CHAPTER 1. AN OVERVIEW

The great body of physical science, a great deal of the essential fact of financial science, and endless social and political problems are only accessible and only thinkable to those who have had a sound training in mathematical analysis, and the time may not be very remote when it will be understood that for complete initiation as an efficient citizen of one of the new great complex world-wide States that are now developing, it is as necessary to be able to compute, to think in averages and maxima and minima, as it is now to be able to read and write.

—H. G. Wells (*Mankind in the Making,* 1911)

1.1 Distinctive Features of the Approach to Sampling and Inference in This Volume

I. What Is Statistical Sampling?

If we want to learn things about some population, we could, in principle, interview or inspect each member of that population. But doing so would almost certainly be costly and time-consuming—perhaps impossibly so if there were constraints on when we needed the information and/or the population is large and/or there are many attributes of the population in which we are interested and/or when some of the population may be hard to reach. So in many contexts, from exit polls and public opinion surveys in politics to quality control checks on factory assembly lines, we commonly look only at a *sample* of cases and use that sample to draw inferences about the larger population. But how do we know how accurately the inferences we draw from our sample are vis-à-vis key characteristics of the population in which we are interested? Answering that (and related) questions is the domain of applied *statistical sampling theory* and the topic of this text.

However, there are several ways in which this volume is quite different from the approach to sampling and inference in the typical undergraduate statistics text.

First, while the concepts we use are broadly applicable in survey research and in other areas where sampling and inference play a key role, we illustrate our results largely with publicly available data drawn from elections with two-candidate or two-party competition, data on ballot initiatives, data on partisan voting patterns in legislatures, or data on voter political attitudes from surveys by reputable organizations. We think that students, regardless of their disciplinary background, will find these real-world examples of interest, and they allow us a more unified approach where

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
Do not copy, post, or distribute

students can see the same (types of) data analyzed from different perspectives. However, we also give some non-political examples where we think readers will have knowledge that will allow them to make intuitive sense of the statistical ideas we present. For example, we introduce an application of statistical sampling theory to link the frequency distribution of GPA scores (grade point average) to student class ranking.

Second, we provide the basic mathematical results underpinning sampling theory in a verbal form in terms of what we refer to as the ten "laws of statistical sampling." While calling them "laws" is distinctive to this volume, nonetheless, these ten laws can be derived as corollaries to fundamental mathematical results such as the "central limit theorem." We present these laws as simple sentences that we hope students will find relatively easy to remember. Relatedly, to further help students remember important ideas in sampling and inference, we offer mnemonics (memory shortcuts) such as the "teeter-totter principle" and the "DMV principle," the meanings of which will be explained later in the text. Central to our exposition is the idea of an "approximately normal distribution." Here, "normal" has a precise technical meaning that is different from our ordinary language use of that term. We specify what it means for a distribution to be "normal" at the beginning of the next chapter and then introduce the idea of approximate normality.

The ten laws we introduce are divided into three groups. The first seven of these laws apply only to sampling processes that involve random sampling. The first five of the seven are some of the most fundamental results in sampling theory. For example, we show that the accuracy of a poll in drawing inferences about the population mean—for example, the vote share for a given candidate in a two-candidate election—increases not with the sample size but with the square root of the sample size. The next two can be thought of as "notes of caution" regarding drawing inferences from samples. The last three deal with what happens when samples are not random draws from a (approximately) normal distribution, and we consider ways in which sample biases may be "corrected." A detailed discussion of the implications and applications of the ten laws is given in Chapter 3. At the end of that chapter, we consider reasons why real-world polling data may be flawed or biased

Each of these laws is phrased in ordinary English rather than in mathematical symbols. While some of the terms used in the statement of these laws—for example, "confidence interval," may be unfamiliar to readers, they will be explained later in the text. As previously noted, the first seven of these laws apply only to sampling processes that involve random sampling, and the first five of the seven are some of the most fundamental results in sampling theory. The next two can be thought of as notes of caution regarding drawing inferences from samples. The last three deal with what happens

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

when samples are not random. In Chapter 3, we discuss each of those laws in more detail—with some helpful examples for many of the laws.

A. Laws Involving Random Sampling

First Law of Statistical Sampling (Law of Large Numbers): As we increase the sample size, *n*, the distribution of sample means gets ever more tightly clustered around the actual population mean.

Second Law of Statistical Sampling (sampling distributions are normal): Regardless of what distribution we begin with or how far away it is from looking like a *normal distribution, the* distribution of sample means from that distribution is normally distributed.

***Third Law of Statistical Sampling* (*inverse square law*):** The degree to which a random sample matches the characteristics of the population from which it is a sample increases in accuracy (reduces *standard error*) inversely with the square root of the size of sample *n*, not with *n* itself.

Fourth Law of Statistical Sampling: For a random sample, for a fixed standard deviation of the population from which the sample was drawn, we know that, as we increase the sample size, the closer, on average, we come to matching the population characteristics. But how close we come is *independent of the size of the population from which we are sampling.*

Fifth Law of Statistical Sampling—Combining Units Reduces Variance (CURV): Ceteris paribus when we aggregate data into larger groupings, in general, we reduce the expected variance.

B. Laws Giving Reasons for Uncertainty

Sixth Law of Statistical Sampling: Close divisions in a survey make estimating majority attitudes or preferences subject to greater uncertainty.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Seventh Law of Statistical Sampling: If you have to multiply probabilities to estimate outcomes, then the error rate generally can be expected to rise.

C. Laws Involving the Consequences of Non-Random Sampling

Eighth Law of Statistical Sampling (the BS = BS principle): Biased (non-random) samples give you bad (unreliable) results.

Ninth Law of Statistical Sampling (the reweighting escape clause): Even if you have a *non-random sample,* you might (under certain circumstances) be able to *reweight the cases* to still be able to draw reliable inferences from your sample.

Tenth Law of Statistical Sampling: If a sample is not random, then the larger the sample size the more we will misstate the true *confidence interval* around our estimate(s) by overestimating the reliability of those estimates.

These ten laws are central to the exposition of this book. They allow us to answer what are arguably the six important questions about sampling (see discussion directly following). We do not try to "prove" these laws. In this non-technical introduction to sampling theory, we necessarily leave precise statements and proofs of the intuitions we state to much more advanced statistics texts that are calculus-based. In this volume, we do not require any mathematics more sophisticated than high school algebra. Instead, we may use EXCEL's random number generation tools to develop simulations that show how these laws work.

Third, in our more intuitive approach, we focus on the questions about polls, especially election polls, we want to use statistics to answer. In particular, we focus on the six question identified:

1. If we observe a particular (randomly sampled) outcome, say a vote share for one of two candidates, *What is the probability that we would have observed the same outcome or one even more extreme in the whole population?*
2. If we observe a particular (randomly sampled) outcome, say a vote share for one of two candidates, *what are the (95%, 99%, etc.)*

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

confidence intervals around that outcome? The *confidence interval* is the range in which we expect most (95%, 99%, etc.) values to lie.

3. If, in a two-candidate election, we observe a particular (randomly sampled) vote share for one of the candidates, *What is the likelihood that the given candidate will win the election?*
4. For a fixed specified sample proportion—for example, about expected votes—*how large must the sample size be before we can conclude with, say, 95% confidence, that we have identified the winner?*
5. For a specified sample size—for example, about expected votes—*How large must the sample proportion be before we can conclude with, say, 95% confidence, that we have identified the winner*?
6. If we have data on two distributions (say, election results in two elections in the same jurisdiction), *What is the likelihood that they are both random draws from the same population?*

We show how to use statistical tools to answer each question. While the statistical tools to answer these six questions are very similar, the questions are distinct, even if it is not readily apparent that they are distinct. For example, though they seem similarly worded, Question 2 and Question 3 are distinct questions. The 95% confidence interval around the sampled outcome (vote share for candidate A)—that is, the zone where we think 95% of the outcomes are most likely to lie—may include some outcomes where the candidate loses; nonetheless, as we later demonstrate, we might still be more than 95% confident about who the winner of the election might be. The distinction between Questions 2 and 3 is one of the features of statistical sampling that can be most confusing to the beginning student. In particular, an election might be labeled "too close to call" in the media, and yet a gambler could still have a very clear idea as to which way to bet and what odds to give. When the reader completes this book, s/he will understand how this seeming contradiction can be possible.

Several of the preceding questions explicitly or implicitly reference a *null hypothesis.* That is a hypothesis against which the observed outcome(s) is (are) to be compared. That null hypothesis will usually be the simplest one possible—that is, what we might call "nothing that interesting is happening"—for example, the observed value appears to be a random draw from a fair coin toss, or the two distributions we are comparing appear to be drawn from the same population, or the population mean is the same as the sample mean. In practice, of course, we may have expectations in advance that the null hypothesis is unlikely to be true—for example, playing dice with a professional gambler, we might think that the dice were

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

loaded to make the gambler's success more likely. Or if we performed an experiment in which we subjected subjects to some particular treatment, we might hope that the treatment worked the way we thought it should rather than have a zero effect. Our "real" expectation about what should happen is often referred to as the "research hypothesis," as distinguished from the "null hypothesis." But throughout this elementary textbook, it is the null hypothesis whose plausibility we evaluate. One way in which we assess plausibility is in terms of *statistical significance* (often referred to as *p value*, the likelihood that the null hypothesis is true given the data); another and even more important way to think about inference is in terms of the *confidence interval* around the sample estimate that identifies the zone of most likely outcomes. However, as we make clear later in the text, ultimately what we are concerned about is the *substantive* importance of the inferences we draw from the data, not merely (or even mostly) their statistical significance.

A fourth feature of this book that is distinctive is that we use EXCEL and EXCEL functions such as =NORMDIST, =BINOMDIST, and =TTEST to answer the six questions previously identified, along with tools provided in the DATA ANALYSIS package in EXCEL—which is a free add-in. Actual proofs of most of the results we give would require advanced math; instead, as noted earlier, we often offer simulation results that students can duplicate on their own using EXCEL's random number generation tools—whose operation we describe. Thus, students can see for themselves how sampling processes work. We view this as one of the most attractive features of this volume from a learning standpoint.

Even though there are more sophisticated statistical packages and programming languages such as R available, we chose to use EXCEL throughout this book for seven reasons that we regard as compelling.

First, EXCEL is readily available as part of Microsoft Office.

Second, EXCEL is commonly used in the business world. Thus, learning to better use EXCEL is likely to continue to be useful to students long after they have left college.

Third, EXCEL is easy to learn, and many students will already have had some exposure to it, so students can start on the statistics immediately without spending a lot of time learning to program.

Fourth, EXCEL allows you to answer key questions in sampling while staying "close to the raw data," which is always in front of you in the EXCEL spreadsheet. While the text does provide formulas for the situations where you do not have access to the raw data but only have overall parameters such as mean and standard deviation, even if you

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

don't like sushi, you should always prefer your data raw rather than having it pre-processed for you. There is no substitute for looking closely at the raw data and not just looking at overall parameters such as mean or standard deviation, since these can be highly misleading if the underlying data are not unimodal or approximately normal.

Fifth, EXCEL has a standard set of graphical tools to view data that are available in Chart Wizard, such as its histogram and scattergram features. While many programs allow for prettier and more complex graphics, using EXCEL's basic tools will give you a good feel for the data. EXCEL l also allows you to create cross cross-tabulation of your data.

Sixth, EXCEL has tools for random number generation that allow us to create simulations of various sampling processes. Simulations are a really important tool for understanding sampling processes.

Seventh, in both academic and business settings, even when EXCEL itself is not being used as the spreadsheet program, the open-source program that is used will be very similar to EXCEL, so making the transition will be easy.

However, although we explain how to use EXCEL whenever we consider a particular question that needs to be answered, this book is not intended as a general primer in EXCEL. The reader interested in such a primer should supplement our discussion with other texts, such as the first several chapters of Salkind and Frey (2016).

A final element of distinctiveness in this volume comes in our concluding chapter, where we provide a unifying perspective on all our results using the most important distribution in statistics, the normal distribution. We summarize how each of the answers to Questions 1 through 3 and Question 6 can be thought of as identifying particular segments of some specified normal distribution. We also point out how Questions 4 and 5 are answered by simulating results for a change in the predicted sampling outcomes from a normal distribution as we vary either the sample size (holding observed vote share fixed) or the observed vote share in a two-candidate contest (holding sample size fixed).

1.2 The Structure of This Book

Now we outline the structure of the book.

In the remainder of this chapter, after providing a standard notation used throughout the volume, we introduce basic concepts in statistics, such as

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

standard deviation, *standard score*, and *odds ratio*, which must be understood before we can proceed to a discussion of sampling and inference. Some readers may already be familiar with these concepts, but a refreshed memory is likely to be useful.

In the next chapter, Chapter 2, we continue our exposition of key concepts needed for answering the six questions that are at the heart of this volume. We begin with an important distinction between distributions that are the actual values that we have sampled and an *ideal distribution,* such as the *normal distribution*. We start with the *normal distribution*, which is central to statistics, and then turn to the *standardized normal distribution.* We link this latter concept to something likely to be familiar to readers, namely "grading on the curve." We also consider *cumulative (normal) distributions*. The latter is a concept that we can link to something with which any reader who has taken an SAT, GRE, or LSAT exam will be familiar—namely, *percentile rank.* Then we discuss other distributions that are what we call *approximately normal*—the *binomial distribution* and the t *distribution*—which are central to understanding sampling theory. In Chapter 2, we discuss the key properties of the *binomial distribution* and present a function in EXCEL that can be used to calculate binomial values. The binomial distribution is especially important in the context of elections, where we are interested in predicting the winner. In this chapter, we provide a simulation to demonstrate how the binomial distribution looks more and more like a normal distribution as sample size increases. And we briefly consider the t *distribution*, which is very similar to the normal distribution. We also very briefly introduce two metrics that can be used to characterize distributions: *skewness* and *kurtosis*. These concepts help us understand differences among unimodal distributions. Skewness (which can be proxied as the difference between the median and the mean of a distribution) can be used as a way to measure partisan bias in legislative outcomes—that is, the extent to which there has been partisan gerrymandering. Kurtosis, which is related to dispersion, has to do with how much of the distribution is in the middle and how much is in the tails. It has implications for understanding phenomena such as electoral tides.

Chapter 3 provides more details about the ten laws of statistical sampling. The reader who understands these laws will have a firm, intuitive grasp on the theoretical bases of applied statistical sampling, even though s/he will need to turn to other textbooks to learn how to design a sampling frame for public opinion surveys or how to develop good questionnaire wording. Most importantly, these abstract laws provide the theoretical underpinnings of the specific EXCEL formulas we give in the subsequent two chapters that show how to answer the six questions introduced earlier in this chapter. To repeat: These six questions are at the heart of this

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

volume. As also noted previously, because proofs of the laws require mathematics of an advanced kind, we show readers how some of these laws work by providing simulations using EXCEL's random number generation tools. We also mention in Chapter 3 what we call the *teeter-totter principle,* namely that the higher in absolute value is the t score, the lower the probability that the null hypothesis is true. We refer to this principle as a "secret of the statistical universe." It is a fundamental insight that, in our view, is not as much emphasized as it should be in statistics textbooks.

Chapter 4 is about the first five of the six key questions we introduced in Chapter 1. Most importantly, it shows how each of these questions can be answered using tools in EXCEL and it provides examples of each law's applicability to polling and other sampling data. The reader who understands these EXCEL formulas and who has been exposed to simple simulations will have been given a basic grounding in the use of computer-based tools in statistics that will facilitate learning to handle statistical analyses via more sophisticated statistics packages or via programming languages such as R.

Chapter 5 shows how to extend the sampling ideas of the previous chapter to make comparisons between distributions to answer our sixth key question—namely, "If we have two samples, how likely are they to have come from the same population?" In this chapter, we introduce the =TTEST function in EXCEL and we explain the *DMV principle*. We show that the likelihood of getting statistically significant results is a function of the difference (D) between the means of the two sample distributions, how many (M) cases there are in each distribution, and the *variance* (squared *standard deviation*) of each of the two samples. The DMV principle can also be regarded as another of the secrets of the statistical universe. In Chapter 5, we also explain the difference between the usual difference of means test and one that is used extensively in studies of before-and-after experiments, where we compare not just the overall means of some variable before and after but what happened to *each* of the individual subjects with respect to that variable. In this chapter, we also consider how the results of a difference of means test may mislead about causality when the observed difference is due to confounding factors such as *selection bias*.

Chapter 6 provides a unifying and simplifying perspective on the various questions we have used statistical sampling to solve. We argue that the combination of theory and empirics in this book and our use of simulations provide a foundation for the understanding of statistical issues in the real world, especially in the context of elections and surveys. We also point out how our emphasis on "unpacking" the formulas used in basic sampling theory can be thought of as identifying the size and/or location of particular segments of a normal distribution with a specified mean and variance, and we show how some of the answers to our six key questions can be viewed

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

as generalizations of the concept of standard score, while others show the comparative statics of what happens when we fix all but one parameter in a formula and allow that parameter (e.g., "*n*," sample size) to vary. In an appendix to Chapter 6, we show how the ideas in the book generalize to perhaps the single most common application of statistics in the social sciences, namely the use of regression models for purposes of prediction and explanation. In the references, we provide suggestions for useful background reading as well as sources of more advanced material for students interested in further pursuing their statistical interests as well as the list of references cited within the text.

1.3 Notation

We use a notation throughout the book that is common to most statistical textbooks, with one exception. We use p_{NH} as the proportion we specify as our null hypothesis to distinguish it from other uses of "*p*" in statistics, though sometimes we will use "*p*" when the meaning is clear from context.

s = standard deviation of the sample

σ = standard deviation of the population

n = sample size (i.e., number of individuals/cases sampled)

N = population size

$\overline{x}$ = sample mean

μ = population mean

$\hat{p}$ = sample proportion. However, when it is clear from context that we are referring to a sample proportion rather than a population value, we usually omit the circumflex and simply use p to refer to the sample proportion.

p_{NH} = the proportion we specify as our null hypothesis against which the actual sample proportion is to be compared

df = degrees of freedom

The precise meaning of these terms will be clarified in later chapters.

Text boxes will be used to highlight important formulas or functions but also to call attention to important warnings about easy-to-make errors in interpreting statistical findings.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

1.4 Basic Metrics

I. Measures of Central Tendency

There are three common measures of central tendency: the median, the mode, and the mean. The *median* is the value such that half (or more) of the values are above the median and half (or more) of the values are below the median.

EXAMPLE 1.1 Median Voter Theorem

In political science, the median is an especially important concept if we can classify voters along a single scale (e.g., a left-right ideological scale). If voters always support the candidate or party that is closest to them ideologically on that dimension, then the candidate who attracts the median voter wins the election (Downs, 1957).

The *mode* is the most common value in a distribution.

The *mean* is just the familiar average.

Here and throughout, we assume that the values we are sampling from are measured on a *quantitative scale,* such as vote share—that is, at a level of measurement where summary statistics like mean and median are meaningful.

II. Measures of Dispersion

The *standard deviation* of a variable is simply a kind of average deviation, a measure of how close most of the observed values are to the mean value of that variable. To find the standard deviation of some set of values for a population, which we denote σ, we subtract each value from the mean and square the result and then take the square root of the sum of those squared values divided by n, the number of cases. In symbols,

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}. \tag{1.1}$$

Note that x with a line over it is used to denote the mean value of the variable X. Because we are subtracting each value from the mean, we are seeing how far away any given value is from the mean. By summing and dividing by n, we are calculating an average distance from the mean. We first square and then take a square root to ensure that all our values are positive numbers. Sometimes we present results in terms of *variance* rather than standard deviation. The *variance* is simply the square of the standard deviation.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

For mathematical reasons beyond the scope of this introductory text, we use the square root of the squared deviation of the individual values from the mean when we calculate the standard deviation rather than calculating the deviation of those values from the mean in terms of absolute values. Also for mathematical reasons, when we calculate the *standard deviation of a sample*, which we denote by s, we divide by $n - 1$ rather than by n. That is essentially because, in most statistical operations, we are using the mean of the distribution, and that uses up one *degree of freedom*, since if we know the mean and $n - 1$ of the values, then we can immediately determine the remaining value (see discussion of *degree of freedom* later in the text).

III. Standard Score

To understand sampling, both the *mean* and the *standard deviation* of a distribution are key concepts. They come together in the form of a *standard score.*

Sometimes we wish to see how much more extreme (i.e., far from the mean) one observation is than another. But GPA, for example, is measured on a different scale than, say, SAT scores. One way to make comparisons across different variables is to convert everything to a single scale. One way we do that is by measuring everything in units of standard deviation. To convert ordinary values to ***standard scores***, we simply normalize by subtracting the mean and dividing by the standard deviation. Standard scores, also called *z scores*, give us measurements in terms of the standard deviation of each distribution relative to its mean.

In other words, to find the ***standard score*** of any observation on some given variable,

Step 1. Find the mean and standard deviation of the variable.

Step 2. For any observation, subtract the mean of the given variable from its score on that variable.

Step 3. Divide the value you got in Step 2 by the standard deviation.

It is easier to understand the formula for calculating standard scores by expressing it symbolically using the standard notation we introduced previously.

Let z_i = the standard score of the ith case (z is the standard notation used by statisticians for standard score values):

$$z_i = \frac{x_i - \bar{x}}{\sigma}. \tag{1.2}$$

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Also note that

- Unlike the standard deviation, a standard score can be either positive or negative.
- Unlike the standard deviation, which is a property of an entire distribution (set) of values, a standard score is a property of a single observation.
- To calculate a standard score for any given observation, we need to have the values for two properties of the *overall* distribution: the distribution mean and the distribution standard deviation.

To standardize an entire distribution, we convert all of its values into *standard scores*.

IV. Percentile Rank

If we find the frequency of occurrence of each value and convert frequencies into proportions or percentages that we can view as probabilities of occurrence and then we order alternatives from lowest to highest in terms of that probability, we can create a continually updated sum of the proportion or percentages we get as we move up the distribution from lowest to highest rank. When we are dealing with proportions or percentages that sum to one, we are creating what is called a *cumulative density distribution*. For a given value x to have a percentile rank of k% for x means that k% of the cases have a parameter value (e.g., vote share for the Democratic candidate as a fraction of the two-party vote in some set of legislative elections) that is *less* than x. An outcome x at the r^{th} percentile rank gives rise to a probability p of getting that value or higher of $p = (100 - r)/100$. Similarly, if someone's test score on a test such as the MCAT or LSAT is in the 94th percentile, that means that 94% of those who took the test *did less well* than that person.

Be careful about how to interpret percentile ranks. The number one performer is the one who came in first, so lower numbers are better. In contrast, high values of percentile rank usually indicate successful performance—that is, higher is better.

The HISTOGRAM function in EXCEL (see appendix to this chapter) can create a graph of any distribution once we specify what EXCEL calls the *bin ranges*—that is, once we group the data into categories (*cell ranges*). The term "bin range" may be left over from agricultural practices of sorting eggs into bins by size of egg. By choosing the cumulative option, the EXCEL HISTOGRAM function can also create a cumulative frequency and/or a cumulative density distribution based on the specified bin ranges.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

V. Odds Ratio

In betting, probability values are often expressed in the form of odds.

To see how this works, consider an example using data on the estimated probability of candidate victory. If we estimate, say, a Republican candidate's probability of victory as .95, that means that we expect the probability of loss at .05. Our best estimate based on that information is that the odds of *victory* of that candidate are 19:1, since .95 is 19 times .05. This form of odds can be thought of as giving a probability ratio known as an *odds ratio*—that is, it is the probability that Candidate A wins relative to the probability that Candidate A loses.

But often odds ratios are expressed as the *odds against*—that is, a long shot in a horse race may have odds of 9 to 1 against in the betting pool. That means that the estimated probability of the horse winning the race is only 10%. But, of course, if the horse does win, anyone who bet on it will be well rewarded, getting nine dollars back for every dollar they bet, along with the return of their original bet.

EXAMPLE 1.2 Odds as Ratio of Ratios

Sometimes, however, we may have odds expressed comparing outcomes for two different events as a ratio of ratios. For example, a September 9, 2021, report from the U.S. Centers for Disease Control (CDC) informed Americans that, in recent months, in the United States, the unvaccinated were 11 times more likely to die of COVID-19 than the vaccinated (Brueck & Musameci, 2021). Here we are looking at a ratio of ratios—that is, the proportion of the unvaccinated who die from COVID-19 divided by the proportion of the vaccinated who died from COVID-19 during that same period.

Note that this is NOT the ratio of the *number* of COVID-19 deaths among the unvaccinated to the *number* of COVID-19 deaths among the vaccinated. We have to be very careful in interpreting ratios using raw numbers, since the proportion of the population that is vaccinated is different from the proportion of the population that remains unvaccinated, and those proportions change over time, and thus the ratio of vaccinated to unvaccinated also changes over time. When we do relative rate comparisons, we must take into account the baseline sizes of the groups being compared. Here this is done by reporting the *proportion* of deaths among one group relative to the *proportion* of deaths in the second group.

But even this comparison may be misleading if the vaccinated and unvaccinated differ not just in whether they were vaccinated but in ways related to vulnerability; for example, if older citizens were more likely to get vaccinated and the older are more likely, ceteris paribus, to die from COVID-19, then the death rate among the vaccinated may seem unduly high relative to that among

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

the unvaccinated *if we do not differentiate death rates by age* as well as by whether or not someone has been vaccinated. See the discussion of *selection bias* effects at the end of Chapter 5.

APPENDIX to Chapter 1: A Few Useful EXCEL Functions and Tools

The EXCEL functions we make most use of are those allowing us to find values for distributions such as the normal distribution and the binomial distribution. We defer introducing these until later in the text, when their usefulness can be more fully explained. However, there are some basic EXCEL functions that are very helpful to know.

=AVERAGE(array): This calculates the average value of the values located in the specified set of cells—that is, it calculates the mean value.

=STDEVP(array): This calculates the standard deviation of the values located in the specified set of cells when those values are assumed to represent the entire population.

=STDEV (array): This calculates the standard deviation of the values located in the specified set of cells when those values are assumed to come from a sample of the entire population.

=STANDARDIZE (x, mean, standard deviation): This standardizes the value x when it is assumed to be taken from a distribution with the specified mean and standard deviation

=RAND(): This is a random number generation tool that generates a random number between 0 and 1.

Many of the EXCEL tools that are most useful for statistical analyses are available only in the DATA ANALYSIS add-in, which is free for users. Adding DATA ANALYSIS to EXCEL may require accessing the Options menu in FILE and searching for add-ins. That tool, when loaded, is reachable from DATA in the main menu.

The EXCEL HISTOGRAM function in DATA ANALYSIS allows us to display data in both tabular and graphical form and can be used to simultaneously create a cumulative distribution for that data from which we can find percentile ranks The HISTOGRAM function requires the user to provide the data to be used to create the histogram and a column of numbers that tells EXCEL to group the data into ordered categories provided by the user—that is, into *cell ranges* (called *bin ranges* by EXCEL). For example,

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

if we were looking at two-candidate elections, we might choose to have bin ranges with five categories (less than 40%, 40% to just under 50%, 50% to just under 60%, 60% and above) to represent potentially competitive and clearly non-competitive elections. The values reported by the HISTOGRAM function for these categories are simply the proportion of cases that fall into the category. However, EXCEL has a particular syntax for creating a histogram. In EXCEL we would enter these percentages as a set of four values (0, 40, 50, 60), and EXCEL would automatically create the fifth category <more> for values over 60. Note that we must be careful in interpreting a histogram created by EXCEL since, for example, the probability we see when the *x*-axis shows a value of 50 in the histogram previously described is the proportion of the ordered data that is between 40% (the immediately lower value) and just under 50%. Figure 1.1 that follows was created using the HISTOGRAM function in EXCEL.

The RANDOM NUMBER GENERATION function in DATA ANALYSIS allows us to create datasets that are random samples from distributions such as the normal and the binomial. To use this function, we are required to provide EXCEL information about the features of the required distribution (e.g., mean and standard deviation). The RANDOM NUMBER

EXAMPLE 1.3 Histogram for a Two-Party Election

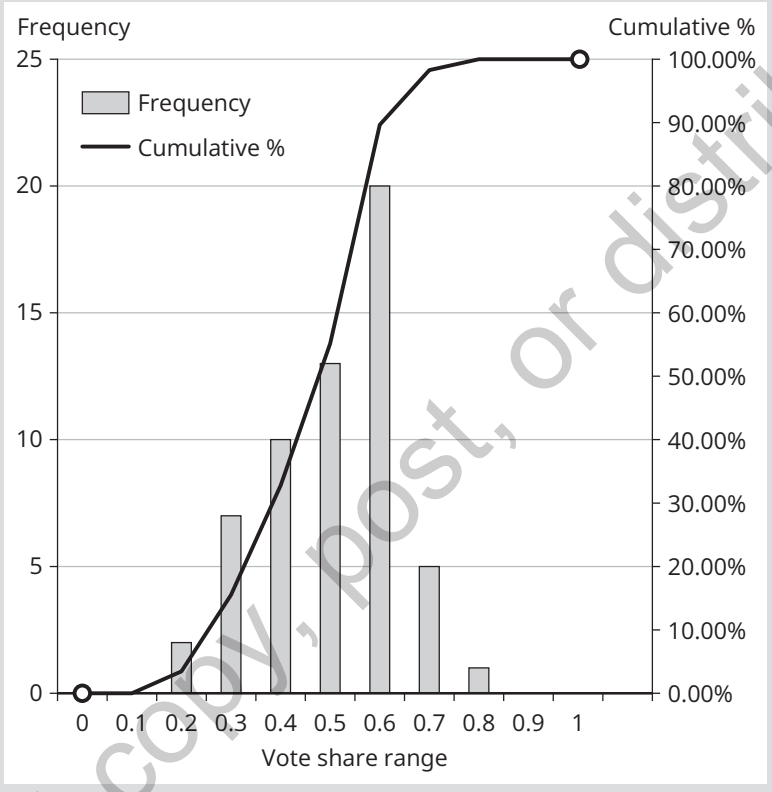
We show in Figure 1.1 a histogram for the distribution of the (two-party) vote share of the late U.S. senator from California, Diane Feinstein, in her 2012 election (the black bars). The graph also shows a cumulative distribution for this data (the gray line).

In Figure 1.1, the categories are data ranges of 10 percentage points. We can see from this graph, for example, that when we include counties with at least a 60% vote share for Senator Feinstein, we will have included 80% of the counties. However, we must be careful in interpreting the graph since the visualization in a histogram (a) depends on the cases (here counties) we are using to report the data, (b) depends upon the "fineness" (cell range size) of the way we group the data, and (c) since counties are highly unequal in population, this histogram does not visually do justice to Senator Feinstein's overall level of support in 2012 in terms of share of the two-party vote, since it turns out to be the case that most of the counties in which she failed to carry a majority of the two-party vote were counties that were among the lowest in population. Thus, the proportion of *counties carried* by a candidate is misleading as to that candidate's actual *vote share* (see Grofman & Cervas, 2024). Also, the reader must be careful in interpreting the height of the bars; they represent the shares of the distribution located within a particular cell range, while the actual values of that range are shown on the *x*-axis.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Figure 1.1 Diane Feinstein (D)—Share of the Two-Party California U.S. Senate Vote in 2012: Histogram and Cumulative Distribution by County (equal weights)



GENERATION also allows us to sample from any distribution whose members we specify—that is, from what is called a *discrete distribution.*

The EXCEL DESCRIPTIVE function in DATA ANALYSIS allows us to quickly calculate all of the most commonly used features of a dataset—for example, mean, standard deviation, range, skewness, kurtosis, and so forth.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

CHAPTER 2. SAMPLING DISTRIBUTIONS

In Chapter 2, we consider a variety of data distributions and their properties, with an emphasis on those that can be regarded as "approximately" normal to the point that we can apply EXCEL formulas to them based on the normal distribution, at least for sufficiently large sample sizes. We show how simulations using EXCEL's random number generator function in its DATA ANALYSIS menu allow us to better understand sampling processes. While our general concern in this volume is with polling, in this chapter, we draw on several examples tied to grading, since these should make intuitive sense to most readers. In particular, we treat test scores as if they are approximately normally distributed, and we link familiar concepts, such as "percentile rank" and "grading on a curve," to standard scores and features of cumulative density functions.

2.1 Ideal Types of Univariate Data Distributions

One important distinction in statistics is between *actual* data distributions and *ideal* data distributions. An actual data distribution is what we observe in the real world. An *ideal data distribution* is a theoretical distribution with known statistical properties. In this text, we are most interested in distributions, both ideal and empirical, that consist of values of some single variable—that is, what are called *univariate distributions*. Among the most common of the ideal *univariate distributions* in elementary statistics are the *normal distribution*, the *binomial distribution*, the t *distribution*, and the *uniform distribution*. A *sampling distribution* is a set of values sampled either from some actual data distribution or from an ideal data distribution.

2.2 The Normal Distribution and the Standardized Normal Distribution

A **normal distribution** is perfectly symmetric (*i.e.*, its mean, median, and mode are identical), and it is *unimodal* and bell-shaped. In a normal distribution, almost exactly 68% of the distribution is between –1 and +1 standard deviation units, and almost exactly 95% of the distribution is between –2 and +2 standard deviation units. Strictly speaking, 95% of the distribution is between +1.96 and –1.96 standard deviation units of the mean.

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
Do not copy, post, or distribute

But to gain an intuitive understanding of the key features of the normal distribution, it is easier to round this value and think of the area in question as plus or minus 2 standard deviation units. The value of plus or minus 2 standard deviation units will also work as an approximation if we are using a *t* distribution (see later in the text) rather than a normal distribution—at least for sufficiently large sample sizes. But if exact accuracy is required, then we must use the exact parameter values for our calculation.

Also, a normal distribution is unlimited from both above and below—that is, it is possible both to get very small values or very high ones, even if such values are very, very unlikely. The area of the distribution that is to the left of the –2 standard deviation point is called the **left tail** of the distribution. The area of the distribution that is to the right of the +2 standard deviation point is called the **right tail** of the distribution. Each tail of a normal distribution contains almost exactly 2.5% of the distribution.

We show in Figure 2.1 a standardized (univariate) normal distribution—that is, a normal distribution whose values have been standardized as described in a previous section. Recall from the preceding section that to create standard scores, we simply take each value, x_i, subtract the distribution's mean from it, and then divide the resultant value by the standard deviation of the distribution. This rescaling gives us values that are expressed in standard deviation units. It shows the area under the curve from −4 to +4, standard deviations, but we will focus on the area between +2 and −2 standard deviations units, especially the probability under the curve from −2 to −1, from −1 to 0, from 0 to 1, and from 1 to 2 standard deviations.

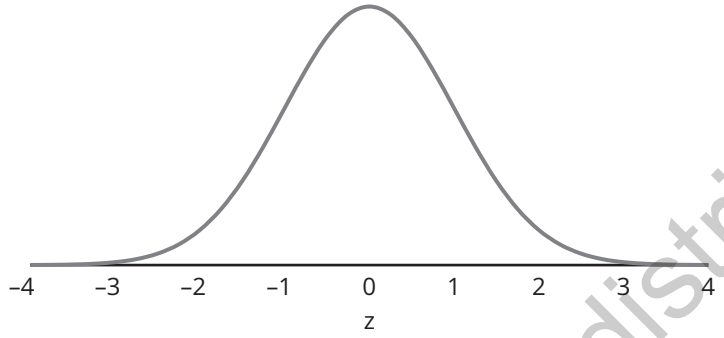
As we prove in Theorem 2.1 in the appendix to this chapter, which is optional reading, when we standardize a normal distribution by converting (*i.e.*, standardizing) the values of that distribution into standard scores, we get a (normal) distribution with a mean of 0 and a standard deviation of 1. A normal distribution with these properties is called a *standardized normal distribution*. Any normal distribution can be converted into a standardized normal distribution. To visualize the properties of a distribution more readily without the complication of looking at actual numerical data ranges, we may choose to standardize that distribution.

Most of the "probability mass" of a normal distribution is in the middle. More precisely, we know that 2.5% is the left of –2 standard deviation units—that is, in the left tail, and that almost exactly 34% of the distribution (= 68%/2) is between –1 and 0 standard deviation units. Since we also know that exactly 50% of the distribution is to the right of 0 (the median), then, by some simple arithmetic, it must be the case that 13.5%

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Figure 2.1 Standardized (univariate) Normal Distribution



is between –2 and –1 standard deviation units. The same arguments get us the percentages on the left side of the distribution (see Figure 2.1). Thus, we get the following:

Normal Distribution by Percentage

–2	–1	0	+1	+2	
←2.5%	←13.5%→	←34%→	←34%→	←13.5%→	2.5%→

Similarly, the cumulative percentages (i.e., percentile ranks) are as follows*: Cumulative percentages* show the area (probability mass) to the left of the designated standard score.

Normal Distribution by Cumulative Percentage

–2	–1	0	+1	+2	
←2.5%	←16%	←50%	←84%	←97.5%	←100%

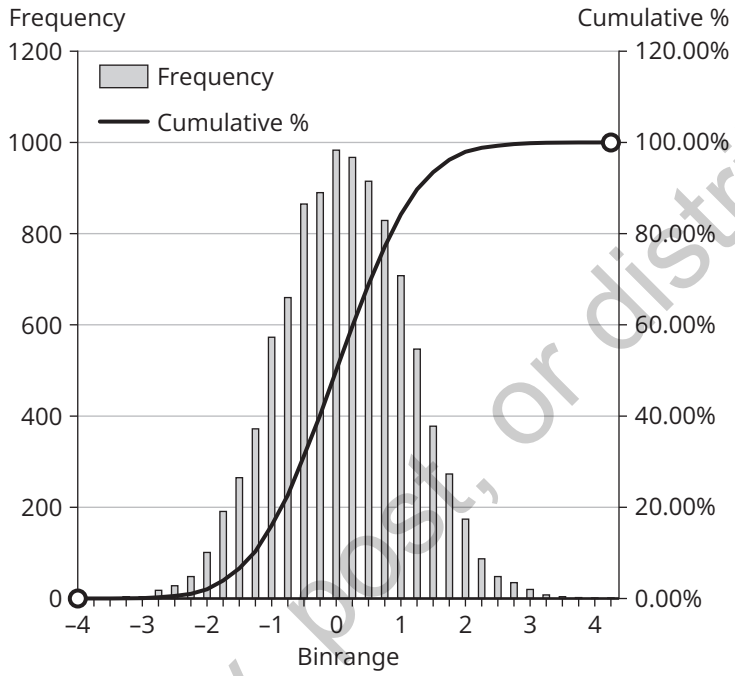
We can easily recognize a normal distribution by its bell-shaped form, and the *cumulative normal distribution* has an equally distinctive graph, an S-shaped curve. That cumulative function is shown in gray in Figure 2.2, which shows the cumulative distribution for a sample of size 10,000 taken from a standardized normal distribution.

The normal distribution is the most important in statistics. For mathematical reasons beyond the scope of this text, most real-world distributions are at least approximately normal.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Figure 2.2 Histogram of Standardized Normal and Cumulative Normal (based on a simulation with $n = 10{,}000$)



EXAMPLE 2.1 Grading on a Curve and Related Uses of Standardization

Normal distributions are often used as the basis for "grading on a curve." The idea of grading on a curve is to judge student performance not in absolute terms but relative to the performance of the other students. When instructors decide to grade on a curve, they, in effect, standardize the grades. While instructors have idiosyncratic ways of setting cutoffs when grading on the curve, often they may act as if the grade distribution was a normal distribution. One way to do this is to give As to students with scores above two standard deviations from the mean, Bs to students between +1 and +2 standard deviations from the mean, Cs to students between −1 and +1 standard deviations from the mean, Ds to students between −1 and −2 standard deviations from the mean, and Fs to students with grades below −2 deviations from

(*Continued*)

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

(Continued)

the mean. This means that 13.5% of the students can expect to get a B, 68% can expect to get a C, 13.5% can expect to get a D, and it is really unlikely to get an A (2.5%) or to flunk the course (2.5%).

The alternative to grading on a curve is to set standards for test performance in advance such that the proportion of students who get any given grade is not fixed—that is, if you do really well on an exam you can get an A, while if do really badly, you probably get a D or an F. Thus, grading on a curve is the good student's bane, because they are always in competition with other good students for a handful of As, and grading on a curve is the bad student's friend because they have to do outstandingly badly relative to the other students to flunk the course.

EXAMPLE 2.2 Grading on a Curve Without Realizing It

Sometimes, we are implicitly doing something like grading on a curve without realizing it. Consider the common practice of rating K–12 schools by the proportion of students (in a given grade) who are below average relative to all students in that grade (though perhaps only relative to those from a given multi-school school district). You may be surprised that only a relatively small fraction of the schools score above average and dismayed by the high proportion of schools below average. But we need to be careful in interpreting such statistics. Once we understand the concept of standardization and of normal distributions, it should be obvious that, for any grade level, *half of all schools will have averages below the school median for that grade level!* Moreover, even if, on average, the students in a given grade level in each school improve their performance next year so all schools improve and the median score goes up, it will still be the case that *half of all schools will have averages below the new median for that grade level!*

2.3 Approximately Normal Distributions

Often in doing statistical tests, we assume that some observed distribution is *approximately normal*, so that we may use information about the properties of the normal distribution (or one of its close relatives) to generate our statistical tests in the expectation that our conclusions will not be that far off from analyses that don't make this simplification. An important feature of

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

the normal distribution, shared by other approximately normal distributions, such as the *binomial distribution* and the *t* distribution, is that it is *unimodal*—that is, the histogram of the distribution of values looks somewhat like a bell, rising to a single peak and falling on either side of that peak.

Some "ideal" distributions, such as the *t* distribution and the binomial distribution, come to look more and more like a normal distribution, as the number of elements we sample from that distribution increases. Later in the text we provide a series of graphs to demonstrate this fact for the binomial distribution. Also, as we later show through simulation results, if we sample from any distribution and look at the distribution of those means—that is, at the *sampling distribution of the mean*—as we increase sample size, that distribution will look increasingly like the normal. But we also find that many distributions that we find in the real world are *approximately normal*—that is, bimodal and roughly symmetric, with mean, median, and mode at or about the same location.

As noted earlier, there are deep mathematical reasons why the normal distribution (or distributions very similar to it) are so commonly found in the real world and why sampling processes involving other ideal distributions tend to lead us back to the normal distribution. But those take us into mathematics well beyond the scope of this introductory text. The best we can do here is to confirm the result by providing simulations using EXCEL that lead us to distributions that look more and more like the normal distribution as we increase the sample size.

2.4 Cumulative Distributions and Finding Percentile Ranks Using EXCEL

When we have a normal distribution, finding the percentile rank for any given value is very simple. If we know the standard score of the value, then we use the =NORMSDIST() function to convert that standard score into the area under the curve up to and including that value—that is, we use EXCEL to locate the values of the cumulative distribution. Because the standard normal is a *density distribution*—that is, the area under it sums to one, using the =NORMSDIST () function in EXCEL gives us a probability that corresponds to the proportion of the curve that lies to the left of that standard score. Once we have found that proportion, then all we need to do is convert it into a percentage. That percentage will be the desired percentile rank. But we can use this method even if we do not begin with a normal distribution but rather begin with one that we can treat as approximately normal.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

EXAMPLE 2.3 Finding Percentile Ranks Based on GPA

We may rank-order students at some college in terms of their GPA (say, from lowest to highest). A student's percentile rank (with respect to GPA) is defined as the percentage of member students whose GPA is the same or lower than his. If GPA is approximately normally distributed, and if we know the mean and the standard distribution of GPA scores, we can readily estimate what GPA corresponds to what percentile rank (and conversely). There are two ways to do this.

In EXCEL, if you have already converted values into standard scores, then you use the function

=NORMSDIST(standard score),

to convert from standard scores to the (estimated) percentile rank. (Results are reported as a fraction from 0 to 1.)

For example, the value that corresponds to =NORMSDIST(0) is .500; the value that corresponds to =NORMSDIST(–1.96) is .025; the value that corresponds to =NORMSDIST(1.96) is .975. In other words, if, say, NORMSDIS(0) =.50, then that tells you that a standard score of 0 corresponds to a percentile rank of 50%. Recall that when we say that 2.5% of a standardized normal distribution is to the right of 2 standard deviation units, this is only an approximation. The more exact value is 1.96.

To go in the other direction—that is, to convert from percentile rank to the (estimated) standard score, you use the function

=NORMSINV(proportion),

where percentile rank is entered as a proportion—that is, is entered as a number from 0 to 1.

What =NORMSINV(proportion) does for a normalized distribution is tell you the standard score that corresponds to a given percentile rank.

Of course, you may want to know the actual value that corresponds to a given percentile rank—for example, what UCI GPA would you need to be at the 75th percentile (*i.e.*, a higher GPA than 75% of your fellow students). To do this, you make use of Eq (1.2)—that is,

$z_i = \dfrac{x_i - \bar{x}}{s}$, and then, after some simple arithmetic to rearrange terms, we get

$$x_i = z_i s + \bar{x}. \tag{2.1}$$

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

We can then use =NORMSINV(proportion) to find the standard score that corresponds to the desired percentile rank, and then we plug in the value of the mean ($\bar{x}$) standard deviation (s) into the formula to find the actual value that corresponds to that percentile rank.

For example, if GPA has a mean of 2.9 and a standard deviation of 1.0 and we want to find the GPA that corresponds to the 75th percentile rank, then we can first use the function =NORMSINV(.75) to find the standard score that corresponds to the 75th percentile rank (learning that =NORMSINV(.75) = 0.67) and then plug that value, along with the mean (2.9) and the standard deviation (1.0) into Equation 2.1 to find that

$$x_i = 0.67 \times 1.0 + 2.9 = 3.57.$$

Thus, under these assumptions, it would take an estimated GPA of 3.57 to put a student into the top quartile of the grade distribution (*i.e.*, the top quarter of the class) at this college.

If you did not want to bother calculating standard scores, a second way to calculate the GPA score is to use the =NORMINV() function to directly find the value that corresponds to a given percentile rank. The =NORMINV() function has three arguments: The first is the percentile rank expressed as a fraction; the second is the mean; and the third is the standard deviation—that is,

=NORMINV(probability, mean, standard deviation).

Thus, to find the GPA value that corresponds to the 75th percentile rank for a GPA distribution with a mean 2.9 and a standard deviation 1.0, we simply find

=NORMINV(.75,2.9,1),

the value of which is, as expected, 3.57.

Another useful EXCEL function is

=NORMDIST(x, mean, standard deviation, cumulative).

This function has four arguments and allows us to directly deal with approximately normal distributions without the need to standardize their values. When we are looking for percentile ranks, we use TRUE as the value of the fourth argument. This tells us the percentile rank that corresponds to a given value. For example, if you have a GPA of 3.57 and want to know where that

(*Continued*)

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

(Continued)

ranks you and you know that the (approximately normal) distribution has a mean of 2.9 and a standard deviation of 1.0, then you specify

=NORMDIST(3.57,2.9,1,TRUE),

which turns out to be equal to 0.75. In other words, under the preceding assumptions, a GPA of 3.57 is estimated to put someone at the 75th percentile rank—exactly as we would expect. But we can also do this calculation for any GPA we might be interested in.

2.5 The Binomial Distribution

I. Characteristics of the Binomial Distribution

The *binomial distribution* is used when we have dichotomous or polychotomous variables. In most statistics texts, the binomial distribution is introduced by considering things like coin flips and dice throws, but for our purposes, we most need to understand how the binomial distribution helps us make sense of two-candidate elections. We may treat, say, a national random sample of voters who stated whether or not they voted for President Trump in 2016 (or 2020 or 2024), taking p, the proportion of those voting for President Trump in the sample, as if it were coming from a normal distribution. But when studying two-candidate elections where a standard deviation was called for, in the EXCEL formula to calculate the normal distribution we would normally insert the standard deviation of a binomial distribution with mean p and take the number of cases in the sample as our value of n. Moreover, to ensure consistency and ease of interpretation, we would focus on the two-party vote—that is, the vote for President Trump divided by the sum of his vote and that of his principal opponent in the Democratic party. However, in some elections, failing to take into account minor party candidates can lead to misleading results (see Grofman & Cervas, 2024).

A. Binomial Mean

The mean, μ, of any dichotomous variable with success probability p is simply the expected proportion of observations that will be successes—that is,

$$\mu = p. \tag{2.2}$$

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Over n trials, the mean μ of any binomial variable with success probability p is simply the expected *number* of successes in n trials—that is,

$$\mu = np. \tag{2.3}$$

If, for some binomial variable, we increase the number of *independent random trials*, also known as *Bernoulli trials*, from say n to N, then the mean μ of that variable will be increased by the factor N/n.

In a notation scheme often used in statistics textbooks, we should be using p with a little hat (circumflex) over it to denote the sample proportion rather than the population proportion. The author finds this more trouble than it is worth, as long as the reader remembers that any given sample has a certain mean number of successes or a given proportion of successes that need not be the same number of successes or proportion of successes that you find in the population as a whole and that the sample is not the population.

B. Binomial Standard Deviation

Over n trials, the standard deviation, σ, of any binomial variable with success probability p is given by

$$\sigma = \sqrt{np(1-p)} \text{ or equivalently } \sigma = \sqrt{npq}. \tag{2.4}$$

Here q is used to represent $(1 - p)$.

If you increase the number of Bernoulli (independent random) trials for some binomial variable from say n to N, then the standard deviation σ of the number of successes will be increased by the factor

$$\sqrt{\frac{N}{n}}. \tag{2.5}$$

C. Binomial Proportions

When we are interested in proportions rather than frequencies, we simply divide by the number of observations to convert from frequencies to proportions. For example, if 6 out of 50 items in a sample have a given value, then the proportion (*i.e.*, fraction) of the sample having that value is equal to .12 of the total number of observations in the sample (= 6/50 = 12%). So once we know any statistical formula that is specified for frequencies, to convert that formula to proportions, all we need do is divide it by n.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

1. Mean of a binomial proportion

   In the preceding, we indicated that the mean of a binomial distribution is *np*. What is the mean proportion for a binomial distribution? That is easy to figure out; it's simply $np/n = p$, *i.e.*,

$$\mu = p. \tag{2.6}$$

2. Standard deviation of a binomial proportion

   We previously noted that the standard deviation of a binomial distribution when there are n trials is $\sqrt{np(1-p)}$. What then is the standard deviation for proportions for a binomial distribution? Well, we simply take the preceding formula for frequencies and divide through by *n*—that is,

$$\sigma = \frac{\sqrt{np(1-p)}}{n} = \sqrt{\frac{p(1-p)}{n}}. \tag{2.7}$$

Using high school algebra, this equivalence is demonstrated in Theorem 2.2 in the appendix to this chapter. That appendix should be viewed as optional material. In Chapter 5, we will have more to say about this important formula.

II. As n Increases, the Binomial Distribution Looks More and More Like a Normal Distribution

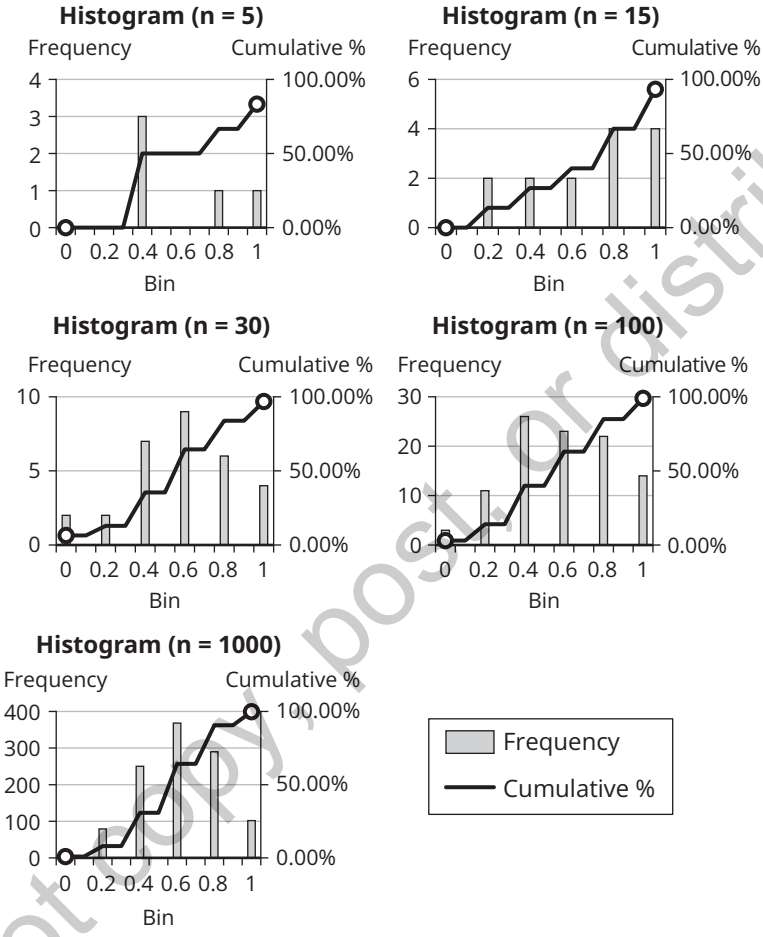
We have previously noted that the binomial distribution is approximately normal. But there is an even stronger result—namely, that the binomial distribution becomes indistinguishable from a normal distribution as *n*, the number of trials, increases. While proving that result requires mathematics well beyond the scope of this text, we can show the intuition underlying the result by presenting the results of a *simulation* (Figure 2.3) in which we look at a binomial distribution with $p = .6$ and use the random number generator feature of the DATA ANALYSIS tool in EXCEL to create *n* random binomial trials—that is, the distribution we could get if we tossed a coin with $p = .6$ *n* times.

Figure 2.3 is an example of a simulation using the idea of random draws from a given distribution, here the binomial. By the time we get to 1000 trials, as can be seen from Figure 2.3, the distribution of outcomes from our binomial distribution looks very much like a normal distribution.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Figure 2.3 Sampling From a Binomial Distribution With $p = .6$



NOTE OF CAUTION: Symbols like <*p*> and <*n*> have multiple meanings in statistics. This is one of many things that beginning students find confusing. Actually, so do I. But you will just have to get used to deciphering symbol meaning from the context. Most commonly, "*n*" refers to the number of data points (cases) in a population or sample, but in the context of simulations, "*n*" can also be used (as it is in Figure 2.3) to refer to the number of times we run the simulation.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

While we can create various types of random distributions using the random number generator from the DATA ANALYSIS tool in EXCEL, some EXCEL formulas will directly create randomness—of which perhaps the most important is =RAND(), which randomly draws values from the *uniform distribution* over the [0, 1] range. We make further use of simulation tools when we consider the "laws of statistical sampling" in the next chapter. Simulations are one of the most powerful tools in statistical modeling (see e.g., discussion in Page, 2018), though illustrations of the full scope of their uses will have to wait for a more advanced course in statistics.

There is a common-sense meaning of randomness as drawing an item from some distribution with equiprobability, but we have to be careful to understand what it is that is being equalized. As noted earlier, if we draw from a distribution where some values are more common than others, then all values do not have the same probability of being chosen when we do a random draw. Also, we need to distinguish random draws from some *actual* distribution (*sampling randomly* from the distribution) from random sampling from some *ideal* distribution, such as a binomial distribution with fixed parameters or a standardized normal distribution.

If $p = .5$, the binomial distribution looks almost exactly like a normal distribution by the time you hit 30 trials. But the simulation we provided in Figure 2.3 has a p value of .6, so there is skewness in the data (see discussion of skewness that follows), and it takes longer for the underlying normality to be clearly visible.

WARNING: The degree to which a binomial distribution resembles a normal depends upon p as well as upon n. When we have p values that are far from .5, the binomial distribution, limited as it is by 0 and 1, is no longer symmetric. The binomial distribution skews right for small p (<.5) and left for large p (>.5). A rough rule of thumb is that you may use the normal approximation to the binomial if $np > 10$. This rule of thumb takes into account the fact that the fit of the normal to the binomial depends not only on n but also on p.

2.6 The *t* Distribution

The distribution more formally known as *Student's* t *distribution*, now better known simply as the t distribution, is very similar to the normal

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Do not copy, post, or distribute

distribution. *Student's t* is named after the early 20th-century statistician, William S. Gossett, who went by the pseudonym "Student." The *t* distribution looks even more like the normal distribution than does the binomial. It is commonly used when we are analyzing data that can be regarded as a sample coming from some underlying population. It works better for small samples than the normal distribution.

As noted earlier, a *z* score is just another name for a *standardized score* (i.e., what you get when you take the original value, subtract the mean of the distribution, and then divide the resultant difference by the standard deviation of the distribution). We've learned, for example, that for a normal distribution, a standard score of −1.96 (i.e., a *z* score of −1.96) corresponds to a percentile rank of 2.5%; while a standard score of +1 (i.e., a *z* score of +1) corresponds to a percentile rank of roughly 84%. A *t* statistic can also be thought of as a form of standard score, except that it is the standard score for a *t* distribution rather than for a normal distribution. The *t* statistic tells us how far away from the mean of the distribution we are (measured in standard deviation units) for data that can be represented as a *t* distribution.

When the sample size is large, a *t* distribution is essentially the same thing as a normal distribution, so that, when the sample size is large, the corresponding *t* statistic will be essentially equal to the *z* score (see Table 2.1 that follows). The *t* statistic is calculated for a given number of degrees of freedom. Its number of degrees of freedom is just $n - 1$. For small sample sizes, there will be some difference between *t* statistics and *z* statistics, but the differences narrow and eventually vanish as the sample size increases. Table 2.1 shows the *t* values that correspond to a percentile rank of 97.5% (a *z* score of +1.96) for selected values of *n* from 5 through 1000. As is apparent, the *t* statistic comes closer and closer to 1.96, and using a value of 2 as an approximation works for a wide range of values of *n*.

Because the normal distribution is intended for use only for sample sizes above 30*, most computer packages report results only in terms of the* t *distribution.* That is because the *t* distribution looks more and more like a normal distribution as sample size increases but also because the *t* distribution is more accurate for small samples—though the mathematical reasons for this statement are beyond the scope of this introductory textbook. What

Table 2.1 The Magnitude of the *t* Statistic Needed for a 95% Level of Confidence as a Function of Sample Size

n	*5*	*10*	*15*	*20*	*25*	*30*	*50*	*70*	*150*	*200*	*500*	*1000*
***t* test (.05)**	2.57	2.23	2.13	2.09	2.06	2.04	2.01	1.99	1.98	1.97	1.96	1.96

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

we can say is that, for sample sizes above 30, the results from positing the data from which the data come is a *t* distribution are very close to those we get from positing a normal distribution, while for sample size in the 100s the results are essentially indistinguishable from the results that come from positing that the distribution from which the data comes from is a normal distribution. Moreover, even for sample sizes below 10, the two distributions look remarkably alike (see Table 2.1). However, for small sample sizes, a *t* distribution has less of its mass in the middle of the distribution than does the normal distribution—that is, it has more probability mass in the tails (see discussion of *kurtosis* the follows).

Just as there are EXCEL functions to calculate *z* scores (standard scores) for a normal distribution, so are there functions =TDIST and =TINV() to calculate *t* statistics.

To calculate a *t* distribution, we sometimes input more information and sometimes less information than for a normal distribution. For =NORMDIST, we need to input four parameters: the value whose probability we are interested in, the mean and the standard deviation of the distribution, and whether or not we are seeking the cumulative function or simply the probability of occurrence of that given value. For =TDIST, we have only three parameters in the function, namely the value whose probability of occurrence we are interested in, the number of *degrees of freedom* (*df*) in the data, which is usually $n - 1$, and whether or not we are seeking the cumulative function or simply the probability of occurrence of that given value. For a *t* distribution, we do not need to tell it how many tails we want, because, in EXCEL, this function is set up to *always* give a *one-tailed test*. If we want a two-tailed test, we need to double the value.

For =NORMINV, to obtain a *z* score that corresponds to a given probability value, we need to input only the *p* value. For =TINV to find the *t* score corresponding to the *z* score (i.e., the standard score of the corresponding normal distribution) in addition to the desired *p* value, we also need to specify the degrees of freedom (*df*) of the *t* distribution. The structure of =TINV () is =TINV (probability, degrees of freedom)—that is, it only requires two inputs. The *df* of a *t* distribution is usually $n - 1$. However, for a *difference of means* test using the *t* distribution (see discussion of difference of means tests in Chapter 5), a different *df* is required.

2.7 Other Approximately Normal Distributions

There are other distributions, such as the *chi-square distribution*, the F *distribution*, and the *Poisson distribution,* that are closely related to the normal and that can also become very nearly normal as their number of

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

degrees of freedom increases. Also relevant in more advanced courses is the *lognormal distribution* (see below).

Some distributions can look like normal distributions reflected in some kind of funhouse mirror, e.g., they have "fatter tails" than the normal distribution, on the one hand, or they have more of their mass concentrated in the middle, on the other hand (measured by *kurtosis* – see below); or they may be skewed away from perfect symmetry (measured by *skewness* – see below). Such distributions still can for certain purposes be treated as approximately normal, though in more advanced statistics classes statistical tools specifically designed for particular distributions (e.g., the *lognormal distribution*) will be introduced. But in this introductory text, it is the normal, the binomial, and the *t* distribution on which we will focus while, for the lognormal, we limit ourselves to a very brief discussion and two examples that follow.

2.8 Skewness and Kurtosis

I. Skewness—Degree of Asymmetry in a Unimodal Distribution

We will give examples of s*kewed distributions* in the next section. **Skewness** (like *kurtosis*—discussed later in this chapter) tends to be largely ignored in most introductory statistics texts but can be another useful piece of information when attempting to characterize a distribution. Skewness answers the question, "Are the data symmetrically distributed about the mean?" If the answer is no, we generally would like to know which side is favored. If the right side (i.e., scores greater than the mean) is favored, the distribution is *positively skewed*.

> In a ***positively skewed distribution,*** the mean is greater than the median.

If, on the other hand, the left side is favored, the distribution is *negatively skewed*.

> In a ***negatively skewed distribution,*** the mean is less than the median.

A. Lognormal Distribution

Perhaps the most important skewed distribution is the *lognormal.* If, instead of *adding* elements to get a mean, we were *multiplying*_elements, we no longer get a normal distribution. Instead, we get a lognormal distribution.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

For a lognormal distribution, it is the logs of the values that are normally distributed. There are statistical tests to determine if the degree of skewness significantly differs from what would be expected from a normal distribution, but these take us outside the scope of this introductory volume.

EXAMPLE 2.4a Real-World Applications of the Lognormal Distribution: Income Inequality

In the study of income, if the mean and the median of an income distribution don't coincide, with the mean considerably above the median, then it is a virtual certainty that there is a great deal of income inequality—that is, where there are many more very poor than very wealthy and where most of the wealth is concentrated in the hands of the super-rich. Lognormal distributions can be used to model such skewed distributions (see especially, discussion of the lognormal distribution in Page, 2018).

EXAMPLE 2.4b Real-World Applications of the Lognormal Distribution: Distribution of Electoral Support

In politics, an application of the lognormal distribution is to the study of party vote share and party seat share sizes in multiparty democracies (Taagepera, 2007).

B. Skewness in the Distribution of Electoral Support at the Constituency Level

EXAMPLE 2.5 Real-World Importance of Skewness as a Test for Gerrymandering

When we observe a difference between the mean and the median of a distribution, this is a potential clue that something interesting is going on. In politics, when political boundary lines in a legislature (or for the U.S. Congress) are gerrymandered by one side, the side controlling the redistricting process, then the *mean* support for the other side is usually greater than the *median* support for the other side. In other words, the party that is victimized by the gerrymandering is not as good at translating its votes into winning districts as the party doing the gerrymandering—that is, there is bias (asymmetry) in the translation of each party's votes into seats. While there are other more sophisticated tools to identify gerrymandering, looking at skewness is a useful "quick and dirty test" (cf., Grofman & Cervas, 2025 forthcoming, "Partisan Gerrymandering . . .").

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

II. Kurtosis—Peakedness in a Unimodal Distribution

Kurtosis is a measure of how peaked a distribution is—that is, how much of the distribution is concentrated very close to the peak at the mode of the distribution. Kurtosis is a measure not very commonly used in introductory statistics classes because the standard deviation presents similar information (at least for unimodal distributions). Ceteris paribus, highly peaked distributions tend to have a lower variance. However, sometimes we want to know how closely a distribution fits with the ideal normal distribution, and knowing the kurtosis of a distribution helps us answer that question.

When a distribution is peaked, more of it is in the middle and less in the tails than is the case for a normal distribution. When a distribution is flattened, less is in the middle and more is in the tails than is the case for a normal distribution. The latter type of distribution, usually referred to as a *leptokurtic distribution*, is sometimes called a *fat-tailed distribution*, and no, we are not making that term up!

In general, the ideal normal distribution has a kurtosis of 3, so we can compare a distribution's kurtosis to that value and know whether or not that distribution is more peaked than, less peaked than, or equally as peaked as the normal distribution. There are statistical tests to determine if the degree of kurtosis significantly differs from what would be expected from a normal distribution, but these take us outside the scope of this introductory volume. The idea that some distributions have *fat tails*—that is, seemingly rare events are less rare than we would expect were the distribution to have been normally distributed—is very important in economics and in other areas such as epidemiology.

EXAMPLE 2.6a Real-World Importance of Kurtosis: Insurance Failures

Fat-tailed distributions can occur when failure rates are correlated—for example, if one insurance company "hedges its bets" by insuring a portion of its mortgages with another insurance company and there is a mortgage crisis, then the failure of the second insurance company can lead to the failure of the first insurance company. The failure to take into account the interconnectedness among insurance companies has been used to account for why almost no one predicted the mortgage failures/insurance failures crisis in the U.S. housing market in 2008, because they were modeling the probability of failure as a series of *independent* events. Because of this interconnectedness, we can get fat-tailed distributions since tipping effects can lead to either few events at each tail or many events at each tail.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

EXAMPLE 2.6b Real-World Importance of Kurtosis: Spread of Infectious Diseases

Another example of a distribution that is likely to have fat tails comes in the spread of infectious diseases when we look at the distribution of disease incidence. Since we catch infectious diseases from others, either directly or through surface or airborne transmission, areas where a non-trivial number are initially infected are likely to end up in a situation where a very high proportion are ultimately infected, and areas that avoid any infection in the early stage of the disease may stay relatively infection free and are likely to be more common than one might expect from a normal distribution.

EXAMPLE 2.6c Real-World Importance of Kurtosis: Catastrophic Events

In situations where the failure of any single link can lead to overall failure—for example, any break in a levee leads to flooding, the probability of failure can also be modeled as a fat-tailed distribution. Such a model has been used to study catastrophic events, such as the devastation of New Orleans by Hurricane Katrina in 2005.

EXAMPLE 2.6d Real-World Importance of Kurtosis: Electoral Tides

But there are applications of fat-tailed distributions in politics as well. For example, imagine that we are trying to predict how likely it is that a given candidate will win the Electoral College. This requires us to look at the likelihood of outcomes in multiple states. If we treated those outcomes as independent events, we could be severely misled. Rather, we might expect that there are electoral tides or exogenous events that will, by and large, have a similar directionality of effect (if not the same magnitude of effect) in different states. In statistics, this is an example of what is called *correlated error*. If there is a tide running in a given direction, then there will be fewer very large victories for one party (against whom the tide is running) and more very large victories for the party benefiting from the tide than we would expect from a normal distribution.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

2.9 Not All Univariate Distributions Are Approximately Normal

I. Examples of Distributions That Are Not Approximately Normal

While unimodal (approximately) normal distributions are the most common, not all distributions qualify as approximately normal. As discussed previously, unimodal distributions can differ from normality in terms of features such as skewness and kurtosis. For example, following we will show California Congressional ADA roll call voting scores (range 0–100) for the Democrats only (negatively skewed) and California Congressional ADA roll call voting scores (range 0–100) for the Republicans only (positively skewed).

In the list that follows, we identify some non-unimodal distributions. In the next section of the text, we provide figures to illustrate several of these non-unimodal distributions.

1. Bimodal and dish-shaped

   Example: California Congressional ADA scores for the entire set of legislators from that state.

2. Bimodal like a two-humped camel

   Example: DW-NOMINATE roll call voting scores from various years, showing the humps moving wider apart since the 1960s as the parties polarize in ideological terms.

3. Trimodal

   Example: Create a variable for vote shares in U.S. House elections for Democratic candidates. Historically, there has been a concentration of values in three parts of the graph: safe seats for Democrats won by overwhelming margins or uncontested, safe seats for Republicans won by overwhelming margins or uncontested, and more competitive seats closer to the middle. More recently, the share of competitive seats has been falling.

4. Uniform distribution (identical probabilities graphed as a flat line)

   Example: In the Vietnam-era draft lottery of 1969, the likelihood of drawing a lottery number for each birthdate by day and month was supposed to be the same for all birthdates. Random selection mattered a lot since the first k birthdates chosen ended up being drafted. For the record, I should note that the author's birthdate was part of that lottery.

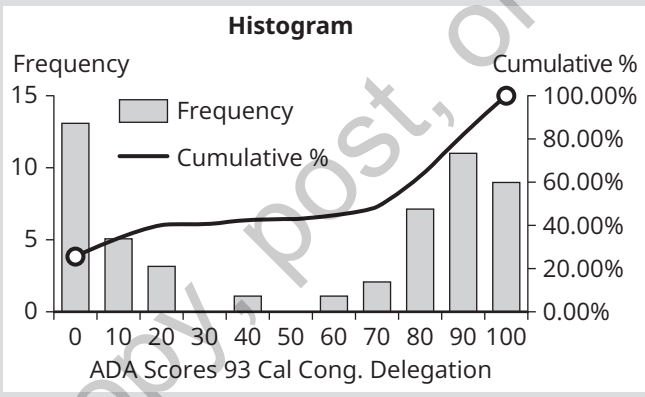
Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

EXAMPLE 2.7 ADA Scores in the U.S. House of Representatives (dish-shaped distribution)

Look at the distribution shown in Figure 2.4 of 1993 ADA (right-left ideology) scores of California members of the U.S. House of Representatives based on roll call votes. ADA scores were created by Americans for Democratic Action, a liberal group, and range from 0 to 100, with 0 being the score associated with the most conservative roll call voting pattern and 100 being the score for the most liberal California congress member. (Note that this distribution is bounded from both above and below.) Figure 2.4 is a bimodal distribution that is dish-shaped rather than bell-shaped.

Figure 2.4 1993 ADA Scores of the California Congressional Delegation



If we were to look at more recent data for California, the data would look even more like having spikes at either end, since there is essentially no California congress member found in the ideological middle.

Now, consider what happens when we look only at Democratic members of the California House delegation or only at Republican members of the delegation, as shown in Figure 2.5a and Figure 2.5b, respectively.

Figure 2.5a is certainly not a normal distribution. It is not even close to symmetric; it is strongly negatively skewed. California representatives, who were Democrats, were generally characterized by quite high liberalism as represented by their high ADA scores.

Similarly, if we look at ADA scores among Republican members of the California House delegation, as shown in Figure 2.4b, we find a distribution again not even close to symmetric, but now it is strongly positively skewed. California's Republican representatives were generally characterized by quite high conservatism, as represented by their low ADA scores.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Figure 2.5a 1993 ADA Scores of the Democratic Members of the California Congressional Delegation

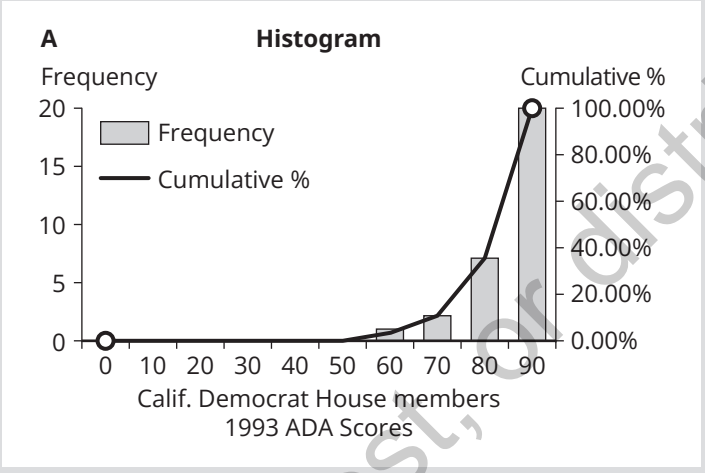
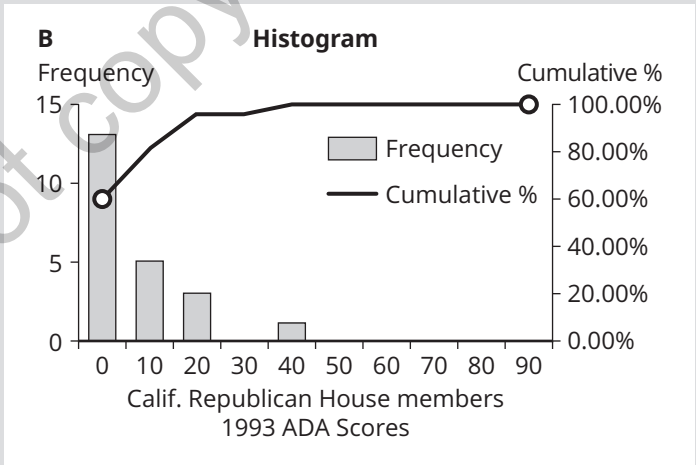


Figure 2.5b 1993 ADA Scores of the Republican Members of the California Congressional Delegation



Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

EXAMPLE 2.8 Combining Overlapping (approximately) Normal Distributions—Student Heights

Now let us consider the distribution of heights among undergraduate students at some college. Would you think that distribution would be approximately normal?

Answering that question turns out to be tricky. We need to take gender into account. The distribution of men's heights might be (approximately) normal, and the distribution of women's heights might be (approximately) normal, but, because there are gender differences in heights that are rather strong, it might seem unlikely that the combined distribution would be approximately normal. The overall distribution we get when we "mix" two different approximately normal distributions with quite different means is not clear. It could be bimodal, with one "hump" around the average woman's height and another "hump" around the average man's height. But even though the modes of each distribution may be rather separated from one another because there is overlap among these two (approximately) normal distributions, the result might not be too far from a normal distribution.

If we combine two normal distributions with different means and variances, there is a mathematical result that uses those values to help us decide how close to a normal we would get for particular combinations of two normal distributions.

> Proposition 1. A sufficient condition for the combined distribution of two normal distributions to be unimodal is that the absolute value of $(\mu_1 - \mu_2)$ is greater than $2*\max(\sigma_1, \sigma_2)$.

When we applied this test to the height distribution of students in one of my undergraduate classes (as self-reported by the students), the condition was satisfied, but when I applied it to the weight distribution in that same class (also as self-reported), it was not.

II. Properties of the Uniform Distribution

The *uniform distribution*, with all values equally likely, is perhaps the most important univariate distribution that is not approximately normal. A uniform distribution assumes that all the values in some set occur with exactly equal probability. If we create a histogram for a random sample from the uniform distribution, using equal intervals for our cell values, the bars in the histogram graph will all be of (approximately) the same height.

> If the uniform distribution runs between a and b, with $a < b$, it has a mean of $(b - a)/2$, and a standard deviation of $(b - a)/\sqrt{12}$.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

For the uniform distribution we get from the EXCEL Function RAND(), where $a = 1$ and $b = 0$, we get a mean of .5 and a standard deviation of .2887.

We show in Figure 2.6 a standardized uniform distribution whose values range from 0 to 1, along with its associated cumulative density distribution simulated using the EXCEL uniform distribution function. Even with an n of 10,000, the graph of standardized values does not have each of the cells of exactly equal size, but the idea that, when we have a uniform distribution, the middle part of its cumulative distribution should look like a straight line with a 45-degree slope nonetheless shows through clearly.

EXAMPLE 2.9 Real-World Applications of the Uniform Distribution

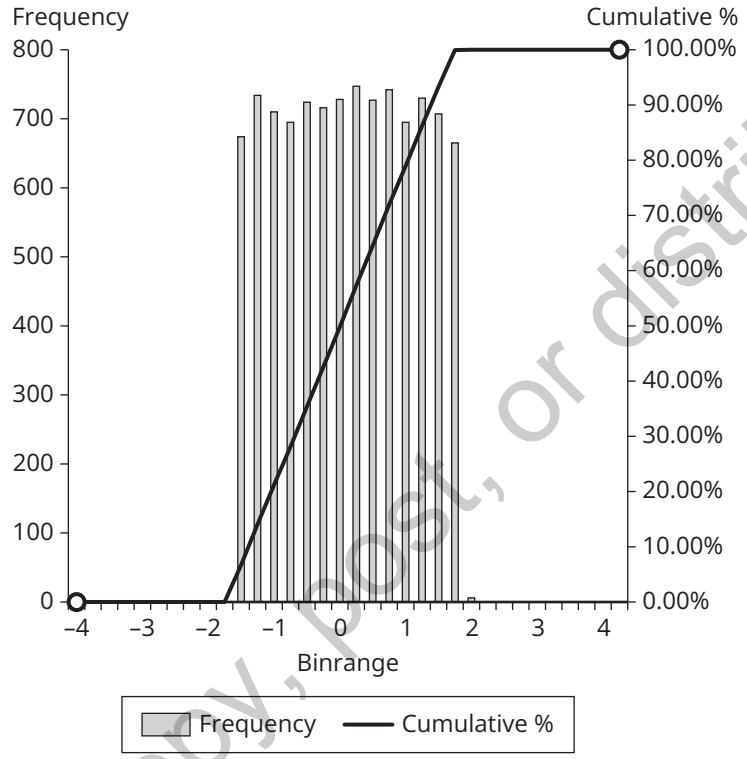
We noted earlier that a *uniform distribution* arises in lotteries, such as the Vietnam-era draft lottery, when we are trying to make sure that everyone (or everything, such as a lottery ticket) entered into the lottery has an equal chance of being chosen. Similarly, a roulette wheel is supposed to have each outcome equally likely. Another way that the uniform distribution is used in the social sciences is in terms of what philosophers of social science call the "principle of indifference"—that is, if we don't know anything about how probable anything is and we can't even separate more probable from less probable outcomes, then it may make the most sense to just assume equiprobability.

A political science application of the concept of equiprobability involves data on how voters rank alternatives or candidates. Ranking methods are of considerable interest in the study of multi-candidate elections. Contemporary election reformers have proposed to replace plurality elections (where the candidate with the most votes wins, even if that vote is not a majority) with methods labeled as *rank-choice voting*, such as the *instant runoff voting family of methods* (IRV). IRV requires voters to rank the candidates, with a majority of the votes required to win. In the most common variant, if no candidate has a majority, candidates with the fewest first-place votes are dropped from eligibility and their votes reallocated to the next highest ranked still eligible candidate on the voter's ranking, with the process continuing until some candidate gets a majority—which must happen when we are down to two candidates. Social choice theorists who study the properties of alternative voting methods, such as the instant runoff, to compare these with results under simple plurality, have done simulations under what they call the *impartial culture* assumption, which assumes that all possible rankings are equally likely.

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Figure 2.6 Histogram of Standardized Uniform (initial values between 0 and 1) and Standardized Cumulative Uniform ($n = 10{,}000$)



APPENDIX to Chapter 2: Theorem Proofs (optional material)

Theorem 2.1 Properties of Standardized Distributions

When we convert a normal distribution into a standard normal distribution by converting (i.e., standardizing) the values of that distribution into standard scores, we get a normal distribution with a mean of 0 and a standard deviation of 1.

Proof:

I. Proof that the mean of a distribution of standard scores (i.e., the mean of any standardized distribution) equals zero.

When we convert the values of a distribution into standard scores, we get, new x_i, which we shall denote x_i' (read *x*-sub-i-prime) such that

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

$$x_i' = \frac{x_i - \bar{x}}{\sigma_x}.$$

The mean of this new standardized distribution is simply

$$\bar{x}_i' = \frac{\sum x_i'}{n} = \frac{\sum (x_i - \bar{x})}{n\sigma_x} = \frac{\sum x_i}{n\sigma_x} - \frac{\sum \bar{x}}{n\sigma_x}.$$

But

$$\frac{\sum x_i}{n} = \bar{x},$$

and

$$\sum \bar{x} = n\bar{x}.$$

And so, combining these results, the mean of a standardized distribution is

$$\bar{x}_i' = \frac{\bar{x}_i}{\sigma_x} - \frac{\bar{x}_i}{\sigma_x} = 0.$$

QED

II. Proof that the standard deviation of a distribution of standard scores (i.e., the standard deviation of any standardized distribution) equals one.

The standard deviation of the distribution x_i' such that

$$x_i' = \frac{x_i - \bar{x}}{\sigma_x},$$

is simply the standard deviation of $\frac{x_i}{\sigma}$ minus the standard deviation of $\frac{\bar{x}}{\sigma}$:

$$= \sigma_{\frac{x_i}{\sigma_x}} - \sigma_{\frac{\bar{x}}{\sigma_x}}.$$

But, since both $\bar{x}$ and σ_x are constants, it follows that

$$\sigma_{\frac{\bar{x}_i}{\sigma_x}} = 0.$$

Therefore, the standard deviation of x_i' is given solely by the standard deviation of $\frac{x_i}{\sigma}$.

But, since σ_x is a constant, it follows that the standard deviation of $\frac{x_i}{\sigma}$ is

$$\sum \sigma_{\frac{x_i}{\sigma_x}} = \frac{\sigma_x}{\sigma_x} = 1.$$

QED

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Theorem 2.2. Standard Deviation of a Binomial Proportion

We previously noted that the standard deviation of a binomial distribution when there are n trials $\sqrt{np(1-p)}$. What then is the standard deviation for proportions for a binomial distribution? Well, we simply take the formula above for frequencies and divide through by n—that is,

$$\sigma = \frac{\sqrt{np(1-p)}}{n} = \sqrt{\frac{p(1-p)}{n}} \tag{2.7}$$

Proof: To see the equivalence shown in Equation 2.7, just recall from your high school algebra class that $\sqrt{n}/n = \sqrt{n}/\sqrt{n^2} = \sqrt{n/n^2} = \sqrt{1/n}$.

You may wonder why the preceding formula, Equation 2.7, gives us a standard deviation since it does not look anything at all like the customary formula for the standard deviation, σ, given in all statistics texts, including this one, namely,

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}. \tag{1.3}$$

Proof Sketch of Equivalence: For a binomial distribution with success rate p, and thus a mean success of np, we can think of there being two groups in the sample, np of whom have value 1 and $n\ (1 - p)$ of whom have value 0. Substituting these values into Equation 1.3, we get the expression for the numerator inside the square root sign equal to

$$\begin{aligned} np*((1-p)*(1-p) + n*(1-p)*p) &= np*((1-p)^2 + (1-p)*p)) \\ &= np*(1 - 2p + p^2 - p^2 + p) \\ &= np(1-p). \end{aligned}$$

Now we put that expression under a square root sign and divide by the denominator of $\sqrt{n}$ to get

$$\sqrt{\mathrm{np}(1-p)}/\sqrt{n} = \sqrt{\left(\frac{n}{\sqrt{n}}\right)p(1-p)} = \sqrt{\frac{p(1-p)}{n}}.$$

Thus, since, $\sqrt{n}/n = \sqrt{1/\sqrt{n}}$, we have shown that, for binomial proportions

$$\sigma = \sqrt{\frac{p(1-p)}{n}}.$$

QED

Do not copy, post, or distribute

Copyright ©2026 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.