

Variance Of Solubility With Different Variable

Radhika Gupta
Roll no. 23GG61R14
Geology and geophysics

November 2023

1 Objective

To know the relationship between the solubility (predicted and observed) and different parameters such as Minimum Degree, Molecular Weight, Number of H-Bond Donors, Number of Rings, Number of Rotatable Bonds and Polar Surface Area.

2 Introduction

Solubility, a critical property in the realm of chemistry and pharmaceutical sciences, refers to the ability of a substance to dissolve in a particular solvent under specific conditions. Understanding the factors that influence solubility is paramount, as it plays a pivotal role in various industrial processes, drug development, environmental science, and material science.

In this report, we delve into the exploration of the relationship between solubility, both predicted and observed, and various molecular parameters. Our investigation focuses on key factors such as Minimum Degree, Molecular Weight, Number of H-Bond Donors, Number of Rings, Number of Rotatable Bonds, and Polar Surface Area. Employing a straightforward approach, we utilize simple scatter plots to visually represent and analyze the correlations between these parameters and solubility. By examining these fundamental molecular characteristics, we aim to uncover insights into the factors influencing solubility, providing a comprehensive understanding of the interplay between molecular structure and solubility outcomes.

3 Methodology

Data Collection

The dataset for this analysis was sourced from a GitHub repository, accessible through the link <https://raw.githubusercontent.com/GLA/processed.csv>.

- The dataset, named `ESOL_delaney-processed.csv`, contains information on molecular properties.
- Two specific columns were extracted for investigation: 'Molecular Weight' and 'Measured Log Solubility in Mols per Litre.'
- A linear regression model was employed to assess the relationship between Molecular Weight and Measured Log Solubility. We assumed a linear association between the two variables.

Plots for different parameters were plotted with respect to solubility criteria.

4 Discussion and Conclusion

The plots are too scattered for all the variables for predicted solubility. However, for parameters such as number of rings, Number of H bonds and for number of rotational bond the range of solubility decreases with increase in these values. Upper limit for these range decreases for Number of rings as well as for number of rotational bonds. So, we can say in general solubility decreases with increase in these parameter.

However the plot for solubility vs Molecular weight the observation is clear cut that the parameter shows the inverse relationship. From the regression line it is clear that predicted solubility is greater than measured one for the molecular weight greater than 160 whereas the measured is more than predicted solubility for molecular weight less than 160.

Linear regression stands as a versatile and foundational statistical technique, offering a straightforward yet potent framework for comprehending and modeling the connections among variables. With its ease of use, interpretive capabilities, and broad applicability, linear regression proves to be an invaluable instrument in academic research and practical problem-solving scenarios.

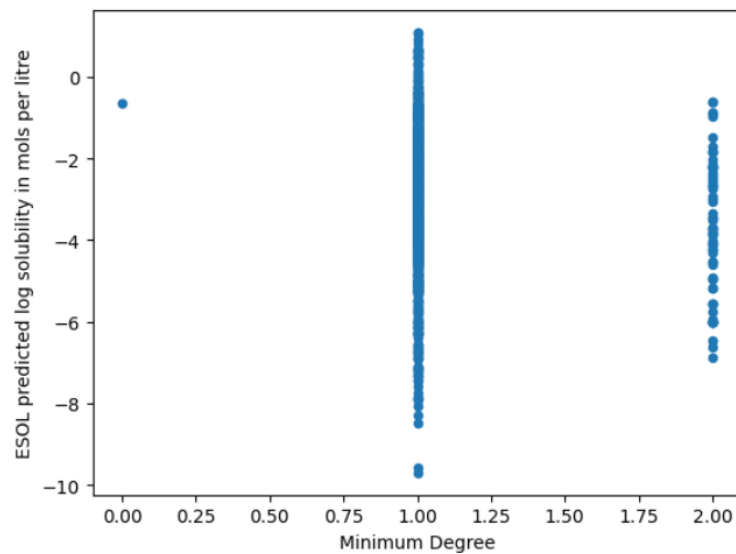


Figure 1: Min. degree vs solubility

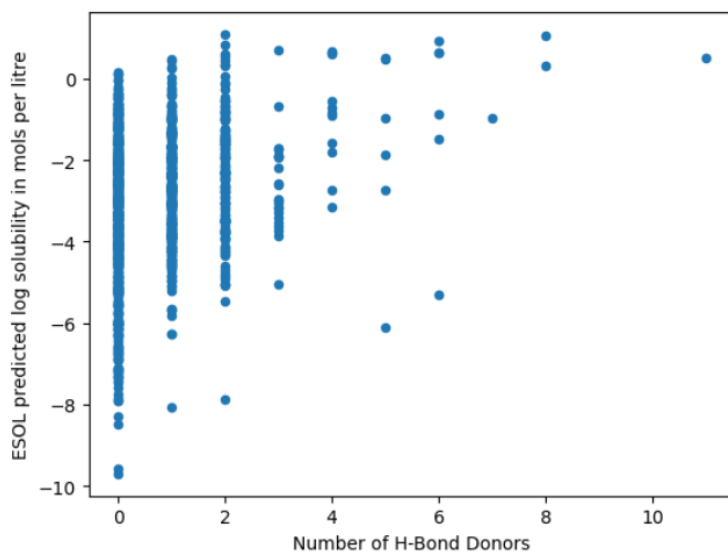


Figure 2: Solubility vs number of H-bond

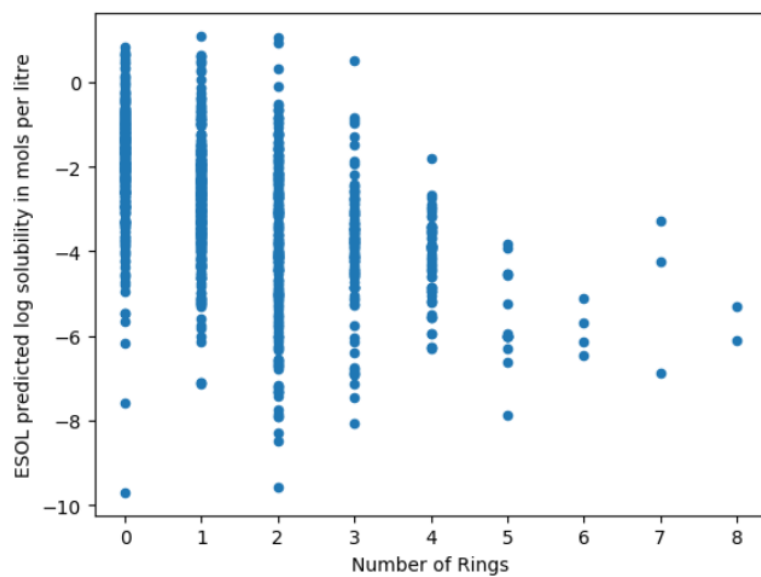


Figure 3: solubility vs number of rings

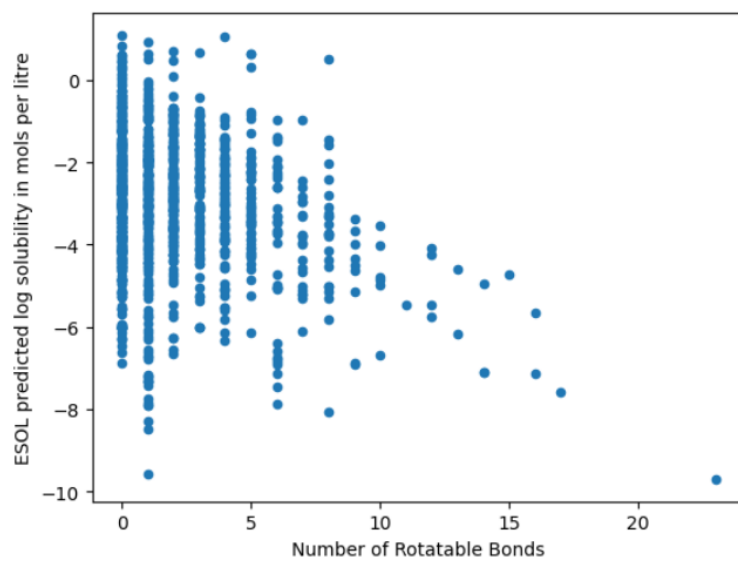


Figure 4: Solubility vs number of rotational bonds

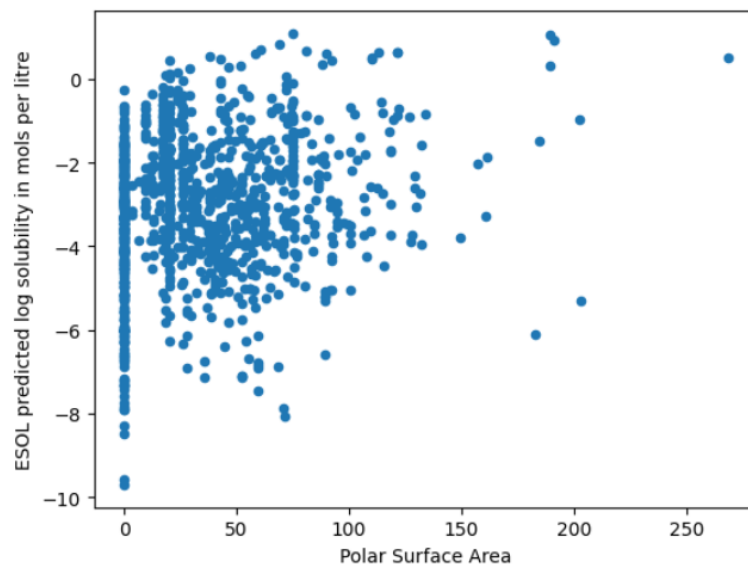


Figure 5: Solubility vs Polar surface area

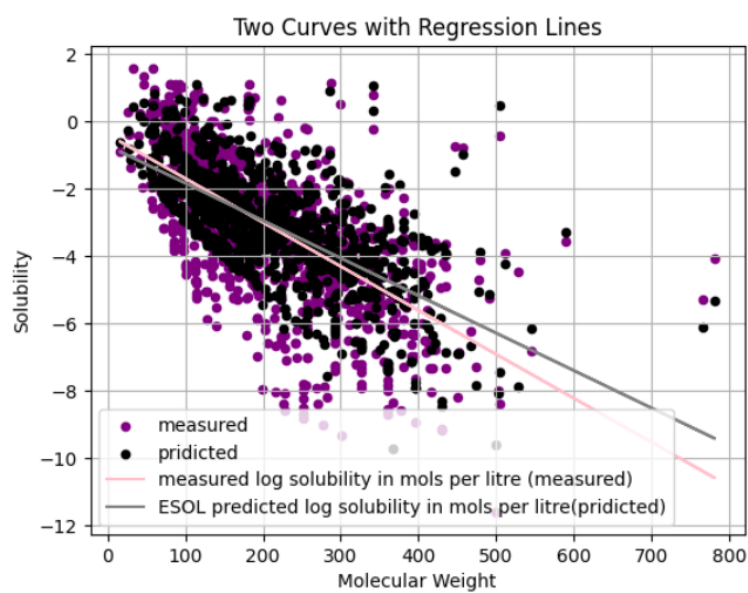


Figure 6: Solubility vs molecular wt

[b]0.4

Histogram with Mean, Median, StdDev, and Mode for measured log solubility in mols per litre

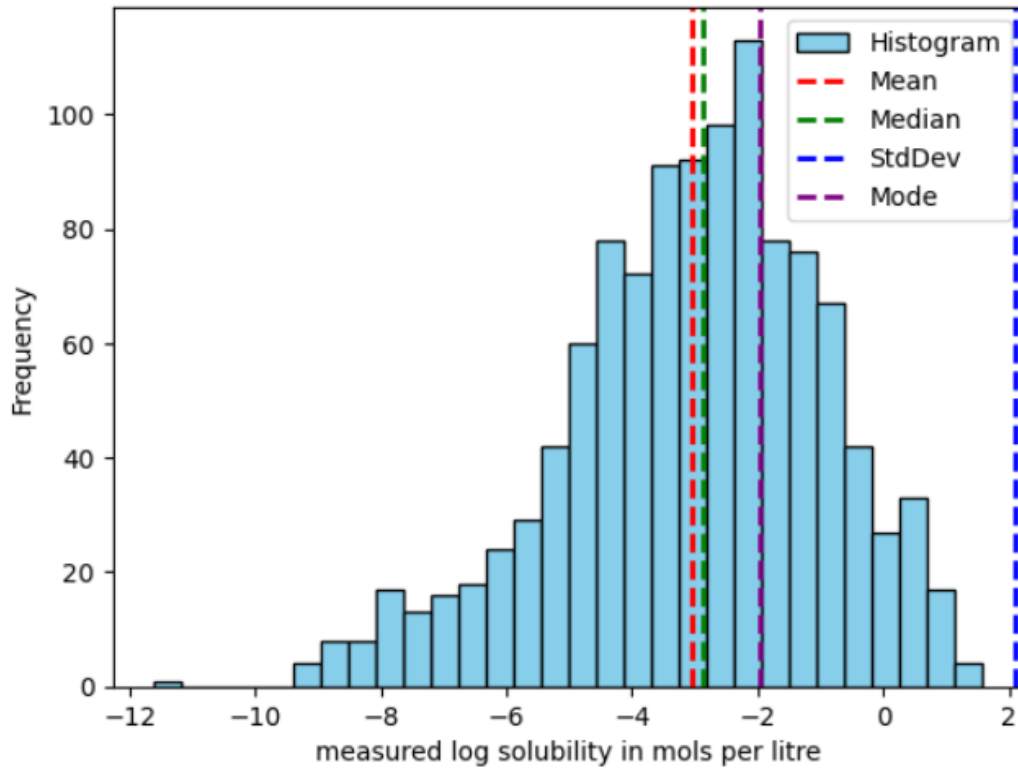


Figure 7: Histogram1

[b]0.4

Histogram with Mean, Median, StdDev, and Mode for measured log solubility in mols per litre

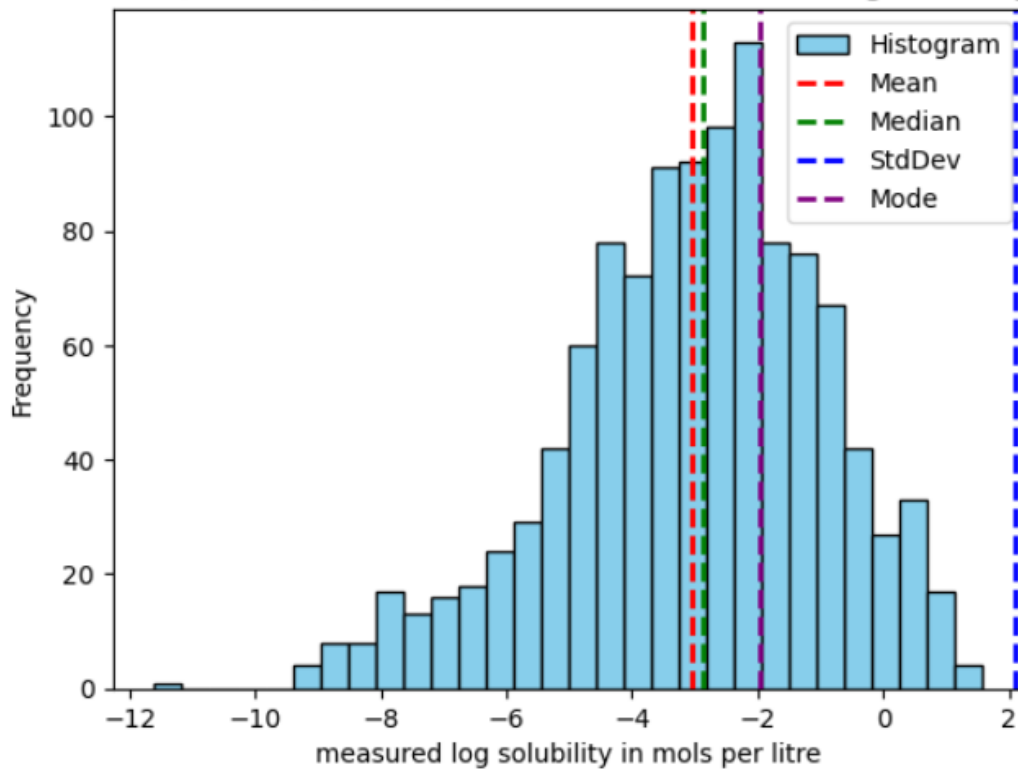


Figure 8: Histogram2

Figure 9: Parameters