Warm Up:
http://bit.ly/SCMSMath1
You may begin this survey, but
DO NOT HIT SUBMIT!

Announcements

◄ Math MAPs are next week on Monday and Tuesday

◄ Flashback Fridays will begin this Friday
  ◄ 10 questions
  ◄ On SchoolNet
  ◄ Graded for accuracy
  ◄ You MAY consult friends, notes, etc.

Unit Map - Statistics

- ~~Tuesday - Histograms and Box Plots~~
- ~~Wednesday - Measures of Central Tendency~~
- **Thursday - Line of Best Fit**
- Friday - Line of Best Fit
- Monday - Math MAP
- Tuesday - Math MAP (continued)
- Wednesday - Friday DC Trip
- Spring Break
- Monday - Statistics Review
- Tuesday - Statistics Test

Tutor.com

Through OneAccess

Homework Review

Line of Best Fit

4/4/2019

NC.M1.S-ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

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

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

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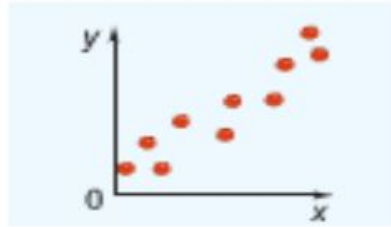
Scatterplots and Trend Lines

A scatter plot is a graph that relates two different sets of data by displaying them as ordered pairs.

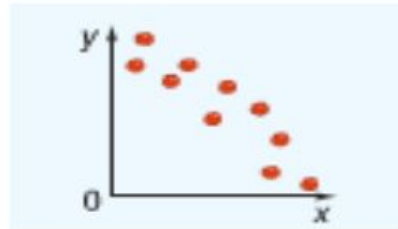You can use scatter plots to find trends in data.

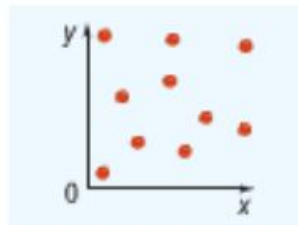| Positive Coorelation	When y tends to increase as x increases
Negative correlation	When y tends to decrease as x increases
No correlation	When x and y are not related

Making a Scatterplot and Describing its Correlation

Temperature The table shows the altitude of an airplane and the temperature outside the plane.

Plane Altitude and Outside Temperature										
Altitude (m)	0	500	1000	1500	2000	2500	3000	3500	4000	4500	5000
Temperature (°F)	59.0	59.2	61.3	55.5	41.6	29.8	29.9	18.1	26.2	12.4	0.6

A Make a scatter plot of the data.

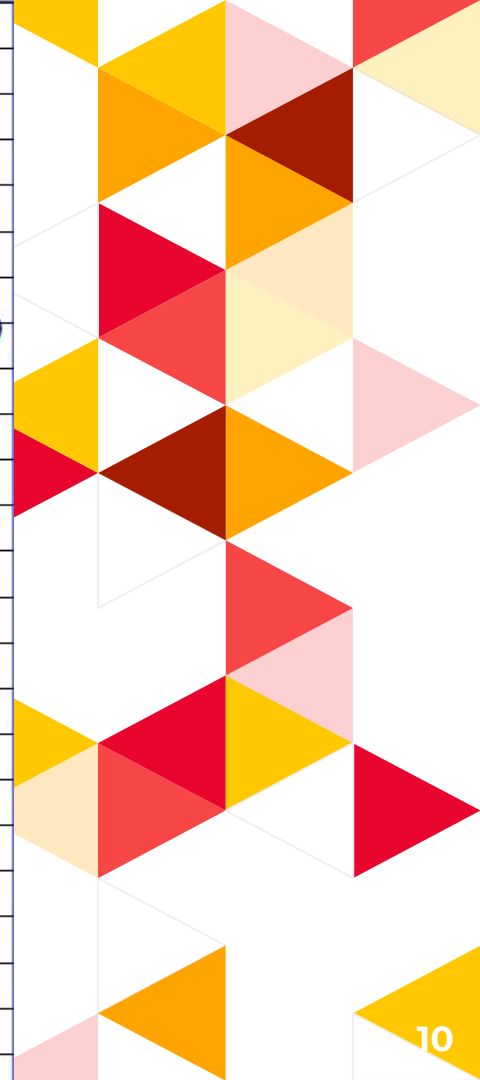Treat the data as ordered pairs. For the altitude of 1500 m and the temperature of 55.5°F, plot (1500, 55.5).

Plane Altitude and Outside Temperature										
Altitude (m)	0	500	1000	1500	2000	2500	3000	3500	4000	4500	5000
Temperature (°F)	59.0	59.2	61.3	55.5	41.6	29.8	29.9	18.1	26.2	12.4	0.6

What type of relationship does the scatter plot show?

60
50
40
30
20
10

1000 2000 3000 4000 5000

Altitude 3000

Analyzing Data

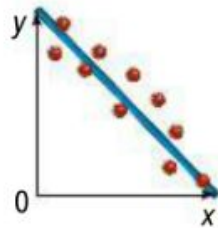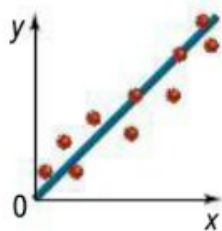|	
Trend Line	A line on a scatter plot, drawn near the points that show a correlation.
Interpolation	Estimating a value between two known values
Extrapolation	Predicting a value outside the range of known values.
Line of Best Fit / Trend Line	The trend line that shows the relationship between two sets of data most accutately
Correlation Coefficient	a number from -1 to 1 that tells you how closely the equation models the data. This is shown on the graphing calculator.

A trend line is a line on a scatter plot, drawn near the points that show a correlation.

A trend line to estimate a value between two know data values or to predict a value outside the range of known data values.

Interpolation is estimating a value between two known values

Extrapolation is predicting a value outside the range of known values.
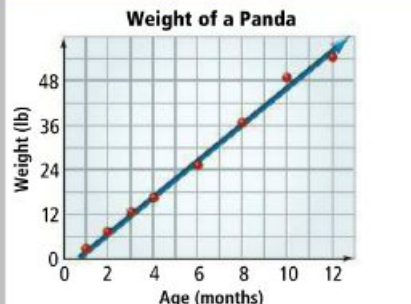
Writing an Equation of a Trend Line

Biology Make a scatter plot of the data at the right. What is the approximate weight of a 7-month-old panda?

Step 1 Make a scatter plot and draw a trend line. Estimate the coordinates of two points on the line.
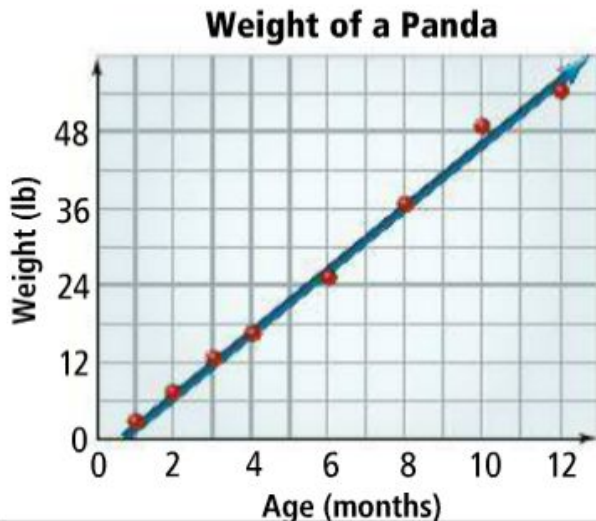
Weight of a Panda

Age (months)	Weight (lb)
1	2.5
2	7.6
3	12.5
4	17.1
6	24.3
8	37.9
10	49.2
12	54.9

Weight of a Panda

Use point-slope form to find the equation

Two points on the trend line are $(4, 17.1)$ and $(8, 37.9)$.

Step 2   Write an equation of the trend line.

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{37.9 - 17.1}{8 - 4} = \frac{20.8}{4} = 5.2 \qquad \text{Find the slope of the trend line.}$$

$$y - y_1 = m(x - x_1) \qquad \text{Use point-slope form.}$$

$$y - 17.1 = 5.2(x - 4) \qquad \text{Substitute 5.2 for } m \text{ and } (4, 17.1) \text{ for } (x_1, y_1).$$

$$y - 17.1 = 5.2x - 20.8 \qquad \text{Distributive Property}$$

$$y = 5.2x - 3.7 \qquad \text{Add 17.1 to each side.}$$

Step 3   Estimate the weight of a 7-month-old panda.

$$y = 5.2(7) - 3.7 \qquad \text{Substitute 7 for } x.$$

$$y = 32.7 \qquad \text{Simplify.}$$

The weight of a 7-month-old panda is about 32.7 lb.

How to Clear your Calculator's Memory

2nd   +   7   >   >   Enter   2

* You will be using this often, because anytime you enter data or graphs your calculator will save this information. You will want to clear out the information between every problem.

AFTER YOU CLEAR YOUR CALCULATOR ALWAYS TURN DIAGNOSTICS ON

2nd   0   DiagnosticOn   Enter

Putting a Scatterplot on your Calculator

1) STAT; EDIT; ENTER
2) Enter x's in L1 column and y's in L2 column
3) Go to Y= and turn StatPlot On
4) Graph

* If you do not see your graph then your window needs adjusted. Use the ZOOM, 9 (Zoom Stat) button.

How to find the trend line (linear regression)

1) STAT, CALC, LinReg, get the equation
2) Go to Y=
3) VARS, Statistics, EQ, RegEQ
4) Graph

How to make a prediction based on your trend line

1) 2nd, Trace, Value, type number you are estimating, Enter

Use your graphing calculator to
1. create a scatter plot
2. find the equation of the line of best fit
3. create a line of best fit on your graph.
4. What is the cost of attending in the
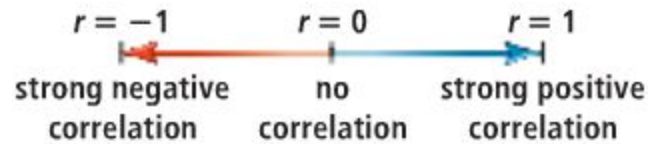   2012-2013 academic year?

Average Tuition and Fees at Public 4-Year Colleges

Academic Year	Cost ($)
2000–2001	3508
2001–2002	3766
2002–2003	4098
2003–2004	4645
2004–2005	5126
2005–2006	5492
2006–2007	5836

Source: The College Board

Correlation coefficient

Correlation coefficient, r, is a number between -1 to 1, that tells you how closely the equation models the data.

The nearer r is to 1 or -1 the more closely the data clusters around the line of best fit.

If r is near 1, the data lie close to a line of best fit with positive slope.
If r is near -1, the data lie close to a line of best fit with negative slope.

(Hint - if you can't find your correlation coefficient, you didn't turn diagnostic on)

Problem 3 Finding the Line of Best Fit

College Tuition Use a graphing calculator to find the equation of the line of best fit for the data at the right. What is the correlation coefficient to three decimal places? Predict the cost of attending in the 2012–2013 academic year.

Step 1 Press **stat**. From the **EDIT** menu, choose **Edit**. Enter the years into L_1. Let $x = 2000$ represent academic year 2000–2001, $x = 2001$ represent 2001–2002, and so on. Enter the costs into L_2.

Step 2 Press **stat**. Choose **LinReg(ax + b)** from the **CALC** menu. Press **enter** to find the equation of the line of best fit and the correlation coefficient. The calculator uses the form $y = ax + b$ for the equation.
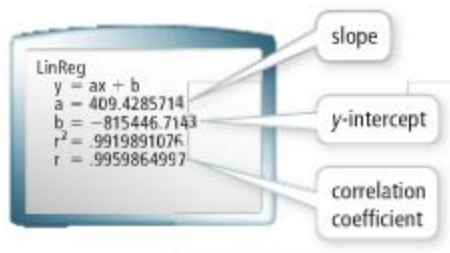
Average Tuition and Fees at Public 4-Year Colleges

Academic Year	Cost ($)
2000–2001	3508
2001–2002	3766
2002–2003	4098
2003–2004	4645
2004–2005	5126
2005–2006	5492
2006–2007	5836

Source: The College Board

```
LinReg
 y = ax + b
 a = 409.4285714          slope
 b = -815446.7143         y-intercept
 r² = .9919891076
 r = .9959864997          correlation
                          coefficient
```

Round to the nearest hundredth. The equation of the line of best fit is $y = 409.43x - 815{,}446.71$. The correlation coefficient is about 0.996.

Step 3 Predict the cost of attending in the 2012–2013 academic year.

$y = 409.43x - 815{,}446.71$ Use the equation of the line of best fit.

$y = 409.43(2012) - 815{,}446.71$ Substitute 2012 for x.

$y \approx 8326$ Simplify. Round to the nearest whole number.

The cost of attending a four-year public college in the 2012–2013 academic year is predicted to be about $8326.

Total U.S. Vehicle Production (millions)

1960	1970	1980	1990	2000
7.9	8.8	8.0	9.8	12.8

You try!

Using your calculator complete the following

1. create a scatter plot

2. create a line of best fit

3. create an equation for the line of best fit.

4. Find the value of the correlation coefficient r and explain what that means for this problem

Causation is when a change in one quantity causes a change in a second quantity. A correlation between quantities does not always imply causation.

Problem 4 Identifying Whether Relationships Are Causal

In the following situations, is there likely to be a correlation? If so, does the correlation reflect a causal relationship? Explain.

A the number of loaves of bread baked and the amount of flour used

There is a positive correlation and also a causal relationship. As the number of loaves of bread baked increases, the amount of flour used increases.

B the number of mailboxes and the number of firefighters in a city

There is likely to be a positive correlation because both the number of mailboxes and the number of firefighters tend to increase as the population of a city increases. However, installing more mailboxes will not *cause* the number of firefighters to increase, so there is no causal relationship.

Got It? 4. In the following situations, is there likely to be a correlation? If so, does the correlation reflect a causal relationship? Explain.
   a. the cost of a family's vacation and the size of their house
   b. the time spent exercising and the number of Calories burned

What if the data looks like it is closer to an exponential?

Instead of finding the LinReg…

STAT, CALC, ExpReg

Trend Lines on the Calculator

Plot the following points in your calculator:

| 5	6	7	8
1	3	9	20

Do the points appear to be a linear trend, or an exponential trend?

Find the line of best fit for the data

Correlation Coefficient to determine if you should be linear or exponential

Residuals

A residual is the difference between the measured value and the predicted value of a regression model. It is important to understand residuals because they show how accurate a mathematical function, such as a line, is in representing a set of data.
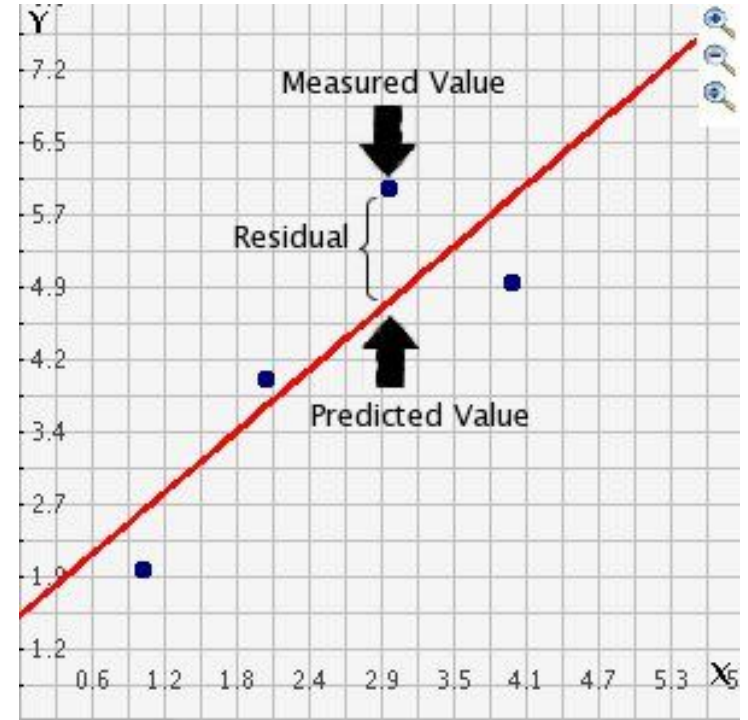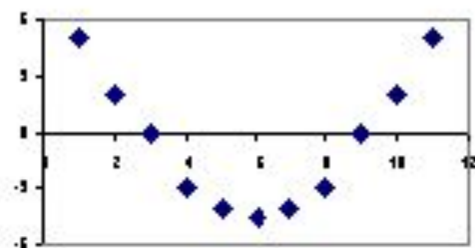
Residual Plot

A residual plot is a graph that shows the residuals on the vertical axis and the independent variable on the horizontal axis. If the points in a residual plot are randomly dispersed around the horizontal axis, a linear regression model is appropriate for the data; otherwise, a non-linear model is more appropriate.
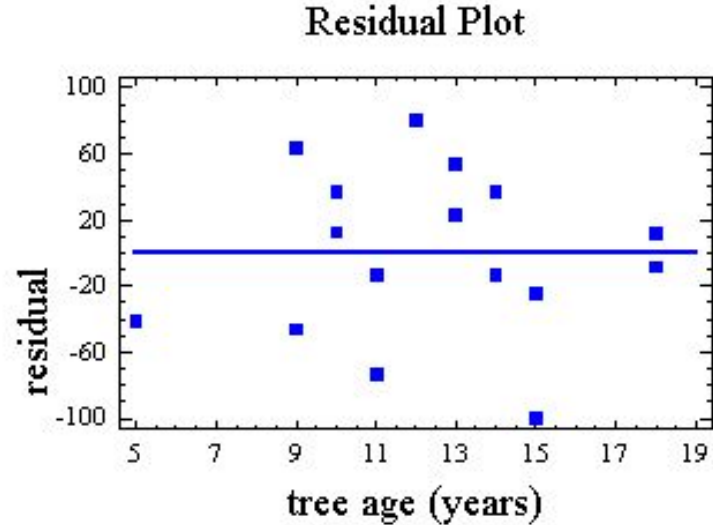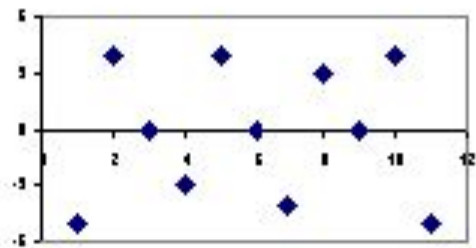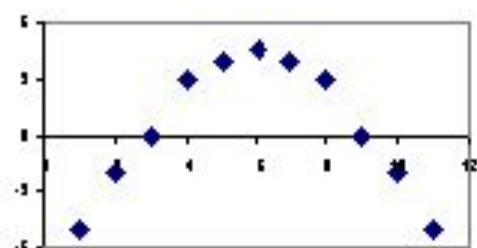

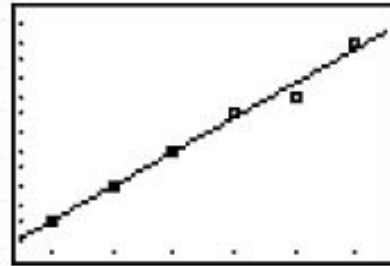
Residual Plot

Random pattern   **Non-random: U-shaped**   **Non-random: Inverted U**

Graphing a Residual Plot on you Calculator

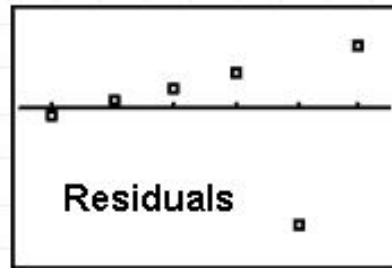Consider this set of data (which is "almost" a perfect linear equation)

Released EOC Questions for this Standard

46 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

37   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.

37 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%

Homework: Prepare for a line of best fit calculator race tomorrow