Linear Regressions

Topic 3 Overview

How is *Linear Regressions* organized?

Students began this module, **Searching for Patterns**, by analyzing patterns in graphs of functions. In *Sequences*, they analyzed patterns in sequences of numbers, including both arithmetic and geometric sequences. In *Linear Regressions*, students focus on the patterns that are evident in certain data sets and use linear functions to model those patterns. Students advance their statistical methods to make predictions about real-world phenomena.

First, students explore a data set, represent it with a scatter plot, and estimate lines of best fit based on observable patterns. Recognizing that the estimates vary widely and provide very different predictions about the model, students learn about the Least Squares Method and how to use technology to determine the regression equation. Students need access to a handheld graphing calculator or to any online calculator or application that can render such equations. Students use the linear regression equations to make and assess predictions, and they differentiate between *extrapolation* and *interpolation*.

Next, students learn to use correlation coefficients to measure the appropriateness of a linear fit. They analyze the formula for the correlation coefficient but use graphing technology to actually calculate the value. Finally, they differentiate between *correlation* and *causation*, recognizing that a correlation between two quantities does not necessarily mean that there is a causal relationship.

What is the entry point for students?

In grade 8, students investigated patterns of association between two quantities. They informally fit straight lines to model data sets, determined the equations of those lines, interpreted the slopes and *y*-intercepts of the lines, and used the equations to make and judge the reasonableness of predictions about the data. They have also informally assessed the fit of proposed linear models. Students have examined linear relationships and recognized that the slope of a line defines its steepness and direction. In *Linear Regressions*, students build upon their foundational experiences and learn formal strategies to write a function to model a data set and to quantify the fit of a function for the data. They use the correlation coefficients to formalize the associations they recognize in the scatter plots. Students move from an intuitive understanding of patterns in bivariate data to a formal understanding of linear regressions.

How does a student demonstrate understanding?

Students will demonstrate understanding of the standards in *Linear Regressions* if they can:

- Determine when a data set should be modeled by a linear function.
- Use technology to determine the least squares regression for a data set.
- Make predictions using the line of best fit.
- Interpret the meaning of the slope and *y*-intercept of a linear regression in terms of the problem context.

- Explain the correlation coefficient as a measure of how well a function fits a data set.
- Understand that the sign of a correlation coefficient indicates the direction of the association and that the magnitude indicates the strength of the fit.
- Use technology to compute the correlation coefficient.
- Analyze a function on a scatter plot and its correlation coefficient to determine whether a function is an appropriate fit for a data set.
- Recognize that correlation does not imply causation.
- Determine whether there is a causal relationship between two correlated variables and defend their reasoning.

Why is *Linear Regressions* important?

Using and analyzing linear regressions to model data is an important bridge between the first two topics in Module 1 and the concepts students will encounter in Module 2, **Exploring Constant Change**. Analyzing linear regressions requires students to contextualize the slope and y-intercept of a function, reminding them of what they learned in grade 8 and preparing them for a more abstract exploration in the next topic. Because students have had more extensive experiences with linear relationships, it is important that they initially learn to fit a function to linear data. However, not all data can be modeled by linear functions. As they continue in this course and in high school mathematics, students will determine and analyze more complicated regressions, including exponentials, quadratics, and higher-order polynomials. The ability to assess the fit of a function is an important skill in choosing the type of function to model a given data set. Using the correct function type leads to more accurate predictions, the ultimate goal of calculating regression curves.

How do the activities in *Linear Regressions* promote student expertise in the mathematical process standards?

All Carnegie Learning topics are written with the goal of creating mathematical thinkers who are active participants in class discourse, so elements of the mathematical process standards should be evident in all lessons. Students are expected to make sense of problems and work towards solutions, reason using concrete and abstract ideas, and communicate their thinking while providing a critical ear to the thinking of others.

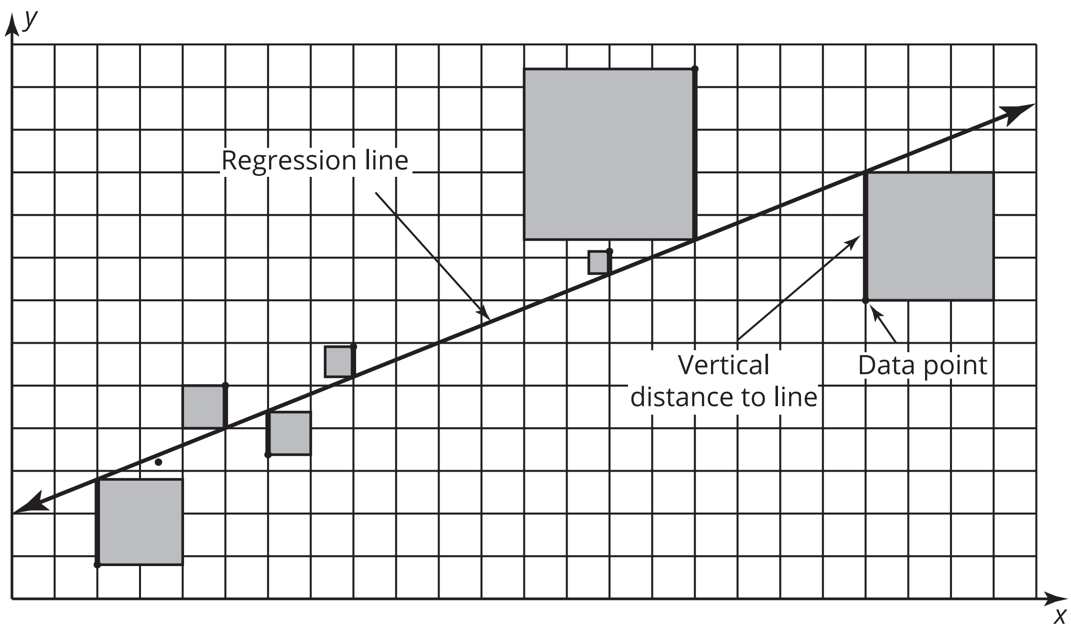
The activities in *Linear Regressions* require students to reason about and describe the patterns of association on scatter plots and the fitness of lines. Modeling with mathematics is prevalent throughout the topic. Scenarios are presented for students to explore, providing them an opportunity to notice patterns and wonder about possible relationships. They use the scatter plots to organize and visualize the data and the regression equations to mathematize the relationship. They use the correlation coefficient to analyze the fit of the linear model. They then test their equations by making predictions and analyzing their reasonableness. In some cases, students will have to return to their linear models and consider a different function type to better fit the data. Finally, they report their results.

Materials Needed

Graphing technology
Uncooked spaghetti

New Tools and Notation

Technology can be used to determine a line of best fit or regression line for a data set. Graphing technology uses the Least Squares Method to calculate the regression line. The least squares line has the smallest possible vertical distances from each given data point to the regression line. The sum of the squares of those distances are at a minimum with this line.



Learning Together

ELPS: 1.A, 1.C, 1.E, 1.F, 1.G, 2.C, 2.E, 2.I, 3.D, 3.E, 4.B, 4.C, 5.B, 5.F, 5.G

Lesson	Lesson Name	TEKS	Days	Highlights
1	Like a Glove: Least Squares Regressions	A.3C A.4C A.12A	2	Students informally determine a line of best fit by visual approximation of a hand-drawn line. They are then introduced to a formal method to determine the linear regression line of a data set using graphing technology; the mathematics behind the calculator function is explained using the related terms *Least Squares Method*, *regression line*, and *centroid*. Students then use the line of best fit to make predictions and distinguish between the terms *interpolation* and *extrapolation*.
2	Gotta Keep It Correlatin': Correlation	A.4A A.4B A.4C	2	This lesson provides several definitions related to correlations. The terms *correlation* and *correlation coefficient* are defined. The formula to compute the correlation coefficient is given; however, students are only required to use technology to determine the value of r or to estimate correlation coefficients from a list of choices. The distinction is then made between the meanings of r and r^2, the coefficient of determination. Students use these terms to make decisions regarding the model that best fits the data. It is suggested that students revisit the modeling process as they solve these problems in context. The terms *causation, necessary condition*, and *sufficient condition* are defined. Examples are provided to help students see the difference between correlation and causation. The terms *common response* and *confounding variable* are defined as relationships often mistaken for causation.

Suggested Topic Plan

*1 Day Pacing = 45 min. Session

Day 1	Day 2	Day 3	Day 4	Day 5
TEKS: A.3C, A.4C, A.12A **LESSON 1** **Like a Glove** **GETTING STARTED** **ACTIVITY 1**	**LESSON 1** continued **ACTIVITY 2** **ACTIVITY 3** **TALK THE TALK**	MATHia® Use LiveLab and Reports to monitor students' progress	TEKS: A.4A, A.4B, A.4C **LESSON 2** **Gotta Keep Correlation'** **GETTING STARTED** **ACTIVITY1** **ACTIVITY2**	**LESSON 2** continued **ACTIVITY 3** **ACTIVITY 4** **TALK THE TALK**

Day 6	Day 7
MATHia® Use LiveLab and Reports to monitor students' progress	**END OF TOPIC ASSESSMENT**

Assessments

There is one assessment aligned to this topic: End of Topic Assessment.