SCHOLASTIC

Grades 7–8 & Advanced

**BONUS: Interactive Data Analysis Tools!**
www.actuarialfoundation.org/cultivatingdata

Expect the Unexpected With Math®

A Program of The Actuarial Foundation

# Cultivating Data

## Poster/Teaching Guide for Grades 7–8 & Advanced

**Aligns with Common Core State and NCTM Standards**


Supplement to Scholastic Magazines. SCHOLASTIC and associated logos are trademarks and/or registered trademarks of Scholastic Inc. All rights reserved. "Expect the Unexpected With Math" is a registered trademark of The Actuarial Foundation. 0-545-38443-5

**What is an actuary?** An actuary is an expert in statistics who works with businesses, governments, and organizations to help them plan for the future. Actuarial science is the discipline that applies math, economics, and statistical methods to assess risk.

DEVELOPED WITH THE ACTUARIAL FOUNDATION®

Preparing for tomorrow's possibilities® *through education*

# DEAR TEACHER,

Welcome to ***Cultivating Data***, a new math program to help grades 7–8 students understand key concepts of data analysis and their graphic representations.

Developed by The Actuarial Foundation with Scholastic, this standards-aligned program provides knowledge- and skill-building materials to address math concepts in a real-world business and decision-making context that middle school students will find meaningful. We hope you enjoy this new program!

Sincerely,
The Actuarial Foundation  Scholastic Inc.

# OVERVIEW OF PROGRAM

**1.** The **POSTER/TEACHING GUIDE** includes **lessons** and **worksheets** on:

- *constructing and interpreting data tables, line plots, box-and-whisker plots,* and *stem-and-leaf plots*;
- *scatterplots with line of best fit and outliers*; and
- *calculating expected value* (advanced lesson)

It also includes a **classroom poster** to reinforce key data analysis vocabulary and graph types.

**2. The ONLINE MATERIALS**, available at **www.actuarialfoundation.org/cultivatingdata**, include:

- **Additional worksheets** to supplement the lessons
- **Standards alignment chart**
- **Supplemental lesson and worksheets** on *measures of central tendency*

**3. ONLINE BONUS:** Also, look for custom **interactive data analysis tools** that teachers can use for whiteboard class presentations and student assignments. You'll find *calculators of mean, median, mode,* and *standard deviation*; a *box-and-whisker plot generator*; and a *scatterplot generator*.

***Program Story Line:*** Entrepreneurial middle school students Jennifer and Lucas Harris have been given ten acres of their family farm to start their own business. They develop and market new products such as organic and exotic vegetables. To make business decisions, such as selecting profitable crops and optimizing promotions, they turn to data analysis tools.

**MORE FREE MATH PROGRAMS:**
**www.actuarialfoundation.org/programs/youth_education.shtml**

# Lesson 1: Representation of Statistical Information

## STANDARDS ALIGNMENT

- **Grade 7:** CCSS.Math.Content.7.SP.B.4
- **Grades 6–8:** NCTM Data Analysis and Probability

## OBJECTIVES:

**Students will be able to use statistical information to construct:**

- line plots;
- stem-and-leaf plots; and
- box-and-whisker plots.

**Time Required:** 60 minutes, plus additional time for worksheets (may be split over two or more days)

**Materials:** Worksheets 1.1 (inside), 1.2 (online), and 1.3 (online)

**Note:** The lesson is based on the assumption that students have a working knowledge of mean, median, mode, and range. If needed, a lesson and worksheets on these topics are available at **www.actuarialfoundation.org/cultivatingdata**.

## DIRECTIONS:

1. Ask the class how many phones they have in their home. Have students respond one at a time and record each answer on the board. (Note that any similar question that works for your particular class will serve the purpose.) Ask if there is a way to organize the data on the board.
2. If it isn't mentioned, suggest a **line plot**. Demonstrate how to construct a line plot. First, find the smallest and largest answers, and then construct a number line with the smallest value at one end and the largest at the other. Draw an "X" on the number line for each answer, with duplicate responses placed above previously recorded responses.
3. Point out how to use the plot to find the **range**, **mode**, and **median**. Review the meaning of these terms as necessary.
4. Indicate that a line plot can be useful when range of the data isn't very wide. To provide an example of data with a wider range, write the points scored by the San Antonio Spurs in games they played in April 2013* on the board: 90, 98, 88, 99, 86, 108, 86, 106, 95, 91, 102, 120, 103. (Feel free to select statistics for a team you and/or your class support!)
5. Demonstrate how to construct a **stem-and-leaf plot** from the data on the board:

| Stem | Leaves |
|---|---|
| 8 | 6 6 8 |
| 9 | 0 1 5 8 9 |
| 10 | 2 3 6 8 |
| 11 | |
| 12 | 0 |

*Source: nba.com/gameline/spurs

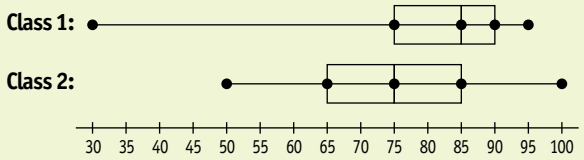
*(continued on next page)*

## Lesson 1: Representation of Statistical Information *(continued from previous page)*

6. Point out how to find the range, mode, and median using the plot and remind the class how to find the mean.
7. Indicate to the class that a **box-and-whisker plot** is another way to display and analyze data. Write test scores for two classes on the board:

   **Class 1:** 30, 75, 75, 80, 80, 85, 85, 85, 90, 95, 95

   **Class 2:** 50, 60, 65, 70, 75, 75, 80, 85, 85, 90, 100
8. Ask how the performance of the two classes can be compared.
9. Demonstrate how to construct a box-and-whisker plot for each class, showing how the five points of **minimum, first quartile, median, third quartile,** and **maximum** are determined and displayed as follows:




10. Discuss how class 1, even though their highest and lowest scores were below class 2, can be considered to have performed better because the middle two quartiles are higher on the number line.
11. Distribute **Worksheets 1.1–1.3** to students over 1–3 days, then review answers with class.

## Lesson 2: Constructing and Analyzing Scatterplots

### STANDARDS ALIGNMENT

- **Grade 8:** CCSS.Math .Content.8.SP.A.1
- **Grades 6–8:** NCTM Data Analysis and Probability

### OBJECTIVES:

**Students will be able to:**

- construct scatterplots;
- find outliers; and
- draw a line of best fit.

**Time Required:** 40 minutes, plus additional time for worksheets (may be split over two or more days)

**Materials:**

- Study Time vs. Test Score: Table and Scatterplot (online)
- Worksheets 2.1 (inside), 2.2 (online), and 2.3 (online)
- Graph paper

### DIRECTIONS:

1. Pose the following problem to your class: Do you think there is a relationship between the time studying for a test and the score earned? Discuss as a class.
2. Ask how one could determine whether or not the relationship exists (e.g., conduct a study in which data on the amount of study time and the resulting grade are captured).
3. Show the **Study Time vs. Test Score table** to the class. Ask if the table clearly shows the relationship between the two variables. Ask if there might be a clearer way to show the relationship (a **scatterplot**).
4. Show the **Study Time vs. Test Score scatterplot** to the class. Point out that the **independent variable** (which could be thought of as the cause) is study time and is on the X-axis. The **dependent variable** (the effect)—in this case the test score—is on the Y-axis.
5. If the class is comfortable creating and reading scatterplots, take a few data points from the table and show where they appear on the scatterplot. If the class has less experience with scatterplots, show how all the points on the table are represented on the scatterplot.
6. Ask if the scatterplot shows a relationship between study time and test score and, if so, how. Point out that the general shape the points make slope upward to the right, showing that, in general, scores increase as study time increases. This pattern is known as a **positive correlation**. If the scatterplot sloped downward to the right, we might say it has a **negative correlation**. An example of a negative correlation might be apparent if we made a scatterplot of hours of television watched versus test scores. Sometimes, no relationship is apparent between the variables, e.g., if we plotted the number of letters in a person's first name versus the number of letters in his or her last name.
7. Call the class's attention to the points on the scatterplot for Sloane (studied 25 minutes and received a score of 98) and Rick (studied 125 minutes and received a score of 77). Ask if these two points appear to follow the trend observed of more study time leading to a higher score. Indicate that these points are called "**outliers**" because they fall outside the pattern formed by the other points. Ask if these two points represent possible data collection errors or if there could be a reasonable explanation for why they fall outside the pattern.
8. Ask if it would be possible to use the scatterplot to predict what a person would score if they studied for, say, 40 or 100 minutes. Show how to draw a **line of best fit** by "eyeballing" the points on the scatterplot. Inform the class that the line doesn't have to go through any of the points, but it's easiest to draw the line by identifying two representative points and connecting them. The resulting line can be used to make predictions, either by finding a point on the line or by determining the formula of the line and plugging in values. If the class is sufficiently advanced, demonstrate how to determine line of best fit with the least squares method or on a graphing calculator.
9. Distribute **Worksheets 2.1–2.3** to students over 1–3 days, then review answers with class.

*(continue to next page for Lesson 3)*

# Advanced Lesson: Expected Value

## STANDARDS ALIGNMENT

- **High School:** CCSS.Math.Content.HSS-MD.B.5b
- **Grades 9–12:** NCTM Data Analysis and Probability

## OBJECTIVES:

**Advanced (beyond grade 8) math students will be able to:**

- Calculate and interpret expected value; and
- Explore using the mean and deviation from it to draw conclusions.

**Time Required:** 40 minutes, plus additional time for worksheet

**Materials:** Worksheet 3.1 (inside)

## DIRECTIONS:

1. Pose the following problem to your class: A middle school volleyball team has 20 players. Four of the players are 5'4" (64") tall, four are 5'6"(66"), four are 5'8" (68"), three are 5'10"(70"), and five are 6'(72"). If the coach picked a player at random, how tall would you expect the player to be? Discuss possible solution methods with the class, including finding the mean by adding all the heights and dividing by 20.
2. Indicate that an **expected value** calculation is another way to determine the solution. Explain that expected value is a calculation that predicts value by summing the values of all possible outcomes, with each outcome multiplied by the probability of its occurring. Demonstrate how to calculate the expected value of the height of a volleyball player picked at random. First, make a frequency table showing the number of times a specific height is represented on the team.

| Height | 64" | 66" | 68" | 70" | 72" |
|---|---|---|---|---|---|
| Frequency | 4 | 4 | 4 | 3 | 5 |

3. Then, show how to calculate expected value by writing the following on the board:

   20%(64") = 12.8"

   20%(66") = 13.2"

   20%(68") = 13.6"

   15%(70") = 10.5"

   25%(72") = 18.0"

   Total = 68.1" (Expected Value)

   First, divide the frequency of each height by the total number of players, resulting in a percentage (for example, 3 players at 70" divided by 20 players = 3/20 = .15 = 15%). Then, multiply the result by each respective height represented on the team (for example, 15%(70") = 10.5". Finally, add the products together to find the expected value.
4. Explain that the total equals the expected value of the height of a player picked at random. Further, explain how expected value is the predicted value of a future event, taking into account each possible outcome multiplied by its probability.
5. Note that in this data set, the expected value is not the height of one specific individual on the team. Indicate that this is not uncommon with statistical measures, e.g., the mean or, in some cases, the median of a group of numbers.
6. Connect expected value to mean as follows with the following example: Consider the numbers 3, 3, 4, 5, 5. The mean is 4 [(3 + 3 + 4 + 5 + 5)/5]. Another way to think about this group is that 3 makes up 40% of the group (2/5), 4 makes up 20%, and 5 makes up 40%. So:

   .4(3) = 1.2

   .2(4) = 0.8

   .4(5) = 2.0

   Expected Value/Mean = 4.0
7. Give this problem to show the practical use of calculating expected value:

   A town carnival includes a game of chance to benefit the local food bank. Each player pays $5 to spin a prize wheel. The wheel is divided into the following segments:

   50%: No Prize

   25%: $1 Prize

   20%: $5 Prize

   5%: $50 Prize

   Ask: "Will the food bank make money on this game? [Yes, the expected value of each play to the charity is $1.25 or $5 – (50%(0) + .25($1) + .20($5) + .05($50)]. Write this calculation on the board and explain as necessary. Note that the expected value of the total profit for the evening for the food bank is $1.25 times the number of people who play the game.

   Note that on any one spin of the wheel, the charity will either win $5 or $4, break even, or lose $45. However, when the outcomes are weighted by the probability of their occurrence, the result is an expected profit of $1.25. The food bank is counting on many people playing the game so that, in total, the actual outcomes of the spins approach their theoretical probabilities.

   For more information about theoretical versus experimental probability, see: **www.actuarialfoundation.org/probabilitychallenge/lab**.
8. Distribute **Worksheet 3.1** to students, then review answers with class.

**EXTENSION:** For those students with sufficient background knowledge and a strong passion for mathematics needing an additional challenge in advanced topics like standard deviation, direct them to the self-study program at: **www.actuarialfoundation.org/pdf/Probstat_SG_2012.pdf**, beginning on page 24.

# More Rutabagas, Please!

Middle school students Jennifer and Lucas Harris's 200-acre family farm has grown traditional crops such as corn and wheat for four generations. The siblings are both interested in business and feel the farm could prosper if they added a few modern touches, like growing organic and exotic crops. Their parents are willing to try out these new ideas, and put Lucas and Jennifer in charge of ten acres to see how successful they are on a more limited scale.

One of their first steps is to introduce themselves and their excellent produce to local chefs. They make a positive impression on celebrity chef Wunderbar, owner of Le Trend Magnifique, the hottest restaurant in town. "Your rutabagas are *très magnifique*!" he exclaimed. "I will place an order with you daily, so they are at the peak of freshness!"

Jennifer and Lucas receive good publicity from Wunderbar, but his orders are hard to predict. Jennifer created a table (below) to show the rutabagas he ordered each day for four weeks.

## WORK THE MATH

*Use separate paper to show your work.*

**Pounds of Rutabagas Ordered by Chef Wunderbar**

| Sunday | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday |
|---|---|---|---|---|---|---|
| 4 | 8 | 0 | 1 | 12 | 16 | 20 |
| 7 | 0 | 0 | 6 | 18 | 20 | 20 |
| 5 | 9 | 2 | 10 | 5 | 18 | 16 |
| 8 | 12 | 3 | 16 | 18 | 20 | 20 |

1. Prepare a **line plot** to display chef Wunderbar's orders for the past 4 weeks.
2. Use the line plot to calculate the **median**, **mode**, and **range** for the data.

## NOW TRY THIS:

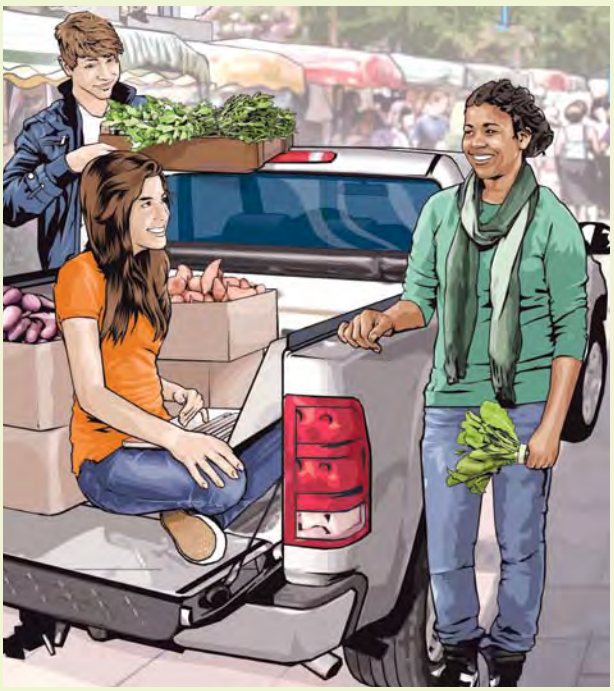
Using the information in the above table and the line plot you prepared, explain why the mode would be much higher than the mean or median order.

# That's a Lotta Lettuce!

Middle school farmers Jennifer and Lucas Harris have a strong entrepreneurial spirit and want to update their operations by using modern business practices and technology.

One of their early ventures is growing specialty baby arugula that is shipped fresh to customers. While Lucas sells locally, Jennifer has been experimenting with offering free shipping on Internet orders. She tested different minimum order amounts to qualify for free shipping.

After eight weeks, she isn't sure whether her program is working. While setting up their stand at the farmers' market, she shares her questions, as well as a table of sales data, with Priscilla, owner of the Jammin' Jam Shack. Priscilla suggests that a scatterplot might be a good way to determine whether or not the free shipping program was working.



## WORK THE MATH

*Use separate paper to show your work.*

| Minimum Purchase to Receive Free Shipping | Total Baby Arugula Sales |
|---|---|
| $50 | $800 |
| $75 | $650 |
| $100 | $400 |
| $30 | $1,400 |
| $25 | $1,600 |
| $40 | $1,250 |
| $0 | $2,500 |
| $60 | $600 |

1. Create a **scatterplot** to display the data in the table.
2. Is there a positive, negative, or no correlation between total baby arugula sales and the minimum purchase amount to receive free shipping? Explain your thinking.
3. Which variable is independent and which is dependent? Explain your thinking.

## NOW TRY THIS:

What do you think might happen if Jennifer and Lucas didn't offer free shipping at all?

# You Say Tomato!

Middle school students Jennifer and Lucas Harris would like to devote some acreage from their thriving new vegetable business to heirloom tomatoes, which they've noticed are big sellers at the farmers' market. "There are some juicy varieties that look great and taste even better!" exclaimed Lucas. "But with all those varieties, how will we decide which seeds to invest in?" Jennifer wondered.

Jennifer and Lucas met with Raymond "Red" Vines, the local representative of the Herbivore Seed Company. "Well," said Red, "I can't predict which tomato variety will give the biggest yield in the future, but I can show you data on how my customers have done over the past four years with my two best-selling seeds."

## WORK THE MATH

*Use separate paper to show your work.*

**Tomato Seed Yield (Tons per Acre)**

| Tomato Variety | Year | Farm 1 | Farm 2 | Farm 3 | Farm 4 | Farm 5 |
|---|---|---|---|---|---|---|
| Red Ready | 2009 | 19 | 15 | 11 | 14 | 20 |
| | 2010 | 16 | 18 | 16 | 16 | 16 |
| | 2011 | 10 | 19 | 11 | 12 | 16 |
| | 2012 | 14 | 12 | 20 | 16 | 18 |
| Crimson Champ | 2009 | 26 | 0 | 12 | 20 | 17 |
| | 2010 | 0 | 14 | 10 | 26 | 28 |
| | 2011 | 30 | 28 | 16 | 17 | 26 |
| | 2012 | 10 | 16 | 17 | 10 | 17 |

1. Prepare a **frequency table** for each variety of tomato.
2. Calculate the **expected value** for each variety.
3. How risky do you think it is for Jennifer and Lucas to rely on Red's data?

## NOW TRY THIS:

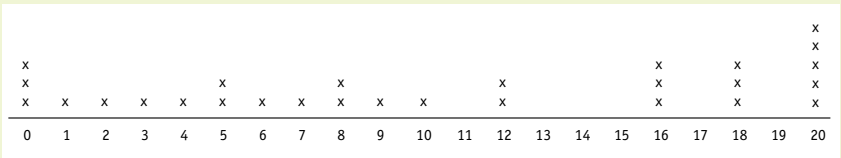
What would you advise Jennifer and Lucas to do about the tomatoes? *Hint:* Be sure to consider not only the expected value of each variety but also how risky the yield for each variety is. If the yield can deviate (vary) widely from the mean, the poor yield in some years could be a financial disaster.

If you would like to learn more about the topic of variance from the mean, including standard deviation, visit **www.actuarialfoundation.org/pdf/Probstat_SG_2012.pdf**, beginning on page 24.

# ANSWER KEY For Online Materials, Visit: www.actuarialfoundation.org/cultivatingdata

## Worksheet 1.1: "More Rutabagas, Please!"

1.




2. **Median:** 9.5 pounds
**Mode:** 20 pounds
**Range:** 20 pounds

**NOW TRY THIS:** Wunderbar's biggest orders come on the weekends, when his restaurant is busier than on weeknights. He places much larger orders (16, 18, or 20 pounds) on those days alone. The lower weekday orders bring down the median.

## Worksheet 1.2: "Love Those Leafy Greens!" (online)

1.

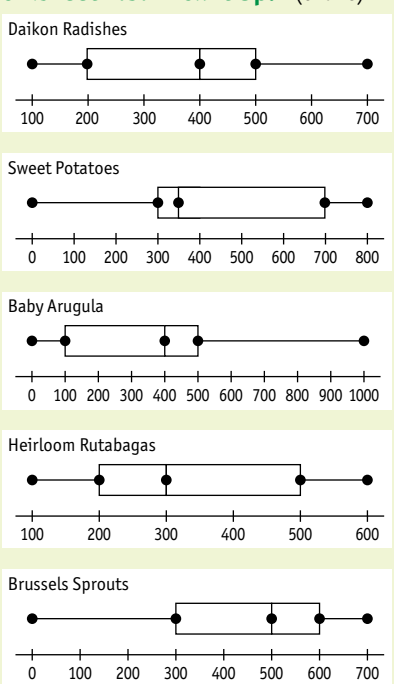
| Stem | Leaves |
|---|---|
| 2 | 44 |
| 3 | 467 |
| 4 | 02567899 |
| 5 | 00000022244566 |
| 6 | 002 |

2. **Median:** 50; **Mode:** 50; **Range:** 38

**NOW TRY THIS:** The median and mode of 50 are reasonable measures of the typical number of customers. Although there are a number of lower values, it can be attributed to slow business during the first few days of operation.

## Worksheet 1.3: "Box It Up!" (online)

1.

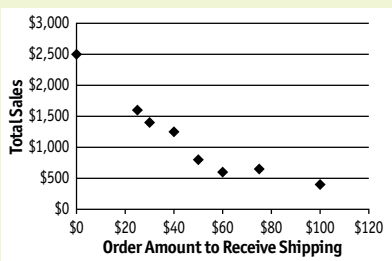



2. Sweet potatoes and Brussels sprouts have the highest interquartile ranges.

**NOW TRY THIS:** Answers will vary, depending on the vegetables selected. Although the box-and-whisker plot takes longer to prepare, it has the advantage of providing information about interquartile ranges.

## Worksheet 2.1: "That's a Lotta Lettuce!"

1.



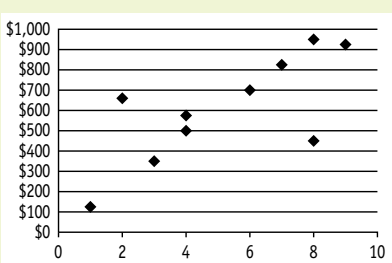

2. There is a negative correlation. The lower the order amount to receive free shipping, the higher total sales are.

3. The order amount for free shipping is the independent variable. Total sales is the dependent variable.

**NOW TRY THIS:** It's impossible to tell for sure. Sales might decline further, but it may also be possible that some customers simply want the finest vegetables and would continue to buy even without free shipping.

## Worksheet 2.2: "Pounding the Pavement!" (online)

1.

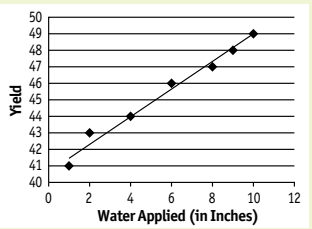



2. Although there are outliers, there is a positive correlation between hours worked and sales.

**NOW TRY THIS:** Answers will vary, but the day with eight hours worked with only $450 in sales might have been a rainy, slow day while the day with only two hours worked and $660 in sales might have been right before July 4, with chefs stocking up for a busy weekend.

## Worksheet 2.3: "The Future of Our Farm" (online)

1. & 2.




3. 3 inches of water = a yield of approximately 43.5
7 inches of water = a yield of approximately 46.5

**NOW TRY THIS:** Answers will vary, but it is important for students to realize that the line of best fit is a model used to make predictions but it can't be followed blindly. Common sense dictates that the relationship between applying water and increasing yields can't continue forever.

## Advanced Worksheet: "You Say Tomato!"

1. Frequency tables:

Red Ready

| Number of Tons | Frequency of Number | Probability of Number = Frequency / Total |
|---|---|---|
| 10 | 1 | 5% |
| 11 | 2 | 10% |
| 12 | 2 | 10% |
| 14 | 2 | 10% |
| 15 | 1 | 5% |
| 16 | 6 | 30% |
| 18 | 2 | 10% |
| 19 | 2 | 10% |
| 20 | 2 | 10% |

Crimson Champ

| Number of Tons | Frequency of Number | Probability of Number = Frequency / Total |
|---|---|---|
| 0 | 2 | 10% |
| 10 | 3 | 15% |
| 12 | 1 | 5% |
| 14 | 1 | 5% |
| 16 | 2 | 10% |
| 17 | 4 | 20% |
| 20 | 1 | 5% |
| 26 | 3 | 15% |
| 28 | 2 | 10% |
| 30 | 1 | 5% |

2. Expected values:
Red Ready = 15.45
Crimson Champ = 17.0

3. Answers will vary, but could include mention that current crop yields might differ from yields in prior years due to drought, excessive rain, extreme temperatures, pests, etc. Additionally, students might mention that a sample of 20 might not be large enough to ensure a high level of comfort.

**NOW TRY THIS:** Crimson Champ has a higher expected value than Red Ready. However, Red Ready has a more "dependable" yield in that, in prior years, yields stayed fairly close to the mean while yields for Crimson Champ deviate more widely from the mean. Additionally, with Crimson Champ, there is a 10% chance of no yield at all. Crimson Champ offers higher potential reward, but at the price of higher risk.