

Resumo do Capítulo 6 com aplicação em R

Universidade Federal da Paraíba - CCEN

Paulo Ricardo Seganfredo Campana - 20210044220

Gabriel de Jesus Pereira - 20200121424

Jéssika Helen de Almeida - 20190041118

10 de novembro de 2023

Gráfico de controle para variáveis

Estamos interessados em monitorar o comportamento de variáveis: características de qualidade que podem ser expressas numericamente, para isso utilizamos as ferramentas dos gráficos de controle de \bar{x} e R e também de \bar{x} e s , gráficos para \bar{x} trazem a informação sobre a média do processo e sua tendência, R atua sobre a amplitude da variável: a diferença entre o maior e o menor elemento de uma amostra, s é uma estatística de variação mais comum na estatística, o desvio padrão, juntas elas são medidas da variação do processo, indicando a qualidade da produção.

Gráficos de Controle para \bar{x} e R

A Base Estatística dos Gráficos

Supondo a princípio que a característica de interesse segue distribuição normal $X \sim \mathcal{N}(\mu, \sigma)$, temos que a distribuição da média aritmética de uma amostra segue também distribuição normal $\bar{X} \sim \mathcal{N}(\mu, \sigma/\sqrt{n})$.

Sendo assim, obtemos diferentes amostras com pequenas quantidades de observações da variável de interesse (recomenda-se em torno de 25 amostras com aproximadamente 5 observações cada), usaremos esse conjunto de dados para a construção dos limites de controle.

A construção do gráfico de controle se baseia na definição de limites de controle inferiores e superiores para a medida de posição \bar{x} e de dispersão R , limites esses que definem pontos de corte para novas amostras coletadas, se em uma nova amostra a estatística calculada para o

gráfico ultrapassar algum dos limites, dizemos que o processo está fora de controle e precisamos investigar alguma causa para a mesma.

Primeiramente para o gráfico de \bar{x} , precisamos de uma medida de posição em que os limites inferiores e superiores de controle serão baseados, como μ diz a respeito da média da característica de interesse, utilizaremos a média das médias amostrais $\bar{\bar{x}}$ por ser um estimador não viesado para o mesmo que faz o uso da separação por amostras.

Para o gráfico de controle de R , uma medida de posição para o mesmo está relacionado a variação da característica, para uma estimativa pontual do desvio padrão populacional podemos usar o desvio padrão amostral $\hat{\sigma}$ ou também a amplitude da amostra $R = x_{\max} - x_{\min}$ como será o caso.

Desse modo temos as seguintes medidas de posição para os gráficos de \bar{x} e R , estes servirão de limite central para os gráficos de controle. Seja uma quantidade m de amostras, cada uma de tamanho n :

$$\bar{\bar{x}} = \frac{1}{m}(\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m) \qquad \bar{R} = \frac{1}{m}(R_1 + R_2 + \dots + R_m)$$

Já para as medidas de dispersão dos parâmetros μ e σ temos que um estimador para σ é $\hat{\sigma} = R/d_2$, dessa forma, seguindo a distribuição normal, sabemos que a média populacional μ tem α de probabilidade de pertencer ao intervalo a seguir, se tomarmos Z_α como 3 para um intervalo de 3 sigma e o estimador para σ acima, chegaremos nos limites de controle para \bar{x} .

$$\begin{aligned} \bar{x} \pm Z_\alpha \frac{\sigma}{\sqrt{n}} \\ \bar{x} \pm 3 \frac{\sigma}{\sqrt{n}} \\ \bar{x} \pm \frac{3}{d_2 \sqrt{n}} \bar{R} \\ \bar{x} \pm A_2 \bar{R} \end{aligned}$$

Sendo assim para \bar{x} e de maneira similar para R , temos as linhas centrais (LC), limites inferiores de controle (LIC) e limites superiores de controle (LSC) para cada gráfico:

$$\begin{array}{ll} \text{LIC} = \bar{\bar{x}} - A_2 \bar{R} & \text{LIC} = \bar{R} D_3 \\ \text{LC} = \bar{\bar{x}} & \text{LC} = \bar{R} \\ \text{LSC} = \bar{\bar{x}} + A_2 \bar{R} & \text{LSC} = \bar{R} D_4 \end{array}$$

Com d_2 , A_2 , D_3 e D_4 constantes que dependem do tamanho da amostra e contidas no apêndice.

A partir de amostras coletadas quando sabemos que o processo está sobre controle, calculamos estes limites e tomamos eles como limites tentativos, então analisamos as medidas de posição de cada amostra para procurar comportamentos não aleatórios ou amostras que fogem dos limites tentativos, se algum destes ocorrerem, isso indica que as amostras não estavam sobre controle quando coletadas e devemos começar o processo novamente até obtermos limites apropriados retirando tais amostras identificadas como fora de controle.

Gráficos Baseados nos Valores de Referência

Quando é possível obter valores de referência para a média μ e o desvio padrão σ do processo, podemos usar estes valores para a criação dos limites de controle de maneira mais simplificada, sem precisar da coleta de amostras sobre controle:

$$LIC = \mu - A\sigma$$

$$LC = \mu$$

$$LSC = \mu + A\sigma$$

$$LIC = D_1\sigma$$

$$LC = d_2\sigma$$

$$LSC = D_2\sigma$$

Interpretação dos Gráficos \bar{x} e R

Um gráfico de controle pode indicar um comportamento fora de controle mesmo que os pontos estejam dentro dos limites de controle. Para isso, os pontos devem exibir comportamento sistemático ou não aleatório. O padrão dos pontos plotados podem fornecer informações valiosas para o diagnóstico do processo, fazendo com que essa informação possa ser utilizada para modificar o processo e reduzir sua variabilidade.

Para a interpretação de padrões do gráfico \bar{x} , deve-se inicialmente determinar se R está, ou não, sob controle. Se os dois gráficos apresentam padrões não aleatórios, a melhor estratégia será primeiramente eliminar as causas atribuíveis do gráfico R . Isso poderá eliminar automaticamente os padrões não aleatórios de \bar{x} .

Existem padrões originados de diversas formas, sendo um deles o “padrão cíclico”. Este padrão no gráfico \bar{x} pode resultar de mudanças ambientais sistemáticas, tais como temperatura, fadiga do operador, flutuação na voltagem ou pressão, rotação de operadores ou máquinas. O gráfico R pode revelar padrões cíclicos por causa do planejamento da manutenção, da fadiga do operador ou do desgaste do equipamento, resultando em variabilidade excessiva.

- **mistura:** Uma mistura é indicada quando os pontos plotados tendem a ficar perto ou levemente fora dos limites de controle, com relativamente poucos pontos perto da linha central.

- **deslocamento no nível do processo:** Pode ser resultado da introdução de novos trabalhadores; mudanças nos métodos, matéria-prima ou máquinas; atenção ou motivação dos operadores. Podem ser observadas melhorias no desempenho do processo em seguida à introdução de um programa gráfico de controle, porque a motivação influencia os operadores.
- **estratificação:** Pode ser causa do cálculo incorreto dos limites de controle.
- **tendência:** Podem ser decorrentes de um desgaste ou deterioração gradual de uma ferramenta ou outro componente crítico do processo.

Gráficos de controle para \bar{x} e s

Como dito anteriormente, podemos usar como estimador para o desvio padrão populacional σ a amplitude da amostra R ou também o desvio padrão amostral s , usaremos este para a construção do gráfico s , que traz a informação da variabilidade da característica de interesse de forma mais direta e mais eficiente para maiores tamanhos de amostra (acima de 10) ou até mesmo quando esse tamanho é variável.

O desvio padrão de cada amostra s é calculado de maneira usual e resumimos a informação de s e cada amostra em \bar{s} , a média dos desvios padrões amostrais.

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

$$\bar{s} = \frac{1}{m} (s_1 + s_2 + \dots + s_m)$$

Porém s não é um estimador não viesado para σ , precisamos de um fator adicional c_4 para sua correção e a variância do estimador é dada também em função de c_4 . De maneira similar a derivação dos limites do gráfico de \bar{x} temos que:

$$\bar{s} \pm Z_\alpha \frac{s}{c_4} \sqrt{1 - c_4^2}$$

$$\bar{s} \pm \frac{3s}{c_4} \sqrt{1 - c_4^2}$$

Resumindo a informação de c_4 em constantes B_3 e B_4 temos os limites de controle do gráfico para s e para \bar{x} utilizando o estimador s ao invés de R :

$$\text{LIC} = \bar{\bar{x}} - A_3 \bar{s}$$

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

$$\text{LSC} = \bar{\bar{x}} + A_3 \bar{s}$$

$$\text{LIC} = \bar{s} B_3$$

$$\text{LC} = \bar{s}$$

$$\text{LSC} = \bar{s} B_4$$

Gráficos de Controle \bar{x} e s com Tamanho de Amostra Variável

A definição dos limites de controle para este caso é feita de maneira similar, porém teremos limites variáveis para cada amostra e utilizaremos de estimadores ponderados para $\bar{\bar{x}}$ e \bar{s} , além de que as constantes A_3 , B_3 e B_4 poderão ser diferentes para cada amostra.

$$\bar{\bar{x}} = \frac{\sum_{i=1}^m n_i \bar{x}_i}{\sum_{i=1}^m n_i} \quad \bar{s} = \sqrt{\frac{\sum_{i=1}^m (n_i - 1) s_i^2}{\sum_{i=1}^m n_i - m}}$$

Outra alternativa para quando temos tamanho das amostras variável é utilizar o tamanho médio ou modal das mesmas quando não há grande variância entre o tamanho das amostras.

Função Característica de Operação

As curvas características descrevem a capacidade dos gráficos \bar{x} e R em detectar deslocamentos na qualidade do processo. A probabilidade de não se detectar esse deslocamento na primeira amostra subsequente é chamado de risco, sendo denotado por β .

$$\beta = \Phi(L - k\sqrt{n}) - \Phi(-L - k\sqrt{n})$$

Para a construção da curva C.O. para o gráfico \bar{x} , é necessário que a plotagem do risco β versus a magnitude do deslocamento que queremos detectar, expressa em unidades de desvio padrão, para vários tamanhos de amostra n . É possível calcular também a probabilidade de que o deslocamento seja detectado na r^a amostra subsequente, sendo calculado simplesmente como

$$\beta^{r-1} (1 - \beta)$$

O número esperado de amostras necessárias para se detectar um deslocamento é calculado como

$$\text{CMS} = \frac{1}{1 - \beta}$$

Para a construção das curvas C.O. para o gráfico R , é necessário empregar a distribuição da amplitude relativa $W = R/\sigma$. A curva CO para R representa a probabilidade de não se detectar o deslocamento para um novo valor de σ .

Aplicação dos gráficos de controle

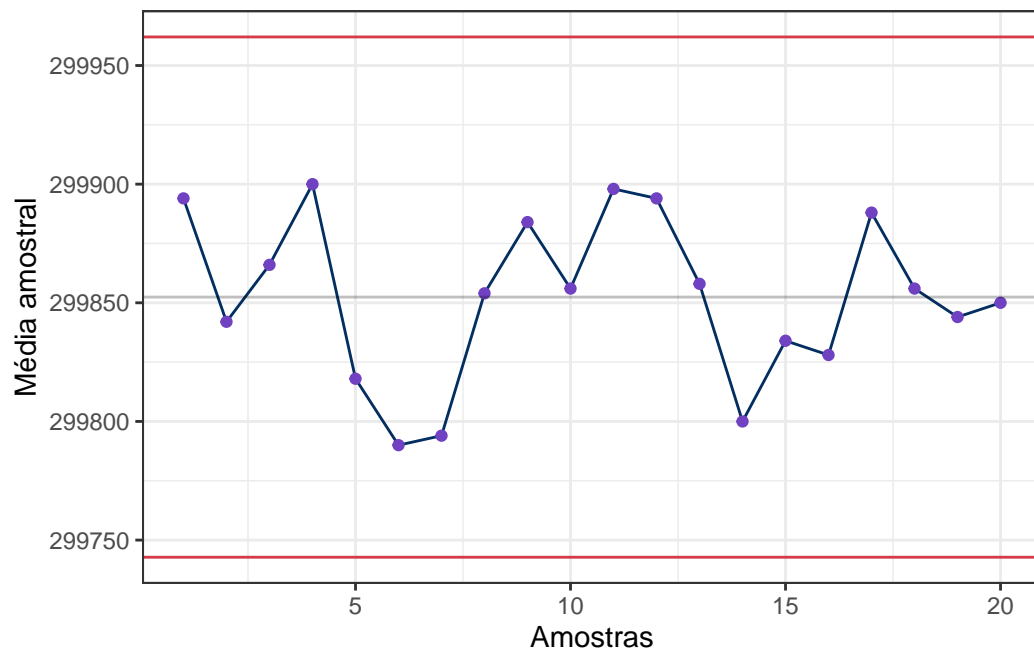
O conjunto de dados `morley` do pacote `datasets` do R contém dados de experimentos de medida da velocidade da luz em quilômetros por segundo, são 20 amostras cada uma com 5 medições, dessa forma, faremos gráficos de controle para \bar{x} e s sobre as medições da velocidade da luz utilizando as fórmulas citadas anteriormente e com as constantes A_3 , B_3 e B_4 do apêndice do livro.

```
library(tidyverse)
data <- morley |>
  mutate(Speed = Speed + 299000) |>
  summarise(
    xbar = mean(Speed),
    s = sd(Speed),
    .by = Run
  )
```

```

A3 <- 1.427
data |>
  ggplot(aes(x = Run, y = xbar)) +
  geom_line(color = "#032e61") +
  geom_point(color = "#7042c2") +
  geom_hline(aes(yintercept = mean(xbar) - A3 * mean(s)), color = "#d63b4a") +
  geom_hline(aes(yintercept = mean(xbar)), alpha = 0.25) +
  geom_hline(aes(yintercept = mean(xbar) + A3 * mean(s)), color = "#d63b4a") +
  labs(x = "Amostrs", y = "Média amostral") +
  theme_bw()

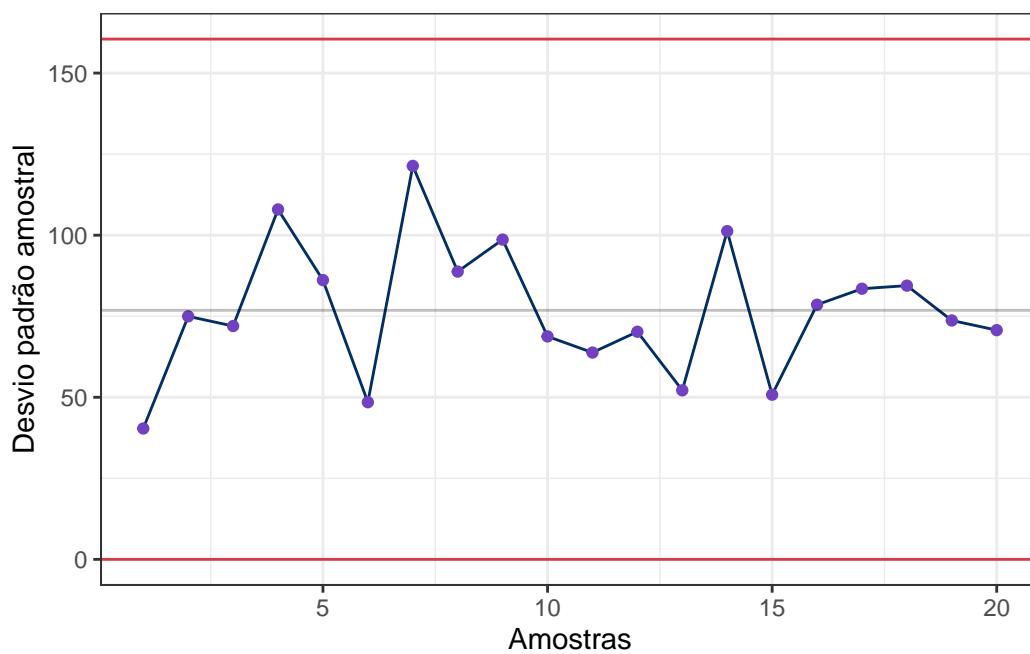
```



```

B3 <- 0
B4 <- 2.089
data |>
  ggplot(aes(x = Run, y = s)) +
  geom_line(color = "#032e61") +
  geom_point(color = "#7042c2") +
  geom_hline(aes(yintercept = mean(s) * B3), color = "#d63b4a") +
  geom_hline(aes(yintercept = mean(s)), alpha = 0.25) +
  geom_hline(aes(yintercept = mean(s) * B4), color = "#d63b4a") +
  labs(x = "Amostras", y = "Desvio padrão amostral") +
  theme_bw()

```



Vemos que ambos os gráficos para \bar{x} e s aparentam estar sob controle pois não há pontos fora dos limites de controle e não identificamos nenhuma característica não aleatória na distribuição das médias e desvios padrões das amostras.