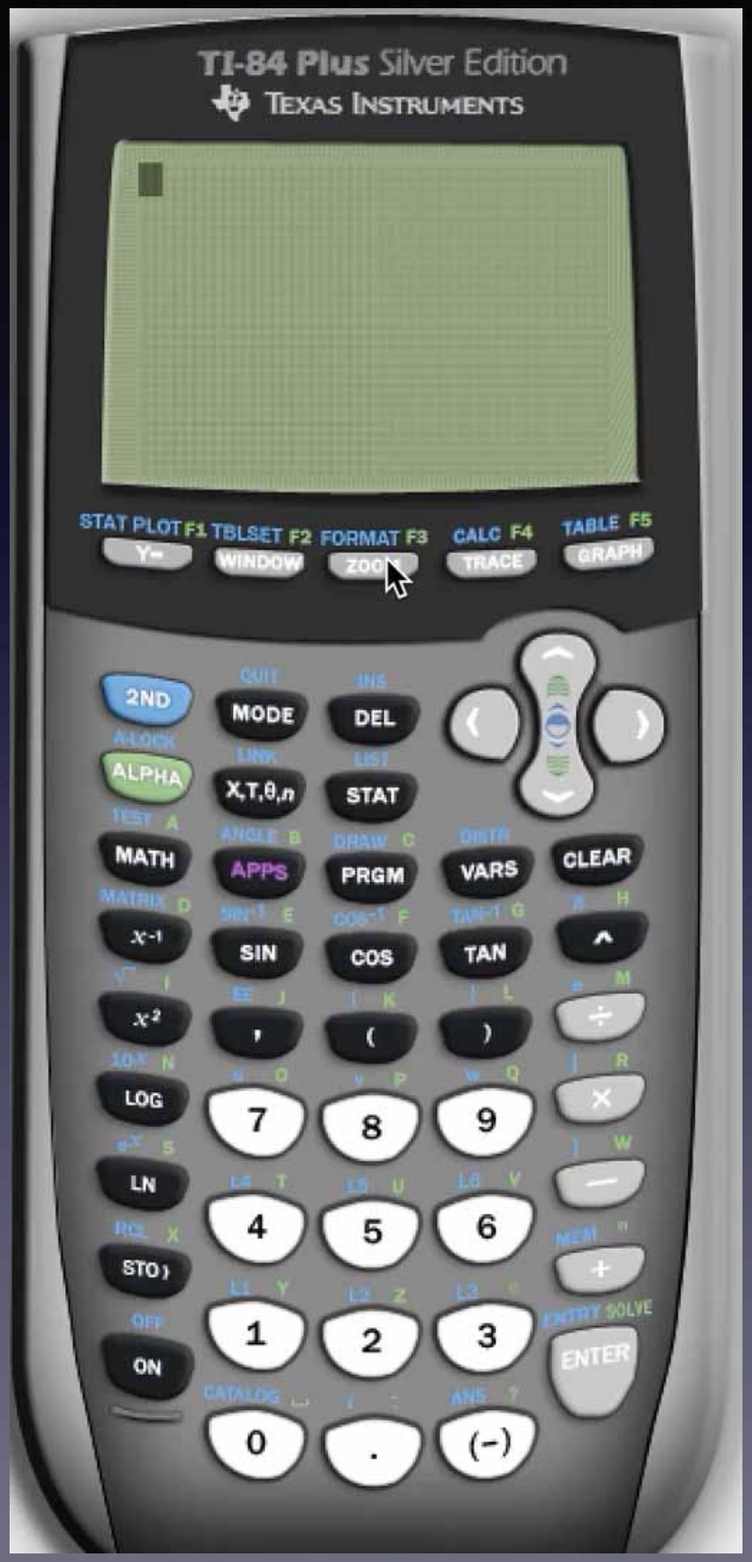
Diagnostics Must Be ON



r - Correlation Coefficient

measures the **strength** and **direction** of the linear relationship between two **quantitative** variables

sigma - just means sum/add

deviations for *x* observations

deviations for *y* observations

Do you recognize these deviations?

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

They are the *z* scores for *x* and *y*

$$r = \frac{1}{n-1}\sum_{i=1}^{n} z_{x_i} z_{y_i}$$

Use your calculator to find *r* -

- Put data in L_1 and L_2. Stat→ Calc→ 8. LinReg(a + bx)→ L_1, L_2, Y_1
- Note: Make sure your diagnostics are on: 2nd →Catalog →Scroll down to Diagnostics On →Enter
- Y_1: Vars→ Y-Vars→ 1. Function→ Y_1

Properties of *r*

- *r* indicates the **strength** and **direction** of a *linear* relationship.
- *r* can only be calculated for graphs with 2 numerical (quantitative) variables.
- *r* is always between −1 and 1, inclusive.
- Graphs with positive slopes have positive *r* values; graphs with negative slopes have negative *r* values.
- *r* remains unchanged if *x* and/or *y* are rescaled.
- *r* remains unchanged if *x* and *y* are interchanged.
- *r* is dimensionless (has no units).
- *r* is not resistant to the effects of outliers.

Interpretation of *r*

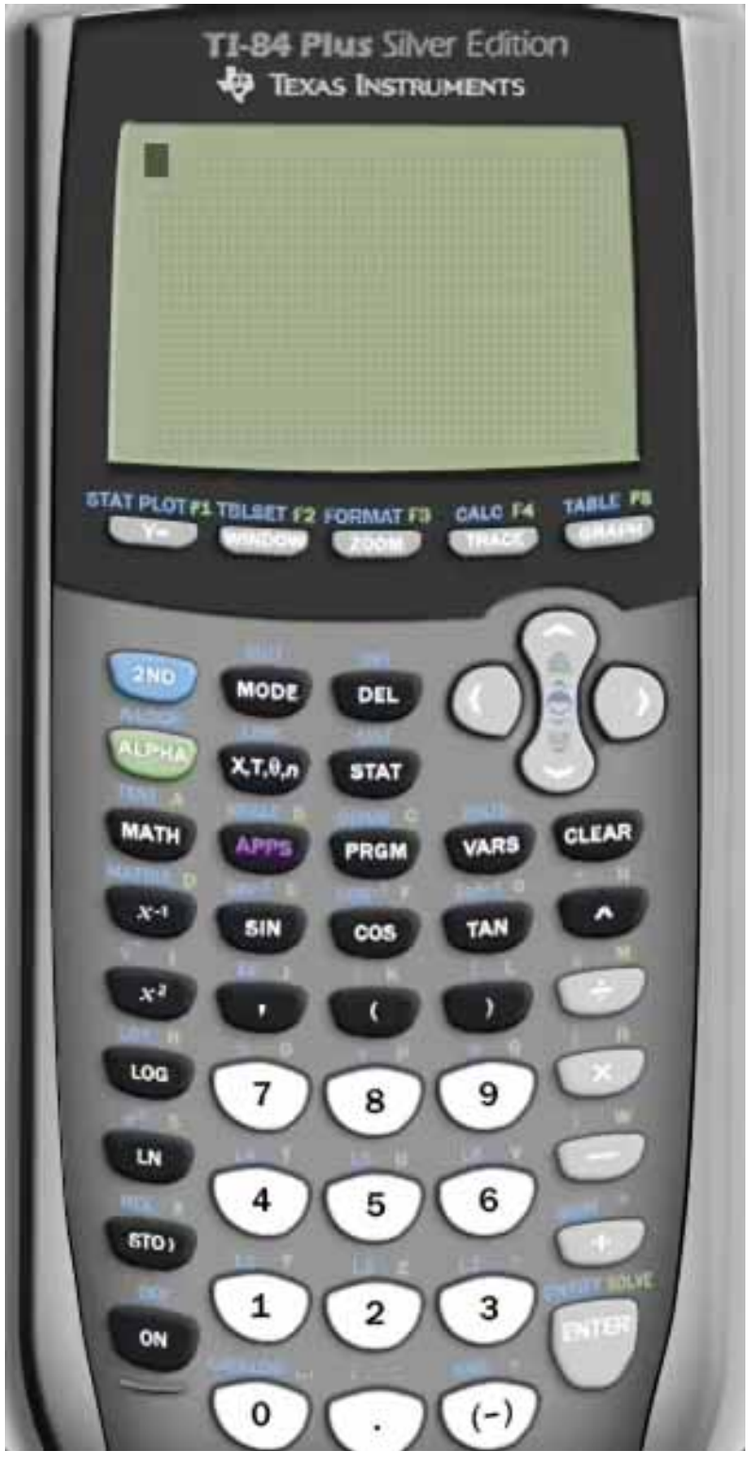
Correlation coefficient (*r*): The correlation coefficient of _____ indicates that there is a (strong, moderate, weak), (positive, negative) linear relationship between (context of *y*) and (context of *x*).

Example: *y* = height of a plant in cm, *x* = age in months, *r* = 0.945
The correlation coefficient of 0.945 indicates that there is a **strong**, **positive**, **linear** relationship between the age of the plant and its height.

Always label your variables!

Ex1 Are more expensive bike helmets safer than less expensive ones? The accompanying data on x = price and y = quality rating for 12 different brands of bike helmets is given below. Quality rating was a number from 0 (worst possible rating) to 100, and was determined based on factors that included how well the helmet absorbed the force of an impact, the strength of the helmet, ventilation, and ease of use.

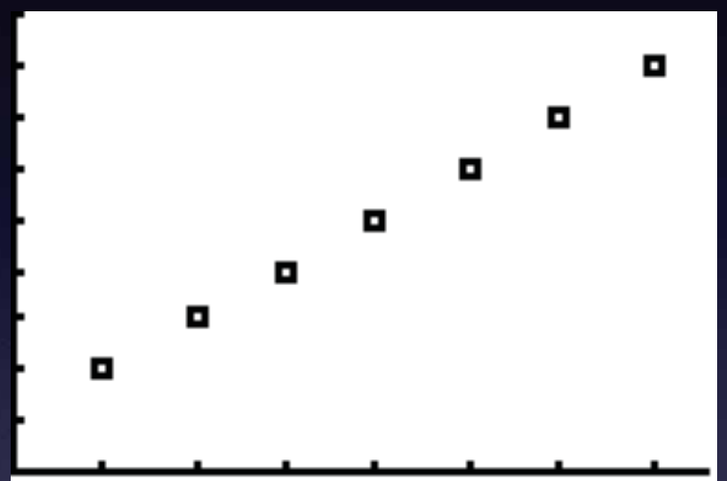
Price	Quality Rating
35	65
20	61
30	60
40	55
50	54
23	47
30	47
18	43
40	42
28	41
20	40
25	32



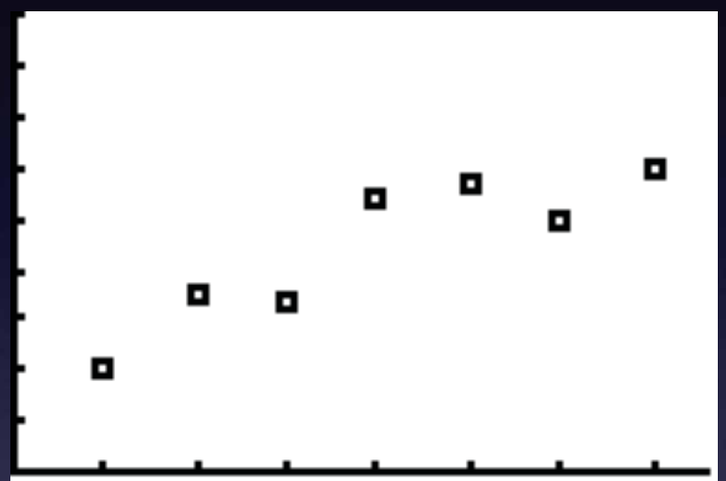
$$r = 0.3034$$

So how does *r* look on a graph?

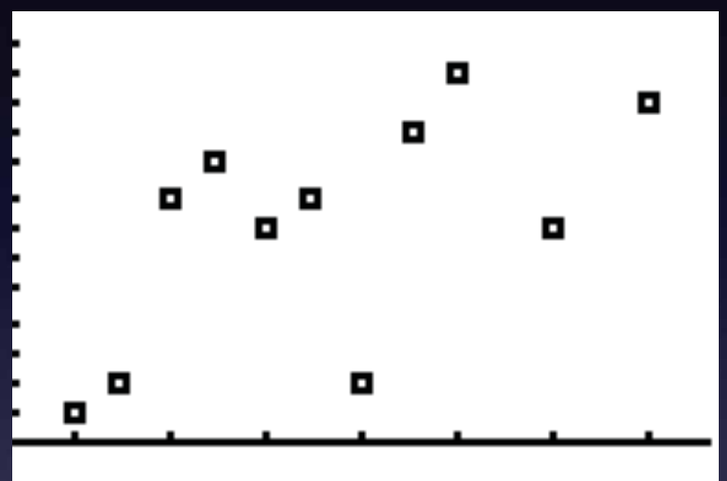
$r = 1$



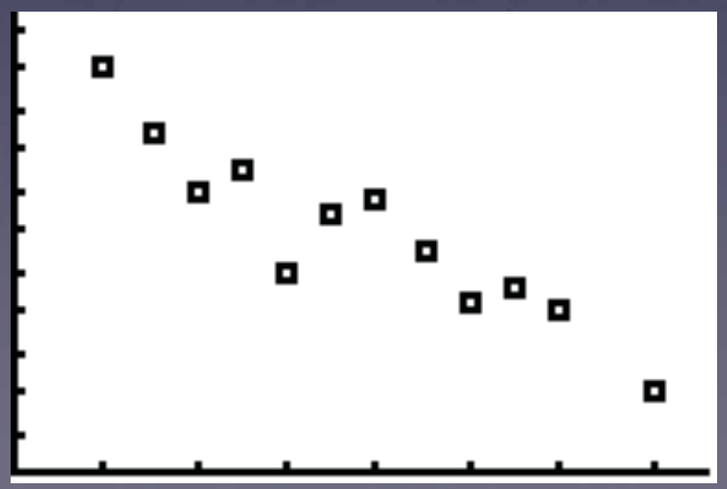
$r = 0.89$



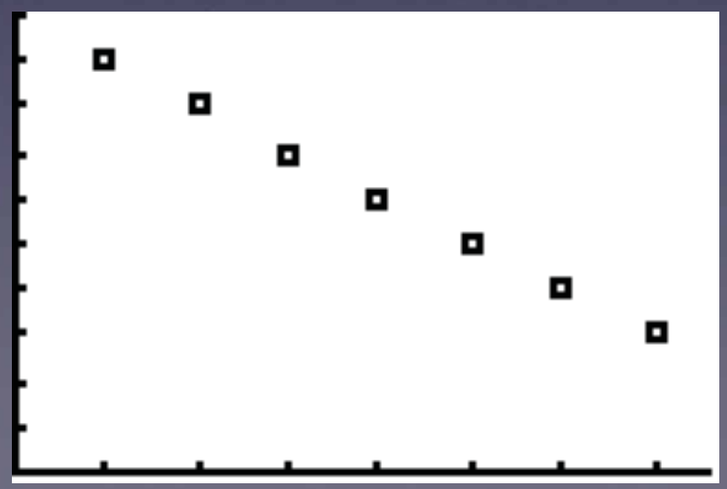
$r = 0.65$



$r = -0.92$



$r = -1$



So remember

r - Correlation Coefficient

measures the **strength** and **direction** of the linear relationship between two **quantitative** variables

Don’t forget to turn your Diagnostic On