

Activity 3: Exploring Least-Squares Regression

The purpose of this investigation is to analyze three different data sets, compute the correlation coefficients, and in each case plot the scatterplot and the least-squares regression line. Draw the graphs by hand and use your calculator to complete this assignment. The tables below present three sets of two-variable data prepared by the statistician Frank Anscombe to illustrate the dangers of calculating without first plotting the data.

Data Set A:

x	10	8	13	9	11	14	6	4	12	7	5
y	8.04	6.95	7.58	8.81	8.33	9.96	7.24	4.26	10.84	4.82	5.68

Data Set B:

x	10	8	13	9	11	14	6	4	12	7	5
y	9.14	8.14	8.74	8.77	9.26	8.10	6.13	3.10	9.13	7.26	4.74

Data Set C:

x	8	8	8	8	8	8	8	8	8	8	19
y	6.58	5.76	7.71	8.84	8.47	7.04	5.25	5.56	7.91	6.89	12.50

Carefully complete each of the following:

1. Using **three separate pieces of graph paper**, draw a careful scatterplot of each set of data. Use the top half of the paper for the graph and use the same scale for each graph.
2. For each set of data, calculate the least squares regression equation and the correlation coefficient (round each to two decimal places). Record this information below the corresponding graph.
3. Carefully add the least squares regression line for each data set to its corresponding graph.
4. Below each graph, discuss how well the least-squares regression line actually fits the data.
5. On a **separate (fourth) piece of paper**, answer the following questions:
 - a) In which of these three cases would you be willing to use the fitted regression line to predict y given that $x = 14$? Explain your answer.
 - b) What is the most important statistical message that is being reinforced by this investigation? Explain your answer.