Begin with the **NEA/Fat Gain** data (found on my website or in the eBook Example 4.16). We are interested in understanding the relationship or association between these two variables and we'll use JMP to help.

1. Draw a scatterplot (use Analyze -> Fit Y by X). Which variable is X and which is Y? Explain why.

2. Describe the association you see in this plot (use Direction, Strength, and Form as your guides…). Be specific, use the context of the problem – this problem is about non-exercise activity increase in people who have been overeating for 8 weeks and the amount of fat that they gain in that time. Don't forget the units of measure!

3. Compute the correlation coefficient r – there are at least two ways of doing this in JMP. I would suggest using the Fit Line option under the red triangle after you've plotted the scatterplot (even though you will need to take the +/- square root of the RSquare value that JMP gives you). How does this number help you describe the direction, strength, and form of this relationship? Explain.

4. If the form looks linear, use least squares to fit the prediction line (under the red triangle, choose Fit Line). Explain in some detail the meaning (in context) of the two numbers (slope and intercept) JMP computed and shows under Linear Fit.

5. What is the RSquare value for this analysis? Explain how this helps you make predictions with the LSRL.

6. Use JMP to do residual plots (under the red triangle next to Linear Fit) and describe any patterns you see (a good linear fit will have a residual plot that fits a normal distribution with mean = 0 – a good scatter of residuals around 0, no patterns when plotted against the original X values.). Do you see any outliers or influential points? If so, remove them (under Rows choose Exclude/Unexclude) and re-fit and describe the differences you see in the various plots and statistics you computed with the points included.

7. Now consider the other two datasets on my website (**manatee.jmp** and **gesell.jmp**) and repeat the same 1-6 analyses as above on those data. Don't forget the contexts: a. In the manatee data, the number of manatees killed by powerboats in Florida is one variable and the other is the number of powerboats registered in Florida (in thousands) for several years; b. In the gesell data, the age (in months) at which a child began to speak is recorded along with the same child's Gesell Adaptive Score on a standardized test, taken when the child was much older. **Notes:** 1. There are outliers and influential points in the **gesell** data so #6 will be more important than in the other two data sets! 2. What about Extrapolation in the **manatee** dataset to estimate the number of powerboats registered in Florida last year? Can you use this number to then predict the number of manatee deaths last year? Can you check your result by finding the actual numbers for last year?