

CADERNOS DO IME – Série Estatística

Universidade do Estado do Rio de Janeiro - UERJ
Rio de Janeiro - RJ - Brasil
ISSN impresso 1413-9022 / ISSN on-line 2317-4536 - v., p.19 - 28, 2014

UMA FUNÇÃO ESCRITA NA LINGUAGEM R PARA OBTER OS PARÂMETROS DO GRÁFICO DE CONTROLE DA AMPLITUDE AMOSTRAL

Nilo Antonio de Souza Sampaio
AEDB – Associação Educacional Dom Bosco
nilo.samp@terra.com.br

Roberto Campos Leoni
Academia Militar das Agulhas Negras
rleon@yahoo.com.br

Rogério Carvalho Mendes Tavora
Academia Militar das Agulhas Negras
rcmqco@ig.com.br

Resumo

O gráfico de controle da amplitude amostral (GC R) é usado para monitorar mudanças na variabilidade de processos. Adotando-se limites de controle 3σ , em média, tem-se um alarme falso a cada 370 amostras no gráfico de controle de \bar{X} . Porém, por ser assimétrica a distribuição amostral da amplitude, o mesmo não ocorre no GC R. Neste artigo, apresentou-se uma função escrita na linguagem R para obter os limites e o poder do GC R a partir da função de probabilidade da amplitude amostral. A abordagem empregada diminui o risco alfa, quando a comparamos com a construção clássica de 3σ para os limites de controle, e detecta melhorias no processo quando ocorre redução da variabilidade. Esta contribuição permite que o usuário possa planejar adequadamente o emprego do GC R.

Palavras-chave: Gráfico de controle da amplitude; Variabilidade; Linguagem R.

1. Introdução

Gráficos de controle por variáveis são amplamente utilizados no monitoramento de variáveis contínuas, destacando-se o gráfico da média, amplitude, variância e desvio-padrão. Uma das mais importantes técnicas de controle e monitoramento da variabilidade de um processo é o gráfico de controle para a amplitude amostral (GC R). Apesar de antigo, o GC R introduzido por Shewhart (1931) ainda é usado e investigado por diversos autores, tais como Amin *et al.* (1997), Chang e Gan (2004), Khoo e Lim (2005), Prajapati e Mahapatra (2009) e Kiani *et al.* (2010).

No gráfico de \bar{X} o uso de limites de controle 3σ , supondo normalidade nos dados, gera um alarme falso, em média, a cada 370 amostras. Porém, o mesmo não acontece no GC R . A distribuição amostral de R é assimétrica, ainda que os dados sejam normais. Assim, limites de controle 3σ no GC R é apenas uma aproximação para o risco que se assume em sinalizar erroneamente um estado de falta de controle na ausência de causas especiais que podem alterar a variabilidade do processo (COSTA *et al.*, 2008).

O *Software R* é um software livre e bastante conhecido na comunidade acadêmica (R DEVELOPMENT CORE TEAM, 2014). Neste artigo, uma função escrita na linguagem **R** foi desenvolvida para calcular o fator de abertura dos limites de controle do GC R , o poder do gráfico e a estatística de desempenho NMA (número médio de amostras até o sinal). A função é baseada no risco α (erro tipo I) para calcular os limites do GC R a partir da distribuição amostral da amplitude relativa. A abordagem empregada é diferente da técnica clássica que utiliza as constantes d_2 e d_3 na construção dos limites de controle (MONTGOMERY, 2008; COSTA *et al.*, 2008).

Este artigo está assim estruturado: na seção 2, o gráfico de controle da amplitude amostral (GC R) e a função distribuição acumulada da amplitude relativa que é aplicada no cálculo para avaliar o desempenho do GC R são apresentados. Na seção 3, estabeleceram-se os riscos associados à avaliação do desempenho do GC R e a estatística de desempenho utilizada na avaliação. Na seção 4, compararam-se os riscos associados ao uso da abordagem tradicional para construção dos limites de controle do GC R com uma abordagem baseada na distribuição amostral da amplitude. Essa abordagem é ilustrada com um exemplo através da aplicação de uma função escrita em linguagem **R** apresentada no apêndice. As conclusões são apresentadas na seção 5.

2. Gráfico de controle da amplitude amostral

GC R é utilizado para monitorar a variabilidade de um processo. Os limites de controle são ajustados em relação à linha média do gráfico, ou seja:

$$LSC_R = \mu_R + k\sigma_R \quad (1)$$

$$LM_R = \mu_R \quad (2)$$

$$LIC_R = \max\{0; \mu_R - k\sigma_R\} \quad (3)$$

sendo: k o fator de abertura dos limites de controle; $\mu_R = d_2\sigma$ e $\sigma_R = d_3\sigma$ a média e o desvio padrão da distribuição da amplitude amostral (d_2 e d_3 são constantes que dependem apenas do tamanho da amostra n) e σ é o desvio padrão do processo em controle. Tippett (1925), Mahoney (1998) e Kao e Ho (2007) apresentam como calcular as constantes $d_2(n)$ e $d_3(n)$ quando a característica de qualidade possui distribuição normal e não normal. Alguns livros relacionados ao controle estatístico de qualidade também apresentam as constantes em forma de tabela (MONTGOMERY, 2008; COSTA *et al.*, 2008).

A distribuição amostral da amplitude $R = X_{\text{MAX}} - X_{\text{MIN}}$ (diferença, em módulo, entre o maior e menor valor de uma amostra) não é normalmente distribuída, possui forma assimétrica, e depende do desvio padrão dos valores da característica de qualidade monitorada. O cálculo da probabilidade de uma amplitude amostral R ser menor (ou maior) que um dado valor R_0 não é trivial, porém é necessário para análise do desempenho do GC R .

A distribuição amostral da variável $W = R/\sigma$, definida como amplitude relativa, depende apenas do tamanho da amostra n e pode ser usada para avaliar o desempenho do GC R (COSTA *et al.*, 2008). Assumindo que a característica de qualidade possui distribuição normal padrão, a função distribuição acumulada de w , $F(w)$ é dada por Lee (2011):

$$F(w) = n \int_{-\infty}^{\infty} \phi(x) [\Phi(x+w) - \Phi(x)]^{n-1} dx \quad (4)$$

sendo: n tamanho da amostra, $\phi(\cdot)$ a função densidade de probabilidade e $\Phi(\cdot)$ a função de distribuição acumulada. Person e Hartley (1942) apresentam valores de $F(w)$.

3. Desempenho do GC R

Testar se um processo está ou não em controle faz com que o uso de um gráfico de controle tenha uma estreita relação com o teste estatístico de hipóteses. Com base em uma amostra, pode-se verificar se há alguma evidência acerca do parâmetro em estudo. A hipótese a ser testada, H_0 , é rejeitada ou não de acordo com os dados amostrais. Em caso de rejeição, evidenciamos uma hipótese alternativa H_1 ; caso contrário, evidencia-se H_0 .

As hipóteses associadas ao GC R são: $H_0: \sigma = \sigma_0$ (processo em controle) e $H_1: \sigma \neq \sigma_0$ (processo fora de controle). Na hipótese H_1 , o desvio padrão do processo sofre aumento de ordem λ , ou seja,

$$\sigma_1 = \lambda \sigma_0 \quad (5)$$

No controle de um processo cada amostra é testada para verificação da hipótese de interesse. Os limites de controle ($LIC_R; LSC_R$) equivalem a região de aceitação do teste. Dentro dos limites não há evidência de H_1 ; caso contrário, há.

Como em um teste de hipóteses, podemos obter os riscos α e β associados ao GC R :

- Risco α (erro tipo I) - equivale a sinalizar erroneamente um estado de falta de controle na ausência de causas especiais ou perturbações no processo;
- Risco β (erro tipo II) - equivale a não sinalizar o estado de falta de controle na presença de causas especiais ou perturbações no processo.

Os riscos α e β são representados respectivamente pelas expressões de probabilidade condicional:

$$\alpha = 1 - \Pr(LIC_R \leq R \leq LSC_R | n = n_0; \sigma = \sigma_0) \quad (6)$$

$$\beta = \Pr(LIC_R \leq R \leq LSC_R | \sigma \neq \sigma_0) \quad (7)$$

O poder do gráfico de controle (Pd) é a probabilidade complementar de β :

$$Pd = 1 - \beta \quad (8)$$

O estabelecimento dos limites de controle é uma decisão crítica, pois afeta os riscos citados. Alargar os limites diminui o risco α , porém aumenta o risco β e vice-versa.

Com o processo isento de causas especiais, o ideal é que todos os pontos da estatística R se situem dentro dos limites de controle. Entretanto, existe o risco α de surgir um ponto na região de ação do gráfico, gerando um alarme falso.

O risco α depende da distribuição da variável aleatória R . Para limites iguais a (1) e (3), a probabilidade de um ponto R cair na região de ação do gráfico é dada por:

$$\alpha = 1 - \Pr(|\max\{0, (d_2 - kd_3)\} \leq w \leq (d_2 + kd_3) | n = n_0) \quad (9)$$

O risco de se assumir que o processo está fora de controle, na ausência de causa especial, ou seja, alarme falso, depende apenas n .

Dada a ocorrência de um aumento no desvio padrão, a probabilidade de um ponto da estatística R acima ou abaixo dos limites de controle é:

$$Pd = \Pr\left(W \geq \frac{d_2 + kd_3}{\lambda} | n = n_0\right) \quad (10)$$

Os riscos α e β são utilizados para avaliar o desempenho do GC R . Quando o processo está fora de controle, a velocidade de detecção é medida pelo número de amostras que o gráfico requer para sinalizar a causa especial (NAS).

$$\Pr(NAS = d) = p(1 - p)^{d-1}, d = 1, 2, \dots \quad (11)$$

O valor esperado do NAS , conhecido como NMA , é um índice usado para avaliar o desempenho do gráfico de controle e segue uma distribuição geométrica de parâmetro p . Como a média da distribuição geométrica de parâmetro p é igual a $1/p$, têm-se:

$$NMA = \sum_{i=1}^{\infty} d \Pr(NAS = d) = 1/p \quad (12)$$

Se a hipótese H_0 for verdadeira ($\lambda = 1$), então $p = \alpha$. Se a hipótese H_1 for verdadeira ($\lambda > 1$), então $p = Pd$.

Com o processo em controle é desejável que o NMA seja grande o bastante para que alarmes falsos demorem a ocorrer. Quando uma causa especial desloca a média do processo, ou seja, $\sigma \neq \sigma_0$, o NMA deve ser pequeno de modo que a detecção da causa especial seja rápida. Com o processo em controle, o NMA é chamado de NMA_0 .

4. Construção do GC R

Nesta seção, apresenta-se uma forma alternativa para construção do GC R . A técnica consiste em utilizar limites de controle considerando um valor pré-determinado para o risco α (alarme falso). Para tal, consideramos a distribuição amostral da amplitude relativa (w). Por exemplo, fixando o valor de α , calculam-se as probabilidades $P[W < w_{\text{inf}}] = \alpha/2$ e $P[W > w_{\text{sup}}] = \alpha/2$. As equações (1) e (3), tornam-se:

$$LSC_R = w_{\text{sup}} \sigma \quad (13)$$

$$LIC_R = w_{\text{inf}} \sigma \quad (14)$$

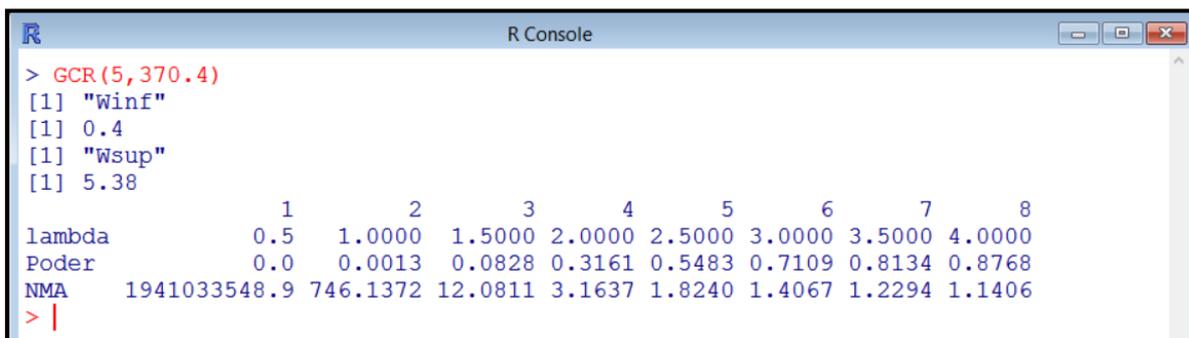
Duas vantagens são verificadas com essa abordagem: o resultado de (14) não pode ser negativo, permitindo detectar melhoria no processo, ou seja, detectar se há redução de variabilidade. A segunda vantagem se refere ao uso de $k=3\sigma$ na equação (3), essa abordagem tradicional leva a valores menores que o LIC_R calculado em (14), ou seja, o risco α é maior na abordagem tradicional (COSTA *et al.*, 2008).

No apêndice, há uma função escrita na linguagem **R** que é utilizada para avaliar o desempenho do GC R considerando as expressões (13) e (14).

4.1. Exemplo

Costa *et al.* (2008) apresenta um processo de produção de leite em que 25 amostras de tamanho $n=5$ são selecionadas para a construção do GC R . Com o processo isento de causas especiais, os limites para o GC R foram estimados utilizando a abordagem tradicional ($k=3$): $LSC_R = (d_2 + 3d_3)\hat{\sigma}_0 = 22,20$; $LM_R = 10,5$; $LIC_R = (d_2 - 3d_3)\hat{\sigma}_0 = -1,20 \Rightarrow LIC_R = 0,00$. Observe que o somatório $d_2 + 3d_3 = 2,326 + 3(0,864) = 4,918$ e $d_2 - 3d_3 = 2,326 - 3(0,864) = -0,266$. O desvio padrão do processo foi estimado pela equação: $\hat{\sigma}_0 = \bar{R} / d_2 = 10,5 / 2,326 = 4,514$.

Tradicionalmente, o fator de abertura $k=3$ é usado no gráfico de controle para que o $NMA_0=370,4$, ou seja, o risco $\alpha = 0,0027$. Porém, esse resultado não é válido para o GC R , pois a distribuição amostral da amplitude não é normal. Aplicando a função apresentada no apêndice, obtêm-se os resultados da Figura 1.

Figura 1. Resultado da aplicação da função GCR (n, NMA_0) apresentada no apêndice.


```

R Console
> GCR(5,370.4)
[1] "Winf"
[1] 0.4
[1] "Wsup"
[1] 5.38

lambda      1      2      3      4      5      6      7      8
Poder      0.0  0.0013 0.0828 0.3161 0.5483 0.7109 0.8134 0.8768
NMA      1941033548.9 746.1372 12.0811 3.1637 1.8240 1.4067 1.2294 1.1406
> |

```

Fonte: os próprios autores.

A Figura 1 fornece os limites: $w_{\text{inf}}=0,4$ e $w_{\text{sup}}=5,38$. De acordo com o exemplo, teríamos $LSC_R = 5,38\hat{\sigma}_0$ e $LIC_R = 0,4\hat{\sigma}_0$. Usar $d_2 + 3d_3=4,918$ não é adequado para o risco $\alpha=0,0027$, pois $w_{\text{sup}}=5,38$ é o fator de abertura que corresponde ao $NMA_0=370,4$.

Observa-se, também na Figura 1 o poder e o NMA do GC R . Os desvios são representados por λ , ou seja, $\sigma_1 = \lambda\sigma_0$. Por exemplo: se $\lambda = 3$, o gráfico demora em média 12,0811 amostras para sinalizar uma causa especial que provoca mudança no desvio padrão do processo.

A função GCR (n, NMA_0) também pode usar o risco α como parâmetro de entrada. Para tal, basta rodar a função invertendo o NMA_0 , ou seja, GCR ($n, 1/\alpha$). Por exemplo: GCR (5, 1/0.002699784) fornece o mesmo resultado que GCR (5, 370.4).

5. Conclusão

Apresentou-se nesse artigo o gráfico de controle da amplitude amostral utilizado para monitorar a variabilidade de um processo. Avaliou-se o poder do gráfico e do risco α com a aplicação de uma função desenvolvida para uso no *software* R. Comparou-se o risco associado ao uso da abordagem tradicional, em que o fator de abertura (k) dos limites do gráfico é igual a 3, com uma abordagem baseada na distribuição amostral da amplitude relativa. Duas vantagens são destacadas se o GC R é projetado com a função (para uso no *software* R) aqui disponibilizada: a diminuição do risco α e a possibilidade de detectar melhorias no processo com a redução da variabilidade.

Referências

- AMIN, R. W.; SCHMID, W.; FRANK, O. The Effects of Autocorrelation on the R-Chart and the S^2 -Chart. **The Indian Journal of Statistics**, Series B (1960-2002), v. 59, n. 2, p. 229-255. 1997.
- CHANG, T. C.; GAN, F. F. Shewhart Chart for Monitoring the Variance Components. **Journal of Quality Technology**, v. 36. n. 3, p. 293-308. 2004.
- COSTA, A. F. B.; EPPRECHT, E. K.; CARPINETTI; L. C. R. **Controle Estatístico de Qualidade**. 2a. ed., São Paulo: Editora Atlas, 334 p. 2008.
- KAO, S. C.; HO, C. Robustness of R-chart to Non Normality. **Communication in Statistics, Simulation Computation**, v. 36, p. 1089–1098. 2007.
- KHOO, M. B. C.; LIM, E. G. An Improved R (Range) Control Chart for Monitoring the Process Variance. **Quality and Reliability Engineering International**, v. 21, p. 43–50. 2005.
- KIANI, M.; PANARETOS, J.; PSARAKIS, S. A New Procedure for Monitoring the Range and Standard Deviation of a Quality Characteristic. **Quality & Quantity**, v. 44, p. 1–24. 2010.
- LEE, PEI-HSI. Adaptive R Charts with Variable Parameters. **Computational Statistics and Data Analysis**, v. 55, p. 2003–2010. 2011
- MAHONEY, J. F. The Influence of Parent Population Distribution on d_2 . **IIE Transactions**, v.30, n.6, p. 563–569. 1998.
- MONTGOMERY, D. C. **Statistical Quality Control: A Modern Introduction**, 6th ed. John Wiley & Sons Inc., New York. 2008.
- PERSON, E. S.; HARTLEY, H. O. The Probability Integral of the Range in Samples of n Observations from a Normal Population. **Biometrika**, v. 32, p. 301–310. 1942.
- PRAJAPATI, D. R.; MAHAPATRA, P. B. A Simple and Effective R Chart to Monitor the Process Variance. **International Journal of Quality & Reliability Management**, v. 26, n. 5, p. 497-512. 2009.
- R DEVELOPMENT CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>, 2014.
- SHEWHART, W. A. **Economic Control of Quality of Manufactured Product**. D. Van Nostrand Co., New York. 1931.
- TIPPETT, L. H. C. On the Extreme Individuals and the Range of Samples taken from a Normal Population. **Biometrika**, v. 17, p. 364–386. 1925.

A FUNCTION WRITTEN IN THE R LANGUAGE FOR THE PARAMETERS OF THE CONTROL CHART OF SAMPLE AMPLITUDE

Abstract

The control chart of sample range (CCSR) is used to monitor changes in process variability. Adopting 3σ limits, on average, it has a false alarm every 370 samples in the control \bar{X} chart. However, because the amplitude sampling distribution is asymmetric, the same does not occur in CCSR. In this article, it has been presented a function written in R language for obtaining the limits and the power of CCSR from probability function of sample amplitude. The used approach reduces the alpha risk, when compared with the classical construction of 3σ for control limits, and detects process improvements when reducing variability occurs. This contribution allows that the user properly plans the use of CCSR.

Keywords: R control chart; Variability; R Language.

APÊNDICE – Código fonte para avaliar o desempenho do GC R

código da função → copie o código a seguir no ambiente R

```

rm(list=ls(all=TRUE))
GCR <- function (n,nma0) {
Fnw=0; wi=0; alfa=1/nma0
while (Fnw<=alfa/2) {
wi=wi+0.01
integrando<- function(z) {((1/(sqrt(2*pi)))*)exp(-(z^2)/2))*(pnorm(z+wi)-pnorm(z))^(n-
1)}
Fnw=integrate(integrando, lower=-Inf, upper=Inf)
Fnw=n*as.numeric(Fnw[1])
}
Fnw=0; ws=0
while (Fnw<=1-alfa/2) {
ws=ws+0.01
integrando<- function(z) {((1/(sqrt(2*pi)))*)exp(-(z^2)/2))*(pnorm(z+ws)-pnorm(z))^(n-
1)}
Fnw=integrate(integrando, lower=-Inf, upper=Inf)
Fnw=n*as.numeric(Fnw[1])
}
print('Winf');print(wi);print('Wsup');print(ws)
# calculo do poder do GC R
lambda=1.5;i=1
Poder=matrix(c(NA),3,8)
while (lambda<=5) {
integrando<- function(z) {((1/(sqrt(2*pi)))*)exp(-(z^2)/2))*(pnorm(z+ws/lambda)-
pnorm(z))^(n-1)}
Fnw=integrate(integrando, lower=-Inf, upper=Inf) #list
Fnw=1-n*as.numeric(Fnw[1]) # poder Pr[W>Ws|n=n0]
Poder[1,i] <-lambda
Poder[2,i] <-Fnw #poder
Poder[3,i] <-1/Fnw #NMA
i=i+1; lambda=lambda+0.5
}
colnames(Poder) <- c("1", "2", "3","4", "5", "6","7", "8")
rownames(Poder) <- c("lambda", "Poder", "NMA")
round(Poder,4)
}

```

Exemplo de execução da função → execute a função escolhendo (n,NMA₀)

GCR(5,370.4)