Name: _____   **Statistics Review Packet**

Statistics Vocabulary Review:

> **Shapes of data distributions – skewed vs. symmetrical**

> **Centers of distributions – mean vs. median**

> **Univariate data**

> **Box Plots – quartiles**

> **Histograms**

> **Measures of Variability – standard deviation, interquartile range**

> **2-way tables – frequency, relative frequency & conditional relative frequency**

> **Bivariate data (qualitative vs. quantitative)**

> **Scatterplot**

> **Line of best fit (linear regression, least-squares line)**

> **Residuals (and residual plot)**

> **Correlation (& correlation coefficient) vs Causation**

Statistics Review

Multiple Choice:
Identify the choice that best completes the statement or answers the question.

_____ **1.** A floral delivery company conducts a study to measure the effect of worker experience on productivity. Tell whether the scatter plot appears to have a linear or non-linear pattern of association. Describe any clustering and identify outliers.
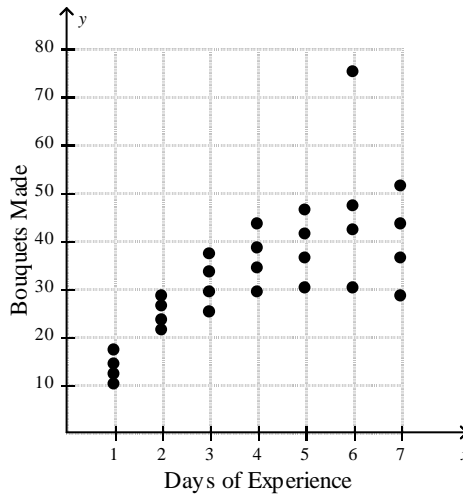
a. The pattern of association appears to be linear.
There appears to be clustering of the data points at 1 and 2 days. After that, the results become less clustered.
There do not appear to be any outliers.

b. The pattern of association appears to be non-linear.
There appears to be clustering of the data points at 6 and 7 days. Before that, the results are less clustered.
There do not appear to be any outliers.

c. The pattern of association appears to be non-linear.
There appears to be clustering of the data points at 1 and 2 days. After that, the results become less clustered.
The point near (6, 75) appears to be an outlier.

d. The pattern of association appears to be linear.
There appears to be clustering of the data points at 1 and 2 days. After that, the results become less clustered.
The point near (6, 75) appears to be an outlier.

2. 25 males were selected at random from a database to determine a leadership score.

Leadership Score (Males)
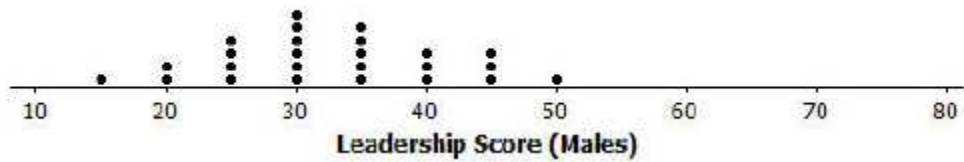
Which of the following best describes the distribution of the data?

a. The distribution is symmetric.

b. The distribution is skewed with the tail to the left.

c. The distribution is skewed with the tail to the right.

d. The distribution is uniform.

3. Which table does *not* show bivariate 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

_____ **4.** Find an equation in slope-intercept form for the line of best fit, and tell what the slope and intercepts represent in terms of the data it models. Give the slope and intercept to the nearest integer.

Cost of Family Vacation
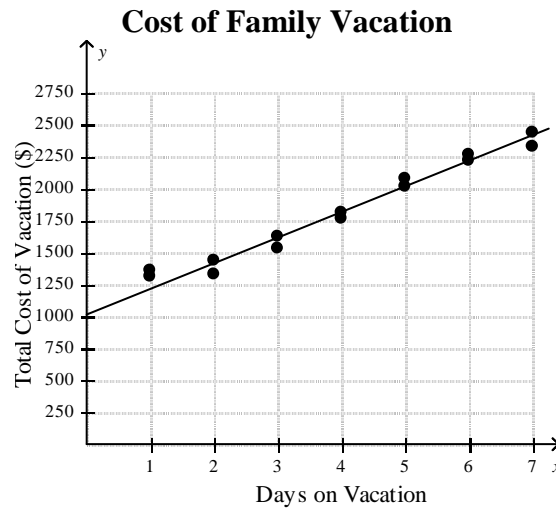
a. The slope of the best-fit line is 200, and the y-intercept is 1000.
The slope, $200 per day, is the typical daily cost, for instance, hotel and meal expenses.
The y-intercept, $1000, does not depend on the number of days the vacation lasts. It is a one-time cost, such as air fare.

b. The slope of the best-fit line is 1000, and the y-intercept is 200.
The slope, $1000 per day, is the typical daily cost; for instance, hotel and meal expenses.
The y-intercept, $200, does not depend on the number of days the vacation lasts. It is a one-time cost, such as air fare.

c. The slope of the best-fit line is 1000, and the y-intercept is 200.
The slope, $1000, does not depend on the number of days the vacation lasts. It is a one-time cost, such as air fare.
The y-intercept, $200 per day, is the typical daily cost; for instance, hotel and meal expenses.

d. The slope of the best-fit line is 200, and the y-intercept is 1000.
The slope, $200, does not depend on the number of days the vacation lasts. It is a one-time cost, such as air fare.
The y-intercept, $1000 per day, is the typical daily cost; for instance, hotel and meal expenses.

4

_____ **5.** What 5-year interval of ages represented in the 2010 histogram of the Kenyan age distribution has the most people?

Histogram of Ages for Kenya

Histogram of Ages for U.S.
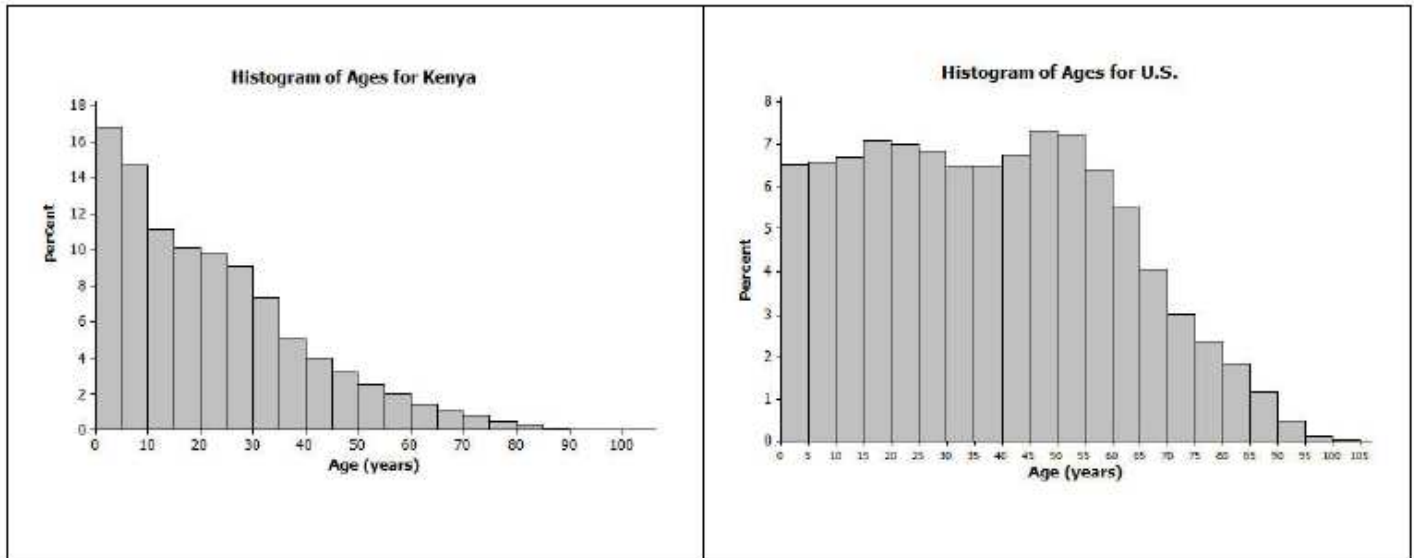
a. 0-5 years old

b. 15-20 years old

c. 50-55 years old

d. 45-50 years old

_____ **6.** The table shows the number of first, second, and third place finishes by members of two teams at a track meet. Of the Panthers, what is the relative frequency who placed first?

	First	Second	Third	Total
Tigers	12	9	8	29
Panthers	8	22	10	40
Total	20	31	18	

a. 0.2

b. 0.7

c. 0.3

d. 0.8

5

_____ **7.** Which relationship can best be described as causal?

a. height and intelligence

b. number of correct answers on a test and test score

c. shoe size and running speed

d. number of students in a class and number of students with brown hair

_____ **8.** A sample of 12 snowboard prices (in dollars) is shown below.

345   375   356   360   405   350   386   343   402   395   370   392

What is the standard deviation to the nearest hundredth?

a. 22.49   **b.** 373.25

c. 21.53   **d.** 4,479

_____ **9.** The freshman class held a canned food drive for 12 weeks. The results are summarized in the table below.

Canned Food Drive Results

Week	1	2	3	4	5	6	7	8	9	10	11	12
Number of Cans	20	35	32	45	58	46	28	23	31	79	65	62

Which number represents the interquartile range of the number of cans of food collected?

a. 30.5   **b.** 29.5

c. 59   **d.** 6

6

_____ **10.** A movie theater recorded the number of tickets sold daily for a popular movie during the month of June. The box-and-whisker plot shown below represents the data for the number of tickets sold, in hundreds.
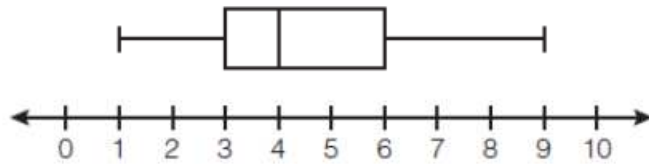
Which conclusion can be made using this plot?

a. The second quartile is 600.

b. The mean of the attendance is 400.

c. The range of the attendance is 300 to 600.

d Twenty-five percent of the attendance is between 300 and 400.

11. What is the correlation coefficient of the linear fit of the data shown below, to the *nearest hundredth*?

(1) 1.00   (3) −0.93
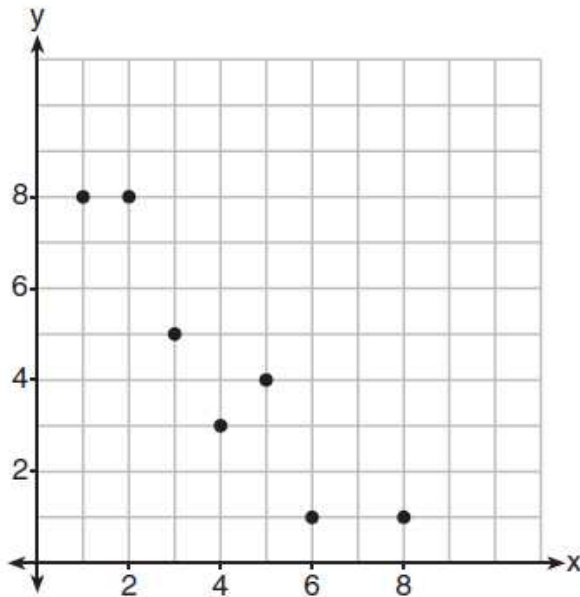
(2) 0.93   (4) −1.00

12. Christopher looked at his quiz scores shown below for the first and second semester of his Algebra class.

Semester 1: 78, 91, 88, 83, 94
Semester 2: 91, 96, 80, 77, 88, 85, 92

Which statement about Christopher's performance is correct?

(1) The interquartile range for semester 1 is greater than the interquartile range for semester 2.

(2) The median score for semester 1 is greater than the median score for semester 2.

(3) The mean score for semester 2 is greater than the mean score for semester 1.

(4) The third quartile for semester 2 is greater than the third quartile for semester 1.

13. Isaiah collects data from two different companies, each with four employees. The results of the study, based on each worker's age and salary, are listed in the tables below.

Company 1

Worker's Age in Years	Salary in Dollars
25	30,000
27	32,000
28	35,000
33	38,000

Company 2

Worker's Age in Years	Salary in Dollars
25	29,000
28	35,500
29	37,000
31	65,000

Which statement is true about these data?

(1) The median salaries in both companies are greater than $37,000.

(2) The mean salary in company 1 is greater than the mean salary in company 2.

(3) The salary range in company 2 is greater than the salary range in company 1.

(4) The mean age of workers at company 1 is greater than the mean age of workers at company 2.

8

14. The table below shows the number of grams of carbohydrates, x, and the number of Calories, y, of six different foods.

Carbohydrates (x)	Calories (y)
8	120
9.5	138
10	147
6	88
7	108
4	62

Which equation best represents the line of best fit for this set of data?
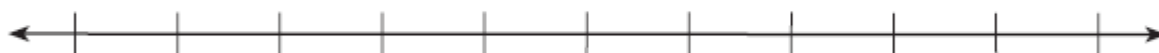
(1) $y = 15x$

(2) $y = 0.07x$

(3) $y = 0.1x - 0.4$

(4) $y = 14.1x + 5.8$

Short Answer

15. Robin collected data on the number of hours she watched television on Sunday through Thursday nights for a period of 3 weeks. The data are shown in the table below.

|	Sun	Mon	Tues	Wed	Thurs
Week 1	4	3	3.5	2	2
Week 2	4.5	5	2.5	3	1.5
Week 3	4	3	1	1.5	2.5

Using an appropriate scale on the number line below, construct a box plot for the 15 values.

16. The table shows the relationship between the time a student spends working out each week and his percent improvement on race times.

Hours Spent Working Out	6	8	10	12	14	16	18
Percent Improvement	18	18	32	27	31	39	37

a) Make a scatter plot for the data.

Percent
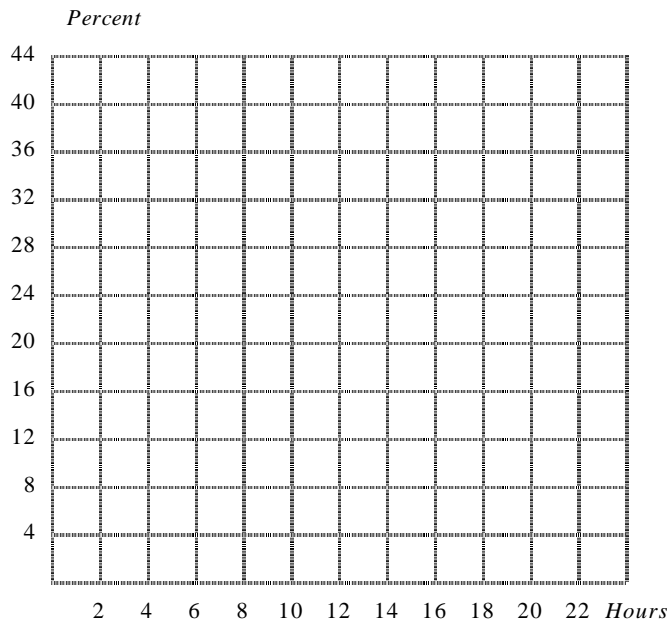
b) Use the statistical features of your calculator to fit a linear function to the data. Calculate and interpret the correlation coefficient (round to the nearest thousandth).

c) Use your equation to predict the number of hours the student would be expected to work out if his percent improvement is 50% (round to the nearest hour).

10

17. Fifty moviegoers were surveyed about their favorite movie types.

- 15 men and 6 women chose "Action" as their favorite type
- 9 men and 10 women chose "Drama" as their favorite type
- 6 men and 4 women chose "Comedy" as their favorite type

**a)** Use the table below to construct a two-way frequency table.

Favorite Movie Types				
	Action	**Drama**	**Comedy**	**Total**
Men				
Women				
Total				

b) Find the relative frequencies to compare and describe the survey.

Favorite Movie Types				
	Action	**Drama**	**Comedy**	**Total**
Men				
Women				
Total				

c) Compare and describe (minimum of 3 statements):

18. Transportation officials collected data on sixty flight delays in the month of December and sixty flight delays in the month of January.

Dot Plot of December Delay Times

Delay Time (minutes)

Delay Time (January)
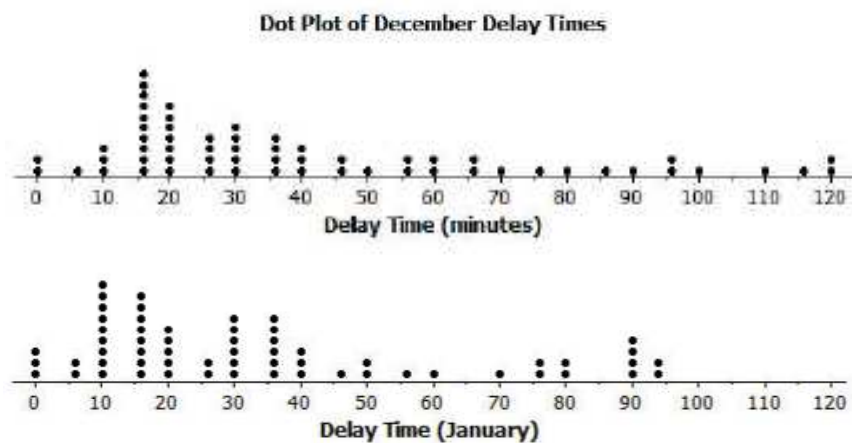
Construct a box plot for each month:


Delay Time (minutes)
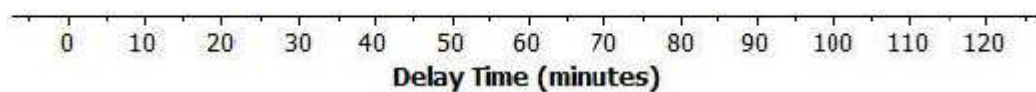
How is the January flight delay distribution different from the December flight delay distribution? Justify your response.

19. Twenty-five students were surveyed about the number of days they played outside in one month. The results of this survey are shown below.

3, 3, 4, 4, 4, 4, 5, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 16, 17, 22, 22, 25

a. On the grid below, create a histogram based on the data.
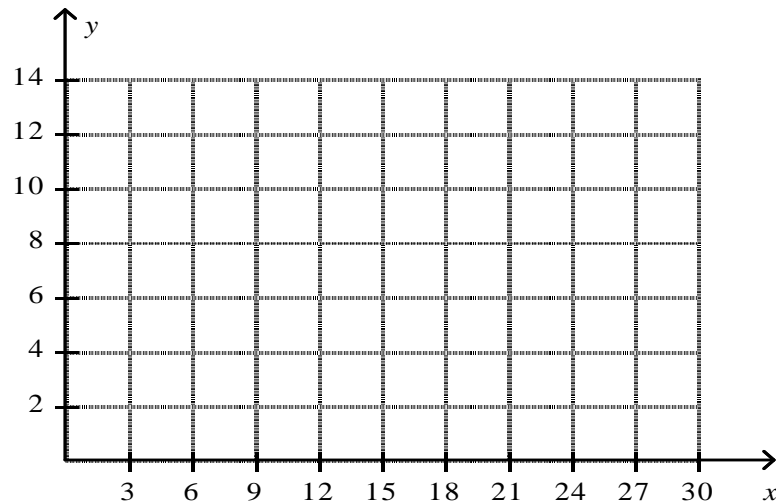
b. Identify the typical number of days spent outside by the twenty-five students.

c. Use the statistical features of your calculator to find the standard deviation of the data set (round to the nearest hundredth).

Standard Deviation: _____

Statistics Review ANSWERS

1) C	2) A	3) C	4) A	5) A
6) A	7) B	8) A	9) A	10) D
11) 3	12) 3	13) 3	14) 4	15) Min = 1, Q1 = 2, Med = 3, Q3 = 4, Max = 5

16. b) y = 1.75x + 7.86 $r \approx 0.901$ It's a linear, strong, positive correlation.

16. c) $50 = 1.75x + 7.86$ $24.08 = x$ Approximately 24 hours.

17. a)

Favorite Movie Types

	Action	Drama	Comedy	Total
Men	15	9	6	30
Women	6	10	4	20
Total	21	19	10	50

b) Find the relative frequencies to compare and describe the survey.

Favorite Movie Types

|	Action	Drama	Comedy	Total
Men	15/50 = 30%	9/50 = 18%	6/50 = 12%	30/50 = 60%
Women	6/50 = 12%	10/50 = 20%	4/50 = 8%	20/50 = 40%
Total	21/50 = 42%	19/50 = 38%	10/50 = 20%	100%

c) **Examples**:

Less than 10% of the movie goers were women whose favorite movie type was comedy.
30% of the movie goers were men whose favorite type was action.
80% of the movie goers had either action or drama as their favorite.

18. December: Min = 0, Q1 = 16, median = 30, Q3 = 60, max = 120
     January: Min = 0, Q1 = 16, median = 30, Q3 = 48, max = 94
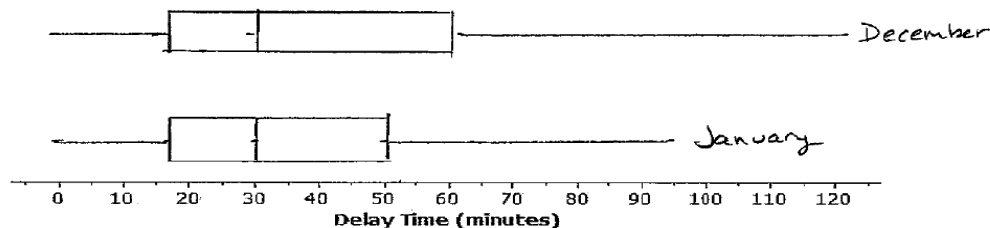
The bottom 50% is identical for both sets of data.
The upper 50% for December has a higher variability than for January.
The range for December is 120 while the range for January is 94.

19. b. 6 because it is the median and falls within the most frequent interval.
19. c. sx = 6.38