Review for Test #7 - Descriptive Statistics
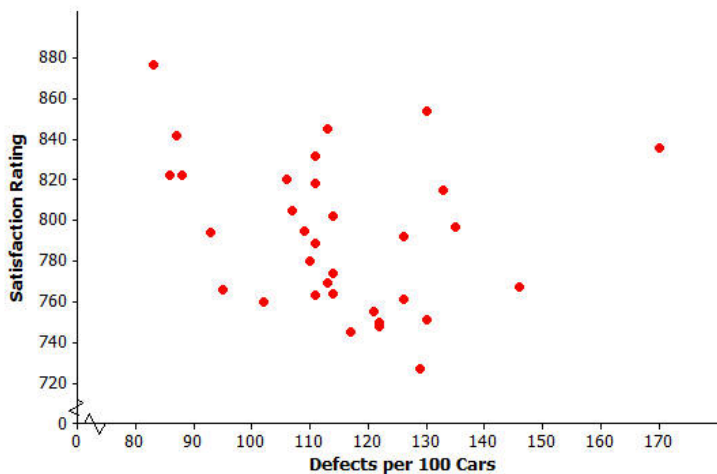
Name _____ Class _____

1.) The scatterplot 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 could be the value of the correlation coefficient for this data set?

    I. r = -0.95   II. r = -0.24   III. r = 0.83   IV. r = 1.00

  b.) Explain why you selected this value.

2.) Which table does not show bivariate numerical data?

(A)

Height (inches)	Weight (pounds)
39	50
48	70
60	90

(B)

Gallons	Miles Driven
15	300
20	400
25	500

(C)

Quiz Average	Frequency
70	12
80	15
90	6

(D)

Speed (mph)	Distance (miles)
40	80
50	120
55	150

3.) Which value of *r* represents data with a strong positive linear correlation between the two variables?

(A) 0.89   (B) 0.34   (C) 1.04   (D) 0.01

4.)   Which graph represents data used in a linear regression that produces a correlation coefficient closest to -1?

(A)

(C)

(B)

(D)
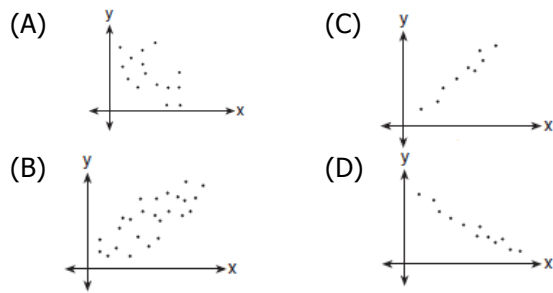
5.)   The box-and-whisker plot below represents the results of test scores in a math class.

```
   0    20   40   60   80   100
```
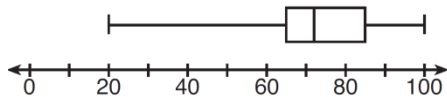
What do the scores 65, 85, and 100 represent?

(A) Q_1, median, Q_3 (B) Q_1, Q_3, maximum   (C) median, Q_1, maximum   (D) minimum, median, maximum

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

Shoe Length (x)	Height (y)
inches	inches
12.6	73
11.8	66
12.2	70
11.6	68
12.2	68
11.4	65
12.8	72
12.2	70
12.6	71
11.8	70

a.)   Use your calculator to construct the scatter plot of this data set. Include the least-squares line on your graph.

b.)   Explain what the slope of the least-squares line indicates about shoe length and height.

c.)   Use your calculator to construct the residual plot for this data set. Explain what information the residual plot tells you about the data.
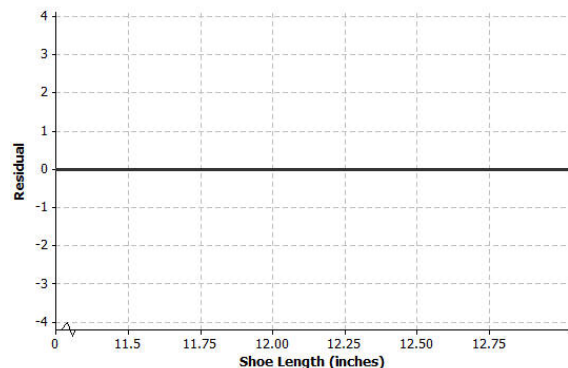
d.)   Calculate the correlation coefficient for this data and interpret this value.

7.) Consider the following results from 100 randomly selected students asked to describe their favorite color:

- Of the 30 female students selected, 10 of them said red, 2 picked blue, 9 picked purple, and the rest of them picked orange.
- Of the male students, 20 picked blue, 15 picked red, 30 picked orange, and the rest picked purple.

a.) Complete the above table for the 100 students who were surveyed on their favorite color.

	Red	Blue	Purple	Orange	Total
Female					
Male					
Total					

b.) Calculate the **relative frequencies** for each of the cells to the nearest hundredth. Place the relative frequencies in the cells of the following table.

	Red	Blue	Purple	Orange	Total
Female					
Male					
Total					

c.) Calculate the **column conditional relative** frequencies for each of the cells to the nearest hundredth. Place the row relative frequencies in the cells.

	Red	Blue	Purple	Orange	Total
Female					
Male					
Total					

d.) What conclusions can be drawn about the association between gender and favorite color? Justify your conclusion using the given data.

8.) Fill in the chart below:

	Best Measure of Center	Best Measure of Variability
Symmetrical Distribution		
Skewed Distribution		

9.)   Here is the scatterplot of age (in years) and finish time (in minutes) of the NYC Marathon.
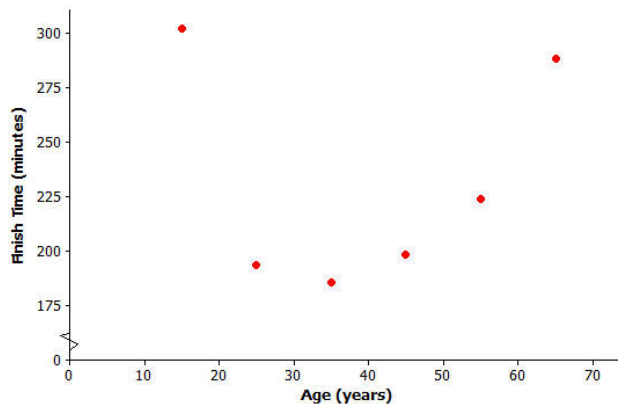

a.)   What type of model (linear, quadratic, or exponential) would best describe the relationship between age
       and finish time? Explain your reasoning.

b.)   According to the data, at what age would the **fastest** time occur?

10.) Describe the difference between univariate and bivariate data. Provide an example for each.

11.) Describe the difference between qualitative (categorical) and quantitative (numerical) data. Provide an example
      for each.

12.) Describe the difference between correlation and causation.

13.) A list is arranged in order from least to greatest. The list is 25, 30, 44, x, 52, 60, 75, 80. If the median is 50,
      what does x equal?

14.) Simplify: $\sqrt{32}$

15.) The following lists the players' heights (in inches) on the Rochester Rhinos Soccer Team. If necessary, round to the nearest hundredth.

67	68	68	68	68	69	69	70	70	70	70	70	70
70	70	71	71	72	72	72	73	73	74	75	75	76

a.) Find the mean: _____

b.) Find the standard deviation: _____

c.) Find the mode: _____

d.) Find the minimum value: _____

e.) Find the lower quartile: _____

f.) Find the median: _____

g.) Find the upper quartile: _____

h.) Find the maximum value: _____

i.) Find the IQR: _____

j.) Draw a box plot:
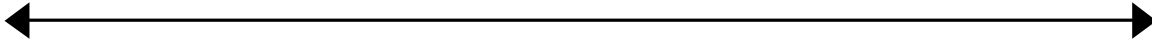
16.) The table below shows the ages of eight cows (in years) and their milk production (in gallons) per week.

a.) Create a linear regression model in your calculator.

b.) Which type of correlation exists for this data?

    i.   Positive linear correlation

    ii. Negative linear correlation

    iii. No correlation

c.) Write the linear regression equation below. Round to the nearest hundredth.

d.) Interpret the slope of the regression line.

e.) What is the y-intercept telling us?

    i.   We can predict cows of age 3 to produce over 40 gallons of milk per week.

    ii. We can predict cows of age 1 to produce over 40 gallons of milk per week.

    iii. It is telling us nothing. Cows of age 0 do not produce milk.

f.) Find the correlation coefficient, rounded to the nearest hundredth. Interpret this value.

g.) Calculate the residual for a cow of age 8, rounded to the nearest hundredth. Interpret this value.

The test results of 66 students were compared with whether or not the students completed a review sheet provided by their teacher prior to taking the test. A conditional relative frequency table was prepared.

	Passed Test	Failed Test	Totals
Completed Review Sheet			51
Did not complete review sheet		10	
Totals	55		

17.) Fill in the missing values in the chart above.

18.) What are the column conditional relative (joint) frequencies, from top to bottom, for the category passed test?

19.) What percentage of students who failed the test did not complete the review sheet?

20.) Fill in the row conditional relative joint frequencies. Round each value to the nearest hundredth.

	Passed Test	Failed Test	Totals
Completed Review Sheet			
Did not complete review sheet			

21.) Explain whether the table in question #20 indicates a statistical association between completing the review sheet and passing the test.

22.) If an association between completing the review sheet and passing the test exists, would it imply that completing a review sheet will improve test scores for **all** students everywhere? Explain.