3. ANÁLISE DE DADOS EXPERIMENTAIS

3.1 – Introdução.

Todo dado experimental deve ser analisado através de algum tipo de procedimento. Um bom experimentalista deve fazer todo o esforço possível para eliminar todos os erros de seu experimento. Este objetivo, no entanto, nunca será plenamente alcançado, cabendo então ao experimentalista a responsabilidade de apresentar uma medida da confiabilidade de seus dados.

Vamos definir erro como sendo:

a diferença entre o valor medido e o valor verdadeiro de uma grandeza

Normalmente não se conhece o valor verdadeiro de uma grandeza, o que torna esta definição difícil de ser aplicada.

Podemos então definir o conceito de incerteza:

incerteza experimental é um valor possível que o erro pode assumir. Define uma faixa onde se estima estar localizado o valor da grandeza medida (dentro de um determinado nível de probabilidade).

Em termos da estimativa da confiabilidade, as experiências podem ser divididas em duas categorias:

- experiências de várias amostragens
- experiências de uma única amostragem

Idealmente gostaríamos de poder repetir medidas várias vezes, usando vários instrumentos e observadores para que a confiabilidade dos resultados pudesse ser determinada usando métodos estatísticos (experiências de várias amostragens). Normalmente, os custos e o tempo associados a este tipo de experimento o tornam proibitivo.

Note que muitos experimentos que parecem, a princípio, ser de várias amostragens são na realidade de amostragem única. Por exemplo, várias medidas com um mesmo voltímetro carregam a incerteza inerente àquele voltímetro.

Nosso objetivo na análise de dados experimentais é responder a três perguntas básicas:

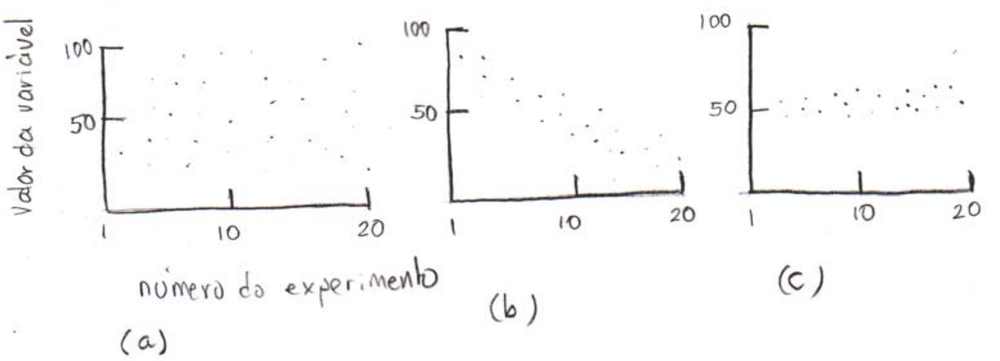
- como estimar e descrever a incerteza em uma determinada variável?
- como calcular a propagação destas incertezas no resultado final?
- como apresentar os resultados de modo a dar, de forma concisa, uma medida da confiabilidade dos resultados?

3.2 – A Natureza das Incertezas Experimentais.

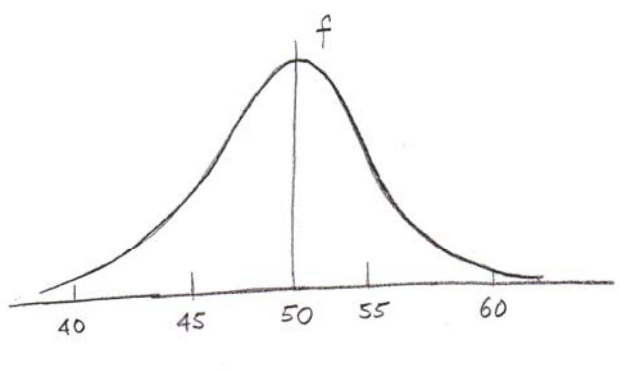
Para chegarmos a um método de descrição das incertezas nas variáveis, precisamos conhecer a natureza das incertezas. Podemos classificar as incertezas experimentais em:

- incertezas aleatórias: são incertezas que fazem com que medidas repetidas apresentem valores diferentes. Por exemplo, atrito no instrumento, flutuações eletrônicas, flutuações na leitura pelo observador, etc.
- incertezas sistemáticas: são incertezas que fazem com que medidas repetidas apresentem aproximadamente o mesmo desvio (positivo ou negativo) sem razão aparente (se a razão fosse conhecida uma correção poderia ser feita). Por exemplo, um pequeno amassado em um tubo de Pitot, ou perdas de calor pelo corpo de um termômetro.

As incertezas experimentais podem ser estudadas tomando-se várias observações do valor da variável. Podemos obter os resultados apresentados na figura abaixo. Os experimentos dos casos (a) e (b) não são *bem controlados*. O experimento do caso (c) é *bem controlado*.



Caso tenhamos um número muito grande de dados disponíveis, podemos construir um histograma ou uma curva de distribuição de freqüência de ocorrência de um certo valor, tal como mostrado esquematicamente na figura abaixo.



Incertezas aleatórias normalmente apresentam uma distribuição como a mostrada na figura. Analisando a distribuição dos resultados podemos fazer as seguintes observações:

- pequenos desvios ocorrem mais freqüentemente
- desvios positivos ou negativos ocorrem com igual freqüência
- não existe limite superior ou inferior para o desvio

Muitos autores assumem a distribuição das incertezas aleatórias como sendo Gaussiana. Isto nem sempre é verdade. A forma da distribuição para um certo conjunto de dados experimentais pode ser verificada através da construção de um histograma.

3.3 – Distribuição Normal ou Gaussiana.

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]}$$

onde *P* é a probabilidade de que, em uma distribuição com média μ e desvio padrão σ, o valor da observação aleatória seja *x*.

Assumindo-se que os dados experimentais sejam bem representados por uma distribuição normal, os dados podem ser *amarrados* por dois números:

- média, o valor mais provável da variável
- desvio padrão, uma medida do espalhamento em torno da média

A média é dada por:

$$\mu = \int_{-\infty}^{+\infty} xP(x)dx$$

A estimativa da média para uma amostra finita é dada por:

$$\bar{x} = \frac{1}{N}\sum_{1}^{N} x_i$$

O desvio padrão é dado por:

$$\sigma^2 = \int_{=\infty}^{+\infty} (x-\mu)^2 P(x)dx$$

A estimativa do desvio padrão para amostras finitas é dada por:

$$\sigma = \left[\frac{1}{N} \sum (x_i - \bar{x})^2 \right]^{1/2}$$ para N grande (> 30, por exemplo)

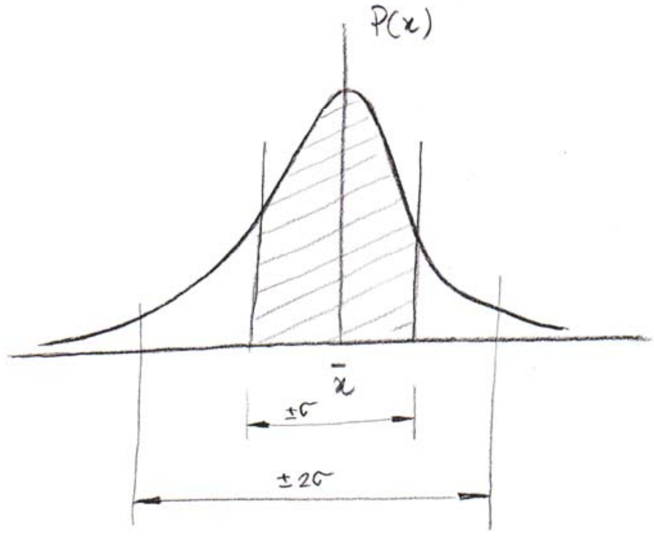
$$\sigma = \left[\frac{1}{N-1} \sum (x_i - \bar{x})^2 \right]^{1/2}$$ para N pequeno

Para uma distribuição normal de média $\bar{x}$ e desvio padrão σ, temos que a probabilidade de uma medida x estar compreendida em uma certa faixa $\pm x_1$ da média é:

$$P = \int_{\bar{x}-x_1}^{\bar{x}+x_1} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx$$

Isto representa a área abaixo da curva normal, como mostrado na figura. Para uma faixa em torno da média dada em termos de números de desvios padrão, a área abaixo da curva pode ser calculada e associada probabilidade de ocorrência de um valor dentro da faixa. Assim, para $\mu - n\sigma < x < \mu + n\sigma$, tem-se:

a) $n = 1, P = 0{,}683$ ou 2,15:1
b) $n = 2, P = 0{,}954$ ou 20:1
c) $n = 3, P = 0{,}997$ ou 356:1



Note a diferença entre os conceitos de métodos estatísticos e avaliação de incertezas no estudo de dados experimentais:

- métodos estatísticos só podem ser usados quando os dados existem!
- avaliação de incertezas é uma previsão

Em uma experiência de amostragem única não há como "fazer estatística". O máximo que pode ser feito é indicar-se o que aconteceria caso o experimento fosse repetido um grande número de vezes.

3.4 – Incerteza na Medição.

Uma notação satisfatória para uma medição de uma variável deve incluir:

- a melhor estimativa do valor verdadeiro da variável medida
- uma indicação da magnitude do desvio esperado para esta estimativa (ou seja, a incerteza)

A melhor estimativa do valor verdadeiro é normalmente dada pelo valor medido (ou pela média, caso existam vários valores medidos da variável).

Uma medida da confiabilidade da medida é dada pela incerteza. A incerteza experimental é dada pela faixa $\bar{x} \pm 2\sigma$, onde a média $\bar{x}$ tem 95,4 % de propabilidade de ocorrência. A escolha de $\pm 2\sigma$ é arbitrária. Então, uma maneira completa de se reportar uma medida é:

$m \pm \delta m$ *(b:1)* ou

$m \pm \delta m/m$ *(b:1)*

onde δm é a incerteza absoluta e $\delta m/m$ é a incerteza relativa.

Por exemplo, uma medida de temperatura seria reportada de maneira completa como 15,7 ±0,5 ^{o}C (20:1).

A determinação da faixa de incerteza δm é tarefa do experimentalista. Esta pode ser estimada por:

a) pré-testes onde são feitas várias medidas e pode-se calcular σ
b) estimativas das incertezas dos instrumentos fornecidas pelo fabricante
c) bom senso. O experimentalista é a pessoa mais indicada para fazer uma boa estimativa da incerteza.

O *Guia para Expressão da Incerteza de Medição* é uma publicação internacional editada pelo *INMETRO*, onde os conceitos e definições sobre incerteza de medidas são apresentados. As recomendações contidas neste guia devem ser obedecidas no Brasil. No guia, as incertezas obtidas como descrito no item a) acima são denominadas incertezas do Tipo A. Aquelas obtidas pelos procedimentos descritos nos itens b) ou c) são denominadas incertezas do Tipo B.

3.5 – Propagação de Incertezas.

Agora que sabemos como estimar a faixa de incerteza para uma variável individual, precisamos avaliar como estas incertezas se propagam em um resultado. Suponha que um resultado R é função de n variáveis independentes

$R = R(x_1, x_2, x_3, \ldots x_n)$

Uma possível maneira de se estimar a incerteza final no resultado R, pode ser obtida através da chamada *combinação da pior situação.* Assim,

$$\delta R = \left|\frac{\partial R}{\partial x_1}\delta x_1\right| + \left|\frac{\partial R}{\partial x2}\delta x2\right| + \ldots + \left|\frac{\partial R}{\partial x_n}\delta x_n\right|$$

onde δR é a incerteza no resultado R, e δx_i é a incerteza em cada variável x_i. As derivadas parciais $\partial R / \partial x_i$ são chamadas de coeficientes de sensibilidade e medem o quão sensível o resultado R é a cada variável x_i.

É pouco provável que as incertezas individuais se combinem da pior maneira possível como propõe a expressão acima. As previsões obtidas com esta expressão são normalmente exageradas quando comparadas com experimentos reais.

Kleine e McClintock (Mechanical Engineering, vol 75, Jan, 1953, pg3) propuseram uma outra forma de calcular a propagação de incertezas experimentais que fornece boa previsão das incertezas.

$$\delta R^2 = \left(\frac{\partial R}{\partial x_1}\delta x_1\right)^2 + \left(\frac{\partial R}{\partial x_2}\delta x_2\right)^2 + \ldots. + \left(\frac{\partial R}{\partial x_n}\delta x_n\right)^2$$

Nesta expressão, os níveis de probabilidade das medidas individuais são preservados na grandeza R. Isto é, se δx_i é conhecido dentro de $\pm 2\sigma$, δR será obtido também dentro de $\pm 2\sigma$. Esta expressão apresenta resultados satisfatórios e é amplamente utilizada.

3.5.1 – Exemplo 1: Medida de velocidade de fluido com tubo de Pitot

Sabemos que a velocidade medida por um tubo de Pitot é dada pela seguinte expressão:

$$V = \sqrt{\frac{2\Delta p}{\rho}}$$

assumindo-se que o fluido comporte-se como um gás perfeito, $\frac{p}{\rho} = RT$, temos:

$$V = \sqrt{\frac{2\Delta p\, R T}{p}}$$

Suponha que a seguinte medida foi realizada:

Δp = 2 ± 0,03 kPa (20:1)
p = 100 ± 2 kPa (20:1)
T = 300 ± 0,2 K (20:1)

Para este conjunto de valores obtemos que *V = 58 m/s*

Suponha que R = *287 Nm/KgK* e que não haja incerteza na constante do gás (isto não é verdade! A constante do gás também possui incerteza).

Aplicando a expressão de *Kleine e McClintock* para a propagação das incertezas obtemos:

$$\delta V = \left\{ \left[\frac{\partial V}{\partial(\Delta p)} \delta(\Delta p) \right]^2 + \left[\frac{\partial V}{\partial T} \delta T \right]^2 + \left[\frac{\partial V}{\partial p} \delta p \right]^2 \right\}^{1/2}$$

onde as derivadas parciais, neste caso, podem ser obtidas de forma analítica como:

$$\frac{\partial V}{\partial(\Delta p)} = \frac{1}{2} \left(\frac{2RT}{p\,\Delta p} \right)^{1/2}, \qquad \frac{\partial V}{\partial p} = -\frac{1}{2} \left(\frac{2\,\Delta p\,R\,T}{p^3} \right)^{1/2}, \qquad \frac{\partial V}{\partial T} = \frac{1}{2} \left(\frac{2\,\Delta p\,R}{p\,T} \right)^{1/2}$$

substituindo e dividindo por *V:*

$$\frac{\delta V}{V} = \left\{ \left[\frac{1}{2} \frac{\delta(\Delta p)}{\Delta p} \right]^2 + \left[\frac{1}{2} \frac{\delta p}{p} \right]^2 + \left[\frac{1}{2} \frac{\delta T}{T} \right]^2 \right\}^{1/2}$$

substituindo-se os valores numéricos:

$$\frac{\delta V}{V} = \left\{ \left[\frac{1}{2} \frac{0{,}03}{2} \right]^2 + \left[\frac{1}{2} \frac{2}{100} \right]^2 + \left[\frac{1}{2} \frac{0{,}2}{300} \right]^2 \right\}^{1/2}$$

$$\frac{\delta V}{V} = \left\{ 5{,}6x10^{-5} + 1x10^{-4} + 1.1x10^{-7} \right\}^{1/2} = 0{,}0125 = 1{,}25\%$$

Uma análise das parcelas da soma acima mostra que as contribuições das incertezas na medida da diferença de pressão e da pressão para a incerteza na velocidade são da mesma ordem de grandeza. Por outro lado, a contribuição da incerteza da temperatura é desprezível em comparação com as outras grandezas. Assim, vê-se que não há necessidade da medição da temperatura com tão baixo nível de incerteza e que os recursos utilizados na aquisição do sensor de temperatura poderiam ter sido melhor utilizados na compra de sensores de pressão e diferença de pressão com menores incertezas. O resultado final da medição é dado por:

V = 58 ± 1,25% m/s (20:1)

3.5.2 – Exemplo 2: Medida da diferença de duas grandezas.

Suponha que o resultado que se deseja, *R,* seja obtido pela diferença da medição de duas grandezas, *x* e *y* da seguinte forma:

$$R = x - y$$

Suponha que as medições forneceram *x =1* e *y = 0,98.* Ainda, suponha que as incertezas relativas nas duas grandezas sejam $\frac{\delta x}{x} = \frac{\delta y}{y} = \pm 1\%$

A incerteza no resultado final *R* pode ser avaliada por:

$$\delta R^2 = \left(\frac{\partial R}{\partial x}\delta x\right)^2 + \left(\frac{\partial R}{\partial y}\delta y\right)^2$$

com $\frac{\partial R}{\partial x} = 1$ *e* $\frac{\partial R}{\partial y} = -1$, tem-se $\delta R^2 = (\delta x)^2 + (-\delta y)^2$ ou

$$\left(\frac{\partial R}{R}\right)^2 = \left(\frac{x}{x-y}\frac{\delta x}{x}\right)^2 + \left(\frac{y}{x-y}\frac{\delta y}{y}\right)^2$$

substituindo-se os valores numéricos:

$$\left(\frac{\partial R}{R}\right) = \left\{\left(\frac{1}{1-0{,}98}\cdot 0{,}01\right)^2 + \left(\frac{0{,}98}{1-0{,}98}\cdot 0{,}01\right)^2\right\}^{1/2} = 0{,}71 = 71\% \text{ !!!}$$

3.6 – Incerteza na Estimativa da Média de uma Amostra Finita.

Vimos que a média e o desvio padrão caracterizam uma distribuição Gaussiana. Estes parâmetros são determinados a partir de um número muito grande de medidas. Entretanto, normalmente, dispõe-se de um número limitado de medidas, tornando-se necessário estimar a incerteza da média obtida através da expressão:

$$\bar{x} = \frac{1}{N}\sum_{1}^{N} x_i$$

Podemos associar a cada medição x_i um desvio padrão σ_i. Assim,

$$R = \bar{x} = \frac{x_1 + x_2 + ... + x_N}{N} = \frac{1}{N}\sum x_i$$

o coeficiente de sensibilidade $\frac{\partial R}{\partial x_i} = \frac{1}{N}$

então,

$$\sigma_R^2 = \sigma_1^2\left(\frac{\partial R}{\partial x_1}\right)^2 + \sigma_2^2\left(\frac{\partial R}{\partial x_2}\right)^2 + ... + \sigma_N^2\left(\frac{\partial R}{\partial x_N}\right)^2$$

tomando-se os σ_i como senso iguais,

$$\sigma_R^2 = \frac{1}{N^2}\left(\sigma_1^2 + \sigma_2^2 + ... + \sigma_N^2\right) = \frac{1}{N^2}\left(N\sigma^2\right) = \frac{\sigma^2}{N}$$

ou

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

Suponha, como exemplo, dez medidas da massa de um corpo em gramas:

86, 85, 84, 89, 85, 89, 87, 85, 83, 85g

A média é $\overline{m} = 85{,}7g$ e o desvio padrão $\sigma_m = 2{,}16g$. Então, o desvio padrão da média é dado por:

$$\sigma_{\overline{m}} = \frac{\sigma_m}{\sqrt{10}} = 0{,}7g$$

Assim, a medida é dada por m = 86,7g ±0,7g. Se utilizarmos a faixa de incertezas como $\pm 2\sigma$, temos:

m = 86,7g ±1,4g

Note que $\sigma_{\bar{x}}$ cai com $1/\sqrt{N}$. Assim, para diminuirmos a incerteza da média por um fator de, digamos, 10, precisamos aumentar o número de medições que compõem a média de um fator de 100. Note também que as incertezas sistemáticas não serão reduzidas com o aumento de N.

3.7 – Rejeição de Dados Experimentais.

Algumas vezes uma certa medida de uma série de medidas parece desviar significativamente das outras medidas. Neste caso, o experimentalista deve decidir se esta medição é o resultado de algum erro grosseiro cometido e deve, portanto, ser descartada. Obviamente, caso seja possível, deve-se repetir os experimentos para confirmar o resultado duvidoso. Caso não seja possível repetir o experimento, seria interessante ter-se um critério que auxiliasse a decisão sobre a rejeição ou não de um determinado ponto experimental. O Critério de *Chauvenet* para a rejeição de pontos ruins pode ser utilizado.

O critério de *Chauvenet* estabelece que uma determinada leitura pode ser rejeitada se a probabilidade de obter um desvio particular em relação à média for menor que 1/2N, onde N é o número de medições realizadas.

A probabilidade de um determinado desvio $x - \mu$ ocorrer é dada por:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

Definindo o desvio de uma medida como $d = x_{susp} - \bar{x}$, onde x_{susp} é a medida suspeita em questão, pelo critério de *Chauvenet*, o máximo desvio aceitável para uma amostra formada por N pontos, é dado por:

$$\frac{1}{2N} = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2}\left(\frac{d_{\max}}{\sigma}\right)^2\right]$$

caso $d > d_{max}$, o ponto em questão pode ser rejeitado.

A tabela abaixo apresenta valores para a razão entre o máximo desvio aceitável e o desvio padrão, para diferentes valores de *N,* obtidos da expressão acima.

Tabela 1

No. de Medidas (N)	d_{max}/σ
2	1,15
3	1,38
4	1,54
5	1,65
6	1,73
7	1,80
10	1,96
15	2,13
25	2,33
50	2,57
100	2,81
500	3,29
1000	3,48

Para eliminar pontos ruins seguindo o critério de *Chauvenet*, procedemos da seguinte forma:

a) mede-se a variável um número *N* de vezes e estima-se média e desvio padrão da disribuição

$$\bar{x} = \frac{1}{N}\sum_{1}^{N} x_i \ , \qquad \sigma = \left[\frac{1}{N-1}\sum\left(x_i - \bar{x}\right)^2\right]^{1/2}$$

b) calcula-se o desvio entre cada medida e a média dividindo-se o resultado por σ

$$\frac{d_i}{\sigma} = \frac{x_i - \bar{x}}{\sigma}$$

c) usando um número de leituras N, comparar o valor de d_i/σ com d_{max}/σ. Se for maior, rejeita-se o ponto e recalcula-se a média e o desvio. O critério somente deve ser aplicado uma única vez à distribuição.

A seguir é apresentado um exemplo da aplicação do critério de *Chauvenet*. Considere os resultados na medição do comprimento de um corpo apresentados na tabela abaixo. Verifique se algum ponto pode ser rejeitado.

Leitura	**Comprimento (cm)**
1	49,36
2	50,12
3	48,98
4	49,24
5	49,26
6	50,56
7	49,18
8	49,89
9	49,33
10	49,39

a) $\bar{x} = \frac{1}{N}\sum_{1}^{N} x_i = 49{,}53$

$$\sigma = \left[\frac{1}{N-1}\sum(x_i - \bar{x})^2\right]^{1/2} = 0{,}495$$

b)

Leitura	$d = x_i - \bar{x}$	$\lvert d_i/\sigma \rvert$
1	-0,171	0,345
2	0,589	1,190
3	-0,551	1,113
4	-0,291	0,588
5	-0,271	0,547
6	1,029	2.079
7	-0,351	0,709
8	0,359	0,725
9	-0,201	0,406
10	-0,141	0,285

c) verifica-se que o desvio da leitura número 6 dividido pelo desvio padrão é o único que ultrapassa o valor de 1,96 estabelecido como o máximo desvio para uma amostra de 10 medidas, como pode ser verificado na Tabela 1. Assim, este ponto pode ser rejeitado. Recalculando-se a média e o desvio padrão obtém-se:

$\bar{x} = 49{,}42$ e $\sigma = 0{,}359$