the alternative hypothesis H_1: $\mu \neq 20.0$. We use the terminology "reject" and "accept" very guardedly, for we recognize that our conclusions are based on probabilities, therefore "believe" and "disbelieve" might be more appropriate terminology, suggesting that we might be wrong.

If we had not been provided with the population standard deviation σ, from which we calculated $\sigma_{\bar{x}}$, we would have had to use the sample standard

deviation s in its place, and estimate the standard error of the mean as $s_{\bar{x}} = s/\sqrt{n}$. Suppose, for example, that the sample standard deviation were $s = 3.2$. The estimated standard error of the mean would then be $s_{\bar{x}} = s/\sqrt{n} = 3.2/\sqrt{100} = .32$, and we would have

$$z = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{19.1 - 20.0}{.32} = -2.81$$

However, since we have estimated $\sigma_{\bar{x}}$ with $s_{\bar{x}}$, this test and the corresponding significance levels will only be *approximate* in nature. Recognizing this, our conclusion is the same as above. Since $z = -2.81$ is beyond the value $z = -1.96$ needed to enter the critical region for $\alpha = .05$, we must reject the hypothesis that $\mu = 20$. That is, rather than believe our sample mean $\bar{x} = 19.1$ was a "fluke" from a population with $\mu = 20$, we prefer, rather, to disbelieve that $\mu = 20$. Consequently, we place more credence in the alternative hypothesis that $\mu \neq 20$. We do not know the exact value of μ but we are fairly confident that it is not 20.

Small sample technique. In situations involving a small sample from a *normal population* with an unknown standard deviation σ, we can use Student's t variable

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \tag{2}$$

to *exactly* test hypotheses about population means. Associated with the t ratio are degrees of freedom $df = n - 1$, as discussed in the Parameter Estimation chapter.

Suppose, in the preceding example, we had tested only $n = 10$ buses and obtained a mean mileage figure of $\bar{x} = 21.2$ with a standard deviation of $s = 3.4$. The estimated standard error of the mean is then $s_{\bar{x}} = 3.4/\sqrt{10} = 1.08$. We then calculate t as

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{21.2 - 20.0}{1.08} = 1.11$$

with degrees of freedom $df = 9$. We look in Appendix Table IV and find that for $df = 9$, a t of 2.26 is required before we enter the area of an unlikely event for a significance level of $\alpha = .05$. In other words, our value of $t = 1.11$ could well have happened *by chance alone* if in fact the hypothesis $\mu = 20.0$ were true. Figure 3 shows the relative location of $t = 1.11$. Consequently, we *will not*

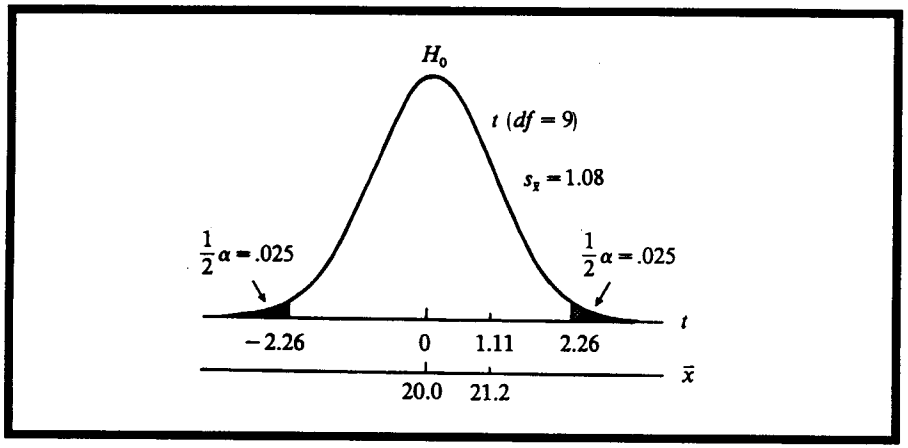


Figure 3 The relative location of $t = 1.11$ ($df = 9$).

reject that hypothesis; we do not have enough evidence to discredit it. We never "accept" H_0, only refuse to reject it.

The preceding examples demonstrate the probabilistic foundations of our conclusions with regard to the believability of the alternative hypotheses that we are testing. As such, we cannot be 100 percent certain that we have reached the correct conclusion. In the following section we will consider the types of errors we are liable to make.

4. Type I and Type II Errors

In the American system of justice an accused person is assumed to be innocent until proven guilty. We could express these alternatives as hypotheses:

H_0: Innocent

H_1: Not innocent (i.e., guilty)

Now just as with the earlier statistical hypotheses, we as a jury might come to believe either of these alternatives. But since we base our conclusions only on sample evidence, we are liable to make mistakes in our judgments. For instance, we could conclude that the accused person is guilty when in fact they are innocent; or we could conclude that the person is innocent when in fact they are guilty. These two types of errors are shown in Table 1a, along with the two possible correct conclusions—acquitting an innocent person and convicting a guilty one.

In the case of deciding between statistical hypotheses, analogous types of errors can arise. If, for example, we have the hypotheses

$$H_0: \mu = a$$

$$H_1: \mu \neq a$$

where a is some specified value, we could *conclude that H_0 is false when it is in fact true*; or, on the other hand, we could *fail to reject H_0 when in fact it is false*. These two types of errors are referred to as *type I* and *type II errors* respectively, and are shown in Table 1 along with the two correct conclusions—not rejecting H_0 when it is true, and concluding that it is false when it is false.

The probability of committing a type I error—rejecting H_0 when it is true—is easy enough to determine. It is simply equal to the significance level α which we use as our criterion for judging whether our sample statistic deviates an unlikely amount from the hypothesized value. For example, if our hypothesis H_0: $\mu = a$ is actually true, there will be a certain number of instances in which our sample mean $\bar{x}$ deviates more than 1.96 standard errors from it; to be more specific, the probability of occurrence of such a happening is $\alpha = .05$. We will be less likely to commit a type I error if we have a more stringent significance level, say $\alpha = .01$, for then we are even more unlikely to obtain such a deviant result when in fact H_0 is true.

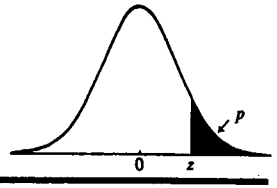
It is not such an easy matter to determine the probability of committing a type II error—i.e., concluding that H_0 is true when in fact it is false. The probability of making a type II error, which we designate with the Greek letter *beta*, β, depends on a number of factors, including (1) the *true* value of the parameter in question; (2) the significance level α we use to evaluate our working hypothesis H_0 and whether we use a one-tailed or two-tailed test; (3) the standard deviation σ of the sampled population; and (4) the size of our

Table 1 Outcomes of (a) a jury trial, and (b) a statistical test of an hypothesis H_0.

(a) Jury Trial			(b) Test of an hypothesis H_0		
Jury verdict	True state of affairs		Conclusion of statistical test	True state of affairs	
	Innocent	Guilty		H_0 is true	H_0 is false
Innocent	Correct judgment	Error	H_0 is not rejected	Correct conclusion	Type II error
Guilty	Error	Correct judgment	H_0 is rejected	Type I error	Correct conclusion

Table III Normal Distribution

The tabled entries represent the proportion p of area under the normal curve above the indicated values of z. (Example: .0694 or 6.94% of the area is above $z = 1.48$). For negative values of z, the tabled entries represent the area less than $-z$. (Example: .3015 or 30.15% of the area is beneath $z = -.52$).

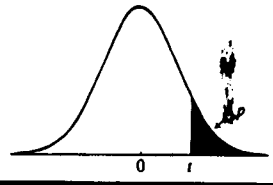


	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2297	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

Adapted with rounding from Table II of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural, and Medical Research*, 6th Edition, Longman Group, Ltd., London, 1974. (Previously published by Oliver & Boyd, Ltd., Edinburgh). Used with permission of the authors and publishers.

Table IV Student's t Distribution

For various degrees of freedom (df), the tabled entries represent the critical values of t above which a specified proportion p of the t distribution falls. (Example: For $df = 9$, a t of 2.262 is surpassed by .025 or 2.5% of the total distribution). By symmetry, negative values of t cut off equal areas p in the left tail of the distribution. Double the p for two-tailed probabilities.



	p (*one-tailed probabilities*)				
df	.10	.05	.025	.01	.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

Adapted from Table III of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultu and Medical Research*, 6th Edition, Longman Group, Ltd., London, 1974. (Previously publishe Oliver & Boyd, Ltd., Edinburgh). Used with permission of the authors and publishers.