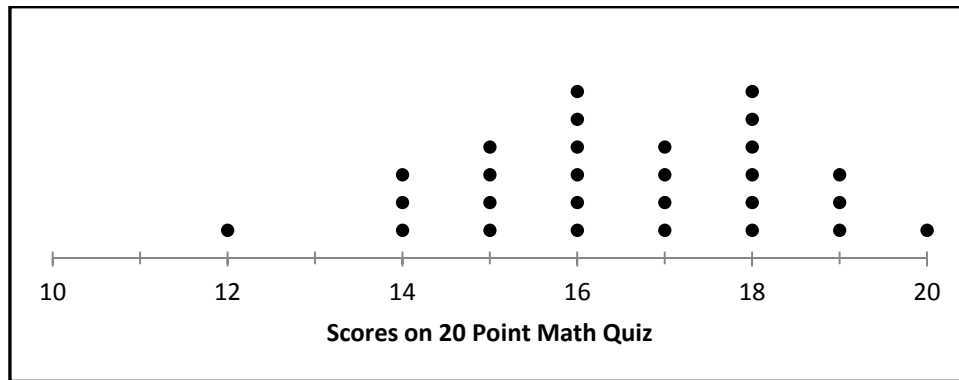


Name: \_\_\_\_\_

Date: \_\_\_\_\_

## UNIT #10 – STATISTICS REVIEW QUESTIONS

### Part I Questions



Ms. Wright's Algebra class consists of 28 students who recently took a 20 point quiz on factoring. The distribution of scores is shown on the dot plot above. Answer questions 1 through 4 based on this graph.

1. Which of the following is closest to the percent of students who scored less than a 16 on the quiz?

- (1) 8%                      (3) 29%  
(2) 50%                     (4) 44%

\_\_\_\_\_

2. Which of the following would be the median score of this data set?

- (1) 15.5                    (3) 16.5  
(2) 16                      (4) 17

\_\_\_\_\_

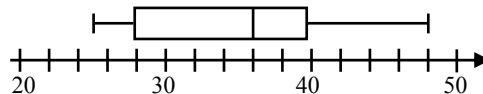
3. Which of the following represents the range of the scores?

- (1) 8                        (3) 12  
(2) 15                      (4) 20

\_\_\_\_\_

4. The high temperatures in March in Dutchess County, New York, are shown in the box plot shown below. Which of the following is the median high temperature?

- (1) 28                      (3) 40  
(2) 36                      (4) 48



\_\_\_\_\_



5. A local babysitting group takes care of 8 children on Saturday morning. The children's ages are shown below.

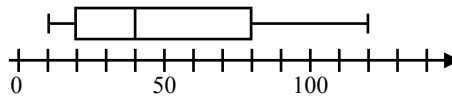
4, 7, 6, 8, 4, 5, 6, 4

Which of the following is the mean age of this group of children?

- (1) 5.5                      (3) 6.4  
 (2) 5.9                      (4) 6.7

6. Which of the following is the interquartile range of the data set shown in the box plot below?

- (1) 110                      (3) 55  
 (2) 60                        (4) 85



7. Which of the following data sets would have a standard deviation closest to zero?

- (1) {1, 3, 4, 6, 9}                      (3) {6, 6, 6, 7, 7}  
 (2) {15, 18, 20, 25}                      (4) {1, 1, 3, 3, 7}

8. A sample of ten large eggs were weighed with the results shown below, in grams.

58, 64, 61, 56, 60, 57, 60, 62, 59, 68

Which of the following is this sample's standard deviation to the nearest tenth of a gram?

- (1) 2.7                      (3) 4.8  
 (2) 3.5                      (4) 5.1

9. A survey of 52 graduating seniors was conducted to determine if there was a connection between the gender of the student and whether they were going on to college. Which of the following represents the proportion of those going to college who are male?

- (1) 0.53                      (3) 0.55  
 (2) 0.31                      (4) 0.16

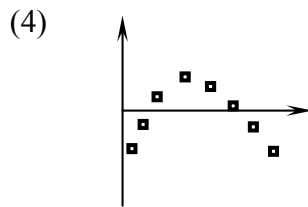
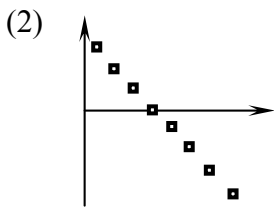
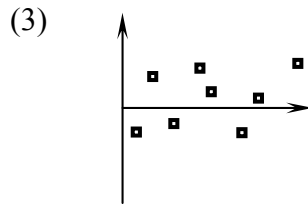
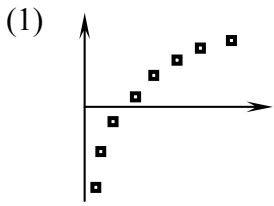
	Gender		Total
	Male	Female	
Going to College	16	13	29
Not Going to College	14	9	23
Total	30	22	52

10. Which of the following correlation coefficient values below indicates the greatest correlation between the two variables?

- (1)  $r = 0.78$                       (3)  $r = 0.12$   
 (2)  $r = -0.54$                       (4)  $r = -0.97$



11. Four different types of regression were used on a data set. Which of the following four residual plots shows the model that is the most appropriate to use?



\_\_\_\_\_

12. Economists create a linear regression equation to predict the price of a gallon of gasoline,  $y$ , based on the price of a barrel of oil,  $x$ . The equation they find is  $y = 0.035x + 0.95$ . One of the data points they use is  $(60, 3.62)$ . What is the residual for this data point?

(1) 0.57

(3) 0.72

(2) 0.68

(4) 0.84

\_\_\_\_\_

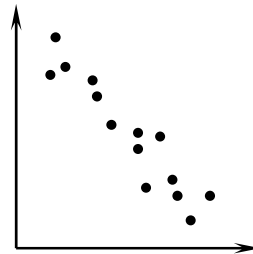
13. Which correlation coefficient is most appropriate for the scatterplot shown below?

(1)  $r = -1.00$

(2)  $r = -0.92$

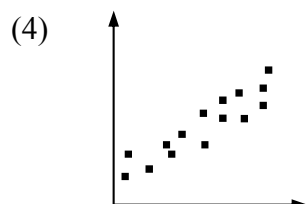
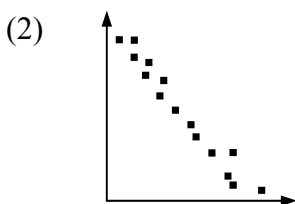
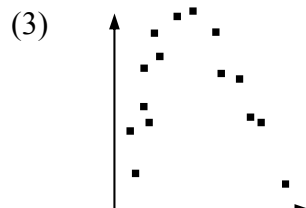
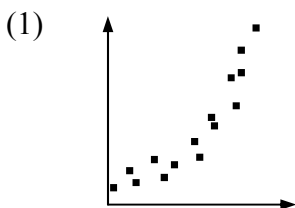
(3)  $r = 0.78$

(4)  $r = 1.00$



\_\_\_\_\_

14. Which of the following scatter plots most likely indicates a quadratic regression equation is most appropriate for the data?

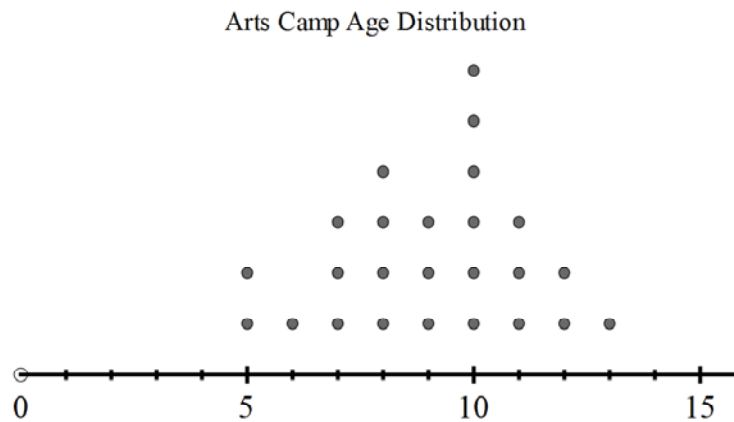


\_\_\_\_\_

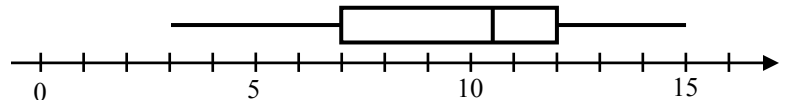


## Free Response Questions

15. A local summer art camp admits kids from ages 5 to age 13. In the summer of 2015, the camp had the following distribution of ages for its 25 participants:



- (a) What percent of the kids attending are less than 8 years old?
- (b) Which is greater, the mean or the median of this data set? By how much?
- (c) What is the interquartile range of this data set? Show how you calculated your answer.
- (d) Sometimes data sets are broken into **quintiles**, which are groups that have one-fifth, or 20%, of the data set in them. Which ages would the upper quintile contain? Explain how you determined your answer.
- (e) The box plot below shows the distribution of ages at a farm camp. Which of the two camps has the greater range of ages? Support your answer.



Age Distribution at Farm Camp



16. A marketing company did a survey of 40 people with ages from 12 to 20 and broke them into the three age categories shown below. Each person tried three colas and specified which their favorite cola was.

		Cola Type			Total
		Cola A	Cola B	Cola C	
Age Range	12 to 14	3	6	1	10
	15 to 17	1	8	4	13
	18 to 20	2	8	7	17
Total		6	22	12	40

- (a) What are the relative frequencies of a person choosing each of the following as their favorite?

Cola A

Cola B

Cola C

- (b) Is a person in the 15 to 17 age range more likely or less likely to choose Cola B as their favorite compared to all those surveyed? Justify your answer.

- (c) Which is more likely: (A) that a person who chooses Cola C as their favorite is in the 18 to 20 age bracket or (B) that a person in the 18 to 20 age bracket chooses Cola C as their favorite? Justify your answer.

- (d) The company will only stock vending machines at the mall with a particular soda if more than 25% of at least one of the age groups chooses it as their favorite. Will any of the three colas be excluded based on this criteria? Justify your choice.



17. A company is tracking the number of times an post to a popular social internet site has been shared. The data is shown below.

Hours, $x$	1	2	3	4	5	6
Shares, $y$	28	44	49	65	88	100

- (a) Determine a linear regression model and an exponential regression model that best fit these data. Round all parameters to the nearest tenth.

Linear,  $y = ax + b$

Exponential,  $y = a(b)^x$

- (b) The correlation coefficient for both models, when rounded to the nearest hundredth, is  $r = 0.99$ . What does this value tell you about the predictive ability of these models?

- (c) Does your linear model from part (a) over predict or under predict the number of shares at 6 hours? Show the work that leads to your answer.

- (d) After 24 hours, the post had been shared 10,952 times. Which model does a better job extrapolating the data trend out to 24 hours? Justify your response.



18. A linear regression was done to relate the age of a person,  $x$ , and the number of optimal hours of sleep,  $y$ . The equation was found to be  $y = -0.12x + 11.5$  with a correlation coefficient of  $r = -0.86$ .

(a) Explain what each of the parameters in the linear model represent, i.e. the  $-0.12$  and the  $11.5$ .

(b) How do you interpret the fact that the correlation coefficient is negative?

19. Severe flu cases are increasing in a local hospital. The number of reported cases is shown over the span of a week in the table below.

Day, $x$	1	2	3	4	5	6	7
Flu Cases, $y$	13	19	24	27	30	32	34

(a) Determine a linear regression model for this data set. Round all parameters to the nearest tenth.

(b) Create a sketch of the residuals below. Discuss whether this plot indicates that a linear model is appropriate or not to fit this data.

