CHAPTER 16

Random Variables

What Is an Actuary?

Actuaries are the daring people who put a price on risk, estimating the likelihood and costs of rare events, so they can be insured. That takes financial, statistical, and business skills. It also makes them invaluable to many businesses. Actuaries are rather rare themselves; only about 19,000 work in North America. Perhaps because of this, they are well paid. If you're enjoying this course, you may want to look into a career as an actuary. Contact the Society of Actuaries or the Casualty Actuarial Society (who, despite what you may think, did not pay for this blurb).

Insurance companies make bets. They bet that you're going to live a long life. You bet that you're going to die sooner. Both you and the insurance company want the company to stay in business, so it's important to find a "fair price" for your bet. Of course, the right price for *you* depends on many factors, and nobody can predict exactly how long you'll live. But when the company averages over enough customers, it can make reasonably accurate estimates of the amount it can expect to collect on a policy before it has to pay its benefit.

Here's a simple example. An insurance company offers a "death and disability" policy that pays $10,000 when you die or $5000 if you are permanently disabled. It charges a premium of only $50 a year for this benefit. Is the company likely to make a profit selling such a plan? To answer this question, the company needs to know the *probability* that its clients will die or be disabled in any year. From actuarial information like this, the company can calculate the expected value of this policy.

Expected Value: Center

NOTATION ALERT:

The most common letters for random variables are X, Y, and Z. But be cautious: If you see any capital letter, it just might denote a random variable.

We'll want to build a probability model in order to answer the questions about the insurance company's risk. First we need to define a few terms. The amount the company pays out on an individual policy is called a **random variable** because its numeric value is based on the outcome of a random event. We use a capital letter, like X, to denote a random variable. We'll denote a particular value that it can have by the corresponding lowercase letter, in this case x. For the insurance company, x can be $10,000 (if you die that year), $5000 (if you are disabled), or $0 (if neither occurs). Because we can list all the outcomes, we might formally call this random variable a **discrete** random variable. Otherwise, we'd call it a **continuous** random variable. The collection of all the possible values and the probabilities that they occur is called the **probability model** for the random variable.

> **A S** ***Activity:* Random Variables.** Learn more about random variables from this animated tour.

Suppose, for example, that the death rate in any year is 1 out of every 1000 people, and that another 2 out of 1000 suffer some kind of disability. Then we can display the probability model for this insurance policy in a table like this:

Policyholder Outcome	Payout x	Probability $P(X = x)$
Death	10,000	$\frac{1}{1000}$
Disability	5000	$\frac{2}{1000}$
Neither	0	$\frac{997}{1000}$

To see what the insurance company can expect, imagine that it insures exactly 1000 people. Further imagine that, in perfect accordance with the probabilities, 1 of the policyholders dies, 2 are disabled, and the remaining 997 survive the year unscathed. The company would pay \$10,000 to one client and \$5000 to each of 2 clients. That's a total of \$20,000, or an average of 20000/1000 = \$20 per policy. Since it is charging people \$50 for the policy, the company expects to make a profit of \$30 per customer. Not bad!

NOTATION ALERT:

The expected value (or mean) of a random variable is written $E(X)$ or μ.

We can't predict what *will* happen during any given year, but we can say what we *expect* to happen. To do this, we (or, rather, the insurance company) need the probability model. The expected value of a policy is a parameter of this model. In fact, it's the mean. We'll signify this with the notation μ (for population mean) or $E(X)$ for expected value. This isn't an average of some data values, so we won't estimate it. Instead, we assume that the probabilities are known and simply calculate the expected value from them.

How did we come up with \$20 as the expected value of a policy payout? Here's the calculation. As we've seen, it often simplifies probability calculations to think about some (convenient) number of outcomes. For example, we could imagine that we have exactly 1000 clients. Of those, exactly 1 died and 2 were disabled, corresponding to what the probabilities would say.

$$\mu = E(X) = \frac{10{,}000(1) + 5000(2) + 0(997)}{1000}$$

So our expected payout comes to \$20,000, or \$20 per policy.

Instead of writing the expected value as one big fraction, we can rewrite it as separate terms with a common denominator of 1000.

$$\begin{aligned}\mu &= E(X)\\ &= \$10{,}000\left(\frac{1}{1000}\right) + \$5000\left(\frac{2}{1000}\right) + \$0\left(\frac{997}{1000}\right)\\ &= \$20.\end{aligned}$$

How convenient! See the probabilities? For each policy, there's a 1/1000 chance that we'll have to pay \$10,000 for a death and a 2/1000 chance that we'll have to pay \$5000 for a disability. Of course, there's a 997/1000 chance that we won't have to pay anything.

Take a good look at the expression now. It's easy to calculate the **expected value** of a (discrete) random variable—just multiply each possible value by the probability that it occurs, and find the sum:

$$\mu = E(X) = \sum xP(x).$$

Be sure that every possible outcome is included in the sum. And verify that you have a valid probability model to start with—the probabilities should each be between 0 and 1 and should sum to one.

FOR EXAMPLE Love and expected values

On Valentine's Day the *Quiet Nook* restaurant offers a *Lucky Lovers Special* that could save couples money on their romantic dinners. When the waiter brings the check, he'll also bring the four aces from a deck of cards. He'll shuffle them and lay them out face down on the table. The couple will then get to turn one card over. If it's a black ace, they'll owe the full amount, but if it's the ace of hearts, the waiter will give them a \$20 Lucky Lovers discount. If they first turn over the ace of diamonds (hey—at least it's red!), they'll then get to turn over one of the remaining cards, earning a \$10 discount for finding the ace of hearts this time.

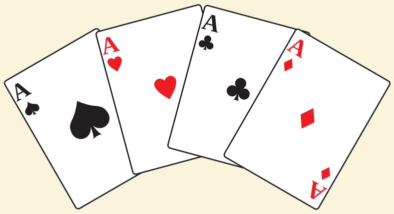
Question: Based on a probability model for the size of the Lucky Lovers discounts the restaurant will award, what's the expected discount for a couple?

Let X = the Lucky Lovers discount. The probabilities of the three outcomes are:

$$P(X = 20) = P(A\heartsuit) = \frac{1}{4}$$

$$P(X = 10) = P(A\diamondsuit\text{, then }A\heartsuit) = P(A\diamondsuit) \cdot P(A\heartsuit | A\diamondsuit) = \frac{1}{4} \cdot \frac{1}{3} = \frac{1}{12}$$

$$P(X = 0) = P(X \neq 20 \text{ or } 10) = 1 - \left(\frac{1}{4} + \frac{1}{12}\right) = \frac{2}{3}.$$



My probability model is:

Outcome	A♥	A♦, then A♥	Black Ace
x	20	10	0
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{2}{3}$

$$E(X) = 20 \cdot \frac{1}{4} + 10 \cdot \frac{1}{12} + 0 \cdot \frac{2}{3} = \frac{70}{12} \approx 5.83$$

Couples dining at the Quiet Nook can expect an average discount of \$5.83.

JUST CHECKING

1. One of the authors took his minivan in for repair recently because the air conditioner was cutting out intermittently. The mechanic identified the problem as dirt in a control unit. He said that in about 75% of such cases, drawing down and then recharging the coolant a couple of times cleans up the problem—and costs only \$60. If that fails, then the control unit must be replaced at an additional cost of \$100 for parts and \$40 for labor.

a) Define the random variable and construct the probability model.
b) What is the expected value of the cost of this repair?
c) What does that mean in this context?

Oh—in case you were wondering—the \$60 fix worked!

First Center, Now Spread . . .

Of course, this expected value (or mean) is not what actually happens to any *particular* policyholder. No individual policy actually costs the company \$20. We are dealing with random events, so some policyholders receive big payouts, others nothing. Because the insurance company must anticipate this variability, it needs to know the *standard deviation* of the random variable.

For data, we calculated the **standard deviation** by first computing the deviation from the mean and squaring it. We do that with (discrete) random variables as well. First, we find the deviation of each payout from the mean (expected value):

Policyholder Outcome	Payout x	Probability $P(X = x)$	Deviation $(x - \mu)$
Death	10,000	$\frac{1}{1000}$	$(10{,}000 - 20) = 9980$
Disability	5000	$\frac{2}{1000}$	$(5000 - 20) = 4980$
Neither	0	$\frac{997}{1000}$	$(0 - 20) = -20$

Next, we square each deviation. The **variance** is the expected value of those squared deviations, so we multiply each by the appropriate probability and sum those products. That gives us the variance of X. Here's what it looks like:

$$Var(X) = 9980^2\left(\frac{1}{1000}\right) + 4980^2\left(\frac{2}{1000}\right) + (-20)^2\left(\frac{997}{1000}\right) = 149{,}600.$$

Finally, we take the square root to get the standard deviation:

$$SD(X) = \sqrt{149{,}600} \approx \$386.78.$$

The insurance company can expect an average payout of \$20 per policy, with a standard deviation of \$386.78.

Think about that. The company charges \$50 for each policy and expects to pay out \$20 per policy. Sounds like an easy way to make \$30. In fact, most of the time (probability 997/1000) the company pockets the entire \$50. But would you consider selling your roommate such a policy? The problem is that occasionally the company loses big. With probability 1/1000, it will pay out \$10,000, and with probability 2/1000, it will pay out \$5000. That may be more risk than you're willing to take on. The standard deviation of \$386.78 gives an indication that it's no sure thing. That's a pretty big spread (and risk) for an average profit of \$30.

Here are the formulas for what we just did. Because these are parameters of our probability model, the variance and standard deviation can also be written as σ^2 and σ. You should recognize both kinds of notation.

$$\sigma^2 = Var(X) = \sum (x - \mu)^2 P(x)$$
$$\sigma = SD(X) = \sqrt{Var(X)}$$

FOR EXAMPLE — Finding the standard deviation

Recap: Here's the probability model for the Lucky Lovers restaurant discount.

Outcome	A♥	A♦, then A♥	Black Ace
x	20	10	0
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{2}{3}$

We found that couples can expect an average discount of $\mu = \$5.83$.

Question: What's the standard deviation of the discounts?

First find the variance: $Var(X) = \sum(x - \mu)^2 \cdot P(x)$

$$= (20 - 5.83)^2 \cdot \frac{1}{4} + (10 - 5.83)^2 \cdot \frac{1}{12} + (0 - 5.83)^2 \cdot \frac{2}{3}$$

$$\approx 74.306.$$

So, $SD(X) = \sqrt{74.306} \approx \$8.62$

Couples can expect the Lucky Lovers discounts to average \$5.83, with a standard deviation of \$8.62.

STEP-BY-STEP EXAMPLE — Expected Values and Standard Deviations for Discrete Random Variables

As the head of inventory for Knowway computer company, you were thrilled that you had managed to ship 2 computers to your biggest client the day the order arrived. You are horrified, though, to find out that someone had restocked refurbished computers in with the new computers in your storeroom. The shipped computers were selected randomly from the 15 computers in stock, but 4 of those were actually refurbished.

If your client gets 2 new computers, things are fine. If the client gets one refurbished computer, it will be sent back at your expense—\$100—and you can replace it. However, if both computers are refurbished, the client will cancel the order this month and you'll lose a total of \$1000.

Question: What's the expected value and the standard deviation of the company's loss?

THINK

Plan State the problem.

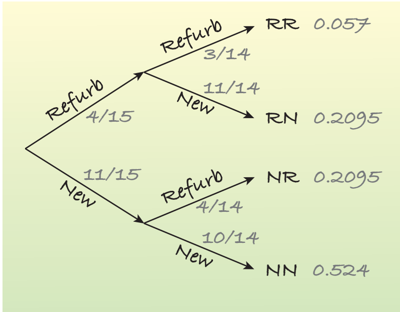
I want to find the company's expected loss for shipping refurbished computers and the standard deviation.

Variable Define the random variable.

Let X = amount of loss.

Plot Make a picture. This is another job for tree diagrams.

If you prefer calculation to drawing, find $P(\mathbf{NN})$ and $P(\mathbf{RR})$, then use the Complement Rule to find $P(\mathbf{NR}\ or\ \mathbf{RN})$.



Model List the possible values of the random variable, and determine the probability model.

Outcome	*x*	$P(X = x)$
Two refurbs	1000	$P(\mathbf{RR}) = 0.057$
One refurb	100	$P(\mathbf{NR} \cup \mathbf{RN}) = 0.2095 + 0.2095 = 0.419$
New/new	0	$P(\mathbf{NN}) = 0.524$

SHOW

Mechanics Find the expected value.

$$E(X) = 0(0.524) + 100(0.419) + 1000(0.057) = \$98.90$$

Find the variance.

$$Var(X) = (0 - 98.90)^2(0.524) + (100 - 98.90)^2(0.419) + (1000 - 98.90)^2(0.057) = 51{,}408.79$$

Find the standard deviation.

$$SD(X) = \sqrt{51{,}408.79} = \$226.735$$

TELL

Conclusion Interpret your results in context.

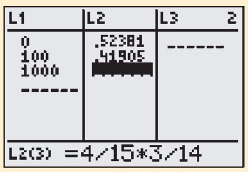
I expect this mistake to cost the firm $98.90, with a standard deviation of $226.74. The large standard deviation reflects the fact that there's a pretty large range of possible losses.

REALITY CHECK

Both numbers seem reasonable. The expected value of $98.90 is between the extremes of $0 and $1000, and there's great variability in the outcome values.

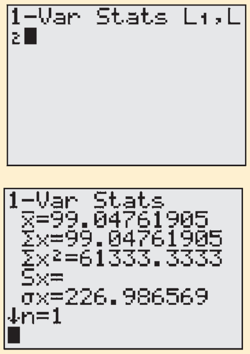
TI Tips

Finding the mean and SD of a random variable



You can easily calculate means and standard deviations for a random variable with your TI. Let's do the Knowway computer example.

- Enter the values of the variable in a list, say, L1: 0, 100, 1000.
- Enter the probability model in another list, say, L2. Notice that you can enter the probabilities as fractions. For example, multiplying along the top branches



of the tree gives the probability of a $1000 loss to be $\frac{4}{15} \cdot \frac{3}{14}$. When you enter that, the TI will automatically calculate the probability as a decimal!

- Under the STAT CALC menu, ask for 1-Var Stats L1,L2.

Now you see the mean and standard deviation (along with some other things). Don't fret that the calculator's mean and standard deviation aren't precisely the same as the ones we found. Such minor differences can arise whenever we round off probabilities to do the work by hand.

Beware: Although the calculator knows enough to call the standard deviation σ, it uses $\bar{x}$ where it should say μ. Make sure you don't make that mistake!

More About Means and Variances

Our insurance company expected to pay out an average of $20 per policy, with a standard deviation of about $387. If we take the $50 premium into account, we see the company makes a profit of 50 − 20 = $30 per policy. Suppose the company lowers the premium by $5 to $45. It's pretty clear that the expected profit also drops an average of $5 per policy, to 45 − 20 = $25.

What about the standard deviation? We know that adding or subtracting a constant from data shifts the mean but doesn't change the variance or standard deviation. The same is true of random variables.[1]

$$E(X \pm c) = E(X) \pm c \qquad Var(X \pm c) = Var(X).$$

FOR EXAMPLE **Adding a constant**

Recap: We've determined that couples dining at the *Quiet Nook* can expect Lucky Lovers discounts averaging $5.83 with a standard deviation of $8.62. Suppose that for several weeks the restaurant has also been distributing coupons worth $5 off any one meal (one discount per table).

Question: If every couple dining there on Valentine's Day brings a coupon, what will be the mean and standard deviation of the total discounts they'll receive?

Let D = total discount (Lucky Lovers plus the coupon); then $D = X + 5$.

$$E(D) = E(X + 5) = E(X) + 5 = 5.83 + 5 = \$10.83$$
$$Var(D) = Var(X + 5) = Var(X) = 8.62^2$$
$$SD(D) = \sqrt{Var(X)} = \$8.62$$

Couples with the coupon can expect total discounts averaging $10.83. The standard deviation is still $8.62.

Back to insurance . . . What if the company decides to double all the payouts—that is, pay $20,000 for death and $10,000 for disability? This would double the average payout per policy and also increase the variability in payouts. We have seen that multiplying or dividing all data values by a constant changes both the mean and the standard deviation by the same factor. Variance, being the square of standard deviation, changes by the square of the constant. The same is true of random variables. In general, multiplying each value of a random variable by a

[1] The rules in this section are true for both discrete *and* continuous random variables.

constant multiplies the mean by that constant and the variance by the *square* of the constant.

$$E(aX) = aE(X) \qquad Var(aX) = a^2Var(X)$$

FOR EXAMPLE **Double the love**

Recap: On Valentine's Day at the *Quiet Nook*, couples may get a Lucky Lovers discount averaging $5.83 with a standard deviation of $8.62. When two couples dine together on a single check, the restaurant doubles the discount offer—$40 for the ace of hearts on the first card and $20 on the second.

Question: What are the mean and standard deviation of discounts for such foursomes?

$$E(2X) = 2E(X) = 2(5.83) = \$11.66$$

$$Var(2x) = 2^2Var(x) = 2^2 \cdot 8.62^2 = 297.2176$$

$$SD(2X) = \sqrt{297.2176} = \$17.24$$

If the restaurant doubles the discount offer, two couples dining together can expect to save an average of $11.66 with a standard deviation of $17.24.

This insurance company sells policies to more than just one person. How can we figure means and variances for a collection of customers? For example, how can the company find the total expected value (and standard deviation) of policies taken over all policyholders? Consider a simple case: just two customers, Mr. Ecks and Ms. Wye. With an expected payout of $20 on each policy, we might predict a total of $20 + $20 = $40 to be paid out on the two policies. Nothing surprising there. The expected value of the sum is the sum of the expected values.

$$E(X + Y) = E(X) + E(Y).$$

The variability is another matter. Is the risk of insuring two people the same as the risk of insuring one person for twice as much? We wouldn't expect both clients to die or become disabled in the same year. Because we've spread the risk, the standard deviation should be smaller. Indeed, this is the fundamental principle behind insurance. By spreading the risk among many policies, a company can keep the standard deviation quite small and predict costs more accurately.

But how much smaller is the standard deviation of the sum? It turns out that, if the random variables are independent, there is a simple Addition Rule for variances: *The variance of the sum of two independent random variables is the sum of their individual variances.*

For Mr. Ecks and Ms. Wye, the insurance company can expect their outcomes to be independent, so (using X for Mr. Ecks's payout and Y for Ms. Wye's)

$$\begin{aligned} Var(X + Y) &= Var(X) + Var(Y) \\ &= 149{,}600 + 149{,}600 \\ &= 299{,}200. \end{aligned}$$

If they had insured only Mr. Ecks for twice as much, there would only be one outcome rather than two *independent* outcomes, so the variance would have been

$$Var(2X) = 2^2Var(X) = 4 \times 149{,}600 = 598{,}400, \text{ or}$$

twice as big as with two independent policies.

Of course, variances are in squared units. The company would prefer to know standard deviations, which are in dollars. The standard deviation of the payout for two independent policies is $\sqrt{299{,}200} = \$546.99$. But the standard deviation of the payout for a single policy of twice the size is $\sqrt{598{,}400} = \$773.56$, or about 40% more.

If the company has two customers, then, it will have an expected annual total payout of \$40 with a standard deviation of about \$547.

FOR EXAMPLE **Adding the discounts**

Recap: The Valentine's Day Lucky Lovers discount for couples averages \$5.83 with a standard deviation of \$8.62. We've seen that if the restaurant doubles the discount offer for two couples dining together on a single check, they can expect to save \$11.66 with a standard deviation of \$17.24. Some couples decide instead to get separate checks and pool their two discounts.

Question: You and your amour go to this restaurant with another couple and agree to share any benefit from this promotion. Does it matter whether you pay separately or together?

Let X_1 and X_2 represent the two separate discounts, and T the total; then $T = X_1 + X_2$.

$$E(T) = E(X_1 + X_2) = E(X_1) + E(X_2) = 5.83 + 5.83 = \$11.66,$$

so the expected saving is the same either way.

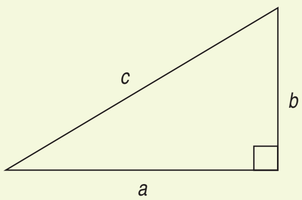
The cards are reshuffled for each couple's turn, so the discounts couples receive are independent. It's okay to add the variances:

$$Var(T) = Var(X_1 + X_2) = Var(X_1) + Var(X_2) = 8.62^2 + 8.62^2 = 148.6088$$
$$SD(T) = \sqrt{148.6088} = \$12.19$$

When two couples get separate checks, there's less variation in their total discount. The standard deviation is \$12.19, compared to \$17.24 for couples who play for the double discount on a single check. It does, therefore, matter whether they pay separately or together.

Pythagorean Theorem of Statistics

We often use the standard deviation to measure variability, but when we add independent random variables, we use their variances. Think of the Pythagorean Theorem. In a right triangle (only), the *square* of the length of the hypotenuse is the sum of the *squares* of the lengths of the other two sides:



$$c^2 = a^2 + b^2.$$

For independent random variables (only), the *square* of the standard deviation of their sum is the sum of the *squares* of their standard deviations:

$$SD^2(X + Y) = SD^2(X) + SD^2(Y).$$

It's simpler to write this with *variances:*

For independent random variables, X and Y, $Var(X + Y) = Var(X) + Var(Y)$.

In general,

- *The mean of the sum of two random variables is the sum of the means.*
- *The mean of the difference of two random variables is the difference of the means.*
- *If the random variables are independent, the variance of their sum or difference is always the sum of the variances.*

$$E(X \pm Y) = E(X) \pm E(Y) \qquad Var(X \pm Y) = Var(X) + Var(Y)$$

Wait a minute! Is that third part correct? Do we always *add* variances? Yes. Think about the two insurance policies. Suppose we want to know the mean and standard deviation of the *difference* in payouts to the two clients. Since each policy has an expected payout of \$20, the expected difference is $20 - 20 = \$0$. If we also subtract variances, we get \$0, too, and that surely doesn't make sense. Note that if the outcomes for the two clients are independent, the difference in payouts could range from $\$10{,}000 - \$0 = \$10{,}000$ to $\$0 - \$10{,}000 = -\$10{,}000$, a spread of \$20,000. The variability in differences increases as much as the variability in sums. If the company has two customers, the difference in payouts has a mean of \$0 and a standard deviation of about \$547 (again).

FOR EXAMPLE Working with differences

Recap: The Lucky Lovers discount at the *Quiet Nook* averages $5.83 with a standard deviation of $8.62. Just up the street, the *Wise Fool* restaurant has a competing Lottery of Love promotion. There a couple can select a specially prepared chocolate from a large bowl and unwrap it to learn the size of their discount. The restaurant's manager says the discounts vary with an average of $10.00 and a standard deviation of $15.00.

Question: How much more can you expect to save at the *Wise Fool*? With what standard deviation?

Let W = discount at the Wise Fool, X = the discount at the Quiet Nook, and D = the difference: $D = W - X$. These are different promotions at separate restaurants, so the outcomes are independent.

$$
\begin{aligned}
E(W - X) &= E(W) - E(X) = 10.00 - 5.83 = \$4.17 \\
SD(W - X) &= \sqrt{Var(W - X)} \\
&= \sqrt{Var(W) + Var(X)} \\
&= \sqrt{15^2 + 8.62^2} \\
&\approx \$17.30
\end{aligned}
$$

Discounts at the Wise Fool will average $4.17 more than at the Quiet Nook, with a standard deviation of $17.30.

For random variables, does $X + X + X = 3X$? Maybe, but be careful. As we've just seen, insuring one person for $30,000 is not the same risk as insuring three people for $10,000 each. When each instance represents a different outcome for the same random variable, it's easy to fall into the trap of writing all of them with the same symbol. Don't make this common mistake. Make sure you write each instance as a *different* random variable. Just because each random variable describes a similar situation doesn't mean that each random outcome will be the same.

These are *random* variables, not the variables you saw in Algebra. Being random, they take on different values each time they're evaluated. So what you really mean is $X_1 + X_2 + X_3$. Written this way, it's clear that the sum shouldn't necessarily equal 3 times *anything*.

FOR EXAMPLE Summing a series of outcomes

Recap: The *Quiet Nook*'s Lucky Lovers promotion offers couples discounts averaging $5.83 with a standard deviation of $8.62. The restaurant owner is planning to serve 40 couples on Valentine's Day.

Question: What's the expected total of the discounts the owner will give? With what standard deviation?

Let $X_1, X_2, X_3, \ldots, X_{40}$ represent the discounts to the 40 couples, and T the total of all the discounts. Then:

$$
\begin{aligned}
T &= X_1 + X_2 + X_3 + \cdots + X_{40} \\
E(T) &= E(X_1 + X_2 + X_3 + \cdots + X_{40}) \\
&= E(X_1) + E(X_2) + E(X_3) + \cdots + E(X_{40}) \\
&= 5.83 + 5.83 + 5.83 + \cdots + 5.83 \\
&= \$233.20
\end{aligned}
$$

Reshuffling cards between couples makes the discounts independent, so:

$$
\begin{aligned}
SD(T) &= \sqrt{Var(X_1 + X_2 + X_3 + \cdots + X_{40})} \\
&= \sqrt{Var(X_1) + Var(X_2) + Var(X_3) + \cdots + Var(X_{40})} \\
&= \sqrt{8.62^2 + 8.62^2 + 8.62^2 + \cdots + 8.62^2} \\
&\approx \$54.52
\end{aligned}
$$

The restaurant owner can expect the 40 couples to win discounts totaling $233.20, with a standard deviation of $54.52.

JUST CHECKING

2. Suppose the time it takes a customer to get and pay for seats at the ticket window of a baseball park is a random variable with a mean of 100 seconds and a standard deviation of 50 seconds. When you get there, you find only two people in line in front of you.

a) How long do you expect to wait for your turn to get tickets?
b) What's the standard deviation of your wait time?
c) What assumption did you make about the two customers in finding the standard deviation?

STEP-BY-STEP EXAMPLE — Hitting the Road: Means and Variances

You're planning to spend next year wandering through the mountains of Kyrgyzstan. You plan to sell your used SUV so you can purchase an off-road Honda motor scooter when you get there. Used SUVs of the year and mileage of yours are selling for a mean of $6940 with a standard deviation of $250. Your research shows that scooters in Kyrgyzstan are going for about 65,000 Kyrgyzstan som with a standard deviation of 500 som. One U.S. dollar is worth about 38.5 Kyrgyzstan som (38 som and 50 tylyn).

Question: How much cash can you expect to pocket after you sell your SUV and buy the scooter?

THINK

Plan State the problem.

I want to model how much money I'd have (in som) after selling my SUV and buying the scooter.

Variables Define the random variables.

Let A = sale price of my SUV (in dollars),
B = price of a scooter (in som), and
D = profit (in som)

Write an appropriate equation.
Think about the assumptions.

$D = 38.5A - B$

✔ **Independence Assumption:** The prices are independent.

SHOW

Mechanics Find the expected value, using the appropriate rules.

$$\begin{aligned} E(D) &= E(38.5A - B) \\ &= 38.5E(A) - E(B) \\ &= 38.5(6{,}940) - (65{,}000) \\ E(D) &= 202{,}190 \text{ som} \end{aligned}$$

Find the variance, using the appropriate rules. Be sure to check the assumptions first!	Since sale and purchase prices are independent, $Var(D) = Var(38.5A - B)$ $= Var(38.5A) + Var(B)$ $= (38.5)^2 Var(A) + Var(B)$ $= 1482.25(250)^2 + (500)^2$ $Var(D) = 92{,}890{,}625$
Find the standard deviation.	$SD(D) = \sqrt{92{,}890{,}625} = 9637.98$ som
TELL **Conclusion** Interpret your results in context. (Here that means talking about dollars.)	I can expect to clear about 202,190 som ($5252) with a standard deviation of 9638 som ($250).
REALITY CHECK Given the initial cost estimates, the mean and standard deviation seem reasonable.	

Continuous Random Variables

A S ***Activity:*** **Numeric Outcomes.** You've seen how to simulate discrete random outcomes. There's a tool for simulating continuous outcomes, too.

A company manufactures small stereo systems. At the end of the production line, the stereos are packaged and prepared for shipping. Stage 1 of this process is called "packing." Workers must collect all the system components (a main unit, two speakers, a power cord, an antenna, and some wires), put each in plastic bags, and then place everything inside a protective styrofoam form. The packed form then moves on to Stage 2, called "boxing." There, workers place the form and a packet of instructions in a cardboard box, close it, then seal and label the box for shipping.

The company says that times required for the packing stage can be described by a Normal model with a mean of 9 minutes and standard deviation of 1.5 minutes. The times for the boxing stage can also be modeled as Normal, with a mean of 6 minutes and standard deviation of 1 minute.

This is a common way to model events. Do our rules for random variables apply here? What's different? We no longer have a list of discrete outcomes, with their associated probabilities. Instead, we have **continuous random variables** that can take on any value. Now any single value won't have a probability. We saw this back in Chapter 6 when we first saw the Normal model (although we didn't talk then about "random variables" or "probability"). We know that the probability that $z = 1.5$ doesn't make sense, but we *can* talk about the probability that z lies *between* 0.5 and 1.5. For a Normal random variable, the probability that it falls within an interval is just the area under the Normal curve over that interval.

A S ***Activity:*** **Means of Random Variables.** Experiment with continuous random variables to learn how their expected values behave.

Some continuous random variables have Normal models; others may be skewed, uniform, or bimodal. Regardless of shape, all continuous random variables have means (which we also call *expected values*) and variances. In this book we won't worry about how to calculate them, but we can still work with models for continuous random variables when we're given these parameters.

The good news is that nearly everything we've said about how discrete random variables behave is true of continuous random variables, as well. When two independent continuous random variables have Normal models, so does their sum or difference. This simple fact is a special property of Normal models and is very important. It allows us to apply our knowledge of Normal probabilities to questions about the sum or difference of independent random variables.

STEP-BY-STEP EXAMPLE

Packaging Stereos

Consider the company that manufactures and ships small stereo systems that we just discussed.

Recall that times required to pack the stereos can be described by a Normal model with a mean of 9 minutes and standard deviation of 1.5 minutes. The times for the boxing stage can also be modeled as Normal, with a mean of 6 minutes and standard deviation of 1 minute.

Questions:

1. What is the probability that packing two consecutive systems takes over 20 minutes?
2. What percentage of the stereo systems take longer to pack than to box?

Question 1: What is the probability that packing two consecutive systems takes over 20 minutes?

THINK

Plan State the problem.

I want to estimate the probability that packing two consecutive systems takes over 20 minutes.

Variables Define your random variables.

Let P_1 = time for packing the first system
P_2 = time for packing the second
T = total time to pack two systems

Write an appropriate equation.

$T = P_1 + P_2$

Think about the assumptions. Sums of independent Normal random variables follow a Normal model. Such simplicity isn't true in general.

✔ **Normal Model Assumption:** We are told that both random variables follow Normal models.

✔ **Independence Assumption:** We can reasonably assume that the two packing times are independent.

SHOW

Mechanics Find the expected value.

$$
\begin{aligned}
E(T) &= E(P_1 + P_2) \\
&= E(P_1) + E(P_2) \\
&= 9 + 9 = 18 \text{ minutes}
\end{aligned}
$$

For sums of independent random variables, variances add. (We don't need the variables to be Normal for this to be true—just independent.)

Since the times are independent,

$$
\begin{aligned}
Var(T) &= Var(P_1 + P_2) \\
&= Var(P_1) + Var(P_2) \\
&= 1.5^2 + 1.5^2 \\
Var(T) &= 4.50
\end{aligned}
$$

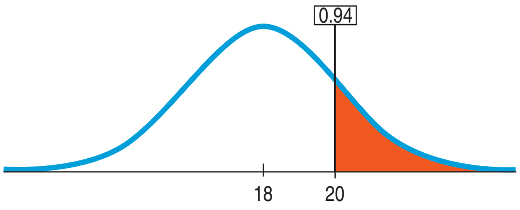
Find the standard deviation.

$$SD(T) = \sqrt{4.50} \approx 2.12 \text{ minutes}$$

Now we use the fact that both random variables follow Normal models to say that their sum is also Normal.

I'll model T with $N(18, 2.12)$.

Sketch a picture of the Normal model for the total time, shading the region representing over 20 minutes.



Find the z-score for 20 minutes.

$$z = \frac{20 - 18}{2.12} = 0.94$$

Use technology or Table Z to find the probability.

$$P(T > 20) = P(z > 0.94) = 0.1736$$

TELL

Conclusion Interpret your result in context.

There's a little more than a 17% chance that it will take a total of over 20 minutes to pack two consecutive stereo systems.

Question 2: What percentage of the stereo systems take longer to pack than to box?

THINK

Plan State the question.

I want to estimate the percentage of the stereo systems that take longer to pack than to box.

Variables Define your random variables.

Let P = time for packing a system
B = time for boxing a system
D = difference in times to pack and box a system

Write an appropriate equation.

$D = P - B$

What are we trying to find? Notice that we can tell which of two quantities is greater by subtracting and asking whether the difference is positive or negative.

The probability that it takes longer to pack than to box a system is the probability that the difference $P - B$ is greater than zero.

Don't forget to think about the assumptions.

✔ **Normal Model Assumption:** We are told that both random variables follow Normal models.

✔ **Independence Assumption:** We can assume that the times it takes to pack and to box a system are independent.

SHOW

Mechanics Find the expected value.

$$\begin{aligned} E(D) &= E(P - B) \\ &= E(P) - E(B) \\ &= 9 - 6 = 3 \text{ minutes} \end{aligned}$$

For the difference of independent random variables, variances add.

Since the times are independent,

$$Var(D) = Var(P - B)$$
$$= Var(P) + Var(B)$$
$$= 1.5^2 + 1^2$$
$$Var(D) = 3.25$$

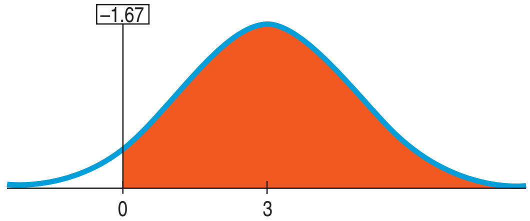
Find the standard deviation.

$$SD(D) = \sqrt{3.25} \approx 1.80 \text{ minutes}$$

State what model you will use.

I'll model D with $N(3, 1.80)$

Sketch a picture of the Normal model for the difference in times, and shade the region representing a difference greater than zero.



Find the z-score for 0 minutes, then use Table Z or technology to find the probability.

$$z = \frac{0 - 3}{1.80} = -1.67$$
$$P(D > 0) = P(z > -1.67) = 0.9525$$

TELL **Conclusion** Interpret your result in context.

About 95% of all the stereo systems will require more time for packing than for boxing.

WHAT CAN GO WRONG?

- **Probability models are still just models.** Models can be useful, but they are not reality. Think about the assumptions behind your models. Are your dice really perfectly fair? (They are probably pretty close.) But when you hear that the probability of a nuclear accident is 1/10,000,000 per year, is that likely to be a precise value? Question probabilities as you would data.
- **If the model is wrong, so is everything else.** Before you try to find the mean or standard deviation of a random variable, check to make sure the probability model is reasonable. As a start, the probabilities in your model should add up to 1. If not, you may have calculated a probability incorrectly or left out a value of the random variable. For instance, in the insurance example, the description mentions only death and disability. Good health is by far the most likely outcome, not to mention the best for both you and the insurance company (who gets to keep your money). Don't overlook that.
- **Don't assume everything's Normal.** Just because a random variable is continuous or you happen to know a mean and standard deviation doesn't mean that a Normal model will be useful. You must *Think* about whether the **Normality Assumption** is justified. Using a Normal model when it really does not apply will lead to wrong answers and misleading conclusions.

To find the expected value of the sum or difference of random variables, we simply add or subtract means. Center is easy; spread is trickier. Watch out for some common traps.

- **Watch out for variables that aren't independent.** You can add expected values of *any* two random variables, but you can only add variances of independent random variables. Suppose a survey includes questions about the number of hours of sleep people get each night and also the number of hours they are awake each day. From their answers, we find the mean and standard deviation of hours asleep and hours awake. The expected total must be 24 hours; after all, people are either asleep or awake.[2] The means still add just fine. Since all the totals are exactly 24 hours, however, the standard deviation of the total will be 0. We can't add variances here because the number of hours you're awake depends on the number of hours you're asleep. Be sure to check for independence before adding variances.
- **Don't forget: Variances of independent random variables add. Standard deviations don't.**
- **Don't forget: Variances of independent random variables add, even when you're looking at the difference between them.**
- **Don't write independent instances of a random variable with notation that looks like they are the same variables.** Make sure you write each instance as a different random variable. Just because each random variable describes a similar situation doesn't mean that each random outcome will be the same. These are *random* variables, not the variables you saw in Algebra. Write $X_1 + X_2 + X_3$ rather than $X + X + X$.

CONNECTIONS

We've seen means, variances, and standard deviations of data. We know that they estimate parameters of models for these data. Now we're looking at the probability models directly. We have only parameters because there are no data to summarize.

It should be no surprise that expected values and standard deviations adjust to shifts and changes of units in the same way as the corresponding data summaries. The fact that we can add variances of independent random quantities is fundamental and will explain why a number of statistical methods work the way they do.

WHAT HAVE WE LEARNED?

We've learned to work with random variables. We can use the probability model for a discrete random variable to find its expected value and its standard deviation.

We've learned that the mean of the sum or difference of two random variables, discrete or continuous, is just the sum or difference of their means. And we've learned the Pythagorean Theorem of Statistics: *For independent random variables*, the variance of their sum or difference is always the *sum* of their variances.

Finally, we've learned that Normal models are once again special. Sums or differences of Normally distributed random variables also follow Normal models.

[2] Although some students do manage to attain a state of consciousness somewhere between sleeping and wakefulness during Statistics class.

Terms

Random variable 366. A random variable assumes any of several different numeric values as a result of some random event. Random variables are denoted by a capital letter such as X.

Discrete random variable 366. A random variable that can take one of a finite number[3] of distinct outcomes is called a discrete random variable.

Continuous random variable 366, 367. A random variable that can take any numeric value within a range of values is called a continuous random variable. The range may be infinite or bounded at either or both ends.

Probability model 366. The probability model is a function that associates a probability P with each value of a discrete random variable X, denoted $P(X = x)$, or with any interval of values of a continuous random variable.

Expected value 367. The expected value of a random variable is its theoretical long-run average value, the center of its model. Denoted μ or $E(X)$, it is found (if the random variable is discrete) by summing the products of variable values and probabilities:

$$\mu = E(X) = \sum xP(x).$$

Variance 369. The variance of a random variable is the expected value of the squared deviation from the mean. For discrete random variables, it can be calculated as:

$$\sigma^2 = Var(X) = \sum (x - \mu)^2 P(x).$$

Standard deviation 369. The standard deviation of a random variable describes the spread in the model, and is the square root of the variance:

$$\sigma = SD(X) = \sqrt{Var(X)}.$$

Changing a random variable by a constant:

372. $E(X \pm c) = E(X) \pm c$ $\quad Var(X \pm c) = Var(X)$

373. $E(aX) = aE(X)$ $\quad Var(aX) = a^2 Var(X)$

Adding or subtracting random variables:

373. $E(X \pm Y) = E(X) \pm E(Y)$

374. *If X and Y are independent,* $Var(X \pm Y) = Var(X) + Var(Y)$.

374. (The Pythagorean Theorem of Statistics)

Skills

THINK

- Be able to recognize random variables.
- Understand that random variables must be independent in order to determine the variability of their sum or difference by adding variances.

SHOW

- Be able to find the probability model for a discrete random variable.
- Know how to find the mean (expected value) and the variance of a random variable.
- Always use the proper notation for these population parameters: μ or $E(X)$ for the mean, and σ, $SD(X)$, σ^2, or $Var(X)$ when discussing variability.
- Know how to determine the new mean and standard deviation after adding a constant, multiplying by a constant, or adding or subtracting two independent random variables.

TELL

- Be able to interpret the meaning of the expected value and standard deviation of a random variable in the proper context.

[3] Technically, there could be an infinite number of outcomes, as long as they're *countable*. Essentially that means we can imagine listing them all in order, like the counting numbers 1, 2, 3, 4, 5, . . .

RANDOM VARIABLES ON THE COMPUTER

Statistics packages deal with data, not with random variables. Nevertheless, the calculations needed to find means and standard deviations of random variables are little more than weighted means. Most packages can manage that, but then they are just being overblown calculators. For technological assistance with these calculations, we recommend you pull out your calculator.

EXERCISES

1. **Expected value.** Find the expected value of each random variable:

a)

x	10	20	30
$P(X=x)$	0.3	0.5	0.2

b)

x	2	4	6	8
$P(X=x)$	0.3	0.4	0.2	0.1

2. **Expected value.** Find the expected value of each random variable:

a)

x	0	1	2
$P(X=x)$	0.2	0.4	0.4

b)

x	100	200	300	400
$P(X=x)$	0.1	0.2	0.5	0.2

3. **Pick a card, any card.** You draw a card from a deck. If you get a red card, you win nothing. If you get a spade, you win $5. For any club, you win $10 plus an extra $20 for the ace of clubs.
   a) Create a probability model for the amount you win.
   b) Find the expected amount you'll win.
   c) What would you be willing to pay to play this game?

4. **You bet!** You roll a die. If it comes up a 6, you win $100. If not, you get to roll again. If you get a 6 the second time, you win $50. If not, you lose.
   a) Create a probability model for the amount you win.
   b) Find the expected amount you'll win.
   c) What would you be willing to pay to play this game?

5. **Kids.** A couple plans to have children until they get a girl, but they agree that they will not have more than three children even if all are boys. (Assume boys and girls are equally likely.)
   a) Create a probability model for the number of children they might have.
   b) Find the expected number of children.
   c) Find the expected number of boys they'll have.

6. **Carnival.** A carnival game offers a $100 cash prize for anyone who can break a balloon by throwing a dart at it. It costs $5 to play, and you're willing to spend up to $20 trying to win. You estimate that you have about a 10% chance of hitting the balloon on any throw.
   a) Create a probability model for this carnival game.
   b) Find the expected number of darts you'll throw.
   c) Find your expected winnings.

7. **Software.** A small software company bids on two contracts. It anticipates a profit of $50,000 if it gets the larger contract and a profit of $20,000 on the smaller contract. The company estimates there's a 30% chance it will get the larger contract and a 60% chance it will get the smaller contract. Assuming the contracts will be awarded independently, what's the expected profit?

8. **Racehorse.** A man buys a racehorse for $20,000 and enters it in two races. He plans to sell the horse afterward, hoping to make a profit. If the horse wins both races, its value will jump to $100,000. If it wins one of the races, it will be worth $50,000. If it loses both races, it will be worth only $10,000. The man believes there's a 20% chance that the horse will win the first race and a 30% chance it will win the second one. Assuming that the two races are independent events, find the man's expected profit.

9. **Variation 1.** Find the standard deviations of the random variables in Exercise 1.

10. **Variation 2.** Find the standard deviations of the random variables in Exercise 2.

11. **Pick another card.** Find the standard deviation of the amount you might win drawing a card in Exercise 3.

12. **The die.** Find the standard deviation of the amount you might win rolling a die in Exercise 4.

13. **Kids again.** Find the standard deviation of the number of children the couple in Exercise 5 may have.

14. **Darts.** Find the standard deviation of your winnings throwing darts in Exercise 6.

15. **Repairs.** The probability model below describes the number of repair calls that an appliance repair shop may receive during an hour.

Repair Calls	0	1	2	3
Probability	0.1	0.3	0.4	0.2

   a) How many calls should the shop expect per hour?
   b) What is the standard deviation?

16. Red lights. A commuter must pass through five traffic lights on her way to work and will have to stop at each one that is red. She estimates the probability model for the number of red lights she hits, as shown below.

X = # of red	0	1	2	3	4	5
$P(X = x)$	0.05	0.25	0.35	0.15	0.15	0.05

a) How many red lights should she expect to hit each day?
b) What's the standard deviation?

17. Defects. A consumer organization inspecting new cars found that many had appearance defects (dents, scratches, paint chips, etc.). While none had more than three of these defects, 7% had three, 11% two, and 21% one defect. Find the expected number of appearance defects in a new car and the standard deviation.

18. Insurance. An insurance policy costs \$100 and will pay policyholders \$10,000 if they suffer a major injury (resulting in hospitalization) or \$3000 if they suffer a minor injury (resulting in lost time from work). The company estimates that each year 1 in every 2000 policyholders may have a major injury, and 1 in 500 a minor injury only.
a) Create a probability model for the profit on a policy.
b) What's the company's expected profit on this policy?
c) What's the standard deviation?

19. Cancelled flights. Mary is deciding whether to book the cheaper flight home from college after her final exams, but she's unsure when her last exam will be. She thinks there is only a 20% chance that the exam will be scheduled after the last day she can get a seat on the cheaper flight. If it is and she has to cancel the flight, she will lose \$150. If she can take the cheaper flight, she will save \$100.
a) If she books the cheaper flight, what can she expect to gain, on average?
b) What is the standard deviation?

20. Day trading. An option to buy a stock is priced at \$200. If the stock closes above 30 on May 15, the option will be worth \$1000. If it closes below 20, the option will be worth nothing, and if it closes between 20 and 30 (inclusively), the option will be worth \$200. A trader thinks there is a 50% chance that the stock will close in the 20–30 range, a 20% chance that it will close above 30, and a 30% chance that it will fall below 20 on May 15.
a) Should she buy the stock option?
b) How much does she expect to gain?
c) What is the standard deviation of her gain?

21. Contest. You play two games against the same opponent. The probability you win the first game is 0.4. If you win the first game, the probability you also win the second is 0.2. If you lose the first game, the probability that you win the second is 0.3.
a) Are the two games independent? Explain.
b) What's the probability you lose both games?
c) What's the probability you win both games?
d) Let random variable X be the number of games you win. Find the probability model for X.
e) What are the expected value and standard deviation?

22. Contracts. Your company bids for two contracts. You believe the probability you get contract #1 is 0.8. If you get contract #1, the probability you also get contract #2 will be 0.2, and if you do not get #1, the probability you get #2 will be 0.3.
a) Are the two contracts independent? Explain.
b) Find the probability you get both contracts.
c) Find the probability you get no contract.
d) Let X be the number of contracts you get. Find the probability model for X.
e) Find the expected value and standard deviation.

23. Batteries. In a group of 10 batteries, 3 are dead. You choose 2 batteries at random.
a) Create a probability model for the number of good batteries you get.
b) What's the expected number of good ones you get?
c) What's the standard deviation?

24. Kittens. In a litter of seven kittens, three are female. You pick two kittens at random.
a) Create a probability model for the number of male kittens you get.
b) What's the expected number of males?
c) What's the standard deviation?

25. Random variables. Given independent random variables with means and standard deviations as shown, find the mean and standard deviation of:
a) $3X$
b) $Y + 6$
c) $X + Y$
d) $X - Y$
e) $X_1 + X_2$

	Mean	SD
X	10	2
Y	20	5

26. Random variables. Given independent random variables with means and standard deviations as shown, find the mean and standard deviation of:
a) $X - 20$
b) $0.5Y$
c) $X + Y$
d) $X - Y$
e) $Y_1 + Y_2$

	Mean	SD
X	80	12
Y	12	3

27. Random variables. Given independent random variables with means and standard deviations as shown, find the mean and standard deviation of:
a) $0.8Y$
b) $2X - 100$
c) $X + 2Y$
d) $3X - Y$
e) $Y_1 + Y_2$

	Mean	SD
X	120	12
Y	300	16

28. Random variables. Given independent random variables with means and standard deviations as shown, find the mean and standard deviation of:
a) $2Y + 20$
b) $3X$
c) $0.25X + Y$
d) $X - 5Y$
e) $X_1 + X_2 + X_3$

	Mean	SD
X	80	12
Y	12	3

29. Eggs. A grocery supplier believes that in a dozen eggs, the mean number of broken ones is 0.6 with a standard deviation of 0.5 eggs. You buy 3 dozen eggs without checking them.

a) How many broken eggs do you expect to get?
b) What's the standard deviation?
c) What assumptions did you have to make about the eggs in order to answer this question?

30. Garden. A company selling vegetable seeds in packets of 20 estimates that the mean number of seeds that will actually grow is 18, with a standard deviation of 1.2 seeds. You buy 5 different seed packets.

a) How many bad seeds do you expect to get?
b) What's the standard deviation?
c) What assumptions did you make about the seeds? Do you think that assumption is warranted? Explain.

31. Repair calls. Find the mean and standard deviation of the number of repair calls the appliance shop in Exercise 15 should expect during an 8-hour day.

32. Stop! Find the mean and standard deviation of the number of red lights the commuter in Exercise 16 should expect to hit on her way to work during a 5-day work week.

33. Tickets. A delivery company's trucks occasionally get parking tickets, and based on past experience, the company plans that the trucks will average 1.3 tickets a month, with a standard deviation of 0.7 tickets.

a) If they have 18 trucks, what are the mean and standard deviation of the total number of parking tickets the company will have to pay this month?
b) What assumption did you make in answering?

34. Donations. Organizers of a televised fundraiser know from past experience that most people donate small amounts ($10–$25), some donate larger amounts ($50–$100), and a few people make very generous donations of $250, $500, or more. Historically, pledges average about $32 with a standard deviation of $54.

a) If 120 people call in pledges, what are the mean and standard deviation of the total amount raised?
b) What assumption did you make in answering this question?

35. Fire! An insurance company estimates that it should make an annual profit of $150 on each homeowner's policy written, with a standard deviation of $6000.

a) Why is the standard deviation so large?
b) If it writes only two of these policies, what are the mean and standard deviation of the annual profit?
c) If it writes 10,000 of these policies, what are the mean and standard deviation of the annual profit?
d) Is the company likely to be profitable? Explain.
e) What assumptions underlie your analysis? Can you think of circumstances under which those assumptions might be violated? Explain.

36. Casino. A casino knows that people play the slot machines in hopes of hitting the jackpot but that most of them lose their dollar. Suppose a certain machine pays out an average of $0.92, with a standard deviation of $120.

a) Why is the standard deviation so large?
b) If you play 5 times, what are the mean and standard deviation of the casino's profit?
c) If gamblers play this machine 1000 times in a day, what are the mean and standard deviation of the casino's profit?
d) Is the casino likely to be profitable? Explain.

37. Cereal. The amount of cereal that can be poured into a small bowl varies with a mean of 1.5 ounces and a standard deviation of 0.3 ounces. A large bowl holds a mean of 2.5 ounces with a standard deviation of 0.4 ounces. You open a new box of cereal and pour one large and one small bowl.

a) How much more cereal do you expect to be in the large bowl?
b) What's the standard deviation of this difference?
c) If the difference follows a Normal model, what's the probability the small bowl contains more cereal than the large one?
d) What are the mean and standard deviation of the total amount of cereal in the two bowls?
e) If the total follows a Normal model, what's the probability you poured out more than 4.5 ounces of cereal in the two bowls together?
f) The amount of cereal the manufacturer puts in the boxes is a random variable with a mean of 16.3 ounces and a standard deviation of 0.2 ounces. Find the expected amount of cereal left in the box and the standard deviation.

38. Pets. The American Veterinary Association claims that the annual cost of medical care for dogs averages $100, with a standard deviation of $30, and for cats averages $120, with a standard deviation of $35.

a) What's the expected difference in the cost of medical care for dogs and cats?
b) What's the standard deviation of that difference?
c) If the costs can be described by Normal models, what's the probability that medical expenses are higher for someone's dog than for her cat?
d) What concerns do you have?

39. More cereal. In Exercise 37 we poured a large and a small bowl of cereal from a box. Suppose the amount of cereal that the manufacturer puts in the boxes is a random variable with mean 16.2 ounces and standard deviation 0.1 ounces.

a) Find the expected amount of cereal left in the box.
b) What's the standard deviation?
c) If the weight of the remaining cereal can be described by a Normal model, what's the probability that the box still contains more than 13 ounces?

40. More pets. You're thinking about getting two dogs and a cat. Assume that annual veterinary expenses are independent and have a Normal model with the means and standard deviations described in Exercise 38.

a) Define appropriate variables and express the total annual veterinary costs you may have.
b) Describe the model for this total cost. Be sure to specify its name, expected value, and standard deviation.
c) What's the probability that your total expenses will exceed $400?

41. Medley. In the 4×100 medley relay event, four swimmers swim 100 yards, each using a different stroke. A college team preparing for the conference championship looks at the times their swimmers have posted and creates a model based on the following assumptions:

- The swimmers' performances are independent.
- Each swimmer's times follow a Normal model.
- The means and standard deviations of the times (in seconds) are as shown:

Swimmer	Mean	SD
1 (backstroke)	50.72	0.24
2 (breaststroke)	55.51	0.22
3 (butterfly)	49.43	0.25
4 (freestyle)	44.91	0.21

a) What are the mean and standard deviation for the relay team's total time in this event?
b) The team's best time so far this season was 3:19.48. (That's 199.48 seconds.) Do you think the team is likely to swim faster than this at the conference championship? Explain.

42. Bikes. Bicycles arrive at a bike shop in boxes. Before they can be sold, they must be unpacked, assembled, and tuned (lubricated, adjusted, etc.). Based on past experience, the shop manager makes the following assumptions about how long this may take:

- The times for each setup phase are independent.
- The times for each phase follow a Normal model.
- The means and standard deviations of the times (in minutes) are as shown:

Phase	Mean	SD
Unpacking	3.5	0.7
Assembly	21.8	2.4
Tuning	12.3	2.7

a) What are the mean and standard deviation for the total bicycle setup time?
b) A customer decides to buy a bike like one of the display models but wants a different color. The shop has one, still in the box. The manager says they can have it ready in half an hour. Do you think the bike will be set up and ready to go as promised? Explain.

43. Farmers' market. A farmer has 100 lb of apples and 50 lb of potatoes for sale. The market price for apples (per pound) each day is a random variable with a mean of 0.5 dollars and a standard deviation of 0.2 dollars. Similarly, for a pound of potatoes, the mean price is 0.3 dollars and the standard deviation is 0.1 dollars. It also costs him 2 dollars to bring all the apples and potatoes to the market. The market is busy with eager shoppers, so we can assume that he'll be able to sell all of each type of produce at that day's price.

a) Define your random variables, and use them to express the farmer's net income.
b) Find the mean.
c) Find the standard deviation of the net income.
d) Do you need to make any assumptions in calculating the mean? How about the standard deviation?

44. Bike sale. The bicycle shop in Exercise 42 will be offering 2 specially priced children's models at a sidewalk sale. The basic model will sell for \$120 and the deluxe model for \$150. Past experience indicates that sales of the basic model will have a mean of 5.4 bikes with a standard deviation of 1.2, and sales of the deluxe model will have a mean of 3.2 bikes with a standard deviation of 0.8 bikes. The cost of setting up for the sidewalk sale is \$200.

a) Define random variables and use them to express the bicycle shop's net income.
b) What's the mean of the net income?
c) What's the standard deviation of the net income?
d) Do you need to make any assumptions in calculating the mean? How about the standard deviation?

45. Coffee and doughnuts. At a certain coffee shop, all the customers buy a cup of coffee; some also buy a doughnut. The shop owner believes that the number of cups he sells each day is normally distributed with a mean of 320 cups and a standard deviation of 20 cups. He also believes that the number of doughnuts he sells each day is independent of the coffee sales and is normally distributed with a mean of 150 doughnuts and a standard deviation of 12.

a) The shop is open every day but Sunday. Assuming day-to-day sales are independent, what's the probability he'll sell over 2000 cups of coffee in a week?
b) If he makes a profit of 50 cents on each cup of coffee and 40 cents on each doughnut, can he reasonably expect to have a day's profit of over \$300? Explain.
c) What's the probability that on any given day he'll sell a doughnut to more than half of his coffee customers?

46. Weightlifting. The Atlas BodyBuilding Company (ABC) sells "starter sets" of barbells that consist of one bar, two 20-pound weights, and four 5-pound weights. The bars weigh an average of 10 pounds with a standard deviation of 0.25 pounds. The weights average the specified amounts, but the standard deviations are 0.2 pounds for the 20-pounders and 0.1 pounds for the 5-pounders. We can assume that all the weights are normally distributed.

a) ABC ships these starter sets to customers in two boxes: The bar goes in one box and the six weights go in another. What's the probability that the total weight in that second box exceeds 60.5 pounds? Define your variables clearly and state any assumptions you make.
b) It costs ABC \$0.40 per pound to ship the box containing the weights. Because it's an odd-shaped package, though, shipping the bar costs \$0.50 a pound plus a \$6.00 surcharge. Find the mean and standard deviation of the company's total cost for shipping a starter set.
c) Suppose a customer puts a 20-pound weight at one end of the bar and the four 5-pound weights at the other end. Although he expects the two ends to weigh the same, they might differ slightly. What's the probability the difference is more than a quarter of a pound?

JUST CHECKING Answers

1. a)

Outcome	X = cost	Probability
Recharging works	\$60	0.75
Replace control unit	\$200	0.25

b) $60(0.75) + 200(0.25) = \$95$

c) Car owners with this problem will spend an average of \$95 to get it fixed.

2. a) $100 + 100 = 200$ seconds

b) $\sqrt{50^2 + 50^2} = 70.7$ seconds

c) The times for the two customers are independent.