

Statistical analyses of flash point and boiling point data for organic compounds

Comments to Peter M. Murphy - lm9080@aol.com

Goal: Flash point and boiling point data for 248 compounds, arranged in a homologous series, are analyzed by various statistical methods in order to investigate the correlation between these two properties and also how each property varies with molecular structure.

Prerequisites: This exercise requires an introductory knowledge of chemistry, safety, and statistical data analysis including (i) generating a scatter plot, (ii) finding means, medians, standard deviations, and confidence intervals, (iii) carrying out least squares linear regression, and (iv) comparing means using the paired t-test. For further background in statistical methods, see "Online Statistics: An Interactive Multimedia Course of Study" which is an introductory-level statistics book at <http://onlinestatbook.com>. For an instructional classroom lesson on data analysis, a portion of the data in Tables 1 and 2 can be used to demonstrate the skills necessary to complete the exercise.

Resources you will need: This exercise can be carried out using either M.S. Excel worksheets or a statistical software package such as Minitab, SAS, or SPSS that is capable of data manipulation, graphic presentation of data, and statistical analyses.

Background:

The National Fire Protection Association (NFPA) provides vast fire safety information including the four-color diamond used on labels to indicate the hazard level of a chemical; see Figure 1. NFPA Health, Flammability, and Reactivity hazards are rated from 0 (none) to 4 (extreme). The Health rating is in the blue section, Flammability in red, and Reactivity in yellow. The white section is reserved for Other Specific Hazards including oxidizers, acids, corrosives, and radioactive materials. Since 1960, the NFPA labeling standard has been a simple, recognizable, and understandable labeling system that provides information regarding important hazards of a material and the severity of these hazards as they relate to handling, fire prevention, exposure, and control. The Flammability rating of an NFPA chemical label provides basic information to fire fighters and other emergency personnel, enabling them to decide whether to evacuate an area or to select the appropriate fire fighting tactics and emergency procedures. The NFPA labeling standard also provides laboratory personnel with the information necessary to select the appropriate level of personal protection when working with a material and the correct method of storage for that material. Proper use of chemical labeling should be common laboratory practice.

HAZARDOUS MATERIALS CLASSIFICATION



Figure 1 : NFPA Label Classifications

Flash point provides a simple and convenient index for the flammability and combustibility of substances. Flash point provides valuable information to those who handle, transport, and store chemicals. Flash point is the minimum temperature at which the vapor present over a liquid forms a flammable mixture when mixed with air. The NFPA Flammability rating is determined by the flash point of the material, with ratings of 4, 3, 2, or 1 for chemicals with flash points below 73°F, below 100°F, below 200°F, or above 200°F, respectively. The NFPA Flammability rating of 0 is reserved for substances that will not burn.

When assessing the safety and flammability of a material or mixture, experimental data on flash point is preferred. But, with so many new compounds being synthesized and mixtures being formulated every year, an estimate of flash point is often sufficient for a preliminary assessment of the flammability of these new materials. Estimates of the flash point of a compound can be based (i) on an assessment of its chemical structure and a comparison to the known flash point of similar chemical compounds or (ii) on a correlation to known physical properties of the material. The normal boiling point of a compound is generally known and this physical property can be used to estimate the compound's flash point. Linear correlations between flash point and boiling point generally obtain excellent correlation coefficients (R^2 between about 0.90 and 0.98) across many different functional group families and across a wide range in boiling point. This correlation is not surprising because a sufficient amount of vapor must be present to form a flammable mixture when in air, and boiling point correlates well to the minimum vapor content of the material necessary for a flammable mixture with air. Compared to the measurement of flash point, the measurement of boiling point is less risky, is inexpensive, uses simple equipment, and requires minimal training. A reliable correlation between boiling point and flash point can provide the information necessary for a preliminary assessment of the flammability of a material or mixture. Advanced QSPR models predict flash point for pure compounds and mixtures based on boiling point, specific gravity, heat of vaporization, and other strategies.

Experimental Data:

The link below (Table 1) contains flash point and normal boiling point data for 248 compounds. The data has been saved in a text-only-tab delimited format, which can be imported into most software packages. The data was gathered from "Yaws' 2003 Handbook of Thermodynamic and Physical Properties of Chemical Compounds" (online version available at <http://www.knovel.com>) and through Internet searches of publicly

available safety and physical property data. Flash point is generally reported in degrees Fahrenheit, because flash point is widely used in chemical shipping and storage by people who do not generally use Celsius for measuring temperature. The normal boiling points were converted to Fahrenheit so both physical properties could be compared in a common scale.

[Table 1: Flash point and boiling point data \(in Fahrenheit\) for 248 organic compounds](#)

Table 2 (below) contains flash points for a series of organic compounds organized by carbon chain length in rows (methyl, ethyl, n-propyl, phenyl, etc.) and functional (end) groups in columns (methyl, hydroxymethyl, chloromethyl, aminomethyl, carboxylic acid, etc.). Table 2 can be cut-n-pasted into the analysis software that you are utilizing.

[Table 2: Flash point data arranged in homologous series of increasing carbon chain length for various functional groups](#)

If using M.S. Excel, the "Data Analysis" functions necessary for these exercises are located in the "Tools" drop down menu. If these functions are not available, follow the instructions for the "Analysis ToolPak" add-in found in the HELP menu.

Exercise:

1. Create an X-Y scatter plot of flash point vs. boiling point for the compounds in Table 1. Perform a least-squares linear regression on the data (a) to obtain an equation for the correlation between these physical properties and (b) to determine the fraction of variance explained by the least-squares line, i.e. R^2 . Is the flash point or boiling point temperature consistently higher for these 248 compounds? Why or why not?
2. For comparing the flash point between two alkyl groups, select any two rows from Table 2. Compare the means and medians of the data for each row. Count the number of higher flash point values between the paired data for each row. Perform a paired sample t-test for the two sets of data. What is the 95% confidence interval for the population difference between the flash points for those two alkyl groups with any possible end group? Do these different statistical analyses agree on the effect of the two alkyl groups you chose on flash point?
3. For comparing the flash point for two functional (end) groups, select any two columns from Table 2. (Shifting the data may be preferred to obtain common carbon chain lengths between the functional prior to these analyses; for example to compare alkanes and 1-alkenes.) Compare the means and medians of the data for each functional (end) group. Count the number of higher flash point values between the paired data for each functional (end) group. Perform a paired sample t-test for the two sets of data. What is the 95% confidence interval for the population difference between the flash points for the two functional (end) groups with any possible alkyl chain? Do these different statistical analyses agree on the effect of the two functional (end) groups you chose on flash point?

Suggestions for improving this web site are welcome. You are also encouraged to submit your own data-driven exercise to this web archive. All inquiries should be directed to the curator: Tandy Grubbs, Department of Chemistry, Unit 8271, Stetson University, DeLand, FL 32720.

wgrubbs@stetson.edu

