CARNEGIE MATH PATHWAYS
WestEd

Corequisite

User Guide

Goal: College-Level Credit

The goal of the Statway Corequisite offering is to provide a one-term, college-level statistics course for all students, with targeted supports built using the Carnegie Math Pathways design principles to provide effective scaffolding for students who need additional support. This offering consists of the one-term, college-level statistics course (Statway College) and a corresponding corequisite component (Statway Corequisite). The corequisite component develops students' statistical and algebraic reasoning skills in support of the learning outcomes of the college-level statistics course. This offering addresses all the learning outcomes of Statway Pathway, and incorporates the same pedagogical strategies.

Main Concepts

Statway College. Statway College covers all major concepts typically addressed in a college-level statistics course. It consists of seven modules:

1) Statistical analysis process, data visualization, measures of center, shape, and spread
2) Probability, probability distribution, and normal distribution
3) Sampling distribution, confidence intervals, and hypothesis testing for sample proportions
4) Sampling distribution, confidence intervals, and hypothesis testing for sample means
5) Two-sample tests, ANOVA, and choosing the appropriate test
6) Correlation and regression
7) Chi-square

Statway Corequisite. Statway Corequisite addresses the foundational concepts and skills necessary for achieving the college-level learning objectives, and is designed to optimize support for the college-level and developmental mathematics learning objectives as needed.

The primary topic of each Statway College and Statway Corequisite lesson is summarized below. This table can help you plan your syllabus and timeline.

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

Module	Statway College Lesson	SW College Topic	Estimated Time	Statway Corequisite Lesson	SW Corequisite Topic	Estimated Time
1	1.1	Introduction to Data & Statistics	50	1.1	Numeracy: conversion and equivalence -fractions, decimals, percentages	50
1	1.2	Statistical Analysis Process (SAS)	50	1.2	Statistical Analysis Process (SAS)	50
1	1.3	Research Questions and variables	50	1.3	Research questions; observations vs experiment; variables	50
1	1.4	Random sampling, Bias, Data Collection, and Conclusions	50	1.4	Proportions; linear expressions; methodology	50
1	1.5 (Optional)	Sampling, bias, random assignment	50	Open	Open	
1	1.6	Data Visualization and distributions	50	1.6	Extra practice creating dot plots and histograms	50
1	1.7	Measures of center, shape, spread	50	1.7	Mean, median, mode explorations with even/odd sample sizes	50
1	1.8	Measures of center, shape, spread	50	1.8	Boxplots & Outliers	50
1	1.9	Measures of center, shape, spread	50	1.9	Standard Deviation; Review; Productive Persistence: Stress Reappraisal	50
2	2.1	Probability basics (large numbers, relative frequency)	50	2.1	Probability: Law of large numbers; one-way tables	50
2	2.2	Probability basics (types of probabilities and rules)	75	2.2A	Conditional probability; two-way tables	50
				2.2B	Conditional probability; two-way tables	50
				2.2C (Optional)	Probability: Simulation	50
2	2.3	Probabilty Distributions: Discrete	50	2.3	Discrete random variables and distributions	50
2	2.4 (Optional)	Binomial Distributions	75	Open	Open	
2	2.5	Probabilty Distributions: Continuous	50	2.5	More probability distributions	50
2	2.6	Normal Distributions	50	2.6	Standard Deviation; Z-scores; Empirical Rule	50
2	2.7	Standard Normal Distribution	50	2.7	Normal Distribution Applications	50
3	3.1	Sampling distirbutions for sample proportions	75	3.1	Parameters and Statistics	60
3	3.2	Confidence intervals for population proportion	60	3.2	Constructing Confidence Intervals	50
3	3.3	Hypothesis test for population proportion	50	3.3	Testing Claims about Population Parameters	60
3	3.4	Hypothesis test for population proportion	60	3.4	Hypothesis Test for a population proportion	60

Module	Statway College Lesson	SW College Topic	Estimated Time	Statway Corequisite Lesson	SW Corequisite Topic	Estimated Time
3	3.5	Hypothesis test for population proportion	50	4.1A (pre)	Normal Distributions of Quantitative Data	50
4	4.1	Sampling distribution for sample means	50	4.1B (post)	Central Limit Theorem for Sample Means	50
4	4.2	Confidence intervals for population mean	50	4.2	T-Distributions & Interval Estimates for Means	50
4	4.3	Hypothesis test for population	50	4.3	Inference on Quantitative Data	50
4	4.4	Hypothesis test for population	50	Open	Open	
5	5.1	Two sample Tests: difference in two sample proportions: distribution (and CI in homework)	50	5.1	Differences Between Sample Proportions	50
5	5.2	Two sample Tests: difference in two sample proportions: hypothesis test	50	5.2	Testing for Differences Between Population Proportions	50
5	5.3 (Optional)	Two sample Tests: paired samples: distribution (and confidence intervals in homework)	50	Open	Open	
5	5.4	Two sample Tests: paired samples: hypothesis test	25	5.4	Hypothesis Tests for Paired Samples	50
5	5.5 (Optional)	Two sample Tests: independent samples: distribution and confidence intervals	40	Open	Open	
5	5.6	Two sample Tests: independent samples: hypothesis test	50	5.6	Testing for Differences Between Population Means	50
5	5.7	Identifying which test is appropriate	50	5.7 (optional)	Summary of Inference Methods	50
5	5.8 (Optional)	ANOVA: Equal sample size	50	Open	Open	
5	5.9 (Optional)	ANOVA: Unequal sample size	50	Open	Open	
6	6.1	Scatterplots	50	6.1	Proportionality	50
6	6.2	Correlation	50	6.2	Linear Models	50
6	6.3	Least Squares Regression	50	6.3	Statistical Models/Math Functions	50
6	6.4 (optional)	Residuals and LSR	50	6.4 (optional)	Special Properties of Least-Squares Regression Line	50
6	6.5 (optional)	Coefficient of Determination	50	6.5 (optional)	Inequalities	50
6	6.6	Correlation Inference	50	Open	Open	
6	6.7	Suitability of Linear Regression Model	50	6.7	Power Models	50
6	6.8 (Optional)	Exponential models	50	6.8 (Optional)	Exponential Models	100
7	7.1	Chi-Square: Introduction	50	7.1	Chi-Square Distribution	50

Module	Statway College Lesson	SW College Topic	Estimated Time	Statway Corequisite Lesson	SW Corequisite Topic	Estimated Time
7	7.2	Chi-Square: degrees of freedom, distribution, p-value: one-way (goodness-of-fit)	50	7.2	Goodness of Fit	50
7	7.3	Chi-Square: independence	50	7.3	Testing for Independence with Two-Way Tables	50
7	7.4	Chi-Square: homogeneity	50	7.4	Chi-Square Test for Homogeneity	50

Materials

Statway College. For the college-level component, Statway College is used as the student workbook (for in-class lessons). Students can also access the online homework platform that includes supplemental materials, exercises, and quizzes that support the learning outcomes in each module.

Statway Corequisite. There is a pool of Statway Corequisite lessons in the Official Curriculum folder from which faculty can choose, depending on their desired learning objectives and student needs. The Statway Corequisite lessons address both statistical and algebraic concepts and thinking needed to support success in Statway College. The Statway Corequisite lessons have a parallel structure to the Statway College lessons. These lessons are to be printed on demand. No formal workbook is published, since Corequisite implementation and needs will vary so widely due do varied implementation strategies. There are no corresponding online materials for Corequisite lessons.

Implementation

Statway College is designed as a 3 transfer-level credit course. Most elect to offer is as a 3-contact hour course, but some elect to offer it as a 4-contact hour course. **Statway Corequisite** is designed to support a 1-, 2-, or 3-contact hour corequisite course depending on student and program needs. The lessons may be printed on demand and taught at any time or in any order throughout the course. The Corequisite lessons were designed to be taught **after** the corresponding Statway College lesson, but can be used in any way an instructor sees fit.

*There are some “open” lessons in Statway Corequisite, where there is no corresponding lesson to the Statway College lesson. These breaks can be used as review sessions, for catching up on longer Statway College lessons, assessments, or any other needs faculty and students have.

**All lessons are designed to be 50 minutes. There are some that are longer or shorter (see lesson list above). Time is always tricky and faculty will need to adjust for longer and shorter lessons, and for varied implementation (e.g. 75 minute class sessions). Faculty mentors can help with this process if needed.

***Varied Corequisite contact-hours (e.g. 1- or 2-contact hour offerings) will also require lesson adjustment. Not all Corequisite lessons can be taught in these administration models. However, the Corequisite materials are designed so every lesson does not need to be taught. Faculty can select among the lessons they feel will be the most beneficial for their students and the Learning Outcomes required in their course.

Example Implementation Strategy

Below is one recommended strategy based on how the lessons were designed. It assumes a 3-hour Statway College course paired with a 3-hour Statway Corequisite course, where the in-class sessions alternate between Statway College and Statway Corequisite. But you are encouraged to tailor the structure of both portions as you see fit for your students.

*Note that Statway College 2.2 has multiple Corequisite lessons that pair with it. Faculty can choose among that Corequisite material, or take advantage of the Open space after 2.4 in any way they see fit.

**Note: The table below shows one recommended order of lessons only. Feel encouraged to adjust and to insert assessments, review days, etc., as needed.

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

Day 1	1.1 College
Day 2	1.1 Corequisite
Day 3	1.2 College
Day 4	1.2 Corequisite
Day 5	1.3 College
Day 6	1.3 Corequisite
Day 7	1.4 College
Day 8	1.4 Corequisite
Day 9	1.5 (Optional) College
Day 10	Open
Day 11	1.6 College
Day 12	1.6 Corequisite
Day 13	1.7 College
Day 14	1.7 Corequisite
Day 15	1.8 College
Day 16	1.8 Corequisite
Day 17	1.9 College
Day 18	1.9 Corequisite
Day 19	2.1 College
Day 20	2.1 Corequisite
Day 21	2.2 College
Day 22*	2.2A Corequisite
Day 22*	2.2B Corequisite
Day 22*	2.2C (Optional) Corequisite
Day 23	2.3 College
Day 24	2.3 Corequisite
Day 25	2.4 (Optional) College
Day 26	Open
Day 27	2.5 College
Day 28	2.5 Corequisite
Day 29	2.6 College
Day 30	2.6 Corequisite
Day 31	2.7 College
Day 32	2.7 Corequisite
Day 33	3.1 College
Day 34	3.1 Corequisite
Day 35	3.2 College
Day 36	3.2 Corequisite
Day 37	3.3 College
Day 38	3.3 Corequisite
Day 39	3.4 College
Day 40	3.4 Corequisite
Day 41	3.5 College
Day 42	4.1A (pre) Corequisite
Day 43	4.1 College
Day 44	4.1B (post) Corequisite
Day 45	4.2 College
Day 46	4.2 Corequisite
Day 47	4.3 College
Day 48	4.3 Corequisite
Day 49	4.4 College
Day 50	Open
Day 51	5.1 College
Day 52	5.1 Corequisite
Day 53	5.2 College
Day 54	5.2 Corequisite
Day 55	5.3 (Optional) College
Day 56	Open
Day 57	5.4 College
Day 58	5.4 Corequisite
Day 59	5.5 (Optional) College
Day 60	Open
Day 61	5.6 College
Day 62	5.6 Corequisite
Day 63	5.7 College
Day 64	5.7 (optional) Corequisite
Day 65	5.8 (Optional) College
Day 66	Open
Day 67	5.9 (Optional) College
Day 68	Open
Day 69	6.1 College
Day 70	6.1 Corequisite
Day 71	6.2 College
Day 72	6.2 Corequisite
Day 73	6.3 College
Day 74	6.3 Corequisite
Day 75	6.4 (optional) College
Day 76	6.4 (Optional) Corequisite
Day 77	6.5 (optional) College
Day 78	6.5 (Optional) Corequisite
Day 79	6.6 College
Day 80	Open
Day 81	6.7 College
Day 82	6.7 Corequisite
Day 83	6.8 (Optional) College
Day 84	6.8 (Optional) Corequisite
Day 85	7.1 College
Day 86	7.1 Corequisite
Day 87	7.2 College
Day 88	7.2 Corequisite
Day 89	7.3 College
Day 90	7.3 Corequisite
Day 91	7.4 College
Day 92	7.4 Corequisite

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

CARNEGIE MATH PATHWAYS
WestEd

6.3

Least Squares Regression

INTRODUCTION

Statistical methods are used in *forensics* to identify human remains based on the measurements of bones. In the 1950s, Dr. Mildred Trotter and Dr. Goldine Gleser measured skeletons of people who died in the early 1900s. From these measurements they developed statistical formulas for predicting a person's height based on the lengths of various bones. These formulas were first used to identify the remains of U.S. soldiers who had been buried in unmarked graves in the Pacific zone during World War II. Modern forensic scientists have made adjustments to Trotter and Gleser's formulas to account for the differences in people living now.

Language Tip
***Forensics* is the use of science to investigate crimes.**

We will use a process similar to Trotter and Gleser's in this problem.[1] Let's see if we can identify a female student based on the length of her forearm. The mystery student has a forearm measurement of 10 inches. She is alive and healthy!

Our task is to determine if the mystery student could be one of three female students. To accomplish our task, we have gathered evidence in the table below. Our evidence lists the heights and ages for three female college students.

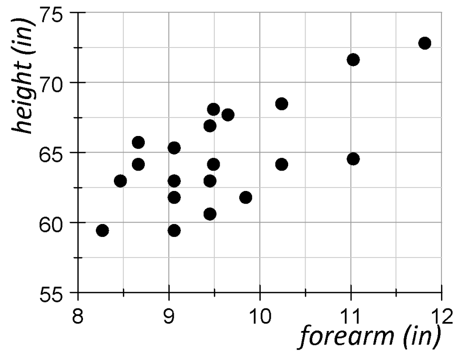
	Jane Doe 1	Jane Doe 2	Jane Doe 3
Age	18	23	33
Sex	Female	Female	Female
Height	5 feet, 5 inches	5 feet, 2 inches	6 feet

First, we need data to help us see if there is a relationship between forearm length and height for female students.

The scatterplot below is a graph of forearm length versus height for 20 female college students taking Introductory Statistics at Los Medanos College in Pittsburg, California, in 2009.

[1]For information on the Terry skeleton collection, see http://anthropology.si.edu/cm/terry.htm. For a more recent example of how forensic scientists are still building on the work of Trotter and Gleser, see R.L. Jantz, "Modification of the Trotter and Gleser Female Stature Estimation Formulae," *Journal of Forensic Science* 38, no. 4 (1993): 758–63.

Statway® Version 3.1, © 2018 WestEd. All rights reserved.



1 Based on the scatterplot, what is a reasonable prediction for the height of the mystery student? Briefly describe or show how you made your prediction.

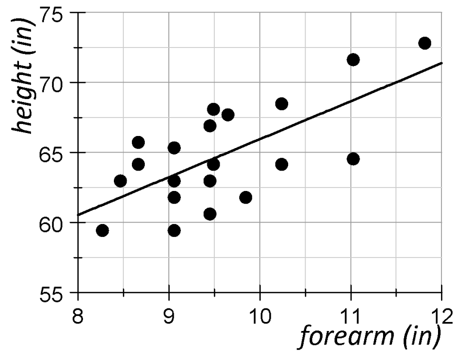
2 There is a lot of variability in the data. This variability can make it difficult to determine if one of these students is the mystery student. One way to help solve the mystery is see if we can eliminate any of the three students. Could any of the three students be eliminated as the mystery student? Explain your reasoning.

NEXT STEPS

Using a Line to Make Predictions

3 The scatterplot shows a positive linear association between forearm and height measurements. The correlation is 0.68, which is fairly strong. Since the correlation is fairly strong, it makes sense to use a linear model to summarize the relationship. In statistics, there is a line that will give us the best description of how height and forearm length are related. We call it the **line of best fit**. The *line of best fit* is the line that represents the center of a linear pattern in a scatterplot. We will learn more about how to find this line in future lessons. For now, we will give you the equation of this line.

Statway® Version 3.1, © 2018 WestEd. All rights reserved.



A The line of best fit is drawn on the scatterplot above. Use the line of best fit to predict the height of the mystery student. *Hint*: use the image above to look for the height where the line of best fit crosses the mystery student's forearm length.

B The equation of the line of best fit is:

predicted height = 2.7 · (*forearm length*) + 39

If we use *x* and *y* to represent these variables, the equation looks more like something from an algebra class.

$$\hat{y} = 2.7x + 39$$

When we use letters to represent variables in the line of best fit, we put a *hat* on the *y* and write $\hat{y}$ instead of *y*. The *hat* is a symbol that tells us that we are making an *estimate*. Results from a line of best fit equation do not represent actual data values.

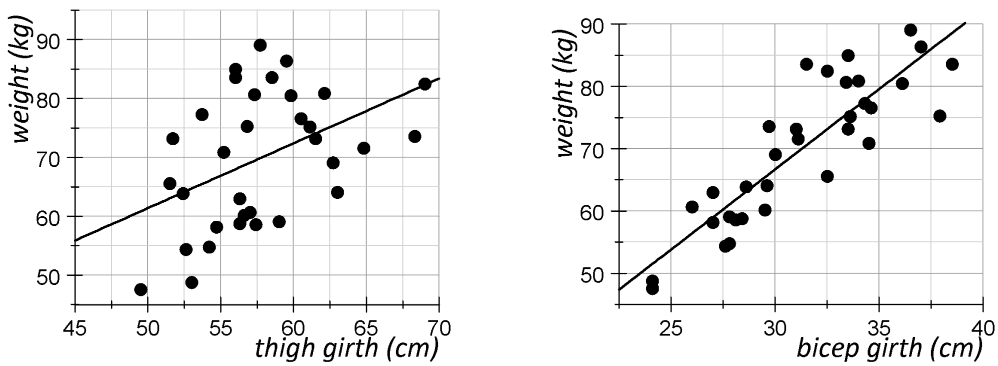
Use the equation to predict the height of the mystery person.

C Is the height of Jane Doe 1, Jane Doe 2, or Jane Doe 3 closest to the predicted height of the mystery student given by the line of best fit?

D Can you be certain that the student you picked in Question 3C is the mystery student? Explain your answer.

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

4 The scatterplots below represent body measurements in centimeters and weight in kilograms for 34 adults who are physically active.[2]



A Based on these data, which do you think is a better predictor of an adult's weight: thigh girth or bicep girth? Why do you think this?

B Adriana has a thigh girth of 57 centimeters and a bicep girth of 25 centimeters. Predict her weight using the graph that you think will give the most accurate prediction. Then plot Adriana's data point on the scatterplot that you used to make her weight prediction.

C The equations of the two lines shown are:

$$\textit{predicted weight} = 6.3 + 1.1 \cdot (\textit{thigh girth})$$

$$\textit{predicted weight} = -10.5 + 2.6 \cdot (\textit{bicep girth})$$

Predict Adriana's weight using the equation that you think predicts weight best.

[2]Grete Heinz et al., "Exploring Relationships in Body Dimensions," *Journal of Statistics Education* 11, no. 2 (2003), accessed 10 July 2014, www.amstat.org/publications/jse/v11n2/datasets.heinz.html.

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

D Of course, we do not really know Adriana's weight. How accurate do you think the prediction of Adriana's weight is? Choose the option that is the most reasonable and explain your thinking.

- Very accurate (within a range of plus or minus 1 kilogram).
- Somewhat accurate (within a range of plus or minus 5 kilograms).
- Not very accurate (within a range of plus or minus 10 kilograms).

5 In previous lessons, we learned about the concept of *correlation*. We learned that the correlation coefficient, *r*, describes the strength and direction of the linear relationship between two quantitative variables. Now we are predicting the value of one variable based on the other.

A Think about the scatterplots in Question 4. Which has a correlation coefficient closer to 1? Explain your answer.

B It is important for our predictions to be accurate. How does the correlation coefficient relate to the accuracy of the prediction?

In statistics, **extrapolation** is the process of using a statistical model, like a line, to make predictions that are outside the range of the available data. The problem with extrapolation is that it yields predictions that have no basis in evidence. Because of this, extrapolation is unreliable for predicting values of a response variable.

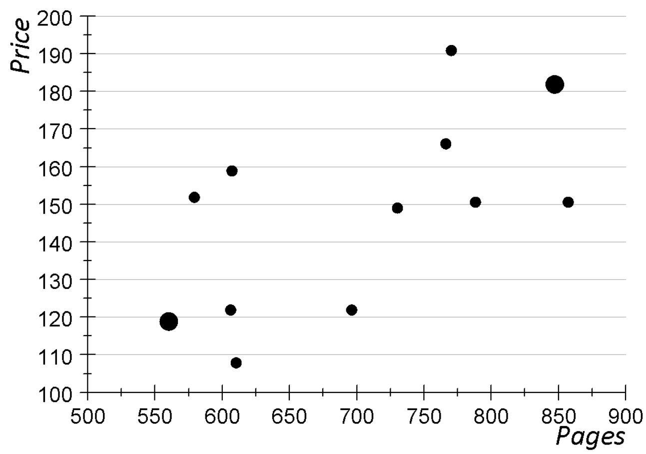
6 Use the equation of the line of best fit to predict the weight of a person with a bicep girth of 0 cm.

7 Do you think that this is a reliable prediction of a person's weight? Explain.

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

8 The predictions for someone who has a bicep girth of 41 cm and a bicep girth of 0 cm, are both examples of extrapolation, but one is much more likely to be accurate. Which extrapolation is most likely to be inaccurate, and why?

9 The textbook scatterplot is graphed again below, but this time we have chosen two points (560, 119.00) and (847, 182.00) that seem to lie in the center of the linear relationship. We have highlighted these points with *large* dots.



Use a straight-edge to draw a line through the (larger) chosen points. Use this line to predict the price of a textbook with 650 pages.

predicted price = _______________

Add the point representing the predicted price of a textbook with 650 pages to the line you drew above.

10 The equation of the line in Question 9 is:

$$\textit{predicted price} = 0.22 \cdot \textit{pages} - 3.92$$

Use this equation to predict the price of the 650 page textbook.

predicted price = _______________

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

YOU NEED TO KNOW

Notice that each least-squares equation above has the form:

$$\textit{predicted } y = a + b \cdot x$$

The numbers a and b in the least-squares equation have the following properties:

- a is called the ***y*-intercept** of the line. The *y*-intercept of a line is the point at which the line crosses the *y*-axis. a is also called the **initial value** because it is the predicted value of the response variable when the explanatory variable is 0.
- b is the **slope** of the line. b describes the change in the response variable when the explanatory variable increases by one unit.

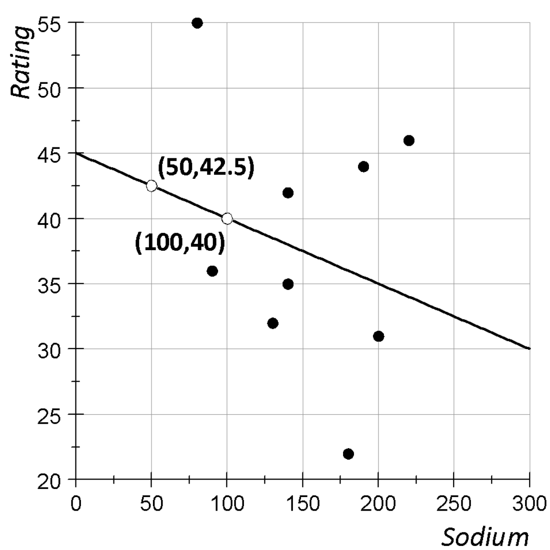
You can interpret a least-squares regression line as:

$$\hat{y} = a + b \cdot x$$

$$\hat{y} = \textit{(initial value)} + \textit{(slope)} \cdot x$$

NEXT STEPS

11 The graph below shows the sodium (salt) content (in milligrams) per serving and the ratings for nine cereals. It also shows the least-squares regression line. The black dots on the graph represent the cereals. The hollow dots (dots that aren't filled in with black) represent points on the regression line. The hollow dots *do not* represent actual cereals. Use the graph to answer the following questions.



Statway® Version 3.1, © 2018 WestEd. All rights reserved.

A What is the *y*-intercept of the least-squares regression line?

B If a cereal has 0 milligrams of sodium per serving, what would be the predicted rating? Use the LSR line to help make your prediction.

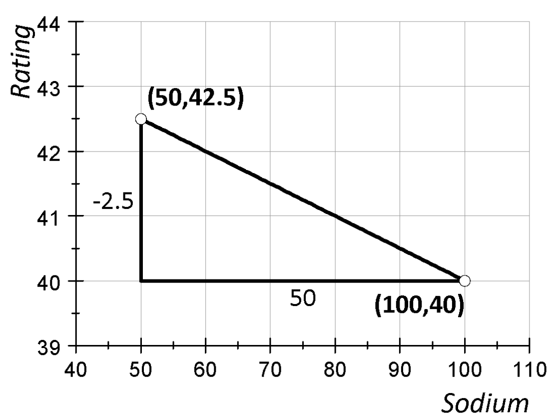
C How does the predicted rating change when the amount of sodium per serving increases from 50 milligrams to 100 milligrams?

The slope of the LSR line tells us how the predicted rating changes when there is a one milligram increase in sodium per serving. The triangle drawn on the graph below is called a *slope* triangle. The formula for slope is sometimes given as “rise over run.” A slope triangle shows the *rise* and *run* in the slope of the regression line.

Language Tip
The *slope* of a line measures the direction and steepness of the line.

The rise is the change in *y*, going up or down. The run is the change in *x*, going left or right.

The graph shows part of the regression line with two points on the line identified. The slope of this line can be calculated by dividing the difference between the *y*-coordinates by the difference between the *x*-coordinates.



$$\text{Slope} = \frac{\text{rise}}{\text{run}} = \frac{\text{change in y}}{\text{change in x}} = \frac{40 - 42.5}{100 - 50} = \frac{-2.5}{50} = -0.05\ \frac{\text{rating points}}{\text{mg}}$$

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

12 The slope tells us what happens to the predicted rating when the amount of sodium increases by 1 milligram.

A If the amount of sodium in the cereal increases by 1 milligram per serving, how does its predicted rating change?

B Knowing the slope allows us to calculate the change in the predicted rating for *any* increase in the amount of sodium per serving. If a cereal's sodium amount increases by 2 milligrams per serving, how does its predicted rating change?

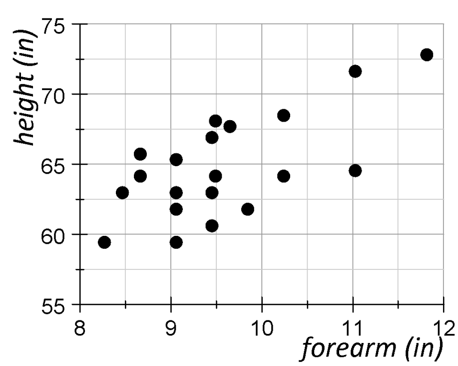
C If a cereal's sodium amount increases by 100 milligrams per serving, how does its predicted rating change?

TRY THESE

Let's return to another data set we studied recently. In this data set, we were given the forearm and height measurements for 20 female college students. These students were taking Introductory Statistics, in 2009, at Los Medanos College in Pittsburg, California. The equation of the least-squares regression line is:

predicted height = 39 + 2.7 · *forearm length*

$$\hat{y} = 39 + 2.7x$$



Statway® Version 3.1, © 2018 WestEd. All rights reserved.

13 What do the numbers 2.7 and 39 in the equation tell you?

14 Suppose two female college students' forearm lengths differ by 4 inches. By how much would their predicted heights differ?

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

Copyright © 2018 WestEd. All rights reserved.

No part of this document may be used or reproduced in any manner without written permission from WestEd. Requests for permission to reproduce any part of this report should be directed to WestEd Publications Center, 730 Harrison Street, San Francisco, CA 94107-1242, 888.293.7833, fax 415.512.2024, permissions@WestEd.org, or http://www.WestEd.org/permissions.

The original version of this work, version 1.0, was created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching. All subsequent versions, including this version 3.1, resulted from the continuous improvement efforts of the Carnegie Networked Improvement Community.

CARNEGIE MATH PATHWAYS
WestEd

Corequisite 6.3

Statistical Models, Mathematical Functions, and Linear Functions

INTRODUCTION

Sarah is planning to start Austin Community College next semester. She plans to take 13 credit hours. Sarah needs to plan her budget. She wants to know how much next semester will cost.

Students from Austin Community College were randomly selected to answer three questions:

(1) How many credit hours are you taking this semester?

(2) How much did your tuition cost this semester?

(3) How much did your books cost this semester?

The data collected from the students is presented in the table below:

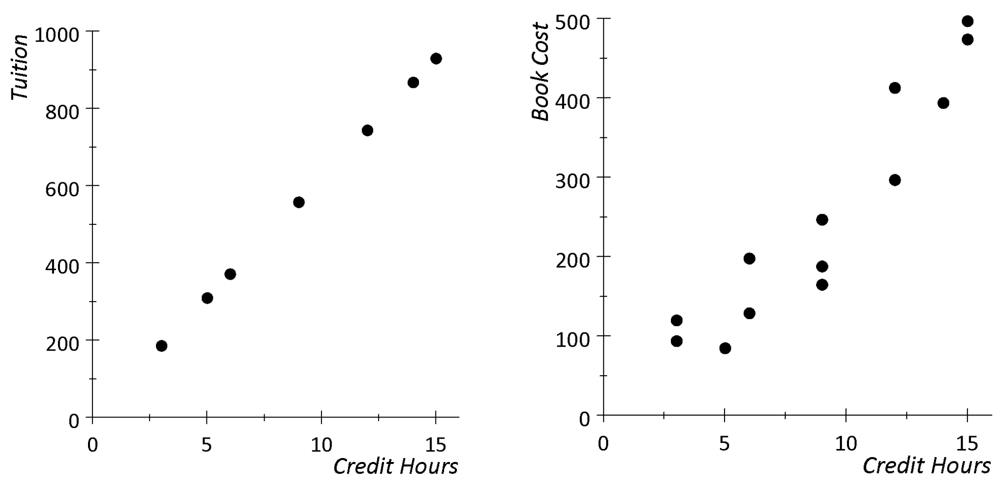
Table A: Credit Hours and Costs

Credit Hours	Tuition ($)	Book Cost ($)
3	186	120
3	186	94
5	310	85
6	372	129
6	372	198
9	558	247
9	558	188
9	558	165
12	744	297
12	744	413
14	868	394
15	930	497
15	930	474

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

1 Sarah needs to plan for the cost of spring semester. How could you use the data to help Sarah plan for the cost?

2 Data from Table A is presented in the scatterplots below. Use the scatterplots to make the following predictions:



A Predict the cost of tuition for a student taking 13 credit hours.

B Predict the cost of books for a student taking 13 credit hours.

3 Look at your predictions for questions 2A and 2B. Do you think one of your predictions is likely to be closer to Sarah's actual cost? If so which one? Explain your reasoning.

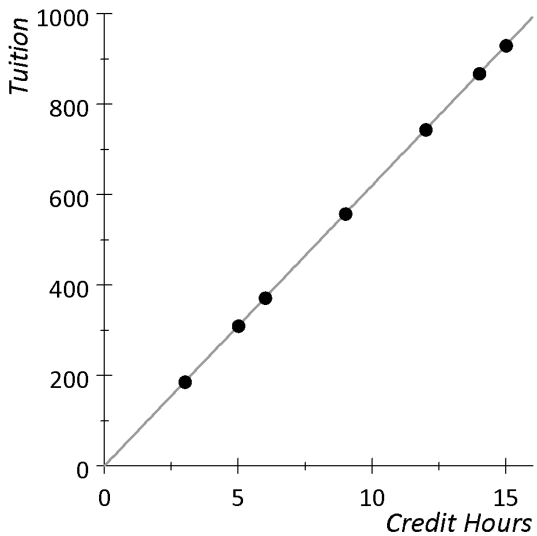
4 Based on your predictions, about how much money should Sarah plan to spend next semester?

NEXT STEPS

The scatterplots below include their *lines of best fit.*

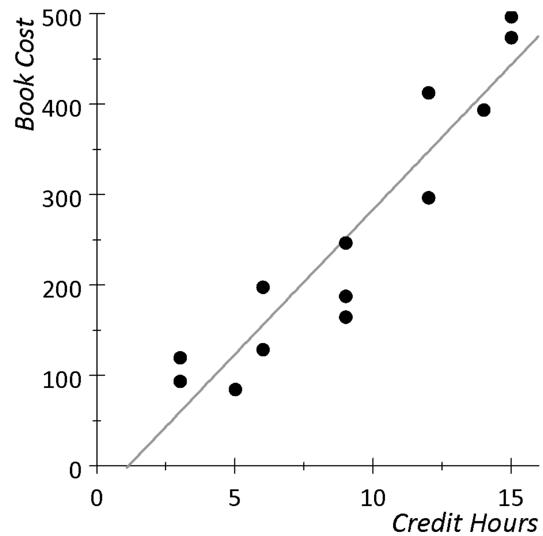
LSR line: $\hat{y} = 62x$

Correlation coefficient: r = 1



LSR line: $\hat{y} = -36 + 32x$

Correlation coefficient: r = 0.94



5 Examine the least squares regression (LSR) models for tuition cost and book cost.

A What are the similarities?

B What are the differences?

6 Use the equations for the LSR lines to make the following predictions:

A Predict the cost of tuition for a student taking 13 credit hours.

B Predict the cost for books for a student taking 13 credit hours.

7 When making predictions, which of these LSR models would you have more confidence in? Explain why you chose that LSR model.

8 How does your answer to Question 7 compare to your answer to Question 3?

YOU NEED TO KNOW

When examining relationships between two variables, we have so far used only **statistical models**. **Mathematical functions** are another useful tool. Here are the definitions of each:

- **Statistical Model**:
  A scatterplot of the data will show a pattern. It's possible that the pattern shows more than one y-value corresponding to a given value of x.

  We can use a LSR model to predict y (the response variable), but the prediction is only an *estimate*. The prediction is likely to be different from any observations with the same x-value (the explanatory variable).

- **Mathematical Function**:
  A scatterplot of the data will show a pattern. Only one y-value can correspond to a given value of x.

  We can use the function to predict an *exact* value for y (the response variable). Predictions are always equal to observations with the same x-value (the explanatory variable).

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

9 Use the scatterplots for tuition cost and textbook cost (before Question 5, above) to answer the following questions:

A Is tuition cost per credit hour represented by a *statistical model* or a *mathematical function*? Explain your reasoning.

B Is textbook cost per credit hour represented by a *statistical model* or a *mathematical function*? Explain your reasoning.

10 For both questions A and B below, read the descriptions of the relationship and then answer the questions: Will the relationships be described by a *statistical model* or a *mathematical function*? Explain.

A The cost to fill a car with regular unleaded gas at the BP station on the corner of Watt Ave. and Fair Oaks Blvd. on May 31, 2012. At this BP station, the price per gallon did not change that day.

B The cost to fill a car with unleaded gas in the United States on May 31, 2012.

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

TRY THESE

Jean lives in a housing co-operative, where she shares food and housing expenses with several roommates. Jean needs to buy meat for her housing co-operative. She has two options:

- She can go to a Fresh-Plus store and pay $4.50 per pound.
- Or, she can go to a warehouse store, like Costco or Sam's Club, and pay $3 per pound.

Fresh-Plus is near her housing co-operative. A trip to Fresh-Plus will cost $0.50 for gas. The warehouse store is further away. A trip to the warehouse store will cost $5 for gas.

Our goal is to write a mathematical *function* (or *formula*). This mathematical function is for the amount of money that Jean would spend at each of the stores.

To do this, let's think about the patterns in *total cost* for each of the two stores.

To help us observe those patterns, we can make a table of costs. In this table, define x and y as the following variables:

- x = *number of pounds of meat purchased*
- y = *cost* of purchasing the meat at Fresh-Plus.

The table below allows x to vary from 1 to 5 pounds.

11 Complete the remaining two calculations in the table. Fill in the value in the Fresh-Plus Cost column.

x	y (computation)	Fresh-Plus Cost
$x = 1$	$y = 0.5 + 4.50$	$5.00
$x = 2$	$y = 0.5 + 2 \cdot 4.50$	$9.50
$x = 3$	$y = 0.5 + 3 \cdot 4.50$	$14.00
$x = 4$		
$x = 5$		

12 Look at the pattern in the formula. How does the cost change each time Jean buys one more pound?

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

13 Imagine that x pounds of meat are purchased from Fresh-Plus.

If we don't know the value of x, we can still represent the cost (y) with a formula—a mathematical function. Look at the y computations in the table. Use these computations to determine the formula for the cost of x pounds of meat.

y = __________

14 Let x continue to be *pounds of meat purchased*, but now let's define y = *cost* for the warehouse store.

Make a table like the one above. Write an equation for y = *cost* of buying x pounds of meat from the warehouse store.

x	**y (computation)**	**Warehouse Cost**
$x = 1$		
$x = 2$		
$x = 3$		
$x = 4$		
$x = 5$		

Formula: y = __________

15 Write the equations for the Fresh-Plus and warehouse store costs. Then, graph them together below.

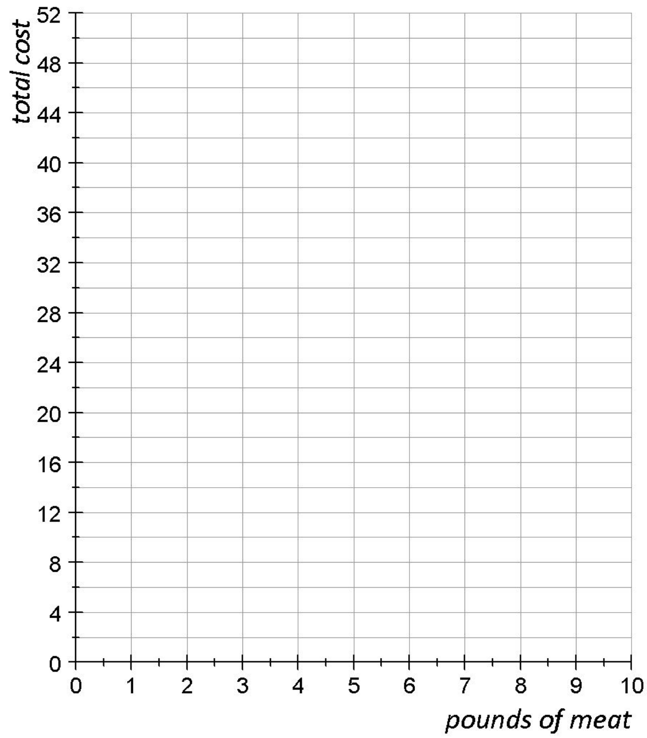
Fresh-Plus Cost

y = ________________

Warehouse Cost

y = ________________

Statway® Version 3.1, © 2018 WestEd. All rights reserved.



16 Do the graphs cross? If not, extend the graphs until they do.

17 What is important about the point where the graphs cross? Describe the costs before the graphs cross and after the lines cross.

18 Under what circumstances would it would make financial sense to make the trip to the warehouse store to buy meat?

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

Finding a Linear Function from Two Points

When the graph of a mathematical function is a line, we call it a **linear function**.

People put money in different types of investments, like savings accounts, government bonds, and the stock market. These types of investments have different types of interest, depending on the investment terms. One type is called simple interest. Simple interest pays you interest on your principal alone. (*Note*: you do not need to know all the details about interest right now, just remember that with simple interest, the rate will not change.)

Suppose you invest a certain amount of money in an account that earns simple interest. At year 2, you have $1200. At year 6, you have $1400. We will find the linear function for account balance after x years. We can find a linear function in four steps.

Steps for finding a linear function:

(1) Decide how to label the explanatory variable and the response variable, and choose units for them.

(2) Find the *slope*. The slope is the rate of change of the response variable per unit change in the explanatory variable. This means how much y changes as x changes by 1 unit. We use the letter m for slope.

(3) Find the *y-intercept* of the response variable. This is the value when the explanatory variable is zero. The y-intercept is the also called the *initial value*. We use the letter b for the y-intercept.

(4) Write the formula for the line, in the form $y = mx + b$, using the values you found for m and b in steps ii and iii above. This formula is our mathematical model.

Below, we apply these steps to our simple interest problem.

(1) We want a function for computing *account balance* at any *time*, so let's define y = account balance, and x = time. In context, it makes sense to measure y in dollars and x in years.

(2) We can find the slope by computing the ratio of changes in each variable.

$$m = \frac{change\ in\ y}{change\ in\ x}$$

$$= \frac{(1400-1200)\ dollars}{(6-2)\ years}$$

$$= \frac{200\ dollars}{4\ years}$$

$$= 50\ dollars/year$$

$$= \$50\ per\ year$$

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

(3) Find the initial value, also known as the *y*-intercept.

We know that m is the slope and b is the y-intercept. Because this is a function, any point on the graph must satisfy this equation. This means we can substitute the coordinates of any point on the line for x and y in the equation and the equation will be true. We can solve for the y-intercept, b, if we make such a substitution.

Two pieces of information are given in the problem. We are told that when $x = 2$ years, the account balance is $y = 1200$.

$$y = mx + b$$

$$1200 = 50 \cdot 2 + b$$

$$1200 = 100 + b$$

Notice the 100 + b on the right side. If we subtract 100 from each side, the 100s on the right will zero out.

We will have solved for b.

$$1200 \mathbf{- 100} = 100 + b \mathbf{- 100}$$

$$1100 = b$$

(4) We now have the slope, m, is 50 and the y-intercept, b, is 1100. Therefore, the formula for our linear function is $y = mx + b$.

$$y = 50x + 1100$$

19 What is the initial value of the investment?

20 The equation found above can give two pieces of information:

- data values of the explanatory variable which represent times and
- data values of the response variable which represent the corresponding account balances.

Is the equation a statistical model of these points, or are they part of a mathematical function?

21 Explain what the slope tells us about the value of the investment over time.

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

YOU NEED TO KNOW

Be careful to subtract values in the right order when computing slope. The slope formula helps with this.

If the points, (x_1, y_1) and (x_2, y_2) are on a line, the slope of the line is

$$m = \frac{\text{change in } y}{\text{change in } x} = \frac{y_2 - y_1}{x_2 - x_1}$$

Once we have seen the units, we don't need to include them at each step as we did with the dollars and years in the last problem.

The units of slope are always units of y per unit of x. For example, the units in question 11 were dollars per year.

TRY THESE

22 Gary borrowed money from his father. He is repaying his father the same amount each month until the loan is paid off.

After 3 months, Gary owes $224. After a total of 7 months, Gary owes $160. Because the loan is being paid at a constant rate, the amount owed (y) is a linear function that depends on the number of months (x).

A Find the *slope* of the linear function. The linear function here gives the amount owed, y, after x months.

B The equation of a linear function is $y = mx + b$. You know the slope, m, for this example. Use the slope and one of the points to find the value of b.

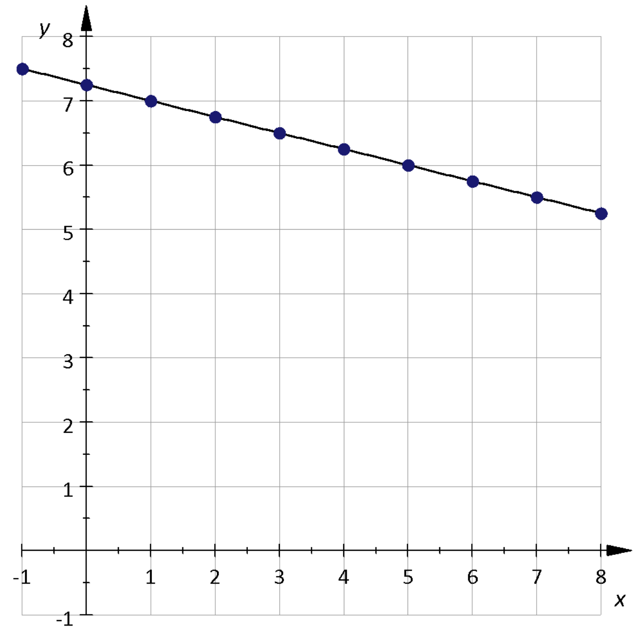
C Give the linear function which gives the amount owed after x months.

D What is the amount owed at 1 year? Use the correct units in your answer.

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

We can find the equation of a linear function by looking at its graph. We just need to identify two points. Try the following problems.

TRY THESE



23 Choose two points on the line. Then use the four steps for finding a linear function to find the equation of the line.

24 Use the equation to predict *y* when *x* = 8.5.

25 Use the graph to estimate *x* if *y* = 6. Then use the formula to check whether your estimate is correct.

Statway® Version 3.1, © 2018 WestEd. All rights reserved.

Copyright © 2018 WestEd. All rights reserved.

No part of this document may be used or reproduced in any manner without written permission from WestEd. Requests for permission to reproduce any part of this report should be directed to WestEd Publications Center, 730 Harrison Street, San Francisco, CA 94107-1242, 888.293.7833, fax 415.512.2024, permissions@WestEd.org, or http://www.WestEd.org/permissions.

The original version of this work, version 1.0, was created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching. All subsequent versions, including this version 3.1, resulted from the continuous improvement efforts of the Carnegie Networked Improvement Community.