EXEMPLO DE APLICAÇÃO DE REGRESSÃO

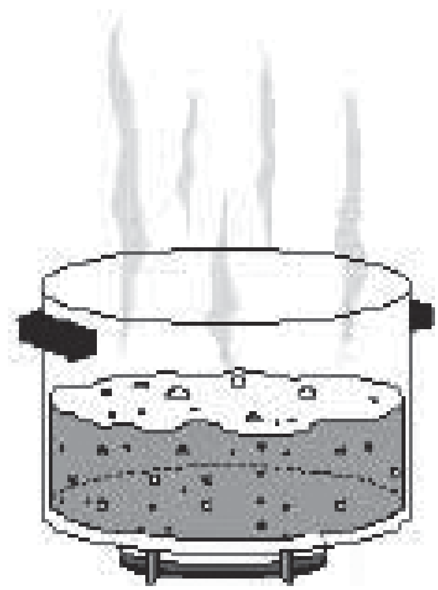
19
aula

META

Apresentar ao aluno um exemplo de aplicação de regressão, que necessite de um termo quadrático

OBJETIVOS

Ao final desta aula o aluno deverá:
analisar o problema prático da determinação do calor de vaporização de um líquido;
testar o ajuste do modelo linear, e fazer a análise da variância e avaliar o gráfico dos resíduos;
modificar o modelo linear, incluindo um termo quadrático;
ajustar os dados ao modelo quadrático, fazer a análise da variância e avaliar o gráfico dos resíduos; e
interpretar quimicamente os resultados.



Vaporização, desenho, autor desconhecido (Fonte: www.ufmg.br)

PRÉ-REQUISITOS

Ajuste por mínimos quadrados e análise do desempenho de ajustes.

INTRODUÇÃO

Os métodos de ajuste que vimos até aqui envolvem pressupostos que vão sendo incorporados e que às vezes acabam sendo esquecidos nos tratamentos de dados. Em atividades de pesquisa científica e também na indústria, é comum as pessoas coletarem dados e aplicarem certos modelos que parecem ser universalmente aceitos, sem fazer a devida análise de desempenho do modelo. O que procuramos fazer nesta aula foi mostrar um exemplo em que vários testes iniciais de desempenho de um modelo (que até visualmente parece perfeitamente satisfatório) apontam para um bom desempenho, mas que, no final da avaliação, se mostra inválido. Com isso, buscamos ilustrar também a aplicação do método de ajuste por mínimos quadrados a um modelo não-linear, que pode ser considerado como uma expansão do modelo linear.

Pesquisa científica, fotografia, autor desconhecido (Fonte: www.sesuweb.mec.gov.br)

VAPORIZAÇÃO

Podemos mostrar que a dependência da pressão de vapor de um líquido com a temperatura obedece à equação de Clausius-Clapeyron (ver nota explicativa):

$$\ln p_{vap} = b_0 - \frac{\Delta_{vap} H}{R}\left(\frac{1}{T}\right)$$

$$y = b_o + b_1 x$$

Observe que a equação lembra a equação de uma reta, com coeficientes b_o e, como b_1, $-(\Delta_{vap}H)/R$, onde R é a constante dos gases. Como y temos $\ln P_{vap}$ e como x temos 1/T.

Com base nisto, um experimento comum em disciplinas de físico-química experimental é a determinação da entalpia de vaporização do líquido ($\Delta_{vap}H$), através de medidas da variação da pressão de vapor com a temperatura. O valor de $\Delta_{vap}H$ será obtido, caso os dados se ajustem satisfatoriamente ao modelo linear, a partir da inclinação da reta.

Suponha que você fez esta prática e obteve, para o tetracloreto de carbono (CCl_4), os dados de pressão de vapor em função da temperatura apresentados na tabela a seguir:

Variação de pressão do vapor de CCl_4 com a temperatura

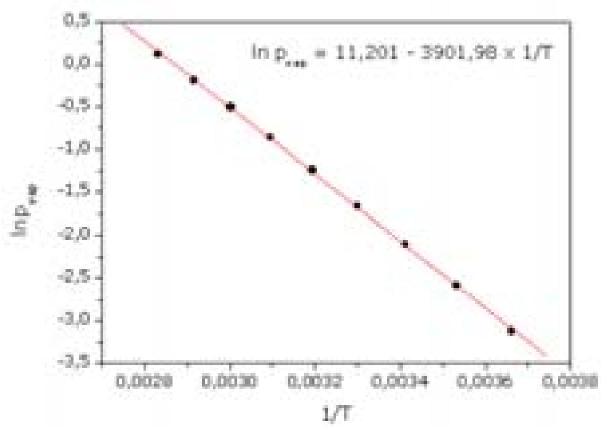
Ensaio	T/K	Pvap/torr
1	273	0,044
2	283	0,075
3	293	0,122
4	303	0,190
5	313	0,288
6	323	0,422
7	333	0,601
8	343	0,829
9	353	1,124

Observando a equação, você percebe que nela aparece o inverso da temperatura (1/T) e o logaritmo neperiano da pressão ln(P_{vap}). Assim, antes de continuar, você precisa fazer estes cálculos com os dados experimentais. Os resultados estão apresentados na tabela abaixo.

Variação de ln da pressão do vapor de CCl_4 com o inverso da temperatura

Ensaio	1/T	Ln(Pvap)
1	0,003663	-3,124
2	0,003534	-2,59
3	0,003413	-2,104
4	0,0033	-1,661
5	0,003195	-1,245
6	0,003096	-0,8627
7	0,003003	-0,5092
8	0,002915	-0,1875
9	0,002833	0,1169

Em seguida, você submeteu os dados a um ajuste linear, pelo método dos mínimos quadrados e também os representou em um gráfico. O gráfico obtido, mostrado a seguir, inclui os pontos observados experimentalmente e a reta proveniente do ajuste.

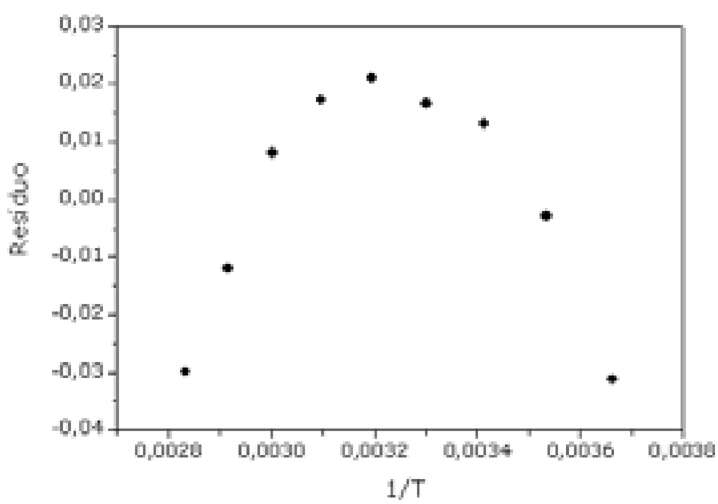


Aparentemente, todos os dados estão praticamente sobre a reta proposta e nossos problemas acabaram... Será? Não é assim tão simples e o olho humano pode ser enganado facilmente nessas análises de desempenho.

Vamos inicialmente realizar a análise da variância, calculando as somas quadráticas e, a partir do número de graus de liberdade, calculamos também as médias quadráticas.

Fonte de Variação	Soma Quadrática	Nº de g.l.	Média Quadrática
Regressão	9,804	1	9,804
Resíduo	0,003	7	$4,658 \times 10^{-4}$
Total	9,807	8	

O coeficiente de determinação R^2 foi de 0,9997, indicativo de um ajuste muito satisfatório. A razão MQ_R/MQ_r obtida foi de $2{,}1 \times 10^4$ que, comparada ao valor do teste $F_{1,7} = 5{,}59$, é um valor extremamente alto. Este valor indicaria que temos uma regressão significativa estatisticamente, não fosse por um detalhe: o teste F pressupõe uma distribuição normal dos resíduos. Contudo, observe o gráfico dos resíduos a seguir.



Observando o gráfico dos resíduos, percebemos que, ao invés de se distribuírem de forma aleatória, estes sugerem nitidamente um padrão geométrico, no qual a região central do gráfico concentra os resíduos positivos, ao passo que os resíduos negativos localizam-se nas extremidades.

Qual conclusão pode ser tirada após esta tentativa quase bem-sucedida de aplicar o modelo linear? O modelo linear precisa ser ampliado, ou seja, precisamos incluir além dos coeficientes b_o e b_1, que é multiplicado por (1/T), um novo coeficiente b_2 que seja multiplicado por $(1/T^2)$. Assim, nosso modelo, que era linear, passa a ser um modelo chamado quadrático. Vamos admitir que a influência da temperatura sobre a pressão de vapor seja adequadamente descrita pela equação:

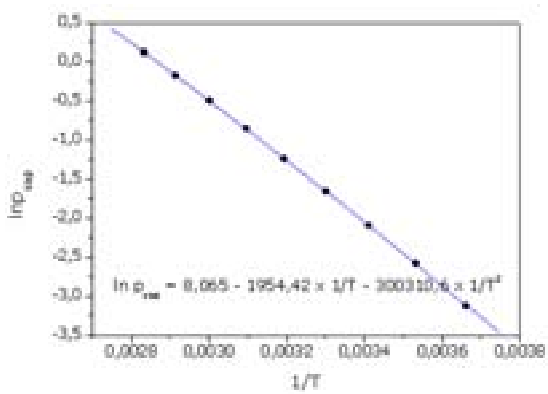
$$\ln P_{vap} = b_o - (\Delta_{vap}H)/R(1/T) + b_2(1/T^2)$$

O ajuste por mínimos quadrados é feito utilizando o mesmo método mostrado para o ajuste linear, mas a ordem das matrizes aumenta:

$$X = \begin{bmatrix} 1 & X1 & X1^2 \\ 1 & X2 & X2^2 \\ \dots & \dots & \dots \\ 1 & Xn & Xn^2 \end{bmatrix} \qquad y = \begin{bmatrix} y1 \\ y2 \\ y3 \\ \dots \\ yn \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta o \\ \beta 1 \\ \beta 2 \end{bmatrix}$$

ATIVIDADES

Faça as operações com as matrizes referentes ao problema, para ajustar por mínimos quadrados o modelo quadrático.
O resultado gráfico é mostrado na figura a seguir, bem como os parâmetros b_o = 8,065, b_1 = -1954,42 e b_2 = 300310,6.

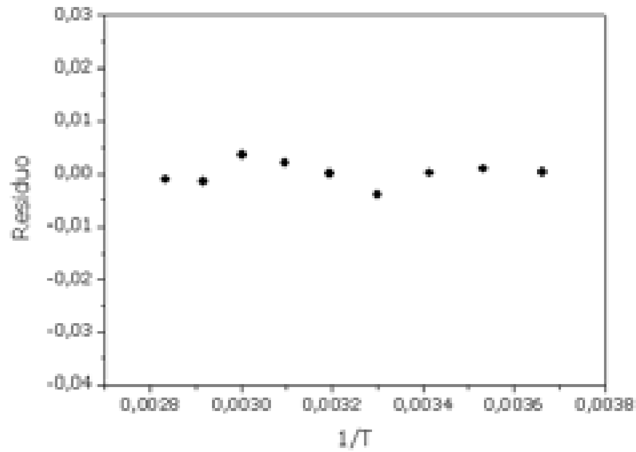


A avaliação do desempenho deste novo modelo deve ser feita pela análise da variância, pelo coeficiente de determinação e pelo gráfico dos resíduos. A novidade aqui é que na análise da variância, a razão MQ_R/MQ_r deve ser comparada ao valor de $F_{2,n-1}$ e não $F_{1,n-2}$ pois agora a ordem do ajuste é 2. Vemos que o valor de R^2 foi igual a 1 e o valor da razão MQ_R/MQ_r foi $2{,}2x10^6$, muito maior do que $F_{2,7}$ =4,74.

ANÁLISE DA VARIÂNCIA - MODELO QUADRÁTICO

Fonte de Variação	Soma Quadrática	Nº de g.l.	Média Quadrática
Regressão	9,807	2	9,807
Resíduo	0,000	6	$4{,}386x10^{-6}$
Total	9,807	8	
			R^2 = 100%; $F_{2,7}$ =5,15 <<< $2{,}2x10^6$

Finalmente, o gráfico de resíduos ficou satisfatório, com uma distribuição aleatória e razoavelmente homogênea em termos de sinais.



A interpretação química que podemos dar a este exercício é que a superioridade do modelo quadrático significa que o calor de vaporização não pode ser considerado constante no intervalo de temperatura estudado.

$$\frac{\Delta_{vap}H}{R} = \frac{d(\ln p_{vap})}{d(1/T)} = -1{,}954 - 600{,}62\left(\frac{1}{T}\right)$$

CONCLUSÃO

Como conclusões desta aula, podemos destacar a importância da análise completa do desempenho de um ajuste. É absolutamente temerário limitar isso a um simples exame visual e realmente muito arriscado somente examinar o coeficiente de determinação, como muita gente faz. Na verdade, mesmo a análise da variância pode guardar armadilhas, ficando de fato a cargo do gráfico de resíduos a demonstração da eficiência do processo.

Análise de dados, fotomontagem, autor desconhecido (Fonte: www.maxsecure.com.br).

RESUMO

Em atividades de pesquisa científica e também na indústria é comum as pessoas coletarem dados e aplicarem certos modelos que parecem ser universalmente aceitos, sem fazer a devida análise de desempenho do modelo. O que procuramos fazer nesta aula foi mostrar um exemplo em que vários testes iniciais de desempenho de um modelo (que até visualmente parece perfeitamente satisfatório) apontam para um bom desempenho, mas que, no final da avaliação, se mostra inválido.

Apresentam-se dados do logaritmo da pressão de vapor de um líquido e do inverso da temperatura. Há uma equação, na teoria de físico-química, que sugere que a relação entre estas variáveis é linear. Com isso em mente, um aluno aplicou o método de ajuste por mínimos quadrados a estes dados, fazendo um ajuste linear. Construiu também o gráfico mostrando o posicionamento dos dados observados frente a reta obtida pelo ajuste. Aparentemente, através de um exame visual, o ajuste foi perfeito.

O coeficiente de determinação R^2 foi de 0,9997, indicativo de um ajuste muito satisfatório. A razão MQ_R/MQ_r obtida foi de $2{,}1 \times 10^4$ que, comparada ao valor do teste $F_{1,7} = 5{,}59$, é um valor extremamente alto. Este valor indicaria que temos uma regressão significativa estatisticamente, não fosse por um detalhe: o teste F pressupõe uma distribuição normal dos resíduos. Ao construir o gráfico dos resíduos, qual não foi sua surpresa ao observar que estes definiram uma parábola, sem qualquer possibilidade de se constatar uma distribuição aleatória.

Assim, concluiu que o modelo linear precisava ser ampliado, com a inclusão de um novo coeficiente b_2 que seja multiplicado por $(1/T^2)$. Assim, nosso modelo, que era linear, passa a ser um modelo chamado quadrático. A interpretação Química que podemos dar a este exercício é que a superioridade do modelo quadrático significa que o calor de vaporização não pode ser considerado constante no intervalo de temperatura estudado.

A EQUAÇÃO DE CLAUSIUS-CLAPEYRON

Se perguntarmos aos alunos qual é a temperatura de ebulição da água na pressão correspondente ao nível do mar, todos (ou a grande maioria) saberão responder. Porém, será que temos como saber qual seria a temperatura de ebulição para valores diferentes de pressão? E se estivermos preocupados com um líquido que não seja a água? Será que é possível calcular estes "novos" valores de temperatura? Há, para isto, equações que nos permitem calcular as condições de mudança de fases (lembre-se das mudanças de estado físico: fusão, vaporização e sublimação, pois é disto que estamos falando), em situações diferentes das condições padrão que encontramos nas tabelas. A equação de Clausius-Clapeyron é a equação que permite calcular, para a transição entre líquido e gás, as pressões de equilíbro a diferentes temperaturas ou vice-versa, ou seja, as temperaturas de equilíbrio a diferentes pressões, para qualquer substância, desde que saibamos a entalpia padrão da transição. Para a fusão e a sublimação, há equações ligeiramente diferentes, mas que levam ao mesmo tipo de informação.

REFERÊNCIAS

Experiments in physical chemistry / Carl W. Garland; Joseph W. Nibler ; **Davia P. Shoemaker.** Boston: McGraw Hill, 2003.

Renato N. Rangel, Práticas de Físico-química, 3. edição, Edgard Blucher, 2006.

BUSSAB, W. O.; MORETIN, P.A. **Estatística básica**. São Paulo, Ed. Atual, 1985.

BARROS NETO, B.; SCARMINIO, I. E. BRUNS. R. E. **Planejamento e otimização de experimentos.** Campinas Editora da Unicamp, 1995.