Unit 7: Statistics with bivariate data

8.4B, 8.5C, 8.5D, 8.5I, 8.11A

Data is all around us! In this unit, students learn to make scatter plots and analyze patterns that they find. They will learn important skills to be able to better interpret the world around them.

- Make scatter plots
- Interpret and describe patterns in scatter plots
- Approximate a line of best fit
- Write an equation for the line of best fit

<table>
<thead>
<tr>
<th>TEKS standards</th>
<th>Common misconceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>8.4B:</strong> Graph proportional relationships, interpreting the unit rate as the slope of the line that models the relationship</td>
<td>“Every point should be on the line”</td>
</tr>
<tr>
<td><strong>8.5C:</strong> Contrast bivariate sets of data that suggest a linear relationship with bivariate sets of data that do not suggest a linear relationship from a graphical representation</td>
<td>“The axes labels don’t matter”</td>
</tr>
</tbody>
</table>

**How to help:** Remind students that a line of best fit is more like a trend line, showing the overall pattern, not a connect-the-dots exercise. The line should be an approximation of the pattern of the data and doesn't actually need to touch ANY of the data points. Have students use dry spaghetti or coffee stirrers to approximate lines of best fit and then draw them on their graphs.

**How to help:** Whenever working on a problem with a graph, model how to start interpreting the graph by looking at the axes labels. Ask students what each axis represents and have them interpret general patterns using the context of the graph. For example, “as time increases, the amount of water in the pool decreases” instead of simply saying “as
one increases, the other decreases.” Graphs can tell us a lot of important information and we need to teach students how to read them thoroughly.

“Correlation equals causation” | Just because two variables might show a trend together, it doesn’t mean one is causing the other. There are almost always other factors at play!

**How to help:** When your students notice a relationship between two variables, have them try to come up with other variables that might also be contributing to that relationship. For more information, see the “Best practices,” below.

<table>
<thead>
<tr>
<th>Standard</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.5D:</td>
<td>Use a trend line that approximates the linear relationship between bivariate sets of data to make predictions</td>
</tr>
<tr>
<td>8.5I:</td>
<td>Write an equation in the form $y = mx + b$ to model a linear relationship between two quantities using verbal, numerical, tabular, and graphical representations</td>
</tr>
<tr>
<td>8.11A:</td>
<td>Construct a scatter plot and describe the observed data to address questions of association such as linear, non-linear, and no association between bivariate data</td>
</tr>
</tbody>
</table>
Unit resources

- For the first video in Lesson 1, use this 8.7.1 Video Notetaking Guide for students to graph along with Sal.
- For the videos in this unit, use the Learning summary video notetaking guide.
- For the articles in this unit, use the Article notetaking guide.
- For the exercises in this unit, use the Blank workspace template.
- To record key terms and information, use the Vocabulary and notation notetaker.

Lesson overview

<table>
<thead>
<tr>
<th>Lesson</th>
<th>Objective</th>
<th>Teaching tips</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lesson 1: Introduction to scatter plots</strong></td>
<td>Students will be able to accurately plot points in a scatter plot and determine the best scaling for the axes.</td>
<td>- <strong>Warm up activity:</strong> Give students a scatter plot and ask them to interpret it (what is this data showing? What is happening at point x on the graph? etc.). This will remind them of scatterplots and prepare them for the first video (and use the specific notetaking guide, above, for the video).&lt;br&gt;- Model how to read problems for important details and then discuss appropriate labeling and scaling.&lt;br&gt;- Use this 8.7.1 Video Notetaking Guide for students to graph along with Sal.</td>
</tr>
<tr>
<td><em>TEKS standard: 8.11A</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lesson 2: Interpreting scatter plots</strong></td>
<td>Students will be able to describe patterns in scatter plots as positive linear, negative linear, nonlinear, or no association and can interpret the relationship of two variables in contexts.</td>
<td>- <strong>Watch the first video together as a class.</strong> There is some new vocabulary: positive linear, negative linear, nonlinear, and no association. Stop the video often to go over vocabulary, answer questions, and check for understanding. Make sure students are taking good notes and include examples of data association patterns in their work.</td>
</tr>
<tr>
<td><em>TEKS standard: 8.5C, 8.11A</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lesson 3: Fitting trend lines to scatter plots</strong></td>
<td>Students will be able to approximate lines of best fit graphically, find the equation of a line of best fit, and interpret the slope and y-intercept of the line of best fit in context.</td>
<td>- <strong>As students begin estimating their own lines of best fit, have them use a piece of dry spaghetti or a coffee stirrer to estimate their line by holding it up to their paper or computer screen.</strong>&lt;br&gt;- <strong>Before the third exercise, show students how to plug an x-value into the equation and solve for y.</strong> You can do one of the problems in this exercise together. See &quot;Best practices,&quot; below, for more.</td>
</tr>
<tr>
<td><em>TEKS standard: 8.4B, 8.5D, 8.5I</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Best practices

Correlation ≠ Causation
When interpreting data, it’s important to remember that correlation does not necessarily imply causation. In other words, if we see a pattern in data, it does not necessarily mean that a trend in one variable is influencing a trend in another. There could be many forces at play underlying any relationships we see.

For example, let’s say you’re reading a research article and it concludes that when ice cream sales increase, the prevalence of migrating songbirds also increases. They conclude that increased ice cream consumption leads to an increased prevalence of migrating songbirds (if you eat more ice cream, you are encouraging more songbirds to migrate).

Now, there could be a correlation between ice cream consumption and the prevalence of migrating songbirds, meaning that there is a relationship between the two. In this case, as the incidents of one occurring increases, so do the incidents of the other. But, we cannot say that one causes the other because we just don’t have enough information. There are many other things that could be at play, like the fact that both ice cream sales and songbird migration increase in the summer when the temperatures are warmer. For another example with more explanation, check out the Khan Academy video Correlation and causality.

More on describing patterns in scatter plots
In this unit, we use the following words to describe the patterns that we see in data in scatter plots. In later courses, students will also interpret strength, outliers, and clusters.

**Linear association:** If the data in a scatter plot form an approximate line, or can be estimated reasonably with a straight line then it is said to have a linear association. These graphs both have a linear trend.

**Nonlinear association:** If the data in a scatter plot show a pattern that is not a straight line then it is said to have a nonlinear association. These graphs have a clear trend, but it is not linear.

**No association:** If the data in a scatter plot show no discernable patterns then we say that the data has no association. These graphs don’t seem to have any pattern at all.

**Positive linear association:** If the data in a scatter plot show a linear association and the y-values increase as the x-values increase, then it has a positive linear association. In these graphs, the pattern of the points shows a linear trend and the line of best fit would have a positive slope.
Negative linear association: If the data in a scatter plot show a linear association and the y-values decrease as the x-values increase, then it has a negative linear association. In these graphs, the pattern of the points shows a linear trend and the line of best fit would have a negative slope.

Finding the equation of the line of best fit, making predictions, and interpreting a graph

Let’s look at an example together. This graph shows a data sample that compares a person’s number of hours spent playing sports in a week to their mood rating. The line of best fit has been provided for us.

To find the equation of the line of best fit, we need the slope and the y-intercept first, then we can substitute those into the slope-intercept form of a linear equation. Let’s look at the steps to do this.

1. First, we need to find two points on the line so we can find the slope. I’ve chosen two points on the line that appear to have integer coordinates, (0, 5) and (2, 8), the red points on this graph (that are not part of the original data in this case). I can choose any two points on the line to find the slope.

   
   \[ m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{8 - 5}{2 - 0} = \frac{3}{2} \]

2. Now that we know the slope, we can find the y-intercept. In this case, I can see that the y-intercept is (0, 5) on the graph, but that won’t always be the case, so let’s do it algebraically. We know so far that the equation is \( y = \frac{3}{2}x + b \) and we know two points on the line. We can plug in either one of those points to solve for \( b \). Let’s use (2, 8).

   \[ y = \frac{3}{2}x + b \]

   \[ (8) = \frac{3}{2}(2) + b \]

   \[ 8 = 3 + b \]

   \[ 5 = b \]

Let’s use the equation to predict someone’s mood rating if they had spent 4 hours playing sports in a given week. The hours playing sports is an x-value, so we’ll plug in 4 for x and solve for y.

   \[ y = \frac{3}{2}(4) + 5 \]

   \[ y = \frac{12}{2} + 5 \]

   \[ y = 6 + 5 \]

   \[ y = 11 \]
This tells us that, using our line of best fit, we predict that if someone plays sports for 4 hours per week, their mood rating would be 11. If you go back and look at the graph and extend the line, that looks pretty right. Let's do some interpreting.

What does the $y$-intercept mean in this situation? In this situation, the $y$-intercept would be $(0, 5)$. This means that if someone were to play zero hours of sports per week, we would predict that they would have a mood rating of 5.

What does the slope mean? The slope in this problem is $\frac{3}{2} = \frac{\Delta y}{\Delta x} = \frac{\text{rise}}{\text{run}} = \frac{\text{mood rating}}{\# \text{ hours playing sports}}$ which means that for every increase of 2 hours in sports play per week, it is expected that the mood rating would go up three points.

CLASSROOM ACTIVITY

Collect your own class data

Make scatter plots from class data! You can collect data from your students on just about anything and then they can make scatter plots that represent them. For example, you could use student wingspan and shoe size or number of siblings and number of pets.

Students can also collect their own numerical data from classmates, family members, or other people. What are they interested in? Then, they can make a graph, find an equation for the line of best fit, and make their own predictions!

GENERAL CLASSROOM IMPLEMENTATION RESOURCES:

- **Weekly Khan Academy Quick Planning Guide**: Use this template to plan your week using Khan Academy.
- **Using Khan Academy in the Classroom**: Learn teaching techniques and strategies to support your students and save time with Khan Academy.
- **Differentiation Strategies for the Classroom**: Discover strategies to support the learning of all students.