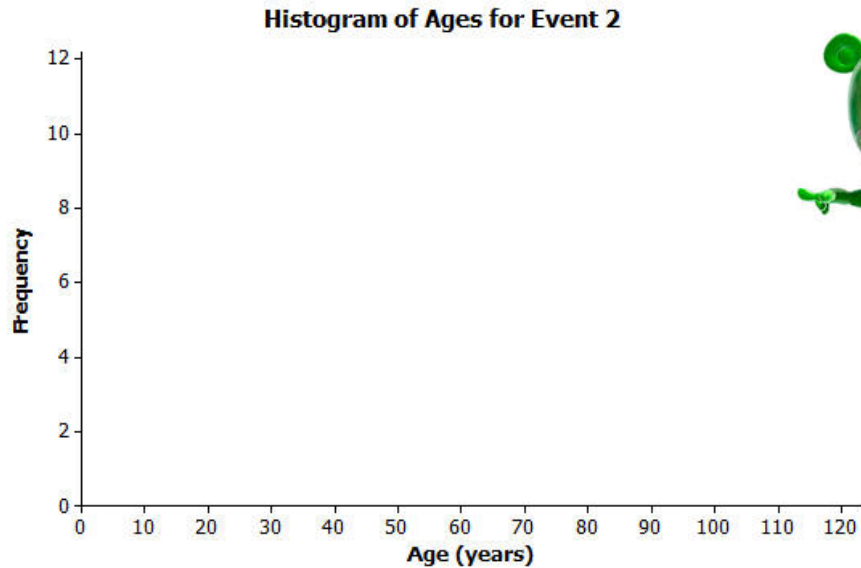


Unit 7 Problem Set Packet - Statistics

Name: _____ Class: _____

•HW 7.1 - Distributions, Shapes, and Centers•

- 1.) Forty people were attending an event. The ages of the people are:
- 6, 13, 24, 27, 28, 32, 32, 34, 38, 42, 42, 43, 48, 49, 49, 49, 51, 52, 52, 53,
53, 53, 54, 55, 56, 57, 57, 60, 61, 61, 62, 66, 66, 66, 68, 70, 72, 78, 83, 97
- a.) Create a histogram of the ages using the provided axes.



- b.) Would you describe your graph of ages as symmetrical or skewed? Explain your choice.
- c.) Identify a typical age of the forty people.
- d.) What event do you think the forty people were attending? Use your histogram to justify your conjecture.
- e.) How would you describe the differences in this histogram and the histogram in today's notes?

For 2 - 6, consider the following scenario.. The company that created a popular video game, "Leaders," plans to release a significant upgrade of the game. Users earn or lose points for making decisions as the leader of an imaginary country. In most cases, repeated playing of the game improves a user's ability to make decisions. The company will launch an online advertising campaign, but at the moment, they are not sure how to focus the advertising. Your goal is to help the company decide how the advertising campaign should be focused. Five videos have been proposed for the following target audiences:

- Video 1: Target females with beginning level scores.
- Video 2: Target males with advance level scores.
- Video 3: Target all users with middle range scores
- Video 4: Target males with beginning level scores.
- Video 5: Target females with advanced level scores.

- 2.) Why might the company be interested in the developing different videos based on user score?
- 3.) Thirty female users and twenty-five male users were selected at random from a database of people who play the game regularly. Each of them agreed to be part of a research study and report their scores. A leadership score is based on a player's answers to leadership questions. A score of 1 to 40 is considered a beginning level leadership score, a score of 41 to 60 is considered a middle level leadership score, and a score of greater than 60 is considered an advanced level leadership score. Use the following data to make a dot plot of the female scores, a dot plot of the male scores, and a dot plot of the scores for the combined group of males and females.

Female scores:

10	20	20	20	30	30	30	40	40	40
50	50	55	65	65	65	65	65	70	70
70	70	76	76	76	76	76	76	76	76



Male scores:

15	20	20	25	25	25	25	30	30	30
30	30	30	35	35	35	35	35	40	40
40	45	45	45	50					



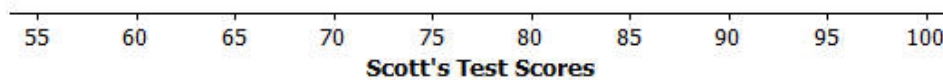
- 4.) What do you think is a typical score for a female user? What do you think is a typical score for a male user? Explain how you determined these typical scores.

- 5.) Why is it more difficult to report a typical score for the overall group that includes both the males and females?

- 6.) Production costs will only allow for two video advertisements to be developed. Which two videos would you recommend for development? Explain your recommendations.

•HW 7.2 - Estimating Centers and Interpreting Mean as a Balance Point•

For 7 - 18, consider another example of balance. Mr. Jackson is a mathematics teacher at Waldo High School. Students in his class are frequently given quizzes or exams. He indicated to his students that an exam is worth 4 quizzes when calculating an overall weighted average to determine their final grade. During one grading period, Scott got an 80% on one exam, a 90% on a second exam, and a 60% and a 70% on two quizzes. How could we represent Scott's test scores? Consider the following number line:



- 7.) What values are represented by the number line?
- 8.) If one "." was used to represent a quiz score, how might you represent an exam score?
- 9.) Represent Scott's exams and quizzes on this number line using "." symbols.
- 10.) Mr. Jackson indicated students should set an 85% overall weighted average as a goal. Do you think Scott met that goal? Explain your answer.
- 11.) Place an X on the number line at a position that you think locates the balance point of all of the "." symbols. Determine the sum of the distances from the X to each "." on the right side of the X.

12.) Determine the sum of the distances from the X to each " \cdot " on the left side of the X .

13.) Do the total distances to the right of the X equal the total distances to the left of the X ?



14.) Based on your answer to question (7), would you change your estimate of the balance point? If yes, where would you place your adjusted balance point? How does using this adjusted estimate change the total distances to the right of your estimate and the total distances to the left?

15.) Scott's weighted average is 81. Recall that each exam score is equal to 4 times a quiz score. Show the calculations that lead to this weighted average.

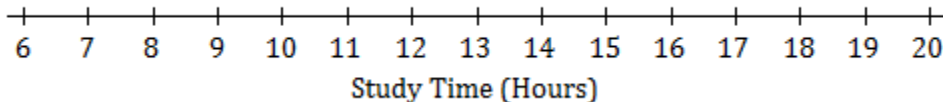
16.) How does the calculated mean score compare with your estimated balance point?

17.) Compute the total distances to the right of the mean and the total distances to the left of the mean. What do you observe?

18.) Did Scott achieve the goal set by Mr. Jackson of an 85% average? Explain your answer.

•HW 7.3 - Deviations from the Mean, Variability for Symmetric Distribution, and Standard Deviation•

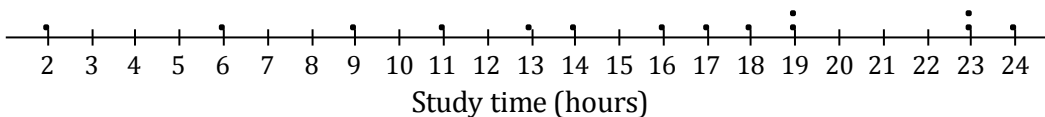
19.) Ten members of a high school girls' basketball team were asked how many hours they studied in a typical week. Their responses (in hours) were: 20, 13, 10, 6, 13, 10, 13, 11, 11, 10. Using the axis given below, draw a dot plot of these values. (Remember, when there are repeated values, stack the dots with one above the other.)



- a.) Calculate the mean study time for these students.
- b.) Calculate the deviations from the mean for these study times, and write your answers in the appropriate places in the table below.

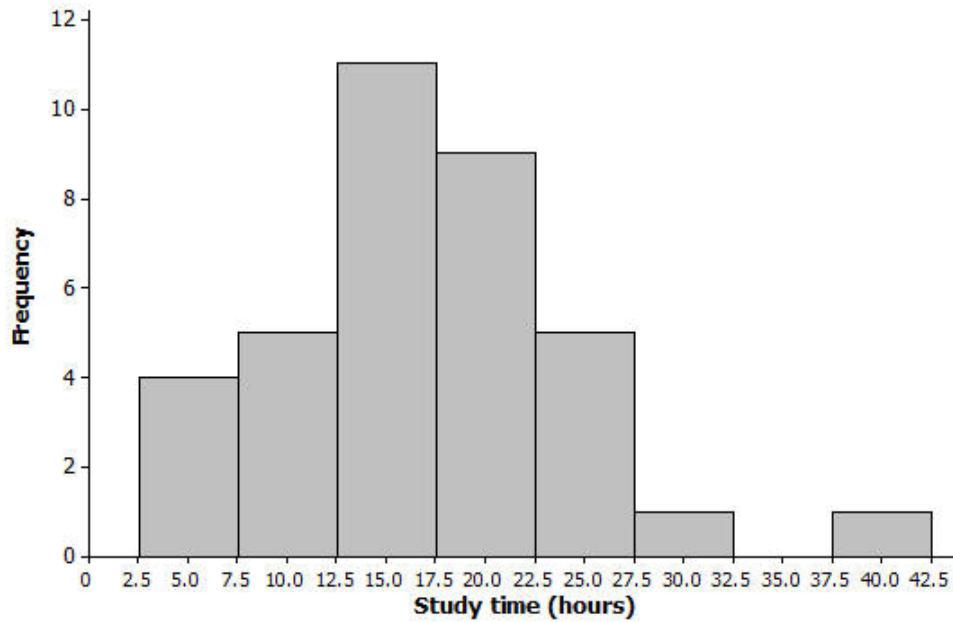
Number of Hours Studied	20	13	10	6	13	10	13	11	11	10
Deviation from the Mean										

c.) The study times for 14 girls from the soccer team at the same school as the one above are shown in the dot plot below.



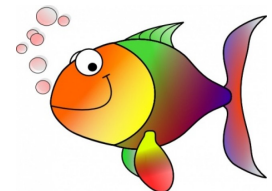
Based on the data, would the deviations from the mean (ignoring the sign of the deviations) be greater or less for the soccer players than for the basketball players?

20.) All the members of a high school softball team were asked how many hours they studied in a typical week. The results are shown in the histogram below.



a.) We can see from the histogram that four students studied around 5 hours per week. How many students studied around 15 hours per week?

b.) How many students were there in total?



c.) Suppose that the four students represented by the histogram bar centered at 5 had all studied exactly 5 hours, the five students represented by the next histogram bar had all studied exactly 10 hours, and so on. If you were to add up the study times for all of the students, what result would you get?

d.) What is the mean study time for these students?

- 21.) A small car dealership has 12 sedan cars on its lot. The fuel efficiency (mpg) values of the cars are given in the table below.

Fuel Efficiency (miles per gallon)	29	35	24	25	21	21	18	28	31	26	26	22
---------------------------------------	----	----	----	----	----	----	----	----	----	----	----	----

Use your calculator to find the mean and standard deviation for the fuel efficiency for these cars.

- 22.) The same dealership has six SUVs on its lot. The fuel efficiencies (in miles per gallon) of these cars are shown below.

21 21 21 30 28 24

Calculate the mean and the standard deviation of these values.

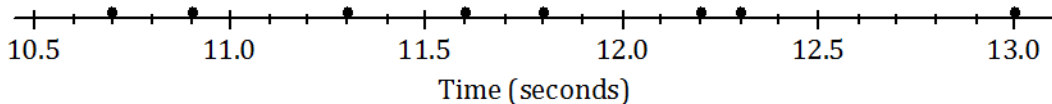


©David Rey * illustrationsOf.com/32674

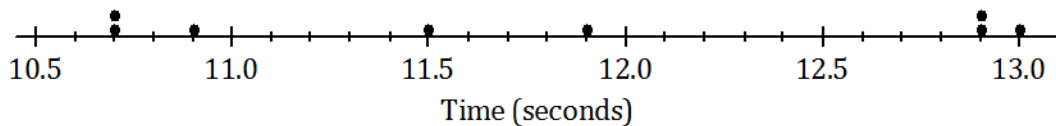
- 23.) Which of the two data sets (Question (21) or Question (22)) has the larger standard deviation? What does this tell you about the two types of cars (sedans and SUVs)?

- 24.) At a track meet there were three men's 100m races. The sprinters' times were recorded to the nearest 1/10 of a second. The results of the three races are shown in the dot plots below.

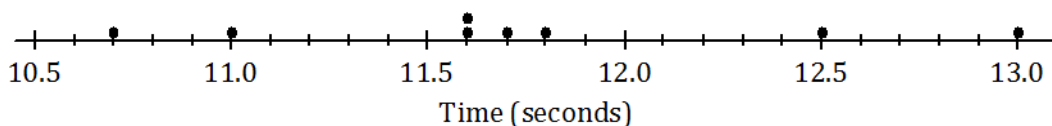
Race 1



Race 2



Race 3



- a.) Remember that the size of the standard deviation is related to the sizes of the deviations from the mean. Without doing any calculations, indicate which of the three races has the smallest standard deviation of times. Justify your answer.
- b.) Which race had the largest standard deviation of times? (Again, don't do any calculations!) Justify your answer.
- c.) Use your calculator to find the mean and the standard deviation for each of the three races. Write your answers in the table below to the nearest thousandth.

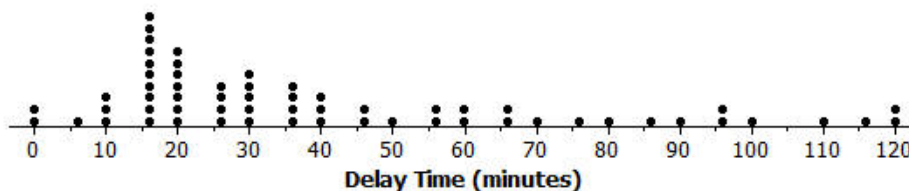
	Mean	Standard Deviation
Race 1		
Race 2		
Race 3		

•HW 7.4 - Measuring Variability for Skewed Distribution, Comparing Distributions•

Consider the following scenario. Transportation officials collect data on flight delays (the number of minutes a flight takes off after its scheduled time).

Consider the dot plot of the delay times in minutes for 60 BigAir flights during December 2012:

Dot Plot of December Delay Times

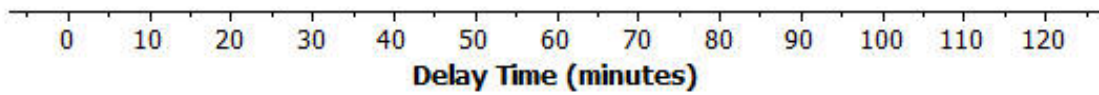


- 25.) How many flights left 60 or more minutes late?
- 26.) Why is this data distribution considered skewed?

27.) Is the tail of this data distribution to the right or to the left? How would you describe several of the delay times in the tail?

min	
Q1	
med	
Q3	
max	

28.) Draw a box below of the flights for December.



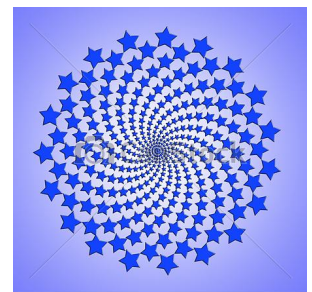
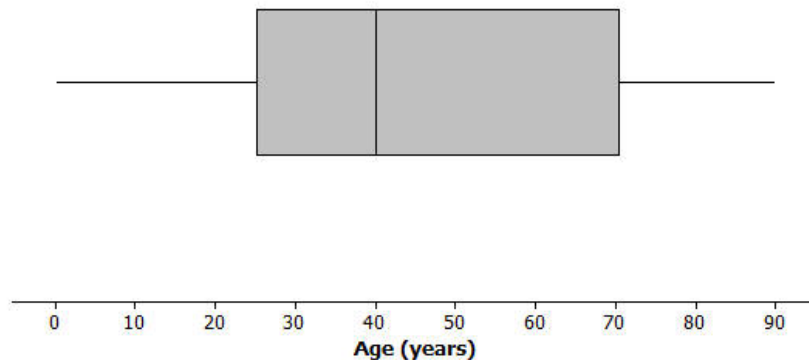
29.) What is the inter-quartile range or IQR of this data set?

30.) The mean of the 60 flight delays is approximately 42 minutes. Do you think that 42 minutes is typical of the number of minutes a BigAir flight was delayed? Why or why not?



The following box plot summarizes ages for a random sample from a made up country named Math Country.

Boxplot of Ages for Sample From Math Country



© Can Stock Photo - csp12749939

- 31.) Make up your own sample of forty ages that could be represented by the box plot for Math Country. Use a dot plot to represent the ages of the forty people in Math Country.



- 32.) Is the sample of 40 ages represented in your dot plot of Math Country the only one that could be represented by the box plot? Explain your answer.

•HW 7.5 - Summarizing Bivariate Categorical Data with Relative Frequencies•

Several students at Rufus King High School were debating whether males or females were more involved in afterschool activities. There are three organized activities in the afterschool program - intramural basketball, chess club, and jazz band. Due to budget constraints, a student can only select one of these activities. The students were not able to ask every student in the school whether they participated in the afterschool program or what activity they selected if they were involved.

- 33.) Write questions that could be included in the survey to investigate the question the students are debating. Questions that could be used for this study include the following:
- 34.) Rufus King High School has approximately 1500 students. Sam suggested that the first 100 students entering the cafeteria for lunch would provide a random sample to analyze. Janet suggested that they pick 100 students based on a school identification number. Who has a better strategy for selecting a random sample? How do you think 100 students could be randomly selected to complete the survey?

35.) Consider the following results from 100 randomly selected students:

- Of the 60 female students selected, 20 of them played intramural basketball, 10 played chess, and 10 were in the jazz band. The rest of them did not participate in the afterschool program.
- Of the male students, 10 did not participate in the afterschool program, 20 played intramural basketball, 8 played in the jazz band, and the rest played chess.

A two-way frequency table to summarize the survey data was started. What label is needed in the empty cell in the first row?

	Intramural Basketball	Chess Club	Jazz Band		Total
Females					
Males					
Total					

36.) Complete the above table for the 100 students who were surveyed.

37.) The table shows the responses to the afterschool activity question for males and females. Do you think there is a difference in the responses of males and females? Explain your answer.

38.) Calculate the relative frequencies for each of the cells to the nearest thousandth. Place the relative frequencies in the cells of the following table. (The first cell has been completed as an example.)

	Intramural Basketball	Chess Club	Jazz Band	Did Not Participate	Total
Females	$\frac{20}{100} = 0.200$				
Males					
Total					

39.) Based on your relative frequency table, what is the relative frequency of students who indicated they played basketball?

- 40.) Based on your table, what is the relative frequency of males who play basketball?
- 41.) If a student were randomly selected from the students at the school, do you think the student selected would be a male or a female?
- 42.) If a student were selected at random from school, do you think this student would be involved in an after-school program? Explain your answer.

•HW 7.6 - Conditional Frequencies and Association•

Consider again the summary of data from the 100 randomly selected students in the Rufus King investigation of after-school activities and gender:

	Intramural Basketball	Chess Club	Jazz Band	Did Not Participate	Total
Females	20	10	10	20	60
Males	20	2	8	10	40
Total	40	12	18	30	100

- 43.) Construct a row conditional relative frequency table for this data. Decimal values are given to the nearest thousandth.

	Intramural Basketball	Chess Club	Jazz Band	Did Not Participate	Total
Females					
Males					
Total					

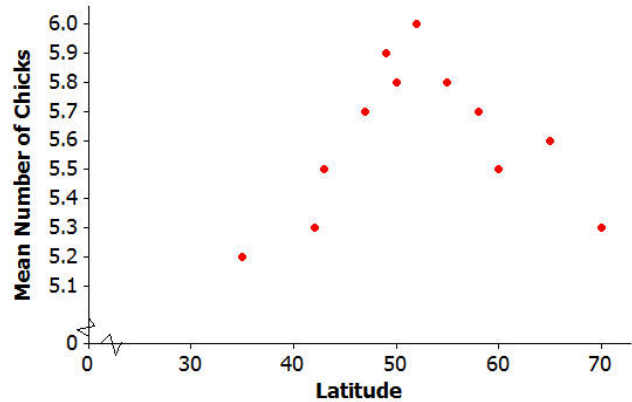
- 44.) For what after-school activities do you think the row conditional relative frequencies for females and males are very different? What might explain why males or females select different activities?
- 45.) If John, a male student at Rufus King High School, completed the after-school survey, what would you predict was his response? Explain your answer.
- 46.) If Beth, a female student at Rufus King High School, completed the after-school survey, what would you predict was her response? Explain your answer.
- 47.) Notice that 20 female students participate in intramural basketball and that 20 male students participate in intramural basketball. Is it accurate to say that females and males are equally involved in intramural basketball? Explain your answer.
- 48.) Do you think there is an association between gender and choice of after-school program? Explain.

Column conditional relative frequencies can also be computed by dividing each frequency in a frequency table by the corresponding column total to create a column conditional relative frequency table. Column conditional relative frequencies indicate the proportions or relative frequencies based on the column totals.

- 49.) If you wanted to know the relative frequency of females surveyed who participated in chess club, would you use a row conditional relative frequency or a column conditional relative frequency?
- 50.) If you wanted to know the relative frequency of band members surveyed who were female, would you use a row conditional relative frequency or a column conditional relative frequency?
- 51.) For the superpower survey data, write a question that would be answered using a row conditional relative frequency.
- 52.) For the superpower survey data, write a question that would be answered using a column conditional relative frequency.

•HW 7.7 - Relationships Between Two Numerical Variables & Modeling Relationships with a Line•

Biologists conducted a study of the nesting behavior of a type of bird called a flycatcher. They examined a large number of nests and recorded the latitude for the location of the nest and the number of chicks in the nest.



53.) What type of model (linear, quadratic or exponential) would best describe the relationship between latitude and mean number of chicks?

54.) One model that could be used to describe the relationship between mean number of chicks and

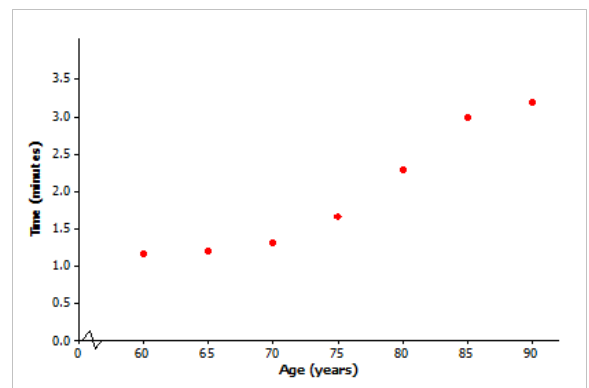
latitude is: $y = 0.175 + 0.21x - 0.002x^2$ where x represents the latitude of the location of the nest and y represents the number of chicks in the nest. Use the quadratic model to complete the following table. Then sketch a graph of the quadratic curve on the scatter plot above.

x	y
30	
40	
50	
60	
70	

55.) Based on this quadratic model, what is the best latitude for hatching the most flycatcher chicks? Justify your choice.

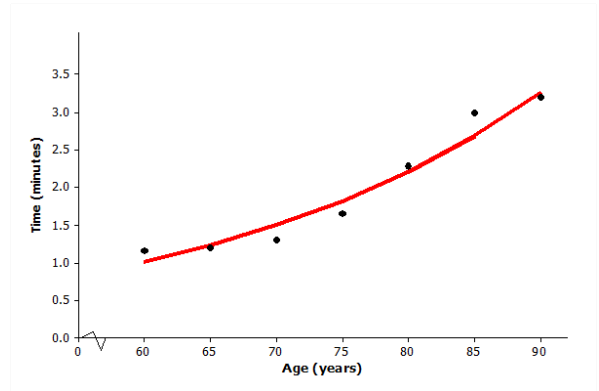
Suppose that social scientists conducted a study of senior citizens to see how the time (in minutes) required to solve a word puzzle changes with age. The scatter plot below displays data from this study.

Let x equal the age of the citizen and y equal the time (in minutes) required to solve a word puzzle for the seven study participants.



56.) What type of model (linear, quadratic, or exponential) would you use to describe the relationship between age and time required to complete the word puzzle?

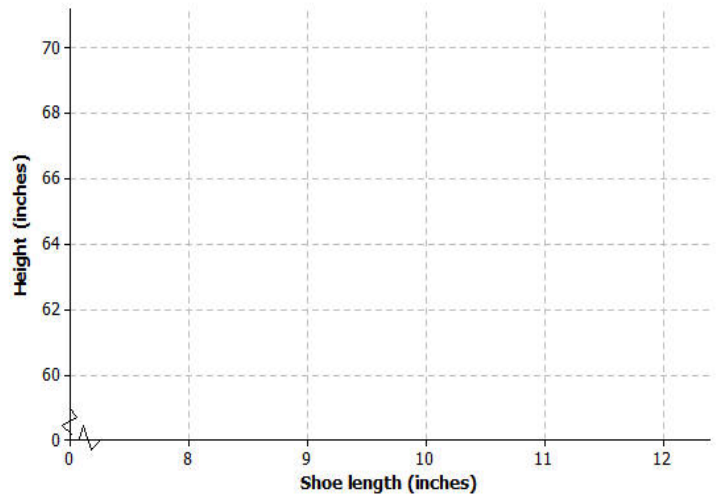
57.) One model that could describe the relationship between age and time to complete the word puzzle is $y = 10^{-1.01+0.017x}$. This exponential curve is shown on the scatter plot to the left. Does this model do a good job of describing the relationship between age and time to complete the word puzzle? Explain why or why not.



58.) Based on this exponential model, what time would you predict for a person who is 78 years old?

59.) Kendra wondered if the relationship between shoe length and height might be different for men and women. To investigate, she also collected data on shoe length (in inches) and height (in inches) for 12 women. Construct a scatter plot of these data.

Shoe Length (Women)	Height (Women)
8.9	61
9.6	61
9.8	66
10.0	64
10.2	64
10.4	65
10.6	65
10.6	67
10.5	66
10.8	67
11.0	67
11.8	70

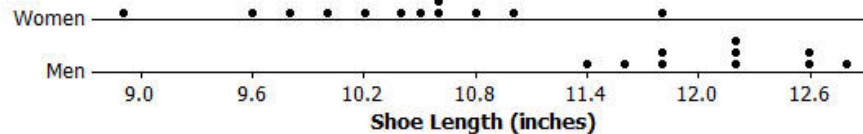


60.) Is there a relationship between shoe length and height for these 12 women?

61.) Find the equation of the least-squares line. (Round values to the nearest hundredth.)

62.) Suppose that these 12 women are representative of adult women in general. Based on the least-squares line, what would you predict for the height of a woman whose shoe length is 10.5 inches? What would you predict for the height of a woman whose shoe length is 11.5 inches?

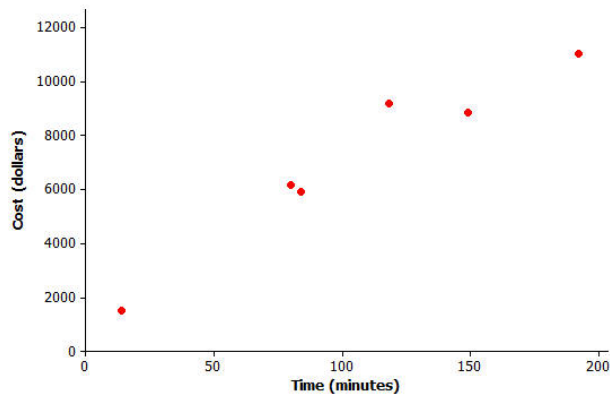
- 63.) One of the women in the sample had a shoe length of 9.8 inches. Based on the regression line, what would you predict for her height?
- 64.) Provide an interpretation of the slope of the least-squares line.
- 65.) Does it make sense to interpret the y-intercept of the least-squares line in this context? Explain why or why not.
- 66.) Below are dot plots of the shoe lengths for women and the shoe lengths for men. Suppose that you found a shoe print and that when you measured the shoe length, you got 10.8 inches. Do you think that a man or a woman left this shoe print? Explain your choice.



•HW 7.8 - Interpreting Residuals from a Line & More Modeling with a Line•

The time spent in surgery and the cost of surgery was recorded for six patients. The results and scatter plot are shown below.

Time (minutes)	Cost (\$)
14	1,510
80	6,178
84	5,912
118	9,184
149	8,855
192	11,023



- 67.) Calculate the equation of the least-squares line relating cost to time. (Indicate slope to the nearest tenth and y-intercept to the nearest whole number.)
- 68.) Draw the least-squares line on the graph above. (Hint: Substitute $x = 30$ into your equation to find the predicted y -value. Plot the point $(30, \text{your answer})$ on the graph. Then substitute $x = 180$ into the equation and plot the point. Join the two points with a straightedge.)

69.) What does the least-squares line predict for the cost of a surgery that lasts 118 minutes? (Calculate the cost to the nearest cent.)

70.) How much do you have to add to your answer to question (3) to get the actual cost of surgery for a surgery lasting 118 minutes? (This is the residual.)

71.) Show your answer to question (4) as a vertical line between the point for that person in the scatter plot and the least-squares line.

72.) Remember that the residual is the actual y -value minus the predicted y -value. Calculate the residual for the surgery that took 149 minutes and cost \$8,855.

73.) Calculate the other residuals, and write all the residuals in the table below.



Time (minutes)	Cost (\$)	Predicted value (\$)	Residual
14	1,510		
80	6,178		
84	5,912		
118	9,184		
149	8,855		
192	11,023		



74.) Suppose that a surgery took 100 minutes.

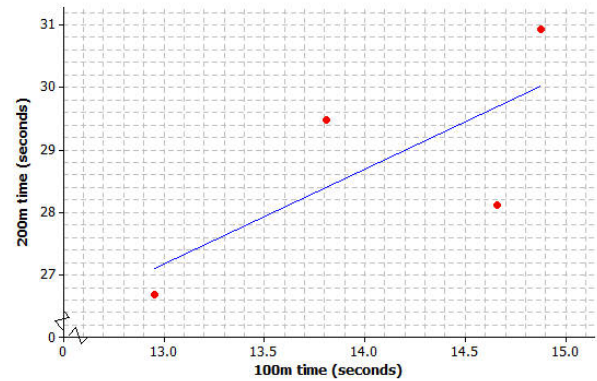
a.) What does the least-squares line predict for the cost of this surgery?

b.) Would you be surprised if the actual cost of this surgery were \$9000? Why or why not?

c.) Interpret the slope of the least-squares line.

75.) Four athletes on a track team are comparing their personal bests in the 100 meter and 200 meter events. A table of their best times and scatter plot of these results (including the least-squares line) is shown below.

Athlete	100m time (seconds)	200m time (seconds)
1	12.95	26.68
2	13.81	29.48
3	14.66	28.11
4	14.88	30.93

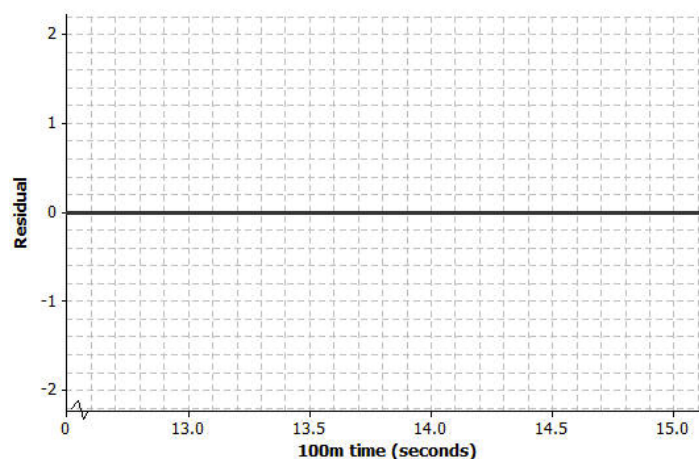


a.) Use your calculator to find the equation of the least-squares line. Round to the nearest thousandth.

b.) Calculate the residuals for the athletes. Write all the residuals in the table given below.

Athlete	100m time (seconds)	200m time (seconds)	Residual
1	12.95	26.68	
2	13.81	29.48	
3	14.66	28.11	
4	14.88	30.93	

c.) Using the axes provided below, construct a residual plot for this data set:



•HW 7.9 - Analyzing Residuals•

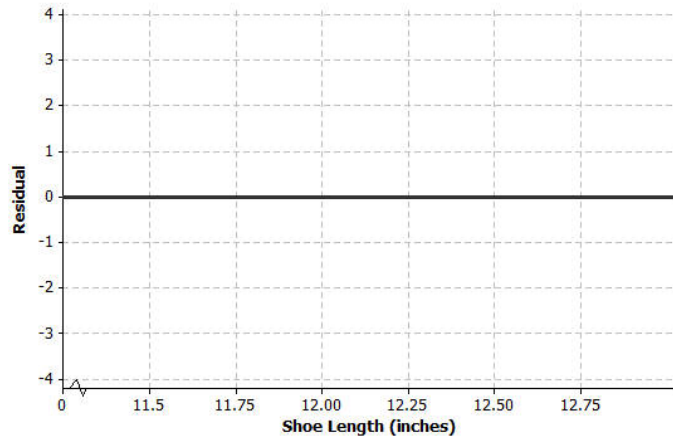
Consider a data set giving the shoe lengths and heights of 10 adult men. This data set is shown in the table below.

Shoe Length (x) inches	Height (y) inches
12.6	74
11.8	65
12.2	71
11.6	67
12.2	69
11.4	68
12.8	70
12.2	69
12.6	72
11.8	71

76.) Use your calculator to construct the scatter plot of this data set. Include the least-squares line on your graph. Explain what the slope of the least-squares line indicates about shoe length and height.

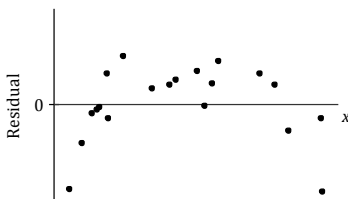
77.) Use your calculator to construct the residual plot for this data set.

78.) Make a sketch of the residual plot on the axes given below. Does the scatter of points in the residual plot indicate a linear relationship in the original data set? Explain your answer.

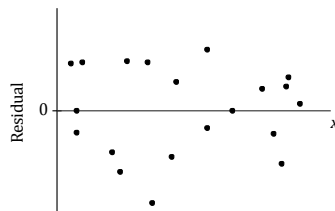


79.) For each of the following residual plots, what conclusion would you reach about the type of relationship (linear or non-linear) between the variables in the original data set? Justify your answer.

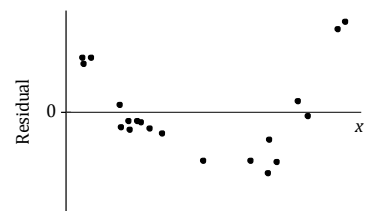
a.



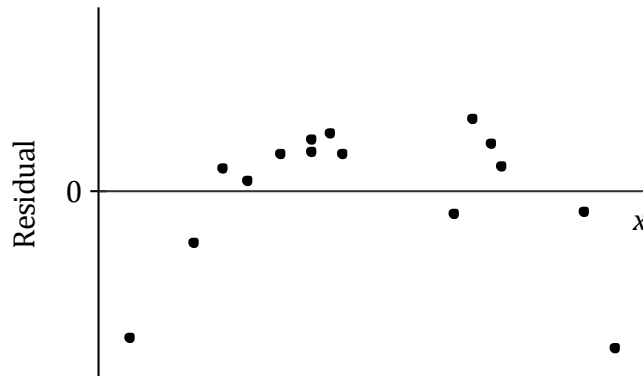
b.



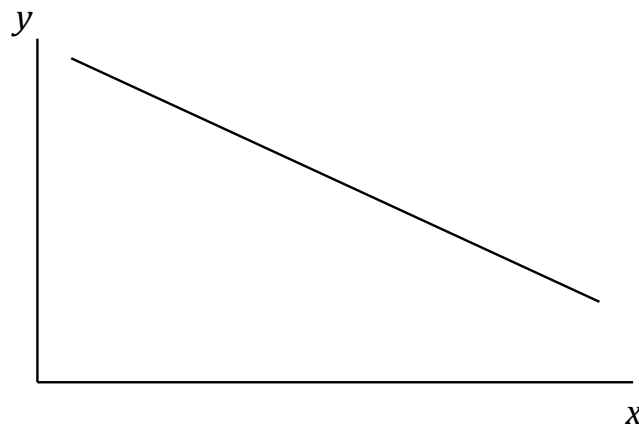
c.



80.) Suppose that after fitting a line, a data set produces the residual plot shown below.



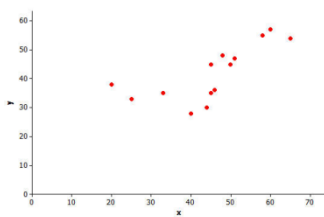
An incomplete scatter plot of the original data set is shown below. The least-squares line is shown, but the points in the scatter plot have been erased. Estimate the locations of the original points and create an approximation of the scatter plot below:



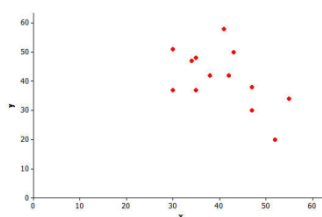
•HW 7-10: Interpreting Correlation and Analyzing Data•

81.) Which of the three scatter plots below shows the strongest linear relationship? Which shows the weakest linear relationship?

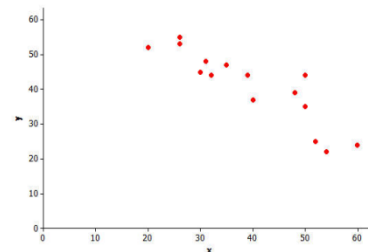
Scatter plot 1



Scatter plot 2



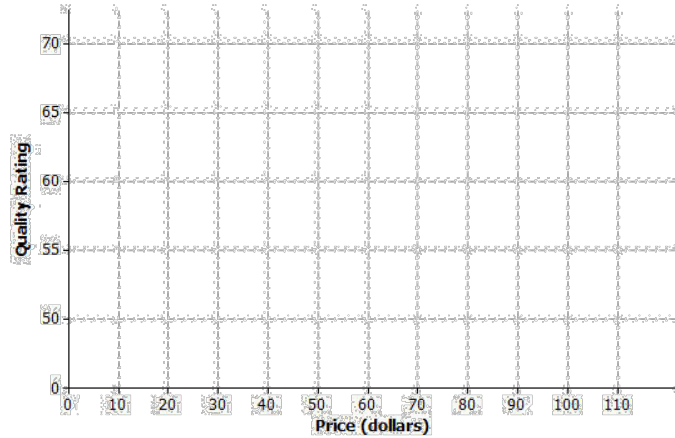
Scatter plot 3



82.) *Consumer Reports* published data on the price (in dollars) and quality rating (on a scale of 0 to 100) for 10 different brands of men's athletic shoes.

a.) Construct a scatter plot of these data using the following grid.

Price (\$)	Quality Rating
65	71
45	70
45	62
80	59
110	58
110	57
30	56
80	52
110	51
70	51



b.) Calculate the value of the correlation coefficient between price and quality rating and interpret this value. Round to the nearest hundredth.

c.) Does it surprise you that the value of the correlation coefficient is negative? Explain why or why not.



d.) Is it reasonable to conclude that higher priced shoes are higher quality? Explain.

e.) The correlation between price and quality rating is negative. Does this mean it is reasonable to conclude that increasing the price causes a decrease in quality rating? Explain.

- 83.) *The Princeton Review* publishes information about colleges and universities. The data below are for six public 4-year colleges in New York. Graduation rate is the percentage of students who graduate within six years. Student-to-faculty ratio is the number of students per full-time faculty member.

School	Number of Full-Time Students	Student-to-Faculty Ratio	Graduation Rate
CUNY Bernard M Baruch College	11,477	17	63
CUNY Brooklyn College	9,876	15.3	48
CUNY City College	10,047	13.1	40
SUNY at Albany	14,013	19.5	64
SUNY at Binghamton	13,031	20	77
SUNY College at Buffalo	9,398	14.1	47

- a.) Calculate the value of the correlation coefficient between graduation rate and number of full-time students. Round to the nearest hundredth.
- b.) Is the linear relationship between graduation rate and number of full-time students weak, moderate or strong? On what did you base your decision?
- c.) True or False? Based on the value of the correlation coefficient, it is reasonable to conclude that having a larger number of students at a school is the cause of a higher graduation rate.
- d.) Calculate the value of the correlation coefficient between graduation rate and student-to-faculty ratio. Round to the nearest hundredth.
- e.) Which linear relationship is stronger: graduation rate and number of full-time students or graduation rate and student-to-faculty ratio? Justify your choice.

•Scatterplots and Lines of Best Fit•

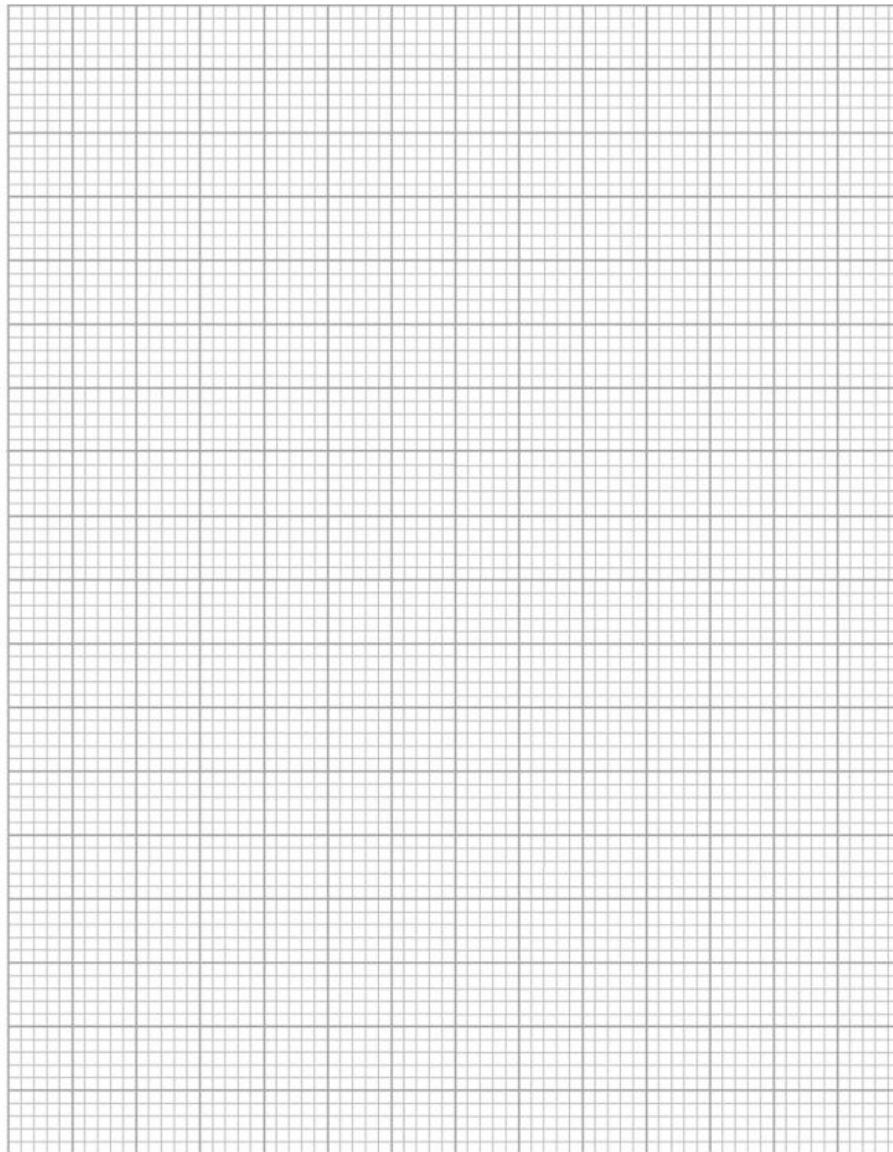
84.) The table below shows the bivariate data on the duration of weekly exercise, x , in hours, and their total cholesterol level, y , in milligrams per deciliter among a group of people.

Weekly Exercise (x hours)	5	4	6	7	1	3	4	1	8	3
Total Cholesterol Level (y mg/dL)	210	230	200	200	250	230	220	260	190	240

- a) Use the graph paper on the next page. Construct the scatter plot and sketch a line of best fit for the given table of bivariate data.
- b) Write an equation of the line of best fit.

c) Interpret the meaning of the slope and y-intercept in context.

85.) Identify the line of best fit that matches the data below.

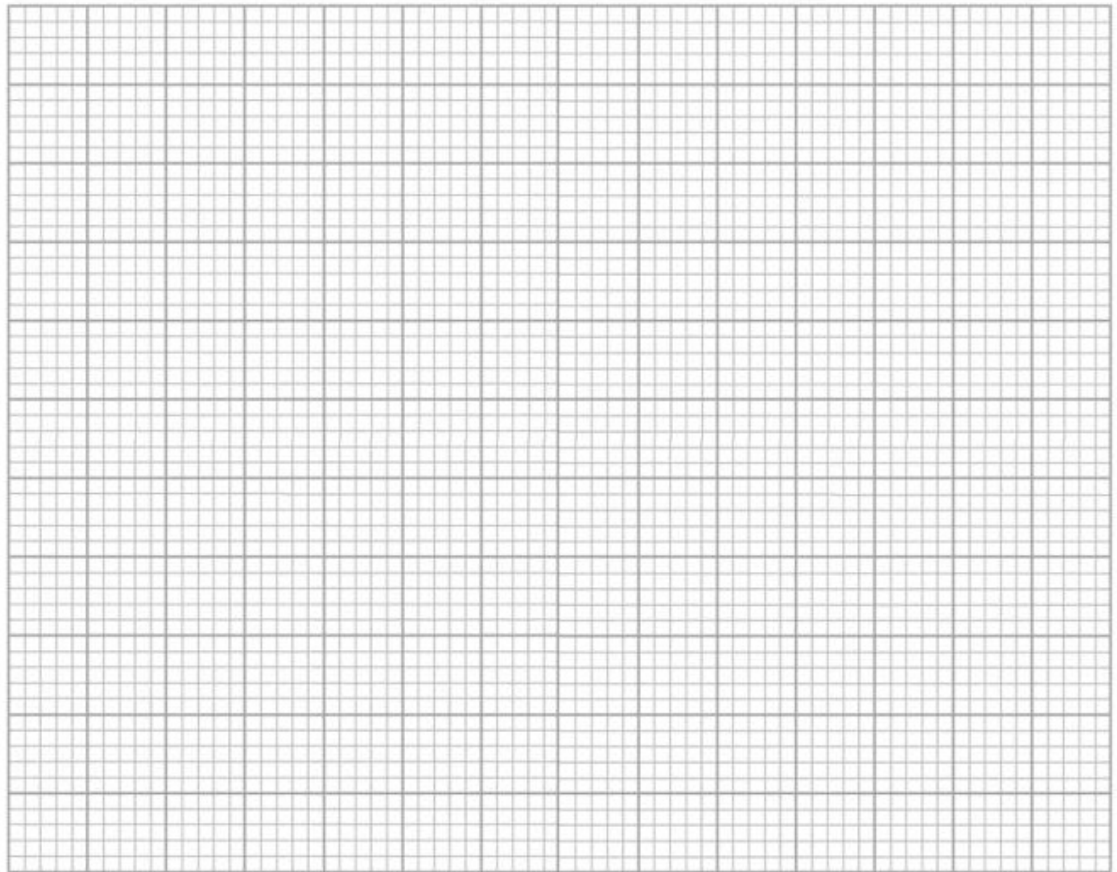


86.) Construct a scatter plot.

Use 1 centimeter on the horizontal axis to represent 1 item for the x interval from 50 to 60. Use 1 centimeter on the vertical axis to represent \$5 for the y interval from 10 to 50.

Number of Items Assembled (x)	53	59	57	50	54	56	51
Cost Per Item (y dollars)	40	25	30	50	38	33	48

Number of Items Assembled (x)	56	52	55	60	58	55
Cost Per Item (y dollars)	32	30	37	18	26	38



b.) Find and draw in the equation for the line of best fit.

c.) Use your equation to find the cost per item if the number of items assembled is 70.