Algebra 1 Notes SOL A.11 (5.6, 5.7) Line of Best Fit Mrs. Grieser

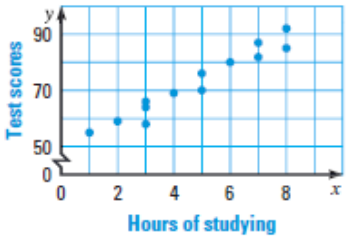
Name: _____ Block: _____ Date: _____

Lines of Fit

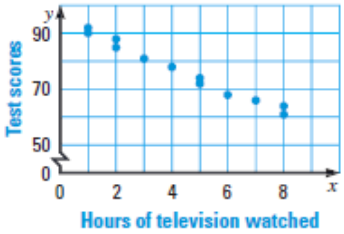
Fitting a line to data is a statistical process used to make predictions about data.

A scatter plot is a graph of unconnected data points used to determine relationships between data.

Data may be correlated. A positive correlation occurs when y-values increase when x-values increase. A negative correlation occurs when y-values decrease as x-values increase. If there is no apparent relationship between the data, then the data is said to have relatively no correlation.



Positive correlation



Negative correlation

Relatively no correlation

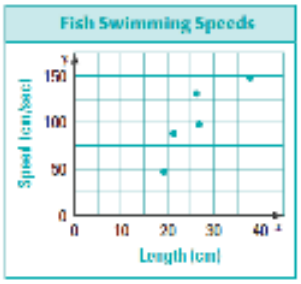
In the first figure, we see a positive correlation between hours of studying and test scores. This means as hours of study increases, test scores increase, too. The second figure shows a negative correlation between hours of television watched and test scores. As television viewing increased, test scores decreased.

The correlation coefficient, r, measures how strong correlation is. The closer r is to 1, the stronger the positive correlation; the closer it is to -1, the stronger the negative correlation. A value of r close to 0 shows relatively no correlation.

Making a Scatter Plot

To make a scatter plot, treat the data as ordered pairs. Plot on a coordinate plane. Describe any correlation seen.

Fish	Pike	Red gurnard	Black bass	Gurnard	Norway haddock
Length (cm)	37.8	19.2	21.3	26.2	26.8
Speed (cm/sec)	148	47	88	131	98



What kind of correlation is shown? _____

Adding a Line of Fit

1) Make a scatter plot of data.
2) Decide whether the data can be modeled (does it show correlation?).
3) "Eyeball" a line that appears to fit the data closely. There should be about the same number of points above or below the line drawn.
4) Write an equation for the line drawn.

Example: Write a Line of Fit for Bird Population Data

The table below shows the number of woodpecker clusters in a part of a forest. Write an equation that models the number of active clusters as a function of the number of years since 1990.

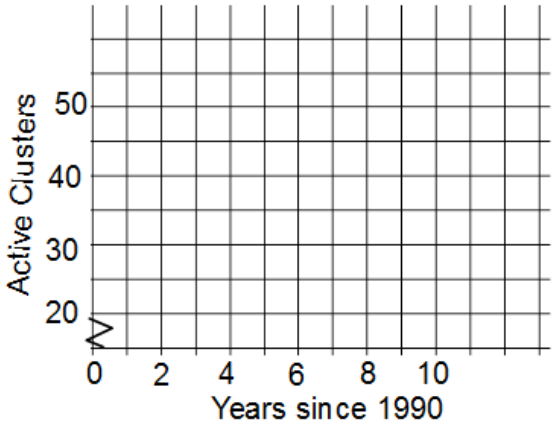
Year	1992	1993	1994	1995	1996	1997	1998	1999	2000
Active clusters	22	24	27	27	34	40	42	45	51

1) Make a scatter plot.

2) Does the plot show a correlation? If so, we can fit a line to the data.

3) Draw a line that appears to fit the points (use a straight tool such as a thin pencil or dry spaghetti strand to help fit the line).

4) Write an equation for the line.
(Find two points on the line, calculate the slope, then write an equation using slope-intercept or point-slope).



Predicting with Linear Models

To get a line of "best fit," a statistical method called linear regression is used. You can use this line to approximate other data points. This is called linear interpolation (for values inside the data range) and linear extrapolation (for values outside of the data range).

The manual process to perform a linear regression is beyond the scope of this course. Luckily we have graphing calculators to do the work for us!

The basic steps on a graphing calculator:

1. Press STAT. Enter the x-value data in L_1. Enter the y-value data in L_2.
2. Press 2^{nd} STAT PLOT. Select Plot 1 and turn it ON, select scatter, enter the Xlist and Ylist values (L_1 and L_2) and your favorite scatter mark (the first one is nice).
3. Press ZOOM 9 (ZoomStat) to see the scatter plot.
4. Press STAT. Select CALC, then choose 4 (LinReg(ax + b)). Before pressing enter, enter L_1, a comma (,), and L_2 (2^{nd} 1 and 2^{nd} 2), a comma, and press VARS. Select Y-VARS, then select 1:Function, then 1:Y_1. Your display should look like: LinReg(ax+b) L_1, L_2, Y_1. Press enter.
5. A display will show you a and b (slope and y-intercept) of the line of best fit. It will also show you r, the correlation coefficient.
6. Press GRAPH.

Use the method above to come up with the line of best fit for the woodpecker example.