

Gráficos de Controle para Variáveis

CE219 - Controle Estatístico de Qualidade

Prof. Cesar Taconeli
taconeli@ufpr.br

Prof. Walmes Zeviani
walmes@ufpr.br

Laboratório de Estatística e Geoinformação
Departamento de Estatística
Universidade Federal do Paraná



Introdução

Tipos de características de qualidade

Na literatura de CEQ, é usual distinguir as características da qualidade em duas categorias:

- ▶ **Variáveis:** Compreendem as **características numéricas**, usualmente correspondente a medidas.
 - ▶ Resistência de placas de alumínio.
 - ▶ Quantidade de café em embalagens.
 - ▶ Diâmetros de anéis de pistão.
 - ▶ Viscosidade da calda de chocolate.
- ▶ **Atributos:** Compreendem as características resultantes de **contagens** nos ou **classificação** dos itens produzidos.
 - ▶ Proporção de clientes insatisfeitos com um produto/serviço.
 - ▶ Número de não conformidades por m^2 de tecido.

Vamos abordar, inicialmente, os gráficos de controle para variáveis.

Estatísticas monitoradas

- ▶ A forma mais convencional para o monitoramento de variáveis do processo consiste na utilização de dois gráficos de controle: um para a média e outro para a dispersão.
- ▶ O monitoramento da média é realizado com base nas médias de amostras (\bar{X}) extraídas periodicamente do processo.
- ▶ Já o monitoramento da dispersão pode ser baseado no desvio padrão, na variância ou até mesmo nas amplitudes amostrais.

Estimação de parâmetros

- ▶ Já definimos, anteriormente, o gráfico de controle \bar{x} de Shewart para a média com limites $3\sigma_{\bar{x}}$ na forma:

$$LSC = \mu_0 + 3\frac{\sigma_0}{\sqrt{n}}$$

$$LC = \mu_0$$

$$LIC = \mu_0 - 3\frac{\sigma_0}{\sqrt{n}}$$

- ▶ Na prática, em geral μ_0 e σ_0 são **desconhecidos**, devendo ser estimados.
- ▶ A **estimação dos parâmetros** do processo baseia-se nos resultados de m amostras preliminares, extraídas do processo no cenário de controle estatístico (fase I).

Duas fases da aplicação de GC

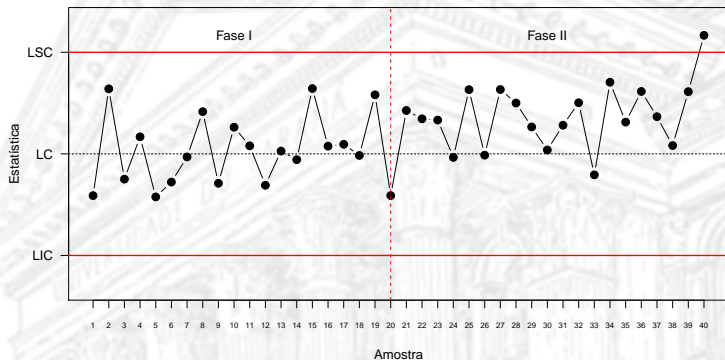


Figura 1. Ilustração das duas fases de um gráfico de controle.



Gráficos de controle para \bar{x} e R

Definição das estatísticas

- ▶ Considere m amostras produzidas na fase I de um gráfico de controle.
- ▶ Sejam $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ as m médias amostrais e R_1, R_2, \dots, R_m as amplitudes amostrais:

$$R(x_1, x_2, \dots, x_n) = \max\{x_1, x_2, \dots, x_n\} - \min\{x_1, x_2, \dots, x_n\}.$$

- ▶ As estimativas de μ_0 e σ_0 baseiam-se em:

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m}{m} \quad \text{e} \quad \bar{R} = \frac{R_1 + R_2 + \dots + R_m}{m}.$$

A forma dos dados coletados

Tabela 1. Estrutura padrão dos dados para a fase 1 de gráficos de controle \bar{x} e R .

Amostra	Observações				\bar{x}	R
1	x_{11}	x_{12}	...	x_{1n}	\bar{x}_1	R_1
2	x_{21}	x_{22}	...	x_{2n}	\bar{x}_2	R_2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
m	x_{m1}	x_{m2}	...	x_{mn}	\bar{x}_m	R_m

A distribuição da amplitude

- ▶ Os gráficos de controle \bar{x} e R baseiam-se na distribuição da amplitude relativa:

$$W = \frac{R}{\sigma}.$$

- ▶ A média e o desvio padrão de W são denotados, respectivamente, por:

$$E(W) = d_2, \quad \text{Var}(W) = d_3,$$

em que d_2 e d_3 são funções, exclusivamente, de n .

- ▶ Um estimador não viciado de σ , baseado nas amplitudes amostrais e na distribuição da amplitude relativa, fica dado por:

$$\hat{\sigma} = \frac{\bar{R}}{d_2}.$$

Limites de controle para a média usando R

- ▶ O gráfico de controle \bar{x} , usando $\bar{\bar{x}}$ e R

$$LSC = \bar{\bar{x}} + 3 \frac{\bar{R}/d_2}{\sqrt{n}}$$

$$LC = \bar{\bar{x}}$$

$$LIC = \bar{\bar{x}} - 3 \frac{\bar{R}/d_2}{\sqrt{n}}$$

Monitoramento da dispersão

- ▶ Para monitorar a dispersão do processo, usaremos o gráfico R . Para isso, precisamos de uma estimativa de σ_R .
- ▶ Supondo normalidade para a característica da qualidade, $\hat{\sigma}_R$ pode ser determinado com base na distribuição da amplitude relativa, por:

$$\hat{\sigma}_R = d_3 \frac{\bar{R}}{d_2}.$$

Limites de controle para dispersão usando \bar{R}

- ▶ O gráfico de controle \bar{R} , para monitorar a dispersão do processo, fica estabelecido por:

$$LSC = \bar{R} + 3d_3 \frac{\bar{R}}{d_2}$$

$$LC = \bar{R}$$

$$LIC = \bar{R} - 3d_3 \frac{\bar{R}}{d_2}$$



Gráficos de controle \bar{x} e s

Gráficos de controle com desvio-padrão

- ▶ Embora os gráficos de controle \bar{x} e R sejam bastante simples, usar as amplitudes amostrais para estimar σ é uma alternativa pouco eficiente, sobretudo para tamanhos de amostras moderados ou grandes.
- ▶ Usar os desvios padrões amostrais é uma alternativa **mais eficiente**, permitindo lidar, ainda, com **amostras de tamanhos variáveis**.
- ▶ As etapas para construção dos gráficos de controle \bar{x} e s são semelhantes às apresentadas para os gráficos \bar{x} e R .

Estimador da dispersão baseado em s

- ▶ Um estimador não viciado de σ^2 , a variância populacional, é a variância amostral, definida por:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

- ▶ No entanto, $s = \sqrt{s^2}$ não é um estimador não viciado de σ , o desvio padrão populacional.
- ▶ Se os dados monitorados têm distribuição Normal, então $E(s) = c_4\sigma$, sendo c_4 uma constante que depende unicamente de n :

$$c_4 = \sqrt{\frac{2}{n-1}} \times \frac{\left(\frac{n}{2} - 1\right)!}{\left(\frac{n-1}{2} - 1\right)!}.$$

Limites de controle para a média usando S

- ▶ Assim, s/c_4 é um estimador não viciado de σ . Se dispomos de s_1, s_2, \dots, s_m , desvios padrões de m amostras aleatórias do processo sob controle, um estimador não viciado de σ é dado por:

$$\hat{\sigma} = \frac{\bar{s}}{c_4},$$

$$\text{sendo } \bar{s} = \frac{s_1 + s_2 + \dots + s_m}{m}.$$

- ▶ Baseado nesse estimador, o gráfico de controle \bar{x} tem limites:

$$LSC = \bar{\bar{x}} + 3 \frac{\bar{s}/c_4}{\sqrt{n}},$$

$$LC = \bar{\bar{x}}$$

$$LIC = \bar{\bar{x}} - 3 \frac{\bar{s}/c_4}{\sqrt{n}}.$$

Limites de controle para dispersão usando S

- ▶ O gráfico de controle s requer o desvio padrão desse estimador que, sob normalidade da característica da qualidade monitorada, é dado por $\sigma_s = \sigma\sqrt{1 - c_4^2}$.
- ▶ Os limites de controle para o gráfico s ficam, portanto, dados por:

$$LSC = \bar{s} + 3\frac{\bar{s}}{c_4}\sqrt{1 - c_4^2}$$

$$LC = \bar{s}$$

$$LIC = \bar{s} - 3\frac{\bar{s}}{c_4}\sqrt{1 - c_4^2}.$$



Tamanhos de amostra variável

Gráficos de controle \bar{x} e s

Diferentes motivos podem justificar amostras de tamanhos variáveis em gráficos de controle, dentre os quais:

- ▶ Produção irregular ou disponibilidade variável de itens para amostragem.
- ▶ Perda de dados amostrais (acidentes, contaminação).
- ▶ A tomada de amostras de diferentes tamanhos pode ter sido planejada (planos de amostragem adaptativa), dentre outros.

Estimativas com amostras de tamanho variável

Para o gráfico \bar{x} , usamos o seguinte estimador para a média do processo:

$$\bar{\bar{x}} = \frac{\sum_{i=1}^m n_i \bar{x}_i}{\sum_{i=1}^m n_i}.$$

Para estimar o desvio padrão do processo temos diferentes opções, dentre as quais:

- ▶ **MVLU-E-R**: Minimum Variance Linear Unbiased Estimator baseado nas amplitudes amostrais:

$$\hat{\sigma} = \frac{1}{f_1 + f_2 + \cdots + f_m} \sum_{i=1}^m \frac{f_i R_i}{d_2(n_i)}, \quad f_i = \frac{[d_2(n_i)]^2}{[d_3(n_i)]^2}.$$

Estimativas com amostras de tamanho variável

- ▶ **MVLUE-S:** Minimum Variance Linear Unbiased Estimator based on sample standard deviations:"

$$\hat{\sigma} = \frac{1}{h_1 + h_2 + \cdots + h_m} \sum_{i=1}^m \frac{h_i s_i}{c_4(n_i)}, \quad h_i = \frac{[c_4(n_i)]^2}{1 - [c_4(n_i)]^2}.$$

- ▶ **RMSDF:** Root-Mean-Square estimator:

$$\hat{\sigma} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_m - 1)s_m^2}{(n_1 + n_2 + \cdots + n_m) - m}}.$$

Limites de controle média e dispersão

- ▶ Os gráficos de controle \bar{x} ficam então definidos conforme as expressões anteriores, mas substituindo $\hat{\sigma} = \bar{R}/d_2$ e $\hat{\sigma} = \bar{s}/c_4$ por algum dos estimadores apresentados.
- ▶ De forma semelhante, os gráficos de controle R e S também terão limites variáveis conforme os diferentes tamanhos amostrais.




Gráfico de controle baseado em medidas
individuais

Situações de amostras de tamanho um

Diversos motivos podem conduzir ao monitoramento de processos com base em amostras unitárias, dentre os quais:

- ▶ Possibilidade de medir todos os itens produzidos, não havendo motivos para compor amostras não unitárias.
- ▶ Produção muito lenta, inviabilizando a formação de subgrupos racionais.
- ▶ Várias medidas são tomadas em uma mesma unidade de produto.
- ▶ Medidas individuais são comuns em processos não industriais, como nas áreas de serviços e negócios, em situações em que a formação de subgrupos é inviável.

Definição das estatísticas

- ▶ Para a construção de gráficos de controle de Shewart para medidas individuais é usual estimar o desvio padrão com base nas amplitudes móveis (*move ranges*):

$$MR_i = |x_i - x_{i-1}|, \quad i = 2, 3, \dots, m.$$

- ▶ A estratégia é utilizar as diferenças de observações consecutivas como amplitudes amostrais, para fins de estimação do desvio padrão do processo.
- ▶ Nesse caso, o estimador de σ fica dado por $\hat{\sigma} = \overline{MR}/d_2$.

Limites de controle para o valor individual

O gráfico de controle para medidas individuais tem limites dados por:

$$LSC = \bar{x} + 3 \frac{\overline{MR}}{d_2}$$

$$LC = \bar{x}$$

$$LIC = \bar{x} - 3 \frac{\overline{MR}}{d_2}$$

Limite de controle para a dispersão

- ▶ Gráficos de controle para as amplitudes móveis podem ser obtidos simplesmente se adaptando o gráfico de controle R , apresentado anteriormente.
- ▶ Uma alternativa é estimar o desvio padrão do processo considerando todas as amostras da fase I como uma única amostra:

$$\hat{\sigma} = \frac{s}{c_4(m)}, \quad s = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (x_i - \bar{x})^2}, \quad \bar{x} = \frac{1}{m} \sum_{i=1}^m x_i.$$

- ▶ Por considerar o conjunto de m amostras, esse estimador pode ser afetado por alterações na média do processo.

Considerações para medidas individuais

Algumas notas sobre gráficos de controle de Shewart para medidas individuais:

- ▶ Por não se trabalhar com médias de amostras, mas sim com medidas individuais, a hipótese de distribuição normal para a característica analisada é fundamental.
- ▶ Amplitudes móveis são correlacionadas, o que, por si só, pode produzir padrões não aleatórios em gráficos de controle.
- ▶ Os gráficos de controle de Shewart para medidas individuais são pouco sensíveis a pequenas alterações nos parâmetros dos processos (veremos opções de gráficos mais poderosos adiante).
- ▶ Gráficos de controle para medidas individuais são mais afetados por possíveis outliers.



A suposição de normalidade

Gráficos de controle e a suposição de normalidade

- ▶ Gráficos de controle de Shewart são relativamente robustos à não normalidade da característica da qualidade sob estudo (os gráficos R e s são menos robustos do que os gráficos \bar{x}).
- ▶ Em geral, os casos em que os gráficos são mais afetados pela não normalidade da variável envolve distribuições acentuadamente assimétricas.
- ▶ Em tais situações, os reais riscos de decisão (dos tipos I e II) podem ser bastante diferentes daqueles obtidos considerando normalidade.

Verificação da normalidade

- ▶ A verificação da suposição de normalidade é particularmente recomendável quando se trabalha com amostras pequenas (eventualmente unitárias).
- ▶ Dentre os métodos para avaliação de normalidade, destacam-se:
 - ▶ Histograma com sobreposição da curva da distribuição normal com parâmetros estimados.
 - ▶ Gráfico de probabilidades Normal.
 - ▶ Testes de hipóteses, como os testes de Shapiro-Wilks ou Lilliefors (em ambos os casos a hipótese nula é a de normalidade).

Gráfico de probabilidades Normal

- ▶ O gráfico de probabilidades Normal é um método gráfico para investigar se os dados que dispomos de fato foram produzidos por uma distribuição Normal.
- ▶ O método consiste, basicamente, na plotagem, em um gráfico de duas dimensões, dos quantis amostrais vs os correspondentes quantis teóricos da distribuição Normal.
- ▶ Método gráfico semelhante pode ser usado para outras distribuições teóricas de probabilidades.

Construção do gráfico

- ▶ **Passo 1:** Ordene os dados amostrais (x_1, x_2, \dots, x_n) do menor ao maior valor $(x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)})$.
- ▶ **Passo 2:** Calcule os quantis da distribuição Normal padrão correspondentes a cada observação $(z_j, j = 1, 2, \dots, n)$, que satisfazem:

$$z_j : \Phi(z_j) = \Pr(Z \leq z_j) = \frac{j - 1/2}{n}.$$

- ▶ **Passo 3:** Plotar os pares $(z_1, x_{(1)}), (z_2, x_{(2)}), \dots, (z_n, x_{(n)})$.

A hipótese de normalidade é verificada se os pontos estiverem dispostos (aleatoriamente) em torno de uma reta.

Cursos de ação para a rejeição da normalidade

- ▶ Se a suposição de normalidade for necessária, mas não verificada, temos, como alternativas:
 - ▶ Usar outra distribuição de referência para a estatística monitorada, se disponível.
 - ▶ Usar gráficos de controle não paramétricos.
 - ▶ Aplicar alguma transformação aos dados, buscando melhor aproximação à normalidade.
- ▶ O método de Box e Cox é amplamente utilizado na determinação de uma transformação (do tipo potência) aos dados, que produza melhor aproximação à distribuição Normal.

Transformação pelo método de Box-Cox

- ▶ Embora em algumas situações transformações possam ser selecionadas empiricamente, um procedimento mais formal para seleção de uma boa transformação pode ser bastante útil.
- ▶ O método de Box-Cox contempla as transformações do tipo potência ($x^* = x^\lambda$), em que λ é um parâmetro a ser determinado.
- ▶ A família de transformações do tipo potência proposta por Box e Cox é definida por:

$$x^* = \frac{x^\lambda - 1}{\lambda \dot{x}^{\lambda-1}} \text{ se } \lambda \neq 0 \quad \text{ou} \quad x^* = \ln(x) \text{ se } \lambda = 0$$

sendo \dot{x} a média geométrica das observações.

Determinação de λ

- ▶ O valor ótimo (ou um conjunto admissível de valores) para λ é obtido com base na maximização da função de (log)verossimilhança para λ , assumindo normalidade:

$$l(\lambda) = -\frac{n}{2} \ln\left(\sum_{i=1}^n (x_i^* - \bar{x}^*)^2\right) + (\lambda - 1) \sum \ln(x_i)$$

- ▶ A estimativa de máxima verossimilhança de λ é o valor desse parâmetro tal que $l(\lambda)$ é máximo.
- ▶ Um intervalo $100(1 - \alpha)\%$ para λ é definido pelo conjunto de valores de λ tal que:

$$l(\hat{\lambda}) - l(\lambda) \leq \frac{1}{2} \chi_{\alpha,1}^2.$$

Transformações particulares

Tabela 2. Casos particulares de transformações decorrentes da aplicação do método Box-Cox.

λ	Transformação
-2	Inversa quadrática
-1	Inversa
0	Logarítmica
1/2	Raiz quadrada
1	Não transformada
2	Quadrática
3	Cúbica

Como proceder com a transformação

Uma vez encontrada uma transformação apropriada aos dados, toda a análise (estimação dos parâmetros, determinação dos limites de controle, monitoramento, etc) deve ser conduzida com base nos dados transformados.