Announcements

- Statistics test tomorrow
- Test corrections due Wednesday
- Last unit - Geometry - begins on Wednesday
- May 30th - Math 1 Exam
- May 27th from 2-5 - Panera Study Day

Math 1 Statistics Review

4/22/2019

NC.M1.S-ID.1 Use technology to represent data with plots on the real number line (histograms, and box plots).

Example: The table below shows the length of a class period for each of the schools listed in a NC school district. Choose and create an appropriate plot to represent the data. Explain your choice of plot.

School	Class period (minutes)	School	Class period (minutes)
Lincoln Middle	45	New Hope Middle	55
Central Middle	65	Sunnyside Middle	50
Oak Grove Middle	70	Pine Grove Middle	60
Fairview Middle	55	Green Middle	65
Jefferson Middle	60	Hope Middle	55
Roosevelt Middle	60		

Example: The following data set shows the number of songs downloaded in one week by each student in Mrs. Jones class: 10, 20, 12, 14, 12, 27, 88, 2, 7, 30, 16, 16, 32, 25, 15, 4, 0, 15, 6, 1, 0, 15, 12, 10, and 7.

a. What are the summary statistics for the data?
b. Construct two different graphs of the data.
c. Describe the distribution of the data, citing both of the plots and the numerical summary statistics.
d. What are the advantages to each data display? Explain.

NC.M1.S-ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets. Interpret differences in shape, center, and spread in the context of the data sets.

The choices below are data sets. In the choices, w is a constant. Each choice has the same mean. Which choice has the greatest standard deviation?

A $w - 2, w - 1, w, w + 1, w + 2$

B $w - 2, w - 2, w, w + 2, w + 2$

C $w - 3, w - 1, w, w + 1, w + 3$

D $w - 3, w, w, w, w + 3$

47 This is a paper/pencil copy of an online technology enhanced item.

Place (click and drag) the data sets below into the appropriate rows in the table.

Symmetric about the Mean	
Skewed Left	
Skewed Right	

15, 25, 35, 45, 55, 115	**15, 75, 85, 95, 105, 115**
15, 25, 35, 45, 55, 65	

The table below shows the area of several states.

State	**Area** (thousands of square miles)
Connecticut	6
Georgia	59
Maryland	12
Massachusetts	11
New Hampshire	9
New York	54
North Carolina	54
Pennsylvania	46

Delaware has an area of 2,000 square miles. Which is true if Delaware is included in the data set?

A The mean increases.

B The range decreases.

C The interquartile range decreases.

D The standard deviation increases.

NC.M1.S-ID.3 Examine the effects of extreme data points (outliers) on shape, center, and/or spread.

The number of points scored by a basketball player in the first eight games of a season are shown below.

15, 35, 18, 30, 25, 21, 32, 16

What would happen to the data distribution if she scored 24, 22, 27, and 28 points in her next four games?

A The data distribution would become less peaked and more widely spread.

B The data distribution would become less peaked and less widely spread.

C The data distribution would become more peaked and less widely spread.

D The data distribution would become more peaked and more widely spread.

A set of nine data points is shown below.

8, 11, 12, 10, 9, 7, 5, 3, 9

Which statement is true if a tenth data point of 45 is added to the data set?

A The mean and median will both increase.

B The mean will increase and the median will decrease.

C The mean will increase and the median will remain the same.

D The mean and median will both decrease.

The table below shows the weights of 8 different bears at a zoo.

Type of Bear	Weight (pounds)
Asiatic Black Bear	225
Black Bear	300
Brown Bear	550
Panda Bear	200
Polar Bear	1,000
Sloth Bear	300
Spectacled Bear	280
Sun Bear	100

If the weight of the polar bear is removed, which statement is true?

A The mean decreases more than the median because the polar bear is a high outlier.

B The mean decreases less than the median because the polar bear is a high outlier.

C The mean decreases more than the median because the high value balances the low value.

D The mean decreases less than the median because the high value balances the low value.

NC.M1.S-ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related. (See A, B, C)

NC.M1.S-ID.6A Fit a least squares regression line to linear data using technology. Use the fitted function to solve problems.

The table below displays the walking heart rate and running heart rate of eight girls in beats per minute (bpm).

Walking Heart Rate	Running Heart Rate
66	128
72	136
74	134
78	138
80	142
84	146
86	148
88	152

Using the linear best-fit model for the data, what is the predicted running heart rate of a girl whose walking heart rate is 100 bpm?

A 161 bpm

B 163 bpm

C 165 bpm

D 167 bpm

NC.M1.S-ID.6b Fit a least squares regression line to linear data using technology. Use the fitted function to solve problems

The table below shows the shoe size and age of 7 boys.

Name	Shoe Size (x)	Age (y)
Tyrone	6	9
Marcel	6	11
Patrick	7	15
Bobby	8	11
Dylan	9	15
Mike	10	16
Jonathan	12	17

Approximately what percent of the boys' ages is more than 1 year different from the age predicted by the line of best fit for the data?

A 14%

B 29%

C 43%

D 57%

The table below shows the number of hours 7 students studied for a math test and the grade each student earned on the test.

Student	Hours Studied (x)	Test Grade (y)
Mary	2.00	84
Jonathan	1.75	86
Susan	2.00	88
Terry	3.00	94
Patrick	3.50	95
Amanda	3.50	93
Darius	2.25	89

How does Amanda's test score compare to the score predicted using the linear best-fit model of data for a student who studied 3.50 hours?

A Amanda scored about 5 points lower than the score predicted for a student who studied 3.50 hours.

B Amanda scored about 5 points higher than the score predicted for a student who studied 3.50 hours.

C Amanda scored about 2 points lower than the score predicted for a student who studied 3.50 hours.

D Amanda scored about 2 points higher than the score predicted for a student who studied 3.50 hours.

NC.M1.S-ID.6b Assess the fit of a linear function by analyzing residuals

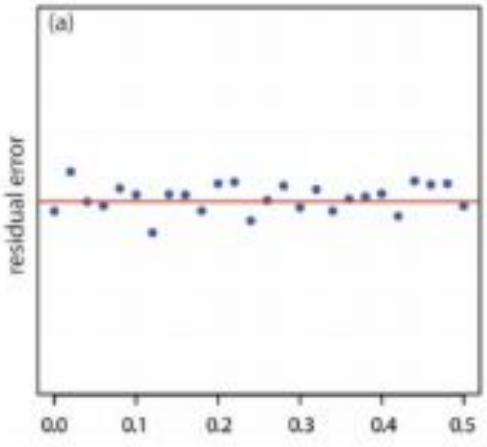
tudents can determine the residual for any value in a data set.

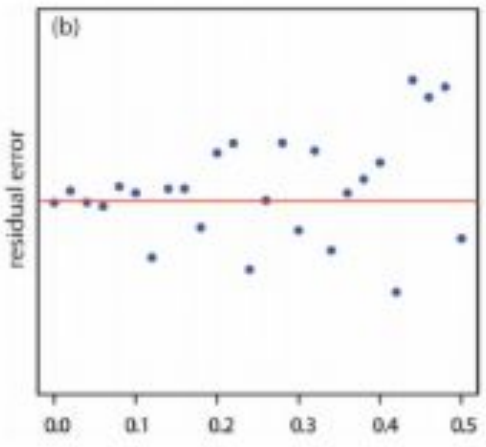
Example: The table to the left displays the annual tuition rates of a state college in the U.S. between 1990 and 2000, inclusively. The linear function $R(t) = 326x + 6440$ has been suggested as a good fit for the data.

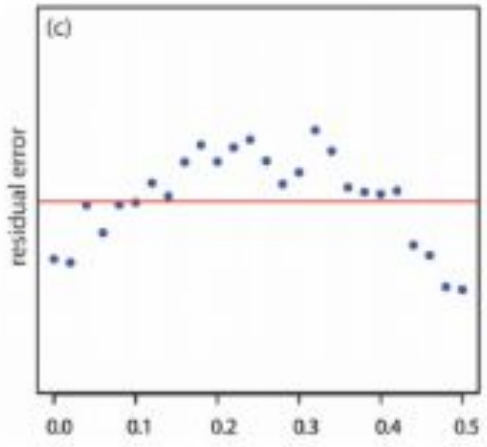
a. Extend the table to find the predicted rates based on the model and the residual values for each year.
b. Create the residual plot for the tuition rates.
c. Use the residual plot to determine the goodness of fit of the function for the data provided in the table.

Year (0 = 1990)	Tuition Rate	Predicted Rate	Residuals
0	6546		
1	6996		
2	6996		
3	7350		
4	7500		
5	7978		
6	8377		
7	8710		
8	9110		
9	9411		
10	9800		

Example: What do the following residual plots tell you about the appropriateness of a linear model for the functions they represent? Explain your responses.







NC.M1.S-ID.6c Fit a function to exponential data using technology. Use the fitted function to solve problems.

Example: What is the exponential function that best models the number of gnats the scientists have gathered after the number of hours listed? How many hours will it take for 200 gnats to gather?

Hours	0	1	2	3	4
Number of gnats	12	20	35	60	80

Example: In an experiment, 300 pennies were shaken in a cup and poured onto a table. Any penny 'heads up' was removed. The remaining pennies were returned to the cup and the process was repeated. The results of the experiment are shown below. Write a function rule suggested by the context. Use the context to explain all values of the function. How are those values reflected in the table?

# of Rolls	0	1	2	3	4	5
# of Pennies	300	164	100	46	20	8

NC.M1.S-ID.7 Interpret in context the rate of change and the Intercept of a linear model. Use the linear model to interpolate and extrapolate predicted values. Assess the validity of a predicted value.

The table below shows the distance a car has traveled.

Minutes	25	50	75	100	125
Distance Traveled (in miles)	20	40	60	80	100

What is the meaning of the slope of the linear model for the data?

A The car travels 5 miles every minute.

B The car travels 4 miles every minute.

C The car travels 4 miles every 5 minutes.

D The car travels 5 miles every 4 minutes.

The table below shows the U.S. average life expectancy at birth, in years, in various decades.

Years since 1930	Life Expectancy at Birth
10	62.9
20	68.2
30	69.7
40	70.8
50	73.7
60	75.4
70	77.0
80	78.7

What is the meaning of the slope of the linear best-fit equation for the data?

A The predicted average life expectancy at birth in 1930 was about 62.7 years.

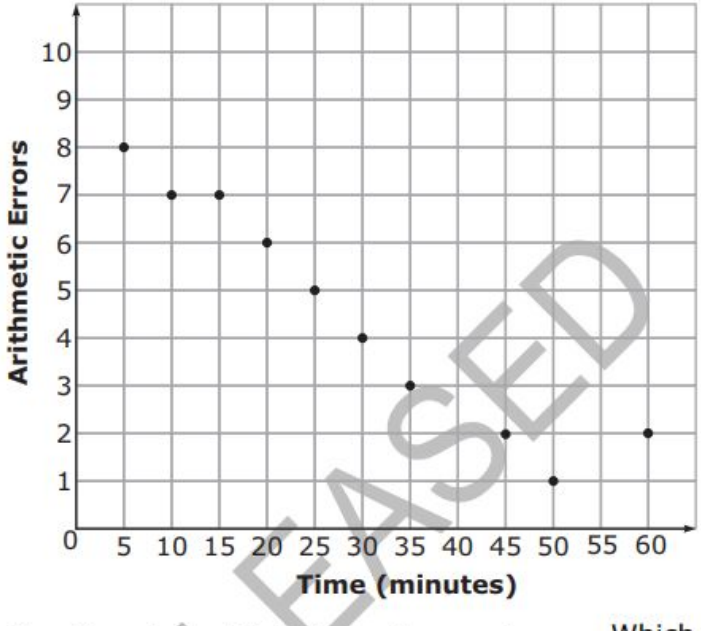
B The predicted average life expectancy at birth in 1930 was about 57.6 years.

C The average life expectancy at birth increases by about 6.7 each year.

D The average life expectancy at birth increases by about 0.2 each year.

NC.M1.S-ID.8 Analyze patterns and describe relationships between two variables in context. Using technology, determine the correlation coefficient of bivariate data and interpret it as a measure of the strength and direction of a linear relationship. Use a scatter plot, correlation coefficient, and a residual plot to determine the appropriateness of using a linear function to model a relationship between two variables.

The scatterplot below shows the number of arithmetic errors 10 students made on a quiz and the amount of time the students took to complete the quiz.



Which describes the relationship between the number of arithmetic errors the students made and the amount of time the students took to complete the quiz?

A There is a strong positive relationship between the variables.

B There is a strong negative relationship between the variables.

C There is a weak positive relationship between the variables.

D There is a weak negative relationship between the variables.

A statistician collected the following data to explore the relationship between two variables, x and y.

x	y
2.3	11.0
4.2	16.5
5.1	19.2
6.4	23.1
8.2	24.3
8.5	29.5

The statistician performed a linear regression and also plotted the residuals.

- Based on the residual plot, the statistician decided to exclude one data point.
- The statistician then performed linear regression on the set of remaining data points.
- The result was that the new linear model fit the remaining data more closely than the original model fit the original data.

Which data point did the statistician exclude?

A (2.3, 11.0)

B (4.2, 16.5)

C (6.4, 23.1)

D (8.2, 24.3)

Marcus measured the height, in inches, y, of plants over the course of 3 weeks. The correlation coefficient between the number of days, x, and the height of the plants is 0.85. Which could be concluded based on the correlation coefficient of the data?

A There is a strong relationship showing that as the number of days increases, the height of the plants increases.

B There is a strong relationship showing that as the number of days increases, the height of the plants decreases.

C There is a weak relationship showing that as the number of days increases, the height of the plants increases.

D There is a weak relationship showing that as the number of days increases, the height of the plants decreases.

NC.M1.S-ID.9 Distinguish between association and causation.

Students will determine if statements of causation are reasonable or not and justify their opinion.

Example: A study found a strong, positive correlation between the number of cars owned and the length of one's life. Larry concludes that owning more cars means you will live longer. Does this seem reasonable? Explain your answer.

Example: Choose two variables that could be correlated because one is the cause of the other; defend and justify the selection of variables.
