

O PROGRAMA SCAP PARA MONITORAMENTO DE PROCESSOS AUTOCORRELACIONADOS E INSPEÇÃO POR ATRIBUTOS

SUELI APARECIDA MINGOTI, JÚLIA PINTO DE CARVALHO

(DEPARTAMENTO DE ESTATÍSTICA-UFGM)

e-mail: sueli@est.ufmg.br , july@est.ufmg.br

RESUMO

Neste artigo apresentamos o módulo *AI (Attribute Inspection)* do programa computacional *SCAP (Statistical Control for Autocorrelated Processes)* que implementa os modelos Binomial Generalizado e de Cadeia de Markov utilizados em controle de processos autocorrelacionados. O programa funciona de modo interativo, acoplado ao software estatístico *Minitab for Windows*, e permite a estimação da correlação através dos estimadores de Madsen (1993), Bhat & Lal (1990) e Mingoti (2002), além da construção de gráficos de controle para o número e proporção de "não conformes" do processo.

1 Introdução

Alguns processos são monitorados através da observação do número ou da proporção de itens "não-conformes". Neste caso, o "*número total de defeituosos*" é visto como tendo uma distribuição Binomial e a construção de gráficos de controle de Shewhart é feita via aproximação pela distribuição Normal, considerando-se independência entre as unidades amostrais em relação ao estado de ser ou não "conforme" (Montgomery,1991). No entanto, a correlação aparece com uma frequência bem maior do que se pensa (Alwan & Roberts,1995), o que exige modelos estatísticos que permitam a introdução da correlação. Em Bhat & Lal (1990) os limites de controle para monitoramento de atributos de processos correlacionados foram desenvolvidos usando a teoria de Cadeia de Markov (Broadbent,1996). Em Lai, Govindaraju & Xie (1998) o modelo Binomial Correlacionado de Madsen (1993), no qual a correlação entre as unidades amostrais é suposta ser igual para todas as unidades, é proposto. No modelo de Cadeias de Markov presume-se a inspeção serial das unidades do processo, e portanto tem-se que manter a informação sobre o estado individual de cada item na sequência exata inspecionada. Já no modelo Binomial Generalizado necessita-se apenas do número total de "não-conformes" de cada amostra observada do processo. Em ambas as metodologias, a estimação do coeficiente de correlação tem um papel fundamental. Bhat & Lal (1990) mostram como estimar o coeficiente de correlação no caso de Cadeias de Markov. Para o modelo Binomial Generalizado o estimador mais conhecido é o de Madsen (1993). Novos estimadores paramétricos e não-paramétricos foram propostos recentemente por Mingoti (2002). Alguns trabalhos interessantes sobre processos correlacionados são: Nelson (1994), Stimson & Mastrangelo (1996) e Mingoti & Fidelis (2001).

2 Modelo de Cadeia de Markov

No modelo de Cadeia de Markov (Bhat & Lal, 1990) a inspeção de itens é feita de forma contínua de modo a ter-se informações necessárias para a estimação das probabilidades de transição da Cadeia. Basicamente, tem a seguinte situação: seja, $\{Y_n, n = 0,1,2,\dots\}$ um Cadeia de Markov com dois estados: 0 (item é conforme) e 1 (item é não conforme) sendo que Y_n caracteriza o estado do n -ésimo item inspecionado. As probabilidades de transição a um passo e a n passos são dadas respectivamente por:

$$p_{ij} = Prob[Y_{n+1} = j / Y_n = i] \quad n = 0, 1, 2, \dots; i, j = 0, 1 \quad (1)$$

$$p_{ij}^{(n)} = Prob[Y_n = j / Y_0 = i] \quad n = 1, 2, \dots; i, j = 0, 1 \quad (2)$$

sendo Y_0 o estado inicial da Cadeia. A matriz de probabilidades de transição a um passo é:

$$P = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 1 - (p(1-\rho)) & p(1-\rho) \\ (1-p)(1-\rho) & p + \rho(1-p) \end{bmatrix} \quad (3)$$

onde $p = \pi_1$ é a probabilidade limite da Cadeia, isto é, $\pi_j = \lim(n \rightarrow \infty) p_{ij}^{(n)}$, $j = 0, 1$, e é

entendida como a probabilidade de que num grande volume de produção, um item do processo seja defeituoso (p), ou seja é a fração de defeituosos do processo. Em Bhat & Lal (1990) é mostrado que sob esta condição de equilíbrio da Cadeia, a correlação de ordem j é dada por: $\rho_j = corr(Y_n, Y_{n+j}) = \rho^j$, $j = 1, 2, \dots$, onde ρ é a correlação de ordem 1 do processo.

Portanto, a correlação entre unidades amostrais decresce exponencialmente à medida que a distância entre as unidades amostrais aumenta na linha de produção, algo que vai de encontro ao que se espera em problemas práticos. Para ser uma Cadeia admissível os parâmetros p e ρ precisam satisfazer as condições:

$$1 - \min\left[\frac{1}{p}; \frac{1}{1-p}\right] < \rho < 1 \quad \text{ou equivalentemente,} \quad \max\left[0, \frac{-\rho}{1-\rho}\right] < p < \min\left[1, \frac{1}{1-\rho}\right] \quad (4)$$

Quando o coeficiente de correlação é igual a zero temos a situação de ensaios de Bernoulli independentes. O modelo de Markov permite que haja correlações negativas, no entanto não permite que haja correlação igual a 1 ou -1. A distribuição exata da variável X_n , número de defeituosos observados em n itens inspecionados é complicada como pode ser observado em Bhat & Lal (1990, 1998). No entanto, pode ser mostrado que assintoticamente X_n tem distribuição aproximadamente Normal com média e variância dadas por:

$$\begin{aligned} E(X_n) &= n p \\ Var(X_n) &= n p(1-p) + 2 p(1-p) \frac{\rho}{1-\rho} \left(n - \frac{1-\rho^n}{1-\rho}\right) \end{aligned} \quad (5)$$

Deste modo, mesmo na presença de correlação o usuário, ao optar pela Cadeia de Markov, ainda tem a opção de se manter dentro da classe da distribuição normal usando os gráficos de controle de Shewhart (Montgomery, 1991).

3 Modelo Binomial Generalizado

Suponha que uma amostra de n itens do processo tenha sido inspecionada. Para cada item define-se a variável: $X_i = 1$ se o item for "não conforme" e $X_i = 0$ se o item for "conforme", e supõe-se que os itens sejam correlacionados de forma que: $corr(X_i, X_j) = \rho$, $i \neq j$, $\rho > 0$. Seja Z_n o número total de itens não conformes observados na amostra de tamanho n do processo. Se considerarmos que a sequência $\{X_i, i=1, 2, \dots\}$ é permutável (Berger, 1985), a

distribuição de probabilidades exata da variável Z_n é a Binomial Generalizada proposta por Madsen (1993), isto é,

$$P[Z_n = k] = \begin{cases} \rho(1-p) + (1-\rho)(1-p)^n, & \text{se } k = 0 \\ (1-\rho) \binom{n}{k} p^k (1-p)^{n-k}, & \text{se } 1 \leq k \leq n-1 \\ \rho p + (1-\rho)p^n, & \text{se } k = n \end{cases} \quad (6)$$

com esperança e variância de Z_n dadas por:

$$\begin{aligned} E(Z_n) &= np \\ \text{Var}(Z_n) &= np(1-p) + n(n-1)p(1-p)\rho = np(1-p)[1 + (n-1)\rho] \end{aligned} \quad (7)$$

A distribuição Binomial Generalizada não pode ser aproximada pela distribuição Normal, a menos que $\rho = 0$. Deste modo, o cálculo de probabilidades exigirá o uso da distribuição exata (6) ou da aproximação de Poisson (Mingoti, 2002).

4 Estimação de Parâmetros

4.1 Cadeia de Markov

Os parâmetros do processo são estimados quando o mesmo está sob controle. Estimadores de máxima verossimilhança de p e ρ no caso da Cadeia de Markov são dados por (Bhat & Lal, 1990):

$$\hat{p}_{ij} = \frac{n_{ij}}{n_{i0} + n_{i1}}, \quad i, j = 0, 1; \quad \hat{p} = \frac{\hat{p}_{01}}{\hat{p}_{01} + \hat{p}_{10}}; \quad \hat{\rho} = 1 - (\hat{p}_{01} + \hat{p}_{10}) \quad (8)$$

onde n_{ij} é o número de transições do estado i para o estado j observado na amostra, $i, j = 0, 1$. A distribuição de $\hat{\rho}$ é aproximadamente normal (Lai, Xie & Govindaraju, 2000).

4.2 Modelo Binomial Generalizado

Para a estimação dos parâmetros do modelo Binomial Generalizado é necessário ter-se m amostras aleatórias de tamanho n da variável Z_n . Denotando-se os valores observados de Z_n para cada amostra i por $Z_{n,i}$, $i=1, 2, \dots, m$, $m > 2$, o estimador do coeficiente de correlação ρ sugerido por Madsen (1993), construído pelo método dos momentos, é dado por:

$$\hat{\rho} = \frac{[s^2 - np(1-p)]}{[n(n-1)p(1-p)]} \quad \text{onde} \quad s^2 = \frac{\sum_{i=1}^m (Z_{n,i} - \bar{Z}_n)^2}{m} \quad \text{sendo} \quad \bar{Z}_n = \frac{\sum_{i=1}^m Z_{n,i}}{m} \quad (9)$$

Caso a fração de "não conformes" p não esteja previamente especificada, ela poderá ser estimada usando a fração amostral de "não conformes" observada dentre as m amostras. A proposta original de Madsen sugere que se use a variância amostral s^2 calculada com o divisor m . No entanto, uma outra alternativa é utilizar-se o estimador não viciado cujo divisor

é $(m-1)$, o qual chamamos de Madsen modificado. Se as m amostras não tiverem o mesmo tamanho, o estimador do coeficiente de correlação terá que ser redefinido de acordo com Madsen (1993). Para amostras contendo n valores define-se como estimativa $\hat{\rho} = 1$.

4.3 Estimadores Alternativos para o Coeficiente de Correlação

Vários estimadores alternativos para o coeficiente de correlação foram propostos por Mingoti (2002). Vamos descrever brevemente apenas o estimador Multinomial e os Bayesianos. Para informações sobre outros estimadores ver Mingoti (2002).

4.3.1 Estimador Multinomial

Suponha que tenhamos m amostras independentes da distribuição Binomial Generalizada $\text{Bin}(n, p, \rho)$. Neste caso a variável aleatória Y_k definida como o número de amostras que apresentam k itens "não conformes", tem distribuição multinomial com parâmetros p_k , dados segundo o modelo da Binomial Generalizado. Usando este modelo pode-se construir o estimador mostrado na equação (11) chamado de Mingoti I (Mingoti (2002)).

$$\hat{\rho}_I = \frac{\hat{p}_0 + \hat{p}_n - (1-p)^n - p^n}{\{1 - [(1-p)^n + p^n]\}}, \quad 0 < p < 1 \quad (11)$$

onde, $\hat{p}_k = \frac{f_k}{m}$, f_k o número de amostras, dentre as m observadas, que apresentam k itens não conformes, $k=0, n, m \geq 2$. Este estimador pode ser estendido para o caso em que se tem observações sequenciais do tipo "0" e "1" e poderá ser usado para de Cadeias de Markov quando a correlação for maior que zero (ver estimador Mingoti II em Mingoti 2002).

4.3.2 Estimadores Bayesianos I

Estimadores Bayesianos (Mingoti,2002) podem ser construídos considerando-se ρ como tendo distribuição à priori $\text{Beta}(\alpha, \beta)$, $\alpha > 0$, $\beta > 0$, ou $\text{Uniforme}(a, b)$, $a \geq 0$, $b \leq 1$, e supondo o modelo Binomial Generalizado dado em (6). Para a distribuição à priori Beta, calculando-se a esperança condicional de ρ dado os valores de Z_n e p , obtém-se o estimador (12) chamado de Estimador Bayesiano Mingoti III em Mingoti (2002).

$$\begin{aligned} E[\rho / Z_n = 0] &= \frac{\alpha [(1-p)(\alpha+1) + \beta(1-p)^n]}{[\alpha(1-p) + \beta(1-p)^n](\alpha + \beta + 1)} \\ E[\rho / Z_n = k] &= \frac{\alpha}{\alpha + \beta + 1}, \quad 1 \leq k \leq n-1 \\ E[\rho / Z_n = n] &= \frac{\alpha [p(\alpha+1) + \beta p^n]}{(\alpha + \beta + 1)[\alpha p + p^n \beta]} \end{aligned} \quad (12)$$

Os valores dos parâmetros (α, β) precisam ser determinados à priori pelo pesquisador, de acordo com o conhecimento prévio que tenha sobre o comportamento do processo. Para maiores detalhes sobre o estimador Bayesiano com distribuição à priori Uniforme ver estimador Mingoti IV em Mingoti (2002). Os estimadores Bayesianos foram desenvolvidos

com base no conhecimento de apenas uma amostra da Binomial Generalizada. Se tivermos $m > 2$ amostras independentes desta distribuição, podemos estimar ρ para cada amostra de tamanho n e no final calcular a média das estimativas encontradas, tendo-se assim uma estimativa de ρ aproveitando-se a informação de todas as m amostras. O interessante dos estimadores Bayesianos é que permitem a estimação de ρ tendo-se apenas uma amostra da Binomial Generalizada, algo que não é possível no estimador de Madsen (1993). No entanto, o pesquisador precisa especificar os valores dos parâmetros (α, β) ou (a, b) das distribuições à priori. Caso o valor de p não esteja pré-especificado ele poderá ser estimado pela fração amostral observada de itens não conformes.

3 Programa SCAP para Monitoramento de Processos Autocorrelacionados e Inspeção por Atributos.

O módulo *AI (Attribute Inspection)* do programa computacional *SCAP* (Mingoti & Carvalho, 2001) foi desenvolvido acoplado ao *software* estatístico *Minitab for Windows*. O usuário pode estimar os parâmetros do processo (p, ρ) tanto no caso de Cadeias de Markov quanto no da Binomial Generalizada para dados sequencias e agrupados. Todos os estimadores descritos neste artigo e apresentados em Mingoti (2002) foram implementados, além de gráficos de controle para a proporção e número de "não conformes" do processo. As estimativas para o coeficiente de correlação podem ser obtidas tanto para o caso em que todas as m amostras têm o mesmo tamanho n de observações quanto para o caso em que as amostras têm tamanhos diferentes. A proporção de defeitos pode ser estimada no programa ou pré-determinada pelo usuário. O programa *SCAP* tem uma boa velocidade de processamento. A seguir vamos apresentar um pequeno exemplo de aplicação usando o programa *SCAP*.

Exemplo: No conjunto de dados do Quadro 1 tem-se $m=10$ amostras de tamanho $n=50$. Para usar o programa computacional *SCAP* o usuário deverá ter o seu arquivo no formato *Minitab for Windows* (ou compatível) de modo que a primeira coluna represente a identificação da amostra e a segunda represente o número de defeitos. O usuário poderá estimar a proporção de itens "não conformes" e o coeficiente de correlação considerando os estimadores construídos de acordo com o modelo Binomial Generalizado.

Quadro 1: Dados-Exemplo

Amostra	NºDefeituosos
1	0
2	2
3	25
4	10
5	0
6	2
7	5
8	0
9	3
10	3

Fornecendo a informação do valor de m e de n ao programa e indicando que os dados de número de defeitos são agrupados e não individuais, o usuário poderá estimar a proporção de defeitos que neste caso seria 0,10. Para estimar o coeficiente de correlação o usuário deverá escolher qual estimador de correlação deseja utilizar. O programa abre uma janela como

mostrado na Figura 1. Se o usuário optasse, por exemplo, pelo estimador proposto por Madsen (1993), o programa forneceria duas estimativas: a primeira, ($r\hat{0}1$), obtida utilizando-se a variância amostral com o divisor ($m-1$) (Madsen modificado), e a segunda, ($r\hat{0}2$), obtida utilizando-se a proposta original de Madsen, isto é calculando-se a variância amostral com o divisor m . O programa *SCAP* dá uma mensagem de aviso e fornece as estimativas de ρ como mostrados na Figura 2. Se tivéssemos optado pelo estimador Multinomial (Mingoti I) teríamos uma estimativa igual a 0,296374.

```
Qual é o método de estimação do coeficiente de correlação?
Opções:
1. Método dos momentos (Madsen);
2. Método de Cadeia de Markov (somente para dados individuais);
3. Método Não-Paramétrico (somente para dados individuais);
4. Método Multinomial (Mingoti I);
5. Método Mingoti II (somente para dados individuais);
6. Método Bayesiano I (Mingoti III);
7. Método Bayesiano II (Mingoti IV).
Digite 1,2,3,4,5,6 ou 7.
```

Figura 1: Opções do *SCAP* para estimação do coeficiente de correlação.

```
!!! OBSERVAÇÃO !!! Considerando o método escolhido, será estimado:
"rô1"= com variância obtida pelo divisor (m-1) (Madsen Modificado), e
"rô2"= com variância obtida pelo divisor(m) (proposto por Madsen,1993);
```

Data Display

```
rô1      0.244646
```

Data Display

```
rô2      0.218141
```

Figura 2: Estimativas da correlação usando-se Madsen original ($r\hat{0}2$) e modificado ($r\hat{0}1$).

Para utilizar os estimadores Bayesianos o usuário precisará indicar os valores dos parâmetros da distribuição à priori de ρ . Por exemplo, se o usuário escolher a opção 6 na Figura 1 (Método Bayesiano I) e uma distribuição Beta com parâmetros $\alpha=\beta=1$ tem-se uma estimativa para ρ igual a 0,383333. No entanto, se o usuário escolher a opção 7 na Figura 1 (Método Bayesiano II) com uma Uniforme com parâmetros $a=0$ e $b=0,5$ tem-se uma estimativa de ρ igual a 0,254993 (ver Figura 3). O interessante neste exemplo, é que os dados do Quadro 1 foram simulados a partir de uma Binomial Generalizada com $n=50$, $\rho = 0,25$ e $p=0,10$. Esta informação é relevante no momento de entendermos o comportamento dos estimadores construídos. O que se percebe neste pequeno exemplo, é que o estimador de Madsen modificado apresenta melhores resultados que o estimador original de Madsen (1993), o estimador Multinomial está superestimando um pouco o verdadeiro valor de ρ , e os

estimadores Bayesianos são úteis se o investigador tiver alguma informação do processo que possa ser utilizada na escolha dos parâmetros da distribuição à priori para ρ . Por exemplo, a escolha de $\alpha=\beta=1$ equivale a uma Uniforme no intervalo $a=0$ e $b=1$, uma distribuição não informativa sobre o comportamento de ρ . No entanto, quando se utiliza uma informação mais refinada, caso no qual escolhe-se a Uniforme no intervalo $a=0$ e $b=0,5$, o resultado da estimação melhora consideravelmente, sendo praticamente igual ao valor verdadeiro da correlação teórica do processo. Este é o papel claro da Estatística Bayesiana, incorporar informações adicionais sobre os parâmetros do processo para melhorar as estimativas finais. O usuário poderá solicitar também um gráfico de controle como o da Figura 4, gerado pelo programa considerando o modelo Binomial Generalizado para os dados do Quadro 1, no caso em que a correlação foi estimada via Madsen original ($\hat{\rho}_2$).

Embora o exemplo tratado neste artigo tenha sido para dados agrupados o programa SCAP funciona de forma similar para tratamento de dados via Cadeia de Markov.

Considerando que o coeficiente de correlação possui uma distribuição Beta com parâmetros alfa e beta, forneça à priori:

O valor da constante alfa (alfa > 0):

DATA> 1

O valor da constante beta (beta > 0):

DATA> 1

Data Display

0.383333

Considerando que o coeficiente de correlação possui uma distribuição Uniforme com parâmetros a e b, forneça à priori:

O valor da constante a ($0 \leq a \leq 1$):

DATA> 0

O valor da constante b ($0 \leq b \leq 1$)

DATA> 0.5

Data Display

0.2549993

Figura 3: Estimativas da correlação usando-se os estimadores Bayesianos.

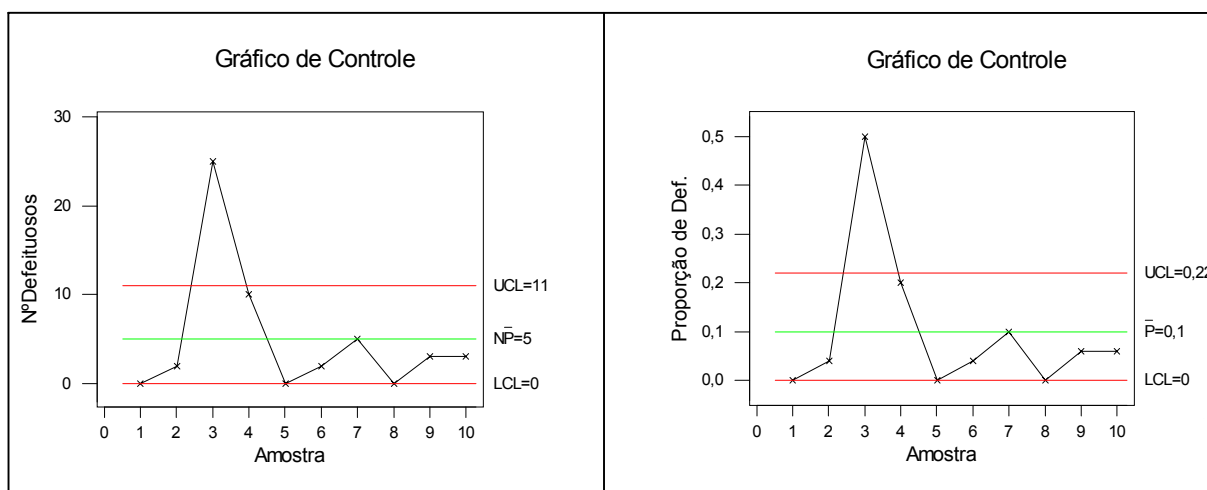


Figura 4: Gráfico de controle para o número e proporção de defeituosos via Distribuição Binomial Generalizada exata.

5 Considerações Finais

O programa *SCAP* estará à disposição para download com manual completo, sem custo para o usuário, no início de 2003 na homepage www.est.ufmg.br.

Agradecimentos

Agradecemos ao CNPq e FAPEMIG pelo apoio financeiro que possibilitou a execução deste trabalho.

Referências Bibliográficas

1. Alwan, L.C., Roberts, H. V. The problem of missplaced control limits. *Applied Statistics, Journal of Royal Statistical Society, Series C*, 44, 3, 269-278,1995.
2. Berger, J. O. *Statistical decision theory and bayesian analysis*. 2nd, New York:Springer Verlag, 1985.
3. Bhat, U.N., Lal, R. Attribute control charts for Markov dependent production processes *IIE Transactions*,22,2,181-188, 1990.
4. Broadbent, S. R.The inspection of a Markov process. *Journal of Royal Statistical Society Series B*, 20,111-119, 1958.
5. Lai, C. D. , Govindaraju, K. e Xie, M. Study of a Markov model for a high-quality dependent process, *Journal of Applied Statistics*, 27, 4, 461-473, 2000.
6. Lai, C. D. , Govindaraju, K. e Xie, M. Effects of correlation on fraction nonconforming statistical process control procedures, *Journal of Applied Statistics*, 25, 4, 535-543, 1998.
7. Madsen, R. W. Generalised binomial distributions, *Communications in Statistics: Theory and Methods*, 22, 3065-3086, 1993.
8. Mingoti, S. A. Técnicas estatísticas para o monitoramento de processos via análise de itens não conformes, *Relatório de Pesquisa*, CNPq, 2002.
9. Mingoti, S. A., Fidelis, M.T. Aplicando a geoestatística no controle estatístico de Processos, *Revista Produto & Produção*, 5, 2, 55-70,2001.
10. Mingoti, S. A., Carvalho, J.P. Monitoramento de processos autocorrelacionados: usando o software Minitab no controle via análise de atributos. *Caderno de Resumos da 9a. Escola de Séries Temporais e Econometria (9a. ESTE)*, Belo Horizonte: ABE/Depto.de Estatística- UFMG, 119, 2001.
11. Montgomery, D. C. *Introduction to statistical quality control*.New York: John Wiley, 1991.
12. Nelson, L. S. A control chart for parts-per-million noncomforming itens. *Journal of Quality Technology*, 19, 239-240, 1994.
13. Stimson, W. A. e Mastrangelo, C. M. Monitoring serially dependent process with attributes data, *Journal of Quality Technology*, 28, 279-288, 1996.