

Name: \_\_\_\_\_

Band: \_\_\_\_\_

Algebra II | Packer Collegiate Institute | 2008-2009

Worksheet on Linear Regressions/Line of Best Fit

I want you to collect and analyze data.

We are going to compare the number of times a single-word entry was viewed in Wikipedia in October 2008 with the number of Google hits that word gets.

**Collecting the data**

Date you're collecting the data: \_\_\_\_\_

Come up with 10 words which have Wikipedia entries (e.g. "monkey"). Don't pick phrases (e.g. "American revolution")! Write them in the first column of the table below.

To find the number of times an article is viewed on Wikipedia in October 2008, please go to <http://stats.grok.se/en/200810/word>. So, for example, to see how many times the "monkey" entry was viewed, type: <http://stats.grok.se/en/200810/monkey>

**Wikipedia article traffic statistics**

[monkey](#) has been viewed 122694 times in 200810. This article ranked 3324 in traffic on en.wikipedia.org.

To find the number of Google hits that a word gets, simply type the word in Google. The approximate number of hits will appear.



<b>Word</b>	<b>L<sub>1</sub>: Times viewed on Wikipedia in October 2008</b>	<b>L<sub>2</sub>: Number of Google Hits</b>
EXAMPLE: monkey	EXAMPLE: 122,694	EXAMPLE: 140,000,000

## Analyzing the Data

Before doing any analysis, answer the following questions:

(1.) I anticipate the data to be [ POSITIVIELY                      NEGATIVELY                      NOT ] correlated because...

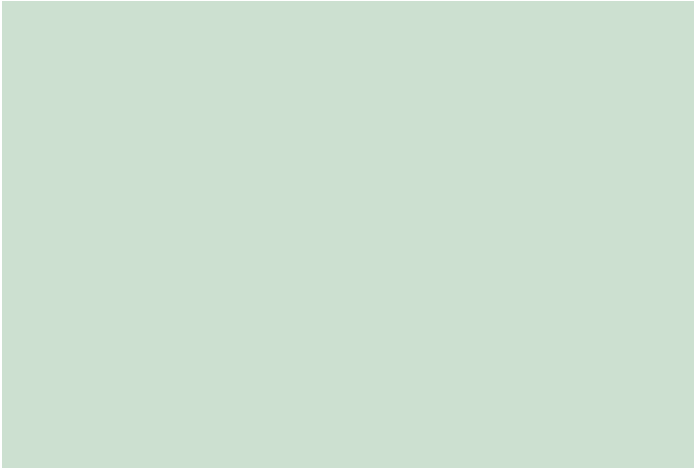
(2.) I anticipate the data to be [STRONGLY LINEAR   WEAKLY LINEAR   NOT LINEAR] because...

Now enter the data into  $L_1$  and  $L_2$  in your calculator.

The window to view this data is:

Xmin:	Ymin:
Xmax:	Ymax:
Xscl:	Yscl:

Plot the data on your calculator. Sketch what you see on your calculator:



The line of best fit is:

The correlation coefficient ( $r$ ) is:

Now that you've analyze the data:

(1.) The data is [ POSITIVIELY                      NEGATIVELY                      NOT ] correlated.  
(2.) The data is [STRONGLY LINEAR   WEAKLY LINEAR   NOT LINEAR]

## Using Your Linear Model To Predict

If the word “pelican” was viewed in Wikipedia 20,533 times in October 2008, predict – using your line of best fit – how many hits you expect it to receive in Google:

As of November 24, 2008, “pelican” has about 17,600,000 Google hits. Did your linear model over- or under-estimate the number of Google hits? (And by a lot or a little?)

Were you *interpolating* or *extrapolating* when you were using your model to analyze “pelican”?

Overall, do you think your linear model is good? Why or why not?

Finally, list one interesting thing you noticed about your data.