Homework 3: Statistics – Regression and Correlation

CHEM4050/5050 Fall 2025

Wednesday, September 24, 2025, at 11:59 PM Central

Contents

1 Trouton's Rule, Regression, and Uncertainty Analysis

Trouton's rule offers a valuable empirical insight: the entropy of vaporization (ΔS_v) for many non-associating liquids is approximately constant, around 85–88 J/mol·K (or $\sim 10.5R$), at their normal boiling points. This rule underscores the consistency in thermodynamic properties during phase transitions and provides a convenient approximation for estimating vaporization entropies when direct measurements are unavailable.

In this problem, you are given data on the boiling points (T_B) and enthalpies of vaporization (H_v) for 24 substances. Using this dataset, complete the following tasks:

1.1 Tasks

- ☐ **Fit a linear regression model**: Model the relationship between H_v and T_B using the equation

$$H_v = a \cdot T_B + b \tag{1}$$

  where a is the slope, and b is the intercept.

- ☐ **Interpret the slope**: The slope a can be interpreted as an approximation of the entropy of vaporization ΔS_v.
- ☐ **Compare to Trouton's Rule**: According to Trouton's rule, $\Delta S_v \approx 10.5R$ (or ~ 88 J/mol-K) for many substances. Discuss how well your data aligns with this approximation.
- ☐ **Compute uncertainty**: Calculate the 95% confidence intervals for both the slope a and the intercept b.

1.2 Submission Guidelines

- ☐ Push a Python script named `troutons_rule.py` to your GitHub repository titled `chem-4050-5050` for this course.
- ☐ When executed, `troutons_rule.py` should:
  - ☐ Import and use the functions `ols_slope`, `ols_intercept`, and `ols` from Lecture 7 for the ordinary least squares (OLS) regression.
  - ☐ Produce a clear plot of H_v vs. T_B with the fitted linear regression line:
    - ☐ Color data points by their `Class`.
    - ☐ Convert H_v to J/mol-K for interpretation and labeling.
    - ☐ Display the equation $H_v = a \cdot T_B + b$ on the plot, along with the numerical values of a (in J/mol-K) and b (in kJ/mol), including their 95% confidence intervals.
    - ☐ Title the plot as "Trouton's Rule."
  - ☐ Save the plot as a `png` file in a folder titled `homework-3-1`.
- ☐ Include clear comments in your code, explaining each key step.
- ☐ Ensure your plots are well-labeled, properly formatted, and saved to the directory `homework-3-1`.

2 Graduate Supplement

In the previous problem, you utilized a linear regression approach to model the relationship between enthalpy of vaporization (H_v) and boiling points (T_B) of various substances, deriving the entropy of vaporization (ΔS_v) as the slope of the regression line. In this graduate supplement, you will approach the same problem using numerical optimization techniques.

2.1 Tasks

- ☐ **Objective Function**: Define a suitable objective function based on the least squares error between the predicted and actual values of H_v. Specifically, minimize the sum of squared residuals

$$\text{Objective}(a, b) = \sum_{i=1}^{n} \left(H_v^{(i)} - (a \cdot T_B^{(i)} + b) \right)^2 \tag{2}$$

  where a and b are the variables to optimize.

- ☐ **Minimization using `scipy.optimize.minimize`**: Implement a script that uses `scipy.optimize.minimize` to find the optimal slope (a) and intercept (b) that minimize the least squares error.
- ☐ **Compare Results**: Compare the slope (a) from the optimization with the slope obtained from the linear regression in the first problem. How do the results differ, if at all?
- ☐ **Interpretation**: Discuss the implications of using an optimization-based approach versus a linear regression approach for this problem. Are there any notable advantages or disadvantages?

2.2 Submission Guidelines

- ☐ Push a Python script named `troutons_rule_optimization.py` to your GitHub repository titled `chem-4050-5050` for this course.
- ☐ When executed, `troutons_rule_optimization.py` should:
  - ☐ Define the objective function as described above.
  - ☐ Use `scipy.optimize.minimize` to find the optimal parameters a and b.
  - ☐ Plot the resulting fit of H_v vs. T_B using the optimized parameters:
    - ☐ Color data points by their `Class`.
    - ☐ Convert H_v to J/mol-K for interpretation and labeling.
    - ☐ Display the equation $H_v = a \cdot T_B + b$ on the plot, along with the numerical values of a and b.
    - ☐ Title the plot as "Trouton's Rule Optimization."
  - ☐ Save the plot as a `png` file in a folder titled `homework-3-2`.
- ☐ Ensure your plots are well-labeled, properly formatted, and saved to the directory `homework-3-2`.
- ☐ Include clear comments in your code, explaining each key step.