

Name \_\_\_\_\_

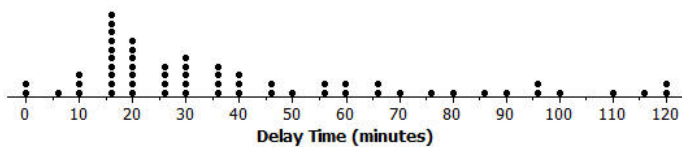
Date \_\_\_\_\_

## Lesson 1: Distributions and Their Shapes

### Exit Ticket

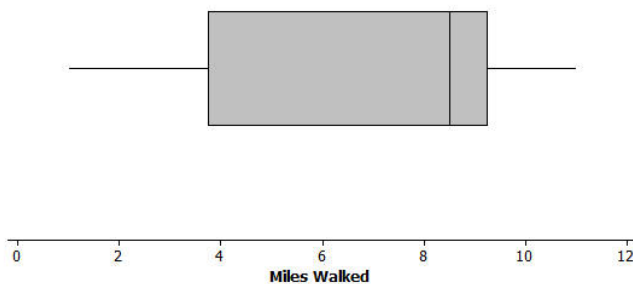
1. Sam said that a typical flight delay for the sixty BigAir flights was approximately one hour. Do you agree? Why or why not?

**Dot Plot of December Delay Times**

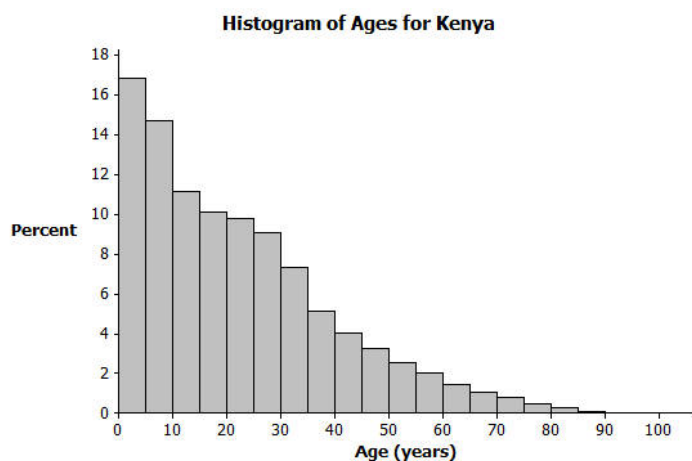


2. Sam said that 50% of the twenty-two juniors at River City High School who participated in the walkathon walked at least ten miles. Do you agree? Why or why not?

**Boxplot of Miles Walked for Juniors**



3. Sam said that young people from the ages of 0 to 10 years old make up nearly one-third of the Kenyan population. Do you agree? Why or why not?



Name \_\_\_\_\_

Date \_\_\_\_\_

## Lesson 2: Describing the Center of a Distribution

### Exit Ticket

Each person in a random sample of ten ninth graders was asked two questions:

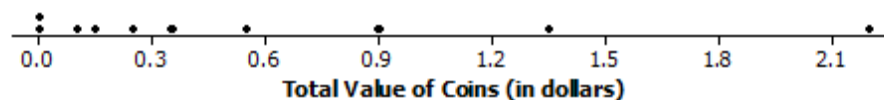
- How many hours did you spend watching TV last night?
- What is the total value of the coins you have with you today?

Here are the data for these ten students:

Student	Hours of TV	Total Value of Coins (in dollars)
1	2	0.00
2	1	0.89
3	0	2.19
4	3	0.15
5	4	1.37
6	1	0.36
7	2	0.25
8	2	0.00
9	4	0.54
10	3	0.10

1. Construct a dot plot of the data on Hours of TV. Would you describe this data distribution as approximately symmetric or as skewed?
2. If you wanted to describe a typical number of hours of TV watched for these ten students, would you use the mean or the median? Calculate the value of the measure you selected.

3. Here is a dot plot of the data on total value of coins.



Calculate the values of the mean and the median for this data set.

4. Why are the values of the mean and the median that you calculated in question 3 so different? Which of the mean and the median would you use to describe a typical value of coins for these ten students?

Name \_\_\_\_\_

Date \_\_\_\_\_

## Lesson 3: Estimating Centers and Interpreting the Mean as a Balance Point

### Exit Ticket

1. Draw a dot plot of a data distribution representing the ages of twenty people for which the median and the mean would be approximately the same.
2. Draw a dot plot of a data distribution representing the ages of twenty people for which the median is noticeably less than the mean.
3. An estimate of the balance point for a distribution of ages represented on a number line resulted in a greater sum of the distances to the right than the sum of the distances to the left. In which direction should you move your estimate of the balance point? Explain.

Name \_\_\_\_\_

Date \_\_\_\_\_

## Lesson 4: Summarizing Deviations from the Mean

### Exit Ticket

Five people were asked approximately how many hours of TV they watched per week. Their responses were as follows.

6   4   6   7   8

1. Find the mean number of hours of TV watched for these five people.
2. Find the deviations from the mean for these five data values.
3. Write a new set of five values that has roughly the same mean as the data set above but that has, generally speaking, greater deviations from the mean.

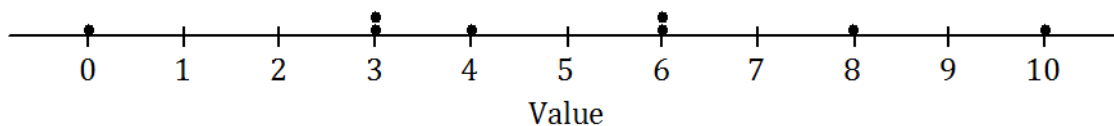
Name \_\_\_\_\_

Date \_\_\_\_\_

## Lesson 5: Measuring Variability for Symmetrical Distributions

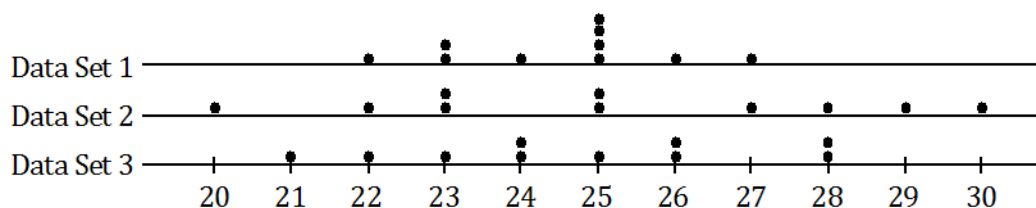
### Exit Ticket

1. Look at the dot plot below.



- a. *Estimate* the mean of this data set.
- b. Remember that the standard deviation measures a typical deviation from the mean. The standard deviation of this data set is either 3.2, 6.2, or 9.2. Which of these values is correct for the standard deviation?

2. Three data sets are shown in the dot plots below.



- a. Which data set has the smallest standard deviation of the three? Justify your answer.
- b. Which data set has the largest standard deviation of the three? Justify your answer.

Name \_\_\_\_\_

Date \_\_\_\_\_

## Lesson 6: Interpreting the Standard Deviation

### Exit Ticket

1. Use the statistical features of your calculator to find the mean and the standard deviation to the nearest tenth of a data set of the miles per gallon from a sample of five cars.

24.9   24.7   24.7   23.4   27.9

2. Suppose that a teacher plans to give four students a quiz. The minimum possible score on the quiz is 0, and the maximum possible score is 10.
- a. What is the smallest possible standard deviation of the students' scores? Give an example of a possible set of four student scores that would have this standard deviation.
- b. What is the set of four student scores that would make the standard deviation as large as it could possibly be? Use your calculator to find this largest possible standard deviation.



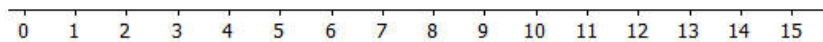
Name \_\_\_\_\_

Date \_\_\_\_\_

## Lesson 7: Measuring Variability for Skewed Distributions (Interquartile Range)

### Exit Ticket

1. A data set consisting of the number of hours each of 40 students watched television over the weekend has a minimum value of 3 hours, a Q1 value of 5 hours, a median value of 6 hours, a Q3 value of 9 hours, and a maximum value of 12 hours. Draw a box plot representing this data distribution.



2. What is the interquartile range (IQR) for this distribution? What percent of the students fall within this interval?
3. Do you think the data distribution represented by the box plot is a skewed distribution? Why or why not?
4. Estimate the typical number of hours students watched television. Explain why you chose this value.

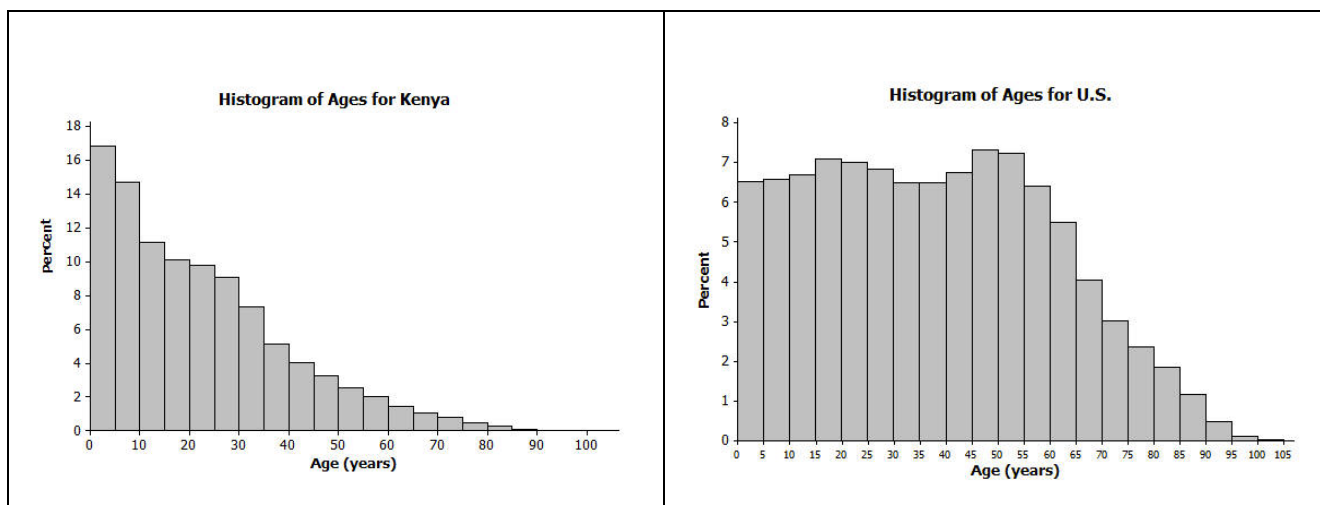
Name \_\_\_\_\_

Date \_\_\_\_\_

## Lesson 8: Comparing Distributions

### Exit Ticket

- Using the histograms of the population distributions of the United States and Kenya in 2010, approximately what percent of the people in the United States were between 15 and 50 years old? Approximately what percent of the people in Kenya were between 15 and 50 years old?

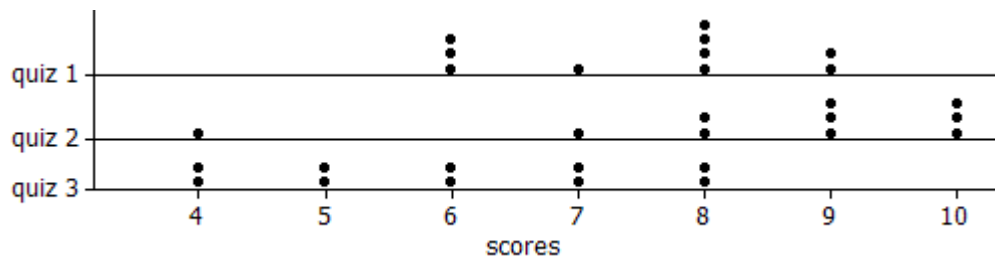


- What 5-year interval of ages represented in the 2010 histogram of the United States age distribution has the most people?
- Why is the mean age greater than the median age for people in Kenya?

Name \_\_\_\_\_

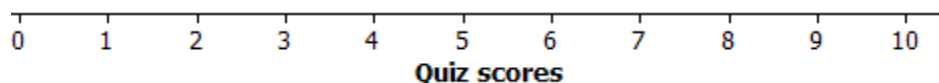
Date \_\_\_\_\_

1. The scores of three quizzes are shown in the following data plot for a class of 10 students. Each quiz has a maximum possible score of 10. Possible dot plots of the data are shown below.



- a. On which quiz did students tend to score the lowest? Justify your choice.
- b. Without performing any calculations, which quiz tended to have the most variability in the students' scores? Justify your choice based on the graphs.

- c. If you were to calculate a measure of variability for Quiz 2, would you recommend using the interquartile range or the standard deviation? Explain your choice.
- d. For Quiz 3, move one dot to a new location so that the modified data set will have a larger standard deviation than before you moved the dot. Be clear which point you decide to move, where you decide to move it, and explain why.
- e. On the axis below, arrange 10 dots, representing integer quiz scores between 0 and 10 so that the standard deviation is the largest possible value that it may have. You may use the same quiz score values more than once.



Use the following definitions to answer questions (f)–(h).

- The *midrange* of a data set is defined to be the average of the minimum and maximum values:  
 $\frac{\text{min} + \text{max}}{2}$ .
- The *midhinge* of a data set is defined to be the average of the first quartile ( $Q_1$ ) and the third quartile ( $Q_3$ ):  $\frac{Q_1 + Q_3}{2}$ .

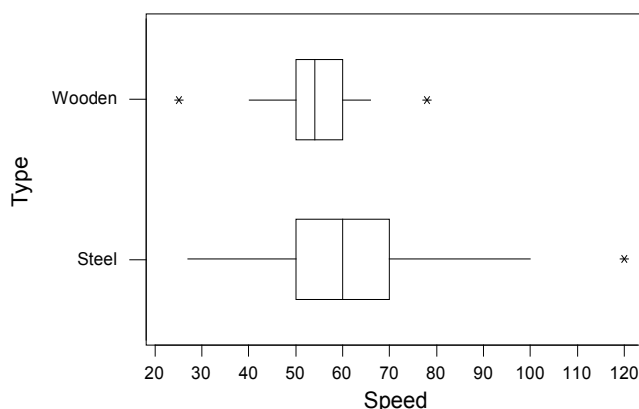
f. Is the midrange a measure of center or a measure of spread? Explain.

g. Is the midhinge a measure of center or a measure of spread? Explain.

h. Suppose the lowest score for Quiz 2 was changed from 4 to 2, and the midrange and midhinge are recomputed. Which will change more?

- A. Midrange
- B. Midhinge
- C. They will change the same amount.
- D. Cannot be determined

2. The box plots below display the distributions of maximum speed for 145 roller coasters in the United States, separated by whether they are wooden coasters or steel coasters.



Based on the box plots, answer the following questions or indicate that you do not have enough information.

- a. Which type of coaster has more observations?

A. Wooden  
B. Steel  
C. About the same  
D. Cannot be determined

Explain your choice:

- b. Which type of coaster has a higher percentage of coasters that go faster than 60 mph?

A. Wooden  
B. Steel  
C. About the same  
D. Cannot be determined

Explain your choice:

- c. Which type of coaster has a higher percentage of coasters that go faster than 50 mph?

A. Wooden  
B. Steel  
C. About the same  
D. Cannot be determined

Explain your choice:

- d. Which type of coaster has a higher percentage of coasters that go faster than 48 mph?

A. Wooden  
B. Steel  
C. About the same  
D. Cannot be determined

Explain your choice:

- e. Write 2–3 sentences comparing the two types of coasters with respect to which type of coaster normally goes faster.

Name \_\_\_\_\_

Date \_\_\_\_\_

## Lesson 9: Summarizing Bivariate Categorical Data

### Exit Ticket

1. A survey asked the question, “How tall are you to the nearest inch?” A second question on this survey asked, “What sports do you play?” Indicate what type of data, numerical or categorical, would be collected from the first question? What type of data would be collected from the second question?

Another random sample of 100 surveys was selected. Jill had a copy of the frequency table that summarized these 100 surveys. Unfortunately, she spilled part of her lunch on the copy. The following summaries were still readable:

	To Fly	Freeze Time	Invisibility	Super Strength	Telepathy	Total
<b>Females</b>	12	15	(c)*	5	(e)*	55
<b>Males</b>	12	16	10	(j)*	3	45
<b>Total</b>	24	31	25	9	(q)*	100

2. Help Jill recreate the table by determining the frequencies for cells (c), (e), (j), and (q).
3. Of the cells (c), (e), (j), and (q), which cells represent joint frequencies?
4. Of the cells (c), (e), (j), and (q), which cells represent marginal frequencies?



Name \_\_\_\_\_

Date \_\_\_\_\_

## Lesson 10: Summarizing Bivariate Categorical Data with Relative Frequencies

### Exit Ticket

Juniors and seniors were asked if they plan to attend college immediately after graduation, seek full-time employment, or choose some other option. A random sample of 100 students was selected from those who completed the survey. Scott started to calculate the relative frequencies to the nearest thousandth.

	Plan to Attend College	Plan to Seek Full-Time Employment	Other Options	Totals
Seniors	$\frac{25}{100} = 0.250$	$\frac{10}{100} = 0.100$		
Juniors				$\frac{45}{100} = 0.450$
Totals	$\frac{60}{100} = 0.600$	$\frac{15}{100} = 0.150$	$\frac{25}{100} = 0.250$	$\frac{100}{100} = 1.000$

- Complete the calculations of the relative frequencies for each of the blank cells. Round your answers to the nearest thousandth.
- A school website article indicated that “A Vast Majority of Students from our School Plan to Attend College.” Do you agree or disagree with that article? Explain why you agree or disagree.
- Do you think juniors and seniors differ regarding after graduation options? Explain.

Name \_\_\_\_\_

Date \_\_\_\_\_

## Lesson 11: Conditional Relative Frequencies and Association

### Exit Ticket

Juniors and seniors were asked if they plan to attend college immediately after graduation, seek full-time employment, or choose some other option. A random sample of 100 students was selected from those who completed the survey. Scott started to calculate the row conditional relative frequencies to the nearest thousandth.

	Plan to Attend College	Plan to Seek Full-Time Employment	Other Options	Totals
Seniors	$\frac{25}{55} \approx 0.455$	$\frac{10}{55} \approx 0.182$	$\frac{20}{55} \approx ???$	$\frac{55}{55} = 1.000$
Juniors	$\frac{35}{45} \approx ???$	$\frac{5}{45} \approx ???$	$\frac{5}{45} \approx 0.111$	$\frac{45}{45} = 1.000$
Totals	$\frac{60}{100} = 0.600$	$\frac{15}{100} = 0.150$	$\frac{25}{100} = 0.250$	$\frac{100}{100} = 1.000$

- Complete the calculations of the row conditional relative frequencies. Round your answers to the nearest thousandth.
- Are the row conditional relative frequencies for juniors and seniors similar, or are they very different?
- Do you think there is a possible association between grade level (junior or senior) and after high school plans? Explain your answer.

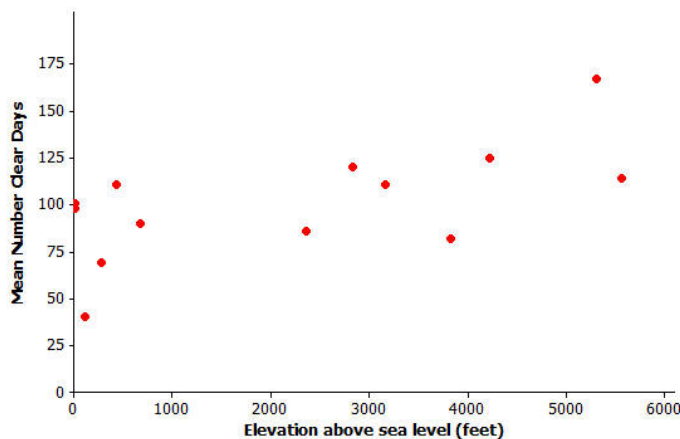
Name \_\_\_\_\_

Date \_\_\_\_\_

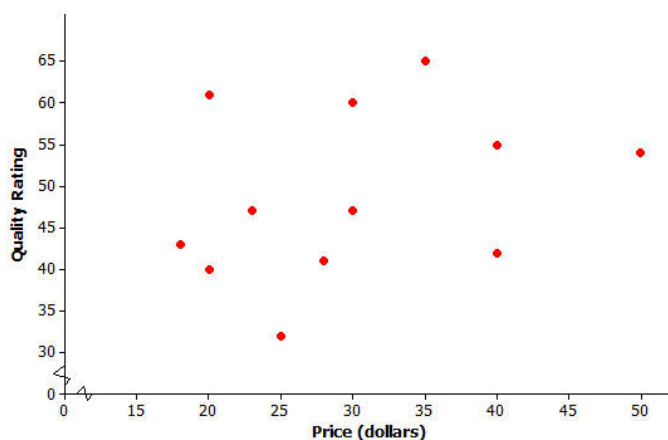
## Lesson 12: Relationships Between Two Numerical Variables

### Exit Ticket

- You are traveling around the United States with friends. After spending a day in a town that is 2,000 feet above sea level, you plan to spend the next several days in a town that is 5,000 feet above sea level. Is this town likely to have more or fewer clear days per year than the town that is 2,000 feet above sea level? Explain your answer.



- You plan to buy a bike helmet. Based on data presented in this lesson, will buying the most expensive bike helmet give you a helmet with the highest quality rating? Explain your answer.



Data Source: [www.consumerreports.org/health](http://www.consumerreports.org/health)

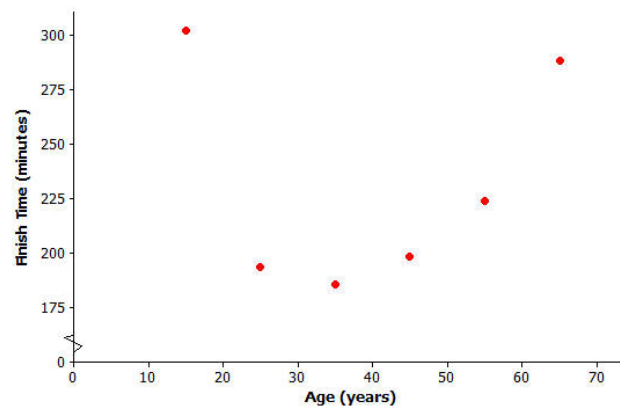
Name \_\_\_\_\_

Date \_\_\_\_\_

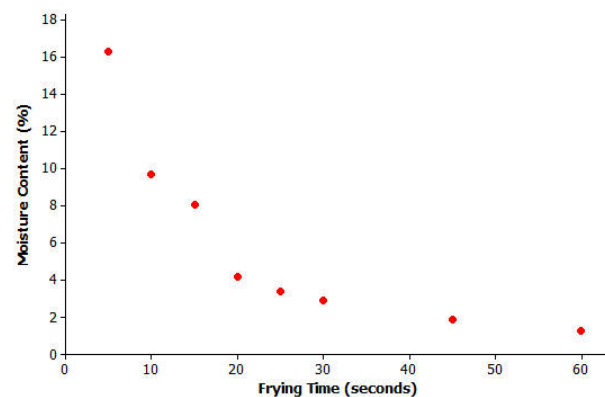
## Lesson 13: Relationships Between Two Numerical Variables

### Exit Ticket

- Here is the scatter plot of age (in years) and finish time (in minutes) of the NY City Marathon that you first saw in an example. What type of model (linear, quadratic, or exponential) would best describe the relationship between age and finish time? Explain your reasoning.



- Here is the scatter plot of frying time (in seconds) and moisture content (as a percentage) you first saw in Lesson 12. What type of model (linear, quadratic, or exponential) would best describe the relationship between frying time and moisture content? Explain your reasoning.



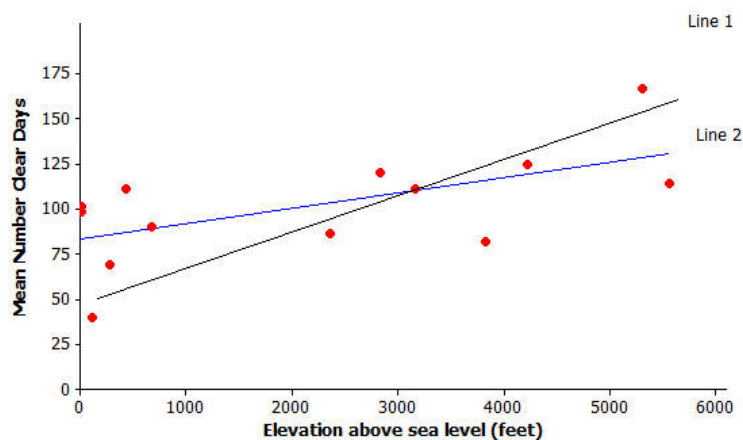
Name \_\_\_\_\_

Date \_\_\_\_\_

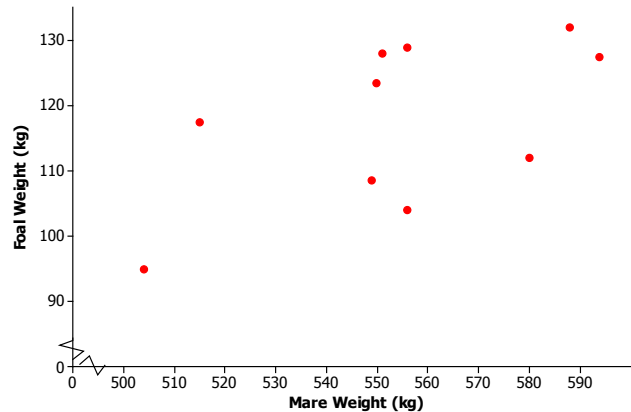
## Lesson 14: Modeling Relationships with a Line

### Exit Ticket

1. The scatter plot below displays the elevation and mean number of clear days per year of 14 U.S. cities. Two lines are shown on the scatter plot. Which represents the least squares line? Explain your choice.



2. Below is a scatter plot of foal birth weight and mare's weight.



- a. The equation of the least squares line for the data is  $y = -19.6 + 0.248x$ , where  $x$  = mare's weight (in kg) and  $y$  = foal's birth weight (in kg).  
What foal birth weight would you predict for a mare that weighs 520 kg?

- b. How would you interpret the value of the slope in the least squares line?

- c. Does it make sense to interpret the value of the  $y$ -intercept in this context? Explain why or why not.

Name \_\_\_\_\_

Date \_\_\_\_\_

## Lesson 15: Interpreting Residuals from a Line

### Exit Ticket

1. Meerkats have a gestation time of 70 days.
  - a. Use the equation of the least squares line from today's class,  $y = 6.643 + 0.03974x$ , to predict the longevity of the meerkat. Remember  $x$  equals the gestation time in days and  $y$  equals the longevity in years.
  - b. Approximately how close might your prediction be to the actual longevity of the meerkat? What was it (from class) that told you roughly how close a prediction might be to the true value?
  - c. According to your answers to parts (a) and (b), what is a reasonable range of possible values for the longevity of the meerkat?
  - d. The longevity of the meerkat is actually 10 years. Use this value and the predicted value that you calculated in part (a) to find the residual for the meerkat.

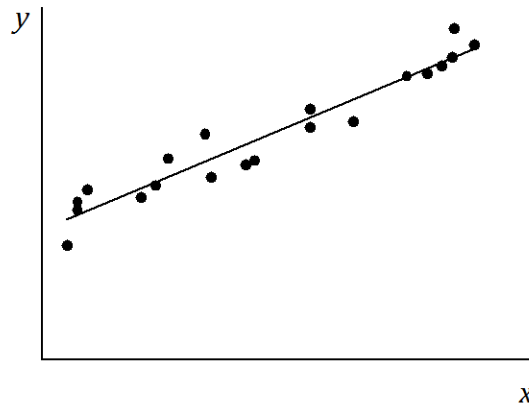
Name \_\_\_\_\_

Date \_\_\_\_\_

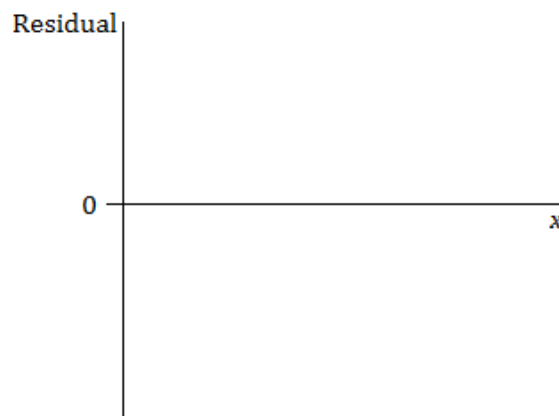
## Lesson 16: More on Modeling Relationships with a Line

### Exit Ticket

1. Suppose you are given a scatter plot (with least squares line) that looks like this:

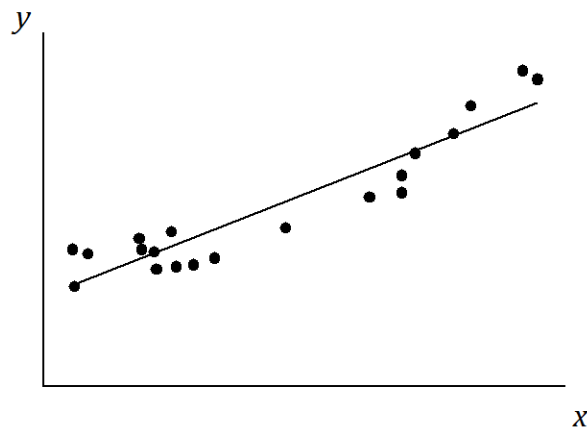


What would the residual plot to look like? Make a quick sketch on the axes given below. (There's no need to plot the points exactly.)

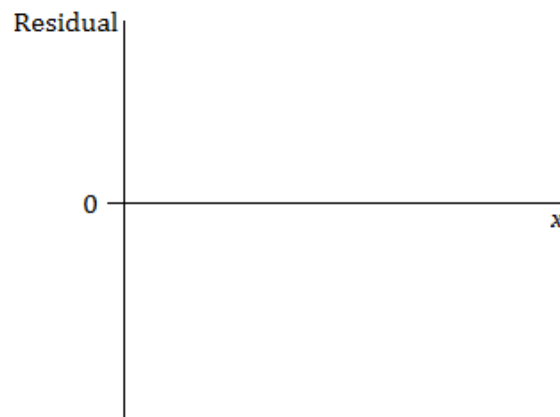




2. Suppose the scatter plot looked like this:



Make a quick sketch on the axes below of how the residual plot would look.



Name \_\_\_\_\_

Date \_\_\_\_\_

## Lesson 17: Analyzing Residuals

### Exit Ticket

1. If you see a random scatter of points in the residual plot, what does this say about the original data set?
2. Suppose a scatter plot of bivariate numerical data shows a linear pattern. Describe what you think the residual plot would look like. Explain why you think this.



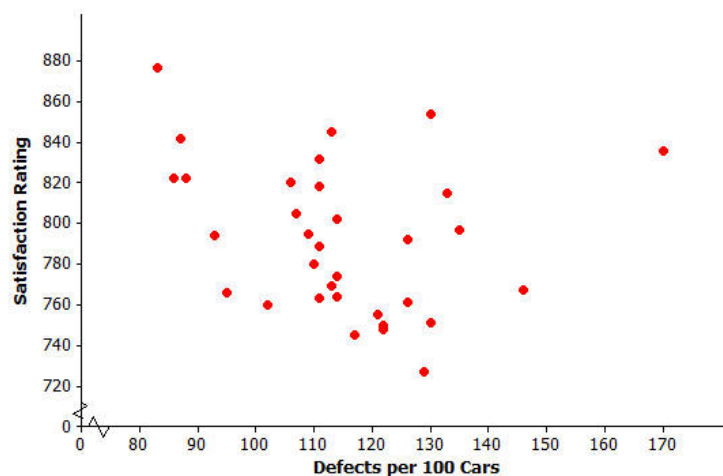
Name \_\_\_\_\_

Date \_\_\_\_\_

## Lesson 19: Interpreting Correlation

### Exit Ticket

The scatter plot below displays data on the number of defects per 100 cars and a measure of customer satisfaction (on a scale from 1 to 1000, with higher scores indicating greater satisfaction) for the 33 brands of cars sold in the United States in 2009.



Data Source: *USA Today*, June 16, 2010 and July 17, 2010

- a. Which of the following is the value of the correlation coefficient for this data set:  $r = -0.95$ ,  $r = -0.24$ ,  $r = 0.83$ , or  $r = 1.00$ ?
- b. Explain why you selected this value.

Name \_\_\_\_\_

Date \_\_\_\_\_

1. A recent social survey asked 654 men and 813 women to indicate how many “close friends” they have to talk about important issues in their lives. Below are frequency tables of the responses.

Number of Close Friends	0	1	2	3	4	5	6	Total
Males	196	135	108	100	42	40	33	654
Females	201	146	155	132	86	56	37	813

- a. The shape of the distribution of the number of close friends for the males is best characterized as
- A. Skewed to the higher values (right or positively skewed).
  - B. Skewed to the lower values (left or negatively skewed).
  - C. Symmetric.
- b. Calculate the median number of class friends for the females. Show your work.
- c. Do you expect the mean number of close friends for the females to be larger or smaller than the median you found in part (b), or do you expect them to be the same? Explain your choice.
- d. Do you expect the mean number of close friends for the males to be larger or smaller than the mean number of close friends for the females, or do you expect them to be the same? Explain your choice.

2. The physician's health study examined whether physicians who took aspirin were less likely to have heart attacks than those who took a placebo (fake) treatment. The table below shows their findings.

	Placebo	Aspirin	Total
Heart attack	189	104	293
No heart attack	10,845	10,933	21,778
Total	11,034	11,037	22,071

Based on the data in the table, what conclusions can be drawn about the association between taking aspirin and whether or not a heart attack occurred? Justify your conclusion using the given data.

3. Suppose 500 high school students are asked the following two questions:

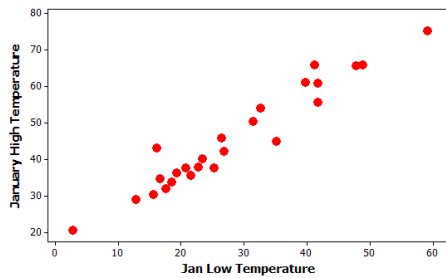
- What is the highest degree you plan to obtain? (check one)
  - ☐ High school degree    ☐ College (Bachelor's degree)
  - ☐ Graduate school (e.g., Master's degree or higher)
- How many credit cards do you currently own? (check one)
  - ☐ None                      ☐ One                      ☐ More than one

Consider the data shown in the following frequency table.

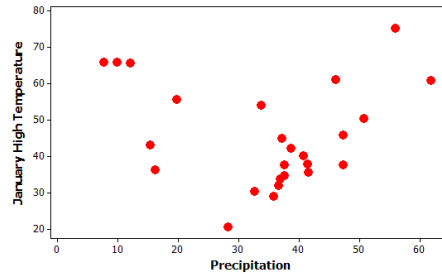
	No Credit Cards	One Credit Card	More than One Credit Card	Total
High school	?		6	59
College	120	240	40	394
Graduate school				47
Total		297		500

Fill in the missing value in the cell in the table that is marked with a “?” so that the data would be consistent with no association between education aspiration and current number of credit cards for these students. Explain how you determined this value.

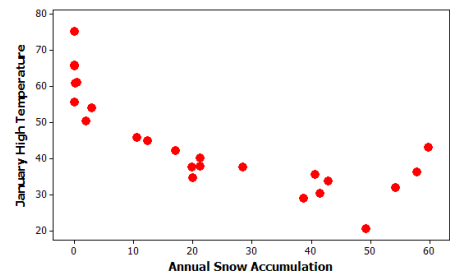
4. Weather data were recorded for a sample of 25 American cities in one year. Variables measured included January high temperature (in degrees Fahrenheit), January low temperature (in degrees Fahrenheit), annual precipitation (in inches), and annual snow accumulation. The relationships for three pairs of variables are shown in the graphs below (Jan. Low Temperature—Graph A; Precipitation—Graph B; Annual Snow Accumulation—Graph C).



Graph A



Graph B



Graph C

- a. Which pair of variables will have a correlation coefficient closest to 0?

- A. Jan. high temperature and Jan. low temperature
- B. Jan. high temperature and precipitation
- C. Jan. high temperature and snow accumulation

Explain your choice:

- b. Which of the above scatterplots would be best described as a strong nonlinear relationship? Explain your choice.



- c. Suppose we fit a least squares regression line to Graph A. Circle one word choice for each blank that best completes this sentence, based on the equation:

*If I compare a city with a January low temperature of 30°F to a city with a higher January low temperature, then the   (1)   January high temperature of the second city will   (2)   be   (3)  .*

(1) actual, predicted

(2) probably, definitely

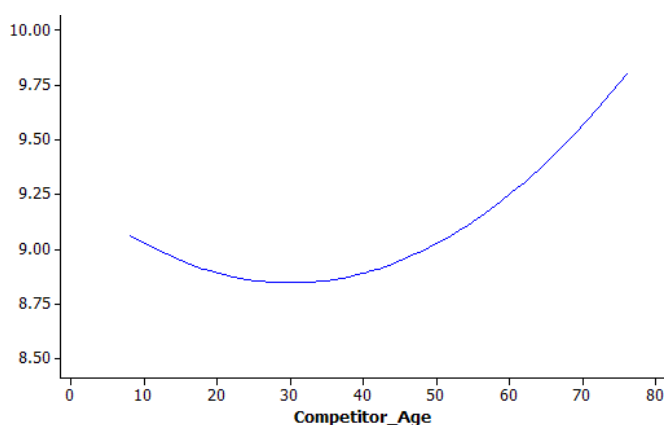
(3) smaller, larger, the same, equally likely to be higher or lower

- d. For the city with a January low temperature of 30°F, what do you predict for the annual snow accumulation? Explain how you are estimating this based on the three graphs above.

5. Suppose times (in minutes) to run one mile were recorded for a sample of 100 runners, ages 16–66 years, and the following least squares regression line was found.

$$\text{Predicted time in minutes to run one mile} = 5.35 + 0.25 \times (\text{age})$$

- a. Provide an interpretation in context for this slope coefficient.
- b. Explain what it would mean in the context of this study for a runner to have a negative residual.
- c. Suppose, instead, that someone suggests using the following curve to predict time to run one mile. Explain what this model implies about the relationship between running time and age, and why that relationship might make sense in this context.



- d. Based on the results for these 100 runners, explain how you could decide whether the first model or the second model provides a better fit to the data.
- e. The sum of the residuals is always equal to zero for the least squares regression line. Which of the following must also always be equal to zero?
- A. The mean of the residuals
  - B. The median of the residuals
  - C. Both the mean and the median of the residuals
  - D. Neither the mean nor the median of the residuals