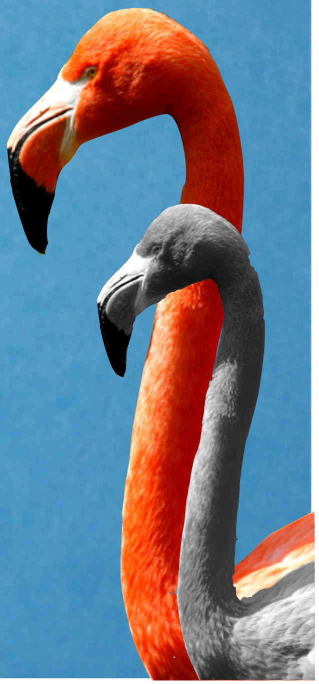
CHAPTER 6
Random Variables

6.2b
Transforming and Combining Random Variables

The Practice of Statistics, 5th Edition
Starnes, Tabor, Yates, Moore



Bedford Freeman Worth Publishers

Transforming and Combining Random Variables

Learning Objectives

After this section, you should be able to:

- ✓ DESCRIBE the effects of transforming a random variable by adding or subtracting a constant and multiplying or dividing by a constant.
- ✓ FIND the mean and standard deviation of the sum or difference of independent random variables.
- ✓ FIND probabilities involving the sum or difference of independent Normal random variables.

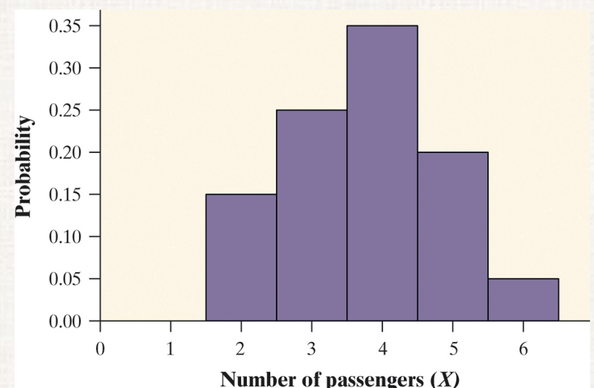
Combining Random Variables

Many interesting statistics problems require us to examine two or more random variables.

Let's investigate the result of adding and subtracting random variables. Let X = the number of passengers on a randomly selected trip with Pete's Jeep Tours. Y = the number of passengers on a randomly selected trip with Erin's Adventures. Define $T = X + Y$. What are the mean and variance of T?

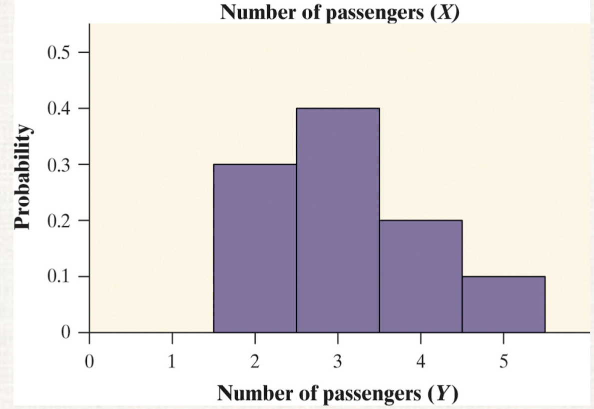
Passengers x_i	2	3	4	5	6
Probability p_i	0.15	0.25	0.35	0.20	0.05

Mean $\mu_X = 3.75$ Standard Deviation $\sigma_X = 1.090$



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 $\sigma_Y = 0.943$



Combining Random Variables

How many total passengers can Pete and Erin expect on a randomly selected day?
Since Pete expects $\mu_X = 3.75$ and Erin expects $\mu_Y = 3.10$, they will average a total of $3.75 + 3.10 = 6.85$ passengers per trip. We can generalize this result as follows:

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

$$E(T) = \mu_T = \mu_X + \mu_Y$$

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.

Combining Random Variables

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.

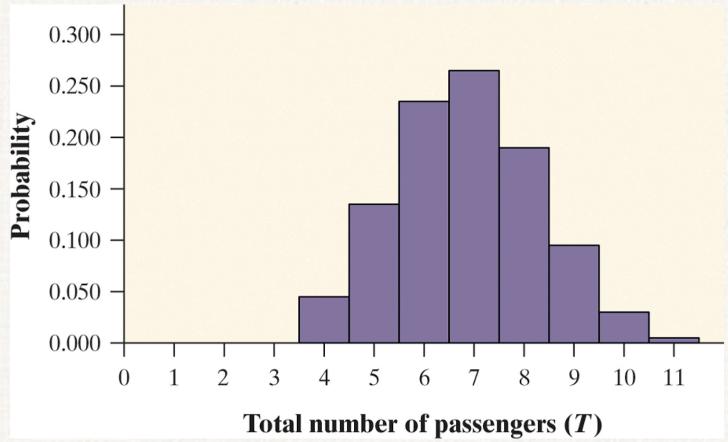
You should always ask whether the assumption of independence seems reasonable.

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.

Combining Random Variables

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	$(0.05)(0.1) = 0.005$



Recall: $\mu_T = \mu_X + \mu_Y = 6.85$

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

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

Note: $\sigma_X^2 = 1.1875$ and $\sigma_Y^2 = 0.89$

What do you notice about the variance of T?

Combining Random Variables

As the preceding example illustrates, when we add two *independent* random variables, their variances add. *Standard deviations do not add*.

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 – El Dorado Community College, again

Recall, the number of units X that a randomly selected El Dorado Community College full-time student is taking in the fall semester has $\mu_X = 14.65$ and $\sigma_X = 2.06$

El Dorado Community College also has a campus downtown, specializing in just a few fields of study. Full-time students at the downtown campus take only 3-unit classes. Let Y = number of units taken in the fall semester by a randomly selected full-time student at the downtown campus. $\mu_Y = 15$ and $\sigma_Y = 2.32$

If you were to randomly select one full-time student from the main campus and one full-time student from the downtown campus and add their number of units, determine the expected value and standard deviation of of the sum ($S = X + Y$).

$$\boldsymbol{\mu_S} = \mu_X + \mu_Y = 14.65 + 15 = \mathbf{29.65}$$

Since it is reasonable to assume X and Y are independent,

$$\sigma_S^2 = \sigma_X^2 + \sigma_Y^2 = (2.06)^2 + (2.32)^2 = 9.626$$

$$\boldsymbol{\sigma_S} = \sqrt{9.626} = \mathbf{3.10}$$

Example: Tuition, fees, and books

Let B = the amount spent on books in the fall semester for a randomly selected full-time student at El Dorado Community College. Suppose that $\mu_B = \$153$ and $\sigma_B = \$32$. Recall from earlier than C = overall cost for tuition and fees for a randomly selected full-time student at El Dorado Community College and that $\mu_C = \$832.50$ and $\sigma_C = \$103$.

Find the mean and standard deviation of the cost of tuition, fees, and books (C + B) for a randomly selected full-time student at El Dorado Community College.

The mean is $\boldsymbol{\mu_{C+B}} = \mu_C + \mu_B = 832.50 + 153 = \mathbf{\$985.50}$

The standard deviation cannot be calculated because the cost for tuition and fees and the cost for books are not independent. Students who take more units will typically have to buy more books.

Example – El Dorado Community College, again

Earlier we defined X = the number of units for a randomly selected full-time student at the main campus and Y = number of units for a randomly selected full-time student at the downtown campus. Also, $\mu_X = 14.65$ and $\sigma_X = 2.06$ and $\mu_Y = 15$ and $\sigma_Y = 2.32$

At the main campus, full-time students pay $50 per unit. At the downtown campus, full-time students pay $55 per unit.

Calculate the mean and standard deviation of the total amount of tuition for a randomly selected full-time student at the main campus and for a randomly selected full-time student at the downtown campus.

Let T = amount collected from a randomly selected full-time student at the main campus. Let U = amount collected from a randomly selected full-time student at the downtown campus.

50X + 55Y = T + U

$\mu_X = 14.65$ and $\sigma_X = 2.06$ and $\mu_Y = 15$ and $\sigma_Y = 2.32$

$\mu_T = 50\mu_X = \$732.50$ $\qquad$ $\sigma_T = 50\sigma_X = \$103$

$\mu_U = 55\mu_Y = \$825$ $\qquad$ $\sigma_U = 55\sigma_Y = \$127.60$

$$\mu_{T+U} = \mu_T + \mu_U = 732.50 + 825 = \mathbf{\$1557.50}$$

$$\sigma^2_{T+U} = \sigma^2_T + \sigma^2_U = 10609 + 16281.76 = 26890.76$$

$$\sigma_{T+U} = \sqrt{26890.76} \approx \mathbf{\$164}$$

On average, the sum of tuition for a randomly selected full-time student at the main campus and a randomly selected full-time student at the downtown campus will be $1557.50. This amount will typically vary about $164.

Combining 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

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

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 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 - difference

Earlier we defined T = tuition collected from a randomly selected full-time student at the main campus and U = tuition collected from a randomly selected full-time student at the downtown campus. The means and standard deviations are

$\mu_T = \$732.50 \quad \sigma_T = \$103 \qquad \mu_U = \$825 \quad \sigma_U = \$127.60$

Suppose we randomly select one full-time student from each of the two campuses. What are the mean and standard deviation of the difference in tuition charges, T – U? Interpret each of these values.

$$\mu_{T-U} = \mu_T - \mu_U = 732.50 - 825 = \mathbf{\$-92.50}$$

This means that on average, full-time students at the main campus pay \$92.50 less in tuition than full-time students at the downtown campus.

Because the students are randomly selected, T and U are independent. Thus, $\sigma^2_{T-U} = \sigma^2_T + \sigma^2_U = 10609 + 16281.76 = 26890.76$

$$\sigma_{T-U} = \sqrt{26890.76} \approx \mathbf{\$164}$$

Although the average difference in tuition for the two campuses is –\$92.50, the difference in tuition will typically vary from the average difference by about \$164.

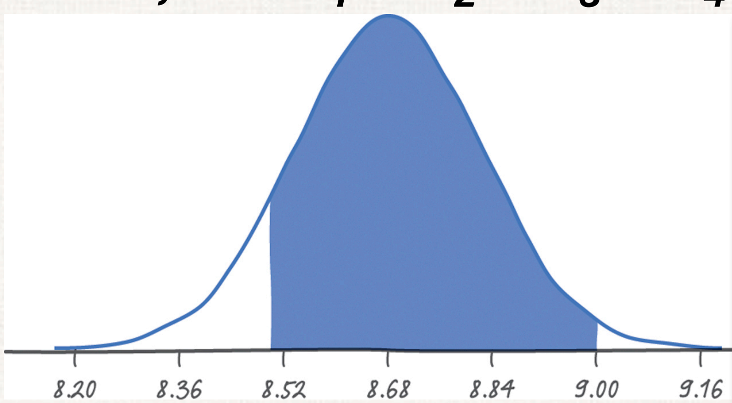
Combining Normal Random Variables

If a random variable is Normally distributed, we can use its mean and standard deviation to compute probabilities.

- *Any sum or difference of independent Normal random variables is also Normally distributed.*

Mr. Starnes likes between 8.5 and 9 grams of sugar in his hot tea. Suppose the amount of sugar in a randomly selected packet follows a Normal distribution with mean 2.17 g and standard deviation 0.08 g. If Mr. Starnes selects 4 packets at random, what is the probability his tea will taste right?

Let X = the amount of sugar in a randomly selected packet.
Then, $T = X_1 + X_2 + X_3 + X_4$. We want to find $P(8.5 \le T \le 9)$.



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

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

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

Example – Apples

Suppose that a certain variety of apples have weights that follow a Normal distribution with a mean of 9 ounces and a standard deviation of 1.5 ounces. If bags of apples are filled by randomly selecting 12 apples, what is the probability that the sum of the weights of the 12 apples is less than 100 ounces?

Let X = weight of a randomly selected apple. Then X_1 = weight of first randomly selected apple, and so on.

We are interested in T = $X_1 + X_2 + \ldots + X_{12}$. Because T is a sum of 12 independent Normal random variables, T follows a Normal distribution with mean $\boldsymbol{\mu_T} = \mu_{X_1} + \mu_{X_2} + \cdots + \mu_{X_{12}} = 9 + 9 + \cdots + 9 = \mathbf{108}$ and variance $\sigma_T^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 + \cdots + \sigma_{X_{12}}^2 = 1.5^2 + 1.5^2 + \cdots + 1.5^2 = 27$. The standard deviation is then $\boldsymbol{\sigma_T} = \sqrt{27} \approx \mathbf{5.2}$. We want to find P(T < 100)

$z = \frac{100-108}{5.2} = -1.54$ P(Z < -1.54) = **0.0618**

There is about a 6.2% chance that the 12 randomly selected apples will have a total weight of less than 100 ounces.

Example – pregnancy

The length of human pregnancies from conception to birth varies according to a distribution that is approximately Normal with mean 266 days and standard deviation 16 days. Choose two pregnancies independently and at random.

What is the expected difference in the lengths of the two pregnancies?

$$\mu_D = \mu_1 - \mu_2 = 266 - 266 = \mathbf{0}$$

What is the standard deviation of the difference in the lengths of the two pregnancies?

$$\sigma_D^2 = \sigma_1^2 + \sigma_2^2 = 16^2 + 16^2 = 512 \qquad \sigma_D = \sqrt{512} \approx \mathbf{22.63}$$

Find the probability that the difference in the lengths of the two pregnancies is greater than 25 days.

*Note, we could have D > 25 OR D < -25

$$z = \frac{25-0}{22.63} = 1.10 \text{ and } z = \frac{-25-0}{22.63} = -1.10$$

P(z > 1.10 or z < -1.10) = 2(.1357) = .2714

Approximately 27% of differences in the lengths of two pregnancies will be greater than 25 days.

Transforming and Combining Random Variables

Section Summary

In this section, we learned how to…

- ✓ DESCRIBE the effects of transforming a random variable by adding or subtracting a constant and multiplying or dividing by a constant.
- ✓ FIND the mean and standard deviation of the sum or difference of independent random variables.
- ✓ FIND probabilities involving the sum or difference of independent Normal random variables.

- ✓ Read p. 369-382 ccc 47, 49, 51, 53, 55, 57-59, 61