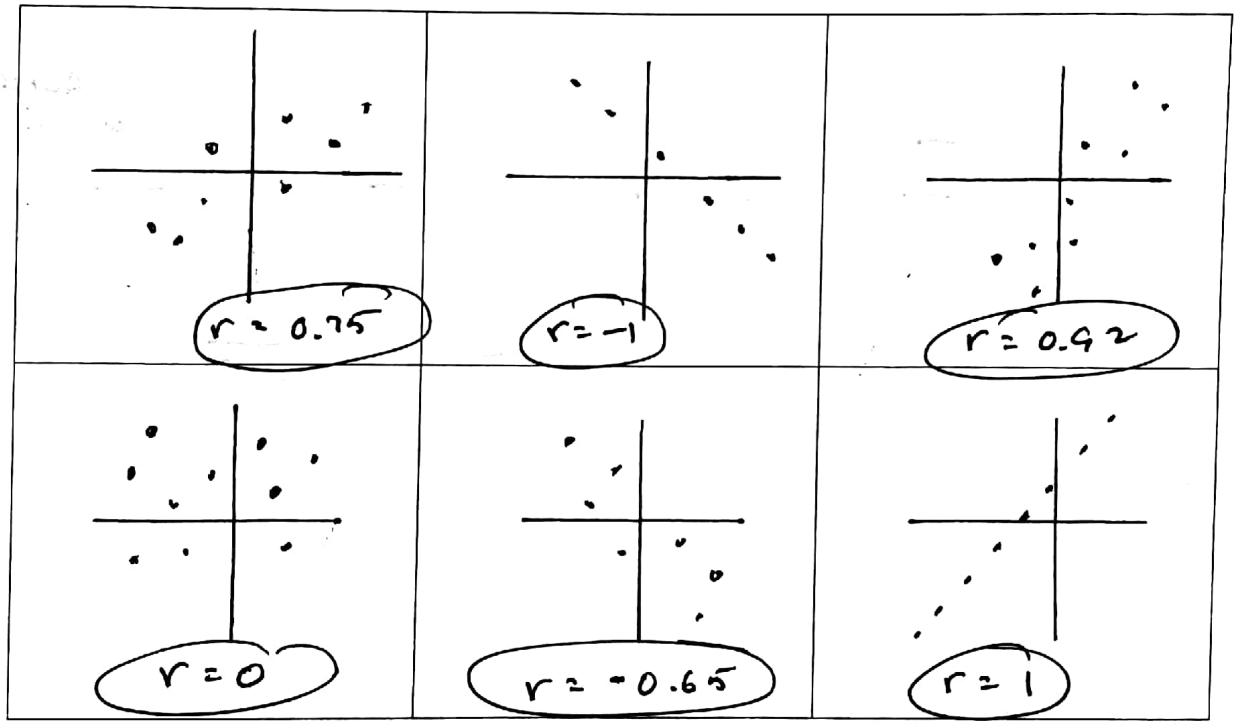
Correlation Coefficient. Another way of determining how good a linear model fits the data is called the **correlation coefficient**. It is denoted by the letter **r**. r measures **how linear a data set is** and has the following characteristics:

1. THE SIGN OF r IS THE SAME AS THE SLOPE OF THE LINE.
2. VALUES RANGE FROM -1 TO 1.
3. WHEN r IS CLOSE TO ± 1, THERE IS STRONG ASSOCIATION BETWEEN THE VARIABLES (NOT CAUSATION).
4. $r = 0 \Rightarrow$ NO LINEAR ASSOCIATION.



To display the r value on the calculator when performing a linear regression, the calculator Diagnostics must be turned on. To do this:

Press **[2nd] [0]** which is CATALOG. Then either press **[D]** and then scroll to **DiagnosticOn** and press **[ENTER] [ENTER]** or scroll from the top to **DiagnosticOn** and press **[ENTER] [ENTER]**.

As long as the Diagnostics are turned on, whenever regression is performed the calculator will provide an r value.

Example. Compute the r value for the New York City example. Interpret the r value in context of the problem.

$r = 0.999 \Rightarrow$ VERY STRONG LINEAR ASSOCIATION BETWEEN HOUSE PRICE & DISTANCE FROM NYC.

Applications

1. The following data set represents the yearly movie attendance (in billions) from 2001 to 2008

Year	2001	2002	2003	2004	2005	2006	2007	2008
Attendance	1.44	1.60	1.52	1.48	1.38	1.40	1.40	1.36

1. Enter the data into your calculator and graph it in a scatter plot. Sketch the graph below. Be sure to label the axes.

ROUGH NEGATIVE LINEAR RELATIONSHIP.

2. How would you describe the graph? What type of association do you see?

3. Use your calculator to determine the equation of the line of best fit. Write your equation below and clearly label your variables.

$\widehat{ATT} = -0.02x + 49.7$

4. What is the real life meaning of the slope of the line?

$\frac{RISE}{RUN} = \frac{ATT}{YR} = \frac{-0.02}{1}$ WE EXPECT FOR EVERY YR, ATTENDANCE WILL ↓ BY 0.02 BILLION PEOPLE.

5. What is the real life meaning of the y-intercept of the line?

PREDICTED ATT. IN YEAR 0 WAS 49.7 BILLION. (NO REAL MEANING)

6. What is the predicted attendance for the year 2010? Show your set up.

$-0.02(2010) + 49.7 \approx$ 1.31 BILLION PEOPLE

7. What year does the equation predict the attendance will surpass 1.8 billion? Round to the nearest year.

$1.8 = -0.02x + 49.7$

$-49.7 \quad -49.7$

$-0.02 \quad -0.02$

USE TBL

1988 1.8443

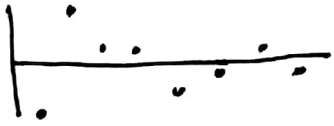
1989 1.8202

1990 1.7962

8. What is the correlation coefficient for the model? What does this say about the data?

$r = -0.72$ MODERATELY NEGATIVE LINEARITY.

9. Use your calculator to construct a residual plot. Sketch the graph below.



NOT AS RANDOM AS WE HOPE.

10. What does the plot tell you about how good the model is?

PROBABLY NOT THE BEST MODEL.

2. The following data set represents the number of stores a company opened from 2000 to 2008

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008
Stores	241	311	396	519	629	721	809	888	971

11. Enter the data into your calculator and graph it in a scatter plot. Sketch the graph below. Be sure to label the axes.

12. How would you describe the graph? What type of association do you see?

LINEAR + POSITIVE

13. Use your calculator to determine the equation of the line of best fit. Write your equation below and clearly label your variables.

$\hat{y} = 94.65x - 189069$

$\hat{y}$ = PREDICTED # OF STORES
x = YEAR

14. What is the real life meaning of the slope of the line?

FOR EACH ADD'L YEAR, IT PREDICTS ABOUT 95 MORE STORES.

15. What is the real life meaning of the y-intercept of the line?

IN YR 0, PREDICTED TO HAVE 189,069 STORES.
(NO REAL MEANING)

16. What is the predicted number of stores for the year 2012? Show your set up.

$\hat{y}(2012) = 94.65(2012) - 189069 = \boxed{1366 \text{ STORES}}$

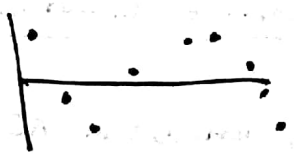
17. What year does the equation predict the 1500 stores will be open? Round to the nearest year.

2013 – 1461
2014 – 1555
} TABLE ⇒ $\boxed{2014}$

18. What is the correlation coefficient for the model? What does this say about the data?

$r = 0.998$ STRONG (+) ASSOCIATION

19. Use your calculator to construct a residual plot. Sketch the graph below.



20. What does the plot tell you about how good the model is?

PATTERN ⇒ MODEL IS BAD.