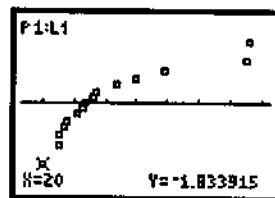


1. Which of the following statements are true?
- I. Regression outliers, if removed, cause a dramatic change in the slope and orientation of the least squares line.
  - II. Determining a conditional distribution for a two-way table consists of calculating percents for either the row or column sums.
  - III. An influential observation can have a strong effect on both the regression line and the correlation between the X and Y variables.
- (a) I and II only  
(b) I and III only  
(c) II and III only  
(d) I, II, and III  
(e) None of the above. The answer is \_\_\_\_\_.

2. The plot shown is a normal probability plot for a set of data. The data value is plotted on the x-axis, and the standardized value is plotted on the y-axis. Which statement is true for this data set?



- (a) The data are clearly normally distributed.  
(b) The data are approximately normally distributed.  
(c) The data are clearly skewed to the right.  
(d) The data are clearly skewed to the left.  
(e) There is insufficient information to determine the shape of the distribution.
3. Which of the following are true statements?
- I. The area under a normal curve is always 1, regardless of the mean and standard deviation.
  - II. The mean is always equal to the median for any normal distribution.
  - III. The interquartile range for any normal curve extends from  $\mu - 1s$  to  $\mu + 1s$ .
- (a) I and II  
(b) I and III  
(c) II and III  
(d) I, II, and III  
(e) None of the above gives the correct set of true responses.
4. A researcher finds that the correlation between the personality traits "greed" and "superciliousness" is  $-0.40$ . What percentage of the variation in greed can be explained by the relationship with superciliousness?
- (a) 0%  
(b) 16%  
(c) 20%  
(d) 40%  
(e) 60%

A political scientist believes that there is a "gender gap" in American voting with women more likely to vote for the Democratic candidate. She therefore interviews a random sample of voters and records the gender of the respondents and the political party of the candidates for whom they voted in the last presidential election. Identify the following variables:

5. Quantitative: *none*
6. Categorical: *gender & political party*
7. Explanatory: *gender*
8. Response: *political party*

9. Scores on the Wechsler Adult Intelligence Scale (a standard IQ test) are approximately normally distributed within age groups. For the 20–34 age group, the mean is 110 and the standard deviation is 25. For the 60–64 age groups, the mean is 90 and the standard deviation is 25. Sarah is 29 and her mother is 62. Sarah scores 135 on the Wechsler test, while Ann scores 120. Who has the better score, relative to her age group?

Ann has a better score since her z-score is higher and so she is in a higher percentile.

*Sarah*

$$z = 1.0$$

*84th percentile*

*Ann*

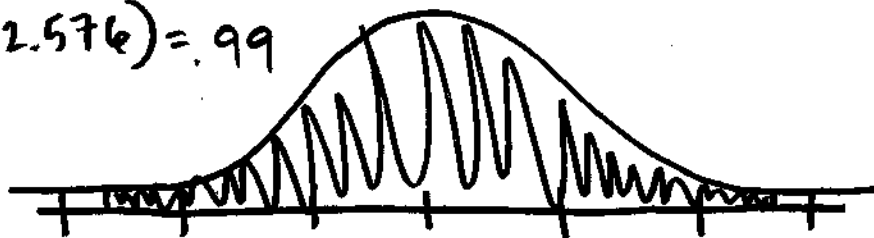
$$z = 1.2$$

*88th percentile*

10. In a normally distributed population, what percent of the population observations lie within 2.576 standard deviations of the mean? Include a sketch to illustrate your answer.

$$\text{normalcdf}(-2.576, 2.576) = .99$$

*99%*



11. The Graduate Record Examinations are widely used to help predict the performance of applicants to graduate schools. The range of possible scores on a GRE is 200 to 900. The psychology department finds that the scores of its applicants on the quantitative GRE are approximately normal with mean 544 and standard deviation 103. What minimum score would a student need in order to score in the top 10% of those taking the test?

$$\text{invNorm}(.9, 544, 103) = 676$$

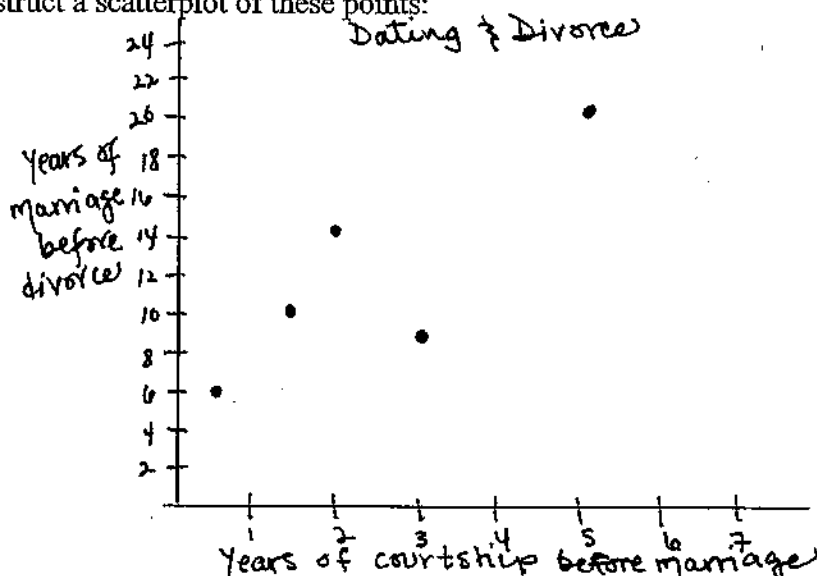
A certain psychologist counsels people who are getting divorced. A random sample of six of her patients provided the following data where

$x$  = number of years of courtship before marriage, and

$y$  = number of years of marriage before divorce.

$x$	3	0.5	2	1.5	5
$y$	9	6	14	10	20

12. Construct a scatterplot of these points:



13. Calculate the least-squares regression line (LSRL). Interpret the slope and y-intercept.

$$\hat{y} = 5.359 + 2.6838x$$

$x \rightarrow$  years of courtship before marriage  
 $y \rightarrow$  years of marriage before divorce

The slope of 2.683 means that for each year of courtship increasing the years of marriage before divorce will increase 2.683 years on average. The y-intercept of 5.359 means that a couple who courts for 0 years before marriage is predicted to be married 5.359 years before getting divorced.

14. What is the correlation between  $x$  and  $y$ ? Interpret this number.

$$r = .849$$

There is a fairly strong, positive, linear relationship between years of courtship before marriage and years of marriage before divorce.

15. Show how the residual for the first data point in the table is calculated.

$$\text{Residual} = y - \hat{y}$$

$$(3, 9)$$

$$\hat{y} = 5.359 + 2.6838(3) = 13.4104$$

$$9 - 13.4104 = -4.4104$$

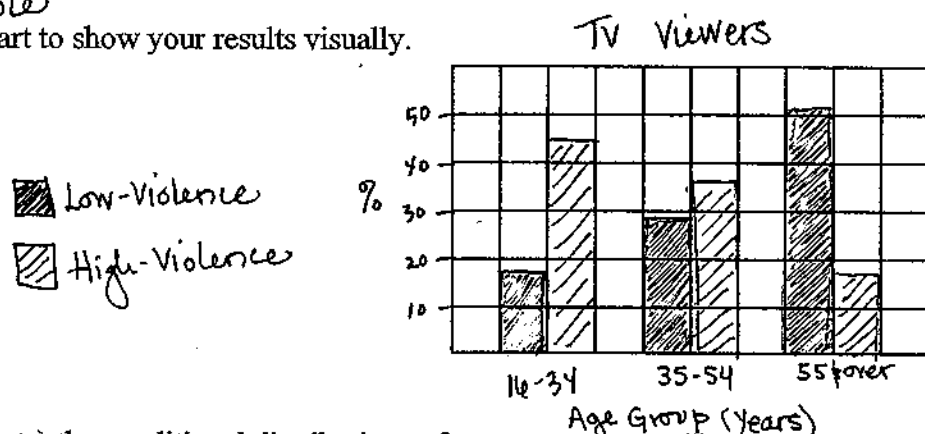
In a study of the relationship between the amount of violence a person watches on TV and the viewer's age, 81 regular TV watchers were randomly selected and classified according to their age group and whether they were a "low-violence" or "high violence" viewer. Here is a two-way table of the results.

		Age Group			
		16-34	35-54	55 & over	
Amount of Violence Watched	Low	8	12	21	41
	High	18	15	7	40
		32%	33%	35%	

16. Compute (in percents) the marginal distribution of age group for all people surveyed.

*in table*

17. Construct a bar chart to show your results visually.



18. Compute (in percents) the conditional distributions of age group among "low-violence" viewers. Then do the same for "high-violence" viewers.

*Low-Violence Viewers:*

$$16-34 \rightarrow 8/41 = 19.5\%$$

$$35-54 \rightarrow 12/41 = 29.3\%$$

$$55 \frac{1}{2} \text{ over} \rightarrow 21/41 = 51.2\%$$

*High-Violence Viewers:*

$$16-34 \rightarrow 18/40 = 45\%$$

$$35-54 \rightarrow 15/40 = 37.5\%$$

$$55 \frac{1}{2} \text{ over} \rightarrow 7/40 = 17.5\%$$