Math 1311
Section 3.4
Linear Regression

In the last section, we used a linear function to model data that had a constant rate of change. In many real world situations, the rate of change may not be exactly constant, but the data falls approximately along a line. Linear regression is a way to find the "best fit" line that models a situation. This linear model can then be used to estimate the function values which are not in our data sample, or extrapolate to what may happen in the near future.

The skills for this section are:

1. Defining a function for the data including a careful choice of our "x" variable so the function is easy to interpret.
2. Entering our data into the STAT lists.
3. Plotting our data points
4. Use the STAT CALC menu to find the equation of the regression line.
5. Add the regression line to our graph.
6. Interpret the results.

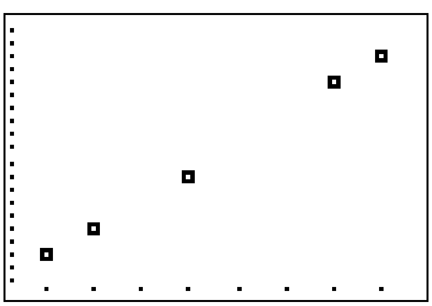
Note that skills 1-3 are the same ones from section 3.3, so the only new part is calculating the regression line and interpreting it.

Let's use a simple set of data first.

Example 1: Plot the data below and find the linear regression line in the form y=ax + b. Graph this line on our data plot and describe the results.

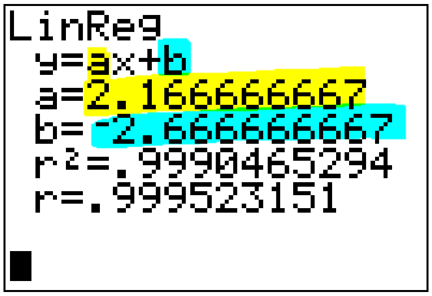
x	3	4	6	9	10
y	4	6	10	17	19

If we look at a plot of the data, we get:



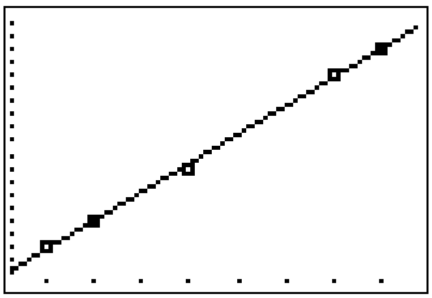
This shows that the data could be well modeled with a linear function.

So if we then find the linear regression for the data we get:



$y = 2.17x - 2.7$

$y = 2.1\overline{6}x - 2.\overline{6}$
Adding this line to the graph of the data, we get:



We see that the model fits the data quite well.

Now, let's add some meaning to our data.

Example 2: The table below shows enrollment, in millions of people, in public colleges in the US from 1996 to 2000. Plot this data in the graphing window. Find the regression line model for this data.

L1 1993 11.2 L2

x	Date	Enrollment in millions
0	1996	11.1
1	1997	11.2
2	1998	11.4
3	1999	11.5
4	2000	11.6

6 2002

x = # of years since 1996

$y = .13x + 11.1$

Explain the meaning of the slope of the regression line.

slope = .13

Enrollment increases by .13 million student per year

Express, using function notation, the enrollment in the US public colleges in 2002. Estimate its value using the regression model.

$E(6) = .13(6) + 11.1 = 11.88$ million of students.

Enrollment in US colleges in 1993 was 11.2 million people. Does it appear that the trend established in the late 1990's was valid as early as 1993?

NO

Example 3: The price of tickets to a certain college's home football games is given in the table below for various years.

a. Show that the table does not represent a linear function.
b. Let x = number of years since 1980, look at a scatterplot and determine if a linear model is appropriate, and then find the linear regression model.
c. Estimate the price in 1984 and in 2004. Which of these estimates is more likely reliable? Explain.

x	Year	Price
1	1981	10
5	1985	13
9	1989	16
14	1994	20
20	2000	25

1984 →

2004 →

x = # of years since 1980

$\frac{13-10}{5-1} = \frac{3}{4} = .75$

$\frac{16-13}{9-5} = \frac{3}{4} = .75$

$\frac{20-16}{14-9} = \frac{4}{5} = \underline{.8}$ not linear

(b) $P(x) = .79x + 9.07$

(c) 1984 → $x = 4$ $P(4) = .79(4) + 9.07 = \$12.23$

2004 → $x = 24$ $P(24) = .79(24) + 9.07 = \$28.03$