

Module 1: Searching for Patterns

TOPIC 3: LINEAR REGRESSIONS

In this topic, students focus on the patterns that are evident in certain data sets and use linear functions to model those patterns. Using the informal knowledge of lines of best fit that was built in previous grades, students advance their statistical methods to make predictions about real-world phenomena. They differentiate between correlation and causation, recognizing that a correlation between two quantities does not necessarily mean that there is also a causal relationship. And, at the end of this topic, students will synthesize what they have learned to decide whether a linear model is appropriate.

Where have we been?

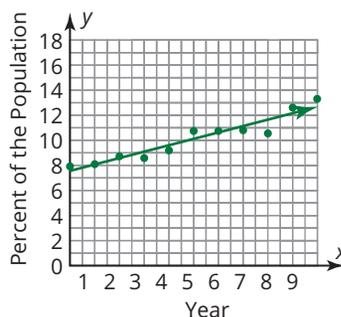
Students have analyzed the shape of data, informally fit lines of best fit to model data sets, determined the equations of those lines, interpreted the slopes and y -intercepts of the lines, and used the equations to make and judge the reasonableness of predictions about the data. Students have also examined linear relationships and recognized that the slope of a line defines its steepness and direction.

Where are we going?

Using and analyzing linear regressions to model data is an important bridge between the first two topics in the first module of this course and the concepts students will encounter in the next module, **Exploring Constant Change**. As students continue in this course and in high school mathematics, they will determine and analyze more complicated regressions, including exponential and quadratic regressions.

Regression Lines

Real-world data points never fit neatly on a line. But you can model the data points using a line, which represents a linear function.



There are an infinite number of lines that can pass through the collection of data points. But there is just one line that models the data with the minimum distances between the data points and the line.

Now Hiring

According to a report published in 2017, the demand for data scientists, data developers, and data engineers will grow nearly 40% over the next few years, paying an average starting salary of over \$80,000 per year.

Currently, approximately 88% of data scientists have a master's degree and 46% have a Ph.D. in math, statistics, computer science, or engineering.

Talking Points

Regressions can be an important topic to know about for college admissions tests.

Here is a sample question:

The data in the table show test scores after certain amounts of study time. Use a linear regression to estimate the score associated with a study time of 20 minutes.

Score	Time (min)
86	45
70	15
90	40
78	35

Time is the independent variable, so the time values must be entered as x 's and score values as y 's in a graphing calculator.

This yields a linear regression equation of $y = 0.61x + 60.51$. A study time of 20 minutes would yield an estimated score of $y = 0.61(20) + 60.51$, or 72.71.

Key Terms

regression line

For a scatter plot of data, the regression line is a mathematical model that can be used to predict the values of a dependent variable based upon the values of an independent variable.

interpolation

Using a linear regression to make predictions within the data set is called interpolation.

extrapolation

To make predictions for values that are outside of the data set is called extrapolation.

correlation

A measure of how well a regression fits a set of data is called correlation.