

# Estratégia de Amostragem Sistemática para Reduzir o Efeito da Autocorrelação no Desempenho do Gráfico de Controle T2 de Hotelling

Roberto Campos Leoni  
rleoni@yahoo.com.br  
AEDB

Marcela Aparecida Guerreiro Machado  
marcela@feg.unesp.br  
UNESP

Antonio Fernando Branco Costa  
fbranco@feg.unesp.br  
UNESP

**Resumo:** Uma hipótese indispensável para a utilização de um gráfico de controle é a independência entre as observações da característica de qualidade a ser monitorada. Entretanto, a crescente automatização no segmento de manufatura revolucionou muitos processos, acentuando coincidentemente a dependência em série entre observações. Nesse artigo, analisou-se o desempenho do gráfico de controle T2 de Hotelling em processos bivariados com observações correlacionadas e autocorrelacionadas. Para tanto, foi necessário desenvolver a matriz de correlação cruzada do vetor de médias amostrais sob a hipótese de que as observações sejam descritas por um vetor autorregressivo de primeira ordem – VAR(1). Usando o método da amostragem sistemática através da seleção de amostras de tamanho  $n$ , procurou-se diminuir o efeito indesejado da autocorrelação presente no processo. Os resultados indicam que a estratégia de amostragem sistemática sempre melhora o desempenho do gráfico, exceto quando apenas uma variável é afetada por uma causa especial e as observações desta variável não são autocorrelacionadas. Durante a formação de uma amostra de tamanho  $n$ , se o intervalo para a seleção dos elementos da amostra for igual a  $s=1$ , o número médio de amostras até o sinal (ARL) reduz em média mais de 30%. Se o intervalo para a seleção dos elementos da amostra for igual a  $s=2$ , o ARL reduz em média mais de 40%.

**Palavras Chave:** Autocorrelação - amostragem - T2 de Hotelling - modelo VAR(1) -

## 1. INTRODUÇÃO

Processos modernos de manufatura apresentam observações multivariadas que além de correlacionadas, também são afetadas pela autocorrelação. A autocorrelação se refere a dependência entre observações de uma mesma característica de qualidade. A correlação se refere a dependência entre as observações das características de qualidade. Pan e Jarret (2011) descrevem muitos destes processos.

Uma das hipóteses fundamentais para a utilização de um gráfico de controle é a independência entre as observações da característica de qualidade a ser monitorada. Entretanto, a crescente automatização no segmento de manufatura revolucionou muitos processos, acentuando coincidentemente a dependência em série entre observações. Uma abordagem típica para estudar o desempenho de gráficos de controle em processos autocorrelacionados é ajustar um modelo de séries temporais às observações do processo. O modelo autoregressivo de primeira ordem, AR(1), e o vetor autoregressivo de primeira ordem, VAR(1), são adotados em muitos estudos que lidam com gráficos de controle univariados e multivariados (ALWAN; ROBERTS, 1988; MONTGOMERY; MASTRANGELO, 1991; TSENG; ADAMS, 1994; RUNGER et al., 1995; FALTIN et al., 1997; ENGLISH et al., 2000; GHOURABI; LIMAM, 2007; CHAN; LI, 1994; CHARNES, 1995; KALGONDA; KULKARNI, 2004; COSTA; CLARO, 2008; PAN; JARRET, 2007; HWARNG; WANG, 2010; HUANG et al., 2013a; HUANG et al., 2013b; FRANCO et al., 2013).

A autocorrelação entre os itens da amostra tem um sério impacto sobre o desempenho dos gráficos de controle. Para reduzir o efeito indesejado de autocorrelação no desempenho do gráfico de  $\bar{X}$ , Costa e Castagliola (2011) propuseram uma estratégia de amostragem denominada *s-kip*. De acordo com esta estratégia, as amostras são obtidas através da seleção de um elemento da linha de produção e, em seguida, pulam-se *s* elementos consecutivos antes de selecionar o próximo. Esta estratégia também é conhecida na área de amostragem como amostragem sistemática (BOLFARINE; BUSSAB, 2005). Mais recentemente, Franco et al. (2014) realizaram o design estatístico e econômico do gráfico de controle de  $\bar{X}$  em processos com observações autoregressivas de primeira ordem usando a estratégia *s-kip*.

O apelo prático de pular observações reside no fato de que o desempenho do gráfico aumenta substancialmente quando algumas observações são ignoradas. Baseando-se nessa premissa, considerou-se no presente artigo a estratégia de amostragem sistemática para melhorar o desempenho do gráfico de controle  $T^2$  de Hotelling. A dedução da matriz de covariância cruzada do vetor de médias amostrais ( $\Gamma_{\bar{X}}$ ) considerando a estratégia de amostragem sistemática não é simples, contudo necessária para obter  $\lambda$ , parâmetro de não centralidade da distribuição qui-quadrado. O poder do gráfico de controle  $T^2$  de Hotelling é função de  $\lambda$ , logo o conhecimento da matriz de covariância cruzada facilita a investigação do efeito da correlação e da autocorrelação no gráfico de controle.

O efeito combinado da autocorrelação e da correlação cruzada no desempenho do gráfico  $T^2$  de Hotelling sob a estratégia de amostragem sistemática é digno de investigação, pois não há referências de trabalhos que utilizam esta estratégia em processos com duas ou mais variáveis. Na seção 2 e Apêndice, obteve-se a matriz de correlação cruzada do vetor de médias amostrais ( $\Gamma_{\bar{X}}$ ) sob a hipótese de que as observações sejam descritas por um vetor autorregressivo de primeira ordem – VAR(1). Na seção 3, investigou-se o efeito da autocorrelação e da correlação cruzada no desempenho do gráfico de controle  $T^2$  de Hotelling sob a estratégia de amostragem sistemática em processos bivariados. Finalmente, na seção 4, apresentamos as principais conclusões.

## 2. O MODELO AUTOREGRESSIVO E A MATRIZ DE COVARIÂNCIA CRUZADA

O modelo vetor autoregressivo tem sido utilizado em estudos recentes para modelar observações multivariadas com autocorrelação e correlação (PAN e JARRET, 2007; HWRANG e WANG, 2010; HUANG et al., 2013a; HUANG et al., 2013b). O modelo VAR(1) é representado por:

$$\mathbf{X}_t - \boldsymbol{\mu} = \Phi(\mathbf{X}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t \quad (1)$$

sendo  $\mathbf{X}_t \sim N_p(\boldsymbol{\mu}, \Gamma)$  o vetor de observações de dimensão  $(p \times 1)$  no instante  $t$  ( $p$  é o número de variáveis),  $\boldsymbol{\mu}$  é o vetor de médias,  $\boldsymbol{\varepsilon}_t$  é um vetor aleatório com observações independentes e distribuição normal multivariada com média zero e matriz de covariância  $\Sigma$  e  $\Phi$  é uma matriz com parâmetros de autocorrelação de ordem  $(p \times p)$ .

De acordo com Kalgonda e Kulkarni (2004), a matriz de correlação cruzada de  $\mathbf{X}_t$  possui a seguinte propriedade:  $\Gamma = \Phi\Gamma\Phi' + \Sigma$ . Após alguma manipulação algébrica, é possível obter a relação:

$$\text{Vec } \Gamma = \left( I_{p^2} - \Phi \otimes \Phi \right)^{-1} \text{Vec } \Sigma \quad (2)$$

onde  $\otimes$  é o operador produto de Kronecker e  $\text{Vec}$  é o operador que transforma a matriz em um vetor empilhando suas colunas.

Para estudar o efeito da autocorrelação e correlação no gráfico de  $T^2$ , considerou-se um processo bivariado ( $p=2$ ):

$$\Phi = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \quad (3)$$

A partir de (2) e (3), segue-se que:

$$\Gamma = \begin{pmatrix} \sigma_X^2 = (1-a^2)^{-1} & \sigma_{XY} = \rho(1-ab)^{-1} \\ \sigma_{XY} = \rho(1-ab)^{-1} & \sigma_Y^2 = (1-b^2)^{-1} \end{pmatrix} \quad (4)$$

A matriz de covariância cruzada do vetor de médias amostrais formadas por subgrupos racionais de tamanho  $n$  selecionados com base na estratégia de amostragem sistemática é desenvolvida no Apêndice:

$$\Gamma_{\bar{X}} = \begin{pmatrix} \zeta_1^2 = \frac{\sigma_X^2}{n} \left[ 1 + \frac{2}{n} \sum_{j=1}^{n-1} (n-j)a^{(s+1)j} \right] & \zeta_{12} = \frac{\sigma_{XY}}{n} \left[ 1 + \frac{1}{n} \sum_{j=1}^{n-1} (n-j)a^{(s+1)j} + \frac{1}{n} \sum_{j=1}^{n-1} (n-j)b^{(s+1)j} \right] \\ \zeta_{21} = \zeta_{12} & \zeta_2^2 = \frac{\sigma_Y^2}{n} \left[ 1 + \frac{2}{n} \sum_{j=1}^{n-1} (n-j)b^{(s+1)j} \right] \end{pmatrix} \quad (5)$$

sendo:  $\sigma_X^2 = \frac{1}{1-a^2}$ ;  $\sigma_Y^2 = \frac{1}{1-b^2}$ ;  $\sigma_{XY} = \frac{\rho}{1-ab}$ . Na ausência da autocorrelação ( $a=b=0$ ),  $\Gamma_{\bar{X}}$  se reduz a  $\Psi_{\bar{X}} = n^{-1}\Sigma$ .

O parâmetro de não centralidade da distribuição de  $T^2$  é função dos elementos da matriz inversa de covariância cruzada. Sob a hipótese de emprego da estratégia de

amostragem sistemática, buscou-se conhecer o comportamento de  $\Gamma_{\bar{X}}^{-1}$  diante do efeito da autocorrelação e da correlação no desempenho do gráfico de controle  $T^2$  de Hotelling.

$$\Gamma_{\bar{X}}^{-1} = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{12} & \theta_{22} \end{pmatrix} \quad (6)$$

A Tabela 1 apresenta valores de  $\theta_{11}$ ,  $\theta_{22}$  e  $\theta_{12}$ , considerando-se:  $n=4$ ,  $\rho \in (0; 0,7)$  e  $a$  e  $b \in (0; 0,2; 0,5; 0,7)$ . Os elementos da diagonal principal de  $\Gamma_{\bar{X}}^{-1}$  aumentam à medida que aumenta a correlação ( $\rho$ ) e a amplitude do intervalo de amostragem ( $s$ ). Contudo, os elementos da diagonal principal de  $\Gamma_{\bar{X}}^{-1}$  diminuem com a presença da autocorrelação ( $a, b$ ). Quando as variáveis são independentes, os elementos da diagonal secundária de  $\Gamma_{\bar{X}}^{-1}$  são iguais a zero e na presença de correlação, eles são negativos.

**Tabela 1:** Elementos da matriz  $\Gamma_{\bar{X}}^{-1}$ ;  $a$  e  $b \in (0; 0,2; 0,5; 0,7)$ ;  $\rho \in (0; 0,7)$ .

		$\rho$													
		0						0,7							
$a$	$b$	0,0	0,0	0,2	0,0	0,5	0,0	0,7	0,0	0,0	0,2	0,0	0,5	0,0	0,7
$b$	$a$	0,0	0,2	0,2	0,5	0,5	0,7	0,7	0,0	0,2	0,2	0,5	0,5	0,7	0,7
$s=0$	$\theta_{11}$	4,00	4,00	2,86	4,00	1,45	4,00	0,75	7,84	7,70	5,60	6,87	2,85	5,86	1,48
	$\theta_{22}$	4,00	2,86	2,86	1,45	1,45	0,75	0,75	7,84	5,50	5,50	2,50	2,85	1,10	1,48
	$\theta_{12}$	0,00	0,00	0,00	0,00	0,00	0,00	0,00	-5,49	-4,51	-3,92	-2,68	-2,00	-1,43	-1,03
$s=1$	$\theta_{11}$	4,00	4,00	3,62	4,00	2,08	4,00	1,00	7,84	7,56	7,09	6,45	4,07	5,58	1,97
	$\theta_{22}$	4,00	3,62	3,62	2,08	2,08	1,00	1,00	7,84	6,83	7,09	3,35	4,07	1,40	1,97
	$\theta_{12}$	0,00	0,00	0,00	0,00	0,00	0,00	0,00	-5,49	-4,93	-4,96	-2,87	-2,85	-1,48	-1,38
$s=2$	$\theta_{11}$	4,00	4,00	3,79	4,00	2,49	4,00	1,23	7,84	7,55	7,44	6,36	4,88	5,45	2,42
	$\theta_{22}$	4,00	3,79	3,79	2,49	2,49	1,23	1,23	7,84	7,16	7,44	3,96	4,88	1,68	2,42
	$\theta_{12}$	0,00	0,00	0,00	0,00	0,00	0,00	0,00	-5,49	-5,05	-5,21	-3,05	-3,42	-1,56	-1,69

### 3. O EFEITO DA AUTOCORRELAÇÃO E DA CORRELAÇÃO NO DESEMPENHO DO GRÁFICO $T^2$ DE HOTELLING

O gráfico de controle  $T^2$  de Hotelling é o mais conhecido esquema de controle utilizado para detectar desvios na média de processos multivariados (HOTELLING, 1947). Quando o vetor de médias  $\mu_0 = (\mu_{01}, \mu_{02})$  e a matriz  $\Gamma_{\bar{X}}$  são conhecidos, a estatística de monitoramento  $T^2$  de Hotelling é representada por:

$$T^2 = (\bar{\mathbf{X}} - \mu_0)' \Gamma_{\bar{X}}^{-1} (\bar{\mathbf{X}} - \mu_0) \quad (7)$$

Com o processo em controle,  $T^2 \sim \chi_p^2$ . Se ocorrer uma causa especial que provoque a mudança no vetor de médias para  $\mu_1 = (\mu_{11}, \mu_{12})$ ,  $T^2 \sim \chi_{(p, \lambda)}^2$  com parâmetro de não centralidade  $\lambda^2 = \delta' \Gamma_{\bar{X}}^{-1} \delta$ , sendo  $\delta = (\delta_1 = \mu_{11} - \mu_{01}, \delta_2 = \mu_{12} - \mu_{02})$ , veja Wu e Makis (2008). Sem perda de generalidade, considerou-se  $\mu_{01} = \mu_{02} = 0$  na avaliação do desempenho do gráfico de  $T^2$ . Na ausência de autocorrelação, o parâmetro de não centralidade é dado por  $\lambda^2 = \delta' \Psi_{\bar{X}}^{-1} \delta$ . Como  $\delta' \Gamma_{\bar{X}}^{-1} \delta < \delta' \Psi_{\bar{X}}^{-1} \delta$ , conclui-se que a autocorrelação reduz o poder o gráfico em detectar uma causa especial.

Para estudar o efeito da correlação e da autocorrelação no desempenho do gráfico de  $T^2$ , assumiu-se um processo bivariado e amostras de tamanho  $n=4$  foram coletadas seguindo os intervalos de amostragem  $s \in (0,1,2)$ . O desempenho dos gráficos foi avaliado pelo número médio de amostras até o sinal (ARL). Quando o processo está em controle, o ARL mede o número médio de amostras até a ocorrência de um alarme falso (ARL<sub>0</sub>). Quando o processo está fora de controle, o gráfico com menor ARL detecta com maior rapidez mudança no processo. Na prática, amostras preliminares são coletadas para estimar a matriz  $\Gamma$  e os parâmetros  $a$ ,  $b$  e  $\rho$ ; veja equação (4). O ARL é dado por:

$$ARL = \left\{ 1 - \left[ \Pr(\chi^2_{(p,\lambda)} < LC) \right] \right\}^{-1} \quad (8)$$

sendo  $LC$  o limite de controle do gráfico de  $T^2$  que é obtido a partir da equação (8) fixando o valor do  $ARL_0=370,4$  e  $\lambda=0$ . Os demais ARLs são apresentados nas Tabelas 2, 3 e 4.

**Tabela 2:** Valores do ARL - Gráfico de  $T^2$ ,  $\delta_1 = \delta_2$ .

(a; b)		(0; 0)		(0; 0,2)			(0; 0,5)			(0; 0,7)		
s	$\delta_1$	0, 1 ou 2	0	1	2	0	1	2	0	1	2	
$\delta_2$	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	
$\rho=0,0$												
,50	,50	27,73	34,43	29,74	28,78	46,56	40,51	37,09	55,16	51,84	49,02	
,75	,75	7,74	9,99	8,40	8,09	14,47	12,17	10,92	17,96	16,58	15,44	
1,00	1,00	3,06	3,89	3,30	3,18	5,63	4,72	4,24	7,07	6,50	6,03	
1,25	1,25	1,68	2,02	1,78	1,73	2,78	2,38	2,18	3,43	3,17	2,96	
1,50	1,50	1,21	1,36	1,25	1,23	1,71	1,53	1,43	2,03	1,90	1,80	
$\rho=0,3$												
,50	,50	39,83	48,07	42,20	40,96	59,46	53,24	49,45	64,48	61,80	59,30	
,75	,75	11,92	15,06	12,79	12,33	19,80	17,16	15,61	22,03	20,84	19,74	
1,00	1,00	4,62	5,87	4,97	4,79	7,86	6,74	6,10	8,83	8,30	7,83	
1,25	1,25	2,34	2,89	2,49	2,41	3,79	3,28	2,99	4,24	3,99	3,77	
1,50	1,50	1,51	1,77	1,58	1,54	2,21	1,95	1,81	2,43	2,31	2,20	
$\rho=0,7$												
,50	,50	55,82	64,12	58,39	56,95	67,08	66,04	63,61	65,51	67,22	67,16	
,75	,75	18,24	21,87	19,34	18,72	23,23	22,75	21,64	22,51	23,29	23,26	
1,00	1,00	7,19	8,76	7,66	7,40	9,36	9,14	8,66	9,04	9,39	9,37	
1,25	1,25	3,48	4,20	3,70	3,57	4,49	4,38	4,16	4,34	4,50	4,49	
1,50	1,50	2,06	2,42	2,16	2,10	2,56	2,51	2,39	2,48	2,56	2,56	
(a; b)		(0; 0)		(0,2; 0,2)			(0,5; 0,5)			(0,7; 0,7)		
s	$\delta_1$	0, 1 ou 2	0	1	2	0	1	2	0	1	2	
$\delta_2$	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	
$\rho=0,0$												
,50	,50	27,73	43,87	31,98	29,90	93,65	64,57	52,09	159,45	129,50	108,90	
,75	,75	7,74	13,43	9,15	8,45	36,68	22,08	16,69	80,52	58,64	45,48	
1,00	1,00	3,06	5,22	3,57	3,31	15,74	8,85	6,54	41,01	27,61	20,29	
1,25	1,25	1,68	2,60	1,89	1,79	7,62	4,25	3,19	21,85	13,98	9,98	
1,50	1,50	1,21	1,63	1,30	1,26	4,19	2,44	1,91	12,35	7,71	5,46	
$\rho=0,3$												
,50	,50	39,83	60,48	45,39	42,68	118,51	85,46	70,57	187,11	156,80	135,09	
,75	,75	11,92	20,25	14,01	12,98	51,43	32,28	24,86	103,84	78,44	62,47	
1,00	1,00	4,62	8,05	5,45	5,04	23,53	13,57	10,09	56,91	39,68	29,85	
1,25	1,25	2,34	3,88	2,70	2,52	11,72	6,53	4,83	31,98	21,04	15,25	
1,50	1,50	1,51	2,25	1,68	1,59	6,43	3,61	2,73	18,70	11,86	8,43	
$\rho=0,7$												
,50	,50	55,82	81,12	62,78	59,41	145,92	109,97	92,97	214,55	185,10	163,19	
,75	,75	18,24	30,04	21,27	19,78	70,22	46,13	36,31	130,24	102,03	83,48	
1,00	1,00	7,19	12,50	8,49	7,85	34,52	20,64	15,55	76,85	55,62	42,94	
1,25	1,25	3,48	6,00	4,08	3,78	17,96	10,17	7,52	45,73	31,13	23,04	
1,50	1,50	2,06	3,33	2,35	2,20	10,02	5,57	4,13	27,90	18,16	13,08	

Considerou-se na análise do desempenho variáveis não correlacionadas,  $\rho = 0$ , e variáveis com baixa ou alta correlação,  $\rho = 0,3$  ou  $0,7$ . Também foram consideradas observações independentes e autocorrelacionadas, ou seja:  $a, b \in (0; 0,2; 0,5; 0,7)$ .

**Tabela 3:** Valores do ARL - Gráfico de  $T^2$ ,  $\delta_1$  ou  $\delta_2=0$ .

(a; b)		(0; 0)			(0; 0,2)			(0; 0,5)			(0; 0,7)		
s	$\delta_2$	0, 1 ou 2	0	1	2	0	1	2	0	1	2		
$\delta_1$		ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL		
<b><math>\rho=0,0</math></b>													
,00	,50	67,32	95,27	75,12	71,36	163,04	126,03	108,04	230,33	201,94	180,35		
,50	,00	67,32	67,32	67,32	67,32	67,32	67,32	67,32	67,32	67,32	67,32		
,00	1,00	9,41	16,19	11,09	10,26	42,87	26,28	20,02	90,63	67,10	52,66		
1,00	,00	9,41	9,41	9,41	9,41	9,41	9,41	9,41	9,41	9,41	9,41		
,00	1,50	2,57	4,31	2,98	2,77	13,05	7,28	5,39	35,03	23,24	16,92		
1,50	,00	2,57	2,57	2,57	2,57	2,57	2,57	2,57	2,57	2,57	2,57		
<b><math>\rho=0,3</math></b>													
,00	,50	60,48	87,08	68,09	64,54	154,61	118,79	101,38	224,59	196,46	175,07		
,50	,00	60,48	60,61	60,75	60,76	61,50	62,03	62,16	62,91	63,40	63,63		
,00	1,00	8,05	13,99	9,57	8,84	38,60	23,63	17,96	85,39	63,18	49,52		
1,00	,00	8,05	8,08	8,10	8,10	8,25	8,35	8,37	8,52	8,62	8,66		
,00	1,50	2,25	3,72	2,61	2,44	11,47	6,46	4,80	32,25	21,45	15,64		
1,50	,00	2,25	2,26	2,26	2,26	2,30	2,32	2,33	2,36	2,38	2,39		
<b><math>\rho=0,7</math></b>													
,00	,50	28,53	46,04	34,59	32,41	107,78	81,44	68,10	192,22	167,20	147,73		
,50	,00	28,53	29,27	30,07	30,10	34,34	37,40	38,16	42,47	45,26	46,61		
,00	1,00	3,15	5,55	3,91	3,63	19,93	12,57	9,57	60,27	45,09	35,34		
1,00	,00	3,15	3,24	3,34	3,34	3,87	4,28	4,39	5,01	5,43	5,64		
,00	1,50	1,23	1,70	1,37	1,31	5,36	3,35	2,61	20,16	13,90	10,30		
1,50	,00	1,23	1,24	1,26	1,26	1,36	1,44	1,46	1,58	1,67	1,72		
(a; b)		(0; 0)			(0,2; 0,2)			(0,5; 0,5)			(0,7; 0,7)		
s	$\delta_2$	0, 1 ou 2	0	1	2	0	1	2	0	1	2		
$\delta_1$		ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL		
<b><math>\rho=0,0</math></b>													
,00	,50	67,32	95,27	75,12	71,36	163,04	126,03	108,04	230,33	201,94	180,35		
,50	,00	67,32	95,27	75,12	71,36	163,04	126,03	108,04	230,33	201,94	180,35		
,00	1,00	9,41	16,19	11,09	10,26	42,87	26,28	20,02	90,63	67,10	52,66		
1,00	,00	9,41	16,19	11,09	10,26	42,87	26,28	20,02	90,63	67,10	52,66		
,00	1,50	2,57	4,31	2,98	2,77	13,05	7,28	5,39	35,03	23,24	16,92		
1,50	,00	2,57	4,31	2,98	2,77	13,05	7,28	5,39	35,03	23,24	16,92		
<b><math>\rho=0,3</math></b>													
,00	,50	60,48	86,92	67,80	64,26	153,08	116,62	99,17	221,26	192,21	170,40		
,50	,00	60,48	86,92	67,80	64,26	153,08	116,62	99,17	221,26	192,21	170,40		
,00	1,00	8,05	13,94	9,51	8,79	37,86	22,87	17,31	82,48	60,27	46,86		
1,00	,00	8,05	13,94	9,51	8,79	37,86	22,87	17,31	82,48	60,27	46,86		
,00	1,50	2,25	3,71	2,59	2,42	11,20	6,23	4,62	30,74	20,16	14,58		
1,50	,00	2,25	3,71	2,59	2,42	11,20	6,23	4,62	30,74	20,16	14,58		
<b><math>\rho=0,7</math></b>													
,00	,50	28,53	45,00	32,87	30,75	95,44	66,03	53,36	161,55	131,52	110,81		
,50	,00	28,53	45,00	32,87	30,75	95,44	66,03	53,36	161,55	131,52	110,81		
,00	1,00	3,15	5,39	3,68	3,42	16,24	9,14	6,76	42,09	28,40	20,91		
1,00	,00	3,15	5,39	3,68	3,42	16,24	9,14	6,76	42,09	28,40	20,91		
,00	1,50	1,23	1,66	1,32	1,28	4,32	2,51	1,96	12,75	7,96	5,65		
1,50	,00	1,23	1,66	1,32	1,28	4,32	2,51	1,96	12,75	7,96	5,65		

Na ausência de autocorrelação ( $a=b=0$ ) ou na presença de duas variáveis com o mesmo nível de autocorrelação ( $a=b \neq 0$ ), alta dependência entre duas variáveis aumenta o desempenho do gráfico de  $T^2$  quando há deslocamento na média de apenas uma variável ( $\delta_1 \neq 0; \delta_2=0$  ou  $\delta_1=0; \delta_2 \neq 0$ ). Por exemplo, se  $a=b=\rho=0$  e  $(\delta_1; \delta_2) = (0; 1,0)$ , o gráfico de  $T^2$  requer, em média,  $ARL=9,41$  amostras para sinalizar o deslocamento na média do processo. Esse número reduz para  $ARL=3,15$  quando ambas as variáveis são altamente correlacionadas ( $a=b=0; \rho=0,7$ ). Se não há independência entre as variáveis ( $a=b=0,5; \rho=0$ ) e  $(\delta_1; \delta_2) = (0; 1,0)$ , o  $ARL = 42,87$ . Quando a correlação é alta ( $a=b=0,5; \rho=0,7$ ), o  $ARL$  decresce para 16,24. Nesses dois últimos casos, a estratégia de emprego da amostragem sistemática ( $s=2$ ) reduz os ARLs, respectivamente, de 42,87 para 20,02 e de 16,24 para 6,76, veja a Figura 1 e Tabela 3.

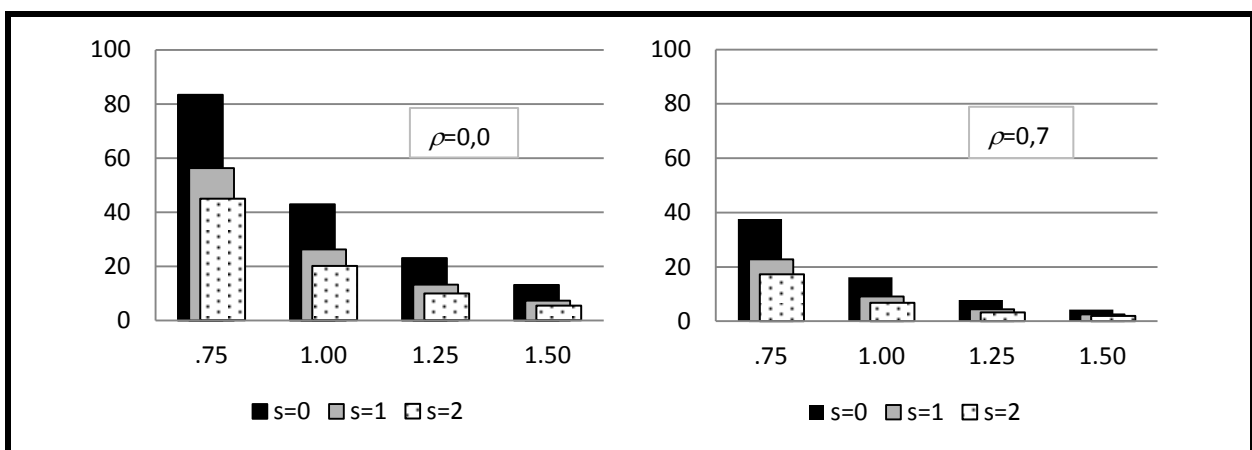


**Tabela 4:** Valores do ARL - Gráfico de  $T^2$ ,  $\delta_1 \neq \delta_2$

(a; b)		(0; 0)				(0; 0,2)				(0; 0,5)				(0; 0,7)			
s	$\delta_1$	$\delta_2$	0, 1 ou 2	0	1	2	0	1	2	0	1	2	0	1	2		
			ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL		
$\rho=0,0$																	
,50	1,00		6,50	9,99	7,42	6,97	20,48	14,41	11,75	33,41	27,67	23,58					
1,00	,50		6,50	7,17	6,72	6,62	8,15	7,69	7,41	8,72	8,51	8,32					
,50	1,50		2,22	3,43	2,51	2,37	8,44	5,28	4,12	17,76	13,15	10,33					
1,50	,50		2,22	2,31	2,25	2,23	2,43	2,37	2,34	2,49	2,47	2,45					
1,00	1,50		1,61	2,10	1,74	1,67	3,54	2,71	2,34	5,25	4,50	3,96					
1,50	1,00		1,61	1,79	1,66	1,64	2,09	1,94	1,85	2,29	2,22	2,15					
$\rho=0,3$																	
,50	1,00		8,76	14,02	10,10	9,43	29,07	20,06	16,14	44,37	36,84	31,38					
1,00	,50		8,76	9,15	8,87	8,81	9,40	9,26	9,12	9,40	9,40	9,35					
,50	1,50		2,56	4,24	2,96	2,76	11,48	6,80	5,15	24,25	17,56	13,54					
1,50	,50		2,56	2,57	2,57	2,57	2,55	2,57	2,57	2,53	2,55	2,56					
1,00	1,50		2,12	2,98	2,34	2,23	5,23	3,88	3,27	7,18	6,18	5,43					
1,50	1,00		2,12	2,31	2,18	2,15	2,51	2,40	2,32	2,56	2,53	2,49					
$\rho=0,7$																	
,50	1,00		8,30	15,70	10,27	9,38	39,85	25,91	19,99	58,22	49,07	41,80					
1,00	,50		8,30	7,56	8,24	8,35	6,92	7,97	8,38	7,20	7,91	8,32					
,50	1,50		1,88	3,35	2,26	2,09	1,75	6,84	4,93	32,26	22,39	16,67					
1,50	,50		1,88	1,77	1,89	1,91	12,74	1,94	2,02	1,89	2,03	2,12					
1,00	1,50		2,56	4,19	2,98	2,77	7,87	5,75	4,69	9,24	8,45	7,60					
1,50	1,00		2,56	2,45	2,55	2,56	2,24	2,46	2,53	2,22	2,38	2,47					

(a; b)		(0; 0)				(0,2; 0,2)				(0,5; 0,5)				(0,7; 0,7)			
s	$\delta_1$	$\delta_2$	0, 1 ou 2	0	1	2	0	1	2	0	1	2	0	1	2		
			ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL	ARL		
$\rho=0,0$																	
,50	1,00		6,50	11,33	7,68	7,10	31,74	18,81	14,13	72,02	51,69	39,67					
1,00	,50		6,50	11,33	7,68	7,10	31,74	18,81	14,13	72,02	51,69	39,67					
,50	1,50		2,22	3,64	2,55	2,38	11,00	6,12	4,53	30,26	19,82	14,33					
1,50	,50		2,22	3,64	2,55	2,38	11,00	6,12	4,53	30,26	19,82	14,33					
1,00	1,50		1,61	2,46	1,80	1,71	7,14	3,99	3,00	20,60	13,13	9,36					
1,50	1,00		1,61	2,46	1,80	1,71	7,14	3,99	3,00	20,60	13,13	9,36					
$\rho=0,3$																	
,50	1,00		8,76	15,13	10,34	9,56	40,53	24,67	18,74	86,86	63,93	49,96					
1,00	,50		8,76	15,13	10,34	9,56	40,53	24,67	18,74	86,86	63,93	49,96					
,50	1,50		2,56	4,30	2,97	2,77	13,02	7,27	5,37	34,97	23,19	16,89					
1,50	,50		2,56	4,30	2,97	2,77	13,02	7,27	5,37	34,97	23,19	16,89					
1,00	1,50		2,12	3,46	2,44	2,28	10,43	5,80	4,30	28,90	18,86	13,61					
1,50	1,00		2,12	3,46	2,44	2,28	10,43	5,80	4,30	28,90	18,86	13,61					
$\rho=0,7$																	
,50	1,00		8,30	14,37	9,80	9,06	38,82	23,52	17,82	84,07	61,60	47,98					
1,00	,50		8,30	14,37	9,80	9,06	38,82	23,52	17,82	84,07	61,60	47,98					
,50	1,50		1,88	2,99	2,14	2,01	8,90	4,95	3,69	25,15	16,24	11,65					
1,50	,50		1,88	2,99	2,14	2,01	8,90	4,95	3,69	25,15	16,24	11,65					
1,00	1,50		2,56	4,29	2,97	2,77	13,00	7,26	5,37	34,93	23,16	16,86					
1,50	1,00		2,56	4,29	2,97	2,77	13,00	7,26	5,37	34,93	23,16	16,86					



**Figura 1:** ARL do gráfico de  $T^2$ ;  $a=0,5$ ;  $b=0,5$ ;  $\delta_1=0$ ,  $\delta_2 \in (0,75;1,00;1,25;1,50)$ .

Efeito contrário é observado se ocorre deslocamento na média das duas variáveis. Por exemplo, se  $a=b=\rho=0$  e  $(\delta_1; \delta_2) = (1,0; 1,0)$ , o gráfico de  $T^2$  requer, em média,  $ARL=3,06$  amostras para sinalizar o deslocamento na média do processo. O ARL aumenta para 7,19 quando as variáveis apresentam alta correlação  $a=b=0; \rho=0,7$ .

Quando ambas são autocorrelacionadas,  $a=b=0,5; \rho=0$  e  $(\delta_1; \delta_2) = (1,0; 1,0)$ , o  $ARL=15,74$ . O ARL aumenta para 34,52 quando há alta correlação  $a=b=0,5; \rho=0,7$ .

Nesses dois últimos casos, a estratégia de emprego da amostragem sistemática ( $s=2$ ) reduz os ARLs, respectivamente de 15,74 para 6,54 e de 34,52 para 15,55, veja a Figura 2 e Tabela 2.

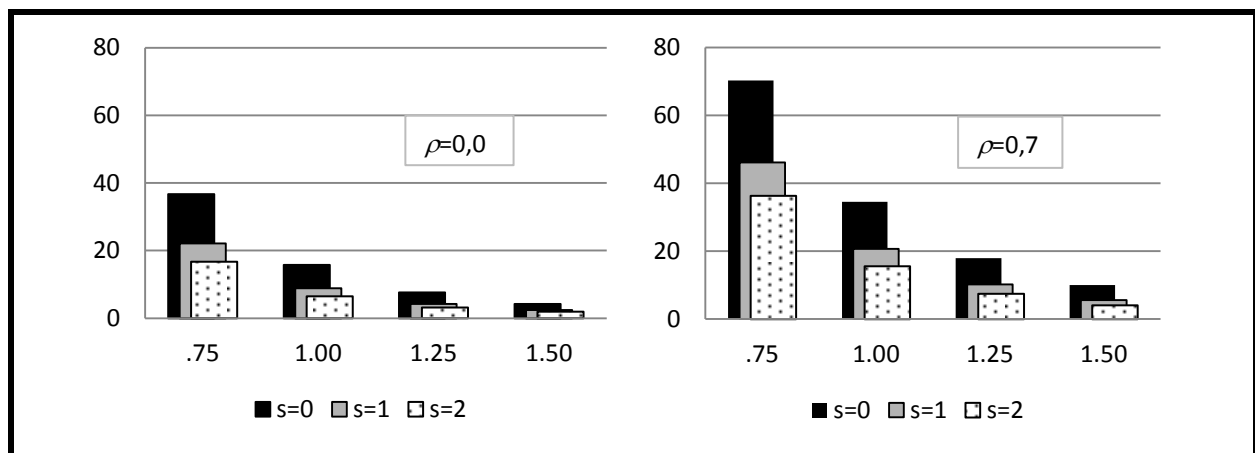


Figura 2: ARL do gráfico de  $T^2$ ;  $a=0,5; b=0,5; \delta_1=\delta_2 \in (0,75;1,00;1,25;1,50)$ .

Se apenas uma variável é autocorrelacionada e apenas uma sofre deslocamento na média, fruto de uma causa especial que atua no processo, ( $a \neq 0; b=0$  ou  $a=0; b \neq 0$  e  $\delta_1 \neq 0; \delta_2=0$  ou  $\delta_1=0; \delta_2 \neq 0$ ), alta dependência entre as duas variáveis aumenta o desempenho do gráfico de  $T^2$ . Por exemplo, se a causa especial modifica a média da variável independente ( $a; b; \delta_1; \delta_2$ ) =  $(0; 0,2; 1,0; 0)$  e  $\rho=0$ , o  $ARL=9,41$ . Esse número diminui para  $ARL=3,24$  quando há alta correlação ( $\rho=0,7$ ). Nesse caso, a estratégia de amostragem sistemática não melhora o desempenho do gráfico de  $T^2$ . Quando a causa especial modifica a média da variável autocorrelacionada ( $a; b; \delta_1; \delta_2$ ) =  $(0; 0,2; 0; 1,0)$  e  $\rho=0$ , o gráfico de  $T^2$  requer, em média, 16,19 amostras para sinalizar. Esse número diminui para  $ARL=5,55$  quando há alta correlação ( $\rho=0,7$ ). Nesse último caso, aplicando a estratégia de amostragem sistemática ( $s=2$ ) os ARLs, diminuem, respectivamente, de 10,26 para 3,63, veja a Figura 3. Para o cenário em que  $b=0,7$ , consulte a Figura 4.



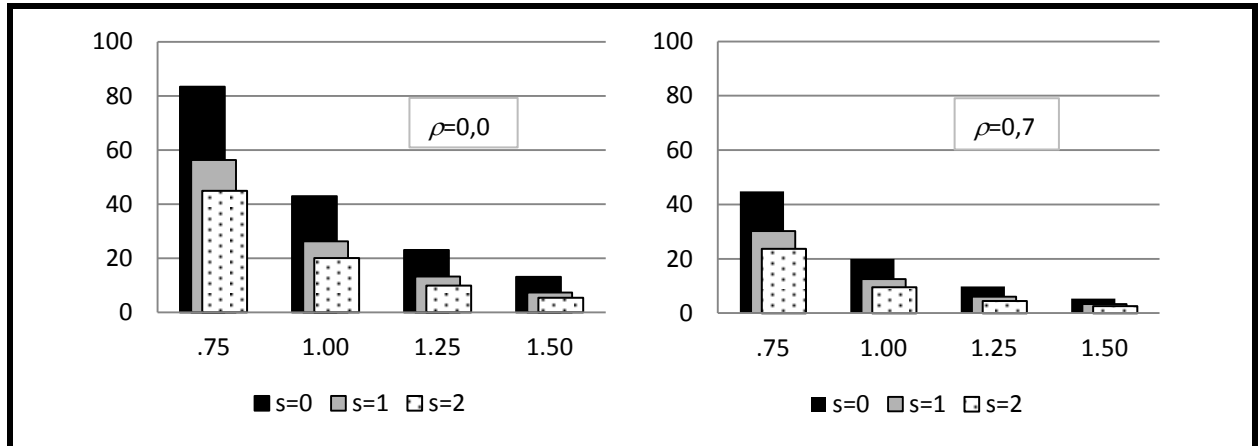


Figura 3: ARL do gráfico de  $T^2$ ;  $a=0$ ;  $b=0,5$ ;  $\delta_1=0$ ;  $\delta_2 \in (0,75;1,00;1,25;1,50)$ .

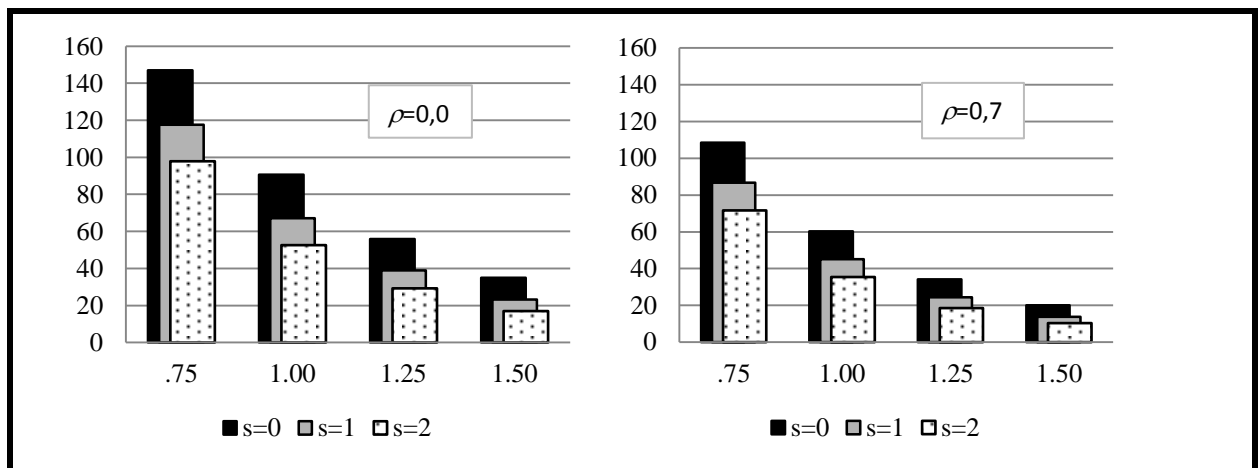


Figura 4: ARL do gráfico de  $T^2$ ;  $a=0$ ;  $b=0,7$ ;  $\delta_1=0$ ;  $\delta_2 \in (0,75;1,00;1,25;1,50)$ .

#### 4. CONCLUSÃO

No presente artigo considerou-se a estratégia de amostragem sistemática para reduzir o efeito da correlação e da autocorrelação no desempenho do gráfico  $T^2$  de Hotelling. A estratégia aumenta o desempenho do gráfico quando a causa especial que atua no processo modifica o valor médio da variável autocorrelacionada. Se as variáveis são autocorrelacionadas e ambas são afetadas por causas especiais que atuam modificando seus valores médios, a correlação, em geral, diminui o desempenho do gráfico.

#### 5. REFERÊNCIAS

- ALWAN, L. C. & ROBERTS, H. V. Time-Series Modeling for Statistical Process Control. Journal of Business & Economic Statistics, v. 6, n. 1, 1988, pp. 87-95.
- BOLFARINE, H. & BUSSAB, W. O. Elementos de amostragem. 1ª Ed. São Paulo: Editora Blucher. p. 274. 2005.
- CHAN, L. K. & LI, G. Y. A multivariate control chart for detecting linear trends. Communications in Statistics—Simulation and Computation, v. 23, 1994, pp. 997-1012.
- CHARNES, J. M. Tests for special causes with multivariate autocorrelated data. Computers and Operational Research, v. 22, 1995, pp. 443-453.

- COSTA A. F. B. & CLARO F. A. E.** Double sampling  $\bar{X}$  control chart for a first order autoregressive-moving average process model. *The International Journal of Advanced Manufacturing Technology*, v. 39, 2008, pp. 521-542.
- COSTA, A. F. C. & CASTAGLIOLA, P.** Effect of measurement error and autocorrelation on the  $\bar{X}$  chart. *Journal of Applied Statistics*, v. 38, n.4, 2011, pp. 661-673.
- ENGLISH, J. R.; LEE, S; MARTIN T.W. & TILMON, C.** Detecting changes in autoregressive processes with  $\bar{X}$  and EWMA charts. *IIE Transactions*, v. 32, n.12, 2000, pp. 1103-1113.
- FALTIN, F.; MASTRANGELO, C.M.; RUNGER, G.C. & RYAN, T.P.** Considerations in the monitoring of autocorrelated and independent data. *Journal of Quality Technology*, v. 29, n. 2, 1997, pp. 131-133.
- FRANCO, B. C.; CASTAGLIOLA, P.; CELANO, G. & COSTA A. F. B.** A new sampling strategy to reduce the effect of autocorrelation on a control chart. *Journal of Applied Statistics*. 2013. DOI:10.1080/02664763.2013.871507.
- FRANCO, B. C.; CELANO, G.; CASTAGLIOLA, P. & COSTA A. F. B.** Economic design of Shewhart control charts for monitoring autocorrelated data with skipping sampling strategies. *International Journal of Production Economics*, v. 151, 2014, pp. 121-130.
- GHOUBABI, M. E. & LIMAM, M.** Residual Responses to Change Patterns of Autocorrelated Processes. *Journal of Applied Statistics*, v. 34, n. 7, 2007, pp. 785-798.
- HOTELLING, H.** Multivariate quality control – illustrated by the air testing of samples bombs sights, in techniques of Statistical Analysis, C. Eisenhard, M.W. Hastay and W.A. Wallis, eds., Mc Graw Hill, New York, 1947, pp. 113-184.
- HUANG, X.; BISGAARD, S. & XU, N.** Model-based Multivariate Monitoring Charts for Autocorrelated Processes. *Quality and Reliability Engineering International*. 2013. DOI: 10.1002/qre.1506
- HUANG, X.; XU, N. & BISGAARD, S.** A Class of Markov Chain Models for Average Run Length Computations for Autocorrelated Processes. *Communications in Statistics - Simulation and Computation*, v. 42, n.7, 2013, pp. 1495-1513.
- HWARNG, H. B. & WANG, Y.** Shift detection an source identification in multivariate autocorrelated process. *International Journal of Production Research*, v. 48, 2010, pp. 835-859.
- KALGONDA, A. A. & KULKARNI, S. R.** Multivariate quality control chart for autocorrelated processes. *Journal of Applied Statistics*, v. 31, 2004, pp. 317-327.
- LOWRY, C. A. & MONTGOMERY, D.C.** A review of multivariate charts. *IIE Transactions*, v. 27, 1995, pp. 800-810.
- MONTGOMERY, D. C. & MASTRANGELO, C. M.** Some Statistical Process Control Methods for Autocorrelated Data. *Journal of Quality Technology*, v. 23, n. 3, 1991, pp. 179-193 with discussion on pp. 194-204.
- PAN, X. & JARRETT, J.** Using vector autoregressive residuals to monitor multivariate processes in the presence of serial correlation. *International Journal of Production Economics*, v. 106, 2007, pp. 204-216.
- PAN, X. & JARRETT, J.** Why and how to use vector autoregressive models for quality control: the guideline and procedures. *Quality and Quantity*, v. 46, n. 3, 2011, pp. 935-948.
- RUNGER, G. C.; WILLEMAIN, T. R. & PRABHU, S. S.** Average Run Lengths for CUSUM Control Charts Applied to Residuals. *Communications in Statistics – Theory and Methods*, v. 25, 1995, pp. 273-282.
- TSENG S. & ADAMS B. M.** Monitoring autocorrelated processes with an exponentially weighted moving average forecast. *Journal of Statistical Computation and Simulation*, v. 50, n. 3-4, 1994, pp. 187-195.

**APÊNDICE – Matriz de covariância cruzada de um vetor de médias amostrais considerando a estratégia de amostragem sistemática.**

Considere que  $n$  elementos de uma amostra são coletados em curto espaço de tempo e o intervalo de amostragem  $t$  é suficientemente grande para eliminar qualquer dependência entre amostras; o vetor de observações da  $j$ -ésima unidade é descrita por um modelo autoregressivo de primeira ordem:

$$\mathbf{X}_j = \sum_{i=1}^j \Phi^{j-i} \varepsilon_i ; j=1,2,\dots,n. \quad (A1)$$

sendo  $\varepsilon_i \sim N_2(\mu; \Sigma)$  com  $\Sigma = (a_{ij})_{2 \times 2}$ ;  $a_{ij} = 1$  se  $i = j$ ;  $a_{ij} = \rho$  se  $i \neq j$  e  $\Phi = \text{diag}(a;b)$ .

De acordo com o modelo autoregressivo,

$$\text{Var}(X_j) = (\Phi^{j-1})\Gamma(\Phi^{j-1})' + (\Phi^{j-2})\Sigma(\Phi^{j-2})' + \dots + \Sigma \quad (A2)$$

Se  $\Gamma = \Phi\Gamma\Phi' + \Sigma$  (KALGONDA e KULKARNI, 2004), a matriz de covariância cruzada do vetor de médias amostrais  $(\Gamma_{\bar{X}})$  é dada por:

$$\Gamma_{\bar{X}} = \text{Var} \left[ \frac{X_1 + X_2 + \dots + X_n}{n} \right] \quad (A3)$$

Quando as amostras são formadas usando a técnica de amostragem sistemática, a matriz de covariância cruzada de um vetor de médias amostrais  $(\Gamma_{\bar{X}})$  é dada por:

$$\Gamma_{\bar{X}} = \frac{1}{n^2} \text{Var} \left[ \sum_{i=0}^{n-1} X_{1+(s+1)i} \right] = \frac{1}{n^2} \left[ \begin{array}{l} \left( I + \Phi^{(s+1)} + \Phi^{2(s+1)} + \dots + \Phi^{(n-1)(s+1)} \right) \Gamma \left( I + \Phi^{(s+1)} + \Phi^{2(s+1)} + \dots + \Phi^{(n-1)(s+1)} \right)' + \\ \left( \Phi^s + \Phi^{2s-1} + \dots + \Phi^{(n-1)s-1} \right) \Sigma_e \left( \Phi^s + \Phi^{2s-1} + \dots + \Phi^{(n-1)s-1} \right)' + \\ \left( \Phi^{s-1} + \Phi^{2s} + \dots + \Phi^{(n-1)s} \right) \Sigma_e \left( \Phi^{s-1} + \Phi^{2s} + \dots + \Phi^{(n-1)s} \right)' + \dots + \Phi \Sigma_e \Phi' + \Sigma_e \end{array} \right] \quad (A4)$$

Baseando-se nas equações (A1), (A2), (A3) e (A4), segue que:

$$\Gamma_{\bar{X}} = \frac{1}{n^2} \left\{ \left( \sum_{i=0}^{n-1} \Phi^{(s+1)i} \right) \Gamma \left( \sum_{i=0}^{n-1} \Phi^{(s+1)i} \right)' + \sum_{k=1}^{n-1} \left\{ \sum_{j=1}^{s+1} \left[ \left( \sum_{i=1}^k \Phi^{(s+1)i-j} \right) \Sigma_e \left( \sum_{i=1}^k \Phi^{(s+1)i-j} \right)' \right] \right\} \right\} \quad (A5)$$

Após manipulação algébrica, a equação (A5) é representada por:

$$\Gamma_{\bar{X}} = \left( \begin{array}{l} \zeta_1^2 = \frac{\sigma_X^2}{n} \left[ 1 + \frac{2}{n} \sum_{j=1}^{n-1} (n-j) a^{(s+1)j} \right] \quad \zeta_{12} = \frac{\sigma_{XY}}{n} \left[ 1 + \frac{1}{n} \sum_{j=1}^{n-1} (n-j) a^{(s+1)j} + \frac{1}{n} \sum_{j=1}^{n-1} (n-j) b^{(s+1)j} \right] \\ \zeta_{21} = \zeta_{12} \quad \zeta_2^2 = \frac{\sigma_Y^2}{n} \left[ 1 + \frac{2}{n} \sum_{j=1}^{n-1} (n-j) b^{(s+1)j} \right] \end{array} \right) \quad (A6)$$

sendo:  $\sigma_X^2 = \frac{1}{1-a^2}$ ;  $\sigma_Y^2 = \frac{1}{1-b^2}$ ;  $\sigma_{XY} = \frac{\rho}{1-ab}$