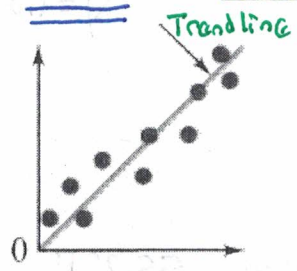
5.1 – Linear Regression

scatter plot → a graph that relates two groups of data using ordered pairs (x, y) (or points)

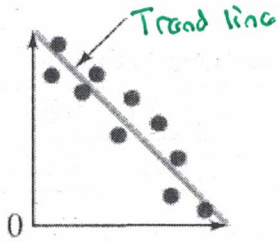
linear regression → a trend line that shows the relationship between two sets of data. that can be used to make PREDICTIONS shown in a scatter plot

- may also ask to find "line of best fit" or data has a "linear" relationship
- Must put data in calculator in list 1 (L_1 / x) and list 2 (L_2 / y) and create a scatter plot [STAT]
- Equation for linear regression (using the calculator) → $y = ax + b$ ($y = mx + b$)
  where you fill "a" part and "b" part with numbers that the calculator produces
- After finding linear regression equation → will also state the equation's correlation coefficient "r"
  The closer "r" is to either −1 or 1, then the **stronger the correlation** (points will cluster together)



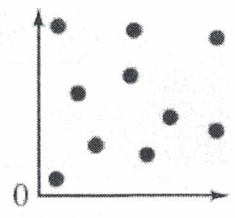
Positive Correlation

r value closer to 1



Negative Correlation

r value closer to −1



No Correlation

r value closer to 0

* No relationship between data set.

Examples: Complete each problem.

1.)

x	y
1	21
2	15
3	12
4	9
5	7

a.) State the correlation coefficient: −.978240074

b.) State the line of best fit: $y = -3.4x + 23$

c.) Is the model a good fit? (Yes) NO Explain: r value is close to −1

d.) Using the regression equation, find y when x = 18 $y = -38.2$

So when given "x"; use table to find y!

e.) Using the regression equation, find x when y = 46.8 $x = -7$

So when given "y", use the "Calc" function. Put y into Y_2=
(#5: Intersection)

2.)

Recreation Expenditures

Year	Dollars (billions)
1993	340
1994	369
1995	402
1996	430
1997	457
1998	489
1999	527
2000	574

a.) Circle the correlation coefficient that represents the given data:

A.) r = – 0.99645 B.) r = 0.99745 (C.) r = 0.99645 D.) r = 0.98645

b.) Circle the equation that BEST represents the given data:

A.) $y = \frac{96}{3}x - 64105$ (B.) $y = \frac{97}{3}x - 64105$ C.) $y = \frac{97}{3}x + 64105$

$y = 32x - 64105$ $y = 32.333333 - 64105$

c.) What amount of money will recreation earn in the year 2011? $917.33 billion

$x = 2011$

$y =$

3.) **Olympic 500-Meter Men's Gold Medal Speed Skating Times**

	0	4	8	12	14	18
Year	1980	1984	1988	1992	1994	1998
Time (seconds)	422	432	404	420	395	382

2013 − 1980 = 33 → X = 2013, Y = ?

a.) Find the linear regression model that fits this data where let x (L1) = numbers of years since 1980

$y = -2.292168675x + 430.560241$

b.) Predict the amount of time in the year 2013.

354.92 seconds

4.) Which of the following tables represents a direct variation? If it is direct variation, state the equation.
Note: Direction Variation equation is represented by the power function $y = k \cdot x^p$ where p is a positive number

a.)

x	y
−2	3.2
1	2.4
4	1.6

NO

$y = -.2666666667x + 2.666666667$

$y = kx$

b.)

x	y
−2	1
3	6
8	11

NO

$y = 1x + 3$

c.)

x	y
4	6
8	12
10	15

YES

$y = 1.5x$

5.) **Average Temperatures in Northern Latitudes**

Latitude (° N)	0	10	20	30	40	50	60	70	80
Temp. (°F)	79.2	80.1	77.5	68.7	57.4	42.4	30.0	12.7	1.0

a.) What is the average temperature for New Bern , NC (Latitude line is 35.1083° N)? 55.2° F

X = 35 Y = ?

b.) What is the Latitude line for a city that has an average temperature of 33.5°F? 55.45° N

X = ? Y = 33.5

6.) Your height and arm span have a linear relationship.
I've selected 6 students from the class and measured each person's height and arm span in inches.
Find the linear regression equation to answer each question.

Student's Name	Height (inches)	Arm Span (inches)
Laura Jenn	62	63
Ny Benjamin	68	71
Cypress Borden	71	72
Logan Desmarais	58	56
Tia Chapman	67	67
Patrick Laureigh	72	72

$y = 1.151376147x - 9.541284404$

a.) Using the regression equation, predict the height of a person with 5 ft arm span. 60.4 in tall

X = ? Y = 5 ft → 60 in

b.) Using the regression equation, predict the arm span of a person who is 7'4" tall. 91.8 in wing span

X = 7'4" Y = ?

↓

88"