

Probabilidade e Estatística 2010/2

Prof. Fernando Deeke Sasse
fernandodeeke@gmail.com

Departamento de Matemática
CCT - UDESC

1. O papel da Estatística na Engenharia

1. O método da engenharia (científico) e o pensamento estatístico
2. Coleta de dados
3. Modelos mecânicos e empíricos
4. Modelos de probabilidade

1-1 Método científico e Pensamento Estatístico

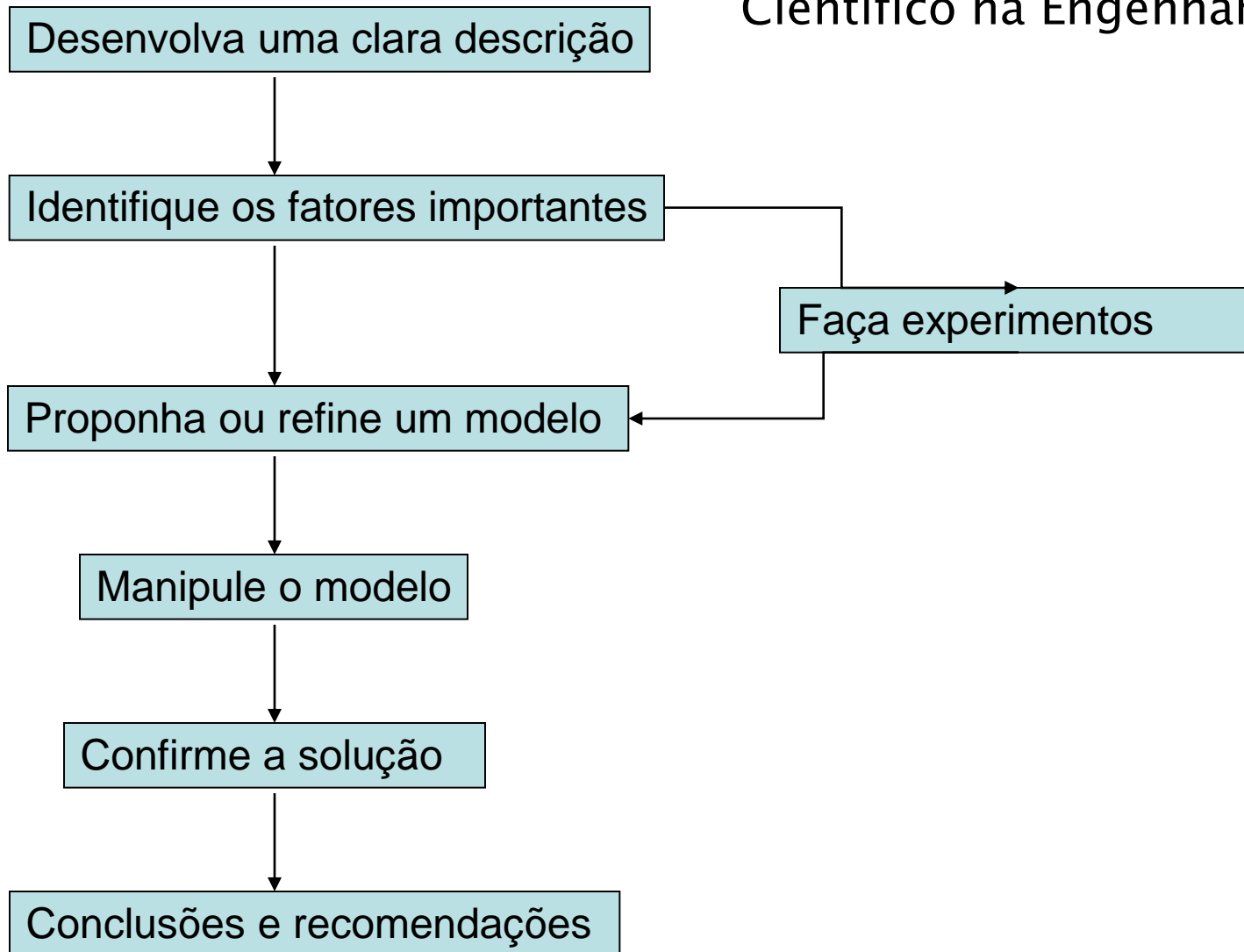
Passos:

1. Desenvolva uma descrição clara e concisa do problema.
2. Identifique, ao menos tentativamente, os fatores importantes que afetam este problema ou que podem desempenhar um papel na sua solução.
3. Proponha um modelo para o problema , usando o conhecimento científico/de engenharia do fenômeno estudado. Estabeleça as limitações ou suposições do modelo.
4. Realize experimentos apropriados e colete dados para testar ou validar o modelo tentativo ou conclusões feitas nos passos 2 e 3.

Passos (cont.) :

5. Refine o modelo com base nos dados observacionais.
6. Manipule o modelo como modo de desenvolver uma solução para o problema.
7. Conduza um experimento apropriado para confirmar que a solução proposta do problema é efetiva e eficiente.
8. Chegue a conclusões ou faça recomendações baseadas na solução do problema.

Resumo do Método Científico na Engenharia



Estatística - lista com a coleta, apresentação, análise e uso de dados para tomar decisões, resolver problemas, elaborar produtos e processos.

Métodos estatísticos – ajudam a descrever e entender variabilidade

successivas observações de um sistema ou fenômeno não produzem exatamente o mesmo resultado.

Exemplo1: Variabilidade no consumo de combustível de um veículo

Potenciais fontes de variabilidade do sistema:

- Tipo de condições de estrada (cidade ou rodovia)
- Mudanças nas condições do veículo ao longo do tempo (pressão dos pneus, compressão, desgaste de válvulas)
- Marca ou categoria de gasolina utilizada
- Condições meteorológicas
- Modo de dirigir
- Outros fatores

Estatística: fornece uma teoria para analisar estas fontes de variabilidade (quais as mais importantes ?) de forma quantitativa.

Exemplo 2 : Confecção de um conector de nylon a ser usado num motor

Problema: Determinar a especificação da espessura da parede nos conectores a serem produzidos

Especificação esperada: 3/32 in

Qual o efeito desta especificação na força de resistência do conector ?

Conector pode falhar quando a força é muito pequena

Força de resistência X de 8 protótipos de espessura 3/32 in. (pounds)

constante

12.6, 12.9, 13.4, 12.3, 13.6, 13.5, 12.6, 13.1.

$$X = \mu + \varepsilon$$

Perturbação randômica

Variável randômica

Diagrama de pontos: evidenciam a **localização** e a **variabilidade**



Média:

```
> with(stats) :  
> X:=[12.6, 12.9, 13.4, 12.3, 13.6, 13.5, 12.6, 13.1]:  
> describe [mean] (X) ;  
13.00000000
```

Engenheiro crê que tal valor da força pode ser muito baixo para atender as especificações desta aplicação.

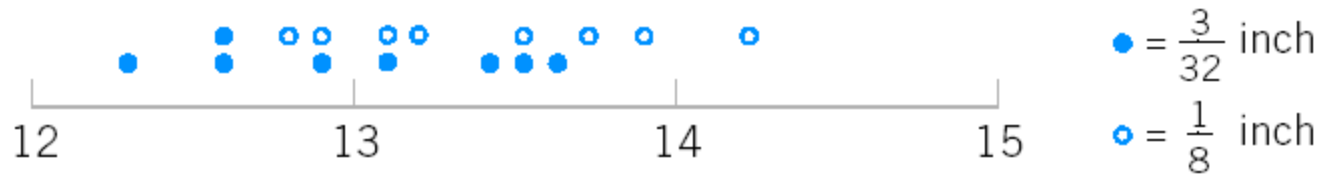


Design alternativo
com espessura
maior: 1/8 in

Força (1/8 in)

12.9, 13.7, 12.8, 13.9, 14.2, 13.2, 13.5, 13.1.

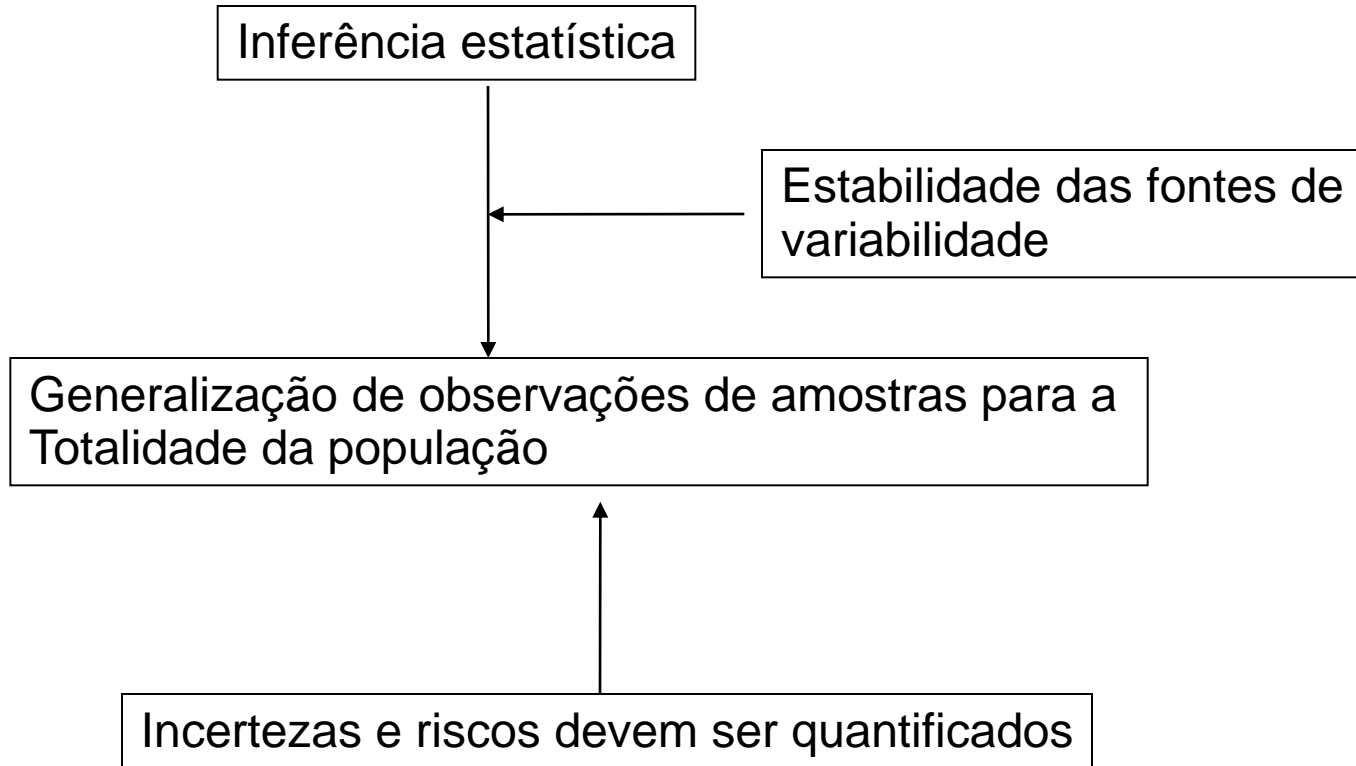
Média : 13.4



Tais resultados parecem indicar que um aumento na espessura Da parede resultou num aumento da força de resistência.

Algumas questões:

- Como sabemos que outra amostra de protótipos não dará diferentes resultados ?
- Uma amostra de oito protótipos é adequada para dar resultados confiáveis ?
- Se usarmos os resultados dos obtidos até agora para concluir que o aumento da espessura da parede aumenta a resistência, quais são os riscos associados com esta decisão ?
- É possível que o aparente aumento da força de resistência observado nos protótipos de maior espessura seja devido somente à variabilidade inerente do sistema e que o aumento da espessura do elemento (e de seu custo) não tenha realmente qualquer efeito na força de resistência ?



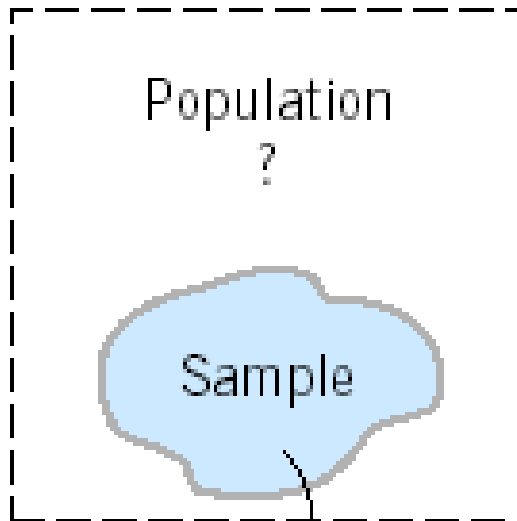
Estudo enumerativo: uma amostra é utilizada para fazer uma inferência sobre a população da qual a amostra é retirada

Estudo analítico: uma amostra é utilizada para fazer uma inferência sobre uma população ainda não existente (futura).

Análise estatística

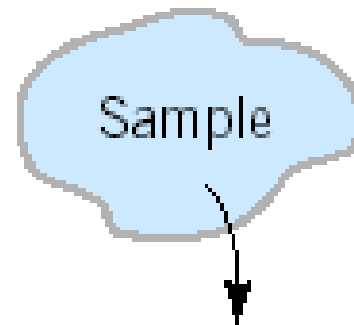
Estabilidade

Time



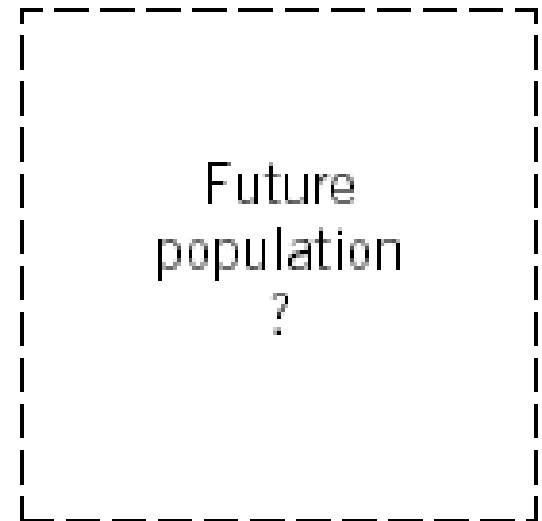
x_1, x_2, \dots, x_n

Enumerative
study



x_1, x_2, \dots, x_n

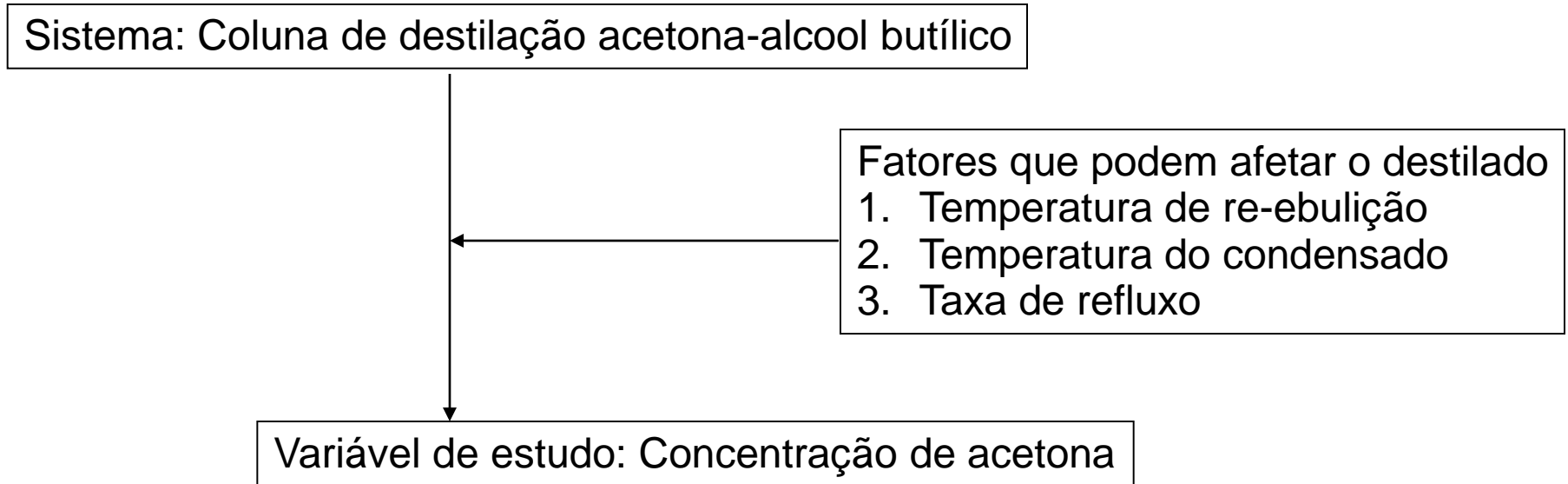
Analytic
study



1-2 Coleta de dados

1. Estudo retrospectivo usando dados históricos
2. Estudo observacional
3. Experimento projetado
4. Observação de processos ao longo do tempo

1-2.1 Estudo retrospectivo



Dados arquivados:

1. Concentração de acetona em uma amostragem a cada hora do produto de saída
1. Gráfico da temperatura de reebulição com o tempo
2. Log da temperatura do condensador
3. Taxa de refluxo nominal a cada hora (normalmente mantido constante)

Possível objetivo de estudo: descobrir as relações entre as duas temperaturas e a taxa de refluxo na concentração de acetona do produto final.

O estudo pode envolver todos os dados arquivados em um período de tempo, ou somente uma amostra


Possíveis problemas deste tipo de estudo:

1. Podemos não ser capazes de ver a relação entre a taxa de refluxo e a concentração de acetona, pois a taxa de refluxo não mudou muito ao longo do tempo considerado.
 2. Os dados arquivados das duas temperaturas (que são arquivados quase continuamente) não correspondem perfeitamente às medidas de concentração de acetona (feitas a cada hora). Não é óbvio como construir uma correspondência aproximada.
-
1. A produção mantém as duas temperaturas tão próximas quanto possível a valores determinados. Como as temperaturas mudam tão pouco, pode ser difícil obter seu impacto real na concentração de acetona.
 2. Dentro dos limitados intervalos de possível variação, a temperatura do condensado tende a subir com a temperatura de reebulição. Conseqüentemente, os efeitos dessas duas variáveis de processo na concentração de acetona podem ser difíceis de serem separados.

Em geral,

- Estudo retrospectivo pode envolver muitos dados, mas poucos com informações realmente úteis.
- Alguns dos dados relevantes podem estar ausentes.
- Podem haver erros de transcrição resultando em valores não usuais.
- Dados sobre outros fatores importantes podem não ter sido coletados e arquivados.

Análises estatísticas de dados históricos algumas vezes identificam fenômenos interessantes, mas explicações sólidas e confiáveis destes fenômenos são frequentemente difíceis de serem obtidas.



Na coluna de destilação, por exemplo, as concentrações específicas de álcool butílico e acetona no fluxo de entrada são um fator muito importante, mas elas não são arquivadas porque as concentrações são muito difíceis de serem obtidas rotineiramente.

1-2.2 Estudo observacional

- O engenheiro observa o processo ou população , perturbando o sistema minimamente, anotando as quantidades de interesse.
- Como tais estudos são usualmente conduzidos em um curto período de tempo, algumas vezes variáveis que não são medidas rotineiramente podem ser incluídas.
- No exemplo da coluna de destilação, o engenheiro elaboraria um formulário para registrar as duas temperaturas e a taxa de refluxo quando as medidas de concentração de acetona são feitas.
- É possível ainda medir as concentrações na entrada, de modo que o impacto destr fator pode ser medido.
- Em geral, um estudo observacional tende a resolver os problemas 1 e 2, mas podem não resolver 3 e 4.

1-2.3 Experimento projetado

O engenheiro faz mudanças nas variáveis controláveis do sistema ou processo, observa os dados de saída resultantes do sistema e então faz uma inferência ou decisão sobre que variáveis são responsáveis pelas observadas mudanças na performance de saída.

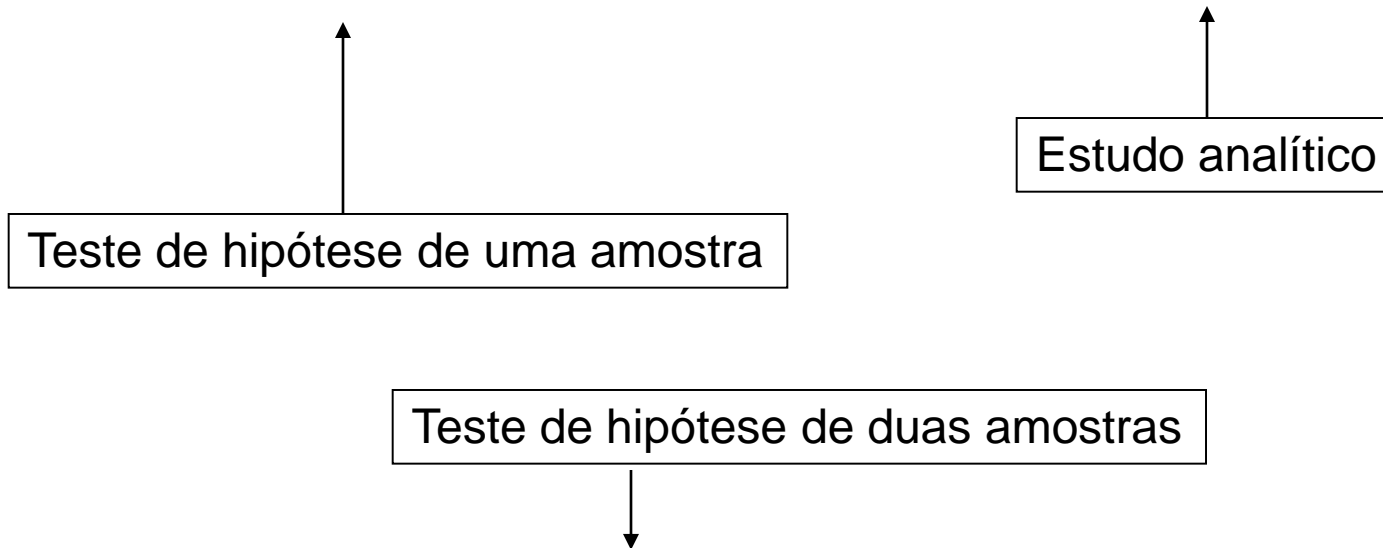
Em geral, quando produtos e processos são planejados e desenvolvidos com experimentos projetados, eles tendo uma melhor performance, maior confiabilidade e menores custos totais.

Hipótese: uma asserção sobre algum aspecto do sistema no qual estamos interessados.

Exemplo da manufatura do conector de nylon:

Teste de hipótese:

Por exemplo, o engenheiro poderia querer saber se a força média de resistência de um design de $3/32$ in poderia exceder a máxima carga esperada para este tipo de aplicação, por exemplo, 12.75 lb.

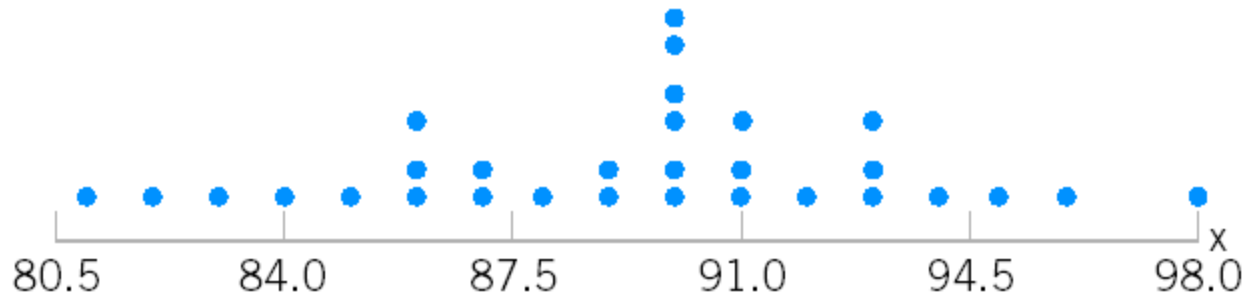


Comparação dos conectores de espessura $3/32$ e $1/8$ in para determinar se o aumento da espessura da parede do conector implica num aumento da força de resistência.

1-2.4 Observação de processos ao longo do tempo

Frequentemente dados são coletados ao longo do tempo

Gráfico de pontos temporal (leituras horárias)

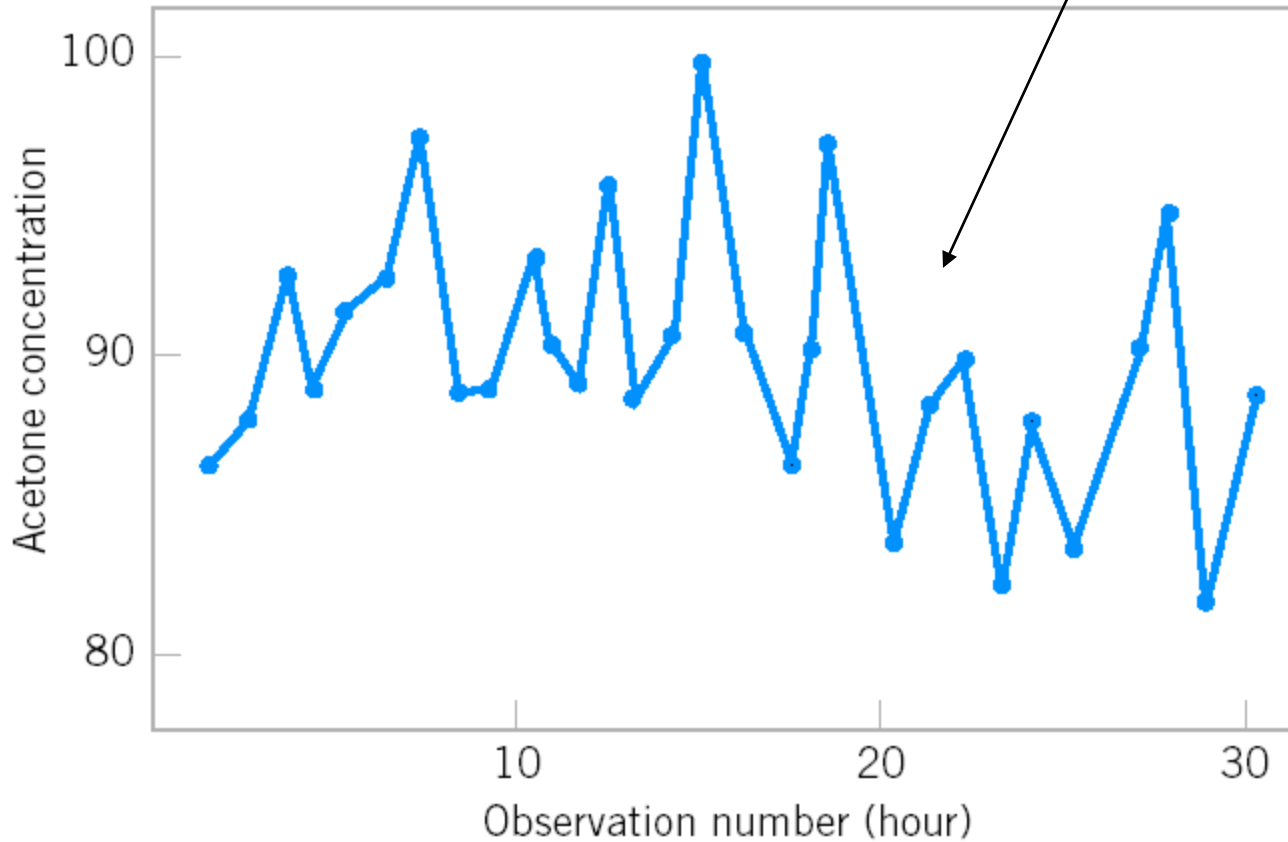


Concentração de acetona

O gráfico mostra grande variação de concentração, mas não explica a razão

Gráfico de série temporal
fornece mais informação

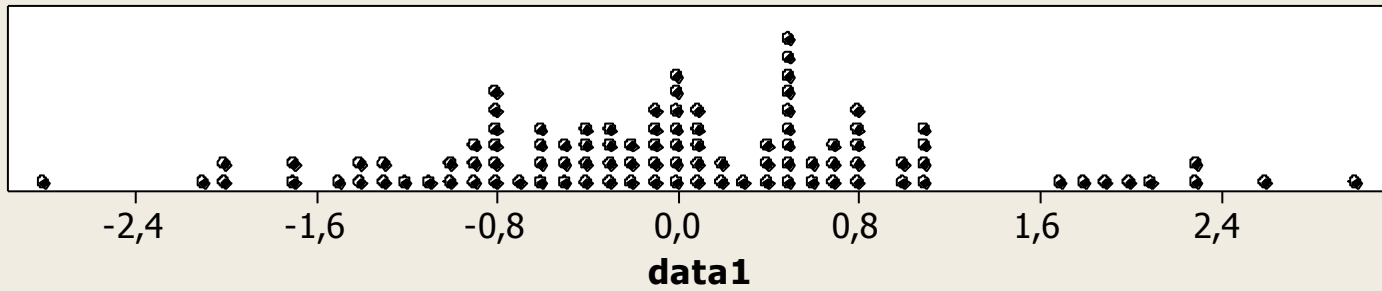
Desvio do nível médio do
processo



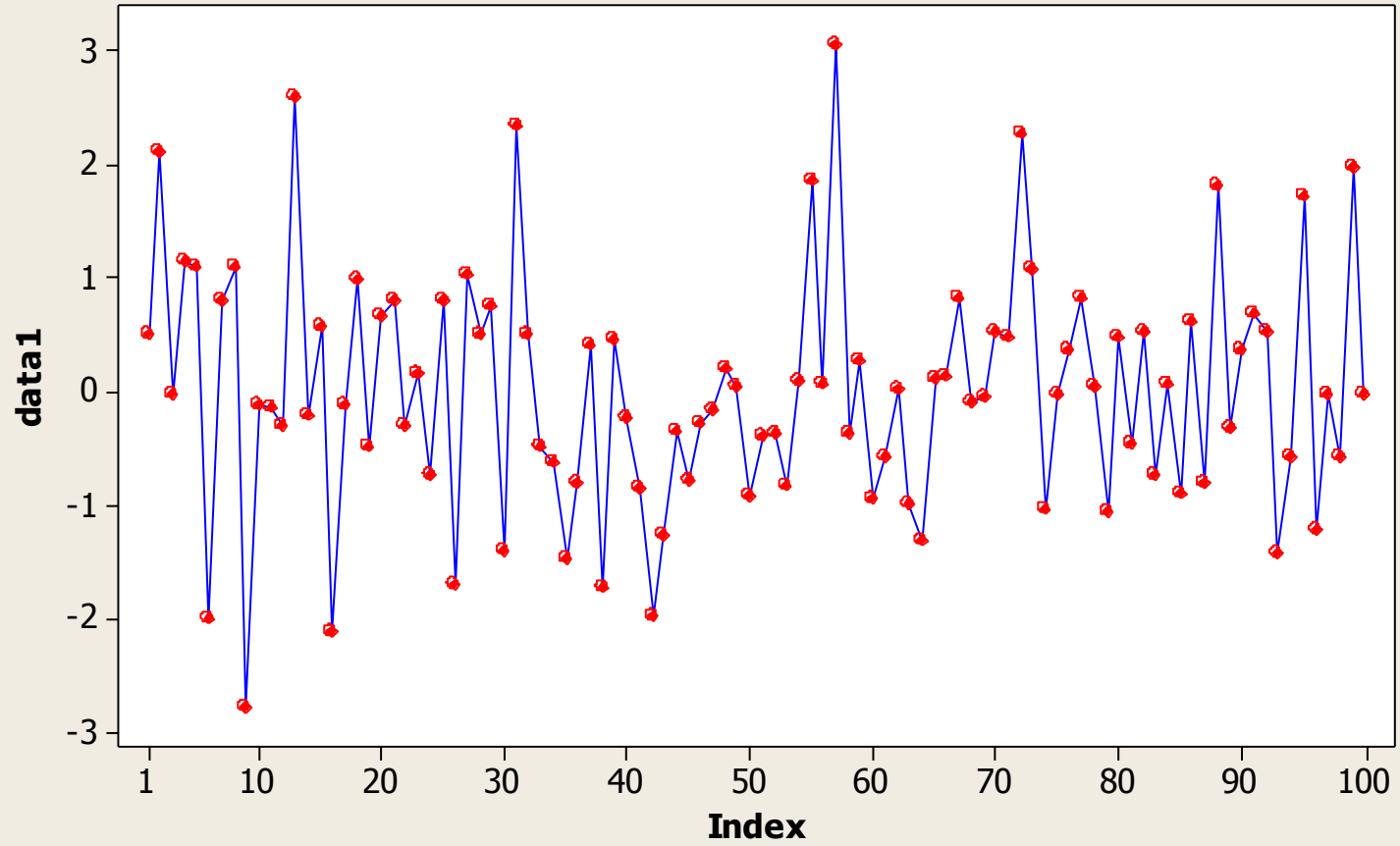
Gráficos no Minitab

```
MTB > name c1 'Strat1'  
MTB > random 100 'Strat1';  
SUBC> normal 0 1.  
MTB > dotplot c1  
MTB > tsplot c1
```

Dotplot of data1

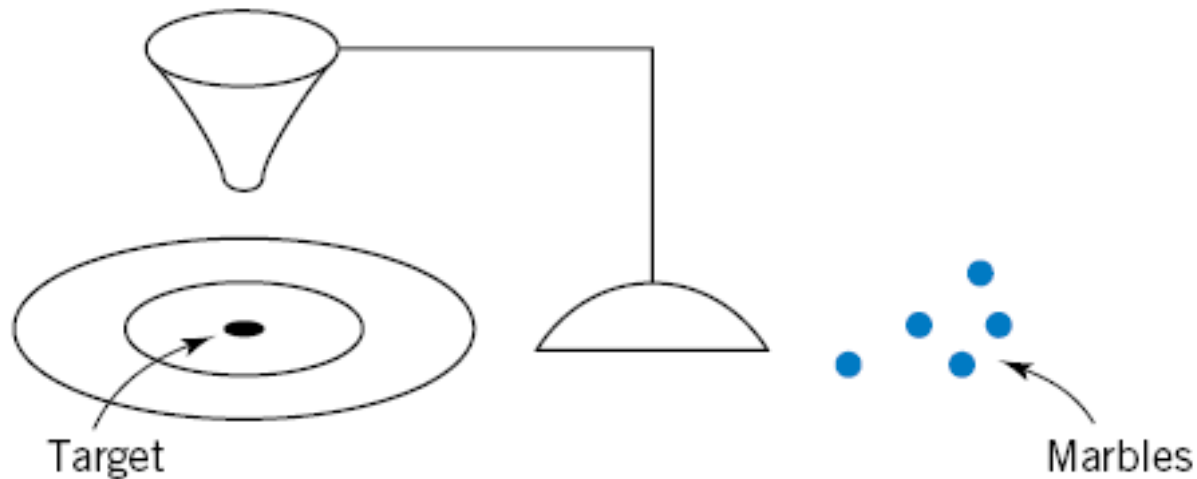


Time Series Plot of data1



Experimento de W. Edwards Deming:

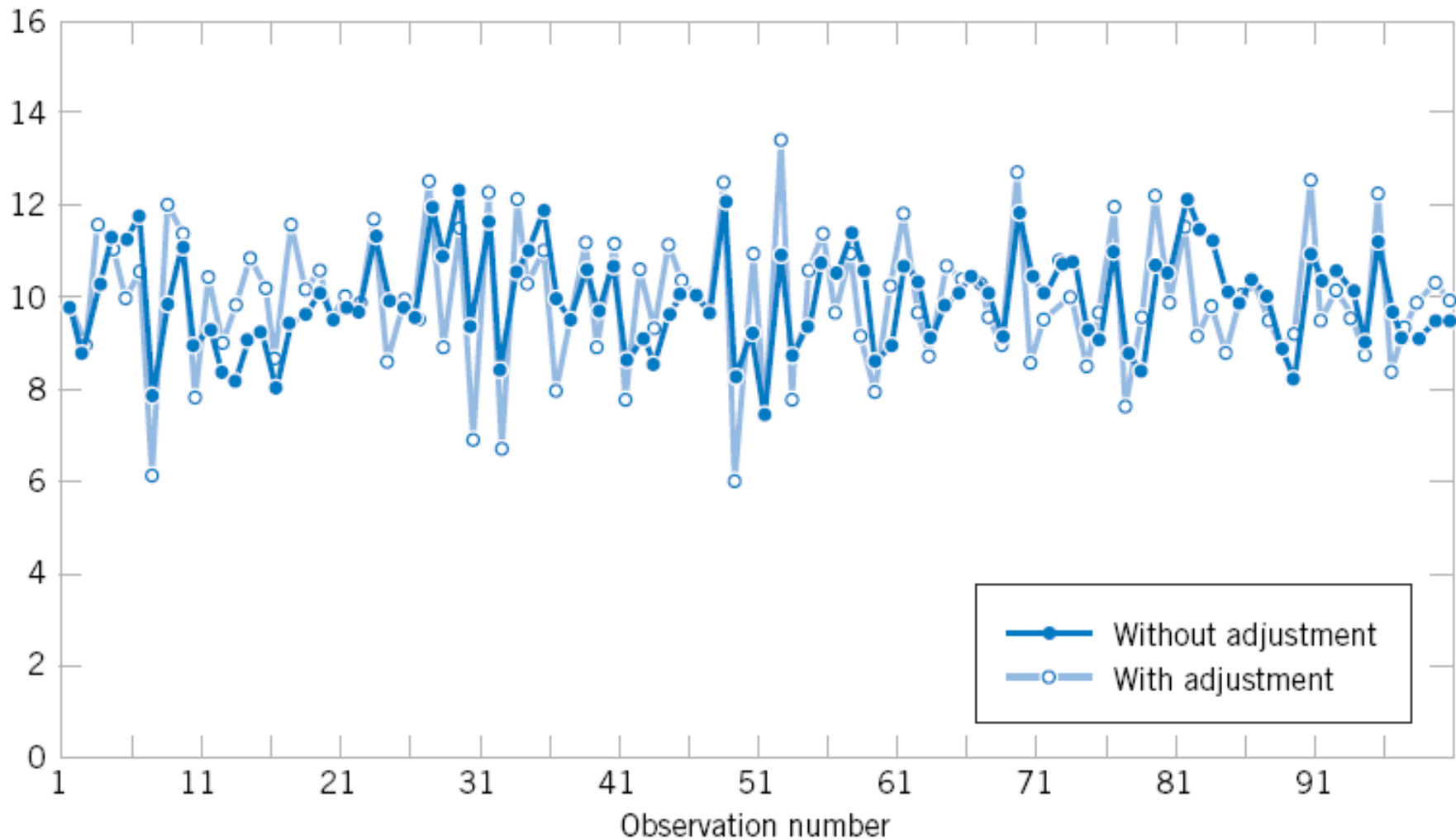
- Bolas de vidro são lançadas sobre uma alvo numa mesa através de um funil.
- O funil foi alinhado da melhor maneira possível com o centro do alvo



Estratégias para conduzir o processo:

1. O funil permanece imóvel e a distância até o alvo é registrada.
2. A primeira bola é lançada e sua localização relativamente ao alvo registrada. O funil é então movido numa direção igual e oposta, numa tentativa de compensar o erro. Tal tipo de ajuste é feito depois que cada bola é lançada.

Resultado em séries temporais (simulação)



Variabilidade da distância ao alvo:

2 vezes maior com a estratégia 2

Os ajustes do funil aumentaram os desvios do alvo.

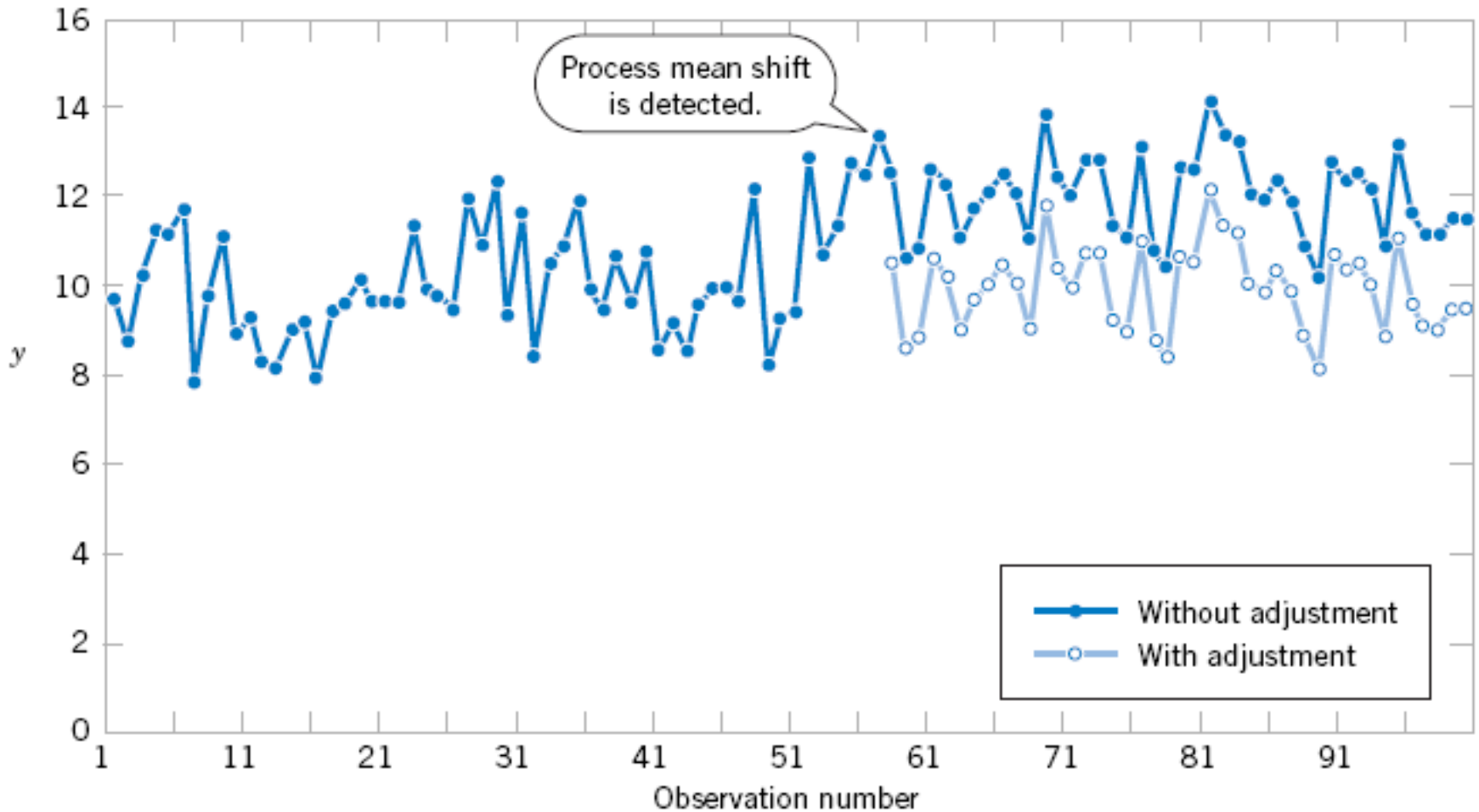
O erro para uma bola não fornece qualquer informação sobre o comportamento da próxima bola.

Ajustes do funil não diminuem futuros erros. Ao contrário, tendem a fazer com que o funil se mova para mais longe do alvo.

Ajustes em processos sujeitos a variações randômicas em geral aumentam a variação do processo

Overcontrol ou tampering

Desvio (shift) de duas unidades introduzido a partir da observação 50 e respectivo ajuste (2 un.) quando um desvio da média é detectado.



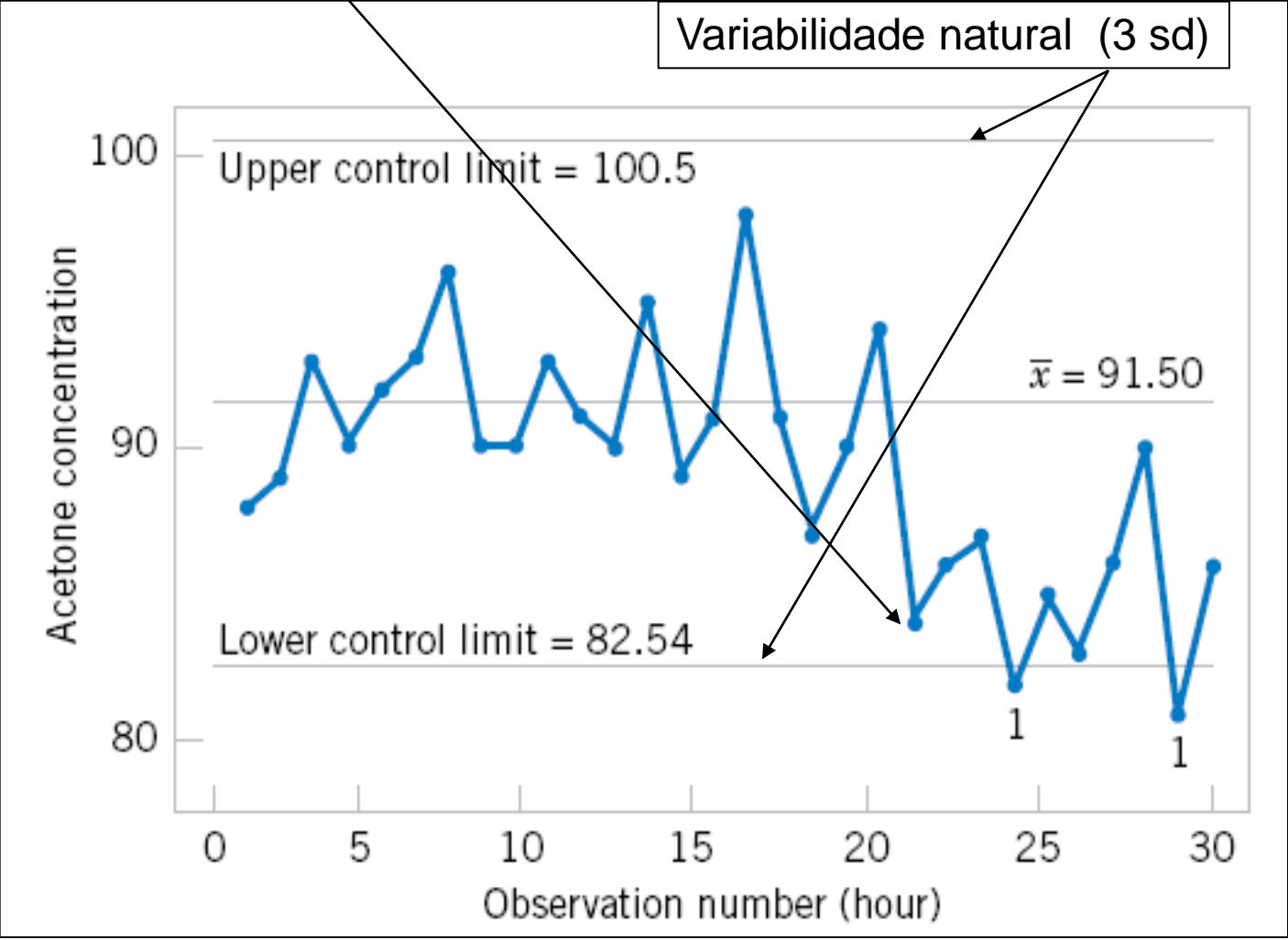
Ajuste diminuiu o erro

Fim da aula 1

Quando aplicar ajustes e em quais quantidades ?
Use uma **carta de controle**

Perturbação não-aleatória afetou o processo

Variabilidade natural (3 sd)



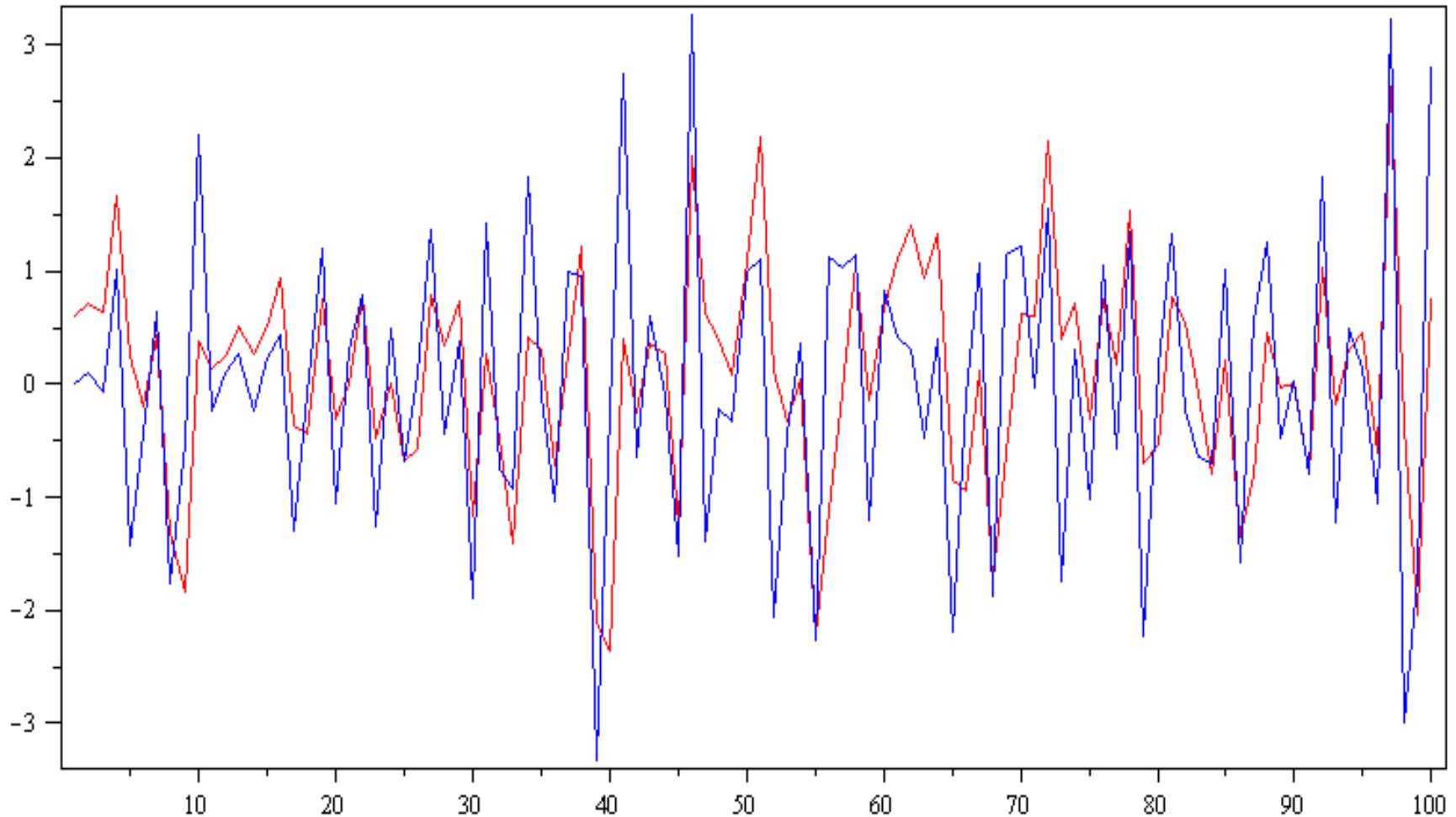
Controle de Processo Estatístico : ramo da estatística que faz uso de cartas de controle



Indica o ponto onde ação corretiva é necessária no processo

```
> s := [stats[random, normald[0]](100)];  
> p1 := [seq([i, s[i]], i = 1 .. 100)];  
> p2 := [[1, 0], seq([i, s[i]-s[i-1]], i = 2 .. 100)];  
> with(plots);  
> plot([p1, p2], axes = boxed, color = [red, blue]);
```

Simulação do Experimento de Deming no Maple



1-3 Modelos Empíricos e Mecanísticos

Suponha que queremos medir a corrente através de um fio de cobre, cujas extremidades estão sujeitas a uma diferença de potencial V .

Modelo mecanístico usual: Lei de Ohm

$$I = \frac{V}{R}$$

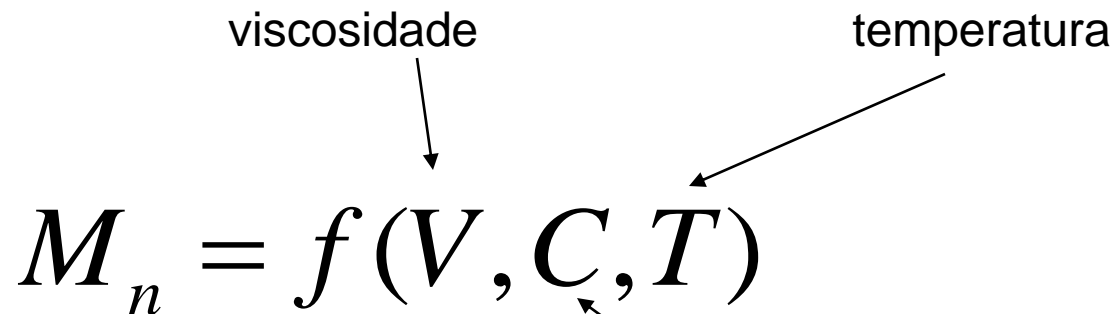
Modelo realístico da corrente observada:

$$I = \frac{V}{R} + \varepsilon$$

Fontes não modeladas de variabilidade que afetam o sistema

Alguns engenheiros trabalham com problemas para os quais não há um modelo mecânico simples ou bem entendido que explique o fenômeno

Exemplo: Determinar o peso molecular médio M_n de um polímero


$$M_n = f(V, C, T)$$


Função desconhecida

Expansão em série de Taylor até primeira ordem:

$$M_n = \beta_0 + \beta_1 V + \beta_2 C + \beta_3 T + \varepsilon$$

Modelo empírico



Fonte de variabilidade

Tal modelo usa o conhecimento científico sobre o fenômeno, mas não é construído diretamente a partir de teorias puras de primeiros princípios da teoria molecular.

Exemplo: Manufatura de Semicondutores

O semicondutor acabado é ligado por fios a um suporte.

Semiconductor die



Variáveis:

- Resistência à tração z (medida da intensidade da força necessária para romper a conexão)
- Comprimento do fio x
- Altura do chip (die) y

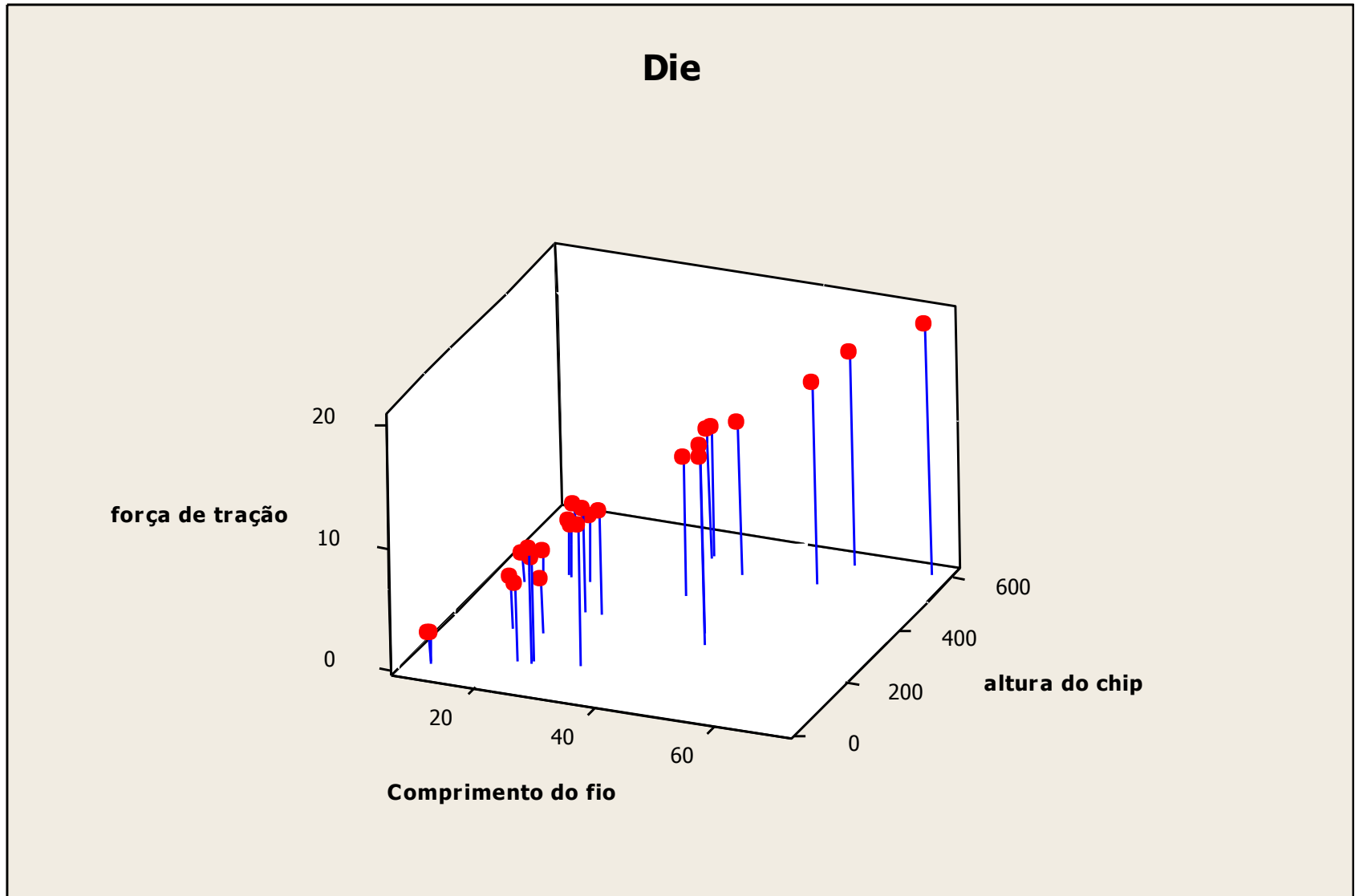
Problema: encontrar um modelo relacionando as três variáveis.

Um modelo mecanístico não é aplicável aqui

Table 1-2 Wire Bond Pull Strength Data

Observation Number	Pull Strength y	Wire Length x_1	Die Height x_2
1	9.95	2	50
2	24.45	8	110
3	31.75	11	120
4	35.00	10	550
5	25.02	8	295
6	16.86	4	200
7	14.38	2	375
8	9.60	2	52
9	24.35	9	100
10	27.50	8	300
11	17.08	4	412
12	37.00	11	400
13	41.95	12	500
14	11.66	2	360
15	21.65	4	205
16	17.89	4	400
17	69.00	20	600
18	10.30	1	585
19	34.93	10	540
20	46.59	15	250
21	44.88	15	290
22	54.12	16	510
23	56.63	17	590
24	22.13	6	100
25	21.15	5	400

Gráfico de pontos em Minitab



Maple Worksheet : mínimos cuadrados

```
>L := [[2, 50, 9.95], [8, 110, 24.45], [11, 120, 31.75], [10, 550, 35.00], [8, 295, 25.02], [4, 200, 16.86], [2, 375, 14.38], [2, 52, 9.60], [9, 100, 24.35], [8, 300, 27.50], [4, 412, 17.08], [11, 400, 37.00], [12, 500, 41.95], [2, 360, 11.66], [4, 205, 21.65], [4, 400, 17.89], [20, 600, 69.00], [1, 585, 10.30], [10, 540, 34.93], [15, 250, 46.59], [15, 290, 44.88], [16, 510, 54.12], [17, 590, 56.63], [6, 100, 22.13], [5, 400, 21.15]]:
```

```
>n := nops(L);
```

25

```
>for i to n do x[i] := L[i][1]; y[i] := L[i][2]; z[i] := L[i][3] end do;
```

25

```
>err2 := Sum((z[k]-a*x[k]-b*y[k]-c)^2, k = 1 .. n):
```

```
> eqna := diff(value(err2), a) = 0;
```

$$4792 a + 154354 b - 16016.94 + 412 c = 0$$

```
> eqnb := diff(value(err2), b) = 0;
```

5

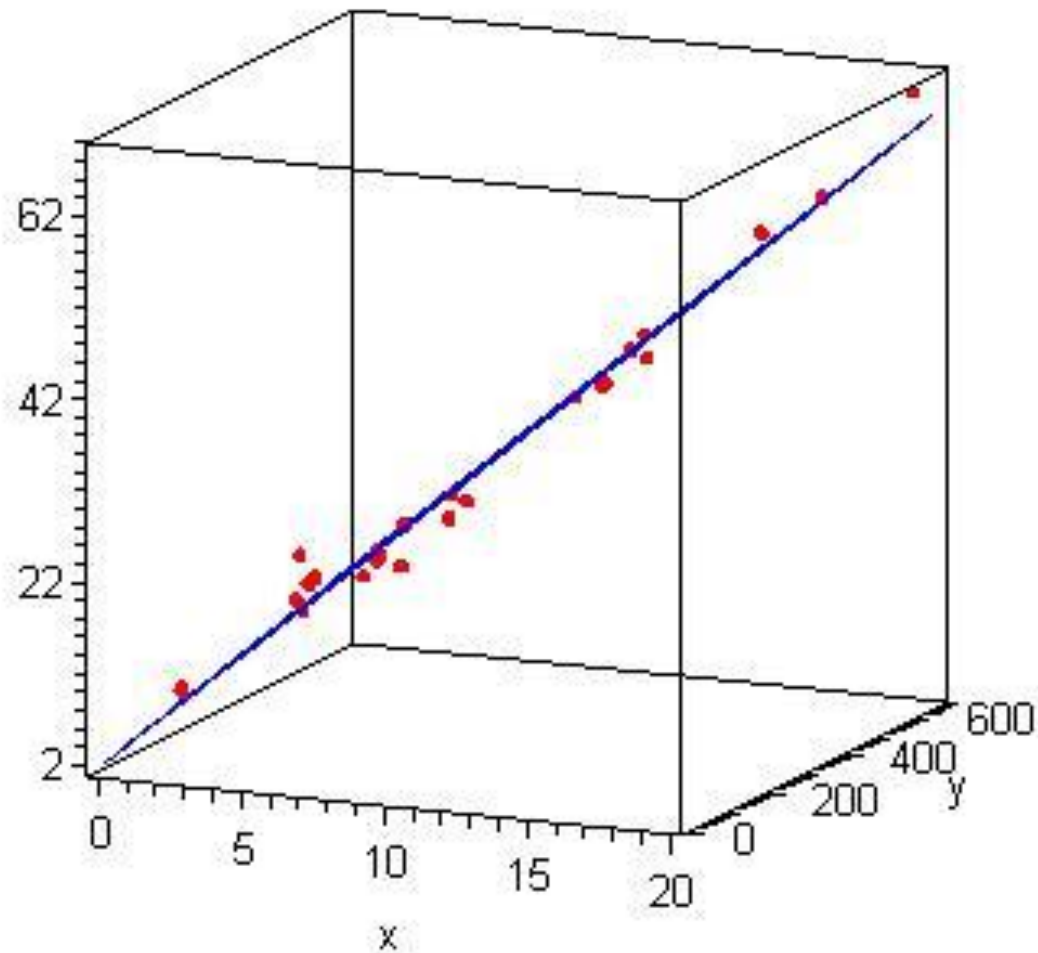
$$154354 a + 7063696 b - 5.4963342 \cdot 10^5 + 16588 c = 0$$

```
> eqnc := diff(value(err2), c) = 0;
```

$$412 a + 16588 b + 50 c - 1451.64 = 0$$

Maple Worksheet : mínimos cuadrados (cont)

```
> sols := solve({eqnb, eqnc, eqna}, {a, b, c});  
      {b = 0.01252781139, a = 2.744269643, c = 2.263791434}  
  
> assign(sols);  
  
> with(plots);  
  
> p1 := plot3d(a*x+b*y+c, x = 0 .. 20, y = 0 .. 650);  
  
> p2 := polyhedraplot(L, polyscale = .2, polytype = hexahedron);  
  
> display([p1, p2], axes = framed, shading = zhue, scaling = unconstrained);
```



```
> plano := z = a*x+b*y+c;
```

$$z = 2.744269643 x + 0.01252781139 y + 2.263791434$$

Probabilidade

Já vimos que decisões frequentemente necessitam ser baseadas somente em um subconjunto de objetos selecionados de uma amostra.

O processo de se partir de uma amostra para se chegar a conclusões sobre uma amostra foi chamado **inferência estatística**.

Exemplo: Consideremos o problema de se estudar a resistividade de uma amostra de 3 wafers semicondutores, selecionados de uma linha de produção.

Baseados nos dados de resistividade coletados nos 3 wafers da amostra desejamos obter conclusões sobre a resistividade de todos os wafers do lote.

Para que a análise seja efetiva, uma análise de quão bem uma amostra representa uma população é necessária.

No presente caso: Se o lote contém wafers defeituosos, quão bem a amostra detectará o problema ?

Necessitamos quantificar o critério “quão bem”.

Necessitamos quantificar os riscos das decisões baseadas em amostras.

Como amostras devem ser selecionadas de modo a produzirem decisões com riscos aceitáveis ?

Modelos de probabilidade : ajudam a quantificar os riscos envolvidos na inferência estatística, ou seja, riscos envolvidos em feitas todos os dias.

Suponhamos que o lote contenha 25 wafers

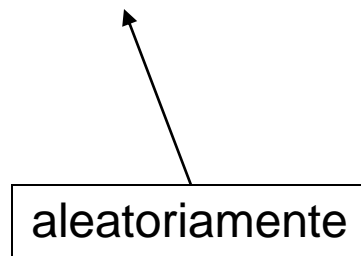
Suponhamos que somente um wafer no lote é defeituoso. Então a amostra poderia detectar ou não este wafer defeituoso.

Um modelo de probabilidade, junto com um método para selecionar a amostra, pode ser utilizado para quantificar os riscos.

Com base nesta análise o tamanho da amostra pode ser aumentado ou diminuído.

Interpretação do risco: suponhamos que uma série de lotes, cada um podendo conter um wafer defeituoso é examinado através de amostragem.

aleatoriamente



A proporção de lotes nos quais um wafer defeituoso está incluído na amostra é a probabilidade de que o wafer defeituoso seja detectado.

Mais especificamente: o limite desta proporção quando o número de lotes tende a infinito é a probabilidade de que o wafer defeituoso seja detectado.

Um **modelo de probabilidade** é usado para calcular esta proporção (pois não queremos medir infinitos lotes).

Termos e conceitos

Analytic study
Designed experiment
Empirical model
Engineering method
Enumerative study
Mechanistic model
Observational study
Overcontrol
Population
Probability model
Problem-solving method
Retrospective study
Sample
Statistical inference
Statistical Process
Control
Statistical thinking
Tampering
Variability

2. Probabilidade

- 2-1 Espaços amostrais e eventos
 - 2-1.1 Experimentos randômicos
 - 2-1.2 Espaços amostrais
 - 2-1.3 Eventos
- 2-2 Interpretações de probabilidade
 - 2-2.1 Introdução
 - 2-2.2 Axiomas de probabilidade
- 2-3 Regras de adição
- 2-4 Probabilidade condicional
- 2-5 Regras de probabilidade total e multiplicação
 - 2-5.1 Regra de multiplicação
 - 2-5.2 Regra da probabilidade total
- 2-6 Independência
- 2-7 Teorema de Bayes
- 2-8 Variáveis randômicas

2-1 Espaços amostrais e eventos

2-1.1 Experimentos randômicos

Exemplo: Medida da corrente em um fio de cobre

Repetições da medida resultam em variabilidade de resultados devido à influência **variáveis não controláveis**

Experimento possui uma componente randômica ou aleatória.

- Mudanças de temperatura ambiente
- Impurezas na composição química do fio
- Correntes de fuga

Em alguns casos essas variações randômicas não podem ser ignoradas (dependendo de nossos objetivos).

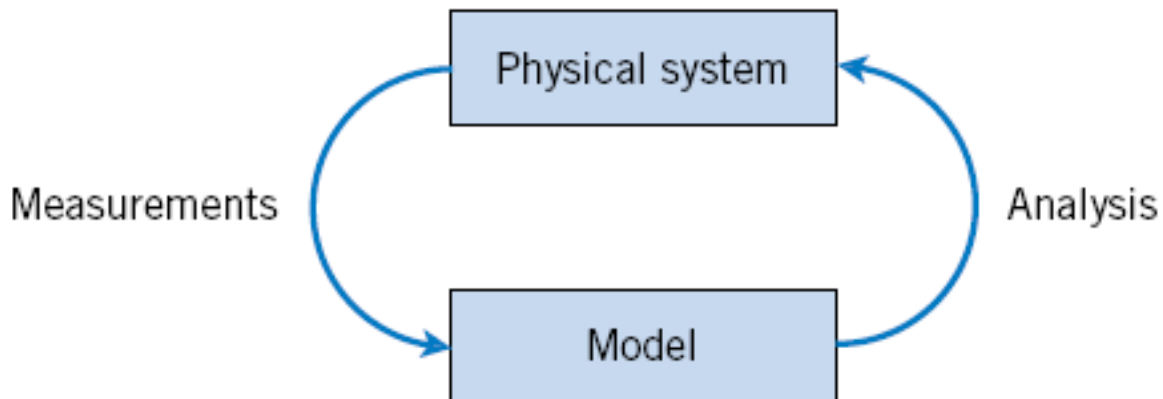
Objetivos: quantificar e modelar os tipos de variações frequentemente encontrados.

Incorporação da variação no nosso pensamento e análises

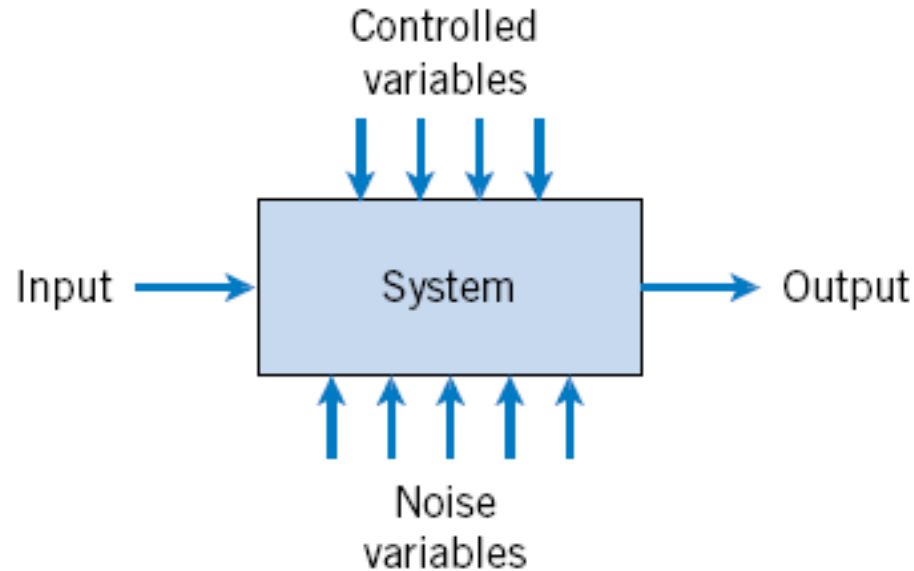


Resultados que não são invalidados pela variação.

Modelos e análises:



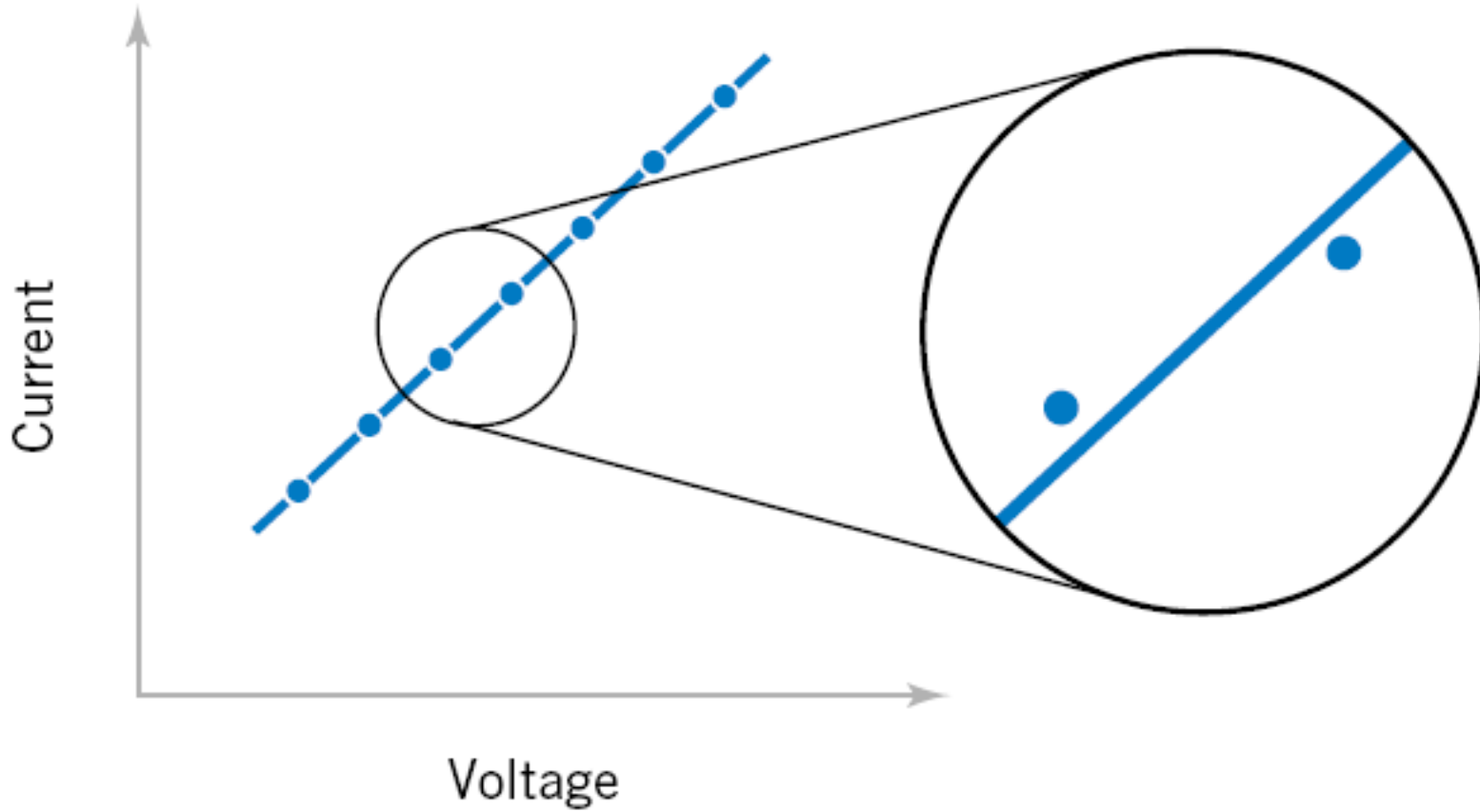
Incorporação de ruído (entradas não controláveis) no modelo



Um experimento que pode resultar em diferentes resultados, mesmo quando ele é repetido do mesmo modo cada vez, é chamado um **experimento aleatório ou randômico**.

Exemplo: Lei de Ohm

$$I = \frac{V}{R} + \varepsilon$$



Projeto de um sistema de comunicações

A capacidade de informação disponível a indivíduos que usam a rede é uma característica importante no design.

Comunicação de voz : linhas externas suficientes devem ser compradas da companhia telefônica.

Supondo que cada linha suporta somente uma conversação, quantas linhas devem ser compradas ?

Se poucas linhas forem compradas, chamadas podem ser postergadas ou perdidas. Se um excessivo número de linhas forem compradas o custo será muito alto.

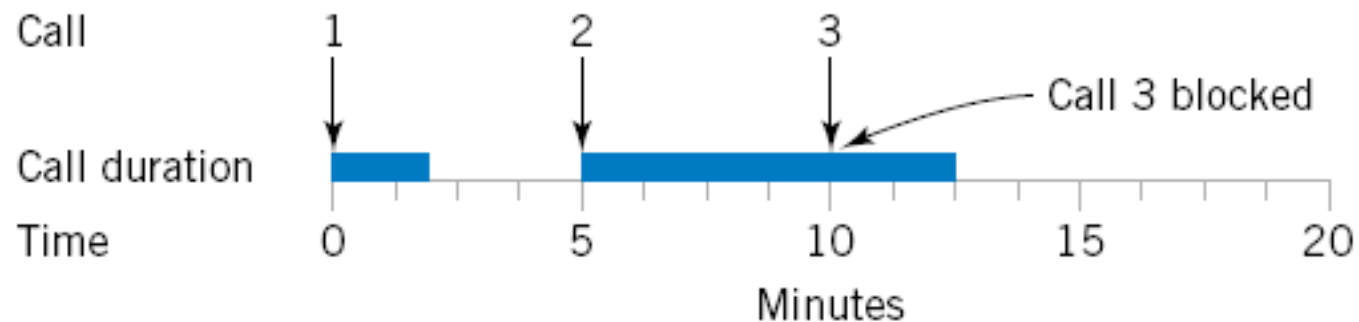
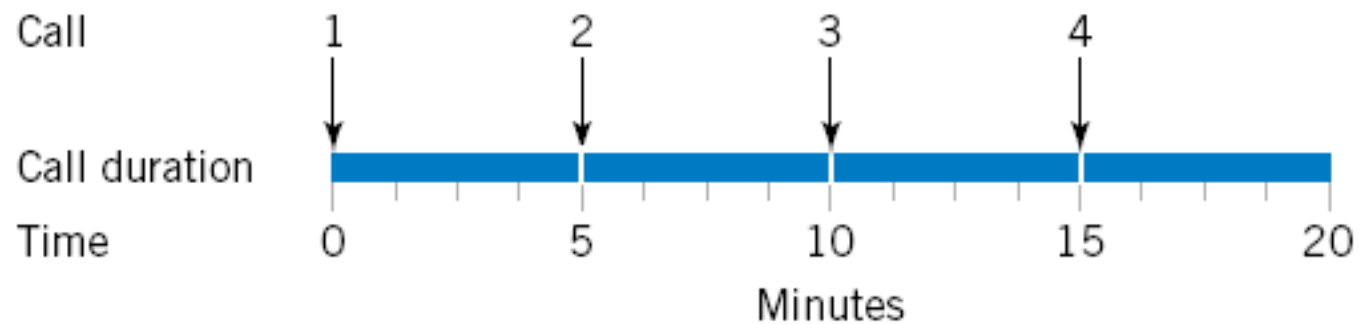
Design e desenvolvimento são necessários para satisfazer as demandas dos clientes **a um custo competitivo.**

Um modelo é necessário para o número de chamadas e para a duração das chamadas.

Mesmo sabendo que na média as chamadas ocorrem a cada 5 min e que elas duram até 5 min não é suficiente.

Se as chamadas chegarem em precisamente intervalos de 5 minutos e durarem precisamente 5 min, uma linha telefônica seria, a princípio, suficiente.

No entanto, uma pequena variação no número de chamadas ou duração resultaria em chamadas bloqueadas.



Uma análise de modelos envolvendo variações é fundamental.

2-1.2 Espaços Amostrais

Para modelar e analisar um experimento randômico devemos conhecer o conjunto dos possíveis resultados do experimento.

Definição: O conjunto de todos os possíveis resultados de um experimento randômico é chamado **espaço amostral do experimento** (S).

Um espaço amostral é definido com base nos objetivos da análise.

Exemplo:

Problema: selecionar um conector plástico moldado,
E medir sua espessura.

Os possíveis valores da espessura dependem da resolução do instrumento de medida e também dos limites superior e inferior da espessura.

Espaço de amostragem: reais positivos (valores negativos de espessura não podem ocorrer).

$$S = R^+ = \{x \mid x > 0\}$$

Se sabemos que todos os conectores têm espessura entre 10 e 11 mm o espaço amostral pode ser definido como

$$S = \{x \mid 10 < x < 11\}$$

Se o objetivo da análise é saber somente se uma parte particular tem espessura baixa, média ou alta, o espaço amostral será

$$S = \{baixa, media, alta\}$$

Se o objetivo é saber se a parte está ou não em conformidade com as especificações de fabricação teremos simplesmente

$$S = \{sim, nao\}$$

Espaço de amostragem discreto: conjunto finito ou infinito contável de possibilidades.

Espaço de amostragem contínuo : possibilidades são representadas por um intervalo, finito ou infinito, dos números reais.

Exemplo 2-2 Dois conectores são selecionados e sua espessura é medida. O espaço de amostragem agora é

$$S = R^+ \times R^+$$

Se o objetivo da análise é considerar somente se as duas partes estão ou não de acordo com as especificações de fabricação. Então

$$S = \{yy, yn, ny, nn\}$$

Se estamos interessados no número de conectores em conformidade na amostra, então

$$S = \{0, 1, 2\}$$

Consideremos um experimento em que a espessura é medida até que um conector não satisfaça as especificações. Então

$$S = \{n, yn, yyn, yyyn, \dots\}$$

Em experimentos randômicos onde os itens são selecionados de um lote, indicaremos se um item selecionado é ou não substituído antes da próxima seleção.

Por exemplo, se o lote consiste de 3 itens {a,b,c} e nosso experimento consiste em selecionar dois itens, **sem substituição**, o espaço de amostragem será representado por

$$S_{SS} = \{ab, ba, bc, cb, ca, ac\}$$

↑
Ordem é mantida

Se o ordenamento não for necessário,

$$S_{SS} = \{\{a, b\}, \{b, c\}, \{a, c\}\}$$

Se os itens selecionados são substituídos antes do próximo, o espaço de amostragem é dado por

$$S_{SS} = \{aa, bb, cc, ab, ba, bc, cb, ca, ac\}$$

Se o ordenamento não for necessário,

$$S_{SS} = \{\{a, a\}, \{b, b\}, \{c, c\}, \{a, b\}, \{b, c\}, \{c, a\}\}$$

Algumas vezes não é necessário especificar o exato item selecionado, mas somente uma propriedade do item .

Exemplo: Suponha que há 5 partes defeituosas e 95 partes boas em um lote. Para estudar a qualidade do lote, dois itens são selecionados, sem substituição.

Notação: b : parte boa
 d : parte defeituosa

É possível que seja suficiente descrever o espaço de amostragem (ordenado) em termos da qualidade de cada parte selecionada como

$$S = \{bb, dd, db, bd\}$$

Nota: esta descrição do espaço de amostragem não mostra que há muito mais pares bb que dd , por exemplo. Isso será levado em conta quando computarmos as probabilidades mais tarde.

Se houver somente uma parte defeituosa no lote, o espaço de amostragem seria

$$S = \{bb, db, bd\}$$

Diagramas de árvores

Exemplo 2-3 Cada mensagem em um sistema de comunicação digital é classificada de acordo com fato dela ser ou não recebida dentro de um tempo especificado no design do sistema.

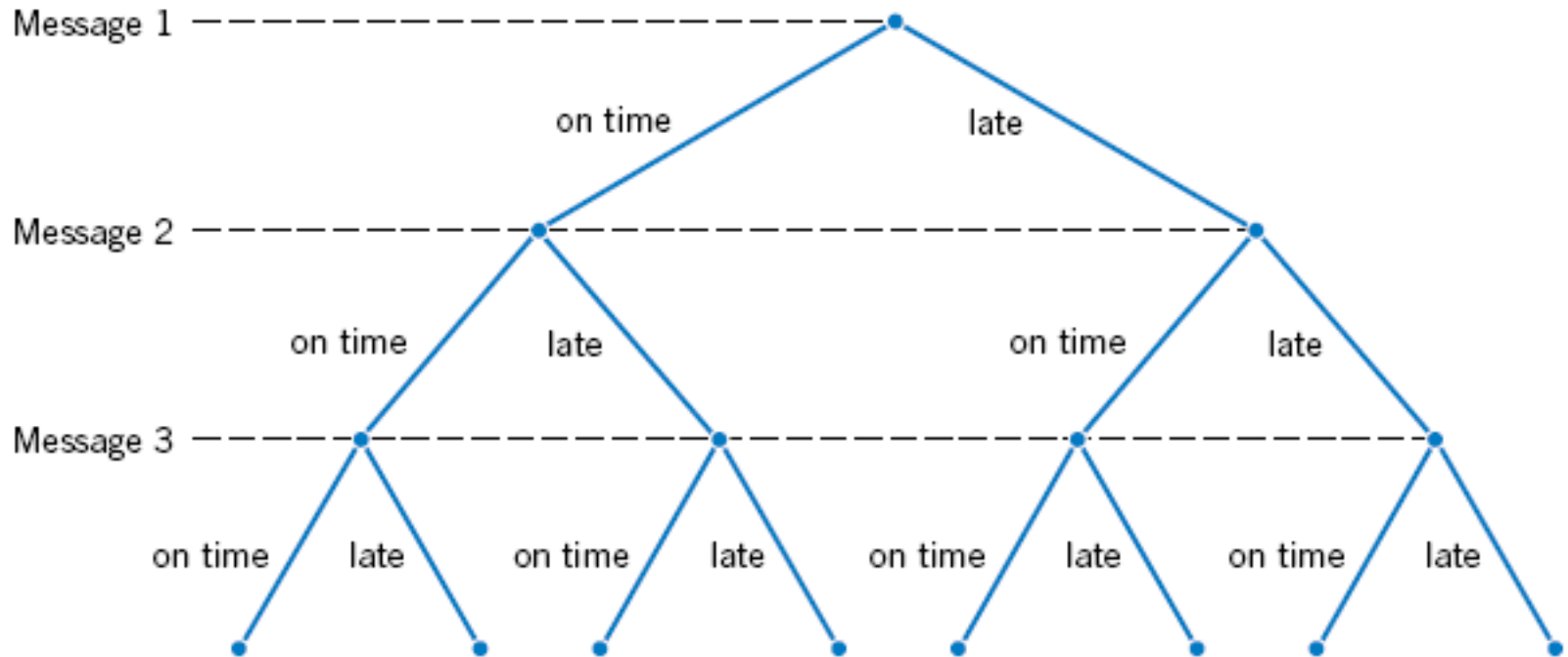
Espaço de amostragem:

$$S = \{lll, ltl, ltt, ttt, ltt, llt, tll, tlt\}$$

l - late

t - on time

Árvore do espaço de amostragem



Exemplo 2.4 Um fabricante de automóveis fornece veículos com opções selecionadas. Cada veículo pode ter as seguintes opções:

Com ou sem transmissão automática

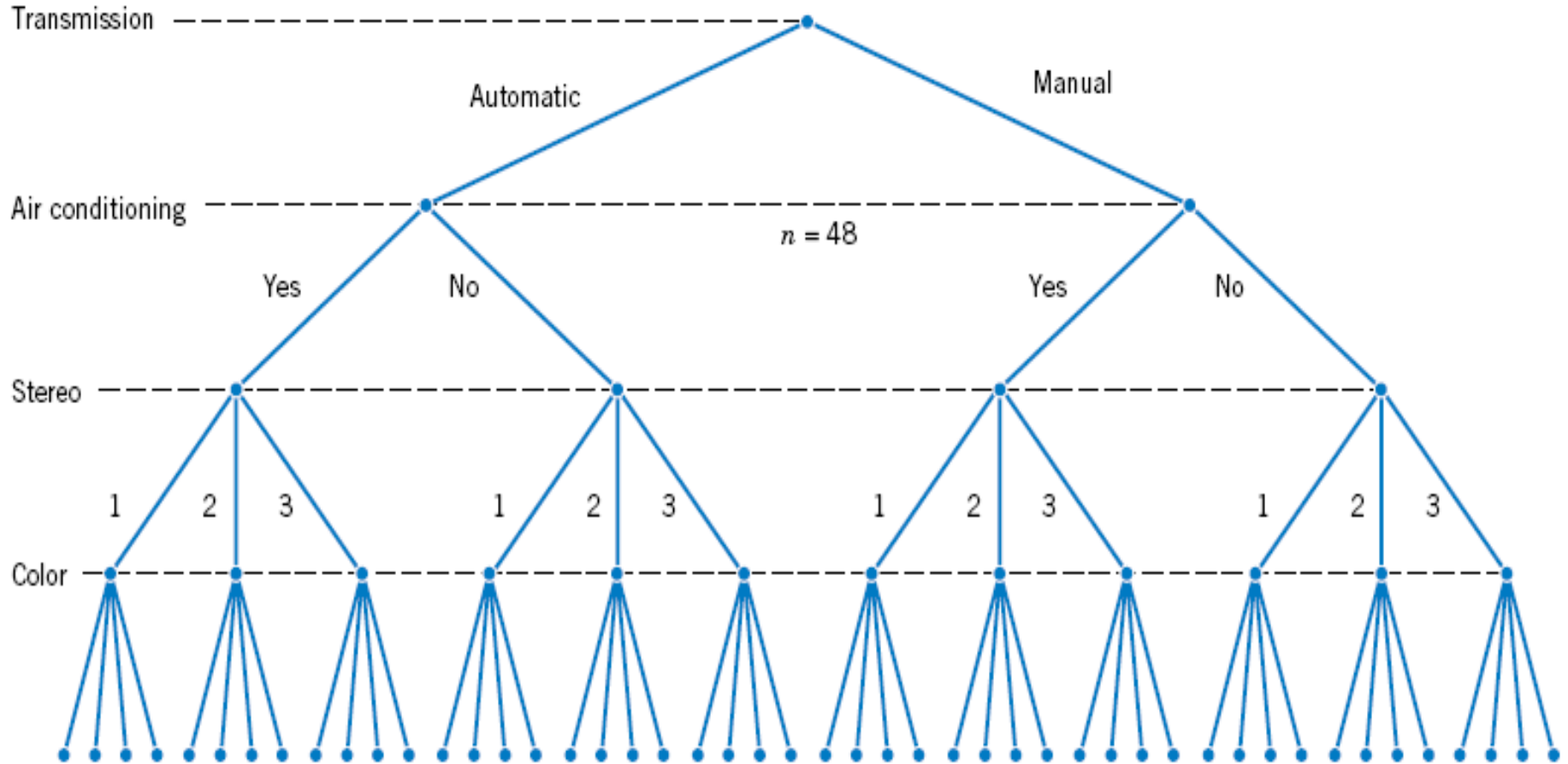
Com ou sem ar-condicionado

Uma de três opções de equipamento de som

Uma de quatro opções de cor exterior

$2 \times 2 \times 3 \times 4 = 48$ elementos no espaço de amostragem

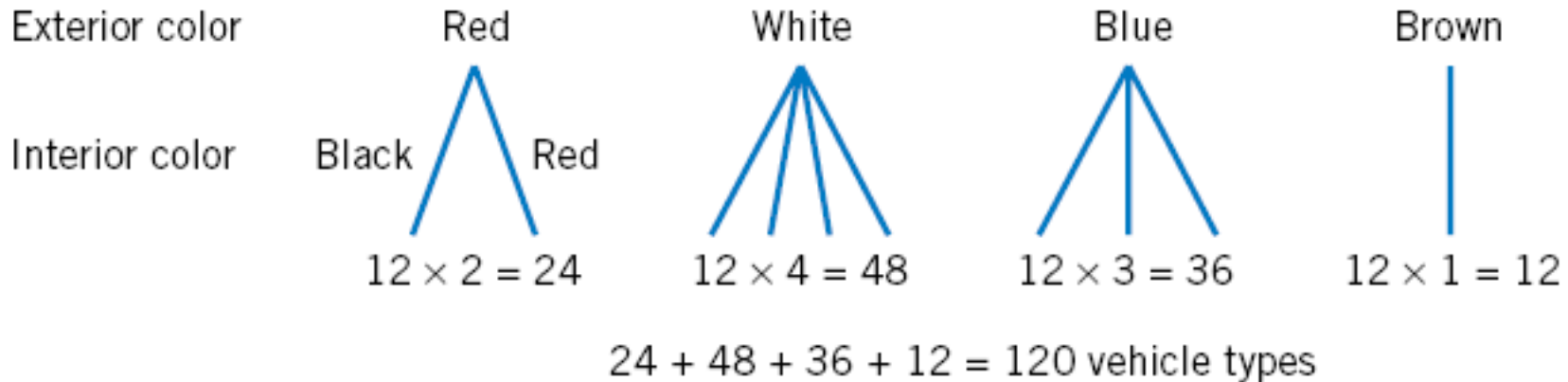
Diagrama de árvore do espaço de amostragem



Exemplo 2.5 : Consideremos uma extensão do exemplo anterior com a adição de mais 4 possíveis opções para cores interiores: azul, vermelho, preto e marron, mas com as seguintes condições:

1. Com vermelho exterior, somente vermelho ou preto interior
2. Com branco exterior, qualquer cor interior
3. Com azul exterior, somente azul, preto ou vermelho interior
4. Com marron exterior, somente marron interior

Notemos que cada cor tem $48/4 = 12$ opções. Temos então $12 \times 2 + 12 \times 4 + 12 \times 3 + 12 \times 1 = 120$ opções.



2-1.3 Eventos

Frequentemente estamos interessados em uma coleção de resultados de um experimento randômico.

Definição. Um evento é um subconjunto do espaço de amostragem de um experimento randômico.

Podemos estar interessados na descrição de novos eventos formados a partir de combinações de eventos existentes.



União, intersecção e complemento de eventos para formar novos eventos.

Exemplo 2.6 Consideremos o espaço de amostragem do Exemplo 2.2

$$S = \{yy, yn, ny, nn\}$$

Suponhamos que o conjunto de todas as possibilidades nas quais uma parte está conforme é denotada por E_1 . Então

$$E_1 = \{yy, yn, ny\}$$

O evento em que ambas partes não estão conformes é

$$E_2 = \{nn\}$$

O evento em que ao menos uma parte não está conforme é

$$E_3 = \{yn, ny, nn\}$$

$$E_1 \cup E_2 = S$$

$$E_1 \cup E_3 = S$$

$$E_1 \cap E_3 = \{yn, ny\}$$

$$E_1' = \{nn\}$$

Exemplo 2.7 Medidas do tempo necessário para completar uma reação química poderiam ser modeladas com o espaço de amostragem

$$S = R^+$$

Sejam os eventos

$$E_1 = \{1 \leq x < 10\}$$

$$E_2 = \{3 \leq x < 118\}$$

Então

$$E_1 \cup E_2 = \{1 \leq x < 118\}$$

$$E_1 \cap E_2 = \{3 \leq x < 10\}$$

$$E_1' = \{x < 10\}$$

Exemplo 2.8 Amostras de plástico policarbonato são analisadas quanto à resistência ao choque e arranhões. Os resultados de 50 amostras são mostrados abaixo:

		<u>shock resistance</u>	
		high	low
scratch resistance	high	40	4
	low	1	5

$$A \cap B = 40$$

$$A \cup B = 45$$

$$A' = 9$$

Evento A: a amostra possui alta resistência a choques

Evento B: a amostra possui alta resistência a arranhões

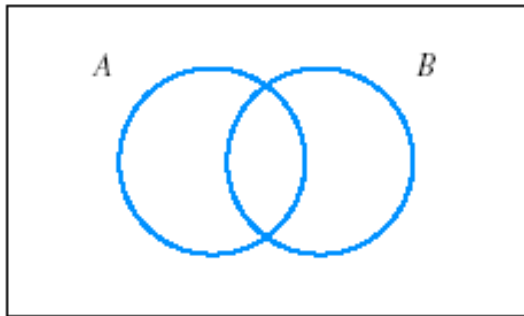
Determine o número de amostras em

$$A \cap B$$

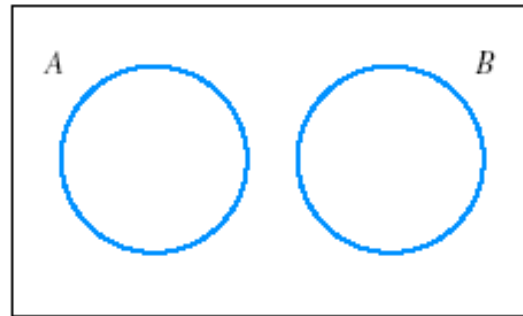
$$A \cup B$$

$$A'$$

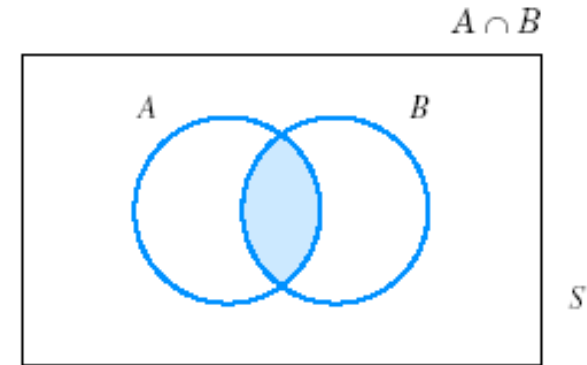
Diagramas de Venn



(a)

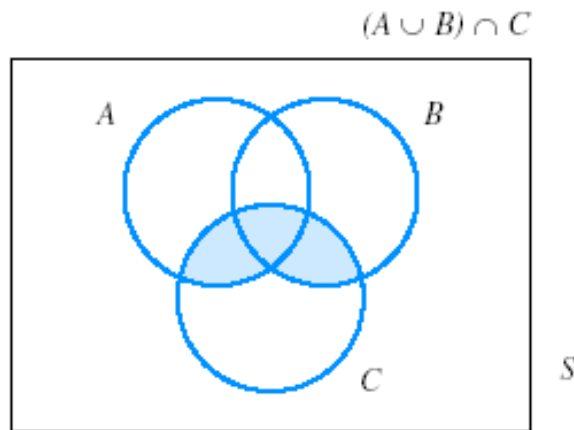


(b)

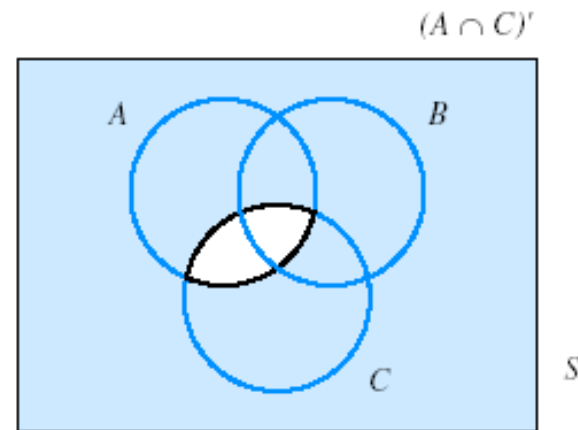


(c)

Sample space S with events A and B



(d)



(e)

Eventos mutuamente exclusivos:

$$E_1 \cap E_2 = \emptyset$$

Propriedades

$$(E')' = E$$

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

$$(A \cap B)' = A' \cup B' \quad (A \cup B)' = A' \cap B'$$

Exercício: Prove estas relações utilizando diagramas de Venn.

2-2 Interpretações de Probabilidade

Consideraremos espaços de amostragem discretos

Probabilidade

Quantificação da chance de um determinado resultado de um experimento randômico ocorrer.

Interpretação objetiva da probabilidade: **Freqüência relativa**

Exemplo: Suponhamos que vamos selecionar um diodo randomicamente de um lote de 100 elementos.

Espaço de amostragem: conjunto de 100 diodos.

Chance de qualquer diodo do lote ser retirado é a mesma.

Como a soma das probabilidades é igual a 1, o modelo de probabilidade para este experimento atribui uma probabilidade 0.01 a cada um dos 100 possíveis resultados.

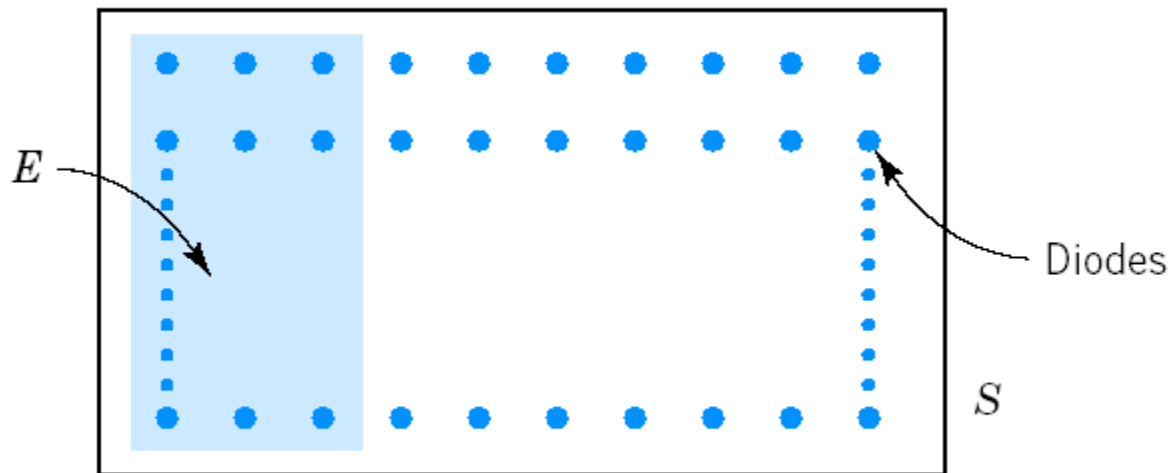
Interpretação da probabilidade: experimento é replicado muitas vezes.

Sempre que o espaço de amostragem consiste de N possíveis resultados que são igualmente prováveis, a probabilidade de cada resultado é $1/N$.

Freqüentemente é necessário atribuir probabilidades a eventos que são compostos por vários elementos do espaço de amostragem.

Exemplo 2.9 Suponhamos que 30% dos diodos em um lote de 100 peças satisfazem as exigências mínima de potência de um cliente específico. Se um diodo é selecionado aleatoriamente, intuitivamente acreditamos que a probabilidade de que as exigências do cliente sejam satisfeitas é 0.3.

Seja E o subconjunto (evento) consistindo de 30 diodos que satisfazem as exigências dos clientes . Como E contém 30 elementos e cada elemento tem probabilidade 0.01, concluímos que a probabilidade de E é 0.3.



$$P(E) = 30(0.01) = 0.30$$

Definição. Para um espaço de amostragem discreto, a probabilidade de um evento E , denotada por $P(E)$, é igual à soma das probabilidades dos elemento em E .

Exemplo 2.10 Um experimento randômico pode resultar em um dos resultados $\{a,b,c,d\}$ com probabilidades 0.1, 0.3, 0.5 e 0.1, respectivamente. Sejam os eventos $A=\{a,b\}$, $B=\{b,c,d\}$, e $C=\{d\}$.

Então

$$P(A) = 0.1 + 0.3 = 0.4$$

$$P(B) = 0.3 + 0.5 + 0.1 = 0.9$$

$$P(C) = 0.1$$

$$P(A') = 0.6$$

$$P(B') = 0.1$$

$$P(A \cup B) = 1$$

$$P(A \cap B) =$$

Exemplo 2.11 Uma inspeção visual de um pedaço de material de wafers de um processo de manufatura de semicondutores resultou na seguinte tabela:

Number of Contamination Particles	Proportion of Wafers
0	0.40
1	0.20
2	0.15
3	0.10
4	0.05
5 or more	0.10

Se um wafer é selecionado aleatoriamente deste processo e examinado, qual é a probabilidade de que ele não contenha partículas de contaminação ?

Se a informação fosse disponível para cada wafer, poderíamos definir o espaço de amostragem como o conjunto de todos os wafers inspecionados e proceder como no exemplo dos diodos. No entanto, tal nível de detalhe não é necessário aqui.

Podemos considerar o espaço de amostragem como consistindo de seis categorias que sumarizam o número de partículas contaminantes em um wafer.

Seja $E = \{0\}$. Então $P(E) = 0.4$.

Qual é a probabilidade de que numa inspeção de um wafer sejam encontradas 3 ou mais partículas contaminantes ?

Seja $E = \{3, 4, 5 \text{ ou mais partículas}\}$. Então

$$P(E) = 0.1 + 0.05 + 0.1 = 0.25$$

Exemplo 2.12 Suponhamos que um lote contém 6 elementos $\{a,b,c,d,e,f\}$ e que duas partes são selecionadas sem substituição.

Seja E o evento onde o primeiro elemento selecionado é a . Então

$$E = \{ab,ac,ad,ae,af\}$$

Notemos que o espaço de amostragem possui $(5)(6) = 30$ elementos, cada um representando um resultado de igual probabilidade. Portanto,

$$P(E) = 5 / 30 = 1 / 6$$

Se E é o evento em que o segundo elemento selecionado é a , ou seja,

$$E = \{ba,ca, da,ea,fa\}$$

Temos novamente

$$P(E) = 5 / 30 = 1 / 6$$

2-2.2 Axiomas de Probabilidade

Probabilidade é um número, atribuído a cada membro de uma coleção de eventos de um experimento aleatório, satisfazendo as seguintes propriedades:

Se S é o espaço de amostragem e E é qualquer evento em um experimento randômico,

$$(i) P(S) = 1$$

$$(ii) 0 \leq P(E) \leq 1$$

(iii) Para dois eventos E_1 e E_2 com $E_1 \cap E_2 = \emptyset$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

2-3 Regras de adição

Exemplo 2.13 Considere 940 wafers em um processo de manufatura de semicondutores

Experimento: um wafer é selecionado aleatoriamente

H – evento de que um wafer tenha alto índice de contaminação

$$P(H) = 358 / 940$$

Table 2-1 Wafers in Semiconductor Manufacturing Classified by Contamination and Location

Location in Sputtering Tool			
Contamination	Center	Edge	Total
Low	514	68	582
High	112	246	358
Total	626	314	

C – evento correspondente à localização do wafer no centro da ferramenta de sputtering. Então

$$P(C) = 626 / 940$$

Além disso,

$$P(H \cap C) = 112 / 940$$

$$P(H \cup C) = 872 / 940$$

$$P(H \cup C) = P(H) + P(C) - P(H \cap C)$$

$$= 358 / 940 + 626 / 940 - 112 / 940 = 872 / 940$$

Sputtering – processo pelo qual átomos são ejetados de um alvo material sólido pelo bombardeamento por íons energéticos. Usado comumente para deposição de filmes finos.

Lei de adição

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Se A e B são mutuamente exclusivos,

$$P(A \cup B) = P(A) + P(B)$$

Exemplo 2.14 Os wafers dos Exemplo 2.13 são agora classificados pelo grau de contaminação.

Number of Contamination Particles	Center	Edge	Totals
0	0.30	0.10	0.40
1	0.15	0.05	0.20
2	0.10	0.05	0.15
3	0.06	0.04	0.10
4	0.04	0.01	0.05
5 or more	0.07	0.03	0.10
Totals	0.72	0.28	1.00

Eventos

E_1 Wafer contém 4 ou mais partículas

E_2 Wafer está na borda

$$P(E_1) = 0.15 \quad P(E_2) = 0.28$$

$$P(E_1 \cap E_2) = 0.04$$

$$\begin{aligned} P(E_1 \cup E_2) &= P(E_1) + P(E_2) - P(E_1 \cap E_2) \\ &= 0.15 + 0.28 - 0.04 = 0.39 \end{aligned}$$

Eventos

E_3 Wafer contém menos de 3 partículas

E_4 Wafer está na borda e contém mais de 4 partículas

$$P(E_3) = 0.75$$

$$P(E_4) = 0.03$$

$$E_3 \cap E_4 = \emptyset$$

$$P(E_3 \cap E_4) = 0$$

$$P(E_3 \cup E_4) = P(E_3) + P(E_4) = 0.78$$

Três ou mais eventos

$$P(A \cup B \cup C) = P[(A \cup B) \cup C] = P(A \cup B) + P(C) - P[(A \cup B) \cap C]$$

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) - P(A \cap B) + P(C) - P[(A \cap C) \cup (B \cap C)] \\ &= P(A) + P(B) - P(A \cap B) + P(C) \\ &\quad - [P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)] \\ &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) \\ &\quad - P(B \cap C) + P(A \cap B \cap C) \end{aligned}$$