

Chapter 6: Random Variables

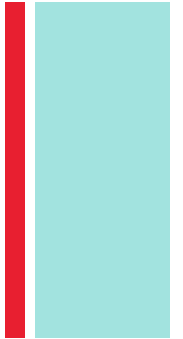
Section 6.2

Transforming and Combining Random Variables

The Practice of Statistics, 4th edition – For AP*
STARNES, YATES, MOORE

+ Chapter 6

Random Variables

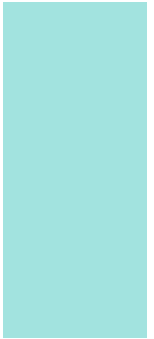


- 6.1 Discrete and Continuous Random Variables
- **6.2 Transforming and Combining Random Variables**
- 6.3 Binomial and Geometric Random Variables



Section 6.2

Transforming and Combining Random Variables



Learning Objectives

After this section, you should be able to...

- ✓ DESCRIBE the effect of performing a linear transformation on a random variable
- ✓ COMBINE random variables and CALCULATE the resulting mean and standard deviation
- ✓ CALCULATE and INTERPRET probabilities involving combinations of Normal 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 linear transformations on the shape, center, and spread of a distribution of data. Recall:

1. Adding (or subtracting) a constant, a , to each observation:

- Adds a to measures of center and location.
- Does not change the shape or measures of spread.

2. Multiplying (or dividing) each observation by a constant, b :

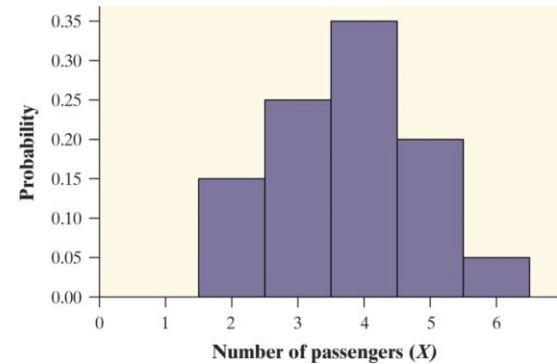
- Multiplies (divides) measures of center and location by b .
- Multiplies (divides) measures of spread by $|b|$.
- Does not change the shape of the distribution.

Linear Transformations

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

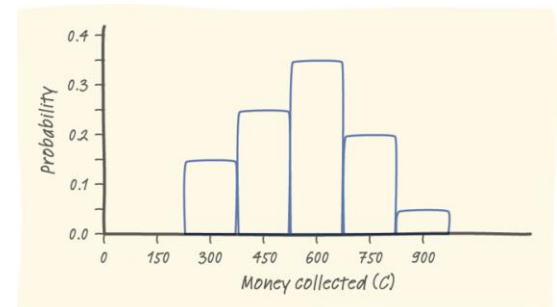
The mean of X is 3.75 and the standard deviation is 1.090.



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 \$163.50.



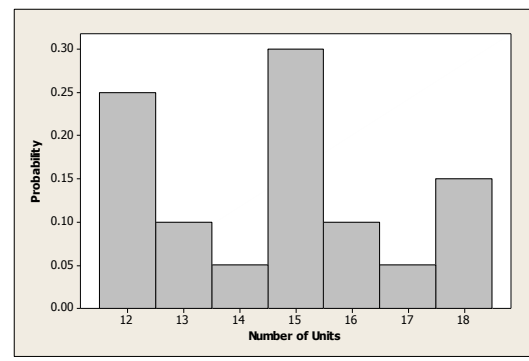
Compare the shape, center, and spread of the two probability distributions.

Linear Transformations

El Dorado Community College considers a student to be full-time if he or she is taking between 12 and 18 units. The number of units X that a randomly selected EDCC full-time student is taking in the fall semester has the following distribution.

# of units (X)	12	13	14	15	16	17	18
Probability	0.25	0.10	0.05	0.30	0.10	0.05	.015

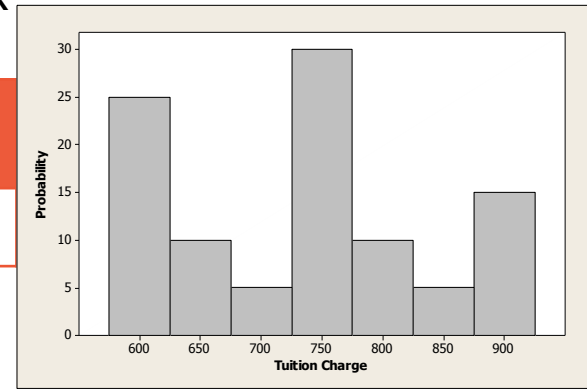
The mean of X is 14.65 and the standard deviation is 2.06.



At EDCC, the tuition for full-time students is \$50 per unit. That is, if T =tuition charge for a randomly selected full-time student, $T=50X$

Tuition Charge (T)	\$600	\$650	\$700	\$750	\$800	\$850	\$900
Probability	0.25	0.10	0.05	0.30	0.10	0.05	.015

The mean of C is \$732.50 and the standard deviation is \$103.



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

■ Linear Transformations

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

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

Multiplying (or dividing) each value of a random variable by a 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.

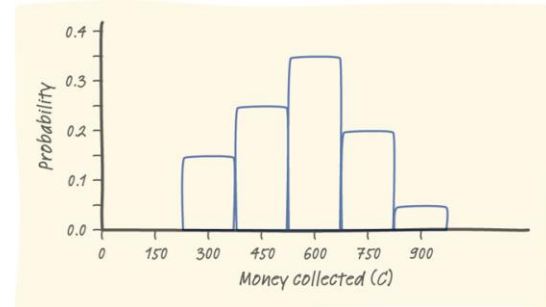
Note: Multiplying a random variable by a constant b multiplies the variance by b^2 .

Linear Transformations

Consider Pete's Jeep Tours again. We defined C as the amount of money 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

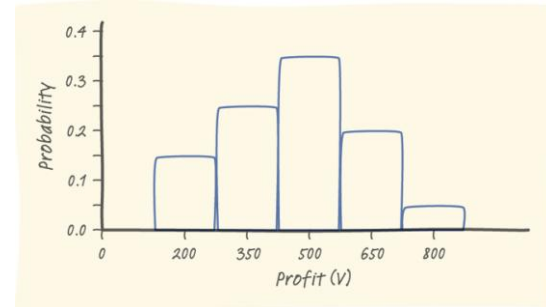
The mean of C is \$562.50 and the standard deviation is \$163.50.



It costs Pete \$100 per trip to buy permits, gas, and a ferry pass. The random variable V describes the profit Pete makes on a randomly selected day.

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 \$163.50.



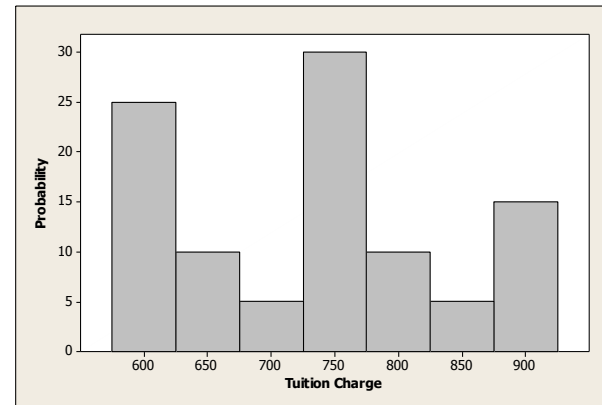
Compare the shape, center, and spread of the two probability distributions.

Linear Transformations El Dorado Community College

Tuition Charge (T)	\$600	\$650	\$700	\$750	\$800	\$850	\$900
Probability	0.25	0.10	0.05	0.30	0.10	0.05	0.155

The mean of C is \$732.50 and the standard deviation is \$103.

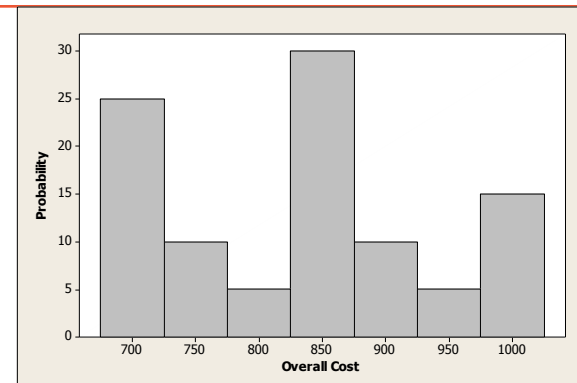
In addition to tuition charges, each full-time student at El Dorado Community College is assessed student fees of \$100 per semester. If C = overall cost for a randomly selected full-time student, $C = 100 + T$.



Tuition Charge (T)	\$700	\$750	\$800	\$850	\$900	\$950	\$1000
Probability	0.25	0.10	0.05	0.30	0.10	0.05	0.155

The mean of V is \$832.50 and the standard deviation is \$103.50.

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



■ Linear Transformations

How does adding or subtracting a constant affect a random variable?

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

■ Linear Transformations

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

Effect on a Linear Transformation on the Mean and Standard Deviation

If $Y = a + bX$ 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 .
- $\mu_Y = a + b\mu_X$.
- $\sigma_Y = |b|\sigma_X$ (since b could be a negative number).

■ Alternate Example - Scaling a Test

- **Problem:** In a large introductory statistics class, the distribution of X = raw scores on a test was approximately normally distributed 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.
- (a) Define the variable Y to be the scaled score of a randomly selected student from this class. Find the mean and standard deviation of Y .
- (b) What is the probability that a randomly selected student has a scaled test score of at least 90?

Solution:

(a) Since $Y = 10 + 4X$,

and

$$\mu_Y = 10 + 4\mu_X = 10 + 4(17.2) = 78.8$$

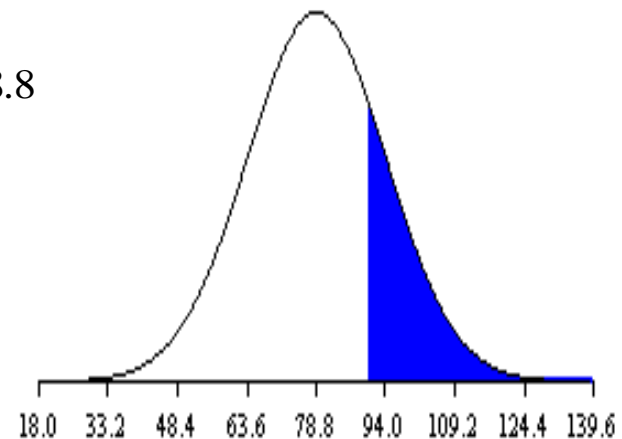
(b) Since linear transformations do not change

the shape, Y has the $N(78.8, 15.2)$ distribution.

The standardized score for a scaled score of 90

is $z = \frac{90 - 78.8}{15.2}$. According to Table A,

$P(z < 0.74) = 0.7704$. Thus, $P(Y \geq 90) = 1 - 0.7704 = 0.2296$.



■ Combining Random Variables

So far, we have looked at settings that involve a single random variable. Many interesting statistics problems require us to examine two or more random variables.

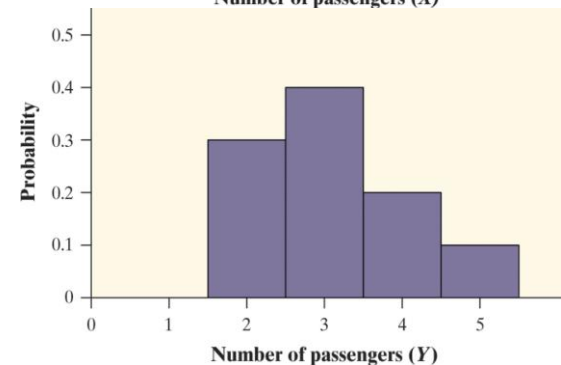
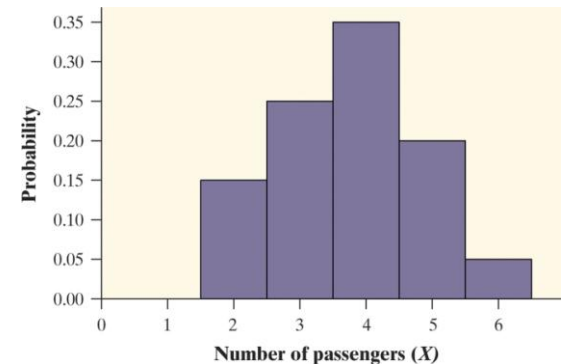
Let's investigate the result of adding and subtracting random variables. Let X = the number of passengers on a randomly selected trip with Pete's Jeep Tours. Y = the number of passengers on a randomly selected trip with Erin's Adventures. Define $T = X + Y$. What are the mean and variance of T ?

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 .

■ Alternate Example – El Dorado Community College

- El Dorado Community College also has a campus downtown, specializing in just a few fields of study. Full-time students at the downtown campus only take 3-unit classes. Let Y = number of units taken in the fall semester by a randomly selected full-time student at the downtown campus. Here is the probability distribution of Y :

Number of Units (Y)	12	15	18
Probability	0.3	0.4	0.3

The mean of this distribution is $\mu_Y = 15$ units, the variance is $\sigma_Y^2 = 5.40$ units² and the standard deviation is $\sigma_Y = 2.3$ units.

If you were to randomly select 1 full-time student from the main campus and 1 full-time student from the downtown campus and add their number of units, the expected value of the sum ($S = X + Y$) would be:

$$\mu_s = \mu_x + \mu_y = 14.65 + 15 = 29.65$$

■ Combining Random Variables

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

Definition:

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.

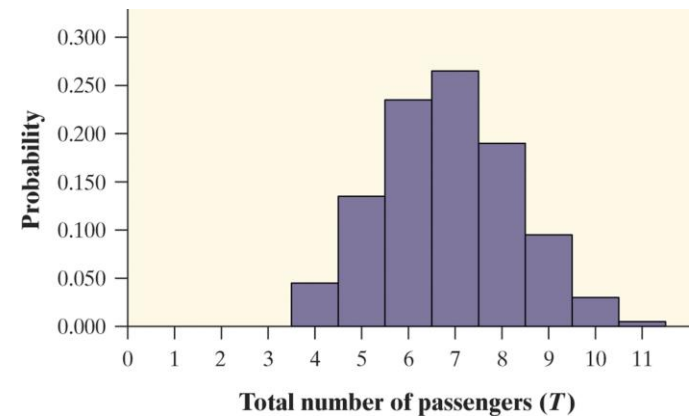
You should always ask whether the assumption of independence seems reasonable.

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

$$\begin{aligned} \sigma_T^2 &= \sum (t_i - \mu_T)^2 p_i \\ &= (4 - 6.85)^2(0.045) + \dots + \\ &\quad (11 - 6.85)^2(0.005) = 2.0775 \end{aligned}$$

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

What do you notice about the variance of T ?

Combining Random Variables – El Dorado Community College

- Let $S = X + Y$ as before. Assume that X and Y are independent, which is reasonable since each student was selected at random. Here are all possible combinations of X and Y .

X	$P(X)$	Y	$P(Y)$	$S=X+Y$	$P(S)=P(X)P(Y)$
12	0.25	12	0.3	24	0.075
12	0.25	15	0.4	27	0.10
12	0.25	18	0.3	30	0.075
13	0.10	12	0.3	25	0.03
13	0.10	15	0.4	28	0.04
13	0.10	18	0.3	31	0.03
14	0.05	12	0.3	26	0.015
14	0.05	15	0.4	29	0.02
14	0.05	18	0.3	32	0.015
15	0.30	12	0.3	27	0.09
15	0.30	15	0.4	30	0.12
15	0.30	18	0.3	33	0.09
16	0.10	12	0.3	28	0.03
16	0.10	15	0.4	31	0.04
16	0.10	18	0.3	34	0.03
17	0.05	12	0.3	29	0.015
17	0.05	15	0.4	32	0.02
17	0.05	18	0.3	35	0.015
18	0.15	12	0.3	30	0.045
18	0.15	15	0.4	33	0.06
18	0.15	18	0.3	36	0.045

$$\mu_S = 24(0.075) + 25(0.03) + \dots + 36(0.045) = 29.65$$

$$\sigma_S^2 = (24 - 29.65)^2(0.075) + (25 - 29.65)^2(0.03) + \dots + (36 - 29.65)^2(0.045) = 9.63$$

Notice that

$$\mu_S = \mu_X + \mu_Y \quad (29.65 = 14.65 + 15)$$

and that

$$\sigma_S^2 = \sigma_X^2 + \sigma_Y^2 \quad (9.63 = 4.23 + 5.40)$$

Here is the probability distribution of S :

S	24	25	26	27	28	29	30	31	32	33	34	35	36
$P(S)$	0.075	0.03	0.015	0.19	0.07	0.035	0.24	0.07	0.035	0.15	0.03	0.015	0.045

■ 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!

■ Alternate 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 that C = overall cost for tuition and fees for a randomly selected full-time student at El Dorado Community College and $\mu_C = \$832.50$ and $\sigma_C = \$103$.
- **Problem:** 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.

Solution: The mean is $\mu_{C+B} = \mu_C + \mu_B = 832.50 + 153 = \985.50 . The standard deviation cannot be calculated since 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.

■ Alternate Example – El Dorado Community College

■ Problem:

- (a) At the downtown campus, full-time students pay \$55 per unit. Let U = cost of tuition for a randomly selected full-time student at the downtown campus. Find the mean and standard deviation of U .
- (b) 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.

Solution:

$$(a) \mu_U = 55(15) = \$825,$$

$$\sigma_U = 55\sigma_Y = 55(2.3) = \$126.50$$

$$(b) \mu_{T+U} = \mu_T + \mu_U = 732.50 + 825 = 1557.50.$$

$$\sigma_{T+U}^2 = \sigma_T^2 + \sigma_U^2 = 10,568 + 16,002 = 26,570$$

$$\text{thus } \sigma_{T+U} = \sqrt{26,570} = \$163$$

■ 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 *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.

■ Alternate Example – El Dorado Community College

- **Problem:** 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, $D = T - U$? Interpret each of these values.

Solution: $\mu_{T-U} = \mu_T - \mu_U = 732.50 - 825 = -92.50$.

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.

$$\sigma_{T-U}^2 = \sigma_T^2 + \sigma_U^2 = 10,568 + 16,002 = 26,570; \text{ thus}$$

$$\sigma_{T-U} = \sqrt{26,570} = \$163$$

Although the average difference in tuition for the two campuses is –\$92.50, the difference in tuition for a randomly selected full-time student from each college will vary from the average difference by about \$163, on average. Notice that the standard deviation is the same for the sum of tuition costs and the difference of tuition costs.

■ Combining Normal Random Variables

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

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?

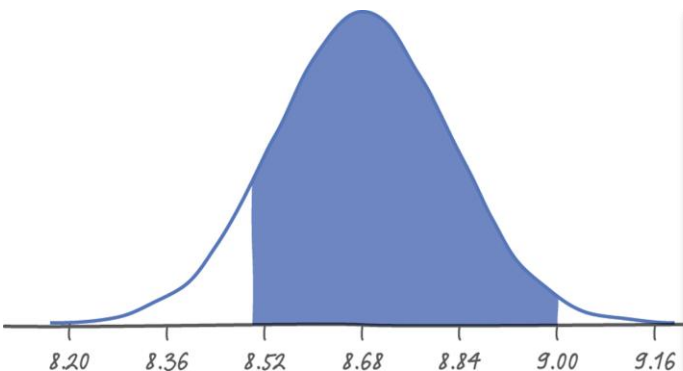
Let X = the amount of sugar in a randomly selected packet.

Then, $T = X_1 + X_2 + X_3 + X_4$. We want to find $P(8.5 \leq T \leq 9)$.

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

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

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



■ Alternate Example - Apples

Suppose that the weights of a certain variety of apples have weights that are $N(9, 1.5)$. 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?

■ *State:* What is the probability that a random sample of 12 apples has a total weight less than 100 ounces?

■ *Plan:* Let X = weight of a randomly selected apple. Then X_1 = weight of first randomly selected apple, etc. We are interested in the total weight $T = X_1 + X_2 + \dots + X_{12}$. Our goal is to find $P(T < 100)$.

■ *Do:* Since T is a sum of 12 independent Normal random variables, T follows a Normal distribution with mean $\mu_T = \mu_{X_1} + \mu_{X_2} + \dots + \mu_{X_{12}} = 9 + 9 + \dots + 9 = 108$ ounces and variance $\sigma_T^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 + \dots + \sigma_{X_{12}}^2 = 1.5^2 + 1.5^2 + \dots + 1.5^2 = 27$. The standard deviation is $\sigma_T = \sqrt{27} = 5.2$ ounces.

■ $P(T < 100) = \text{normalcdf}(-99999, 100, 108, 5.2) = 0.0620$. *Note: to get full credit on the AP exam when using the calculator command normalcdf, students must clearly identify the shape (Normal), center (mean = 108) and spread (standard deviation = 5.2) somewhere in their work.*

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

Alternate Example – Speed Dating

Suppose that the height M of male speed daters follows is $N(70, 3.5)$ and the height F of female speed daters follows a Normal distribution with a mean of 65 inches and a standard deviation of 3 inches. What is the probability that a randomly selected male speed dater is taller than the randomly selected female speed dater with whom he is paired?

State: What is the probability that a randomly selected male speed dater is taller than the randomly selected female speed dater with whom he is paired?

Plan: We'll define the random variable $D = M - F$ to represent the difference between the male's height and the female's height. Our goal is to find $P(M > F)$ or $P(D > 0)$.

Do: Since D is the difference of two independent Normal random variables, D follows a Normal distribution with mean $\mu_D = \mu_M - \mu_F = 70 - 65 = 5$ inches and variance $\sigma_D^2 = \sigma_{X_M}^2 + \sigma_F^2 = 3.5^2 + 3 = 21.25$. Thus, the standard deviation is $\sigma_D = \sqrt{21.25} = 4.61$ inches.

Thus, $P(D > 0) = \text{normalcdf}(0, 99999, 5, 4.61) = 0.8610$. *Note: to get full credit on the AP exam when using the calculator command normalcdf, students must clearly identify the shape (Normal), center (mean = 5) and spread (standard deviation = 4.61) somewhere in their work.*

Conclude: There is about an 86% chance that a randomly selected male speed dater will be taller than the female he is randomly paired with. Or, in about 86% of speed dating couples, the male will be taller than the female.



Section 6.2

Transforming and Combining Random Variables

Summary

In this section, we learned that...

- ✓ Adding a constant a (which could be negative) to a random variable increases (or decreases) the mean of the random variable by a but does not affect its standard deviation or the shape of its probability distribution.
- ✓ Multiplying a random variable by a constant b (which could be negative) multiplies the mean of the random variable by b and the standard deviation by $|b|$ but does not change the shape of its probability distribution.
- ✓ A **linear transformation** of a random variable involves adding a constant a , multiplying by a constant b , or both. If we write the linear transformation of X in the form $Y = a + bX$, the following about are true about Y :
 - ✓ **Shape:** same as the probability distribution of X .
 - ✓ **Center:** $\mu_Y = a + b\mu_X$
 - ✓ **Spread:** $\sigma_Y = |b|\sigma_X$



Section 6.2

Transforming and Combining Random Variables

Summary

In this section, we learned that...

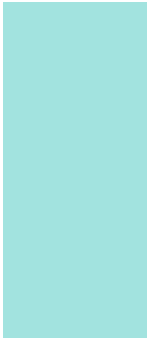
- ✓ If X and Y are any two random variables,

$$\mu_{X \pm Y} = \mu_X \pm \mu_Y$$

- ✓ If X and Y are **independent random variables**

$$\sigma_{X \pm Y}^2 = \sigma_X^2 + \sigma_Y^2$$

- ✓ The sum or difference of independent Normal random variables follows a Normal distribution.



+ Looking Ahead...

In the next Section...

We'll learn about two commonly occurring discrete random variables: binomial random variables and geometric random variables.

We'll learn about

- ✓ **Binomial Settings and Binomial Random Variables**
- ✓ **Binomial Probabilities**
- ✓ **Mean and Standard Deviation of a Binomial Distribution**
- ✓ **Binomial Distributions in Statistical Sampling**
- ✓ **Geometric Random Variables**