Chapter 6
Random Variables

Section 6.2
Transforming and Combining Random Variables

Linear Transformations

In Section 6.1, we learned that the mean and standard deviation give us important information about a random variable. In this section, we'll learn how the mean and standard deviation are affected by transformations on random variables.

In Chapter 2, we studied the effects of transformations on the shape, center, and spread of a distribution of data. Recall what we discovered:

1. **Adding (or subtracting) a constant: Adding the same number *a* (either positive, zero, or negative) to each observation:**

*Adds *a* to measures of center and location (mean, median, quartiles, percentiles).

***Does not* change shape or measures of spread (range, IQR, *standard deviation*).**

2. **Multiplying (or dividing) each observation by a constant *b* (positive, negative, or zero):**

*Multiplies (divides) measures of center and location (mean, median, quartiles, percentiles) by *b*.

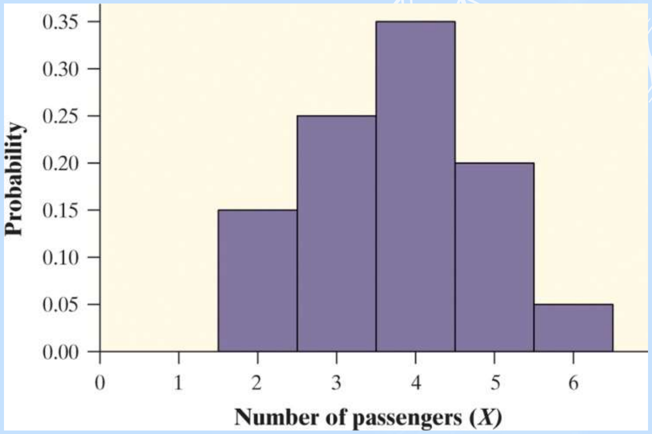
*Multiplies (divides) measures of spread (range, IQR, standard deviation) by $| b |$.

*Does not change the shape of the distribution.

Example 1: Pete's Jeep Tours offers a popular half-day trip in a tourist area. There must be at least 2 passengers for the trip to run, and the vehicle will hold up to 6 passengers. Define *X* as the number of passengers on a randomly selected day.

Passengers x_i	2	3	4	5	6
Probability p_i	0.15	0.25	0.35	0.20	0.05

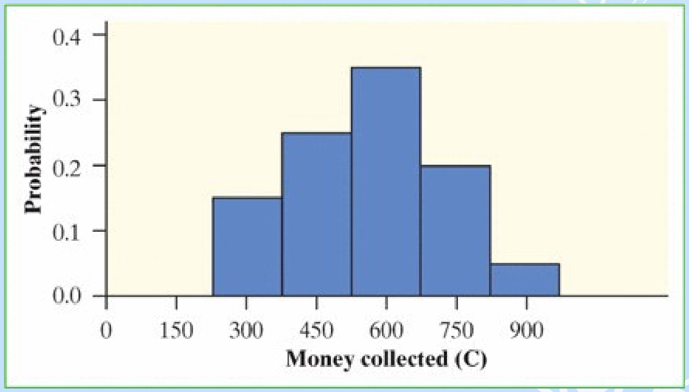
The mean of *X* is 3.75 and the standard deviation is 1.089.



Pete charges $150 per passenger. The random variable *C* describes the amount Pete collects on a randomly selected day.

Collected c_i	300	450	600	750	900
Probability p_i	0.15	0.25	0.35	0.20	0.05

The mean of *C* is $562.50 and the standard deviation is $163.46.



Compare the shape, center, and spread of the two probability distributions.

Shape: The two probability distributions have the same shape.

Center: The mean of X is $\mu_X = 3.75$. The mean of C is $\mu_C = 562.50$, which is (150)(3.75). That is, $\mu_C = 150\,\mu_X$.

Spread: The standard deviation of X is $\sigma_X = 1.089$. The standard deviation of C is $\sigma_C = 163.46$, which is (150)(1.089). That is, $\sigma_C = 150\,\sigma_X$.

How does multiplying or dividing by a constant affect a random variable?

Effect on a Random Variable of Multiplying (or Dividing) by a Constant:

Multiplying (or dividing) each value of a random variable by a positive number b:

*Multiplies (divides) measures of center and location (mean, median, quartiles, percentiles) by b.

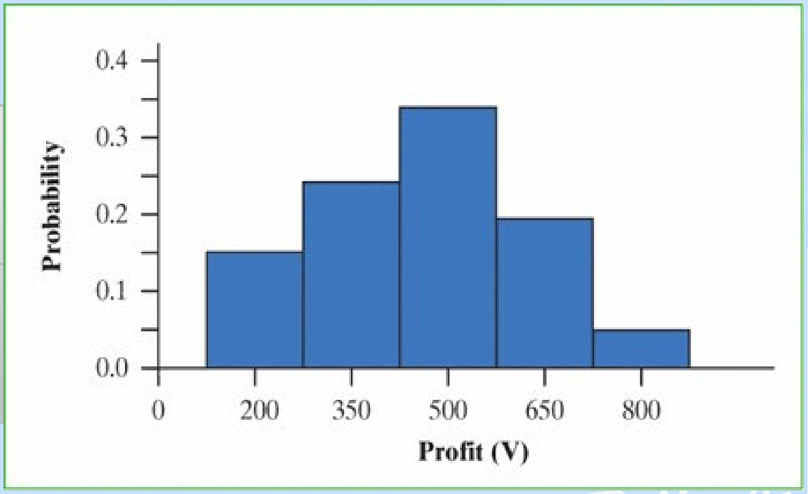
*Multiplies (divides) measures of spread (range, *IQR,* standard deviation) by b.

*Does not change the shape of the distribution.

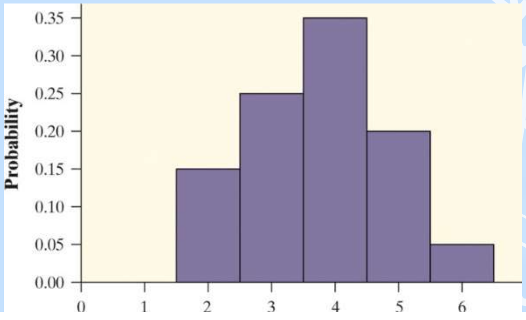
Note: Multiplying a random variable by a constant b multiplies the variance by b^2.

Example 2: It costs Pete \$100 to buy permits, gas, and a ferry pass for each half-day trip. The amount of profit V that Pete makes from the trip is the total amount of money C that he collects from passengers minus \$100. That is, $V = C - 100$. If Pete has only two passengers on the trip ($X = 2$), then $C = 2(150) = 300$ and $V = 200$. From the probability distribution of C, the chance that this happens is 0.15. So the smallest possible value of V is \$200; its corresponding probability is 0.15. If $X = 3$, then $C = 450$ and $V = 350$, and the corresponding probability is 0.25. The probability distribution of V is

Profit v_i	200	350	500	650	800
Probability p_i	0.15	0.25	0.35	0.20	0.05



The mean of V is \$462.50 and the standard deviation is \$163.46.
Standard deviation doesn't change for +/-.!



Effect on a Random Variable of adding (or subtracting) a Constant

Adding the same number *a* (which could be negative) to each value of a random variable:

*Adds *a* to measures of center and location (mean, median, quartiles, percentiles).

Does not* change shape or measures of spread (range, *IQR*, *standard deviation).

Example 3: A large auto dealership keeps track of sales made during each hour of the day. Let X = the number of cars sold during the first hour of business on a randomly selected Friday. Based on previous records, the probability distribution of X is as follows:

Cars sold:	0	1	2	3
Probability:	0.3	0.4	0.2	0.1

The random variable X has mean $\mu_X = 1.1$ and standard deviation $\sigma_X = 0.943$.

a) Suppose the dealership's manager receives a $500 bonus from the company for each car sold. Let Y = the bonus received from car sales during the first hour on a randomly selected Friday. Find the mean and standard deviation of Y.

$Y = 500X$. $\mu_Y = 500(1.1) = \$550$. $\sigma_Y = 500(0.943) = \$471.50$.

b) To encourage customers to buy cars on Friday mornings, the manager spends $75 to provide coffee and doughnuts. The manager's net profit *T* on a randomly selected Friday is the bonus earned minus this $75. Find the mean and standard deviation of *T*.

$T = Y - 75$. $\mu_T = 550 - 75 = \$475$. $\sigma_T = \$471.50$.

(standard deviation doesn't change for adding or subtracting values)

Try: El Dorado Community College has a mean probability of $832.50 and a standard deviation of $103. If the college charges a $100 student fee, what will the new mean and standard deviation be?

Whether we are dealing with data or random variables, the effects of a linear transformation are the same.

Effects of a Linear Transformation on the Mean and Standard Deviation

If $\boldsymbol{Y = a + bX}$ is a linear transformation of the random variable X, then

*the probability distribution of Y has the same shape as the probability distribution of X.

Y = the **mean of Y** *$\mu_Y = a + b\,\mu_X$.

and the **Standard deviation** = |slope or rate|mean of X

*$\sigma_Y = |\ b\ |\ \sigma_X$ (since b could be a negative number).

Example 4: One brand of bathtub comes with a dial to set the water temperature. When the "babysafe" setting is selected and the tub is filled, the temperature X of the water follows a normal distribution with a mean of 34°C and a standard deviation of 2°C.

a) Define the random variable Y to be the water temperature in degrees Fahrenheit (recall that $F = (9/5)C + 32$) when the dial is set on "babysafe." Find the mean and standard deviation of Y.

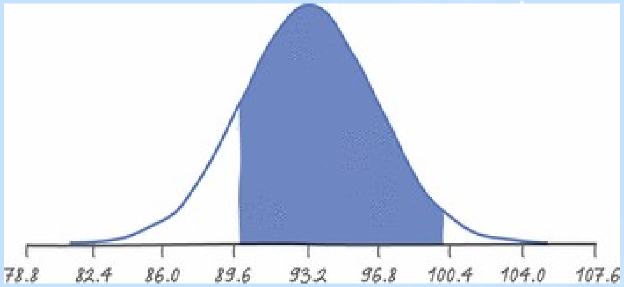
According to the formula for converting Celsius to Fahrenheit, $y = (9/5)x + 32$. We could also write this in the form $y = 32 + (9/5)x$. The mean of Y is

$$\mu_Y = 32 + (9/5)\mu_X = 32 + (9/5)(34) = 93.2°\text{F}.$$

$$\sigma_Y = |\, b \,|\, \sigma_X$$

The standard deviation of Y is $\sigma_Y = (9/5)\,\sigma_X = (9/5)(2) = 3.6°\text{F}$.

b) According to Babies R Us, the temperature of a baby's bathwater should be between 90°F and 100°F. Find the probability that the water temperature on a randomly selected day when the "babysafe" setting is used meets the Babies R Us recommendation. Show your work.



The linear transformation doesn't change the shape of the probability distribution, so the random variable *Y* is normally distributed with a mean of 93.2 and a standard deviation of 3.6. We want to find $P(90 \le Y \le 100)$. The shaded area in the figure above shows the desired probability. To find this area,

Normalcdf(lower: 90, upper:100, μ: 93.2, σ: 3.6) = 0.7835

There's about a 78% chance that the water temperature meets the recommendation on a randomly selected day.

Try: In a large introductory stats class, the distribution of raw scores on a test X follows a normal distribution with a mean of 17.2 and a standard deviation of 3.8. The Professor decides to scale the scores by multiplying the raw scores by 4 and adding 10.

Write an equation to represent Y: $\boldsymbol{Y = a + bX}$

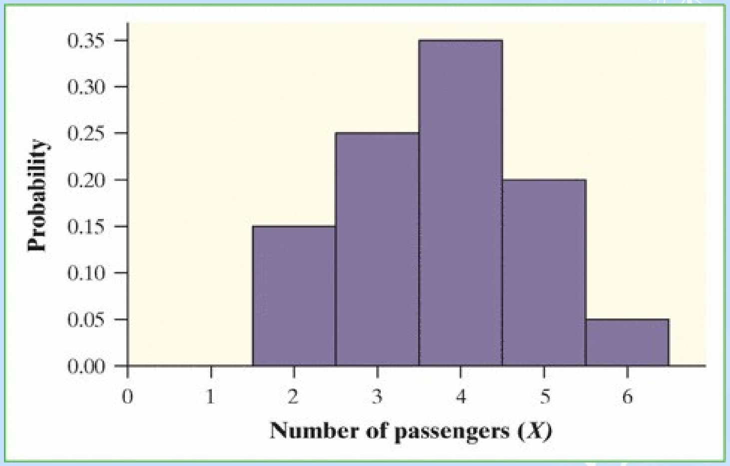
Then Find the mean for the new scores or Y:

What is the standard deviation for y: $\boldsymbol{\sigma_Y = |\, b \,|\, \sigma_X}$

Then find the probability P(Y$\geq$90)

Example 5: Earlier, we examined the probability distribution for the random variable X = the number of passengers on a randomly selected half-day trip with Pete's Jeep Tours. Here's a brief recap:

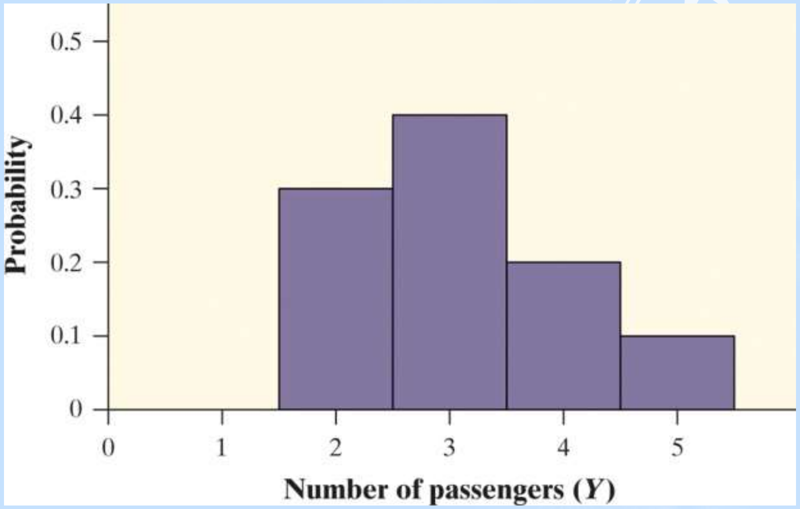
Passengers x_i	2	3	4	5	6
Probability p_i	0.15	0.25	0.35	0.20	0.05



Mean $\mu_X = 3.75$ Standard Deviation $\sigma_X = 1.0897$

Pete's sister Erin, who lives near a tourist area in another part of the country, is impressed by the success of Pete's business. She decides to join the business, running tours on the same days as Pete in her slightly smaller vehicle, under the name Erin's Adventures. After a year of steady bookings, Erin discovers that the number of passengers Y on her half-day tours has the following probability distribution. The figure below displays this distribution as a histogram.

Passengers y_i	2	3	4	5
Probability p_i	0.3	0.4	0.2	0.1



Mean $\mu_Y = 3.10$ Standard Deviation $\sigma_Y = 0.943$

How many total passengers T will Pete and Erin have on their tours on a randomly selected day? To answer this question, we need to know about the distribution of the random variable $T = X + Y$.

How many more or fewer passengers D will Pete have than Erin on a randomly selected day? To answer this question, we need to know about the distribution of the random variable $D = X - Y$.

Mean of the Sum of Random Variables

For any two random variables X and Y, if $T = X + Y$, then the expected value of T is

$$\boldsymbol{E(T) = \mu_T = \mu_X + \mu_Y.}$$

In general, the mean of the sum of several random variables is the sum of their means.

How much variability is there in the total number of passengers who go on Pete's and Erin's tours on a randomly selected day? To determine this, we need to find the probability distribution of T.

The only way to determine the probability for any value of T is if X and Y are **independent random variables.**

If knowing whether any event involving X alone has occurred tells us nothing about the occurrence of any event involving Y alone, and vice versa, then X and Y are **independent random variables.**

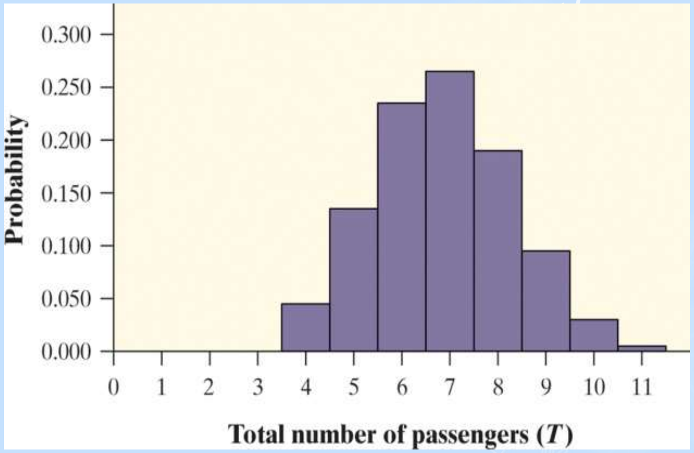
Probability models often assume independence when the random variables describe outcomes that appear unrelated to each other.

You should always ask whether the assumption of independence seems reasonable.

In our investigation, it is reasonable to assume X and Y are independent since the siblings operate their tours in different parts of the country.

Example 6: Let $T = X + Y$. Consider all possible combinations of the values of X and Y.

x_i	p_i	y_i	p_i	$t_i = x_i + y_i$	p_i
2	0.15	2	0.3	4	(0.15)(0.3) = 0.045
2	0.15	3	0.4	5	(0.15)(0.4) = 0.060
2	0.15	4	0.2	6	(0.15)(0.2) = 0.030
2	0.15	5	0.1	7	(0.15)(0.1) = 0.015
3	0.25	2	0.3	5	(0.25)(0.3) = 0.075
3	0.25	3	0.4	6	(0.25)(0.4) = 0.100
3	0.25	4	0.2	7	(0.25)(0.2) = 0.050
3	0.25	5	0.1	8	(0.25)(0.1) = 0.025
4	0.35	2	0.3	6	(0.35)(0.3) = 0.105
4	0.35	3	0.4	7	(0.35)(0.4) = 0.140
4	0.35	4	0.2	8	(0.35)(0.2) = 0.070
4	0.35	5	0.1	9	(0.35)(0.1) = 0.035
5	0.20	2	0.3	7	(0.20)(0.3) = 0.060
5	0.20	3	0.4	8	(0.20)(0.4) = 0.080
5	0.20	4	0.2	9	(0.20)(0.2) = 0.040
5	0.20	5	0.1	10	(0.20)(0.1) = 0.020
6	0.05	2	0.3	8	(0.05)(0.3) = 0.015
6	0.05	3	0.4	9	(0.05)(0.4) = 0.020
6	0.05	4	0.2	10	(0.05)(0.2) = 0.010
6	0.05	5	0.1	11	(0.05)(0.1) = 0.005



We can construct the probability distribution by listing all combinations of X and Y that yield each possible value of T and adding the corresponding probabilities. Here is the result.

Value t_i:	4	5	6	7	8	9	10	11
Probability p_i:	0.045	0.135	0.235	0.265	0.190	0.095	0.030	0.005

The mean of T is

$$\mu_T = \sum t_i p_i = (4)(0.045) + (5)(0.135) + \ldots + (11)(0.005) = 6.85.$$

Recall that $\mu_X = 3.75$ and $\mu_Y = 3.10$. Our calculation confirms that

$$\mu_T = \mu_X + \mu_Y = 3.75 + 3.10 = 6.85$$

Value t_i:	4	5	6	7	8	9	10	11
Probability p_i:	0.045	0.135	0.235	0.265	0.190	0.095	0.030	0.005

What about the variance of *T*? It's

$$\sigma_T^2 = \sum (t_i - \mu_T)^2 p_i$$

$$= (4-6.85)^2(0.045) + (5-6.85)^2(0.135) + \ldots + (11-6085)^2(0.005) = 2.0775$$

Recalling that we see that 1.1875 + 0.89 = 2.0775.That is,

$$\sigma_T^2 = \sigma_X^2 + \sigma_Y^2$$

To find the standard deviation of *T*, take the square root of the variance

$$\sigma_T = \sqrt{2.0775} = 1.441$$

Variance of the Sum of Random Variables

For any two *independent* random variables X and Y, if $T = X + Y$, then the variance of T is $\sigma_T^2 = \sigma_X^2 + \sigma_Y^2$

In general, the variance of the sum of several *independent* random variables is the sum of their variances.

Remember that you can add variances only if the two random variables are independent, and that you can NEVER add standard deviations!

Example 7: A college uses SAT scores as one criterion for admission. Experience has shown that the distribution of SAT scores among its entire population of applicants is such that

SAT Math score X:	$\mu_X = 519$	$\sigma_X = 115$
SAT Critical Reading score Y:	$\mu_Y = 507$	$\sigma_Y = 111$

What are the mean and standard deviation of the total score $X + Y$ among students applying to this college?

$T = X + Y$

The mean overall SAT score is

$$\mu_T = \mu_X + \mu_Y = 519 + 507 = 1026$$

The variance and standard deviation of the total *cannot be computed* from the information given. SAT Math and Critical reading scores are not independent, because students who score high on one exam tend to score high on the other also.

Example 8: Earlier, we defined X = the number of passengers that Pete has and Y = the number of passengers that Erin has on a randomly selected day. Recall that

$$\mu_X = 3.75, \sigma_X = 1.0897 \qquad \mu_Y = 3.10, \sigma_Y = 0.943$$

Pete charges \$150 per passenger and Erin charges \$175 per passenger. Calculate the mean and the standard deviation of the total amount that Pete and Erin collect on a randomly chosen day.

Let W = the total amount collected. Then $W = 150X + 175Y$. If we let $C = 150X$ and G = 175 Y, then we can write W as the sum of two random variables: $W = C + G$. We can use what we learned earlier about the effect of multiplying by a constant to find the mean and standard deviation of C and G.

For $C = 150X$, $\mu_c = 150\mu_X = 150(3.75) =$ \$562.50 and $\sigma_c = 150(1.0897) =$ \$163.46

For $G = 175Y$, $\mu_c = 175(3.10) =$ \$542.50 and $\sigma_G = 175(0.943) =$ \$165.03

We know that the mean of the sum of two random variables equals the sum of their means:

$\mu_W = \mu_C + \mu_G = 562.50 + 542.50 = 1105$

On average, Pete and Erin expect to collect a total of \$1105 per day.

Because the number of passengers *X* and *Y* are independent random variables, so are the amounts of money collected *C* and *G*. Therefore, the variance of W is the sum of the variances of *C* and *G*.

$$\sigma_W^2 = \sigma_C^2 + \sigma_G^2 = (163.46)^2 + (165.03)^2 = 53{,}954.07$$

To get the standard deviation, we take the square root of the variance:

$$\sigma_W = \sqrt{53{,}954.07} = 232.28$$

The standard deviation of the total amount they collect is $232.28.

Differences of random variables

We can perform a similar investigation to determine what happens when we define a random variable as the difference of two random variables. In summary, we find the following:

Mean of the Difference of Random Variables

For any two random variables X and Y, if $D = X - Y$, then the expected value of D is

$$\mu_D = E(D) = \mu_X - \mu_Y$$

In general, the mean of the difference of several random variables is the difference of their means. *The order of subtraction is important!*

Variance of the Difference of Random Variables

For any two *independent* random variables X and Y, if $D = X - Y$, then the variance of D is

$$\sigma_D^2 = \sigma_X^2 + \sigma_Y^2$$

In general, the variance of the difference of two independent random variables is the sum of their variances.

Example 9: We have defined several random variables related to Pete's and Erin's tour businesses. For a randomly selected day, C = amount of money that Pete collects G = amount of money that Erin collects

Here are the means and standard deviations of these random variables:

$$\mu_C = 562.50 \qquad \mu_G = 542.50$$
$$\sigma_C = 163.46 \qquad \sigma_G = 165.03$$

Calculate the mean and the standard deviation of the difference $D = C - G$ in the amounts that Pete and Erin collect on a randomly chosen day. Interpret each value in context.

We know that the mean of the difference of two random variables is the difference of their means. That is, $\mu_D = \mu_C - \mu_G = 562.50 - 542.50 = 20.00$

On average, Pete collects $20 more per day than Erin does. Some days the difference will be more than $20, other days it will be less, but the average difference after lots of days will be about $20.

Because the number of passengers *X* and *Y* are independent random variables, so are the amounts of money collected *C* and *G*. Therefore, the variance of *D* is the sum of the variances of *C* and *G*:

$$\sigma_D^2 = \sigma_C^2 + \sigma_G^2 = (163.46)^2 + (165.03)^2 = 53{,}954.07$$
$$\sigma_D = \sqrt{53{,}954.07} = 232.28$$

The standard deviation of the difference in the amounts collected by Pete and Erin is $232.28. Even though the average difference in the amounts collected is $20, the difference on individual days will typically vary from the mean by about $232.

So far, we have concentrated on finding rules for means and variances of random variables. If a random variable is Normally distributed, we can use its mean and standard deviation to compute probabilities.

An important fact about Normal random variables is that *any sum or difference of independent Normal random variables is also Normally distributed.*

Example 10: Mr. Starnes likes sugar in his hot tea. From experience, he needs between 8.5 and 9 grams of sugar in a cup of tea for the drink to taste right. While making his tea one morning, Mr. Starnes adds four randomly selected packets of sugar. Suppose the amount of sugar in these packets follows a Normal distribution with mean 2.17 grams and standard deviation 0.08 grams. What's the probability that Mr. Starnes's tea tastes right?

Let X = the amount of sugar in a randomly selected packet. Then, X_1 = amount of sugar in Packet 1, X_2 = amount of sugar in Packet 2, X_3 = amount of sugar in Packet 3, and X_4 = amount of sugar in Packet 4. Each of these random variables has a Normal distribution with mean 2.17 grams and standard deviation 0.08 grams. We're interested in the total amount of sugar that Mr. Starnes puts in his tea, which is given by $T = X_1 + X_2 + X_3 + X_4$.

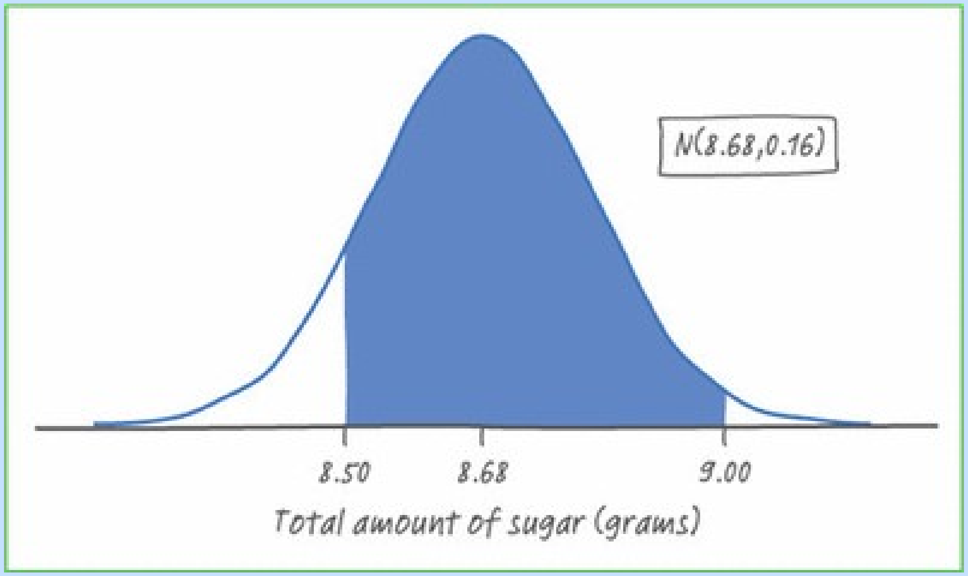
The random variable T is a sum of four independent Normal random variables. So T follows a Normal distribution with mean

$\mu_T = \mu_{X1} + \mu_{X2} + \mu_{X3} + \mu_{X4} = 2.17 + 2.17 + 2.17 + 2.17 = 8.68$ grams and variance

$\sigma_T^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 + \sigma_{X_3}^2 + \sigma_{X_4}^2 = (0.08)^2 + (0.08)^2 + (0.08)^2 + (0.08)^2 = 0.0256$ The standard deviation of T is

$$\sigma_T = \sqrt{0.0256} = 0.16$$

We want to find the probability that the total amount of sugar in Mr. Starnes's tea is between 8.5 and 9 grams. The figure below shows this probability as the area under a Normal curve.



$P(8.5 \le T \le 9)$

normalcdf(lower:8.5, upper:9, μ:8.68, σ:0.16) gives an area of 0.8470.

$$z = \frac{8.5 - 8.68}{0.16} = -1.13 \quad \text{and} \quad z = \frac{9 - 8.68}{0.16} = 2.00$$

$P(-1.13 \le Z \le 2.00) = 0.9772 - 0.1292 = 0.8480$

There is about an 85% chance Mr. Starnes's tea will taste right.

Example 11: The diameter *C* of a randomly selected large drink cup at a fast-food restaurant follows a Normal distribution with a mean of 3.96 inches and a standard deviation of 0.01 inches. The diameter *L* of a randomly selected large lid at this restaurant follows a Normal distribution with mean 3.98 inches and standard deviation 0.02 inches. For a lid to fit on a cup, the value of *L* has to be bigger than the value of *C*, but not by more than 0.06 inches. What's the probability that a randomly selected large lid will fit on a randomly chosen large drink cup?

We'll define the random variable $D = L - C$ to represent the difference between the lid's diameter and the cup's diameter.

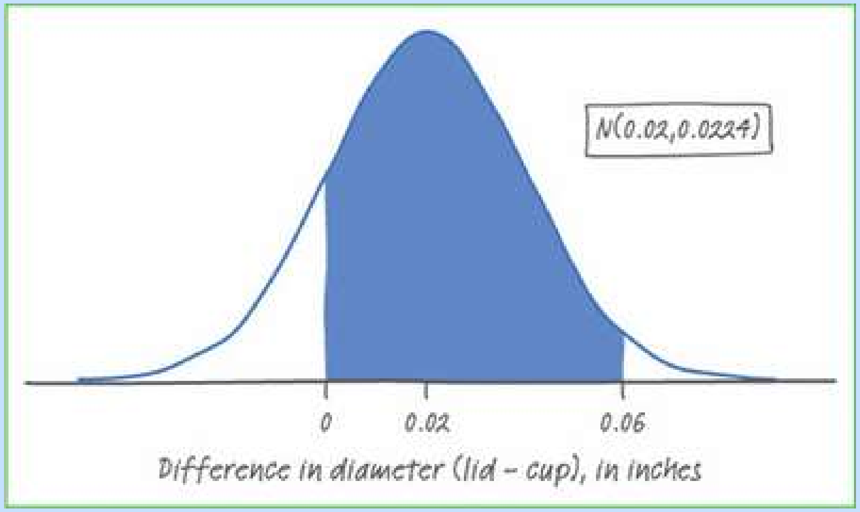
The random variable *D* is the difference of two independent Normal random variables. So *D* follows a Normal distribution with mean

$$\mu_D = \mu_L - \mu_C = 3.98 - 3.96 = 0.02$$

and variance $\sigma_D^2 = \sigma_3^2 + \sigma_C^2 = (0.02)^2 + (0.01)^2 = 0.0005$

The standard deviation of *D* is $\sigma_D = \sqrt{0.0005} = 0.0224$

We want to find the probability that the difference D is between 0 and 0.06 inches. The figure below shows this probability as the area under a Normal curve.



$$z = \frac{0 - 0.02}{0.0224} = -0.89 \quad \text{and} \quad z = \frac{0.06 - 0.02}{0.0224} = 1.79$$

$$P(-0.89 \le Z \le 1.79) = 0.9633 - 0.1867 = 0.7766$$

There's about a 78% chance that a randomly selected large lid will fit on a randomly chosen large drink cup at this fast-food restaurant. Roughly 22% of the time, the lid won't fit. This seems like an unreasonably high chance of getting a lid that doesn't fit. Maybe the restaurant should find a new supplier!