Fitting a Line to Data

1. A large company is expanding its workforce and needs to hire some new administrative assistants. The company wants to know what the relationship is between the amount of experience that its current administrative assistants possessed when they were hired and their starting salary with the company. The data for ten randomly selected administrative assistants is given below.

Experience (in months)	Starting Salary (in $1000)
4	25
13	34
3	22
0	21
10	33
7	27
22	38
15	35
5	26
20	36

   a)   What kind of the relationship would you expect between length of experience and starting salary for the randomly selected administrative assistants.

   b)   Construct a scatterplot of the data. Which variable should be the independent variable and which is the dependent variable? Remember to include scales and labels for your axes.

   c)   Does the scatterplot confirm your description in part (a)? Explain your answer.

   d)   Suppose you wanted to hire an administrative assistant who had 17 months experience. Predict what the starting salary would be. Describe how you used the scatterplot to help you. How does your prediction compare with other student's predictions?

   Since there seems to be an almost linear relationship between starting salary and amount of job experience in months, a line can be drawn on the scatterplot to summarize this relationship. Such a line helps us to predict the value of the variable on the vertical axis (the dependent variable) from the value of the variable on the horizontal axis (the independent variable).

   On your scatterplot in part (b), draw a line that you think summarizes or *fits* the data. This line should go through the middle of the set of data points.

e)   Use the line you drew to predict the approximate starting salary for an assistant with 12 months experience.

f)   According to the data, the actual salary of an assistant with 12 months experience was $33,000. How close was your prediction? Calculate this by subtracting the predicted value from the actual value. This difference in called a <u>residual</u>. Did you over-predict or under-predict your estimate? Justify your answer.

g)   Compare your prediction to those made by others in your class. Who was the closest to the actual value? What can we say about the residual of the person with the closest prediction?

h)   Look at the line drawn by the person who had the smallest residual value. Do you think that this person had the line that best fits the data? Explain your answer.

i)   Instead of using the graph of the fitted line to predict a starting salary based on experience, you can use an equation for the line to do it. Pick two ordered pairs on the line you drew on the scatterplot of length of experience and starting salary. Use your ordered pairs to write an equation of the line. Write your answer in the form of $y = mx + b$.

j)   What does the variable y represent in your equation? What does the variable x represent in your equation?

k)   What is the slope of this line? Interpret the slope in the context of this problem.

l)   Use your equation from part (i) to predict the starting salary for an assistant who has 17 months experience.

m)   Compare your prediction from part (l) to your estimate from the scatterplot in part (d). Are they reasonably close? Do you think having the equation for the fitted line makes it easier to predict a value? Explain your answer.