

Ferramentas para o Controle de Qualidade

Denise A. Botter

08/08/2013

Exemplo: Em uma linha de empacotamento de leite, se medirmos o volume de cada saquinho, vamos descobrir que nenhum deles contém exatamente a mesma quantidade de leite.

Do ponto de vista comercial, tanto faz se o saquinho contém um pouco a mais ou um pouco a menos de leite do que o especificado na embalagem.

No entanto, se o **pouco a mais** de leite não for assim tão pouco, o saquinho poderá estourar facilmente; se, o **pouco a menos**, não for assim tão pouco, o produtor poderá ser multado. A **especificação** determina que o volume de cada saquinho seja igual a 1000 ml de leite.

Na prática, contudo, espera-se que a média dos volumes dos saquinhos fique em torno de 1000 ml, com pequena variabilidade.

O valor especificado de 1000 ml é o **valor-alvo** da **variável aleatória X**, quantidade de leite em cada saquinho. A Tabela 2.1 apresenta a quantidade de leite de 100 saquinhos de um mesmo lote.

A média dos volumes dos 100 saquinhos está bem próxima do valor-alvo ($\bar{X} = 999,8$), e a variabilidade do processo é pequena ($S = 4,32$).

Tabela 2.1: Valores da Variável X

998,8	994,9	1001,0	1005,1	1004,8	1006,9	991,3	999,1	1004,4	995,7
997,2	993,2	992,6	996,1	996,9	991,5	997,7	998,4	1000,5	998,5
998,7	998,5	1005,4	999,7	999,3	997,9	1007,9	1003,5	1009,5	997,4
1006,6	993,6	1002,2	1003,6	1007,7	999,7	997,9	1002,7	998,5	1003,0
994,2	996,6	993,9	998,5	999,9	1000,1	998,7	1008,8	993,0	997,1
989,7	1005,8	994,9	997,4	1003,0	1001,9	1003,5	1002,4	994,5	995,5
1002,8	1001,3	996,2	999,0	1000,5	1002,2	1000,6	996,4	1007,5	1001,9
1000,3	1003,3	1003,4	997,5	996,3	1004,4	995,2	993,8	1002,8	1002,6
1008,8	1005,8	1005,2	1000,5	1000,0	1001,8	999,9	995,8	992,9	1003,3
1001,8	1002,5	1000,9	995,9	1005,0	998,8	996,6	996,7	998,3	998,2

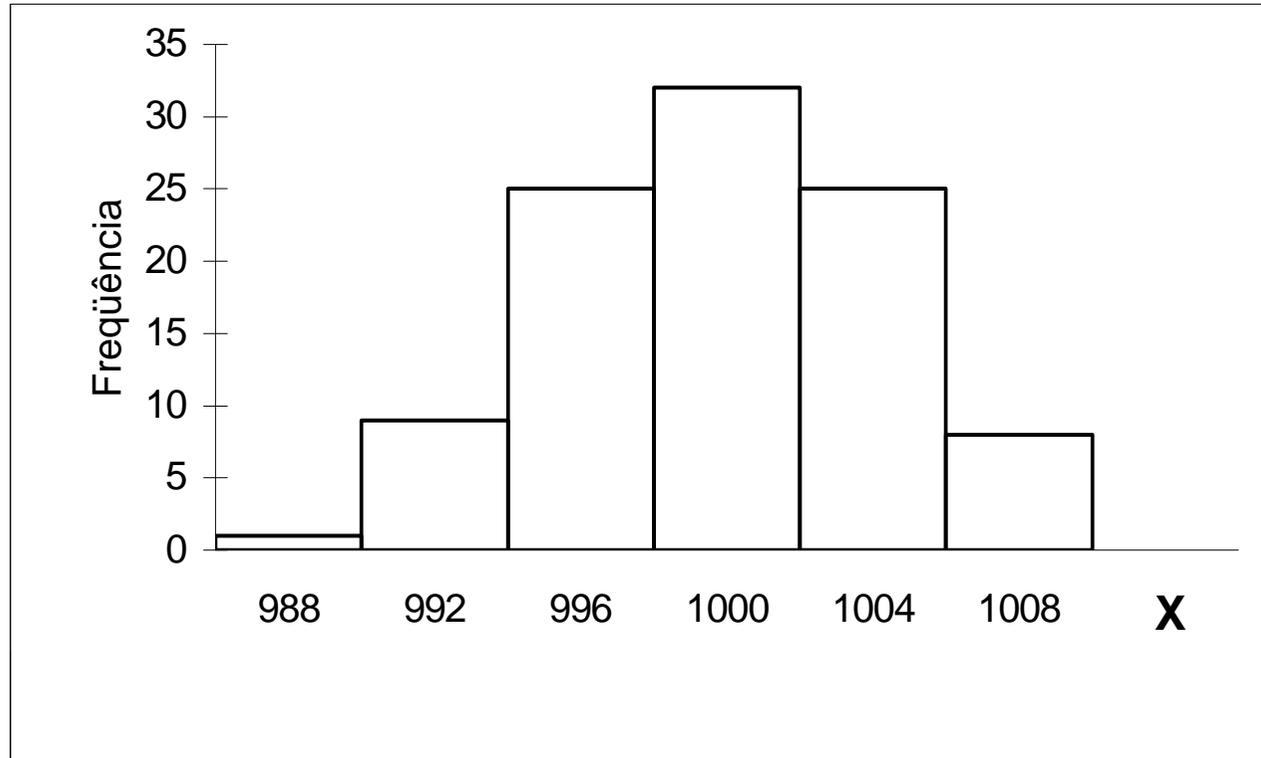


Figura 2.1: Histograma dos Valores de X da Tabela 2.1

Conforme a Figura 2.1, a distribuição de X parece ser bem representada pela distribuição normal.

2.1. Variações aleatórias e variações identificáveis de um processo

A expressão **variabilidade de um processo** tem a ver com as diferenças existentes entre as unidades produzidas. Se a variabilidade for grande, é fácil observar as diferenças entre as unidades produzidas; se, por outro lado, a variabilidade for pequena, tais diferenças serão difíceis de observar.

Em qualquer processo de produção sempre haverá uma certa quantidade de **variabilidade inerente ou natural**. Essa variabilidade natural ou "ruído de fundo" é o efeito cumulativo de muitas causas pequenas, essencialmente inevitáveis. No sistema do controle estatístico da qualidade, essa variabilidade natural é, em geral, chamada de **variações ou causas aleatórias**. Um processo que opera apenas com variações aleatórias está **sob controle estatístico**.

No nosso exemplo, variações de temperatura, de densidade do leite e a precisão do mecanismo que corta o fluxo do leite para o saquinho são algumas das causas aleatórias de variabilidade do processo.

A Figura 2.2 ilustra um processo sujeito apenas a causas aleatórias: a característica X apresenta variabilidade que pode ser representada por uma distribuição de probabilidades (usualmente normal ou bem aproximada por uma distribuição normal) que se mantém estável (isto é, com média e dispersão inalteradas) ao longo do tempo.

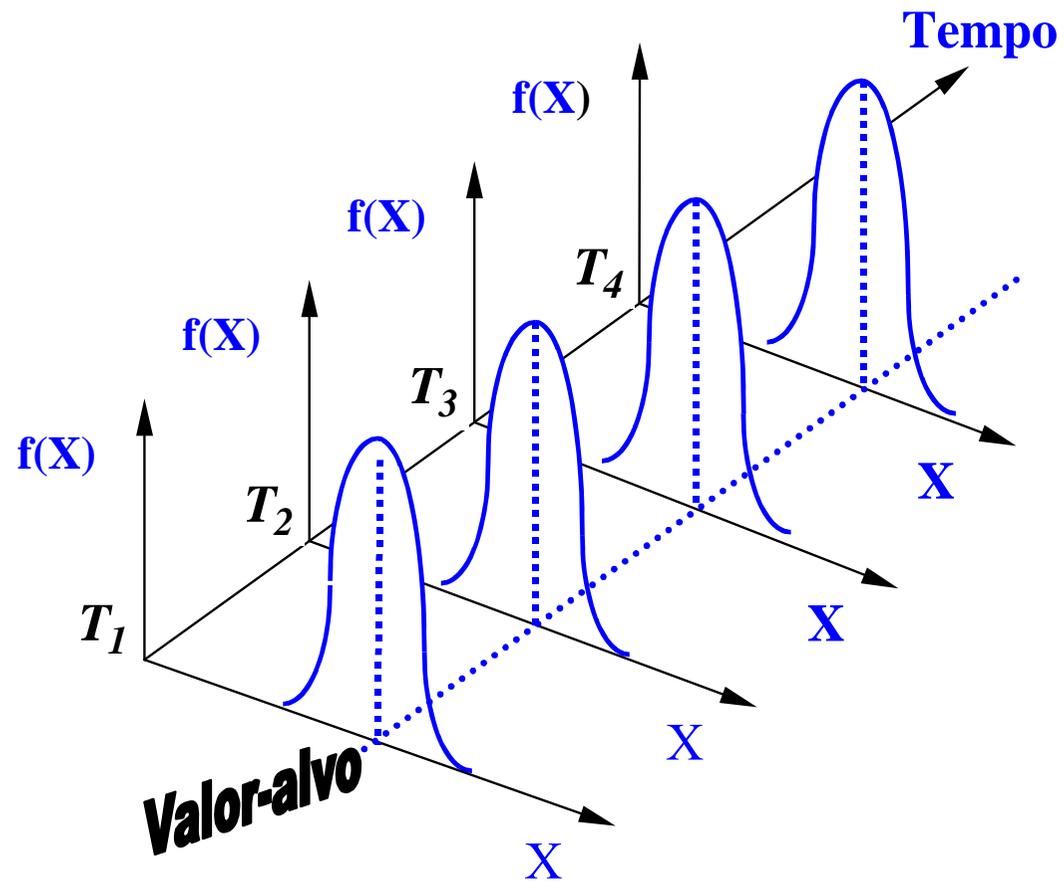


Figura 2.2: Processo Isento de Causas Especiais

Outros tipos de variabilidade podem estar presentes na saída de um processo. Esta variabilidade é, em geral, muito grande quando comparada com o ruído de fundo e representa, usualmente, um nível inaceitável de desempenho do processo. Essas fontes de variabilidade são as **causas identificáveis ou especiais**. Elas têm o efeito de deslocar a distribuição da variável aleatória X (tirando sua média do valor-alvo) e/ou aumentar sua dispersão. Uma causa especial é um problema ou modo de operação anormal do processo, que pode, portanto ser corrigido ou eliminado. Surge, em geral, de três fontes: máquinas ajustadas ou controladas de maneira inadequada, erros do operador ou de matéria-prima defeituosa.

Quando, além das causas aleatórias de variabilidade, causas especiais estiverem presentes, dizemos que o processo está operando **fora de controle**.

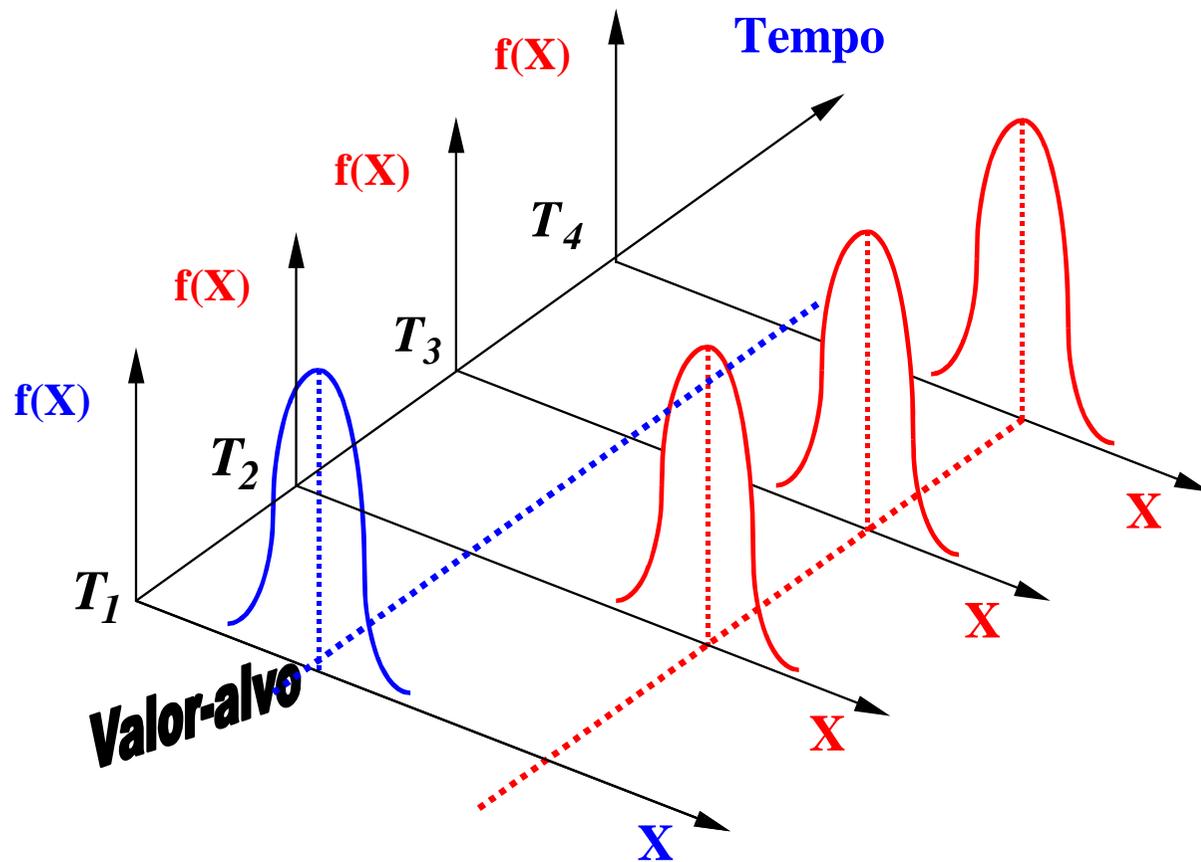


Figura 2.3: Causa Especial Altera a Média do Processo

A Figura 2.3 ilustra o caso em que a causa especial desloca a distribuição de X .

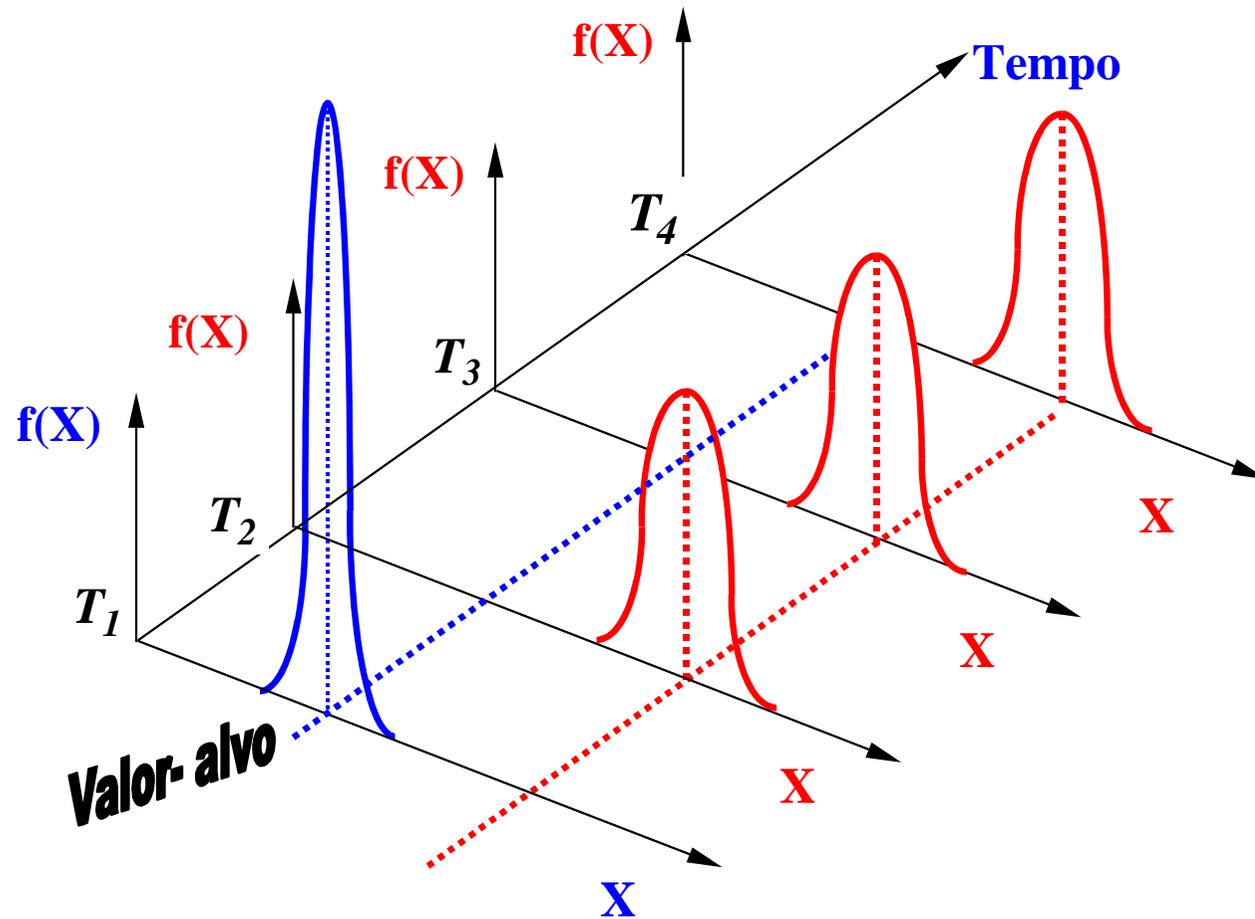


Figura 2.4: Causa Especial Altera a Média e Aumenta a Variabilidade do Processo

A Figura 2.4 ilustra o caso em que a causa especial não só desloca a distribuição de X , como também aumenta sua dispersão.

Embora as causas especiais possam sempre ser eliminadas, certos casos demandam correções significativas no processo. No exemplo dos saquinhos de leite, tomemos como causa especial uma alteração indesejada da pressão de operação nas tubulações do sistema de empacotamento de leite. A Tabela 2.2 apresenta a quantidade de leite de 100 saquinhos de um mesmo lote, extraído do processo após uma alteração da pressão de operação. A média dos volumes dos saquinhos já não está tão próxima do valor-alvo ($\bar{X} = 1004,9$) e a dispersão do processo aumentou nitidamente ($S = 8,55$). A Figura 2.5 apresenta o histograma dos valores de X da Tabela 2.2. Com base nesse histograma, já não podemos dizer que a distribuição de X pareça ser bem representada pela distribuição normal.

Tabela 2.2: Valores da Variável X- processo sob a influência de causas especiais

1010,2	1002,3	1003,8	1000,2	1008,8	992,1	1008,9	999,4	1011,3	1014,0
1010,5	995,0	994,0	1011,2	1008,1	1008,3	1017,6	1005,3	1003,8	1019,6
995,0	1010,2	999,9	1009,5	1017,9	1012,9	1008,5	1003,1	1010,5	1009,5
994,1	991,2	1001,6	1002,1	1010,5	1009,0	992,3	1002,3	1012,7	1006,9
994,8	989,1	1002,5	1008,7	1014,6	1004,9	1002,2	1007,3	1002,4	1011,7
980,2	999,4	1002,0	1011,9	997,8	997,5	986,6	1014,4	1024,0	1006,9
992,0	1004,4	1005,3	1003,2	1016,5	1015,3	1003,3	992,6	1013,1	1016,1
997,2	994,5	1006,9	1012,8	1014,5	1021,7	1007,2	996,1	1008,8	1000,2
1004,5	998,7	1002,4	1012,9	1011,1	1007,8	994,2	1012,0	1017,8	1018,4
988,2	991,1	1004,3	1010,6	1009,9	1011,3	989,9	1002,9	997,5	1002,0

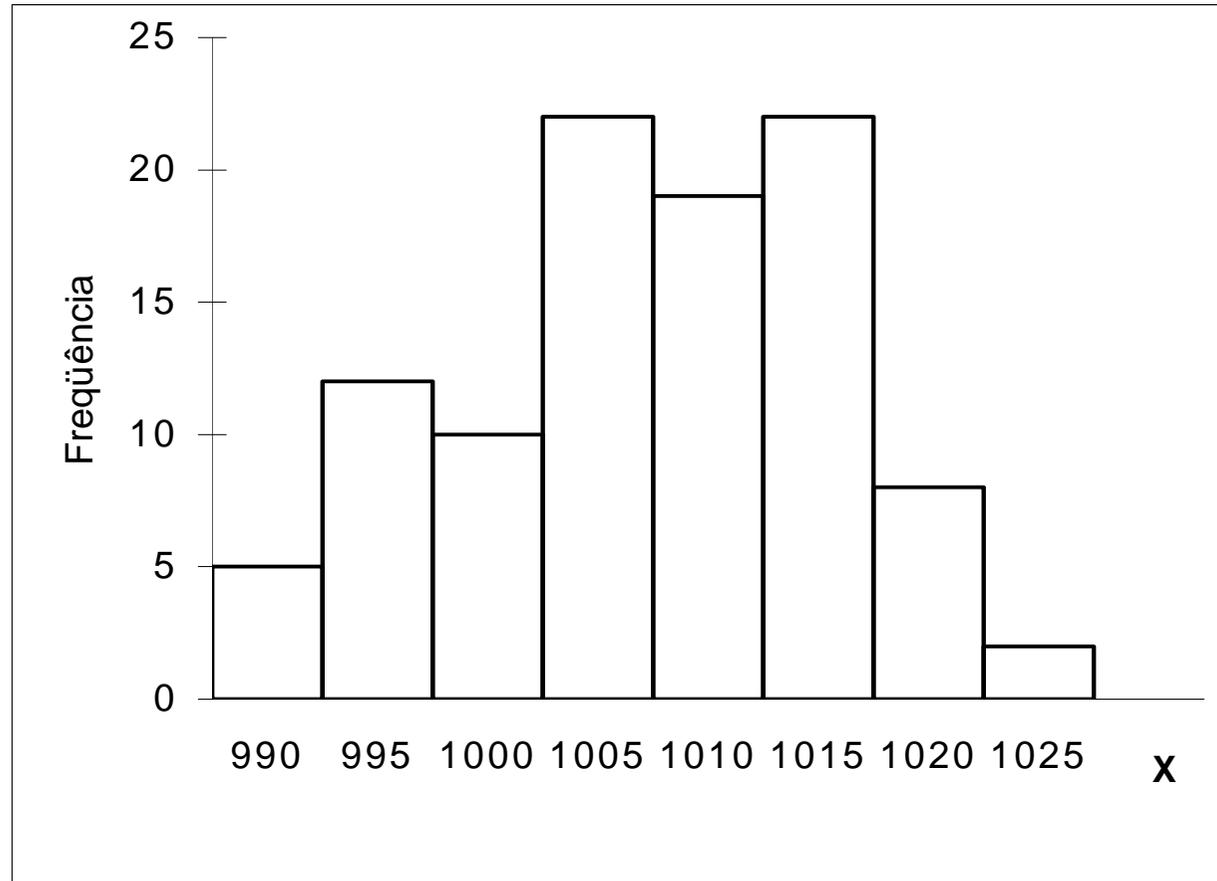


Figura 2.5: Histograma dos Valores de X da Tabela 2.2

2.2 A base estatística dos gráficos de controle

Os processos devem ser constantemente monitorados, para detectar a presença de causas especiais. Detectada essa presença, deve-se proceder a uma investigação para identificar a(s) causa(s) especial(is) e intervir para eliminá-la(s).

No exemplo do leite, o volume X do produto em cada saquinho deve ser sempre monitorado, para se evitar a ocorrência de excessos (que aumentam o risco de os saquinhos estourarem) ou de falta (que incorre em multa para a empresa).

A principal ferramenta utilizada para monitorar os processos e sinalizar a presença de causas especiais é o **gráfico de controle**.

Os gráficos de controle de \bar{X} e R , também conhecidos como gráficos da **média e da amplitude**, servem para monitorar processos cuja característica de qualidade de interesse X é uma grandeza mensurável: o diâmetro de um eixo, o volume de um saquinho de leite, o teor de carbono de uma liga metálica, a concentricidade de um cilindro etc.

O monitoramento é realizado através da análise periódica de amostras: a cada intervalo de tempo h retira-se uma amostra de n itens para análise. Por exemplo, a cada meia hora de produção ($h = 30$ min), selecionam-se, aleatoriamente, cinco saquinhos ($n = 5$), cujos volumes são medidos. Para cada amostra, é calculada a média amostral \bar{X} dos valores medidos e a amplitude amostral R (diferença entre o maior e o menor valores da amostra). Os valores de \bar{X} e de R das diversas amostras são marcados, respectivamente, nos gráficos da média e da amplitude.

A quantidade de leite dos saquinhos de 15 amostras (cada uma de tamanho $n = 5$), bem como a média \bar{X}_i e a amplitude R_i de cada amostra foram anotadas e estão apresentadas na Tabela 2.3.

Tabela 2.3: Valores de X_{ij} , \bar{X}_i e R_i

Amostra (i)	Elemento (j) da amostra (i)					\bar{X}_i	R_i
	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}		
1	1001,7	1004,0	1004,8	996,3	1004,3	1002,2	8,4
2	999,7	1000,3	1003,2	993,9	998,9	999,2	9,2
3	990,9	1004,0	1003,0	1004,0	1002,0	1000,8	13,1
4	1000,7	1007,3	998,1	995,5	994,9	999,3	12,4
5	1000,7	998,3	998,9	997,8	1001,9	999,5	4,1
6	998,6	993,7	1002,8	995,5	994,1	997,0	9,1
7	1002,7	1010,5	990,5	992,5	1003,0	999,8	19,9

Em seguida, os valores de \bar{X}_i e R_i foram plotados nos gráficos de controle da Figura 2.6. Esses gráficos possuem **limites de controle**: um **Limite Superior de Controle** (LSC) e um **Limite Inferior de Controle** (LIC), que mais adiante veremos como são determinados.

Enquanto os pontos nos gráficos vão distribuindo-se aleatoriamente em torno da **Linha Média** (LM), não se deve interferir no processo, ou seja, a variação exibida é decorrente de causas aleatórias. Contudo, se um ponto cai acima de LSC ou abaixo de LIC, como aconteceu com o 15o valor de \bar{X} (ver a Figura 2.6), há evidência de que o processo está fora de controle, sujeito à alguma causa especial (o afastamento excessivo desse ponto em relação à LM é improvável se o processo estiver sujeito apenas às causas aleatórias).

Para facilitar a visualização, costuma-se unir os pontos amostrais no gráfico de controle por meio de segmentos de reta.

Se os pontos apresentam um comportamento sistemático ou não-aleatório, mesmo que estejam entre os limites de controle, há indicação de que o processo esteja fora de controle.

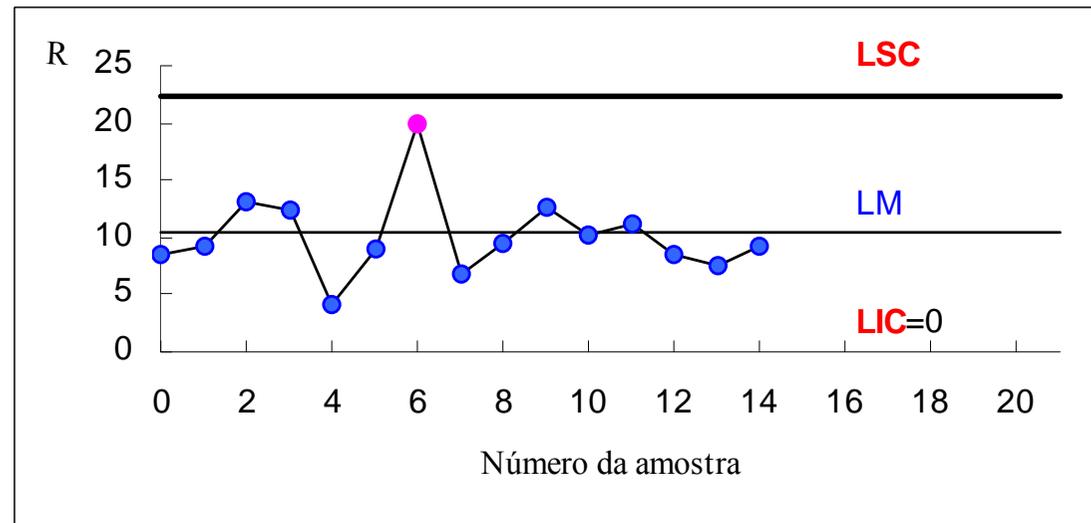
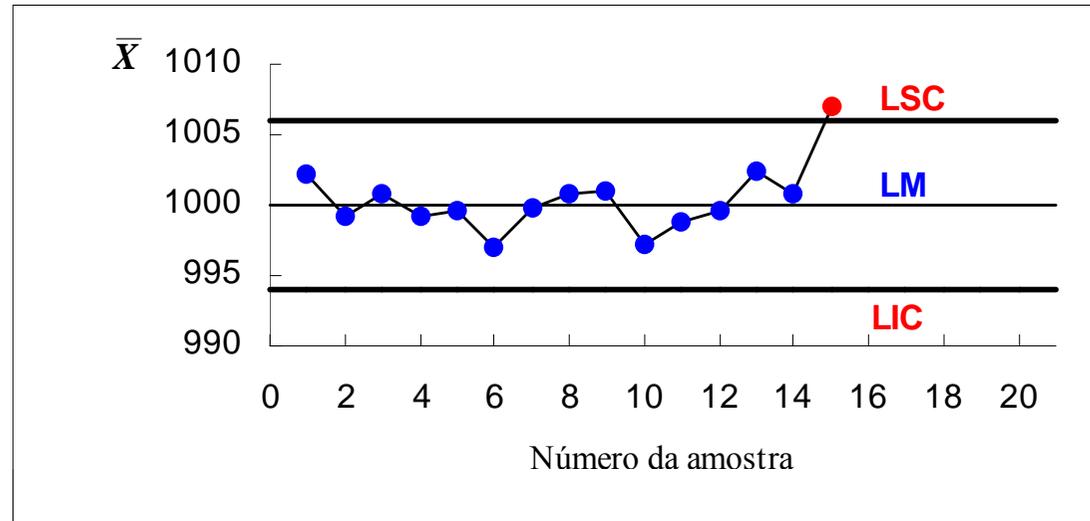


Figura 2.6: Gráficos de Controle \bar{X} e R

2.3 Condições para construção e uso dos gráficos de controle

Os limites dos gráficos de controle são determinados com base na média e no desvio padrão da distribuição da variável X quando o processo está isento de causas especiais.

A média deve sempre coincidir com o valor-alvo especificado, se este estiver disponível (ver exemplo dos saquinhos de leite). Há situações, contudo, em que esse valor é desconhecido.

Exemplo: Num restaurante *fast food*, a variável X é o tempo de atendimento. O valor-alvo será o tempo médio de atendimento nas melhores condições possíveis: o **valor médio em controle** do processo. Esse valor não é definido a priori: ele só será estimado com precisão depois que se tiver o registro dos tempos de atendimento de muitos clientes e, então, passará a ser o valor-alvo para o processo.

Para construir os gráficos de controle precisamos:

- Estimar o desvio padrão do processo; Estimar a média do processo (se não houver o valor-alvo), ou Avaliar se a estimativa da média do processo está próxima do valor-alvo pré-estabelecido.

Estes parâmetros devem ser estimados com base em valores de X obtidos de unidades produzidas durante o período em que o processo permaneceu isento de causas especiais. Assim é necessário intervir no processo antes mesmo da construção dos gráficos de controle.

A suposição de normalidade de X para a construção de gráficos de controle não é essencial. Há como adaptar estes gráficos para as situações de não-normalidade da variável de interesse.

Para a construção dos gráficos de controle é necessário que as observações da variável X sejam independentes.

Quanto aos tipos de variabilidade exibidos pelo processo temos: comportamento estacionário, estacionário mas auto-correlacionado e não-estacionário (ver Figuras 2.7 e 2.8).

- **Estacionário:** Dados variam em torno de uma média fixa sem tendência. Os valores passados dos dados não são de ajuda para predizer valores futuros (ruído branco, dados não-correlacionados).
- **Estacionário mas auto-correlacionado:** Observações sucessivas são dependentes, isto é, um valor acima da média tende a ser seguido por outro valor acima da média; um valor abaixo da média é seguido por outro do mesmo tipo, em geral. Série de dados mostra um padrão de "sequências" moderadamente longas em cada lado da média.

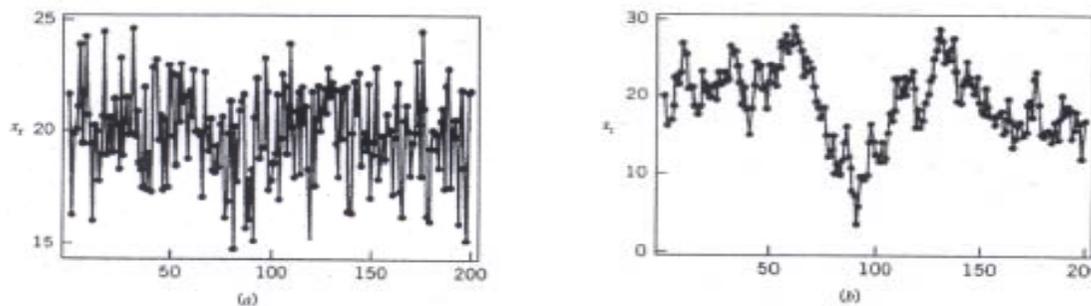


Figura 2.7 : Comportamento (a) Estacionário (b) Auto-correlacionado

•**Não-estacionário**: Dados apresentam tendência.

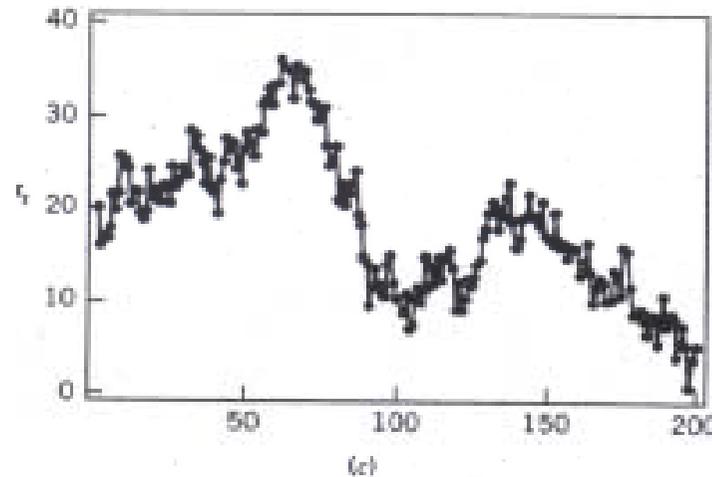


Figura 2.8 : Comportamento (c) Não-estacionário

Os gráficos de Shewhart são mais eficazes com um processo estacionário. Se os dados são correlacionados há algumas alternativas para monitorar o processo.

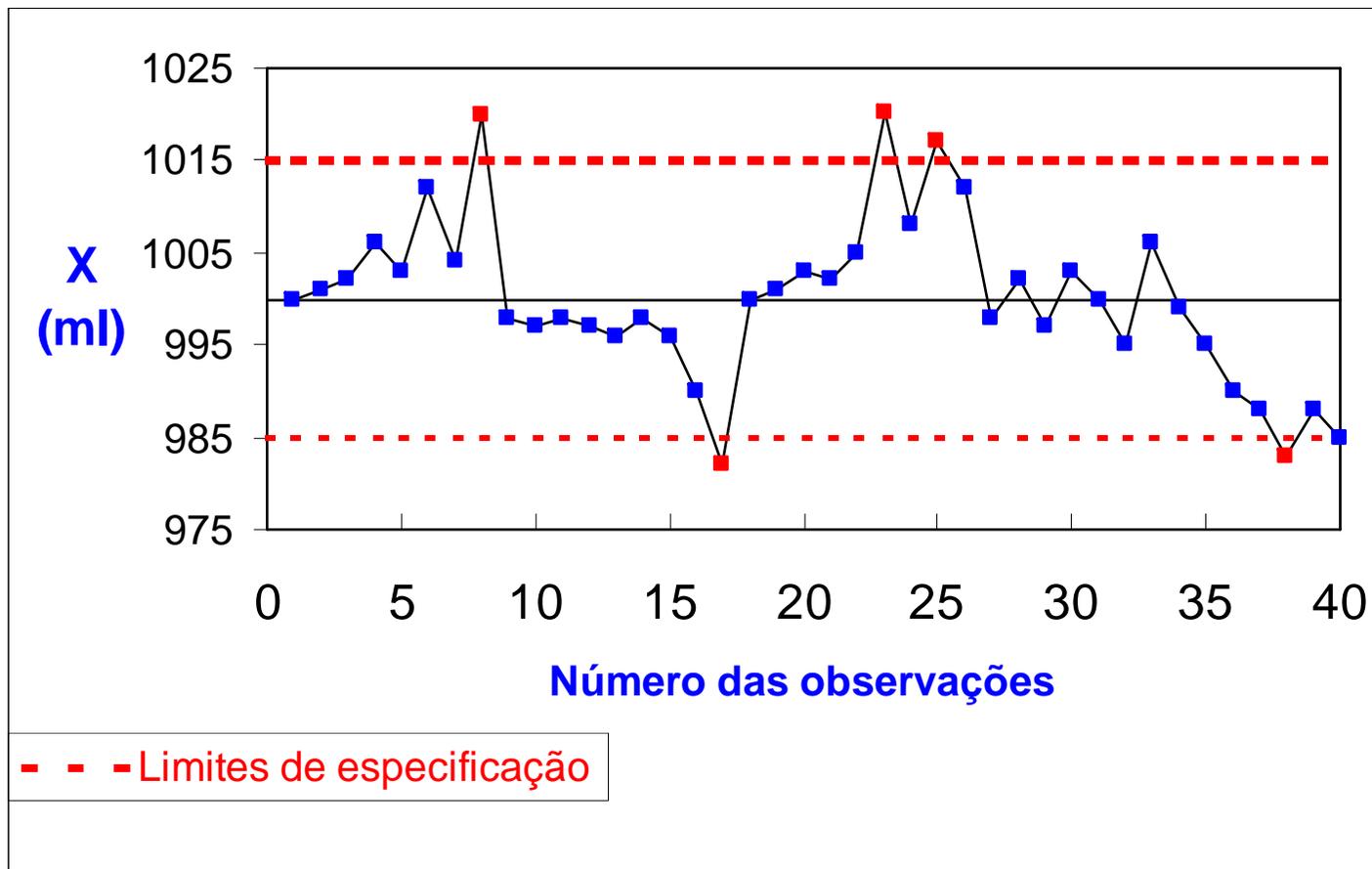
2.4 Etapa inicial: conhecendo, estabilizando e ajustando o processo

Antes da construção e da utilização dos gráficos de controle, deve haver uma etapa inicial de aprendizagem onde procura-se conhecer os fatores que afetam a característica da qualidade X .

Esta é uma etapa difícil e importante, pois ajuda a promover grandes melhorias de qualidade.

Exemplo: Saquinhos de leite. Vamos estudar o processo medindo o volume X de um saquinho a cada 15 minutos de produção.

O gráfico da Figura 2.9 apresenta, no eixo vertical, os valores da variável aleatória X e, no eixo horizontal, os números de ordem das medidas, na sequência cronológica. A Figura 2.9 mostra que a variável X não tem um comportamento estável em torno do valor-alvo 1000 ml.



**Figura 2.9: Volume dos Saquinhos de Leite
(processo instável)**

Se fosse possível “fotografar” a distribuição de X em vários instantes de tempo, possivelmente teríamos um quadro semelhante ao da Figura 2.10, ou seja, a cada instante de tempo teríamos distribuições de X diferentes.

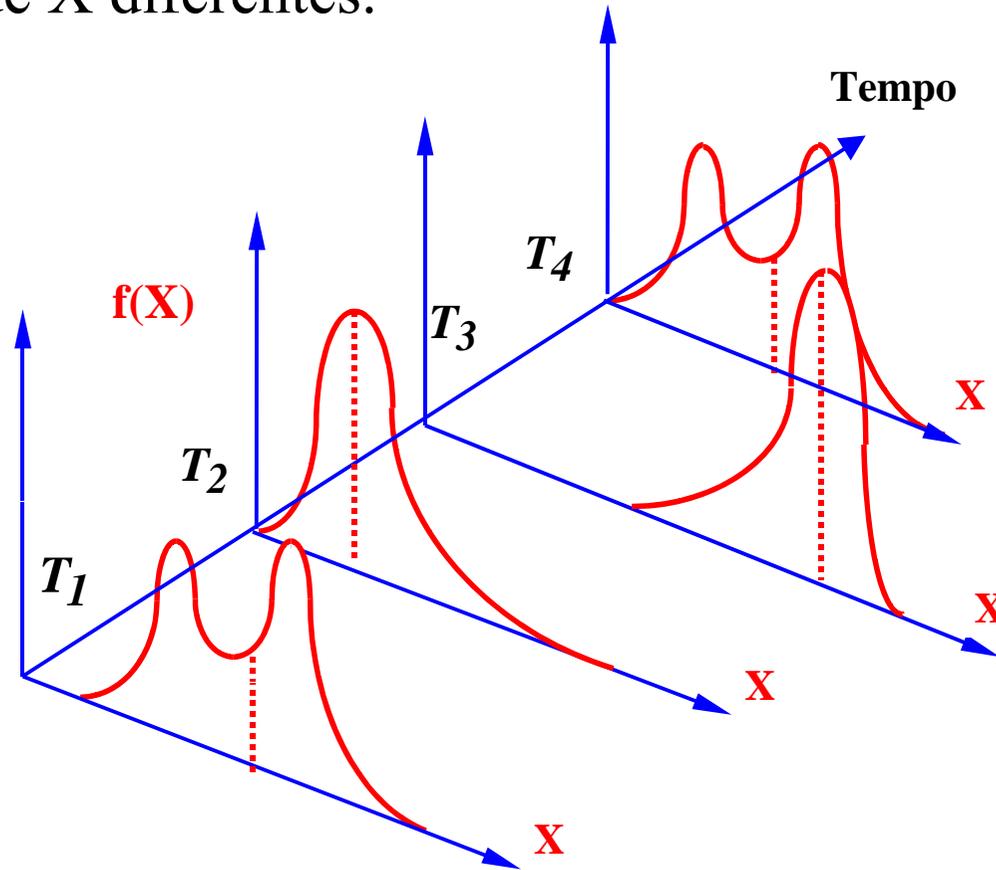


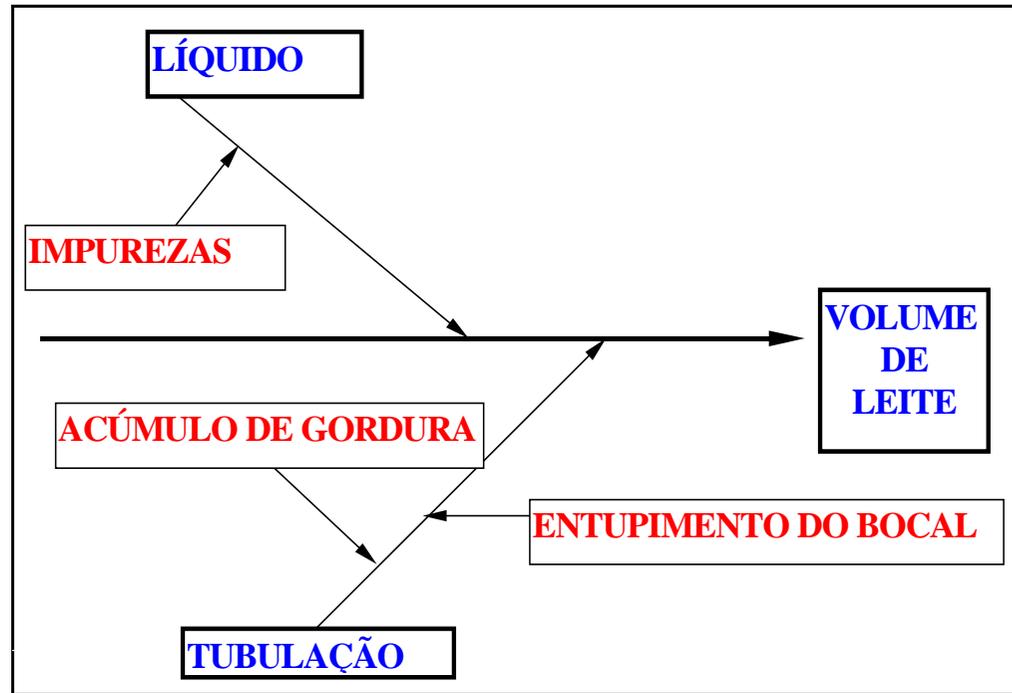
Figura 2.10: Distribuição do Volume dos Saquinhos de Leite ao Longo do Tempo (processo instável)

O processo está sob efeito de uma série de causas especiais, que em geral é possível identificar e eliminar. Após esta etapa é que procedemos à construção dos gráficos de controle, ou seja, estes devem ser construídos apenas com o processo sob controle.

A Figura 2.11 apresenta um **Diagrama de Causa e Efeito** de algumas causas especiais que podem afetar o volume X dos saquinhos de leite. Este diagrama deve ser elaborado pelo pessoal diretamente envolvido com o processo.

Uma vez diagnosticadas as causas especiais, o próximo passo consiste em eliminá-las (ver Tabela 2.4).

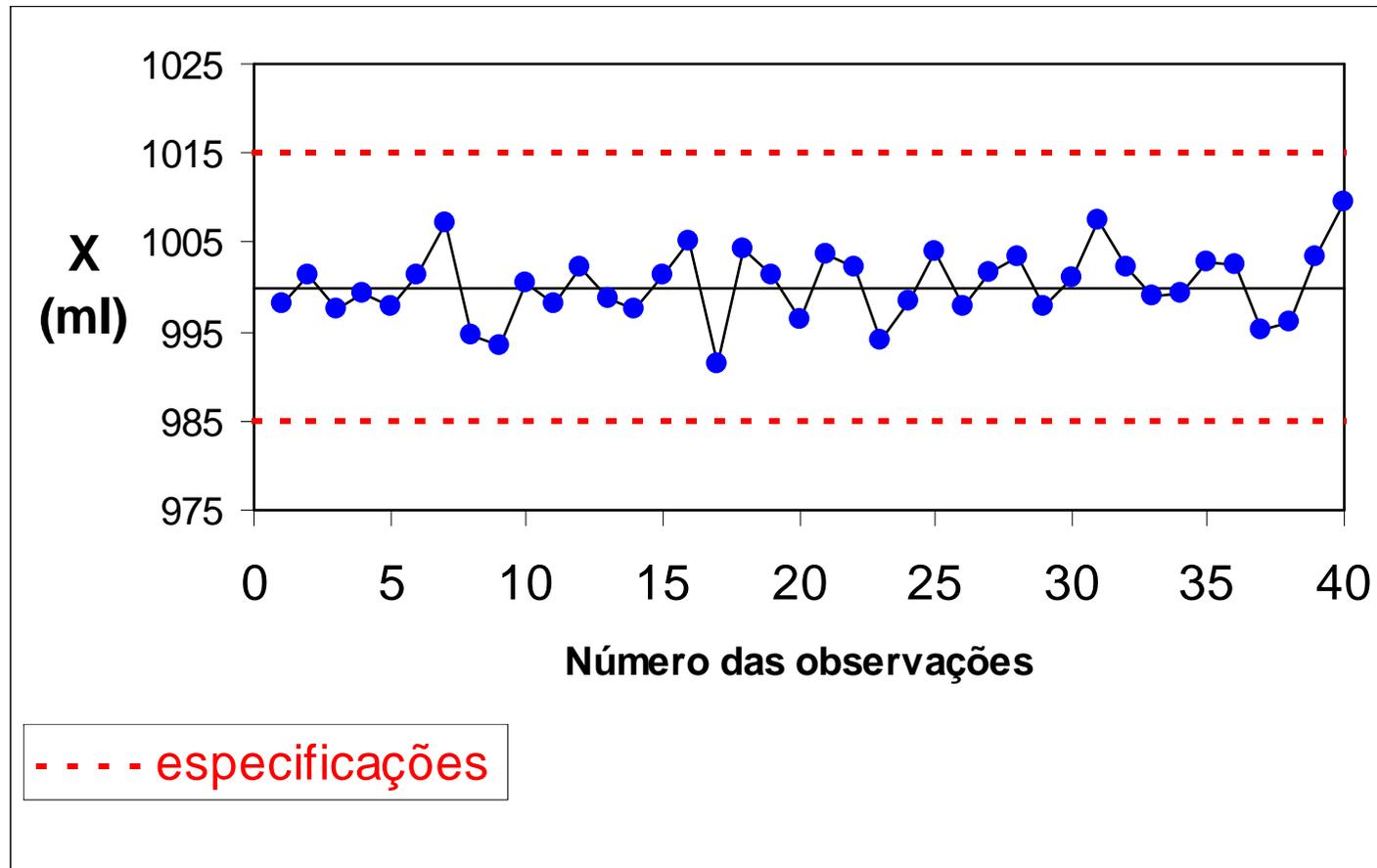
Com a eliminação das causas especiais, os valores de X passam a distribuir-se de maneira totalmente aleatória em torno do valor-alvo de 1000 ml, caracterizando um processo ajustado e estável (ver Figura 2.12).



**Figura 2.11: Diagrama de Causa e Efeito
(causas especiais que afetam o volume de leite)**

Tabela 2.4: Causas Especiais e Medidas Corretivas/Preventivas

<i>Causa especial</i>	<i>Medida corretiva/preventiva</i>
Gordura na tubulação Entupimento do bocal Impurezas no leite	Limpeza mensal da tubulação Troca semanal do bocal Utilização de filtros



**Figura 2.10: Volume dos Saquinhos de Leite
(processo estável e ajustado)**

Tudo isso caracteriza uma amostra aleatória de pontos de uma mesma distribuição, com média e desvio padrão constantes.

Nesta etapa inicial conseguimos, então, uma melhoria da qualidade do processo de empacotamento, além de uma **lista de verificação** (ver Tabela 2.4). Esta lista deve ser usada toda vez que o processo apresentar um comportamento não aleatório, ou seja, com a finalidade de trazer o processo ao estado de controle estatístico.

2.5 Subgrupos racionais

Estabilizado o processo, podemos construir gráficos de controle para monitorá-lo. A função destes gráficos é sinalizar a presença de causas especiais que venham a ocorrer.

Se a característica da qualidade X é contínua, o usual é monitorar o processo por meio dos gráficos de controle da média, \bar{X} , e da amplitude, R .

Para isto precisamos conhecer a distribuição de X , sua média μ e seu desvio padrão σ . Como μ e σ são, em geral, desconhecidos, precisamos estimá-los.

Exemplo: Saquinhos de leite. Com o processo em controle, tomamos uma amostra de 100 valores de X (ver valores da Tabela 2.1). O histograma da Figura 2.1 leva-nos a acreditar que a distribuição de X seja normal.

Se temos certeza de que o processo fica em controle durante a retirada da amostra, como estamos supondo, basta adotar \bar{X} (média aritmética dos 100 valores) como estimativa da média μ , e S^2 como estimativa de σ^2 , onde S^2 é dado por:

$$S^2 = \frac{\sum_{i=1}^{100} (X_i - \bar{X})^2}{99}.$$

Na prática, contudo, não sabemos se durante a retirada da amostra o processo ficou realmente em controle. Para lidar com essa questão, surgiu o conceito de **subgrupos racionais**, que significa retirar pequenas amostras a intervalos de tempo regulares.

No exemplo dos saquinhos de leite, ao invés de retirarmos os 100 saquinhos de uma só vez, retiramos amostras menores, distanciadas no tempo; por exemplo, uma amostra de 4 ou 5 saquinhos a cada meia hora. Cada amostra ou subgrupo racional é formado por unidades produzidas quase no mesmo instante.

Caso alguma causa especial afete o processo (com alteração na média, por exemplo), dificilmente ela ocorrerá *durante* a formação do subgrupo. Dessa forma é muito difícil que a amostra seja formada por unidades vindas de populações diferentes.

Exemplo: Saquinhos de leite. Vamos supor que a causa especial ocorra entre as retiradas de 2 amostras, e que tal causa altere a média do processo de 1000 ml para 1010 ml. Dentro de cada amostra não haverá aumento de variabilidade. Haverá, contudo, aumento de variabilidade *entre* amostras, pois os valores de \bar{X} irão variar.

Logo, a variância do processo deve ser estimada com base na dispersão dos valores *dentro* das amostras (que independe da média do processo).

2.6 Estimando a variabilidade do processo

Existem várias maneiras de estimar σ , o desvio padrão do processo. Veremos a seguir quatro delas.

2.6.1 Estimador S_A

$$S_A = \frac{1}{c_4} \sqrt{\frac{\sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{\bar{X}})^2}{mn - 1}}$$

- X_{ij} é o j -ésimo elemento do i -ésimo subgrupo;
- n é o tamanho dos subgrupos;
- m é o número de subgrupos;
- $\bar{\bar{X}} = \sum_{i=1}^m \sum_{j=1}^n \bar{X}_i / m$ (média global; média das médias dos subgrupos), e
- c_4 é uma constante.

Exemplo: Na Tabela 2.5, onde há valores de X em 8 subgrupos de tamanho 5 ($m = 8$ e $n = 5$), X_{42} corresponde ao segundo valor de X do quarto subgrupo, ou seja, $X_{42} = 1002,1$. Na coluna 7 da Tabela 2.5, estão os valores das médias dos subgrupos ($\bar{X}_i = \sum_{i=1}^n X_{ij} / n$)

Tabela 2.5: Valores de $X_{ij} \sim N(1000,4)$, \bar{X}_i , R_i e S_i

Subgrupo (i)	Elemento (j) do subgrupo (i)					\bar{X}_i	R_i	S_i
	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}			
1	992,9	1006,7	1002,7	1005,4	998,3	1001,2	13,8	5,6
2	1001,3	995,3	999,0	999,1	996,5	998,2	6,0	2,4
3	1001,2	1001,4	999,0	997,8	994,2	998,7	7,2	2,9
4	993,3	1002,1	998,7	993,6	996,6	996,9	8,8	3,7
5	996,8	1006,4	1006,9	994,5	998,4	1000,6	12,4	5,7
6	1000,9	1004,2	999,2	997,8	997,9	1000,0	6,4	2,7
7	1000,2	1002,6	998,3	1006,4	1005,8	1002,7	8,1	3,5
8	1003,3	996,1	1000,5	995,2	1005,8	1000,2	10,6	4,6

O estimador S_A considera as m amostras de n unidades como uma única grande amostra, com mn unidades. Ora, esse estimador é tendencioso, pois seu valor esperado é dado por $c_4\sigma$, onde c_4 é uma função de mn . Logo, S_A/c_4 é um estimador não viesado de σ .

2.6.2 Estimador S_B : baseado no desvio padrão das médias dos subgrupos.

$$S_B = \left[\frac{1}{c_4} \sqrt{\frac{\sum_{i=1}^m (\bar{X}_i - \bar{\bar{X}})^2}{m-1}} \right] \sqrt{n}$$

Se \bar{X} é a média dos valores de uma amostra aleatória de tamanho n de uma v. a. X , então $\bar{\bar{X}}$ também é uma v. a., e as variâncias das distribuições de X e de \bar{X} mantêm a relação: $\sigma_{\bar{X}}^2 = \sigma^2 / n$ e seus desvios padrões mantêm, portanto, a relação $\sigma = \sigma_{\bar{X}} \sqrt{n}$. A expressão entre colchetes é o clássico estimador do desvio padrão de \bar{X} , dividido por c_4 para correção de viés. O fator c_4 é função de m , número de parcelas do somatório. Ao multiplicarmos o termo entre colchetes por raiz de n , obtemos um estimador para σ .

Tabela 2.5: Valores de $X_{ij} \sim N(1000,4)$, \bar{X}_i , R_i e S_i

Subgrupo (i)	Elemento (j) do subgrupo (i)					\bar{X}_i	R_i	S_i
	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}			
1	992,9	1006,7	1002,7	1005,4	998,3	1001,2	13,8	5,6
2	1001,3	995,3	999,0	999,1	996,5	998,2	6,0	2,4
3	1001,2	1001,4	999,0	997,8	994,2	998,7	7,2	2,9
4	993,3	1002,1	998,7	993,6	996,6	996,9	8,8	3,7
5	996,8	1006,4	1006,9	994,5	998,4	1000,6	12,4	5,7
6	1000,9	1004,2	999,2	997,8	997,9	1000,0	6,4	2,7
7	1000,2	1002,6	998,3	1006,4	1005,8	1002,7	8,1	3,5
8	1003,3	996,1	1000,5	995,2	1005,8	1000,2	10,6	4,6

2.6.3 Estimador S_C : baseado nos desvios padrões amostrais S_i dos m subgrupos.

$$S_C = \frac{\bar{S}}{c_4}$$

onde $\bar{S} = \sum_{i=1}^m S_i / m$ e

$$S_i = \sqrt{\frac{\sum_{j=1}^n (X_{ij} - \bar{X}_i)^2}{n-1}}$$

Tabela 2.5: Valores de $X_{ij} \sim N(1000,4)$, \bar{X}_i , R_i e S_i

Subgrupo (i)	Elemento (j) do subgrupo (i)					\bar{X}_i	R_i	S_i
	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}			
1	992,9	1006,7	1002,7	1005,4	998,3	1001,2	13,8	5,6
2	1001,3	995,3	999,0	999,1	996,5	998,2	6,0	2,4
3	1001,2	1001,4	999,0	997,8	994,2	998,7	7,2	2,9
4	993,3	1002,1	998,7	993,6	996,6	996,9	8,8	3,7
5	996,8	1006,4	1006,9	994,5	998,4	1000,6	12,4	5,7
6	1000,9	1004,2	999,2	997,8	997,9	1000,0	6,4	2,7
7	1000,2	1002,6	998,3	1006,4	1005,8	1002,7	8,1	3,5
8	1003,3	996,1	1000,5	995,2	1005,8	1000,2	10,6	4,6

Qualquer dos valores S_i/c_4 é uma estimativa não viesada de σ . Porém, para amostras pequenas ($n = 3, 4$ ou 5), a variância dessas estimativas é grande; portanto ao se adotar \bar{S} , a média de todos os m valores de S_i , em vez de qualquer deles, no numerador de S_C , obtém-se uma estimativa de σ mais precisa. O fator c_4 é função do tamanho n dos subgrupos.

2.6.4 Estimador S_D : baseado na amplitude amostral R .

$$S_D = \bar{R} / d_2$$

onde $\bar{R} = \sum_{i=1}^m R_i / m$,

sendo R_i a amplitude do i -ésimo subgrupo: diferença entre o maior e o menor valores do subgrupo. O fator d_2 é função de n .

Exemplo: Para o terceiro subgrupo da Tabela 2.5, o maior valor é $X_{32} = 1001,4$ e o menor é $X_{35} = 994,2$; portanto, $R_3 = 1001,4 - 994,2 = 7,2$.

O valor esperado de R , amplitude de amostras com n elementos obtidos da v.a. normal X , de média μ e desvio padrão σ , é $E(R) = d_2\sigma$. Como \bar{R} é um estimador não viesado de $E(R)$, \bar{R}/d_2 é um estimador não viciado de σ .

Tabela 2.5: Valores de $X_{ij} \sim N(1000,4)$, \bar{X}_i , R_i e S_i

Subgrupo (i)	Elemento (j) do subgrupo (i)					\bar{X}_i	R_i	S_i
	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}			
1	992,9	1006,7	1002,7	1005,4	998,3	1001,2	13,8	5,6
2	1001,3	995,3	999,0	999,1	996,5	998,2	6,0	2,4
3	1001,2	1001,4	999,0	997,8	994,2	998,7	7,2	2,9
4	993,3	1002,1	998,7	993,6	996,6	996,9	8,8	3,7
5	996,8	1006,4	1006,9	994,5	998,4	1000,6	12,4	5,7
6	1000,9	1004,2	999,2	997,8	997,9	1000,0	6,4	2,7
7	1000,2	1002,6	998,3	1006,4	1005,8	1002,7	8,1	3,5
8	1003,3	996,1	1000,5	995,2	1005,8	1000,2	10,6	4,6

Se a média do processo se altera entre a retirada da primeira e da *m-ésima* amostras, as estimativas S_A e S_B são muito afetadas. Afeta S_A , porque ela é função da dispersão de todos os pontos, e afeta S_B , porque ela é baseada nas diferenças entre as médias amostrais. As estimativas S_C e S_D são mais confiáveis, pois, por serem baseadas apenas na dispersão dos valores *dentro* das amostras, são insensíveis a causas especiais.

Exemplo: Saquinhos de leite. Os valores de X das Tabelas 2.5 e 2.6 foram obtidos de uma distribuição normal com média 1000 e desvio padrão 4, com exceção do segundo subgrupo (segunda linha) da Tabela 2.6, cujos valores vieram de uma normal com média 1010 e desvio padrão igual a 4. Estamos supondo que, na Tabela 2.6, a segunda amostra foi obtida quando uma causa especial estava em ação, o que aumentou o volume médio em 10 ml.

Tabela 2.5: Valores de $X_{ij} \sim N(1000,4)$, \bar{X}_i , R_i e S_i

Subgrupo (i)	Elemento (j) do subgrupo (i)					\bar{X}_i	R_i	S_i
	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}			
1	992,9	1006,7	1002,7	1005,4	998,3	1001,2	13,8	5,6
2	1001,3	995,3	999,0	999,1	996,5	998,2	6,0	2,4
3	1001,2	1001,4	999,0	997,8	994,2	998,7	7,2	2,9
4	993,3	1002,1	998,7	993,6	996,6	996,9	8,8	3,7
5	996,8	1006,4	1006,9	994,5	998,4	1000,6	12,4	5,7
6	1000,9	1004,2	999,2	997,8	997,9	1000,0	6,4	2,7
7	1000,2	1002,6	998,3	1006,4	1005,8	1002,7	8,1	3,5
8	1003,3	996,1	1000,5	995,2	1005,8	1000,2	10,6	4,6

$$S_A = \frac{1}{c_4} \sqrt{\frac{\sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{X})^2}{mn - 1}}$$

$$S_B = \left[\frac{1}{c_4} \sqrt{\frac{\sum_{i=1}^m (\bar{X}_i - \bar{X})^2}{m - 1}} \right] \sqrt{n}$$

$$S_C = \frac{\bar{S}}{c_4}$$

$$S_A = 4,1$$

$$S_C = 4,1$$

$$S_B = 4,2$$

$$S_D = 3,9$$

$$S_D = \bar{R} / d_2$$

Tabela 2.6: Valores de $X_{ij} \sim N(1000,4)$, \bar{X}_i , R_i e S_i

Subgrupo (i)	Elemento (j) do subgrupo (i)					\bar{X}_i	R_i	S_i
	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}			
1	992,9	1006,7	1002,7	1005,4	998,3	1001,2	13,8	5,6
2	1008,2	1009,3	1010,8	1008,4	1010,8	1009,5	2,6	1,3
3	1001,2	1001,4	999,0	997,8	994,2	998,7	7,2	2,9
4	993,3	1002,1	998,7	993,6	996,6	996,9	8,8	3,7
5	996,8	1006,4	1006,9	994,5	998,4	1000,6	12,4	5,7
6	1000,9	1004,2	999,2	997,8	997,9	1000,0	6,4	2,7
7	1000,2	1002,6	998,3	1006,4	1005,8	1002,7	8,1	3,5
8	1003,3	996,1	1000,5	995,2	1005,8	1000,2	10,6	4,6

$$S_A = \frac{1}{c_4} \sqrt{\frac{\sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{X})^2}{mn - 1}}$$

$$S_B = \left[\frac{1}{c_4} \sqrt{\frac{\sum_{i=1}^m (\bar{X}_i - \bar{X})^2}{m - 1}} \right] \sqrt{n}$$

$$S_C = \frac{\bar{S}}{c_4}$$

$$S_D = \bar{R} / d_2$$

$$S_A = 5,1$$

$$S_B = 8,7$$

$$S_C = 4,0$$

$$S_D = 3,8$$

As estimativas S_A e S_B são muito afetadas, de modo que superestimam o desvio padrão σ . Já, as estimativas S_C e S_D são mais robustas (menos sensíveis à causas especiais) e devem ser preferidas. Para escolher entre S_C e S_D olhamos suas *eficiências*. Um estimador não viesado é tanto mais eficiente quanto menor for sua variância. Para subgrupos grandes, n maior ou igual a 10 unidades, S_C é mais eficiente que S_D (notar que S_C usa todas as observações do subgrupo enquanto que S_D usa somente duas). Para n menor do que 10, a eficiência de S_C é quase igual à de S_D , sendo que S_D é mais simples de ser calculado. Assim, vamos adotar S_D como estimativa de σ (é o estimador mais usado no controle estatístico de processos).

2.7 Amostragem estratificada

Devemos sempre conhecer a origem dos nossos dados. Eles devem vir da mesma fonte. No caso dos saquinhos de leite, se existirem vários bocais de enchimento, é necessário construir um gráfico de controle para cada bocal (isso é chamado de *estratificação* das amostras).

- Se as amostras contiverem saquinhos enchidos por bocais diferentes, ao aparecer uma causa especial, ficará difícil saber qual bocal está afetado;
- Se um bocal estiver desregulado para mais e outro para menos, os efeitos se compensarão e o gráfico de controle da média não acusará a causa especial presente. Já, o gráfico R irá acusar um aumento na variabilidade quando esta não existe.
- Se os problemas ocorrerem antes do ajuste do processo, durante a etapa inicial, serão obtidas estimativas erradas dos parâmetros do processo. Uma única média será estimada (na realidade, há mais de uma média para o processo, uma para cada bocal) e a variância estimada será muito grande, gerando gráficos de controle com limites muito largos, que dificultarão a detecção e a eliminação da causa especial.

Exercício: Sejam os dados da Tabela 2.7. Obter e interpretar os valores de S_A , S_B , S_C , e S_D .

Tabela 2.7: Valores de X_{ij} , \bar{X}_i , R_i e S_i

Subgrupo (i)	Elemento (j) do subgrupo (i)					\bar{X}_i	R_i	S_i
	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}			
1	439,5	453,3	449,3	452	444,9			
2	447,9	441,9	445,6	445,7	443,1			
3	447,8	448	445,6	444,4	440,8			
4	439,9	448,7	445,3	440,2	443,2			
5	443,4	453	453,5	441,1	445			
6	447,5	450,8	445,8	444,4	444,5			
7	446,8	449,2	444,9	453,0	452,4			
8	449,9	442,7	447,1	441,8	452,4			