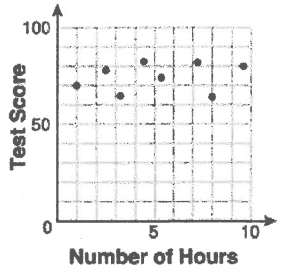
Algebra 1 – Statistics
Statistics Review 2

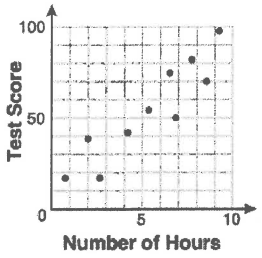
Name: Key

1. The four scatter plots below show the relationship between the number of hours of television watched the night before a test and the test score. Which of the scatter plots would have an r-value closest to -1? Strong negative

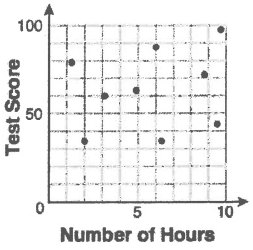
Ans: 4



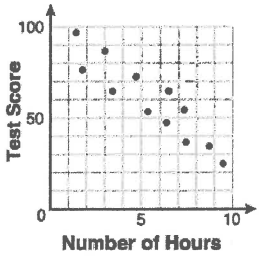
(1)



(2)



(3)



(4)

Answer questions 2 – 4 based on the two-way table below that compares gender to animal preference from your class survey data:

		Animal preference		
		Cat	Dog	Totals
Gender	Male	5	13	18
	Female	2	12	14
	Totals	7	25	32

2. What percent of male students prefer cats? 3
   (1) 5.0%
   (2) 15.6%
   (3) 27.8%
   (4) 71.4%

   of 18 males, 5 prefer cats: $\frac{5}{18} = 28\%$

3. What percent of the students prefer dogs? 4
   (1) 25.0%
   (2) 37.5%
   (3) 40.6%
   (4) 78.1%

   of all students (32), 25 prefer dogs: $\frac{25}{32} = 78.1\%$

4. What percent of dog-preferring students are female? 3
   (1) 12.0%
   (2) 37.5%
   (3) 48.0%
   (4) 85.7%

   of 25 students who prefer dogs, 12 are female
   $12/25 = 48\%$

5. The line of best fit for a particular model has an r-value of 0.87. How would you describe the correlation of the two variables describes by this model? 2
   (1) Weak positive
   (2) Strong positive
   (3) Strong Negative
   (4) No Correlation

   0.87 is close to 1, so is on the stronger side.

6. A principal is trying to figure out if the school should offer more art classes or not by asking a random sample of people what they think. Which is the best source of an unbiased sample of people?
(1) All of the yearbook editors
(2) Every tenth student leaving the cafeteria
(3) A randomly selected art class
(4) All of the teachers in the math department

2

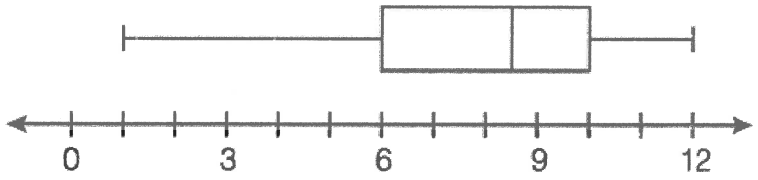
7. What is the mean and sample standard deviation of the following heights?
69, 61, 66, 64, 60, 71, 73, 64

(1) Mean = 66 & SD = 4.4
(2) Mean = 65 & SD = 4.7
(3) Mean = 66 & SD = 4.7
(4) Mean = 65 & SD = 4.4

Use the calculator
1-var stats
$\bar{x} = 66$
$S_x = 4.66 \rightarrow 4.7$ — sample sd
$\sigma_x = 4.4$ — population sd

3

8. Which of the following is *true* about the box plot below?



I. The distribution is skewed to the right. Left
II. The mean is 8.5. median = 8.5
III. Twenty-five percent of the data falls between 1 and 6.

(1) I, only (2) III, only (3) I and II (4) I and III

2

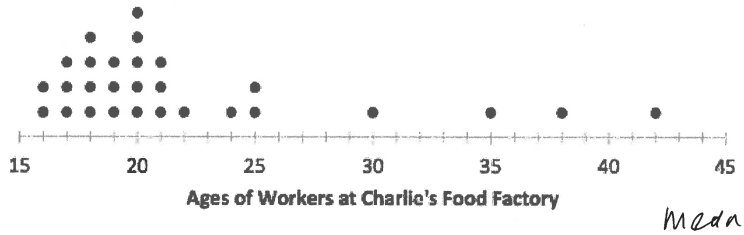
9. What is the IQR for these values, which represent the number of singles a sample of artists has had in the top 10?
17, 11, 2, 2, 7, 0, 2, 1, 0, 2

Use the calculator: 1-var stats
$IQR = Q_3 - Q_1 = 7 - 1 = 6$

(1) 2 (2) 6 (3) 7 (4) 17

2

10. Which value would be most appropriate to describe the variation of this set of data?



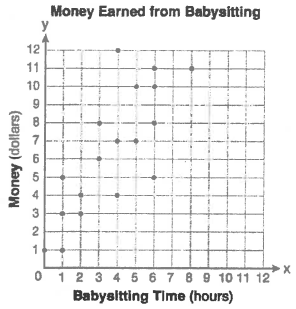
Ages of Workers at Charlie's Food Factory

(1) Mean = 22
(2) SD = 6.59
(3) IQR = 5
(4) Median = 20

Mean and median are measures of center, not variation. Since dotplot is skewed to the left, the SD should not be used.

11. Which r-value would be most appropriate for the scatter plot below?

2



positive and fairly strong.

(1) 0.10 (2) 0.78 (3) -0.92 (4) -0.1

1

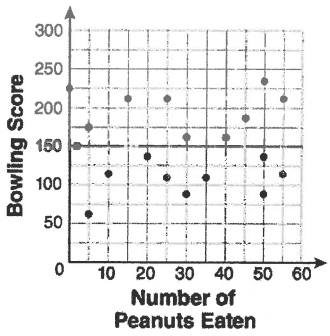
12. Consider the line of best fit for the relationship shown in the above scatter plot, between babysitting time and money earned below, where x is the amount of time babysitting and y is the amount of money earned. Which is the best interpretation of the slope of this line?

$$y = 1.19x + 2$$

Change in money earned per hour.

(1) The amount of money earned increases at a rate of \$1.19 per hour.
(2) The amount of money earned increases at a rate of \$2 per hour.
(3) The amount of time babysitting increases at a rate of 2 hours per dollar
(4) The amount of time babysitting increases at a rate of 1.19 hours per dollar.

13. The scatter plot below represents the relationship between the number of peanuts a student eats and the student's bowling score.



very scattered, mostly horizontal.
No relationship.

1

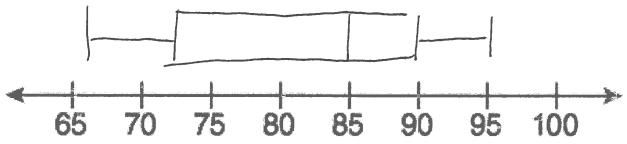
What conclusion about the scatter plot is valid?
(1) There is no relationship between eating peanuts and bowling score.
(2) Students who eat more peanuts have higher bowling scores.
(3) Students who eat more peanuts have lower bowling scores.
(4) No bowlers eat peanuts.

14. The test scores from Mr. Smith's Algebra 1 class are shown below.

72, 73, 66, 71, 82, 85, 95, 85, 86, 89, 91, 92

use calculator: 1 var stats

a) Construct a box plot to display this data.



min = 66
Q_1 = 72.5
Med = 85
Q_3 = 90
Max = 95

b) Is the distribution symmetric or skewed? If it is skewed, in which direction is it skewed?

Skewed to the left (left of center is longer than right of center)

c) Based on your answer to part (b), which measure of center/spread should be used to describe these data: mean/SD or median/IQR?

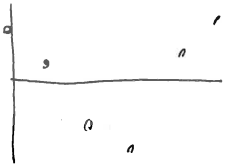
Median/IQR (see 10)

15. Ron creates a linear model to describe the relationship between time and a particular insect population, shown in the table below. Use this to answer questions 15-17.

Time (days)	0	1	3	4	6	7
Population (in thousands)	25	27	31	34	45	49
predicted population	23	26.48	33.44	36.92	43.88	47.36
Residuals = Actual - predicted	2	0.52	-2.44	-2.92	1.12	1.64

a) Given the linear model, $y = 3.48x + 23$, find the residuals and put these in the table.

b) Draw a sketch of the residuals.



c) Is the linear model appropriate? Explain why or why not.

No. The residuals show a curved pattern.

16. Use the following information to answer questions (a) and (b). Two police stations collected data to find the relationship between the number of cars on the road in one day and the number of cars that were speeding on that road that day. The Pleasantville Police Department and the Little Red Police Department both found a line of best fit to represent their data.

Pleasantville PD: $y = 0.32x + 0.2$; $r = 0.95$ Little Red PD: $y = 0.35x - 10.3$; $r = 0.97$

a) Explain which model you would use to predict the number of speeders on a road and why.

I would use Little Red PD's model. It has an r-value of 0.97 which suggests a stronger relationship.

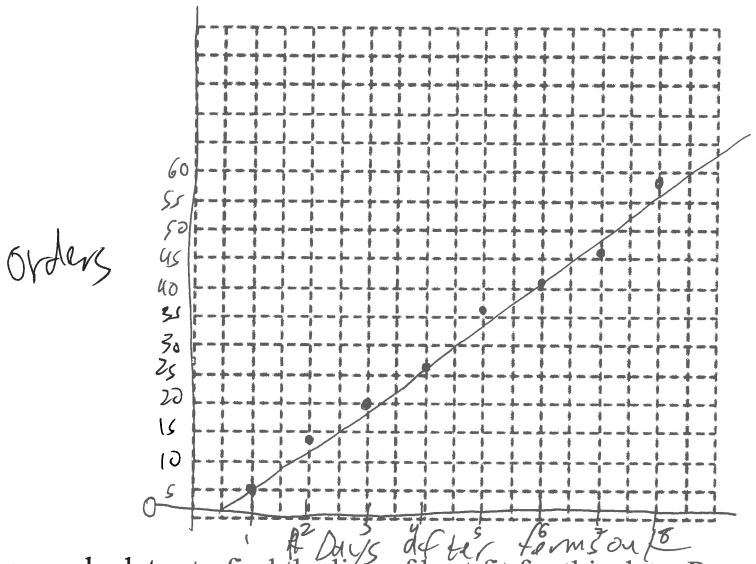
b) Use the model you chose to predict the number of cars speeding on this road on a day when 670 cars drove down the road.

$y = 0.35(670) - 10.3 = 224.2$ predicted ≈ 224 speeders

17. The yearbook club is keeping track of how many yearbook orders are submitted after the day that the order forms were passed out. The data is shown in the table below.

Days after Order Forms were Passed Out (x)	1	2	3	4	5	6	7	8
Yearbook Orders (y)	5	14	20	26	36	41	46	58

a) Using the number of days since the order forms were passed out as the independent variable, construct a scatter plot of these data below. Be sure to label your axes and scale. Draw a rough sketch of the line of best fit after you plot your data.



b) Use your calculator to find the line of best fit for this data. Round all coefficients to the nearest tenth.

$y = 7.2x - 1.6$

c) What does the y-intercept of your equation mean in terms of the number of days and ~~views~~ orders?

On day zero, there were −1.6 orders.
(This is meaningless!)

d) What does the slope of your equation mean in terms of the number of days and views?

Orders are increasing by about 7.2 per day.

e) How many orders placed does your model predict for 17 days after the order forms were passed out?

$y = 7.2(17) - 1.6 = 120.8$

≈ 121 orders.