Exploring Linear Regression Part III – Residuals and r^2

In Part I, you found the least squares regression line for netbook weights compared to length of battery life.

In Part II, we want to determine if the line of best fit is a good model for the data.

1. Open the document **Netbook** from the Statistics folder that you saved in the previous activity.
2. Insert a Calculator Page → ctrl I, *Add Calculator Page*.
3. On this page you will calculate the predicted values of the battery life given the netbook weight. Press var and select *stat.RegEqn()*. Type **weight** in the parentheses, arrow outside of the parentheses and press ctrl var and type **prdctbattery.** You should get the line *stat.RegEqn(weight)→prdctbattery*.
4. Insert a Lists & Spreadsheet page → ctrl I, *Add Lists & Spreadsheet*.
5. In the heading for Column A press var, and select **prdctbattery**. Press enter and the column should fill with the predicted battery life values based in the netbook weights. List these values on the worksheet in the $\hat{y}_i$ column with the netbook weights and battery life values from the data set.

How do these values compare with the actual battery life for the given weight?

6. The next calculation to perform is to find the **residual**. Calculate these values in the List & Spreadsheet by naming a new list **residual** and in the formula bar, press enter**.** Use the var menu to select the variables **battery** and **prdctbattery** separately. Enter the expression **battery - prdctbattery** in the formula bar. Press enter and the values of the residuals will appear in the list. Enter these values for the residual list in the table below.

What do these values tell you? Be specific.

7. Insert a new Data & Statistics page and plot **weight** on the horizontal axis and **residual** on the vertical axis.

Describe this graph. What does this tell you about the values of battery life and the y-values of the line of best fit?

8. In the table and the List & Spreadsheet, calculate **residual^2**. Name this column **residualsq**. Add these values together.

Write the sum of the residuals^2 here.

9. On your scatterplot of **weight** vs. **battery**, click on the line of best fit. Then press (menu), *Analyze*, *Residuals*, *Show Residual Squares*.

Describe what happened. How does the *sum of squares* compare with your answer to #8? Is this what you expected? Explain.

Now we need to look at the differences between the battery life values and the mean battery life. What does this difference tell us?

10. In your spreadsheet containing **weight** and **battery**, use these lists and the statistics variable $\overline{y}$ to calculate $(y_i - \overline{y})$ and $(y_i - \overline{y})^2$. Name the lists **variation** and **variationsq**. Complete the table below using these values.

Add all of the values in the $(y_i - \overline{y})^2$ column. What does this value mean?

Take sum of the residuals^2 and divide it by the sum of $(y_i - \overline{y})^2$. Record this value below. What does this <u>percentage</u> mean?

Calculate 1 – <u>percentage</u> (from above) **and explain what it means.**

Based on the value you just calculated and the plot of the residuals, how well does the model fit the data?

Netbook Weight	Battery Life	$\hat{y}_i$	Residual	Residual^2	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
2.8	6.00					
2.9	7.75					
2.8	7.25					
2.4	5.50					
2.9	8.25					
2.9	9.50					
2.9	7.75					
2.7	7.75					
2.8	8.00					
2.8	7.00					
3.2	6.50					
2.7	6.25					
2.9	5.25					
2.7	5.00					
2.6	3.75					
2.8	4.50					
2.4	2.75					
2.4	2.75					
2.6	2.50					
2.5	2.50					
2.5	2.00					
2.5	2.75					