CHAPTER 10

INTRODUCING PROBABILITY.
Randomness and Probability

We call a phenomenon *random* if individual outcomes are uncertain but there is nonetheless a regular distribution of outcomes in a large number of repetitions. The *probability* of any outcome of a random phenomenon is the proportion of times the outcome would occur in a very long series of repetitions.
Probability is a measure of how likely an event is to occur. Match one of the probabilities that follow with each statement of likelihood given. (The probability is usually a more exact measure of likelihood that is the verbal statement.)
0 0.01 0.45 0.50 0.55 0.99 1
a) The event is impossible. It can never occur.
b) This event is certain. It will occur on every trial.
c) The event is very likely, but once in a while it will not occur in a long sequence of trials.
d) This event will occur slightly less often than not.
a) An impossible event has probability 0.
b) A certain event has probability 1.
c) 0.99 would correspond to an event that is very likely but will not occur once in a while in a long sequence of trials.
d) An event with probability 0.45 will occur slightly less often than it occurs.
Probability Models

The *sample space* \( S \) of a random phenomenon is the set of all possible outcomes. An *event* is an outcome or a set of outcomes of a random phenomenon. That is, an event is a subset of the sample space. A *probability model* is a mathematical description of a random phenomenon consisting of two parts: a sample space \( S \) and a way of assigning probabilities to events.
Sample space

In each of the following situations, describe a sample space $S$ for the random phenomenon.

a. A basketball player shoots three free throws. You record the sequence of hits and misses.

b. A basketball player shoots three free throws. You record the number of baskets she makes.
Solution

H = hit and M = miss.
b. \( S = \{0,1,2,3\}\)
Colors of M & Ms

If you draw an M & M candy at random from a bag of the candies, the candy you draw will have one of the seven colors. The probability of drawing each color depends on the proportion of each color among all candies made. Here is the distribution for milk chocolate M & Ms:

<table>
<thead>
<tr>
<th>Color</th>
<th>Purple</th>
<th>Yellow</th>
<th>Red</th>
<th>Orange</th>
<th>Brown</th>
<th>Green</th>
<th>Blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>?</td>
</tr>
</tbody>
</table>
Colors of M & M’s (cont.)

a) What must be the probability of drawing a blue candy?
b) What is the probability that you do not draw a brown candy?
c) What is the probability that the candy you draw is either yellow, orange, or red?
Solution

a. Probability of Blue = \( P(\text{Blue}) = 1 - 0.9 = 0.1 \)
b. \( P(\text{Not Brown}) = 1 - P(\text{Brown}) = 1 - 0.1 = 0.9 \)
c. \( P(\text{Yellow or Orange or Red}) = 0.2 + 0.1 + 0.2 = 0.5 \)
1. The probability $P(A)$ of any event $A$ satisfies $0 \leq P(A) \leq 1$.
2. If $S$ is the sample space in a probability model, then $P(S) = 1$.
3. For any event $A$, $P(A \text{ does not occur}) = 1 - P(A)$.
4. Two events $A$ and $B$ are disjoint if they have no outcomes in common and so can never occur simultaneously. If $A$ and $B$ are disjoint, $P(A \text{ or } B) = P(A) + P(B)$. This is the addition rule for disjoint events.
Although the rules of probability are just basic facts about percents or proportions, we need to be able to use the language of events and their probabilities. Choose an American adult at random. Define two events:
A = the person chosen is obese.
B = the person chosen is overweight, but not obese.
According to the National Center for Health Statistics, \( P(A) = 0.34 \) and \( P(B) = 0.33 \).
a) Explain why events A and B are disjoint.
b) Say in plain language what the event ” A or B ” is. What is \( P(A \text{ or } B) \)?
c) If C is the event that the person chosen has normal weight or less, what is \( P(C) \)?
Solution

a) Event B specifically rules out obese subjects, so there is no overlap with event A.

b) A or B is the event ”The person chosen is overweight or obese”. 

\[ P(A \text{ or } B) = P(A) + P(B) = 0.34 + 0.33 = 0.67. \]

c) \[ P(C) = 1 - P(A \text{ or } B) = 1 - 0.67 = 0.33. \]
A probability model with a finite sample space is called finite. To assign probabilities in a finite model, list the probabilities of all the individual outcomes. These probabilities must be numbers between 0 and 1 that add to exactly 1. The probability of any event is the sum of the probabilities of the outcomes making up the event. Finite probability models are sometimes called discrete probability models.
Weighty behavior

Choose an adult in the United States at random and ask, "How many days per week do you lift weights?" Call the response $X$ for short. Based on a large sample survey, here is a probability model for the answer you will get:

<table>
<thead>
<tr>
<th>Days</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.73</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
</tr>
</tbody>
</table>

a) Verify that this is a legitimate finite probability model.
b) Describe the event $X < 4$ in words. What is $P(X < 4)$?
c) Express the event "lifted weights at least once" in terms of $X$. What is the probability of this event?
a) This is a legitimate probability model because the probabilities sum to 1.
b) The event \( \{ X < 4 \} \) is the event that somebody lifts weights 3 or fewer days per week.
\[
P(X<4) = 0.73 + 0.06 + 0.06 + 0.06 = 0.91.
\]
c) This is the event \( \{ X \geq 1 \} \).
\[
P(X \geq 1) = 1 - P(X = 0) = 1 - 0.73 = 0.27.
\]
Another way:
\[
P(X \geq 1) = P(X =1) + P(X=2)+ \ldots +P(X=7) = 0.27
\]
Continuous Probability Model

A continuous probability model assigns probabilities as areas under a density curve. The area under the curve and above any range of values is the probability of an outcome in that range.
Random numbers

Let $Y$ be a random number between 0 and 1 produced by an idealized random number generator. Find the following probabilities:

a) $P(Y \leq 0.6)$.
b) $P(Y < 0.6)$.
c) $P(0.4 \leq Y \leq 0.8)$. 
Solution

a) $P(Y \leq 0.6) = 0.6.$
b) $P(Y < 0.6) = 0.6.$
c) $P(0.4 \leq Y \leq 0.8) = 0.4.$
Adding random numbers

Generate two random numbers between 0 and 1 and take $X$ to be their sum. The sum $X$ can take any value between 0 and 2. The density curve of $X$ is the triangle shown below.

a) Verify by geometry that the area under the curve is 1.
b) What is the probability that $X$ is less than 1?
c) What is the probability that $X$ is less than 0.5?
Solution

a) The area of a triangle is \[ \frac{bh}{2} = \frac{(2)(1)}{2} = 1. \]
b) \[ P(X < 1) = 0.5. \]
c) \[ P(X < 0.5) = 0.125. \]
Figure
The Medical College Admission Test

The Normal distribution with mean $\mu = 25$ and standard deviation $\sigma = 6.4$ is a good description of the total score on the Medical College Admission Test (MCAT). This is a continuous probability model for the score of a randomly chosen student. Call the score of a randomly chosen student $X$ for short.

a) Write the event ”the student chosen has a score of 35 or higher?" in terms of $X$.
b) Find the probability of this event.
Solution

a) This is $P(X \geq 35)$.
b) $P(X \geq 35) = P(Z \geq \frac{35-25}{6.4}) = P(Z \geq 1.56) = 1 - 0.9406 = 0.0594$. 
Random Variable

A *random variable* is a variable whose value is a numerical outcome of a random phenomenon. The *probability distribution* of a random variable $X$ tells us what values $X$ can take and how to assign probabilities to those values.
Running a mile

A study of 12,000 able-bodied male students at the University of Illinois found that their times for the mile run were approximately Normal with mean 7.11 minutes and standard deviation 0.74 minute. Choose a student at random from this group and call his time for the mile $Y$.

a) Say in words what the meaning of $P(Y \geq 8)$ is. What is this probability?
b) Write the event ”the student could run a mile in less than 6 minutes” in terms of values of the random variable $Y$. What is the probability of this event?
Solution

a) $Y \geq 8$ means the student runs the mile in 8 minutes or more.

\[ P(Y \geq 8) = P(Z \geq \frac{8-7.11}{0.74}) = P(Z \geq 1.20) = 1 - 0.8849 = 0.1151 \]

b) $P(Y < 6) = P(Z < \frac{67.11}{0.74}) = P(Z < 1.50) = 0.0668.$
Birth order

A couple plans to have three children. There are 8 possible arrangements of girls and boys. For example, GGB means the first two children are girls and the third child is a boy. All 8 arrangements are (approximately) equally likely.

a) Write down all 8 arrangements of the sexes of three children. What is the probability of any one of these arrangements?

b) Let $X$ be the number of girls the couple has. What is the probability that $X = 2$?

c) Starting from your work in a), find the distribution of $X$. That is, what values can $X$ take, and what are the probabilities for each value?
Solution

a) BBB, BBG, BGB, GBB, GGB, GBG, BGG, GGG. Each has probability $\frac{1}{8}$.

b) Three of the eight arrangements have two (and only two) girls, so

$$P(X = 2) = \frac{3}{8} = 0.375.$$  

c)

<table>
<thead>
<tr>
<th>Value of X</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>1/8</td>
<td>3/8</td>
<td>3/8</td>
<td>1/8</td>
</tr>
</tbody>
</table>
Binomial Probability Distribution

\[ f(x) = \frac{n!}{x!(n-x)!} p^x (1 - p)^{(n-x)} \]

- \( f(x) \) is the probability of \( x \) successes in \( n \) trials.
- \( n \) is the number of trials.
- \( p \) is the probability of a success on any one trial.
- \( 1 - p \) is the probability of a failure on any one trial.
- \( n! = n(n-1)(n-2)...(3)(2)(1) \).
- \( 0! = 1 \).
More about Binomial Distributions

\[ E(X) = \mu = np. \]
\[ Var(X) = \sigma^2 = np(1 - p). \]
CHAPTER 11

SAMPLING DISTRIBUTIONS.
A *parameter* is a number that describes the population. In statistical practice, the value of a parameter is not known because we cannot examine the entire population. A *statistic* is a number that can be computed from the sample data without making use of any unknown parameters. In practice, we often use a statistic to estimate an unknown parameter.
State whether each boldface number below is a parameter or a statistic.

Your local newspaper contains a large number of advertisements for unfurnished one-bedroom apartments. You choose 10 at random and calculate that their mean monthly rent is $540 and that the standard deviation of their rents is $80.
Solution

Both $540 and $80 are statistics (related to our sample of 10 apartments.)
Indianapolis voters

State whether each boldface number below is a parameter or a statistic.

Voter registration records show that 68% of all voters in Indianapolis are registered as Republicans. To test a random-digit dialing device, you use the device to call 150 randomly chosen residential telephones in Indianapolis. Of the registered voters contacted, 73% are registered Republicans.
68% is a parameter (related to the population of all registered voters in Indianapolis); 73% is a statistic (related to the sample of registered voters among those called).
Law of Large Numbers

Draw observations at random from any population with finite mean \( \mu \). As the number of observations drawn increases, the mean \( \bar{x} \) of the observed values gets closer and closer to the mean \( \mu \) of the population.
The *population distribution* of a variable is the distribution of values of the variable among all the individuals in the population. The *sampling distribution* of a statistic is the distribution of values taken by the statistic in all possible samples of the same size from the same population.
Example: Multiple-choice examination

On a multiple-choice examination, each question has five possible choices, only one of which is correct. A student’s exam score is the average number of questions that he answers correctly. For example, if he answers 3 out of 10 correctly, then his score is \( \frac{3}{10} \) or 0.3.

A student who has not studied at all merely guesses on every exam question. Each of his guesses can be likened to drawing a single observation from an infinite parent population of which 80% are zeros and 20% are ones. Taken together, his examination answers can be viewed as a random sample drawn from this population. The number of correct answers is a binomial random variable with \( p = 0.20 \) (and \( n \) = number of questions), and his exam score can be viewed as the mean of this sample.
If the exam consists of only one question, then 20% of the time his score (the mean of a sample consisting of one observation) will equal one, and 80% of the time it will be zero. If the exam consists of 10 questions, however, his score (the mean of a sample consisting of 10 observations) is not very likely to be either zero or one. The sample mean is more likely to equal 0.1, 0.2, or 0.3, implying one, two, or three correct answers out of 10 questions.
Let’s take it one step further and suppose that the exam consists of 100 questions. Then his score (the mean of a sample consisting of 100 observations) is even less likely to equal zero or one - that could only happen if he answered either 0 out of 100 questions correctly or 100 out of 100 correctly. Moreover, the chance that the sample mean is 0.1 or less is not likely either, since that corresponds to 10 or fewer correct answers. Likewise, the chance that the sample mean is 0.3 or more is equally unlikely since this corresponds to 30 or more correct answers. So we can see in the above situation that as a sample size increases, the dispersion of the sample mean decreases. That conclusion supports the result presented below, which indicates that the variance of the sample mean decreases as \( n \) increases.
Mean and Standard Deviation of a Sample Mean

Suppose that \( \bar{x} \) is the mean of an SRS of size \( n \) drawn from a large population with mean \( \mu \) and standard deviation \( \sigma \). Then the sampling distribution of \( \bar{x} \) has mean \( \mu \) and standard deviation \( \frac{\sigma}{\sqrt{n}} \).
Juan makes a measurement in a chemistry laboratory and records the result in his lab report. The standard deviation of student’s lab measurements is $\sigma = 10$ milligrams. Juan repeats the measurement 4 times and records the mean $\bar{x}$ of his 4 measurements.

a) What is the standard deviation of Juan’s mean result? (That is, if Juan kept on making 4 measurements and averaging them, what would be the standard deviation of all his $\bar{x}$’s?)

b) How many times must Juan repeat the measurement to reduce the standard deviation of $\bar{x}$ to 2? Explain to someone who knows no Statistics the advantage of reporting the average of several measurements rather than the result of a single measurement.
Solution

a) \( \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{4}} = 5 \) mg.

b) Solve \( \frac{\sigma}{\sqrt{n}} = 2 \), or \( \frac{10}{\sqrt{n}} = 2 \), so \( \sqrt{n} = 5 \), or \( n = 25 \).

The average of several measurements is more likely than a single measurement to be close to the population mean.
Central Limit Theorem

Draw an SRS of size $n$ from any population with mean $\mu$ and finite standard deviation $\sigma$. The Central Limit Theorem (CLT) says that when $n$ is large the sampling distribution of the sample mean $\bar{x}$ is approximately Normal:

$$
\bar{x} \text{ is approximately } N \left( \mu, \frac{\sigma}{\sqrt{n}} \right).
$$

The Central Limit Theorem allows us to use Normal probability calculations to answer questions about sample means from many observations.
Multiple-choice examination (again)

To demonstrate the Central Limit Theorem, let’s take 300 samples of size 10 (that is, \( n = 10 \)) from a Binomial Distribution with \( p = 0.20 \), calculate the mean of each of these 300 samples, and draw the histogram of these 300 means. That histogram should approximate the sampling distribution of \( \bar{x} \). Next, let’s repeat the foregoing procedure and draw histograms for 300 sample means from samples of size 25, 50, and 100. As \( n \) increases, the distribution of 300 observations of \( \bar{x} \) comes closer to a Normal distribution.
Sampling distribution \((n = 10)\)
Sampling distribution \((n = 25)\)
Sampling distribution ($n = 50$)

$n=50$ (number of questions)
Sampling distribution ($n = 100$)
Auto accidents

The number of accidents per week at a hazardous intersection varies with mean 2.2 and standard deviation 1.4. This distribution takes only whole-number values, so it is certainly not Normal.

a) Let $\bar{x}$ be the mean number of accidents per week at the intersection during a year (52 weeks). What is the approximate distribution of $\bar{x}$ according to the central limit theorem?

b) What is the approximate probability that $\bar{x}$ is less than 2?

c) What is the approximate probability that there are fewer than 100 accidents at the intersection in a year? (Hint: Restate this event in terms of $\bar{x}$)
Solution

a) By the Central Limit Theorem, $\bar{x}$ is roughly Normal with mean $\mu = 2.2$ and standard deviation $\sigma/\sqrt{n} = 1.4/\sqrt{52} = 0.1941$.

b) $P(\bar{x} < 2) = P(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < \frac{2 - 2.2}{0.1941}) = P(Z < -1.0303) = 0.1515$.

c) Let $x_i$ be the number of accidents during week $i$.

$P(\text{Total} < 100) = P(\sum_{i=1}^{52} x_i < 100) = P\left(\frac{\sum_{i=1}^{52} x_i}{52} < \frac{100}{52}\right) = P(\bar{x} < 1.9230) = P(Z < -1.4270) = 0.0768$
An insurance company knows that in the entire population of millions of apartment owners, the mean annual loss from damage is $\mu = $75 and the standard deviation of the loss is $\sigma = $300. The distribution of losses is strongly right-skewed: most policies have $0$ loss, but a few have large losses. If the company sells 10,000 policies, can it safely base its rates on the assumption that its average loss will be no greater than $85$?
Solution

The Central Limit Theorem says that, in spite of the skewness of the population distribution, the average loss among 10,000 policies will be approximately $\mathcal{N}(75, \frac{300}{\sqrt{10000}}) = \mathcal{N}(75, 3)$.

Now $P(\bar{x} > 85) = P \left( Z > \frac{85 - 75}{3} \right) = P(Z > 3.33) = 1 - 0.9996 = 0.0004$.

We can be about 99.96% certain that average losses will not exceed $85$ per policy.
Another Example

Assume that the average adult weighs 140 pounds and that the standard deviation is 25 pounds. Five people enter an elevator that has a capacity of 750 pounds. What is the chance that their combined weight exceeds capacity?
The chance that their total weight exceeds 750 is the same as the chance that their average weight exceeds $\frac{750}{5} = 150$ pounds. From the CLT, $\bar{x}$ has approximately a Normal Distribution with mean equal to 140 pounds and standard deviation equal to $\frac{25}{\sqrt{5}} = 11.18$. Therefore, the chance of exceeding capacity is approximately equal to

\[ P(\bar{x} > 150) = P \left( Z > \frac{150 - 140}{11.18} \right) \]

\[ = P(Z > 0.89) = 1 - 0.8133 = 0.1867 \]
CHAPTER 14

CONFIDENCE INTERVALS: THE BASICS
If you buy a jumbo bag of Kit Kats snack size, you will find the following information:

Nutrition Facts
Serving Size 3 two-piece bars (42 g).
If each jumbo bag contains 14 servings, the net weight should be
(14)(42 g) = 588 g.
However, the net weight on the label is 569 g.
How can we explain that difference?
Statistical Inference

*Statistical Inference* provides methods for drawing conclusions about a population from sample data.
Simple Conditions for Inference about a mean

1. We have an SRS from the population of interest. There is no nonresponse or other practical difficulty.
2. The variable we measure has an exactly Normal Distribution $N(\mu, \sigma)$ in the population.
3. We don’t know the population mean $\mu$. But we do know the population standard deviation $\sigma$. 
NAEP test scores

Young people have a better chance of good jobs and good wages if they are good with numbers. How strong are the quantitative skills of young Americans of working age? One source of data is the National Assessment of Educational Progress (NAEP) Young Adult Literacy Assessment Survey, which is based on a nationwide probability sample of households.

Suppose that you give the NAEP test to an SRS of 1000 people from a large population in which the scores have mean 280 and standard deviation $\sigma = 60$. The mean $\bar{x}$ of the 1000 scores will vary if you take repeated samples.
NAEP test scores (cont.)

a) The sampling distribution of $\bar{x}$ is approximately Normal. It has mean $\mu = 280$. What is the standard deviation?

b) Sketch the Normal curve that describes how $\bar{x}$ varies in many samples from this population. Mark the mean $\mu = 280$ and the values one, two, and three standard deviations on either side of the mean.

c) According to the 68-95-99.7 rule, about 95% of all the values of $\bar{x}$ fall within ______ of the mean of this curve. What is the missing number? Call it $m$ for ”margin of error”. Shade the region from the mean minus $m$ to the mean plus $m$ on the axis of your sketch.

d) Whenever $\bar{x}$ falls in the region you shaded, the true value of the population mean, $\mu = 280$, lies in the confidence interval between $\bar{x} - m$ and $\bar{x} + m$. Draw the confidence interval below your sketch for one value of $\bar{x}$ inside the shaded region and one value of $\bar{x}$ outside the shaded region.

e) In what percent of all samples will the true mean $\mu = 280$ be covered by the confidence interval $\bar{x} \pm m$?
Solution

a) The standard deviation of $\bar{x}$ is $\frac{\sigma}{\sqrt{1000}} = 1.8974$

b) See below.

c) $m = 2(1.8974) \approx 3.8$

d) The confidence intervals drawn may vary, of course, but they should be $2m = 7.6$ units wide.

e) 95%.
\[ \mu \pm \sigma / \sqrt{n} \]
\[ \mu \pm 2\left(\sigma / \sqrt{n}\right) \]
$\mu \pm 3(\sigma / \sqrt{n})$
A level $C$ confidence interval for a parameter has two parts:

a) An interval calculated from the data, usually of the form

\[
\text{estimate} \pm \text{margin of error}
\]

b) A confidence level $C$, which gives the probability that the interval will capture the true parameter value in repeated samples. That is, the confidence level is the success rate for the method.
Interpreting Confidence Level

The confidence level is the success rate of the method that produces the interval. We don’t know whether the 95% confidence interval from a particular sample is one of the 95% that capture $\mu$ or one of the unlucky 5% that miss.

To say that we are 95% confident that the unknown $\mu$ lies between 26.2 and 27.4 is shorthand for “We got these numbers using a method that gives correct results 95% of the time.”
Confidence Interval for the mean of a Normal Population

Draw an SRS of size $n$ from a Normal population having unknown mean $\mu$ and known standard deviation $\sigma$. A level $C$ confidence interval for $\mu$ is

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

The critical value $z^*$ depends on confidence level $C$ and can be found using table A. At the bottom of table C one can find $z^*$ for some popular confidence levels.
Find a critical value

The critical value $z^*$ for confidence level 85% is not in Table C. Use software or Table A of Standard Normal probabilities to find $z^*$. Include in your answer a sketch with $C = 0.85$ and your critical value $z^*$ marked on the axis.
Measuring conductivity

The National Institute of Standards and Technology (NIST) supplies "standard materials" whose physical properties are supposed to be known. For example, you can buy from NIST an iron rod whose electrical conductivity is supposed to be 10.1 at 293 kelvins. (The units for conductivity are microsiemens per centimeter. Distilled water has conductivity 0.5). Of course, no measurement is exactly correct. NIST knows the variability of its measurements very well, so it is quite realistic to assume that the population of all measurements of the same rod has the Normal distribution with mean \( \mu \) equal to the true conductivity and standard deviation \( \sigma = 0.1 \). Here are 6 measurements on the same standard iron rod, which is supposed to have conductivity 10.1:

10.08 9.89 10.05 10.16 10.21 10.11

NIST wants to give the buyer of this iron rod a 90% confidence interval for its true conductivity. What is this interval?
Solution

We will estimate the true conductivity, $\mu$ (the mean of all measurements of its conductivity), by giving a 90% confidence interval.

The mean of the sample is $\bar{x} = 10.0833$ microsiemens per centimeter. For 90% confidence, the critical value is $z^* = 1.645$ (from Table C). Hence, a 90% confidence interval for $\mu$ is $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$, i.e. $10.0833 \pm 1.645 \frac{0.1}{\sqrt{6}}$, which yields: 10.0161 to 10.1504 microsiemens per centimeter.

We are 90% confident that the iron rod’s true conductivity is between 10.0161 and 10.1504 microsiemens per centimeter.
Problem

In an effort to estimate the mean amount spent per customer for dinner at a major Atlanta restaurant, data were collected for a sample of 49 customers. Assume a population standard deviation of $5.

a. At 95% confidence, what is the margin of error?

b. If the sample mean is $24.80, what is the 95% confidence interval for the population mean?
Solution

In this case, \( \mu = \) mean amount spent per customer for dinner for all customers at a major Atlanta restaurant.

a. margin of error = \( z^* \frac{\sigma}{\sqrt{n}} = 1.96 \left( \frac{\sigma}{\sqrt{49}} \right) = 1.4 \)

b. \( \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} \)

(24.80 - 1.4, 24.80 + 1.4))

(23.40, 26.20) 95% Confidence Interval for \( \mu \).
IQ test scores

Here are the IQ test scores of 31 seventh-grade girls in a Midwest school district:
114 100 104 89 102 91 114 114 103 105
108 130 120 132 111 128 118 119 86 72
111 103 74 112 107 103 98 96 112 112 93

a) These 31 girls are an SRS of all seventh-grade girls in the school district. Suppose that the standard deviation of IQ scores in this population is known to be \( \sigma = 15 \). We expect the distribution of IQ scores to be close to Normal. Make a stemplot of the distribution of these 31 scores (split the stems) to verify that there are no major departures from Normality. You have now checked the ”simple conditions” to the extent possible.

b) Estimate the mean IQ score for all seventh-grade girls in the school district, using a 99% confidence interval.
a) Stemplot

| 7 | 2 4 |
| 7 |
| 8 |
| 8 | 6 9 |
| 9 | 1 3 |
| 10 | 0 2 3 3 3 4 |
| 10 | 5 7 8 |
| 11 | 1 1 2 2 2 4 4 4 |
| 11 | 8 9 |
| 12 | 0 |
| 12 | 8 |
| 13 | 0 2 |

The two low scores (72 and 74) are both possible outliers, but there are no other apparent deviations from Normality.
b) 99% Confidence Interval

We will estimate $\mu$ by giving a 99% confidence interval. With $\bar{x} = 105.8387$, and $z^* = 2.576$ (from Table C), our confidence interval for $\mu$ is given by

$$105.8387 \pm 2.576 \frac{15}{\sqrt{31}}$$

We are 99% confident that the mean IQ of seventh-grade girls in this district is between 98.8987 and 112.7786.
Problem

Playbill magazine reported that the mean annual household income of its readers is $119,155. Assume this estimate of the mean annual household income is based on a sample of 80 households, and based on past studies, the population standard deviation is known to be $\sigma = \$30,000$.

a. Develop a 90% confidence interval estimate of the population mean.

b. Develop a 95% confidence interval estimate of the population mean.

c. Develop a 99% confidence interval estimate of the population mean.
Solution

a. (113, 620.73; 124, 689.26) 90% Confidence Interval for $\mu$.
b. (112, 580.96; 125, 729.03) 95% Confidence Interval for $\mu$.
c. (110501.41; 127,808.58) 99% Confidence Interval for $\mu$. 
Body mass index (BMI) is used to screen for possible weight problems. It is calculated as weight divided by the square of height, measuring weight in kilograms and height in meters. For data about BMI, we turn to the National Health and Nutrition Examination Survey (NHANES), a continuing government sample survey that monitors the health of the American population. An NHANES report gives data for 654 women aged 20 to 29 years. The mean BMI in the sample was $\bar{x} = 26.8$. We treated these data as an SRS from a Normally distributed population with standard deviation $\sigma = 7.5$.

a) Give three confidence intervals for the mean BMI $\mu$ in this population, using 90%, 95%, and 99% confidence.

b) What are the margins of error for 90%, 95%, and 99% confidence? How does increasing the confidence level change the margin of error of a confidence interval when the sample size and population standard deviation remain the same?
Solution

a)

<table>
<thead>
<tr>
<th>Confidence level</th>
<th>$z^*$</th>
<th>Margin of error</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>1.645</td>
<td>0.4824</td>
<td>26.318 to 27.282</td>
</tr>
<tr>
<td>95%</td>
<td>1.96</td>
<td>0.5748</td>
<td>26.225 to 27.375</td>
</tr>
<tr>
<td>99%</td>
<td>2.576</td>
<td>0.7555</td>
<td>26.045 to 27.555</td>
</tr>
</tbody>
</table>

b) The margins of error increase as confidence level increases.
The last problem described NHANES survey data on the body mass index (BMI) of 654 young women. The mean BMI in the sample was $\bar{x} = 26.8$. We treated these data as an SRS from a Normally distributed population with standard deviation $\sigma = 7.5$.

a) Suppose that we had an SRS of just 100 young women. What would be the margin of error for 95% confidence?
b) Find the margins of error for 95% confidence based on SRSs of 400 young women and 1600 young women.
c) Compare the three margins of error. How does increasing the sample size change the margin of error of a confidence interval when the confidence level and population standard deviation remain the same?
Solution

<table>
<thead>
<tr>
<th>n (sample size)</th>
<th>Margin of error</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1.47</td>
</tr>
<tr>
<td>400</td>
<td>0.735</td>
</tr>
<tr>
<td>1600</td>
<td>0.3675</td>
</tr>
</tbody>
</table>

c) Margin of error decreases as $n$ increases (Specifically, every time the sample size $n$ is quadrupled, the margin of error is halved).
CHAPTER 15

TESTS OF SIGNIFICANCE: THE BASICS
Who wants to be a millionaire?

Let’s say that one of you is invited to this popular show. As you probably know, you have to answer a series of multiple choice questions and there are four possible answers to each question. Perhaps, you also have seen that if you don’t know the answer to a question you could either ”jump the question” or you could ”ask the audience”. Suppose that you run into a question for which you don’t know the answer with certainty and you decide to ”ask the audience”. Let’s say that you initially believe that the right answer is A. Then you ask the audience and only 2% of the audience shares your opinion. What would you do? Change your initial belief or reject it?
Measuring conductivity

The National Institute of Standards and Technology (NIST) supplies a "standard iron rod" whose electrical conductivity is supposed to be exactly 10.1. Is there reason to think that the true conductivity is not 10.1? To find out, NIST measures the conductivity of one rod 6 times. Repeated measurements of the same thing vary, which is why NIST makes 6 measurements. These measurements are an SRS from the population of all possible measurements. This population has a Normal distribution with mean \( \mu \) equal to the true conductivity and standard deviation \( \sigma = 0.1 \).
Measuring conductivity

a) We seek evidence against the claim that \( \mu = 10.1 \). What is the sampling distribution of the mean \( \bar{x} \) in many samples of 6 measurements of one rod if the claim is true? Make a sketch of the Normal curve for this distribution. (Draw a Normal curve, then mark on the axis the values of the mean and 1,2, and 3 standard deviations on either side of the mean).

b) Suppose that the sample mean is \( \bar{x} = 10.09 \). Mark this value on the axis of your sketch. Another rod has \( \bar{x} = 9.95 \) for 6 measurements. Mark this value on the axis as well. Explain in simple language why one result is good evidence that the true conductivity differs from 10.1 and why the other result gives no reason to doubt that 10.1 is correct.
Solution

a) If $\mu = 10.1$, then the sampling distribution of $\bar{x}$ is approximately Normal with mean $\mu = 10.1$ and standard deviation 
\[ \frac{\sigma}{\sqrt{n}} = \frac{0.1}{\sqrt{6}} = 0.041. \]
b) The plot provided shows sampling distribution described in a). On the plot are the two indicated values of $\bar{x} = 10.09$ and $\bar{x} = 9.95$. We see that a value of $\bar{x} = 9.95$ would be very unusual (more than 3 standard deviations away from 10.1). Hence, a value of $\bar{x} = 9.95$ would provide reason to doubt that $\mu = 10.1$, whereas $\bar{x} = 10.09$ would not.
Graph
Null and Alternative Hypotheses

The claim tested by a statistical test is called the *null hypothesis*. The test is designed to assess the strength of the evidence against the null hypothesis. Usually the null hypothesis is a statement of "no effect" or "no difference".

The claim about the population that we are trying to find evidence for is the *alternative hypothesis*. The alternative hypothesis is *one-sided* if it states that a parameter is larger than or smaller than the null hypothesis value. It is *two-sided* if it states that the parameter is different from the null value (it could be either smaller or larger).
Measuring conductivity

State the null and alternative hypotheses for the study of electrical conductivity described above. (Is the alternative hypothesis one-sided or two-sided?).
Solution

\[ H_0 : \mu = 10.1 \text{ vs } H_a : \mu \neq 10.1. \text{ This is a two-sided test because we wonder if the conductivity differs from } 10.1. \]
A test statistic calculated from the sample data measures how far the data diverge from what we would expect if the null hypothesis $H_0$ were true. Large values of the statistic show that the data are not consistent with $H_0$.

The probability, computed assuming that $H_0$ is true, that the test statistic would take a value as extreme or more extreme than that actually observed is called the $P$-value of the test. The smaller the $P$-value, the stronger the evidence against $H_0$ provided by the data.
Tests for a population mean

There are four steps in carrying out a significance test:
1. State the hypotheses.
2. Calculate the test statistic.
3. Find the P-value.
4. State your conclusion in the context of your specific setting.

Once you have stated your hypotheses and identified the proper test, you or your calculator can do Steps 2 and 3 by following a recipe. Here is the recipe for the test we will use in our examples.
Z test for a population mean $\mu$

Draw a simple random sample of size $n$ from a Normal population that has unknown mean $\mu$ and known standard deviation $\sigma$. To test the null hypothesis that $\mu$ has a specified value, $H_0 : \mu = \mu_0$ calculate the one-sample $z^*$ statistic

$$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

In terms of a variable $Z$ having the standard Normal distribution, the P-value for a test of $H_0$ against

$H_a : \mu > \mu_0$ is $P(Z \geq z^*)$.

$H_a : \mu < \mu_0$ is $P(Z \leq z^*)$.

$H_a : \mu \neq \mu_0$ is $2P(Z \geq |z^*|)$. 
Here are 6 measurements of the electrical conductivity of an iron rod:
10.08 9.89 10.05 10.16 10.21 10.11
The iron rod is supposed to have conductivity 10.1. Do the measurements give good evidence that the true conductivity is not 10.1?
The 6 measurements are an SRS from the population of all results we would get if we kept measuring conductivity forever. This population has a Normal distribution with mean equal to the true conductivity of the rod and standard deviation 0.1. Use this information to carry out a test of significance.
Solution

Let \( \mu \) be the rod’s true conductivity (the mean of all measurements of its conductivity).

1. State hypotheses. \( H_0 : \mu = 10.1 \) vs \( H_a : \mu \neq 10.1 \).

2. Test statistic. \( z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{10.0833 - 10.1}{0.1 / \sqrt{6}} = -0.4090 \)

3. P-value. (Using Table A) \( 2P(Z \geq |z^*|) = 2P(Z \geq |-0.40|) = 2P(Z \geq 0.40) = 2(1 - 0.6554) = 2(0.3446) = 0.6892 \).

4. Conclusion. This sample gives little reason to doubt that the true conductivity is 10.1. That is, there is virtually no evidence that the true conductivity of the rod differs from 10.1. Random chance easily explains the observed sample mean.
A student group claims that first-year students at a university must study 2.5 hours per night during the school week. A skeptic suspects that they study less than that on the average. A class survey finds that the average study time claimed by 269 students is $\bar{x} = 137$ minutes. Regard these students as a random sample of all first-year students and suppose we know that study times follow a Normal distribution with standard deviation 65 minutes. Carry out a test of $H_0 : \mu = 150$ against $H_a : \mu < 150$. What do you conclude?
Solution

$\mu =$ average study time for all first-year students at this university.

1. State hypotheses. $H_0 : \mu = 150$ min vs $H_a : \mu < 150$ min.

2. Test statistic. $z^* = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{137 - 150}{65/\sqrt{269}} = -3.28$

3. P-value. $P(Z < z^*) = P(Z < -3.28) = 0.0005.$

4. Conclusion. This is very strong evidence that students study less than 2.5 hours per night.
Sweetening colas

Diet colas use artificial sweeteners to avoid sugar. These sweeteners gradually lose their sweetness over time. Manufacturers therefore test new colas for loss of sweetness before marketing them. Trained tasters sip the cola along with drinks of standard sweetness and score the cola on a "sweetness score" of 1 to 10. The cola is then stored for a month at high temperature to imitate the effect of four months storage at room temperature. Each taster scores the cola again after storage. This is a matched pairs experiment. Our data are the differences (score before storage minus score after storage) in the tasters scores. The bigger these differences, the bigger the loss of sweetness.
Suppose we know that for any cola, the sweetness loss scores vary from taster to taster according to a Normal distribution with standard deviation $\sigma = 1$. The mean $\mu$ for all tasters measures loss of sweetness, and is different for different colas.

The following are the sweetness losses for a new cola, as measured by 10 trained tasters: 2.0 0.4 0.7 2.0 -0.4 2.2 -1.3 1.2 1.1 2.3. Are these data good evidence that the cola lost sweetness in storage?
Solution

\( \mu = \) mean sweetness loss for the population of all tasters.

1. State hypotheses. \( H_0 : \mu = 0 \) vs \( H_a : \mu > 0 \).
2. Test statistic. \( z^* = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{1.02 - 0}{1/\sqrt{10}} = 3.23 \)
3. P-value. \( P(Z > z^*) = P(Z > 3.23) = 0.0006 \).
4. Conclusion. We would very rarely observe a sample sweetness loss as large as 1.02 if \( H_0 \) were true. The small P-value provides strong evidence against \( H_0 \) and in favor of the alternative \( H_a : \mu > 0 \), i.e., it gives good evidence that the mean sweetness loss is not 0, but positive.
Statistical Significance

If the P-value is as small or smaller than $\alpha$, we say that the data are *statistically significant at level* $\alpha$. 
Significance from a table

A test of $H_0: \mu = 0$ against $H_a: \mu > 0$ has test statistic $z^* = 1.876$. Is this significant at the 5% level ($\alpha = 0.05$)? Is it significant at the 1% level ($\alpha = 0.01$)?
Solution

$z^* = 1.876$ lies between 1.645 and 1.960 (See Table C). So the P-value lies between the corresponding entries in the ”One-sided P” row, which are $P = 0.05$ and $P = 0.025$, i.e. $0.025 < P – value < 0.05$.

This $z^*$ is significant at the $\alpha = 0.05$ level and is NOT significant at the $\alpha = 0.01$ level.
Significance from a table (again)

A test of $H_0 : \mu = 0$ against $H_a : \mu \neq 0$ has test statistic $z^* = 1.876$. Is this significant at the 5% level ($\alpha = 0.05$)? Is it significant at the 1% level ($\alpha = 0.01$)?
Solution

\[ z^* = 1.876 \] lies between 1.645 and 1.960 (See Table C). So the P-value lies between the corresponding entries in the "Two-sided P" row, which are \( P = 0.10 \) and \( P = 0.05 \), i.e. \( 0.05 < P - \text{value} < 0.10 \).

This \( z^* \) is NOT significant at the \( \alpha = 0.05 \) level and is NOT significant at the \( \alpha = 0.01 \) level.
Testing software

You have computer software that claims to generate observations from a standard Normal distribution. If this is true, the numbers generated come from a population with \( \mu = 0 \) and \( \sigma = 1 \). A command to generate 100 observations gives outcomes with mean \( \bar{x} = -0.2213 \). Assume that the population \( \sigma \) remains fixed. We want to test
\[
H_0 : \mu = 0 \\
H_a : \mu \neq 0
\]
a) Calculate the value of the \( z^* \) test statistic.
b) Use Table C: is \( z^* \) significant at the 5% level (\( \alpha = 0.05 \))? 
c) Use Table C: is \( z^* \) significant at the 1% level (\( \alpha = 0.01 \))? 
d) Between which two Normal critical values \( z \) in the bottom row of Table C does \( z^* \) lie? Between what two numbers does the P-value lie? Does the test give good evidence against the null hypothesis?
Solution

a) \( z^* = \frac{-0.2213-0}{1/\sqrt{100}} = -2.213 \).

b) Compare \( z^* = -2.213 \), well, actually \( |z^*| = 2.213 \), with the \( z^* \) row in Table C. It lies between 2.054 and 2.326. So the P-value lies between the corresponding entries in the ”Two-sided P” row, which are \( P = 0.04 \) and \( P = 0.02 \). Since P-value \( < \alpha = 0.05 \), the test is significant at the \( \alpha = 0.05 \) level.

c) From b), we know that \( 0.02 < \text{P-value} < 0.04 \), which implies that P-value \( > \alpha = 0.01 \). Therefore, the test is not significant at the \( \alpha = 0.01 \) level.

d) See b). The test gives good evidence against the null hypothesis. We can believe this software is not generating values from a standard Normal distribution.
Where the data come from matters

When you use statistical inference, you are acting as if your data are a random sample or come from a randomized comparative experiment.
Cautions about Confidence Intervals

*The margin of error doesn’t cover all errors.* The margin of error in a confidence interval covers only random sampling errors. Practical difficulties such as undercoverage and nonresponse are often more serious than random sampling error. The margin of error does not take such difficulties into account.
Cautions about Significance Tests

How small a $P$ is convincing? The purpose of a test of significance is to describe the degree of evidence provided by the sample against the null hypothesis. The $P$-value does this. But how small a $P$-value is convincing evidence against the null hypothesis? This depends mainly on two circumstances:

a) How plausible is $H_0$? If $H_0$ represents an assumption that the people you must convince have believed for years, strong evidence (small $P$) will be needed to persuade them.

b) What are the consequences of rejecting $H_0$? If rejecting $H_0$ in favor of $H_a$ means making an expensive changeover from one type of product packaging to another, you need strong evidence that the new packaging will boost sales.
Significance depends on the alternative hypothesis. You may have noticed that the P-value for a one-sided test is one-half the P-value for the two-sided test of the same null hypothesis based on the same data. The two-sided P-value combines two equal areas, one in each tail of a Normal curve. The one-sided P-value is just one of these areas, in the direction specified by the alternative hypothesis. It makes sense that the evidence against $H_0$ is stronger when the alternative is one-sided, because it is based on the data plus information about the direction of possible deviations from $H_0$. If you lack this added information, always use a two-sided alternative hypothesis.
Significance depends on sample size. Significance depends both on the size of the effect you observe and on the size of the sample. Understanding this effect is essential to understanding significance tests.
SAMPLE SIZE FOR DESIRED MARGIN OF ERROR. The $Z$ confidence interval for the mean of a Normal population will have a specified margin of error $m$ when the sample size is

$$n = \left( \frac{z^* \sigma}{m} \right)^2$$
Example 14.1 (page 287) assumed that the body mass index (BMI) of all American young women follows a Normal distribution with standard deviation $\sigma = 7.5$. How large a sample would be needed to estimate the mean BMI $\mu$ in this population within $\pm 1$ with 95% confidence?
Solution

\[ n = \left( \frac{(1.96)(7.5)}{1} \right)^2 = 216.09 \]

Take \( n = 217 \).
Number skills of young men

Suppose that scores on the mathematics part of the National Assessment of Educational Progress (NAEP) test for high school seniors follow a Normal distribution with standard deviation $\sigma = 30$. You want to estimate the mean score within $\pm 10$ with 90% confidence. How large an SRS of scores must you choose?
Solution

\[ n = \left( \frac{(1.645)(30)}{10} \right)^2 = 24.354 \]

Take \( n = 25 \).