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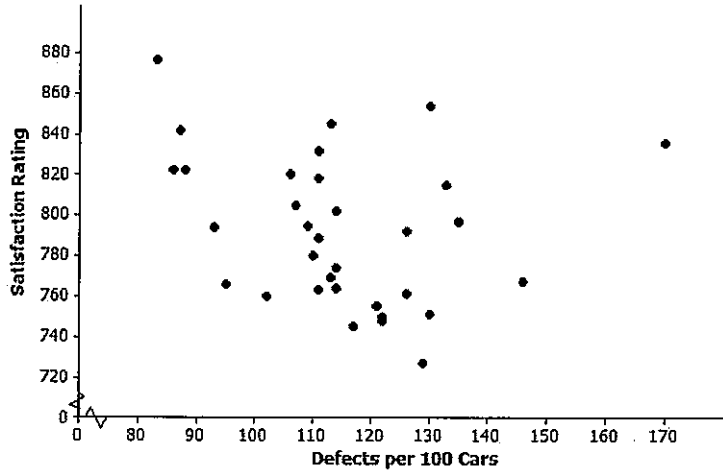
Data Source: *USA Today,* June 16, 2010 and July 17, 2010

   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.

      The trend is generally negative and the correlation is weak.

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

   (A)

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

   (B)

Gallons X	Miles Driven Y
15	300
20	400
25	500

   (C) *only 1 variable*

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 *almost no correlation*

   b/t 0.7 and 1.   *weak correlation*   *too high*

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

(A) *graph with y and x axes, scattered downward points*

(C) *graph with y and x axes, upward curving points*

(B) *graph with y and x axes, upward scattered points*

(D) *graph with y and x axes, downward curving points* — circled
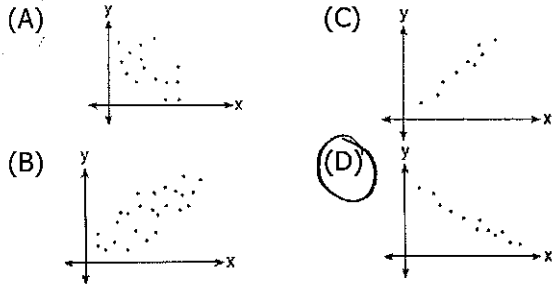
strongest
one that's going down

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

box-and-whisker plot labeled min, Q, med, Q₃, max on number line from 0 to 100
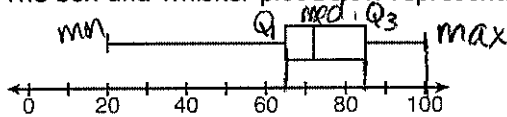
What do the scores 65, 85, and 100 represent?

(A) Q_1, median, Q_3 (B) Q_1, Q_3, maximum — circled   (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. (Thousandth)

$$y = 4.669x + 12.707$$

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

For every additional inch of shoe length, the height increases by 4.669"

c.) Use your calculator to construct the residual plot for this data set. Explain what information the residual plot tells you about the data.

The residual plot indicates that there is a linear relationship because it is scattered.

residual plot with Residual on y-axis (from -4 to 4) and Shoe Length (inches) on x-axis (from 11.5 to 12.75)
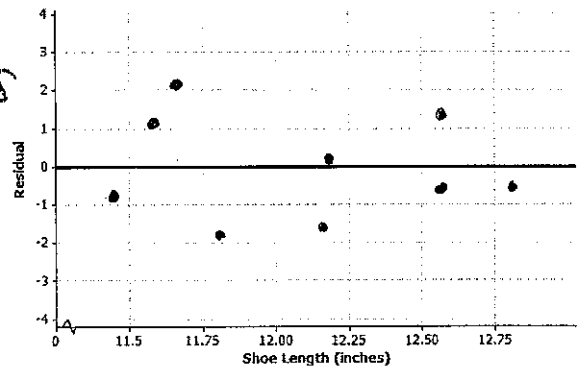
d.) Calculate the correlation coefficient for this data and interpret this value.

$r \approx 0.85$ (rounded to hundredth— you may have rounded differently

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	10	2	9	9	30
Male	15	20	5	30	70
Total	25	22	14	39	100

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	$\frac{10}{100}=0.10$	$\frac{2}{100}=0.02$	$\frac{9}{100}=0.09$	$\frac{9}{100}=0.09$	$\frac{30}{100}=0.30$
Male	$\frac{15}{100}=0.15$	$\frac{20}{100}=0.20$	$\frac{5}{100}=0.05$	$\frac{30}{100}=0.30$	$\frac{70}{100}=0.70$
Total	$\frac{25}{100}=0.25$	$\frac{22}{100}=0.22$	$\frac{14}{100}=0.14$	$\frac{39}{100}=0.39$	$\frac{100}{100}=1.00$

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	$\frac{10}{25}=0.40$	$\frac{2}{22}\approx0.09$	$\frac{9}{14}\approx0.64$	$\frac{9}{39}\approx0.23$	
Male	$\frac{15}{25}=0.60$	$\frac{20}{22}\approx0.91$	$\frac{5}{14}\approx0.36$	$\frac{30}{39}\approx0.77$	
Total	$\frac{25}{25}=1$	$\frac{22}{22}=1.00$	$\frac{14}{14}=1.00$	$\frac{39}{39}=1.00$	

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

There is an association between gender and favorite color because the conditional relative frequencies are different among each gender.

8.) Fill in the chart below:

	Best Measure of Center	Best Measure of Variability
Symmetrical Distribution	mean	standard deviation
Skewed Distribution	median	IQR

9.) Here is the scatterplot of age (in years) and finish time (in minutes) of the NYC Marathon.
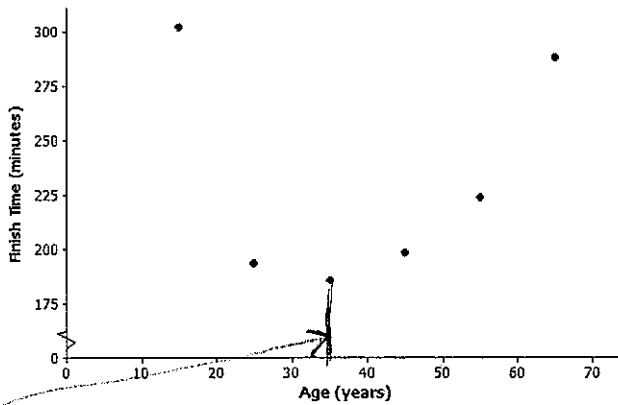
a.) What type of model (linear, quadratic, or exponential) would best describe the relationship between age and finish time? Explain your reasoning.

quadratic — It's shaped as a parabola

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

(lowest time)   35 yrs.

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

univariate - 1 variable -ex: quiz grades
bivariate - 2 variables - ex: quiz grades vs. study time

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

qualitative - descriptive data - ex: eye color
quantitative - numerical data - ex: height

12.) Describe the difference between correlation and causation.

correlation - a relationship b/t 2 variables ⊛ correlation ≠ causation
causation - when one variable makes another variable happe[n]

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?

25 30 44 | x 52 | 60 75 80

median falls halfway between these 2 values

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

32
1 | 32
4 | 8
16 | 2

$\sqrt{32}$

$\sqrt{16}\sqrt{2}$

$4\sqrt{2}$

$\dfrac{x+52}{2} = 50$

$x + 52 = 100$

$x = 48$

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: __70.81__

b.) Find the standard deviation: __2.40__

c.) Find the mode: __70__ —(occurs most often)

d.) Find the minimum value: __67__

e.) Find the lower quartile: __69__

f.) Find the median: __70__

g.) Find the upper quartile: __72__

h.) Find the maximum value: __76__

i.) Find the IQR: __3__ (72-69)

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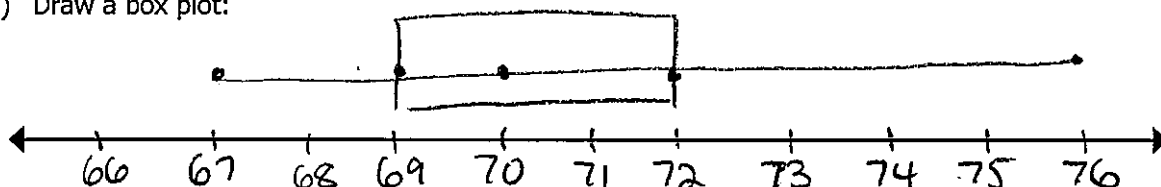
16.) The table below shows the ages of eight cows (in years) and their milk production (in gallons) per week.

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 ⟵ (circled)

    iii. No correlation

Cow's Age	Gallons of Milk per Week
4	37.2
5	35.2
6	33.6
7	33.2
7	32.8
8	33.1
9	31.6
10	29.8

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

$$y = -1.07x + 40.79$$

d.) Interpret the slope of the regression line.

$$m = \frac{\Delta y}{\Delta x} = \frac{\Delta gall}{\Delta age} = \frac{-1.07}{1}$$ → When a cow grows by 1 year in age, it produces 1.07 _less_ gallons of milk each week.

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. ⟵ (circled)

y-int. says initial amt.

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

$$r \approx -0.96$$ There is a strong neg correlation b/t age and milk production

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

✱ actual - predicted

Actual = 33.1

predicted ⇒ y = -1.07(8) + 40.79 ≈ 32.23

33.1 - 32.23 = 0.87

The test results of ⓺⓺ 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	50	1	51
Did not complete review sheet	5	10	15
Totals	55	11	66

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?

$$\frac{50}{55} \approx 0.91 \qquad \frac{5}{55} \approx 0.09$$

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

$$\frac{10}{11} \approx 0.91 = 91\%$$

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

	Passed Test	Failed Test	Totals
Completed Review Sheet	$\frac{50}{51} \approx 0.98$	$\frac{1}{51} \approx 0.02$	$\frac{51}{51} = 1.00$
Did not complete review sheet	$\frac{5}{15} \approx 0.33$	$\frac{10}{15} \approx 0.67$	$\frac{15}{15} = 1.00$

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

yes — there is a major difference b/t the conditional relative frequencies.

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.

No — just because there is a correlation doesn't mean there's a cause-and-effect relationship.

*secret bonus:
Write "I'm awesome" on the front of your review sheet. You are awesome, because you followed directions and checked your work.