Case: Process

Story

The dataset process.txt contains measurements of air flow, water temperature, and acid concentration of a process loss. The aim of this case is to explain the process loss as a function of the other variables.

Data

Variable	Description
loss	loss from process
airflow	air flow
watertemp	water temperature
acidconc	acid concentration

Exercise

1. Plot the variables and make a graphical assessment. Which variables could be helpful in explaining process loss?
2. Using simple linear regression, assess whether air flow, water temperature and acid concentration have an influence on process loss.
3. Now use a multiple linear regression to assess the effects of air flow, water temperature and acid concentration on process loss.

   Notice what happens to the significance of the variables: One of the variables was borderline significant in the simple linear regression, but is not significant in the multiple linear regression. How do you explain this?
4. Determine a reasonable model for process loss based on the variables available. Use model diagnostics/residual analyses, transformation of variables, polynomials and interactions to aid your model search.
5. Perform model diagnostics on your final model and subject it to criticism.
6. Write up a mathematical expression summarizing your final model.
7. Summarize the significance of the variables included in the model.
8. Save the code that performs your analysis in a script and add plenty of comments to your code.