

# Investigating Molecular Structure and the 1-Octanol-Water Partition Coefficient

Comments to Peter M. Murphy - [lm9080@aol.com](mailto:lm9080@aol.com)

**Goal:** The 1-octanol-water partition coefficient ( $K_{ow}$ ), which expresses the ratio of the concentration of a compound dissolved in 1-octanol to the concentration of that compound dissolved in water in an immiscible mixture of 1-octanol and water, is statistically analyzed for 196 organic compounds in order to assess how molecular structure is related to this property.

**Prerequisites:** This exercise requires an introductory knowledge of chemistry and statistical data analysis, including (i) generating histograms and (ii) finding means, medians, standard deviations, and confidence intervals. For further background in statistical methods, see "Online Statistics: An Interactive Multimedia Course of Study", an introductory-level statistics book" at <http://onlinestatbook.com> . For an instructional classroom lesson on data analysis, a portion of the data in Table 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 1-octanol-water partition coefficient ( $K_{ow}$ ) is the equilibrium ratio of the concentration of a material in 1-octanol to the concentration of that material in water in an immiscible mixture of 1-octanol and water. The published values of  $\log_{10}$  of  $K_{ow}$ , i.e.  $\text{Log}(K_{ow})$ , range from about -4 for saccharides to +10 for phthalate esters. Differences in  $\text{Log}(K_{ow})$  of 1.0 indicate a ten-fold difference in the equilibrium ratio of concentrations.  $\text{Log}(K_{ow})$  provides a quantitative data set for introducing students to the principles of partitioning of organic materials between two immiscible solvents, which leads to an understanding of solubility, of the hydrophilic-hydrophobic nature of materials, and of the separation of chemicals by extraction. The  $\text{Log}(K_{ow})$  of organic compounds is a thermodynamic property which is widely used for modeling the biological activity and environmental fate of organic chemicals, including estimating soil-water partition coefficients, dissolved organic matter-water partition coefficients, lipid solubility, the formation of micelles, transport across membranes, environmental risk assessment (including bioconcentration, aqueous toxicity, and biodegradation), and as a valuable tool in drug and pesticide design. The data in Tables 1 and 2 below will be used to analyze the effect on  $\text{Log}(K_{ow})$  of increasing the carbon chain length by  $-\text{CH}_2-$  . A commonly used rule-of-thumb is that  $\text{Log}(K_{ow})$  increases by 0.5 for each additional  $-\text{CH}_2-$  in a homologous series.

## Experimental Data:

Table 1 contains the published  $\text{Log}(K_{ow})$  values for 196 organic compounds, primarily from the LOGKOW© database; available on-line at <http://logkow.cisti.nrc.ca/logkow/index.jsp> ; provided by : Sangster Research Laboratories, P.O. Box 49562, CSP du Musée, 5122 Cote des Neiges , Montréal, Québec, Canada , H3T 2A5. In Table 1, the  $\text{Log}(K_{ow})$  values are arranged with rows of increasing alkyl chain length from hydrogen through 1-decyl and with columns for the various functional (end) groups. Many entries in Table 1 are blank because no reliable published  $\text{Log}(K_{ow})$  data was found for these compounds, though these compounds are certainly known.

Table 2 contains the difference in  $\text{Log}(K_{ow})$  values between 175 homologous pairs of compounds in Table 1. Click on the following link and save each data set on your computer prior to performing the statistical analyses in the exercise. Rearrange the data in Table 2 into a single column for data analysis.

[Tables 1 and 2 \(click to access data\)](#)

If you are using M.S. Excel, the "Data Analysis" functions 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. The 95% Confidence Interval for both the overall population and the population mean depend on the standard deviation, but the Confidence Interval for the population mean also depends the number of data points.

95% Confidence Interval for the population = mean +/- (1.96)(standard deviation)

95% Confidence Interval for the population mean = mean +/- (1.96)(standard deviation)/( $\sqrt{n}$ )

**Exercise:**

1. Graphical Representation. Construct a histogram for the data in Table 2. Are there any outliers?
2. Descriptive Statistics. What is the mean, median, range, and standard deviation for the increase in Log( $K_{ow}$ ) for the 175 homologous pairs of compounds represented in Table 2.
3. Inferential Statistics. What are the 95% Confidence Intervals of the increase in Log( $K_{ow}$ ) when the carbon chain length is increased by  $-CH_2-$  (i) for the population mean of all homologous pairs, and (ii) for all individual homologous pairs?
4. Compare the commonly used rule-of-thumb that Log( $K_{ow}$ ) increases by 0.5 for each additional  $-CH_2-$  in a homologous series with (i) the mean for the collection of pairs of compounds in Table 2, and (ii) the 95% Confidence Interval for the population mean of all possible compounds.

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](mailto:wgrubbs@stetson.edu)

