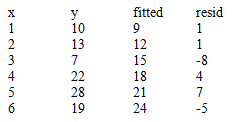
Practice Problems: Residual Analysis

The least squares estimate from fitting a line to the data points in residuals.txt are and (you can check this claim, of course).

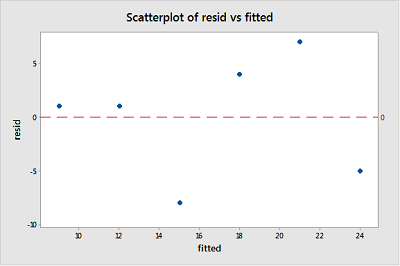
1. Copy the data into, say, columns C1 and C2 of a Minitab worksheet.
2. Using the least squares estimates, create a new column that contains the predicted values, , for each – you can use Minitab’s calculator to do this. Select Calc >> Calculator… In the box labeled “Store results in variable,” specify the new column, say C3, where you want the predicted values to appear. In the box labeled “Expression,” type C1. Select OK. The predicted values, , should appear in column C3. You might want to label this column “fitted.” You might also convince yourself that you indeed calculated the predicted values by checking one of the calculations by hand.
3. Now, create a column, say C4, that contains the residual values – again use Minitab’s calculator to do this. Select >> Calc >> Calculator… In the box labeled “Store results in variable,” specify the new column, say C4, where you want the residuals to appear. In the box labeled “Expression,” type C2 – C3. Select OK. The residuals, , should appear in column C4. You might want to label this column “resid.” You might also convince yourself that you indeed calculated the residuals by checking one of the calculations by hand.
4. Create a “residuals versus fits” plot, that is, a scatter plot with the residuals () on the vertical axis and the fitted values () on the horizontal axis. (see Minitab Help Section – [Creating a basic scatter plot](https://onlinecourses.science.psu.edu/stat501/node/115)). Around what horizontal line (residual = ???) do the residuals “bounce randomly?” What does this horizontal line represent?

# Answers

Here are the data with fitted values and residuals:



And, here is a scatterplot of these residuals vs. fitted values:



Given the small size, it appears that the residuals bounce randomly around the residual = 0 line. The horizontal line where *resid* = 0 (red dashed line) represents potential observations would fall exactly on the fitted regression line.