

O USO DA REGRESSÃO LINEAR E ANÁLISE DE REGRESSÃO COM O EXCEL NA QUÍMICA.

Priscila Mitsue Kussaba Kayano, Marisa Veiga Capela, Jorge Manuel Vieira Capela. – Probabilidade e Estatística – Bacharelado em Química - Departamento de Físico-Química – Instituto de Química – Campus de Araraquara.

A utilização de recursos computacionais em estatística tem hoje o seu lugar de destaque, graças à popularização da informática como meio de comunicação, pesquisa e trabalho. Estudar uma relação entre uma variável dependente e uma ou mais variáveis independentes é uma tarefa comum em análise estatística. Pode-se tentar atingir este objetivo através de modelos de regressão. Com base em um conjunto de dados experimentais, os parâmetros da função de um modelo de regressão podem ser estimados.

O modelo de regressão inclui um componente determinístico e um aleatório, e pode ser representado da seguinte forma:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

onde $\beta_0 + \beta_1 X_i$ descreve a relação entre X e Y, ε_i são os erros aleatórios, Y é a variável resposta ou dependente, X é a variável explicativa ou independente, β_1 é a inclinação da reta e representa a variação esperada em Y por cada variação unitária em X e β_0 é a interseção de Y, valor médio para X igual a zero.

A equação que melhor ajusta os valores reais aos previstos é dado pela seguinte equação: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$, onde \hat{Y}_i é o valor previsto de Y para a observação i e X_i é o valor de X para a observação i.

O método dos mínimos quadrados é a técnica de estimação mais popular e considera os desvios de Y_i em relação a seu valor esperado (\hat{Y}_i):

$$d_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i).$$

Quando os desvios ou resíduos forem os menores possíveis, significa que encontramos a melhor linha reta que se ajusta aos dados, ou seja:

$$\hat{\beta}_0 = \frac{\sum X_i \sum Y_i - \frac{\sum X_i \sum Y_i}{N}}{\sum X_i^2 - \frac{(\sum X_i)^2}{N}} \quad e \quad \hat{\beta}_1 = \frac{\sum Y_i}{N} - \hat{\beta}_0 \frac{\sum X_i}{N}.$$

O método de regressão linear utilizado no tratamento de dados experimentais em Química pode ser facilmente utilizado com o auxílio do computador.

Utilizamos o Excel por ser uma planilha eletrônica de cálculo extremamente popular, acessível e que pode ser utilizada com facilidade. A utilização do Excel permite tornar mais concreta a elaboração dos algoritmos, seus usos e visualizações, possibilitando ainda modificar alguns aspectos de um determinado raciocínio estatístico com a visualização instantânea dos efeitos destas mudanças.

O objetivo deste trabalho é apresentar e interpretar o método dos mínimos quadrados e os conceitos da análise de regressão com a utilização do Excel para a realização dos cálculos.

Uma análise de dados (tabela 1) da decomposição do acetaldeído (etanal, CH_3CHO), foi realizada utilizando-se a planilha eletrônica Excel com a finalidade de interpretar os cálculos e encontrar a energia de ativação e o fator de frequência utilizando-se a equação:

$$k = A e^{\frac{-E_a}{RT}},$$

onde a variável dependente k é a constante de velocidade da reação, A é o fator de frequência, R é a constante universal dos gases ($R = 8,31451 \text{ J K mol}^{-1}$), E_a é a energia de ativação e T é a temperatura na escala Kelvin. Esta equação é conhecida como Equação de Arrhenius e pode ser linearizada da seguinte forma:

$$\ln k = \ln A - \frac{E_a}{RT}.$$

Na figura 1 observa-se os resultados fornecidos pelo Excel, onde $1/T$ é a variável independente e $\ln k$ a variável dependente, então estima-se $E_a = 188 \text{ KJ mol}^{-1}$ e $A = 1,1 \times 10^{12} \text{ L mol}^{-1} \text{ s}^{-1}$.

Os dados que relacionam duas variáveis, são apresentados em um gráfico bidimensional denominado diagrama de dispersão, tornando assim, mais fácil a interpretação visual dos dados, fornecendo uma idéia do tipo de relação existente entre as variáveis.

Com base nos resultados observamos que a análise de regressão linear realizada pelo Excel oferece vários resultados: calcula o intervalo de 95% de confiança (ou outro) para os parâmetros da regressão, realiza o teste F, teste t, calcula o p valor, o coeficiente de correlação (r múltiplo = 0,999) que mede o grau de associação linear de duas variáveis, resultados de resíduos (resíduos, resíduo padrão), análise de variância, gráfico de resíduos, etc, com isso temos uma análise estatística da regressão linear bastante completa.

Tabela 1 - Dados da variação da constante de velocidade com a temperatura do acetaldeído

T(K)	k(L/mols)
700	0,011
730	0,035
760	0,105
790	0,343
810	0,789
840	2,17
910	20
1000	145

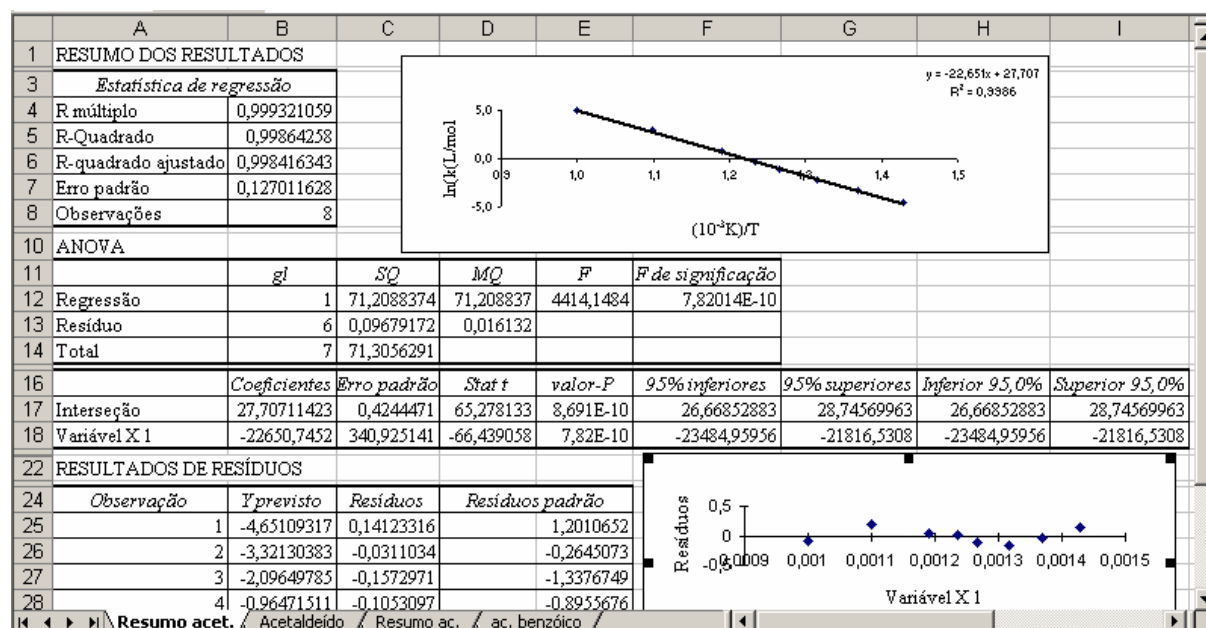


Figura 1 – Resultados fornecidos pelo Excel da Regressão Linear e Análise de regressão da decomposição do acetaldeído.

Concluimos que o emprego de planilhas eletrônicas de cálculo para a resolução de problemas envolvendo regressão linear, em conjunto com um bom conhecimento dos conceitos da análise de regressão, é uma alternativa interessante na compreensão da variabilidade de dados em Química.

Referências Bibliográficas

- BUSSAB, W.O. E MORETTIN, P.A. Estatística Básica. 5^a. ed. São Paulo: Saraiva, 2004.
- LAPPONI, J.C. Estatística usando o Excel. São Paulo: Editora Lapponi, 2000.
- LEVINE, m.1., BERENSON, M.L., STEPHAN, D. Estatística: teoria e aplicações. Usando Microsoft Excel. Rio de Janeiro: Livros Técnicos e Científicos Editora S.^a, 2000.
- MEYER, P.L. Probabilidade – Aplicações à estatística. 2^a ed. Rio de Janeiro: Livros Técnicos e Científicos Editora S.^a, 2000.
- TRIOLA, M.F. Introdução à Estatística, 7ed. Livros Técnicos e Científicos Editora, 1999.