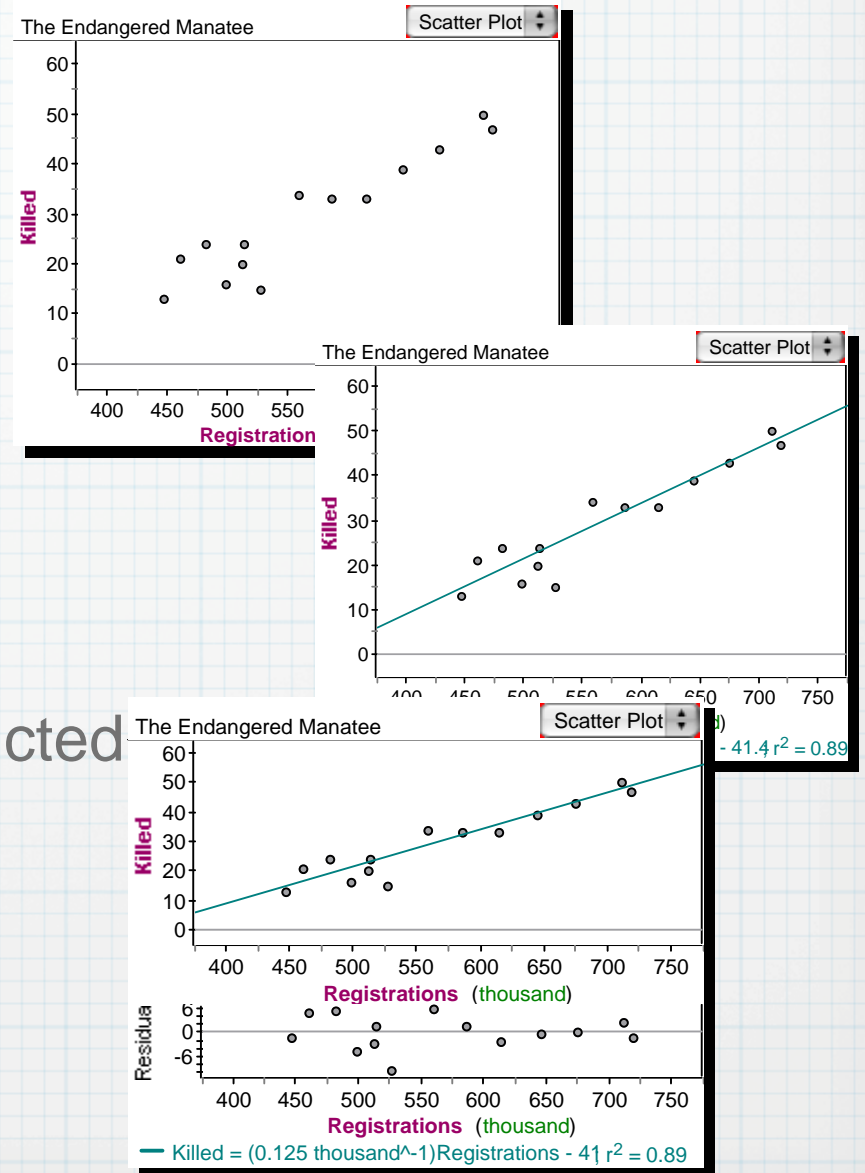


# Review of Regression Basics

- When describing a Bivariate Relationship:
  - Make a Scatterplot
    - Strength, Direction, Form
  - Model:  $\hat{y} = a + bx$ 
    - Interpret slope in context
  - Make Predictions
    - Residual = Observed - Predicted
  - Assess the Model
    - Interpret “r”
    - Residual Plot



# Reading Minitab Output

Regression Analysis: Fat gain versus NEA

The regression equation is

FatGain = \*\*\*\*\* + \*\*\*\*\*(NEA)

Predictor	Coef	SE Coef
Constant	3.5051	0.3036
NEA	-0.0034415	0.0008

S=0.739853      R-Sq = 60.6%

T  
11.54  
P  
0.000

The Intercept is also known  
as a "constant"  
 $a=3.5051$

The Slope and Intercept are  
"coefficients" in the LSRL.

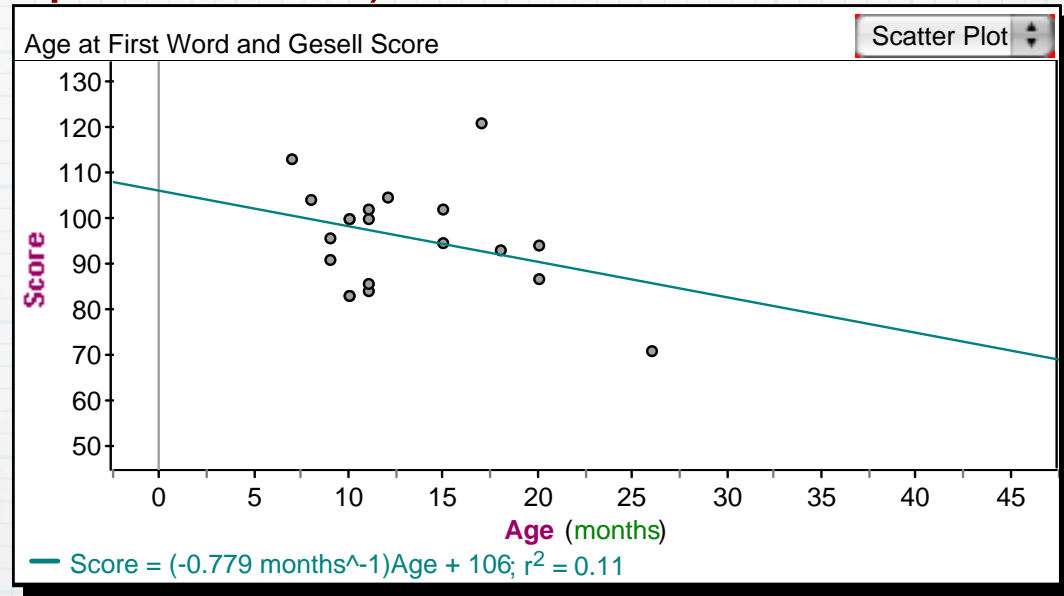
The Slope is the coefficient of  
the explanatory variable.  
 $b=-0.0034415$

Regression equations aren't always as easy to spot as they  
are on your TI-84. Can you find the slope and intercept  
above?  **$\text{FatGain} = 3.5051 - 0.0034415(\text{NEA})$**

# Outliers/Influential Points

	Child	Age	Score
1	1	15 months	95
2	2	26 months	71
3	3	10 months	83
4	4	9 months	91
5	5	15 months	102
6	6	20 months	87
7	7	18 months	93
8	8	11 months	100
9	9	8 months	104
10	10	20 months	94
11	11	7 months	113
12	12	9 months	96
13	13	10 months	83
14	14	11 months	84
15	15	11 months	102
16	16	10 months	100
17	17	12 months	105
18	18	42 months	57
19	19	17 months	121
20	20	11 months	86
21	21	10 months	100

Does the age of a child's first word predict his/her mental ability? Consider the following data on (age of first word, Gesell Adaptive Score) for 21 children.

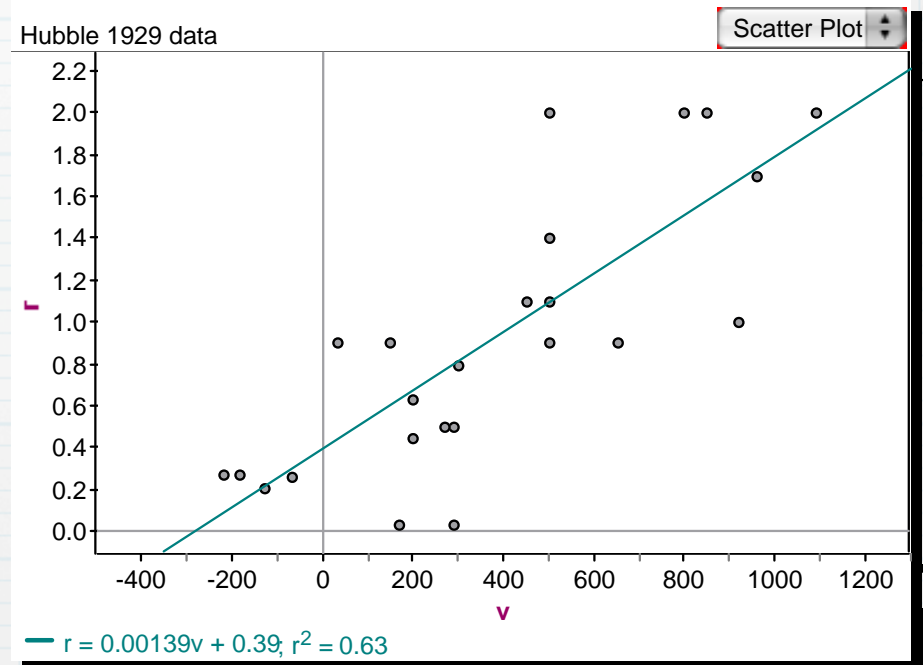
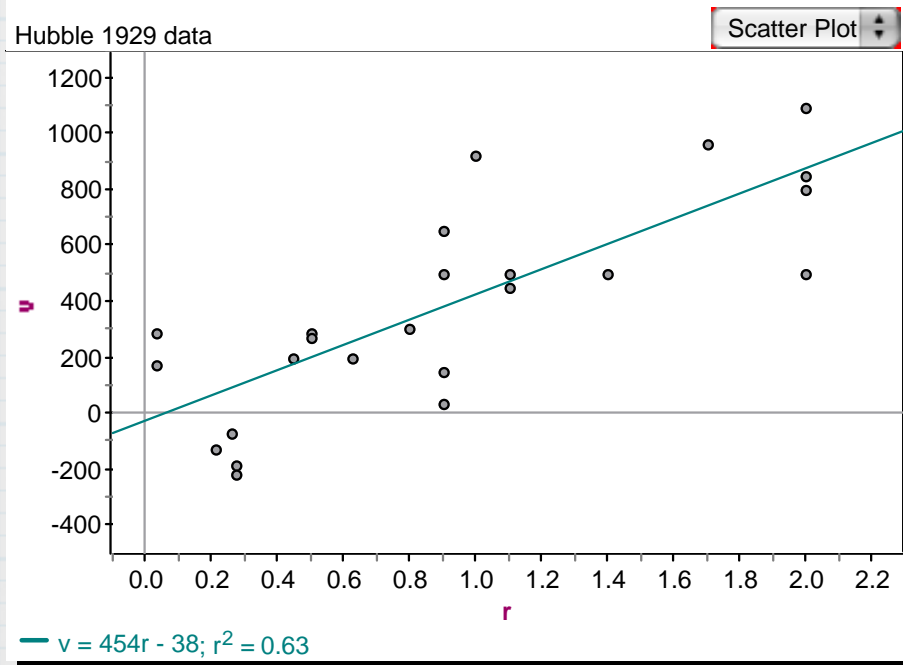


Does the highlighted point markedly affect the equation of the LSRL? If so, it is "influential".

Test by removing the point and finding the new LSRL.

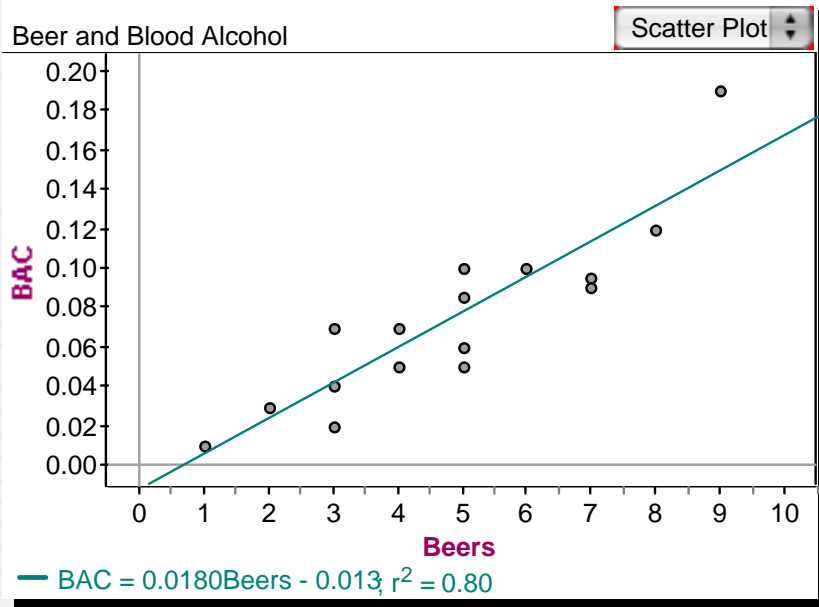
# Explanatory vs. Response

- The Distinction Between Explanatory and Response variables is essential in regression.
- Switching the distinction results in a different least-squares regression line.



- Note: The correlation value,  $r$ , does NOT depend on the distinction between Explanatory and Response.

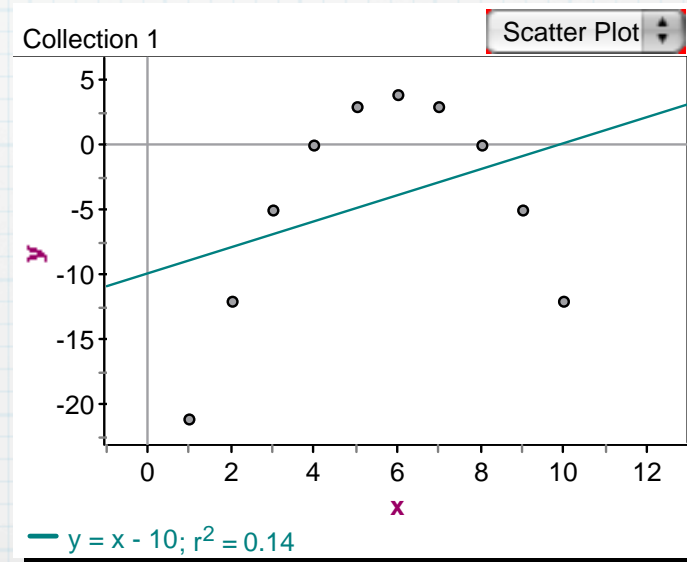
# Correlation



■ The correlation,  $r$ , describes the strength of the straight-line relationship between  $x$  and  $y$ .

■ Ex: There is a strong, positive, LINEAR relationship between # of beers and BAC.

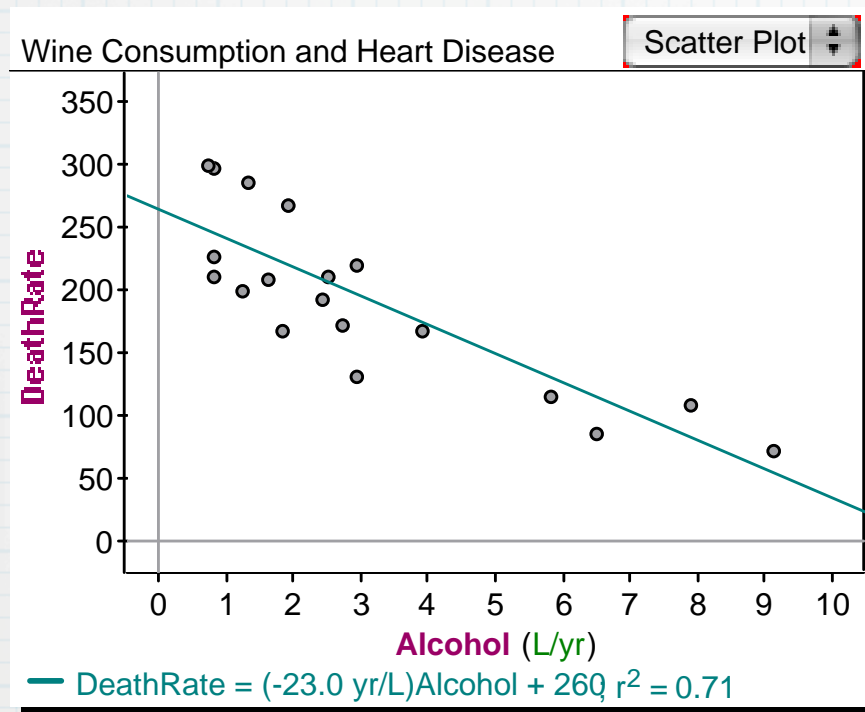
- There is a weak, positive, linear relationship between  $x$  and  $y$ . However, there is a strong nonlinear relationship.
- $r$  measures the strength of linearity...





# Coefficient of Determination

■ The coefficient of determination,  $r^2$ , describes the percent of variability in  $y$  that is explained by the linear regression on  $x$ .



■ 71% of the variability in death rates due to heart disease can be explained by the LSRL on alcohol consumption.

■ That is, alcohol consumption provides us with a fairly good prediction of death rate due to heart disease, but other factors contribute to this rate, so our prediction will be off somewhat.

# Cautions

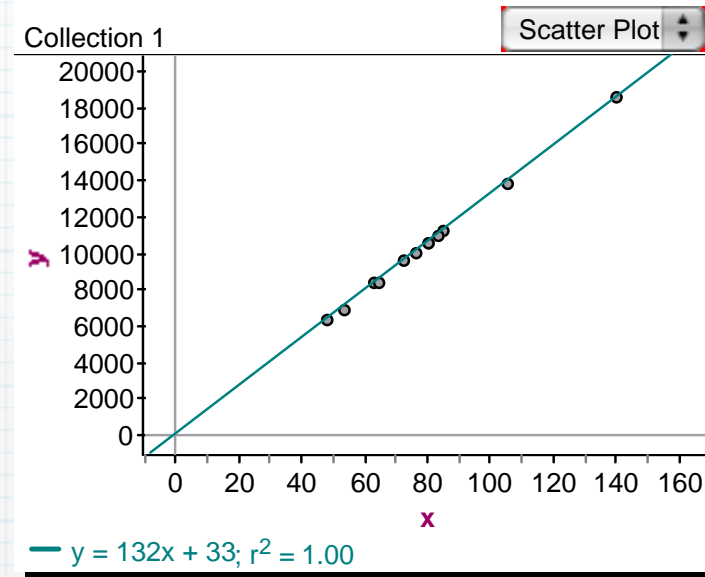
- Correlation and Regression are NOT RESISTANT to outliers and Influential Points!
- Correlations based on “averaged data” tend to be higher than correlations based on all raw data.
- Extrapolating beyond the observed data can result in predictions that are unreliable.

# Correlation vs. Causation

■ Consider the following historical data:

Collection 1

	Year		
1	1860	63	8376
2	1865	48	6406
3	1870	53	7005
4	1875	64	8486
5	1880	72	9595
6	1885	80	10643
7	1890	85	11265
8	1895	76	10071
9	1900	80	10547
10	1905	83	11008
11	1910	105	13885
12	1915	140	18559



■ There is an almost perfect linear relationship between x and y. ( $r=0.999997$ )

- $x$  = # Methodist Ministers in New England
- $y$  = # of Barrels of Rum Imported to Boston
- **CORRELATION DOES NOT IMPLY CAUSATION!**



# Summary

Plot your data.  
Scatterplot

Interpret what you see:  
direction, form, strength, outliers

Numerical summary?  
 $\bar{x}$ ,  $\bar{y}$ ,  $s_x$ ,  $s_y$ , and  $r$ ?

Mathematical model?  
Regression line?

How well does it fit?  
Residuals and  $r^2$

