A cadeia de Markov no monitoramento, via inspeção por atributos, de processos autocorrelacionados: o programa *SCAP- módulo AI*

Sueli Aparecida Mingoti (UFMG) - sueli@est.ufmg.br
Júlia Pinto de Carvalho (UFMG) - jc100_est@hotmail.com

Resumo

Neste artigo apresentamos a técnica de cadeia de Markov para monitoramento de processos autocorrelacionados que está implementada no módulo AI (Attribute Inspection) do programa computacional SCAP (Statistical Control for Autocorrelated Processes). O programa funciona de modo interativo, acoplado ao software estatístico Minitab for Windows, e permite a estimação do coeficiente de correlação entre unidades amostrais através dos estimadores de Bhat & Lal (1990) e Mingoti (2002), além da construção de gráficos de controle para o número e proporção de "não conformes" do processo. Além da cadeia de Markov o programa SCAP- modúlo AI - também tem disponível o modelo Binomial Generalizado.

Palavras Chave: Cadeia de Markov, Inspeção por atributos, Programa SCAP, Processos autocorrelacionados, Estimação de correlação.

1. Introdução

Gráficos de Controle de Shewhart têm sido utilizados no monitoramento da qualidade de processos quando a inspeção é feita por atributos. Neste caso, considera-se que as unidades amostrais do processo são independentes entre si, em relação ao estado de ser ou não "conforme", e o "número total de defeituosos" é visto como tendo uma distribuição Binomial que pode ser aproximada pela distribuição Normal (Montgomery,1996). No entanto, a correlação entre unidades amostrais aparece frequentemente exigindo o uso de modelos estatísticos mais elaborados (Alwan & Roberts,1995). Alguns modelos alternativos são propostos entre eles o de cadeia de Markov (Bhat & Lal, 1990; Broadbent,1958) no qual presume-se a inspeção serial das unidades do processo e a obtenção da informação sobre o estado individual de cada item na sequência exata inspecionada, e o modelo Binomial Generalizado de Madsen (1993) proposto por Lai, Govindaraju & Xie (1998) no qual 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) e implementados no programa *SCAP - Statistical Control for Autocorrelated Processes*.

O propósito deste artigo é mostrar a implementação do modelo de cadeia de Markov disponível no programa *SCAP*. Detalhes sobre a implementação do modelo Binomial Generalizado podem ser encontrados em Mingoti & Carvalho (2002). Algumas referências interessantes sobre processos autocorrelacionados são: Nelson (1994), Stimson & Mastrangelo (1996) e Mingoti & Fidelis (2001).

2. Modelo de cadeia de Markov

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

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

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

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

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

onde $p = \pi_1$ é a probabilidade limite da cadeia, isto é, $\pi_j = lim(n \to \infty)\; p_{ij}^{(n)}, \; j = 0,1$, 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 defeitusosos do processo. Em Bhat & Lal (1990) é mostrado que sob esta condição de equilibrio da cadeia, a correlação de ordem j é dada por: $\rho_j = corr(Y_n, Y_{n+j}) = \rho^j \;, \; j = 1,2,...$, onde ρ é a correlação de ordem 1 do processo. Assim, 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 que a cadeia seja admissível os parâmetros p e ρ precisam satisfazer as condições:

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

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, 1988). No entanto, pode ser mostrado que assintoticamente X_n tem distribuição aproximadamente Normal com média e variância dadas por:

$$E(X_n) = n\,p$$

$$Var(X_n) = n\,p(1-p) + 2\;p\,(1-p)\frac{\rho}{1-\rho}\left(n - \frac{1-\rho^n}{1-\rho}\right) \qquad (5)$$

Portanto, a opção pela cadeia de Markov para tratamento de dados correlacionados, permite que o usuário mantenha-se dentro da classe da distribuição normal para construção dos gráficos de controle de Shewhart.

2.1 Estimação clássica de parâmetros

Os estimadores de máxima verossimilhança de p e ρ dados por Bhat & Lal (1990) são respectivamente:

$$\hat{p}_{ij} = \frac{n_{ij}}{n_{i0} + n_{i1}} \quad , \; i, j = 0,1 \; ; \quad \hat{p} = \frac{\hat{p}_{01}}{\hat{p}_{01} + \hat{p}_{10}} \quad ; \quad \hat{\rho} = 1 - (\hat{p}_{01} + \hat{p}_{10}) \qquad (6)$$

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) com variância dada por:

$$Var(\hat{\rho}) = \frac{1}{n}\left[\frac{(1-\rho)\,[\,p(1-p)(1-3\rho)+\rho}{p(1-p)}\right] \qquad (7)$$

2.2 Estimadores alternativos para o coeficiente de correlação

Alguns estimadores alternativos para o coeficiente de correlação no caso de inspeção serial foram propostos por Mingoti (2002). Nesta seção vamos descrevê-los brevemente.

2.2.1 Estimador multinomial para observações sequenciais

O estimador multinomial é construido supondo-se que o usuário tenha as observações individuais da sequência de "0" e "1" de inspeção e o modelo de probabilidades dado em (8), isto é,

$$\begin{aligned}
&P[X_i = 1, X_{i+1} = 1] = \rho p + (1-\rho)\, p^2 \\
&P[X_i = 1, X_{i+1} = 0] = (1-\rho)\, p(1-p) \\
&P[X_i = 0, X_{i+1} = 0] = \rho(1-p) + (1-p)^2 (1-\rho) \\
&P[X_i = 0, X_{i+1} = 1] = (1-\rho)\, p(1-p)
\end{aligned} \qquad (8)$$

onde X_i é igual a 1 se o ítem é não conforme e 0 caso contrário. Se temos n ensaios de X_i podemos formar os pares (X_i, X_{i+1}) na sequência em que os ítens foram inspecionados, $i=1,2,\ldots, n-1$. O número total de pares que podem ser formados é igual a (n-1). Se f_{10} e f_{01} denotam respectivamente as frequências amostrais observadas de pares do tipo $(X_i = 1, X_{i+1} = 0)$ e $(X_i = 0, X_{i+1} = 1)$ então, o estimador para ρ, chamado de Mingoti II (Mingoti,2002) é dado como na equação (9) sendo que o valor de p pode ser pré-especificado ou então, estimado pela fração amostral observada de itens não conformes.

$$\hat{\rho} = 1 - \frac{f_{10} + f_{01}}{2(n-1)\,p(1-p)} \qquad (9)$$

Basicamente o estimador (9) descreve o relacionamento das observações amostrais que estão a uma distância de 1 unidade e portanto, é um estimador da correlação serial de ordem 1. Este estimador é uma extensão dos estimadores derivados por Mingoti (2002) no caso do modelo Binomial Generalizado quando se tem as observações sequencias de "0" e "1", e se constitue numa alternativa para estimação do coeficiente de correlação de cadeias de Markov no caso em que a correlação é considerada maior que zero.

2.2.2 Estimador não-paramétrico

O estimador não-paramétrico proposto por Mingoti (2002) dado em (10), é uma adaptação do coeficiente não-paramétrico de concordância de Kendall (Sprent,1989). Se temos uma amostra de *n* itens do processo e para cada item é guardada a informação de X_i podemos utilizar o seguinte estimador para ρ :

$$\hat{\rho} = \frac{(f_{00} + f_{11}) - [(n-1) - f_{00} + n_{11}]}{n-1} = \frac{2[(f_{00} + f_{11})] - (n-1)}{n-1} \qquad (10)$$

onde f_{11} e f_{00} representam a frequência amostral de pares do tipo $(X_i = 1, X_{i+1} = 1)$ e $(X_i = 0, X_{i+1} = 0)$ respectivamente. Este estimador tem a vantagem de não depender do conhecimento da distribuição original de X_i podendo ser utilizado tanto para dados gerados via cadeia de Markov quanto para dados gerados via distribuição Binomial Generalizada, além de situações mais gerais.

2.2.3 Caso de várias amostras sequencias do processo

Os estimadores para ρ usando os métodos de cadeia de Markov, multinomial e não-paramétrico, foram desenvolvidos com base no conhecimento de apenas uma amostra de tamanho *n* do processo. Se tivermos *m>2* amostras independentes todas de tamanho *n*, podemos estimar ρ através dos procedimentos (6), (9) ou (10) para cada amostra individualmente e no final podemos calcular a média das estimativas encontradas, obtendo-se assim uma estimativa para o coeficiente de correlação do processo aproveitando-se a informação de todas as *m* amostras.

3. O programa *SCAP* para monitoramento de processos autocorrelacionados e cadeias de Markov

O programa computacional *SCAP - módulo AI (Attribute Inspection)* foi desenvolvido por Mingoti & Carvalho (2001) com a finalidade de implementar os modelos estatísticos de cadeia de Markov e Binomial Generalizado para o monitoramento de processos autocorrelacionados. O usuário pode estimar os parâmetros p *e* ρ do processo tanto no caso de dados sequencias quanto agrupados e pode construir os gráficos de controle de Shewart para a proporção e número de "não conformes" do processo. A proporção de defeitos pode ser estimada no programa ou pré-especificada pelo usuário. O programa *SCAP* funciona acoplado ao *software* estatístico *Minitab for Windows* e tem uma boa velocidade de processamento. A seguir vamos apresentar um exemplo de aplicação usando o programa *SCAP*. A macro que implementa os procedimentos de estimação é chamada de *scapai.mac*.

Exemplo: Considere os dados apresentados no Quadro 1 que representam erros de leitura em discos rígidos de micro computadores (Lai, *et. al.*, 2000). Tem-se uma inspeção serial com *n=208* discos rígidos observados sendo que o dígito "1" representa a existência de erro e "0" a não existência de erro de leitura. Na planilha de dados (*Worksheet*) do *software Minitab for Windows*, esses dados deverão estar dispostos somente em uma coluna, havendo assim somente uma amostra para análise. A análise destes dados via *SCAP* usando o procedimento clássico de cadeia de Markov para estimação dos parâmetros p *e* ρ descrito na seção 2.1 (equação 6), está apresentada na Figura 1 e resulta nas estimativas iguais a 0,419380 para *p* e 0,158422 para ρ. Caso o usuário tivesse optado pelos métodos de Multinomial (Mingoti II) e não-paramétrico para a estimação do coeficiente de correlação ρ encontraria os valores 0,1568243 e 0,178744, respectivamente.

```
MTB > % scapai
Executing from file: scapai.MAC

Quantos são os subgrupos analisados (amostras)?
DATA> 1

Os subgrupos (amostras) possuem o mesmo tamanho (Y/N)?
y

Qual o tamanho de cada subgrupo (amostra)?
DATA> 208

Para cada subgrupo (amostra), seus dados estão apresentados em:
1. Total de defeitos;
2. Dados individuais.
Digite 1 ou 2.
DATA> 2

Para dados apresentados individualmente, cada coluna deve conter os dados de uma
amostra. Deseja continuar (Y/N)?
y

Você deseja estimar a probabilidade de defeito "p" (Y/N)?
y

Qual é o método de estimação de "p"?
1. Fração de defeituosos na amostra;
2. Cadeia de Markov (somente para dados individuais).
Digite 1 ou 2.
DATA> 2

Data Display

p    0,419380

Você deseja estimar o valor do coeficiente de correlação (Y/N)?
y

Qual é o método de estimação do coeficiente de correlação?
1. Método dos momentos;
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.
DATA> 2

Data Display

ro    0,158422
```

Figura 1: Comandos e resultados do *SCAP* para análise dos dados do exemplo do quadro 1

"1": item conforme; "0": ítem não conforme (as linhas devem ser lidas de cima para baixo e as colunas da esquerda para a direita).

1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0
0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1
0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1	0	1
0	0	0	0	0	0	1	1	0	0	1	0	1	1	0	1	1	1	0	0	1
1	0	1	1	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	0	1
1	0	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	1	0	0	1
1	1	1	0	1	0	1	0	0	0	1	0	0	1	0	1	0	1	1	1	0
1	0	1	0	0	1	0	0	1	0	1	1	0	1	0	0	0	1	0	1	
0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0	

Fonte: LAI,C.D.; XIE,M. & GOVINDARAJU,K.(2000).

Quadro 1: Dados de erros de leitura em discos rígidos de microcomputadores

4. Considerações finais

Embora o exemplo tratado neste artigo seja de dados sequenciais o programa *SCAP* funciona de forma similar para o tratamento de dados agrupados via Binomial Generalizada (Mingoti & Carvalho, 2002). O usuário poderá ainda, solicitar o cálculo dos limites de controle de Shewhart usando as estimativas de p e ρ obtidas por qualquer um dos métodos descritos neste artigo e fazer um gráfico.

O programa *SCAP - módulo AI -* com respectivo manual estará disponível em breve para *download* gratuito no endereço: www.est.ufmg.br.

Agradecimentos

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

Referências

ALWAN, L.C., ROBERTS, H. V. (1995) - The Problem of Missplaced Control Limits. *Applied Statistics, Journal of Royal Statistical Society, Series C* Vol.44, n.3, p.269-278.

BHAT, U.N., LAL, R. (1990) - Attribute Control Charts for Markov Dependent Production Processes. *IIE Transactions* Vol.22, n.2, p.181-188.

BHAT, U. N., LAL, R. (1988) - Number of Success in Markov Trials. *Advanced Applied Probability* Vol.20, p. 677-680.

BROADBENT, S. R. (1958) - The Inspection of a Markov Process. *Journal of Royal Statistical Society Series B* Vol. 20, p.111-119.

LAI, C. D., GOVINDARAJU, K. & XIE, M. (2000) - Study of a Markov Model for a High-Quality Dependent Process. *Journal of Applied Statistics* Vol.27, n.4, p.461-473.

LAI, C. D., GOVINDARAJU, K. & XIE, M. (1998) - Effects of Correlation on Fraction Nonconforming Statistical Process Control Procedures. *Journal of Applied Statistics* Vol. 25, n. 4, p.535-543

MADSEN, R. W. (1993) - Generalized Binomial Distributions. *Communications in Statistics: Theory and Methods* Vol. 22, p. 3065-3086.

MINITAB INC. (1999) - User's guide Minitab for windows, version 13.1, Pennsilvania.

MINGOTI, S. A., CARVALHO, J.P. (2002) - O Programa *SCAP* para Monitoramento de Processos Autocorrelacionados e Inspeção por Atributos. *Atas da III Jornada Regional de Estatística e II Semana de Estatística* , Universidade Estadual de Maringá, p.124-132

MINGOTI, S. A. (2002) - Técnicas Estatísticas para o Monitoramento de Processos via Análise de Itens Não Conformes, *Relatório de Pesquisa, CNPq*.

MINGOTI, S. A., FIDELIS, M.T. (2001) - Aplicando a Geoestatística no Controle Estatístico de Processos. *Revista Produto & Produção* Vol. 5, n.2, 55-70.

MINGOTI, S. A., CARVALHO, J.P. (2001) - 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, p.119.

MONTGOMERY, D. C.(1996) - *Introduction to Statistical Quality Control*. New York: John Wiley.

NELSON, L. S. (1994) - A Control Chart for Parts-Per-Million Noncomforming Itens. *Journal of Quality Technology* Vol.19, p. 239-240.

SPRENT,P.- (1989) *Applied Nonparametric Statistical Methods*. New York: Chapman Hall.

STIMSON, W. A. & MASTRANGELO, C. M. (1996) - Monitoring Serially Dependent Process with Attributes Data, *Journal of Quality Technology* Vol. 28, p. 279-288.