

Investigating Linear Regression: Do Numbers Give You the Whole Picture?

Kenn Pendelton

CALCULATORS: Casio: *fx-9750G Plus* • Casio: *CFX-9850G Series*

Teaching Notes/Lesson Plan

Setting the Background

Six data sets are given on the addendum sheet. You may wish to enter those data sets in your master calculator Lists as follows.

Data Set	File Drawer	X-List	Y-List
1	1	L1	L2
2	1	L3	L4
3	1	L5	L6
4	2	L1	L2
5	2	L3	L4
6	2	L5	L6

Only one data set should be given to each group of students. Transfer a File Drawer and delete the Lists that the group will not use. So that all groups can use the same graph location (GPH1 set to use List 1 for x and List 2 for y), copy the X-list to List 1 and the Y-list to List 2 if they are not already there. To copy the X-list: **MENU; LIST(4); cursor to the title of List 1; OPTN; LIST (F1); List (F1); enter the number of the List in which the X-list is found; EXE** . Follow the same process to copy the Y-list in List 2. Alternately, each group can set up and select the graph appropriate for using the lists in which their x - and y -lists are found.

Have each group of students find the linear regression equation ($y = ax + b$) and the correlation coefficient (r) for their set of data. The calculators should be set up to use List 1 for x and List 2 for y . **MENU; STAT(2); CALC (F2); REG (F3); X (F1)**; the linear equation is the first degree equation). Copy a , b , and r rounded to hundredths.

When members of a group agree on the results, have one member of each group write the number of the data set, the regression equation, and the correlation coefficient on the board. A brief discussion is likely to ensue. You may also want students to write a few sentences about what they think may be true about the data sets.

Next, have the groups sketch the scatter plot and the graph of the regression equation. **EXIT; EXIT; GRPH (F1); GPH1 (F1)** (the scatter plot is shown); **X(F1)** (to graph the linear regression equation); **DRAW (F6)**. After consensus has been reached, have one member of each group sketch the scatter plot and the graph of the regression equation beneath the information written previously.

Investigating Linear Regression: Do Numbers Give You the Whole Picture?

continued

CALCULATORS: Casio: *fx-9750G Plus* • Casio: *CFX-9850G Series*

Teaching Notes/Lesson Plan

All data sets produce the exact same linear regression equation and correlation coefficient when the values are rounded to hundredths, but the scatter plots are all quite different.

Discussions can ensue around the following points: for which sets is a linear model appropriate/not appropriate; would the residuals for data set 2 suggest a more appropriate model even if the graph were not shown; do there appear to be highly influential points in any of the data sets; what would be the prediction for a , b , and r , if the influential point was removed from the data set. The last point can be studied by having the group whose data set is in question delete the point from the set, recompute the values of a , b , and r (which is impossible for data set 4), and resketch the scatter plot and graph of the regression equation.

The following points should be made in summation.

- Numbers alone are not sufficient to determine whether a linear regression equation is an appropriate model for given data. Look at the whole "Picture."
- The residuals can sometimes provide a clue as to whether the linear regression equation is the most appropriate model.
- Elimination of an influential point can cause an increase OR decrease on the values for the correlation coefficient, the slope of the regression equation, and/or the y -intercept.

Investigating Linear Regression: Do Numbers Give You the Whole Picture?

continued

CALCULATORS: Casio: *fx-9750G Plus* • Casio: *CFX-9850G Series*

Teaching Notes/Lesson Plan

Addendum: Regression Data Sets

Data Set #1	
List 1 (x)	List 2 (y)
4	4.765
5	5.69
6	7.406
7	5.505
8	6.417
9	7.586
10	7.869
11	8.017
12	10.295
13	8.728
14	10.325

Data Set #2	
List 1 (x)	List 2 (y)
4	3.535
5	5.143
6	6.289
7	7.074
8	7.729
9	8.238
10	8.699
11	8.929
12	9.034
13	9.041
14	8.842

Data Set #3	
List 1 (x)	List 2 (y)
4	5.26
5	5.66
6	6.06
7	6.46
8	6.86
9	7.26
10	7.66
11	8.06
12	8.41
13	11.65
14	9.275

Data Set #4	
List 1 (x)	List 2 (y)
8	5.764
8	6.231
8	6.494
8	6.73
8	6.83
8	7.03
8	7.422
8	7.072
8	8.037
8	8.5
19	12.517

Data Set #5	
List 1 (x)	List 2 (y)
4	3.995
4.5	5.939
5	4.235
5.5	6.649
6	5.112
6.5	7.532
7	5.649
7.5	8.004
8	6.4135
8.5	8.059
20	12.702

Data Set #6	
List 1 (x)	List 2 (y)
4	6.2
4.5	7.1
5	7.1
5.5	6.7
6	5.7
6.5	5.2
7	5
8	5.1
9	6.38
10.5	7.5
25	16.55