



PROJECT-SET

Statistics Education for Teachers

Text messaging is time consuming! What gives?



Overview of Lesson

This activity allows students to perform a hands-on investigation in which they explore relationships between two variables. Students use real data that they download from the Census at School project website. Each student works on a question of choice and uses their downloaded sample data accordingly. They then make a scatterplot to view the relationship of interest and estimate the regression line. After each student has interpreted their regression line, they pool their findings as a class to explore the variability in the slopes, intercepts, and correlations that they found. As a class, they construct approximations to the sampling distributions for each of these statistics and use the sampling distributions to make assertions about the values of the population parameters.

GAISE Components

This activity follows four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report*. The four components are: Formulate Question, Collect Data, Analyze Data, and Interpret Results. This is a GAISE level C lesson that incorporates the following process components:

- I. Formulate Question
 - Students pose their own questions of interest
- II. Collect Data
 - Students make design for differences
- III. Analyze Data
 - Quantification of association; fitting of models for association
- IV. Interpret Results
 - Interpret measures of strength of association
 - Interpret models of association

Common Core State Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
4. Model with mathematics
5. Use appropriate tools strategically.

Common Core State Standards Grade Level Content

- **High School**

- S-ID **Interpreting Categorical and Quantitative Data**

- Summarize, represent, and interpret data on two categorical and quantitative variables

- 6. Represent data on two quantitative variables on a scatterplot, and describe how the variables are related
 - a. Fit a function to the data; use functions fitted to data to solve problems in the context of the data.

- b. Fit a linear function for a scatter plot that suggests a linear association.
Interpret linear models
7. Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data
8. Compute (using technology) and interpret the correlation coefficient of a linear fit

NCTM Principles and Standards for School Mathematics

Data Analysis and Probability Standards for Grades 9-12

Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them

- understand the meaning of measurement data and categorical data, of univariate and bivariate data, and of the term variable
- understand histograms, parallel box plots, and scatterplots and use them to display data.

Select and use appropriate statistical methods to analyze data

- for bivariate measurement data, be able to display a scatterplot, describe its shape, and determine regression coefficients, regression equations, and correlation coefficients using technological tools
- identify trends in bivariate data and find functions that model the data or transform the data so that they can be modeled

Prerequisites

Prior to completing this activity, students should be able to identify the population and sample in any given situation involving random sampling. They should also understand that information from a sample is used to draw conclusions about the entire population. They should have a basic understanding of how to construct a dotplot and do an informal analysis of the shape, center, and spread. They should be able, with the aid of software or a calculator, to find the least-squares regression line.

Learning Targets

After completing the activity, students will have an understanding of how to carry out a regression and create approximations to sampling distributions for the slope, intercept, and the correlation of a regression. They will be familiar with the process of taking repeated samples of the same size, and constructing dotplots.

Time Required

The individual activity is designed to take 45 minutes. After that is completed, the class activity is designed for another 60 minute time frame.

Materials Required

For this activity, student should be in a computer lab. Each student should have access to a computer with Excel. Students will also need some type of software or graphing calculator capable of estimating regression equations.

Instructional Lesson Plan

The GAISE Statistical Problem-Solving Procedure

Many text messages are sent throughout each day. What factors could be related to the number of text messages one sends in a day? In this activity, we will explore the relationship between the number of text messages one sends in a day and a few other potential explanatory factors. Each student will work individually and then share their results with the class. Once each student has obtained their results, we will work together as a group to draw further conclusions.

1. Formulate Question(s)

Before beginning the activity, the teacher may wish to review the concepts of population, sample, population parameter and sample statistic, reinforcing student understanding of these foundational concepts for the lesson. Define variables \hat{y} , β , and α . Also, teachers should ensure that students have had practice estimating the least-squares regression line using either some type of software or a graphing calculator. If students have not had adequate practice constructing and analyzing dotplots, this process should also be reviewed. The questions of interest for the activity are:

1. Does the number of hours you spend hanging out with friends in a day increase or decrease the number of text messages you send?
2. Does the number of hours you spend doing homework in a day increase or decrease the number of text messages you send?
3. Does the number of text messages you receive in a day increase or decrease the number of text messages you send?

Students will be asked to choose the statistical question they would like to focus on.

Note for Teachers: For small classes (less than 30 students) give students only two questions to choose from. Teachers make sure that the class is evenly divided among the questions. Ideally, you would like to have at least 15/20 people or more working on the same question. If the class size is too small to achieve that many people working on the same question, then either reduce the number of questions or have each student choose two out of the three questions to explore.

2. Collect Data

In order to collect the data to be used during this activity, students will be instructed to go to the Census for School website and download a random sample. Students will then have to briefly clean up the data in Excel in order to make it easily workable for the lesson.

To answer the question you choose, you are going to download and work with a real data set. To download the data set, go to the following website:

<http://www.amstat.org/censusatschool/>

Each student will individually carry out the following steps:

1. Click on Random Sampler

2. Accept the Terms & Conditions

3. Select a sample size of 100 from All States and 9, 10, 11, and 12 grade levels. Include All Genders and All Years of data collection.

4. Download the data into Excel.

5. Open the data in Excel. You will see a large number of variables (labeled in each column).

6. Delete all the columns except for the following:

Gender, Text Messages Sent Yesterday, Text Messages Received Yesterday, Hanging out with Friends Hours, Doing Homework Hours

Here is a snapshot of a sample data set showing the first 50 subjects.

	Gender	Text_Messages_Sent_Yesterday	Text_Messages_Received_Yesterday	Hanging_Out_With_Friends_Hours	Doing_Homework_Hours
1	Female	10	10	6	4
2	Male	80	80	45	0.005
3	Male	30	30	40	2
4	Male	1	1	1	
5	Female				
6	Female	5	5	20	5
7	Female	18	23	1	18
8	Male	80	84	53	7
9	Female	4	5	40	6
10	Male	30	28	8	8
11	Female	60	60	25	5
12	Male	100	100	15	2
13	Male	0	0	12	8
14	Male	6	6	10	14
15	Female	20	25	10	6
16	Male	0	0	54	0
17	Female				
18	Male	0	4	15	5
19	Female	500	400	7	1
20	Male				
21	Male	55	156	30	10
22	Male	50	50	35	2
23	Female	15	15	8	25
24	Female	10	10	10	11
25	Male	5	5	45	2
26	Male	3	3	14	1
27	Male	25	27	1	1
28	Male	43	44	2	13
29	Male	15	15	1	6
30	Male	378	380	42	7
31	Female	50	55	8	35
32	Female	0	0	49	2
33	Male	248	230	5	15
34	Male	5	5	35	4
35	Female	500	600	10	2
36	Female	67	34		
37	Male	20	20	12	16
38	Female			7	0
39	Male				
40	Female	0	0	20	1
41	Male	8	12	18	15
42	Female	750	1000	60	1
43	Female	200	200	30	15
44	Female	60	68	30	10
45	Male	150	200	10	2
46	Male	6	11	6	8
47	Male	150	250	60	9
48	Female	30	30	10	4
49	Male	150	152	4	6
50	Male	0	0	2	35
51	Male				

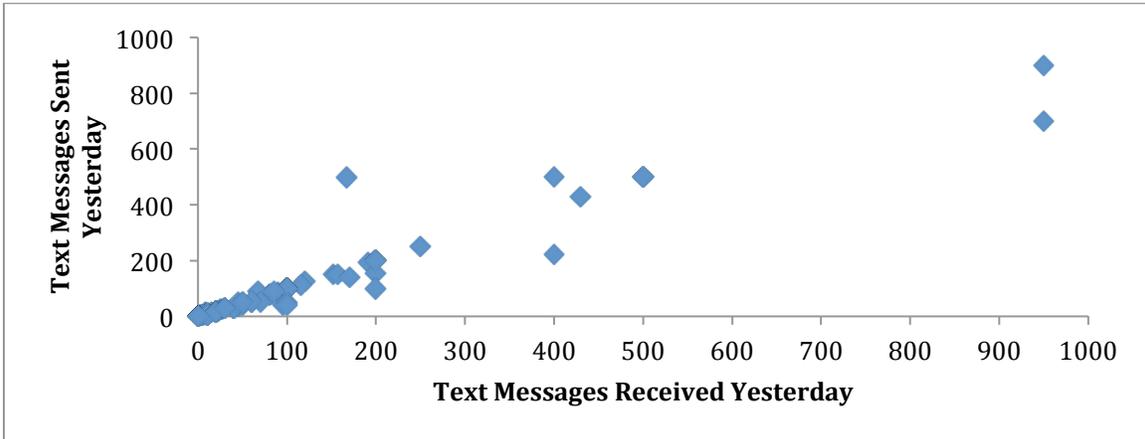
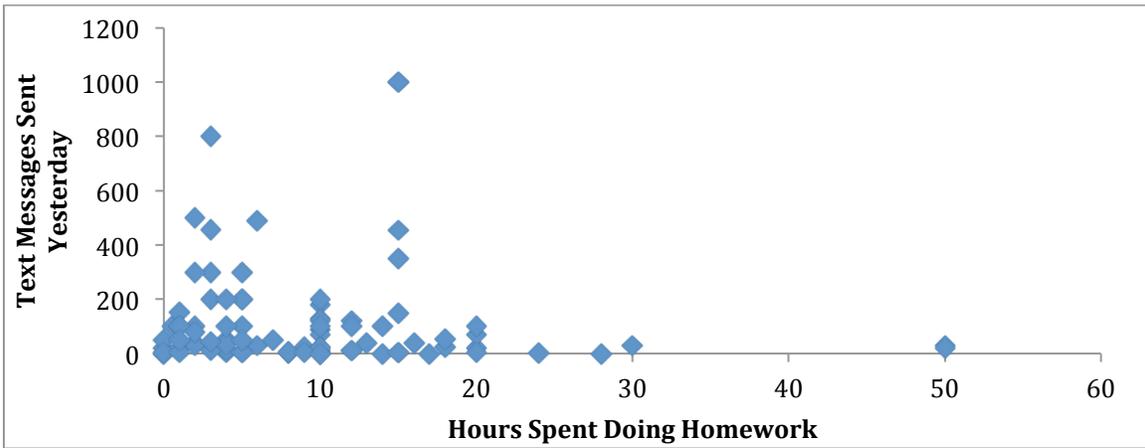
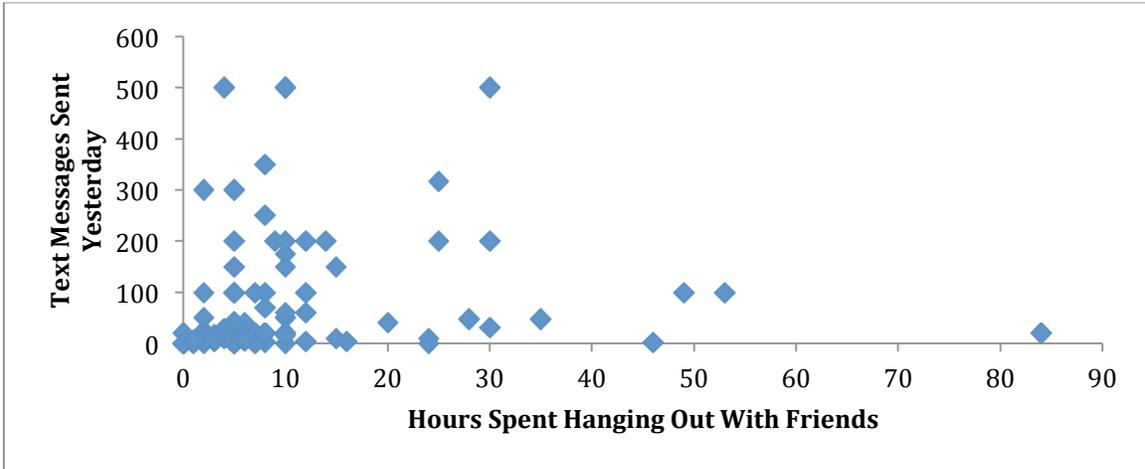
Is this a realistic value for the number of text messages received?

Note for Teachers: The class should discuss whether the data are realist or whether a case-wise deletion of certain observations might be necessary. In the case where data seem incorrect, the student or class may determine that seleting that observation from their data set is advised.

3. Analyze Data

Using each of their individual data, students will then be asked to analyze it on their own. Each student will plot a scatterplot, examine whether there seems to be a linear relationship, point out any outliers, compute the regression line, and compute the correlation coefficient.

Here are snapshots of sample scatterplots showing sample data for each of the three questions.



Using technology students will estimate the regression lines:

If the student chose question (1) above, then they will estimate the following model:

$$\text{Number of Text Messages Sent} = \alpha + \beta(\text{Hours Spent with Friends}) + \varepsilon$$

If the student chose question (2) above, then they will estimate the following model:

$$\text{Number of Text Messages Sent} = \alpha + \beta(\text{Hours Spent doing Homework}) + \varepsilon$$

If the student chose question (3) above, then they will estimate the following model:

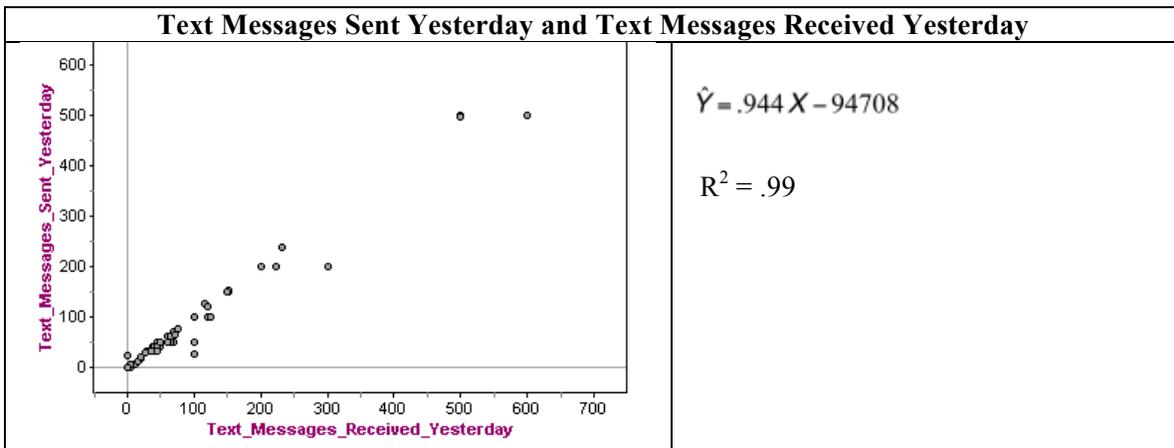
$$\text{Number of Text Messages Sent} = \alpha + \beta(\text{Number of Text Messages Received}) + \varepsilon$$

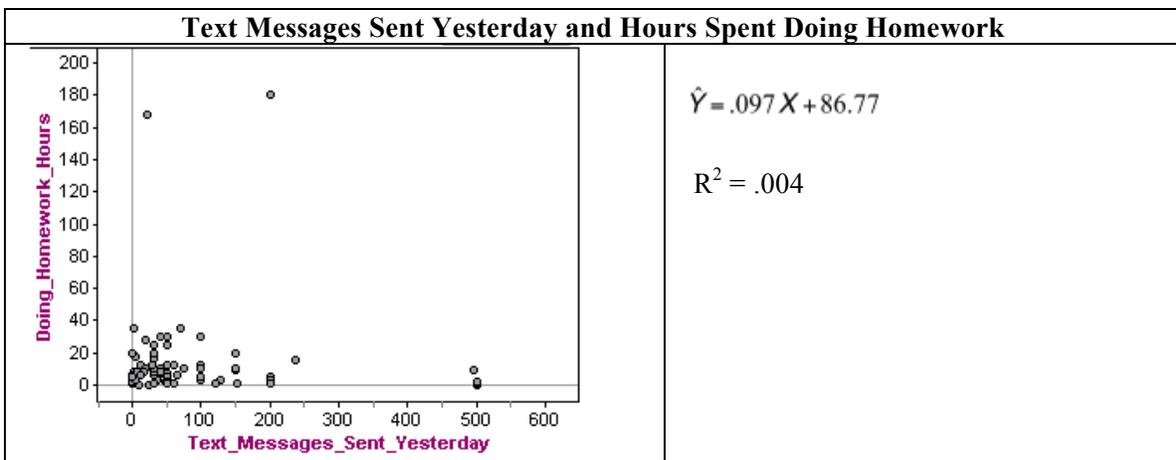
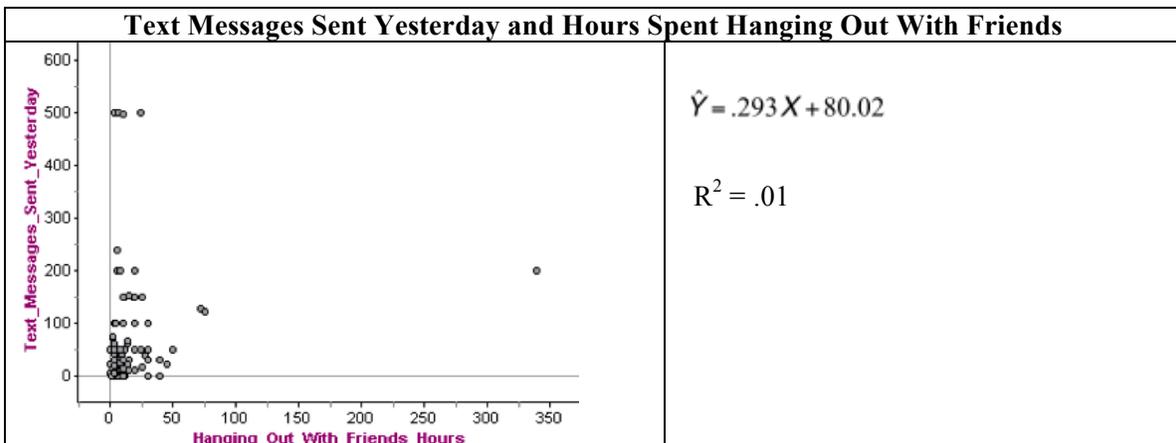
Note for Teachers: The error term is included in this model to represent all other unknown factors that may contribute to the variation of Number of Text Messages Sent. When the model is actually estimated for a given sample data set, the error term no longer needs to be included.

This gives the least squares regression line as:

$$\hat{Y} = \underline{\hspace{2cm}} X + \underline{\hspace{2cm}}$$

Here are three snapshots of sample scatterplots showing graphs and equations of the least squares regression lines for each of the three questions, along with the value of R^2 for each. The value of the correlation coefficient will be calculated using R^2 .

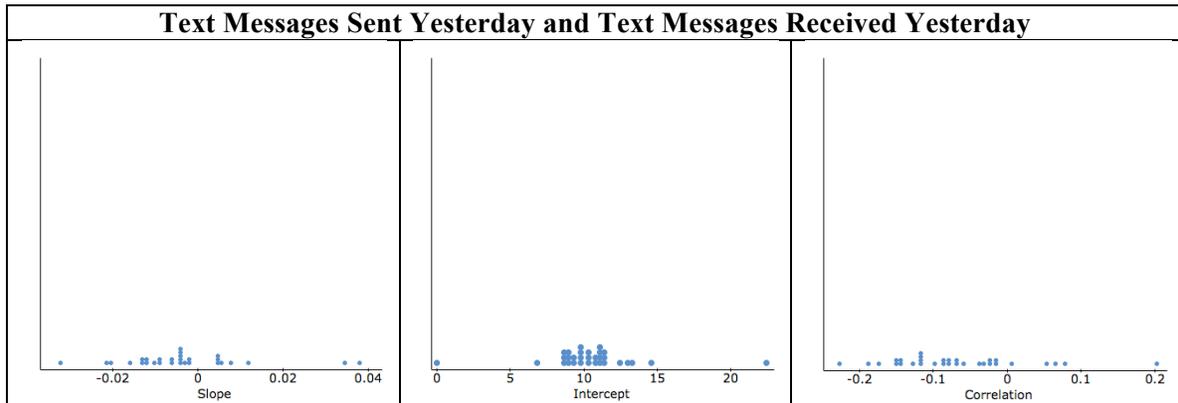




4. Interpret Results

Students will be asked to interpret the intercept and the slope in the context of their question. Once their individual activities have been completed, they will work together as a class. Teachers will arrange the class into groups according to the questions they worked on. Together the groups will write in the equation of the line they found specifying the slope, intercept, and correlation coefficient they found. At this point, the class should about 100 slopes, intercepts, and correlations *for each* question. It is very important to **save the data sets** from which each regression was generated in order to have a way to check if mistakes were made while making the dotplots for #2 in the group activity. A recommended way for each student to save their own data in Excel is included on the ProjectSet website (titled: “Example Census Group Activity Data”). As a class, the students will plot these slopes, intercepts, and correlations in dotplots (a total of 9 dotplots will be created: three dotplots for each of the three questions). Students will then use these dotplots to make inferences about the population parameters for the slopes, intercepts, and correlations.

Here are some example dotplots for the homework versus text message question for the slope, intercept, and correlation coefficient:



Possible Extensions

1. Use the Census at School data to explore relationships between different pairs of variables.
2. Break up each sample by gender and repeat the activity for each gender separately. Compare whether there are differences between males' and females' text messaging experiences.

References

1. Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report, ASA, Franklin et al., ASA, 2007 <http://www.amstat.org/education/gaise/>.
2. Census for School Project <http://www.amstat.org/censusatschool/>



PROJECT-SET

Statistics Education for Teachers

Activity



Text messaging is time consuming! What gives?

Introduction

Many of us send lots of text messages throughout a day. What factors could be related to the number of text messages one sends in a day? In this activity, we will explore the relationship between the number of text messages one sends in a day and a few other potential explanatory factors. Each student will work individually and then share their results with the class. Once each student has obtained their results, we will work together as a group to draw further conclusions.

Choose one of the following questions to explore:

1. Does the *number of hours you spend hanging out with friends* in a day increase or decrease the number of text messages you send?
2. Does the *number of hours you spend doing homework* in a day increase or decrease the number of text messages you send?
3. Does the *number of text messages you receive* in a day increase or decrease the number of text messages you send?

To answer the question you choose, you are going to download and work with a real data set. To download the data set, go to the following website:

<http://www.amstat.org/censusatschool/>

Before we begin the individual activity, let's read about the Census at School project as a class on their home page. In this activity, we will be drawing random samples of students to analyze their results. In addition, at the end of the activity, you may also complete the survey, so your data will be included in future activities.

You will each individually carry out the following steps:

1. Click on Random Sampler
2. Accept the Terms & Conditions
3. Select a sample size of 100 from All States and 9, 10, 11, and 12 grade levels. Include All Genders and All Years of data collection.
4. Download the data into Excel.

5. Open the data in Excel. You will see a large number of variables (labeled in each column).

6. Delete all the columns **except** for the following:

- Gender
- Text Messages Sent Yesterday
- Text Messages Received Yesterday, Hanging out with Friends Hours
- Doing Homework Hours

7. Depending on which question you choose above, plot a scatterplot to visually see the data. Does the relationship appear to be linear? Are there any outliers? What are some possible explanations for why there could be outliers? Should you eliminate the outliers in your data set? Why or why not?

8. Depending on which question you choose above, estimate the least squares regression line for your downloaded data. To carry out this step, you may use any technology available to you in the classroom.

If you choose question (1) above, you will estimate the following model:

$$\text{Number of Text Messages Sent} = \alpha + \beta(\text{Hours Spent with Friends}) + \varepsilon$$

If you choose question (2) above, you will estimate the following model:

$$\text{Number of Text Messages Sent} = \alpha + \beta(\text{Hours Spent doing Homework}) + \varepsilon$$

If you choose question (3) above, you will estimate the following model:

$$\text{Number of Text Messages Sent} = \alpha + \beta(\text{Number of Text Messages Received}) + \varepsilon$$

This gives the least squares regression line as:

$$\hat{Y} = \underline{\hspace{2cm}} X + \underline{\hspace{2cm}}$$

Interpret the slope in the context of the question.

Interpret the y-intercept in the context of the question.

Interpret the correlation coefficient in the context of the question.



HERE AND WAIT FOR OTHERS TO FINISH



Class Activity

Divide the class into three groups according to the question they chose to work on. Each group needs to collect about 100 samples and estimate the regression equation for each sample. To guide you, you may want to use the following table for their question. For each regression equation you found, **save** the slope value, intercept value, and correlation value in a spreadsheet.

1. Collect the regression line information found by each person in the class in a table (depending on the size of your class, you may not use every row):

Question Chosen: Number of Text Messages Sent vs. _____

Student	Regression Line Equation	Slope	Intercept	Correlation
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				

2. Create a dotplot for all of the slopes found for each of the questions (i.e., three dotplots will be created – one for each question).

3. For this question, each student should look at the dotplot that pertains to their question. This dotplot represents an approximation to the **sampling distribution** of the slope of the regression line.

a) What do you notice about the dotplot?

b) What is the range of the slope? What seems to be the most common slope?

c) If you had to guess what the slope of the regression line was for the entire population, what would you guess? Explain why.

4. Repeat #2 and #3 above for the intercept.

5. Repeat #2 and #3 above for the correlation.