CALCULATING CONCENTRATIONS USING BEER-LAMBERT'S LAW AND LINEAR REGRESSION

Nguyen Thi Thu Hoa
Trường Đại học Thủ đô Hà Nội

Abstract: *The article demonstrates an indirect method for determining the concentration of a substance in a solution by applying Beer-Lambert's law and linear regression.*
In chemistry and biochemistry, quantitative analysis of chemical concentrations is a fundamental task that often cannot be implemented through direct weighing. Therefore, we used Beer-Lambert's law and regression techniques to construct a calibration curve that models the linear relationship between absorbance and concentration based on spectrophotometric data. Afterthat, the resulting linear regression model is then used to predict the unknown concentration given its measured absorbance.

Keywords: *Beer-Lambert's law, linear regression, absorbance, spectrophotometry, spectroscopy, concentration, calibration curve*

Nhận bài ngày 25.03.2025; gửi phản biện, chỉnh sửa, duyệt đăng ngày 30.05.2025
Liên hệ tác giả: Nguyen Thi Thu Hoa; email: hoantt@daihocthudo.edu.vn

1. INTRODUCTION

Concentration determination plays a crucial role in chemistry, biochemistry, environmental science, and medicine. In many cases, direct methods are impractical or even impossible, especially when dealing with highly dilute solutions, light-sensitive substances, or samples in limited quantitites (Akash, n.d.; StudyCorgi, 2023). For these situations, one of the most widely used indirect methods is spectrophotometry - using color intensity to compute the concentration of a solution. In particular, Beer-Lambert's law states a linear relationship between the solution's absorbance and its concentration, assuming constant path length and molar absorptivity (O'Haver, 2015). Previous research has applied this relationship in combination with statistical methods - particularly linear regression – to generate calibration curves from standard solutions and subsequently determine unknown concentrations (Dekking, 2005; Bluman, 2012).

While the theoretical basis and experimental procedure are well established, fewer studies have focused on integrating multiple computational tools to illustrate and enhance the teaching of Beer–Lambert's law. In particular, the combined use of GeoGebra, Python, and R for visualizing, analyzing, and validating spectrophotometric data has not been extensively discussed in the literature. This article explains how Beer-Lambert's law and linear regression work together in practice and demonstrates a practical workflow to implement this method using spectrophotometric data with Geogebra, Python and R.

1.1 Beer-Lambert's law

In figure 1, the spectrophometer is used to determine how much light is absorbed by a colored chemical dissolved in the solution. The monochromator takes the light from the light source and splits it into different colors or wavelengths. Then, the aperture acts as a color filter or wavelength selector. The absorbance $A = -\log_{10}\left(\frac{I}{I_0}\right) = -\log_{10} T$ where I_0 is the intensity of the original incident light, I is the intensity of the light that passes through the

sample cuvette and $T = \frac{I}{I_0}$ is the transmittance. The spectrometer measures T, then calculates A.

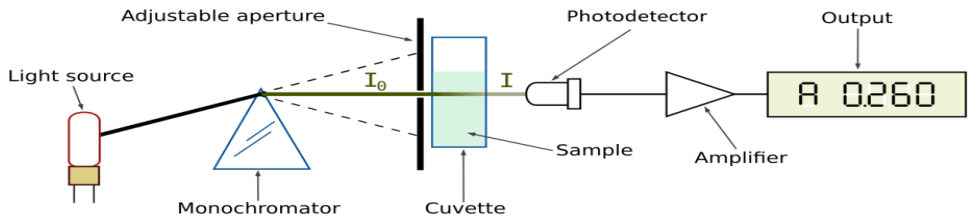


Figure 1: Spectrophotometer components

According to the Beer-Lambert's law, $A = \epsilon \cdot c \cdot l$ where $\epsilon\ [M^{-1}cm^{-1}]$ is the molar absorptivity or molar extinction coefficient, $c\ [M]$ is the concentration of the substance, and $l\ [cm]$ is the path length of the cuvette. When ϵ and l are constant, this equation indicates a linear relationship between absorbance and concentration.

The phET interactive simulation in figure 2 can help us visualize and interact with the components of Beer's law.

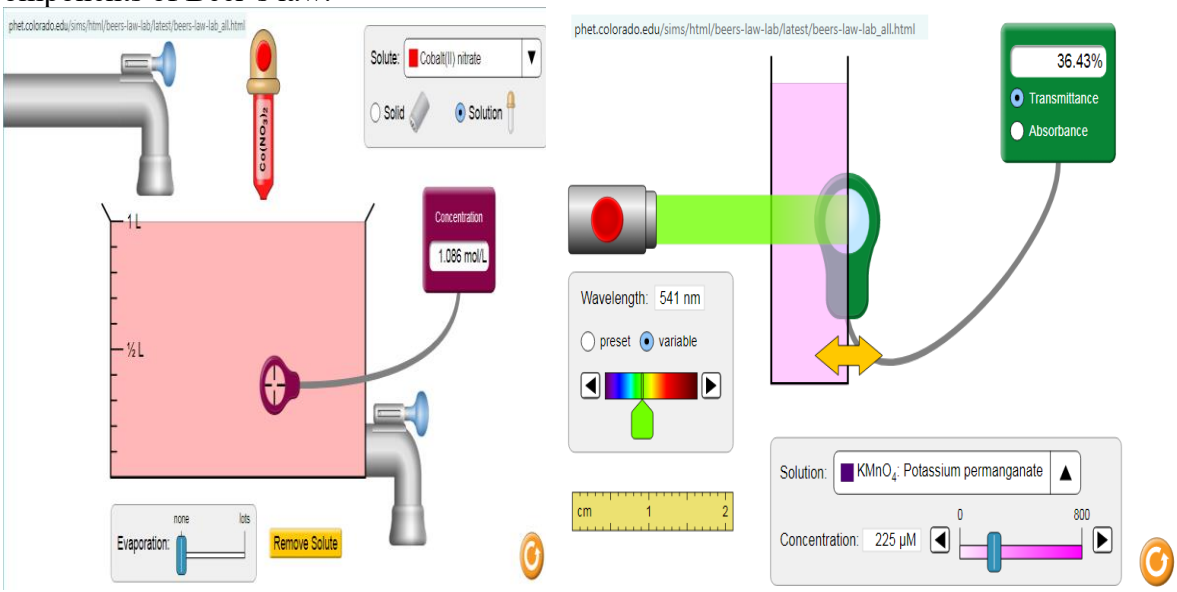


Figure 2. The PhET Beer's law simulation

1.2 Linear Regression

Suppose two random variables under our study are X and Y. For a sample of size n, we obtain data points $\{(x_i, y_i)\}_{i=1}^n$. We set $\Sigma x = \sum_{i=1}^n x_i, \Sigma y = \sum_{i=1}^n y_i, \Sigma xy = \sum_{i=1}^n x_i y_i$,

$$\Sigma x^2 = \sum_{i=1}^{n} x_i^2, \Sigma y^2 = \sum_{i=1}^{n} y_i^2 \quad \bar{x} = \frac{\Sigma x}{n}, \bar{y} = \frac{\Sigma y}{n}, s_{xy} = \Sigma xy - \frac{\Sigma x \cdot \Sigma y}{n} = \Sigma xy - n \cdot \bar{x} \cdot \bar{y},$$

$$s_{xx} = \Sigma x^2 - n \cdot \bar{x}^2, \qquad s_{yy} = \Sigma y^2 - n \cdot \bar{y}^2$$

Spearman's rank correlation coefficient

$$r = \frac{n \cdot (\sum_{i=1}^n x_i \cdot y_i) - (\sum_{i=1}^n x_i) \cdot (\sum_{i=1}^n y_i)}{\sqrt{\left[n(\sum_{i=1}^n x_i^2) - (\sum_{i=1}^n x_i)^2\right] \cdot \left[n(\sum_{i=1}^n y_i^2) - (\sum_{i=1}^n y_i)^2\right]}} = \frac{s_{xy}}{\sqrt{s_{xx} \cdot s_{yy}}} \in [-1, 1]$$

Based on the Beer-Lambert's law, for spectrophotometric data of absorbance and concentrations, r would be close to 1 since there is a strong positive linear relationship between the variables.

Assume that the desired linear regression line is $\hat{y} = \beta_1 \cdot x + \beta_0$ then the sum of squared residuals is $SSE = \sum_{i=1}^{n}(y_i - \hat{y_i})^2 = \sum_{i=1}^{n}(y_i - \beta_1 \cdot x_i - \beta_0)^2$.

Using the least squares method, the minimum of SSE is found when

$$\frac{\partial\, SSE}{\partial \beta_1} = \frac{\partial\, SSE}{\partial \beta_0} = 0$$

$$\Rightarrow \sum_{i=1}^{n}(y_i - \beta_1 \cdot x_i - \beta_0) \cdot x_i = 0, \sum_{i=1}^{n}(y_i - \beta_1 \cdot x_i - \beta_0) = 0$$

$$\Rightarrow \beta_0 = \bar{y} - \beta_1 \cdot \bar{x}\,,\ \beta_1 = \frac{s_{xy}}{s_{xx}}$$ (see [2], [5])

The coefficient of determination $R^2 = 1 - \frac{SSE}{s_{yy}}$ measures the goodness of fit of the linear model.

For spectrophotometric data, the slope β_1 of the calibration curve corresponds to the product of the molar absorptivity and the path length: $\beta_1 = \epsilon \cdot l$. Since in most experiments, the path length l is fixed (typically $l = 1\ cm$), the slope β_1 gives an empirical estimate of how strongly the solute absorbs light at the chosen wavelength. The steeper slope means higher sensitivity, that is, small changes in concentration result in large changes in absorbance. Ideally, the y-intercept β_0 should be 0 (i.e., when the cuvette has only distilled water, the absorbance should be 0). In reality, β_0 reflects the instrumental noise or residual absorbance from solvents, cuvettes, or impurities.

2. COMPUTING CONCENTRATIONS

2.1. Experimental setup and method

2.1.1. Wavelength selection and justification

The molar absorptivity ϵ is wavelength - dependent. For the same substance, different wavelengths of light are absorbed differently (see Figure 3). To maximize sensitivity and linearity in our calibration curve, we would carry out the following steps:

- Step 1: Plot absorbance vs wavelength
- Step 2: Identify the peak absorbance wavelength λmax
- Step 3: Avoid nearby wavelengths where the compound does not absorb strongly.

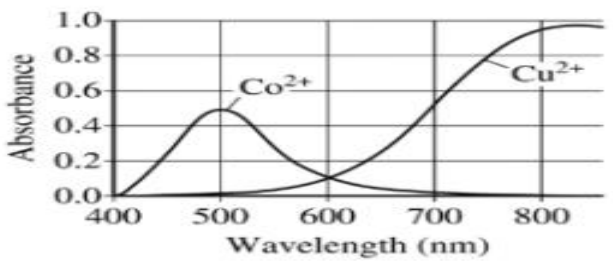


Figure 3. Set frequency in the spectrophotometer

In figure 3, in a solution that contains both Co^{2+} and Cu^{2+}, we only want to determine the concentration of Co^{2+}. Based on the graph, Co^{2+} ions absorb over a wide range of different wavelengths and its relatively high absorbance value is at $500\ nm$ with very little interference from Cu^{2+}. Therefore, we set frequency $\lambda \max = 500\ nm$ in the spectrophotometer.