Practice Set: Residual Analysis

## The Basic Idea (Continued)

In the practice problems in the previous section, you created a “residuals vs. fits” plot *by hand* for the data contained in [residuals.txt](https://onlinecourses.science.psu.edu/stat501/sites/onlinecourses.science.psu.edu.stat501/files/data/residuals.txt). Now, create a “residuals vs. predictor” plot, that is, a scatter plot with the residuals () on the axis and the predictor () values on the axis. (See Minitab Help: [Creating a basic scatter plot](https://onlinecourses.science.psu.edu/stat501/node/115)). In what way – if any – does this plot differ from the residuals vs. fits plot you obtained previously?

## Using residual plots to help identify other good predictors

To assess physical conditioning in normal individuals, it is useful to know how much energy they are capable of expending. Since the process of expending energy requires oxygen, one way to evaluate this is to look at the rate at which they use oxygen at peak physical activity. To examine the peak physical activity, tests have been designed where individuals run on a treadmill. At specified time intervals, the spped at which the treadmill moves and the grade of the treadmill both increase. The individual is then systematically run to maximum physical capacity. The maximum capacity is determined by the individual; the persons stops when unable to go any further. A researcher subjected 44 healthy individuals to such a treadmill test, collecting the following data:

* a measure of x=oxygen consumption, defined as volume of oxygen used per minute per kilogram of body weight.
* how long, in seconds, the individual lasted on the treadmill
* age, in years, of the individual

The data set [treadmill.txt](https://onlinecourses.science.psu.edu/stat501/sites/onlinecourses.science.psu.edu.stat501/files/data/treadmill.txt) contains the data on the 44 individuals.

1. Fit a simple linear regression model using Minitab’s fitted line plot treating as the response and as the predictor . (See Minitab’s Help Section: [Creating a fitted line plot](https://onlinecourses.science.psu.edu/stat501/node/116)). Does there appear to be a linear relationship between and ?
2. Fit a simple linear regression model using Minitab’s fitted line plot treating as the response and as the predictor . Does there appear to be a linear relationship between and ?
3. Fit a simple linear regression model using Minitab’s fitted line plot treating as the response and as the predictor . Does there appear to be a linear relationship between and ?
4. Now, fit a simple linear regression model using Minitab’s regression command treating as the response and as the predictor . In doing so, request a “residuals vs. age” plot. (See Minitab Help Section: [Creating residual plots](https://onlinecourses.science.psu.edu/stat501/node/118)). Does the “residuals vs. age” plot suggest that would be an additional good predictor to add to the model to help explain some of the variation in ?
5. Now, fit a simple linear regression model using Minitab’s regression command treating as the response and as the predictor . In doing so, request a “residuals vs. ” plot. Does the “residuals vs. ” plot suggest that would be an additional good predictor to add to the model to help explain some of the variation in ?
6. Summarize what is happening here.

# Brief Solutions

## The Basic Idea (Continued)

In what way – if any – does this plot differ from the residuals vs. fits plot you obtained previously?

The only difference between the plots is the scale on the horizontal axis.

## Using residual plots to help identify other good predictors

1. Yes, there appears to be a strong linear relationship between and based on the scatterplot and -squared
2. Yes, there appears to be a moderate linear relationship between and based on the scatterplot and -squared
3. Yes, there appears to be a moderate linear relationship between and based on the scatterplot and -squared
4. After fitting a simple linear regression model with as the response and as the predictor , the “residuals vs. ” plot does not suggest that would be an additional good predictor to add to the model to help explain some of the variation in since there does not appear to be a strong linear trend in this plot.
5. After fitting a simple linear regression model with as the response and as the predictor , the “residuals vs. ” plot suggests that could be an additional good predictor to add to the model to help explain some of the variation in since there is a moderate linear trend in this plot.
6. Of the two predictors, has a strong linear association with than . So, there is no benefit to adding to a model including . However, there is some benefit to adding to a model including . If you do this and fit a multiple linear regression model with both and as the predictors, then it turns out that is significant (at the 0.05 level) but is not. In summary, the “best” model includes just , but a model with and is better than a model with just