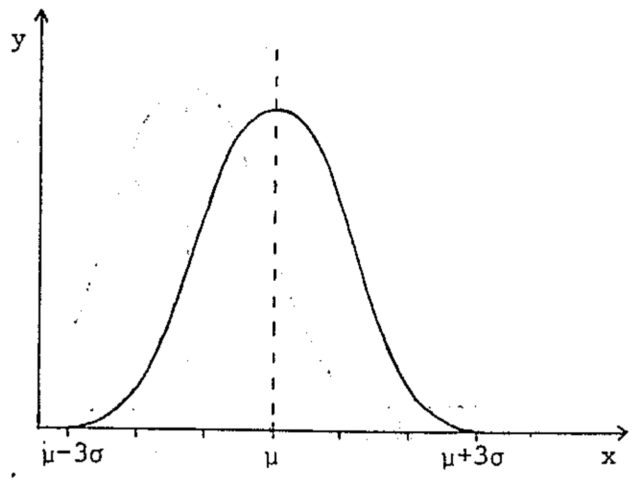
QUÍMICA DAS SOLUÇÕES 1988/89

TRATAMENTO DE DADOS

Distribuição de Gauss

Para um número infinito de medidas



Equação da curva :

$$y = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-(x-\mu)^2/2\sigma^2\}$$

onde μ é a média

σ é o desvio padrão

$$\int_{-\infty}^{+\infty} y\,dx = 1$$

Tabela I. Área dentro da curva da dist. de Gauss entre $\mu-z\sigma$ e $\mu+z\sigma$

z	0,67	1,00	1,29	1,64	1,96	2,00	2,58	3,00	3,30
Área	0,500	0,683	0,800	0,900	0,950	0,955	0,990	0,997	0,999

Definição:

$$\sigma = \left[\frac{\Sigma\,(x_i-\mu)^2}{N}\right]^{1/2}$$

A variância σ^2 é uma propriedade aditiva

i.e. $\sigma^2_{tot} = \sigma_1^2 + \sigma_2^2 + \; .... + \sigma_n^2$

Número finito de medidas

Média $\bar{x} = \dfrac{\Sigma x_i}{N}$ $\qquad \Sigma n_i \to \infty \qquad \bar{x} \to \mu$

Desvio padrão $s = \left[\dfrac{\Sigma(x_i - \bar{x})^2}{N - 1}\right]^{1/2}$ $\qquad \Sigma n_i \to \infty \qquad s \to \sigma$

onde N-1 é o número de graus de liberdade. (Quando $N>20$, $s \approx \sigma$).

Erro padrão da média $s_m = sN^{-1/2}$

Nota: quando se usa uma máquina de calcular, em geral o "SD" é s e a "VAR" é σ^2. Se houver qualquer dúvida, consultar o manual de instruções.

Erros

Para simplicidade, dividimos os erros em dois tipos - os determinados ou sistemáticos que têm sinal e os indeterminados que não têm sinal. Podem-se escrever as regras seguintes :

(i) Erros determinados (não muito usual)

$$y = a + b - c \qquad \Delta y = \Delta a + \Delta b - \Delta c$$

$$y = \frac{a.b}{c} \qquad \frac{\Delta y}{y} = \frac{\Delta a}{a} + \frac{\Delta b}{b} - \frac{\Delta c}{c}$$

$$y = a^x \qquad \frac{\Delta y}{y} = x.\frac{\Delta a}{a}$$

$$y = \ln a \qquad \Delta y = \frac{\Delta a}{a}$$

(ii) Erros indeterminados (usual)

Não se sabe o sinal do erro, portanto soma-se os valores absolutos referentes aos erros individuais.

p.e. $y = a + b - c \qquad |\Delta y| = |\Delta a| + |\Delta b| + |\Delta c|$

Se se sabe o desvio padrão de cada parcela então

$$y = a + b - c \qquad s_y = \left[s_a^2 + s_b^2 + s_c^2\right]^{1/2}$$

$$y = \frac{a.b}{c} \qquad \frac{s_y}{y} = \left[\left(\frac{s_a}{a}\right)^2 + \left(\frac{s_b}{b}\right)^2 + \left(\frac{s_c}{c}\right)^2\right]^{1/2}$$

$$y = a^x \qquad \frac{s_y}{y} = x.\frac{s_a}{a}$$

$$y = \ln a \qquad s_y = \frac{s_a}{a}$$

Limites de confiança e intervalos de confiança

É útil saber se $\bar{x}$ é próximo de μ e ver se μ está entre certos limites especificados com uma certa confiança (ou probabilidade).

(i) Quando $s \approx \sigma$, da distribuição de Gauss este critério segue facilmente.

$$\mu = \bar{x} \pm \frac{z\sigma}{\sqrt{N}}$$

onde z é dado pelos valores na Tabela I.

Exemplo: Calcular os limites a 90% de confiança para uma análise de Fe na água que dá para N=25, $\bar{x}$=50,0ppm e $s(\sigma)$ = 1,0ppm

$$\mu = 50 \pm \frac{1,64 \times 1,0}{\sqrt{25}}$$

$$= 50,0 \pm 0,3$$

i.e. com 90% de probabilidade, μ é entre 49,7 e 50,3 ppm.

(ii) Quando $s \neq \sigma$, (N pequeno), usa-se a formula com modificações:

$$\mu = \bar{x} \pm \frac{ts}{\sqrt{N}}$$

onde t é o t de student. Aplica-se do mesmo modo usando valores de t tabelados (veja Tabela II). Nota-se que quando $N \to \infty$, $s \to \sigma$ e $t \to z$.

Rejeição de dados

Dados devem ser rejeitados com cautela. Verificar os cálculos (se houver) e pensar bem nas razões que podem conduzir a este resultado. Um teste existe para analisar estatisticamente o problema de rejeição - o teste de Q

$$Q = \frac{\text{(Dado para rejeitar?)} - \text{(Dado mais perto)}}{\text{(Dado maximo)} - \text{(Dado minimo)}}$$

Este valor de Q é então comparado com valores de Q_{crit} (Tabela III). Se $Q>Q_{crit}$ o dado é rejeitado.

Exemplo: Um minério de Fe_2O_3 contém percentagens de Fe de 30,2; 30,4; 31,2; 30,6; e 30,3. Deve-se rejeitar o valor de 31,2 ao nível de confiança de 96% ?

$$Q = \frac{31,2 - 30,6}{31,2 - 30,2} = 0,6$$

$$Q_{crit} = 0,73$$

Portanto não se deve rejeitar o dado experimental.

Comparação entre dados experimentais e teóricas

Compara-se a média medida e a média verdadeira

$$\bar{x} - \mu = \pm \frac{ts}{\sqrt{N}}$$

Se $|\bar{x} - \mu| > |\frac{ts}{\sqrt{N}}|$ a um dado nível de confiança, uma diferença significativa existe.

Comparação entre duas séries de medidas experimentais

Isto é especialmente útil para a comparação de séries de dados duma análise de rotina. Escreve-se

$$\mu = \bar{x}_1 \pm \frac{ts}{\sqrt{N_1}} = \bar{x}_2 \pm \frac{ts}{\sqrt{N_2}}$$

$$\text{ou} \qquad \bar{x}_1 - \bar{x}_2 = \pm\, ts\sqrt{\frac{N_1 + N_2}{N_1 N_2}}$$

Muitas vezes s já está estabelecido de experiências prévias (N grande e, $s \approx \sigma$) e então podemos substituir s por σ e t por z (Tabela I). Senão calcular s de s_1 e s_2 como está descrito na página 2. (Graus de liberdade $= N_1+N_2-2$)

Exemplo: Análise do conteúdo do efluente duma fábrica de electrólise de NaCl dá níveis de Cl_2 de 30,0ppm (N=5). Com uma nova técnica para a determinação de Cl_2 obtemos uma média de 30,9ppm (N=4). Ao nível de 95% de confiança podemos confiar na nova técnica? Dado: σ=0,20ppm

$$z\sigma\sqrt{\frac{N_1+N_2}{N_1N_2}} = 1,96 \times 0,2\sqrt{\frac{9}{20}}$$

$$= 0,26 \text{ ppm}$$

Portanto não temos confiança na nova técnica.

Outros testes úteis

(i) F - teste

Compara a precisão de medidas, quando as séries de medidas devem ser idênticas (p.e. comparar os resultados de duas pessoas com amostras do mesmo composto).

$$F_{exp} = \left(\frac{s_1}{s_2}\right)^2$$

Se $F_{exp} < F_{crit}$ (Tabela IV) a precisão é boa.

(ii) χ^2 - teste

Mede a diferença entre as medidas experimentais e as esperadas a partir da teoria das probabilidades.

$$\chi^2 = \Sigma \frac{(f_i - F_i)^2}{F_i}$$

onde f_i é a frequência experimental duma medida e F_i é a frequência teórica (Tabela V), onde o somatório é feito sobre todas as séries de observações, N. Procura-se o valor de P da tabela para este valor de N. Quanto menor P menor a diferença.

Limite de detecção, Δx_{min}

$$\Delta x_{min} = \bar{x}_1 - \bar{x}_b = t s_b \sqrt{\frac{N_1+N_b}{N_1 N_b}}$$

Branco: N_b, $\bar{x}_b$, s_b
Amostra: N_1, $\bar{x}_1$

Esta fórmula prevê o limite de detecção a partir de dados obtidos em ensaios brancos. O limite de detecção baixa com

— aumento do número de ensaios brancos ($N_b \uparrow$ e $s_b \downarrow$(?))

— aumento do número de análises do composto ($N_1 \uparrow$)

Tabela II Values of t for Various Levels of Probability

Degrees of Freedom	Factor for Confidence Interval, %				
	80	90	95	99	99.9
1	3.08	6.31	12.7	63.7	637
2	1.89	2.92	4.30	9.92	31.6
3	1.64	2.35	3.18	5.84	12.9
4	1.53	2.13	2.78	4.60	8.60
5	1.48	2.02	2.57	4.03	6.86
6	1.44	1.94	2.45	3.71	5.96
7	1.42	1.90	2.36	3.50	5.40
8	1.40	1.86	2.31	3.36	5.04
9	1.38	1.83	2.26	3.25	4.78
10	1.37	1.81	2.23	3.17	4.59
11	1.36	1.80	2.20	3.11	4.44
12	1.36	1.78	2.18	3.06	4.32
13	1.35	1.77	2.16	3.01	4.22
14	1.34	1.76	2.14	2.98	4.14
∞ $(t = z)$	1.29	1.64	1.96	2.58	3.29

Tabela III Critical Values for Rejection Quotient Q

Number of Observations	Q_{crit} (Reject if $Q_{exp} > Q_{crit}$)		
	90% Confidence	96% Confidence	99% Confidence
3	0.94	0.98	0.99
4	0.76	0.85	0.93
5	0.64	0.73	0.82
6	0.56	0.64	0.74
7	0.51	0.59	0.68
8	0.47	0.54	0.63
9	0.44	0.51	0.60
10	0.41	0.48	0.57

* Reproduced from W. J. Dixon, *Ann. Math. Stat.*, 22, 68 (1951).

Tabela IV Critical Values for F at the Five Percent Level

Degrees of Freedom (Denominator)	Degrees of Freedom (Numerator)							
	2	3	4	5	6	12	20	∞
2	19.00	19.16	19.25	19.30	19.33	19.41	19.45	19.50
3	9.55	9.28	9.12	9.01	8.94	8.74	8.66	8.53
4	6.94	6.59	6.39	6.26	6.16	5.91	5.80	5.63
5	5.79	5.41	5.19	5.05	4.95	4.68	4.56	4.36
6	5.14	4.76	4.53	4.39	4.28	4.00	3.87	3.67
12	3.89	3.49	3.26	3.11	3.00	2.69	2.54	2.30
20	3.49	3.10	2.87	2.71	2.60	2.28	2.12	1.84
∞	3.00	2.60	2.37	2.21	2.10	1.75	1.57	1.00

Tabela V CHI-SQUARE DISTRIBUTION

N−1	P								
	0.995	0.99	0.95	0.90	0.50	0.10	0.05	0.01	0.005
1	0.00004	0.00016	0.0039	0.0158	0.455	2.71	3.84	6.63	7.78
2	0.0100	0.0201	0.1026	0.211	1.39	4.61	5.99	9.21	10.6
3	0.0717	0.115	0.352	0.584	2.37	6.25	7.81	11.3	12.8
4	0.207	0.297	0.711	1.064	3.36	7.78	9.49	13.3	14.9
5	0.412	0.554	1.15	1.61	4.35	9.24	11.1	15.1	16.7
6	0.676	0.872	1.64	2.20	5.35	10.6	12.6	16.8	18.5
7	0.99	1.239	2.17	2.83	6.35	12.0	14.1	18.5	20.3
8	1.34	1.65	2.73	3.49	7.34	13.4	15.5	20.1	22.0
9	1.73	2.09	3.33	4.17	8.34	14.7	16.9	21.7	23.6
10	2.16	2.56	3.94	4.87	9.34	16.0	18.3	23.2	25.2
15	4.60	5.23	7.26	8.55	14.3	22.3	25.0	30.6	32.8
20	7.43	8.26	10.85	12.44	19.3	28.4	31.4	37.6	40.0
30	13.79	14.95	18.49	20.60	29.3	40.3	43.8	50.9	53.7

Regressão linear

Método dos mínimos quadrados - minimiza o quadrado dos desvios duma linha recta definida por

$$y = mx + b$$

Os cálculos dão os seguintes resultados:

Inclinação $$m = \frac{\Sigma x_i \Sigma y_i - N\Sigma x_i y_i}{(\Sigma x_i)^2 - N\Sigma x_i^2}$$

Intercepção $$b = \bar{y} - m\bar{x}$$

Coeficiente de correlação $$r = \frac{m\sigma_x}{\sigma_y}$$

Erro na inclinação dado por $$\sigma_m^2 = \left(\frac{m}{r}\right)^2 \cdot \frac{1 - r^2}{N - 2}$$

Erro na intercepção dado por $$\sigma_b^2 = \sigma_m^2 \cdot \frac{\Sigma x_i^2}{N}$$

Bibliografia

Skoog e West Cap.3

Laitinen e Harris Cap.26

(.... e qualquer livro de estatística)