Examples of Tasks from CCSS Edition Course 2, Unit 4

Getting Started

The tasks below are selected with the intent of presenting key ideas and skills. **Not every answer is complete**, so that teachers can still assign these questions and expect students to finish the tasks. If you are working with your student on homework, please use these solutions with the intention of increasing student understanding and independence. A list of questions to use as you work together, prepared in English and Spanish, is available. Encourage students to refer to their class notes and Math Toolkit entries for assistance. Comments in red type are not part of the solution.

As you read these selected homework tasks and solutions, you will notice that some very sophisticated communication skills are expected. Students develop these over time. This is the standard for which to strive. See Research on Communication.

The Statistics and Probability page might help you follow the conceptual development of the ideas you see in these examples.

Main Mathematical Goals for Unit 4

Upon completion of this unit, students should be able to:

- describe the shape of a cloud of points on a scatterplot and to describe the association between the two variables.
- interpret the coefficients of the regression equation, learn some properties of the regression line, and understand that a regression line is an appropriate way to summarize the bivariate relationship only if the points form an elliptical cloud.
- compute and interpret Pearson's correlation and to understand that a strong correlation does not imply that one variable causes the other.
- determine whether a point is influential on the correlation and on the equation of the least squares regression line.

What Solutions are Available?

Lesson 1: Investigation 1—Applications Task 3 (p. 271), Connections Task 7 (p. 275), Review Task 18 (p. 278)
Investigation 2—Applications Task 5 (p. 273), Connections Task 9 (p. 275), Review Task 19 (p. 278)

Lesson 2: Investigation 1—Applications Task 1 (p. 305), Connections Task 11 (p. 313), Reflections Task 18 (p. 316), Review Task 28 (p. 320)
Investigation 2—Applications Task 3 (p. 307), Review Task 30 (p. 321)
Investigation 3—Applications Task 6 (p. 310), Connections Task 15 (p. 315), Reflections Task 20 (p. 316), Review Task 32 (p. 321)
Investigation 4—Applications Task 7 (p. 310), Review Task 34 (p. 321)

Selected Homework Tasks and Expected Solutions

(These solutions are for tasks in the CCSS Edition book.
For homework tasks in books with earlier copyright dates, see Helping with Homework.)

Lesson 1, Investigation 1, Applications Task 3 (p. 271)

a. Two possible observations are:

There is a moderate positive association; that is, if one country had a larger population than another country in 2010, it tends to be projected to have a larger population in 2050.

Both Brazil and Nigeria show the largest changes in rank, moving down 3 and up 3, respectively.

b. *Hint:* To answer this question, look at the table on page 261 in Problem 3. That problem walks you through how to find the sum of the differences squared. Then, use that answer and Spearman's rank correlation, r_s, formula on page 261 to calculate the answer.

To be completed by the student.

c. You would expect the correlation between the 2010 ranking and the projected 2025 ranking to be larger than that between 2010 and 2050 because there is less time for population patterns to change. For example, the four largest countries maintained the same rank from 2010 to 2025 but are expected to change by 2050.

The rank correlation between the 2010 ranking and the 2025 projected ranking is 0.939.

Lesson 1, Investigation 1, Connections Task 7 (p. 275)

The distance formula was developed in Unit 3 of this course. Your student should have this formula in their toolkit.

a. $\sqrt{(3-(-5))^2 + (4-1)^2} = \sqrt{8^2 + 3^2} \approx 8.54$ units

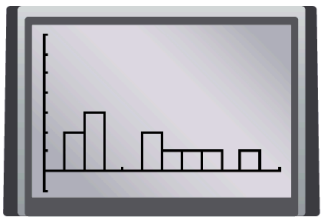
b. To be completed by the student.

c. Both involve a sum of squared differences. Both are meant to measure distance—one between two points in the plane and one between two rankings.

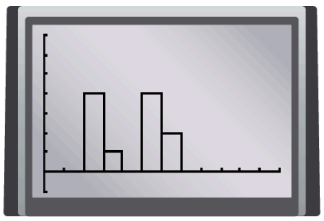
Lesson 1, Investigation 1, Review Task 18 (p. 278)

These data are located in *CPMP-Tools* under Statistics, Data Analysis, Data>Unit 4 Regression and Correlation> Hamburger Nutrition II.

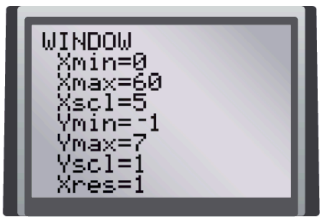
a. There are only 11 hamburgers, so students might see the most detail by making dot plots or histograms.

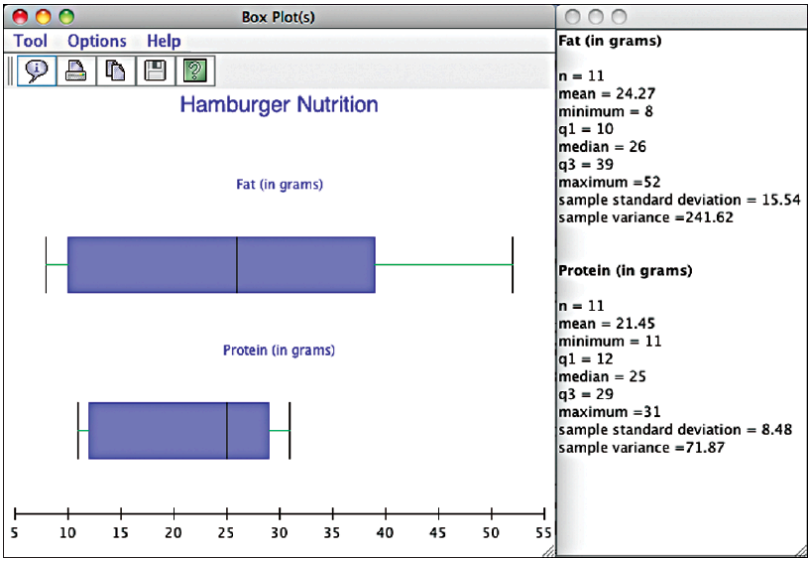


Protein (in grams)



Fat (in grams)





b–d. To be completed by the student.

Lesson 1, Investigation 1, Review Task 19 (p. 278)

a. Two possible dimensions for the cloth are 2 yards by 1 yard, or 0.25 yards by 8 yards.

b–c. To be completed by the student.

Lesson 1, Investigation 2, Applications Task 5 (p. 273)

Hint: As you do this task, remember that the city with the lowest crime rate, Minneapolis, has a rank of 1.

a. Health Care is the variable graphed on the x-axis in the third column. It is the variable graphed on the y-axis in the third row.

b, d, f. To be completed by the student.

c. Education and health care have the strongest positive correlation. Some possible reasons: Cities that fund education as a high priority may also fun health care more generously. People who are better educated have the money to pay for better health care. People who are better educated insist on better health care.

e. *Hint:* To find the missing rank correlation, r_s, use Spearman's formula found on page 261 in the student book.

To be completed by the student.

g. The entry in row i and column j of the correlation matrix gives the rank correlation between the two variables graphed in the scatterplot matrix in row i and column j.

Lesson 1, Investigation 2, Connections Task 9 (p. 275)

Students are looking for a pattern that will find the value for any number of variables k.

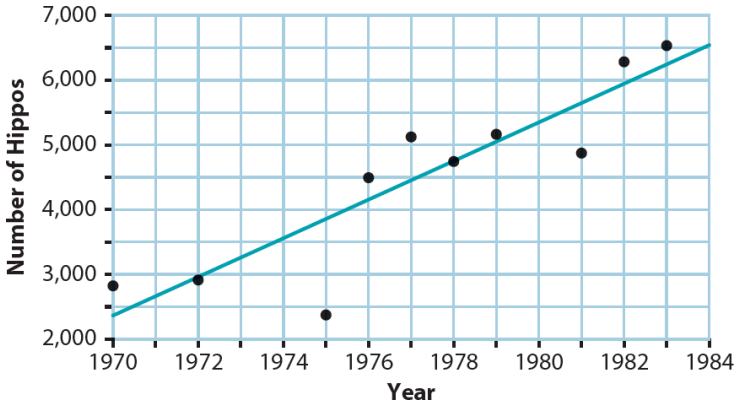
To find the number of scatterplot matrices for 3 variables, you need to calculate the area of a 3×3 square, which is 9, however, the main diagonal (the three boxes going from top-right to bottom-left) are not scatterplots because they are comparing a variable to itself. So to find the number of scatterplots, you would go $3^2 - 3$, or 6 scatterplots.

The variables of 4, 5, and k are left for the student to complete.

Lesson 2, Investigation 1, Applications Task 1 (p. 305)

The data for this task has been included in the *CPMP-Tools* software. Students can access the data and find the regression equation or they can put the data into their calculator.

a. The regression equation is _____.



b. To be completed by the student.

c. *Hint:* To find the residual, you need to take the actual value for 1975 and subtract the predicted value from that.

To find the residual for 1975, take a good estimate from the scatterplot of the actual point (2,300) minus the predicted (3,800) to get –1,500. Use the regression equation to evaluate y when $x = 1975$ to find the predicted number on hippos. Now look at the table on page 305 and find that the actual number of hippos in 1975 is 2,342. The residual is –1,495. This means that there were 1,495 fewer hippos in 1975 than were predicted by the regression line.

d. To be completed by the student.

Lesson 2, Investigation 1, Connections Task 11 (p. 313)

You can get a nice picture of the squared error by using *CPMP-Tools*. Put the data into the spreadsheet and click on this icon  .

a. Use the equation to find the predicted y values for each of the x values in the list. Then find the residuals and the squared residuals. Calculate the sums. The table below shows the computations.

x	y	Predicted $y = x + 0.75$	Residual	Squared Residual
1	3	1.75	1.25	1.5625
2	2	2.75	−0.75	0.5625
3	5	3.75	1.25	1.5625
6	5	6.75	−1.75	3.0625
		Total	0	6.75

b–c. To be completed by the student.

Lesson 2, Investigation 1, Reflections Task 18 (p. 316)

a. This question has to do with the true interpretation of slope. At first glance, students may believe that all are correct, but with closer analysis they can begin to eliminate two of the choices. The definition of slope is the rate of change between any two points; the change in the y values divided by the change in the x values. So, the correct interpretation of slope needs to compare the x and y values of the two points.

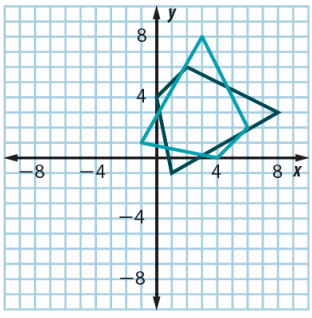
The first interpretation is not correct because it does not compare two students, but rather one student over time. The remainder of the solution is left to the student.

b. *Hint:* Look at the scatterplot and estimate the values of the residuals. Then estimate the error in prediction.

To be completed by the student.

Lesson 2, Investigation 1, Review Task 28 (p. 320)

a. $\begin{bmatrix} 4 & 6 & 3 & -1 \\ 0 & 2 & 8 & 1 \end{bmatrix}$



b–c. To be completed by the student.

Lesson 2, Investigation 2, Applications Task 3 (p. 307)

These data are located in *CPMP-Tools* under Statistics, Data Analysis, Data>Unit 4 Regression and Correlation>Canines.

a. Yes, as there is no curvature, however, the pattern is not very elliptical. Another worry is the influence of the outlier.

b. *Hint:* Remember when you are telling what the slope indicates, you need to represent it as a comparison of two pieces of data.

To be completed by the student.

c. The predicted lifespan of a Cape fox is 14.712 years. There are two reasons not to have much faith in this prediction.

To be completed by the student.

d. The error in prediction is quite large: *observed maximum longevity – predicted maximum longevity* = 14.712 – 7 = 7.712 years.

e–f. To be completed by the student.

Lesson 2, Investigation 2, Review Task 30 (p. 321)

a, c, e. To be completed by the student.

b. $5x(x - 5)$

d. $6x(3 - 2x)$

Lesson 2, Investigation 3, Applications Task 6 (p. 310)

These data are located in *CPMP-Tools* under Statistics, Data Analysis, Data>Unit 4 Regression and Correlation>Seal Sizes.

a. A good estimate would be between 0.7 and 0.9.

b. If it is allowed to use a tool to find the correlation coefficient, you can find this on *CPMP-Tools*. If you need to find it by hand, then use the formula on page 291. To help you check your calculations, the mean of $x = 8.76$, mean of $y = 795$, $s_x = 1.324$, $s_y = 375.169$, and $n = 5$.

The remainder is to be completed by the student.

c. From the plot, it is clear that a curve, not a line, might be a better model for the sizes of seals. This makes sense as, from the geometry of the situation, you would expect the relationship to be cubic because seals are somewhat similar in shape and linear measures are related by a cubic equation to measures depending on volume, such as weight.

d–e. To be completed by the student.

Lesson 2, Investigation 3, Connections Task 15 (p. 315)

Students have transformed measurements in a previous course and found that if you multiply each data point by the same value, then the new mean and standard deviation are also multiplied by that value. However, if the same value is added to each piece of data, then the new mean is the previous mean plus the added value, but the standard deviation is unchanged.

a. Multiply each mean height by 2.54.

b. The value of r does not change if values are all multiplied by the same constant.

c. The y-intercept is multiplied by 2.54, but the slope does not change.

d. To be completed by the student.

Lesson 2, Investigation 3, Reflections Task 20 (p. 316)

This task serves as a warning to always plot the data you are working with before you conclude that a line is or is not the best fit for the data. The correlation coefficient can be weak or strong, but until the graph is plotted it is not obvious whether a line will fit the data well or not. In this particular task, the plot shows the data are going to be best fit by some type of curve.

a. To be completed by the student.

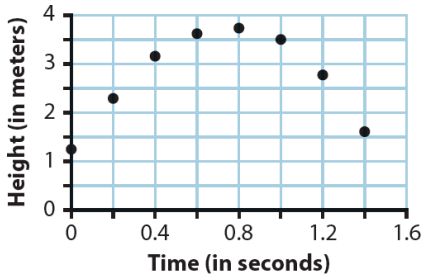
b. **i.** When the time is equal to zero, the height is 1.1 meters, so $h_0 = 1.1$.

ii. To be completed by the student.

iii. $h = 1.1 + 7.3t - 4.9t^2$

iv. To be completed by the student.

c. As time increases, the height increases, reaches a maximum, then decreases. The equation is quadratic and the pattern on the graph below is parabolic. The correlation is very weak because it only indicates the degree to which points cluster about a line; it gives no indication of the degree to which they follow another relationship. There is no reason to expect a high correlation using a formula that measures how closely the points cluster about a line rather than about a parabola.



Lesson 2, Investigation 3, Review Task 32 (p. 321)

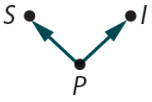
a. Equivalent

b. Equivalent

c–f. To be completed by the student.

Lesson 2, Investigation 4, Applications Task 7 (p. 310)

a. **i.** *Schooling* → *Income* or *S* → *I*

ii. 

b. **i.** The researchers chose twins for the study. The twins in the study probably grew up in the same household and so had identical economic status. By comparing the income of the twin with more education to that of the twin with less education, researchers can eliminate the effect of the parent's economic status. Lurking variables controlled include such things as quality of parenting, quality of local schools, number of siblings, and other variables, as well as economic status.

ii. To be completed by the student.

Lesson 2, Investigation 4, Review Task 34 (p. 321)

a. About 83.4 people per square mile

b–c. To be completed by the student.

