

# Aplicação do coeficiente $RV$ em Controlo Estatístico da Qualidade

Adelaide Maria Figueiredo

Faculdade de Economia e LIAAD-INESC TEC Porto, Universidade do Porto, *adelaide@fep.up.pt*

Fernanda Otilia Figueiredo

Faculdade de Economia da Universidade do Porto e CEAUL, Universidade de Lisboa, *otilia@fep.up.pt*

**Palavras-chave:** Carta de Controlo, Coeficiente  $RV$ , Controlo Estatístico da Qualidade, Monitorização de Processos, Simulação de Monte Carlo

**Resumo:** Em situações reais a avaliação da qualidade global de um produto ou de um serviço depende simultaneamente de várias características de qualidade, pelo que o desenvolvimento de cartas de controlo para dados multivariados é crucial. Vamos considerar a carta de controlo proposta em Figueiredo e Figueiredo [6], baseada no coeficiente  $RV$  definido em [5], para monitorizar a estrutura de covariâncias de um processo multivariado. Para processos normais bivariados, estudaremos o desempenho da carta através do ARL (*average run length*), analisaremos também a distribuição do coeficiente  $RV$  quando o processo está sob controlo e estudaremos ainda várias características da distribuição do RL (*run length*) quando o processo está fora de controlo.

## 1 Introdução

As cartas de controlo são as ferramentas usualmente utilizadas para a monitorização de processos em Controlo Estatístico da Qualidade. As cartas de controlo foram introduzidas por Shewhart em 1924 nos Bell Laboratories para a monitorização de processos industriais,

mas atualmente são aplicadas nas mais diversas áreas, entre elas, na Saúde e Medicina ([13]) e Genética, Ambiente e Finanças ([10]). As cartas de controlo são representações gráficas que têm por objetivo ajudar a tomar decisões sobre o estado do processo que está a ser monitorizado, isto é, ajudar a decidir se está sob controlo ou fora de controlo. Na literatura têm sido propostas várias cartas de controlo para monitorizar o vetor de valores esperados de um processo multivariado como por exemplo, a carta baseada na estatística  $T^2$  de Hotelling [8], e variantes desta carta. Adicionalmente, diversas cartas para controlar a variabilidade de um processo multivariado têm sido propostas, tais como a carta baseada na variância generalizada  $|S|$ , variantes desta carta ([1] e outras), cartas baseadas no máximo das variâncias amostrais ou no máximo das amplitudes amostrais das  $p$  características em estudo ([3], [4]), cartas propostas em [7], [11], [12], [14], entre outras. Existem também esquemas de controlo para monitorizar simultaneamente o vetor de valores esperados e a matriz de covariâncias do processo ([2], [15], etc.)

Neste trabalho iremos recorrer ao coeficiente  $RV$  (proposto em [5]) para desenvolver uma carta de controlo para monitorizar a estrutura de covariâncias associada a um conjunto de características de um processo de controlo multivariado. Para o efeito, com base no coeficiente  $RV$  iremos comparar a matriz de covariâncias das  $p$  variáveis em estudo, associada a um conjunto de amostras de referência retiradas do processo quando o processo está sob controlo com a matriz de covariâncias dessas variáveis obtida num novo instante de tempo. Prosseguindo o trabalho [6], iremos neste estudo explorar a distribuição do coeficiente  $RV$  no contexto referido, de modo a podermos decidir devidamente sobre o estado do processo, i.e., se está sob controlo ou fora de controlo. Note-se que a distribuição exata do coeficiente  $RV$  não é conhecida e apenas existem aproximações a esta distribuição em determinados casos particulares. Em seguida, iremos avaliar o desempenho do procedimento para dois processos normais bivariados, com base em várias características da distribuição do RL, a qual será previamente analisada para cada processo específico.

O trabalho está estruturado do seguinte modo: na secção 2 descrevemos a carta- $RV$ ; na secção 3 discutimos o desempenho da carta para processos normais bivariados.

## 2 Carta de controlo baseada no coeficiente $RV$

O coeficiente  $RV$  proposto por Escoufier [5] permite medir a semelhança entre duas matrizes semi-definidas positivas. Iremos utilizar este coeficiente como medida de semelhança entre duas matrizes de covariâncias na carta que vamos propor. O coeficiente  $RV$  entre as matrizes de covariâncias  $V_k$  e  $V_{k'}$  é definido por

$$RV(V_k, V_{k'}) = \frac{\langle V_k, V_{k'} \rangle_{HS}}{\|V_k\|_{HS} \|V_{k'}\|_{HS}} = \frac{Tr(V_k V_{k'})}{\sqrt{Tr(V_k^2) Tr(V_{k'}^2)}},$$

onde  $\langle V_k, V_{k'} \rangle_{HS} = Tr(V_k V_{k'})$  representa o produto escalar de Hilbert-Schmidt entre  $V_k$  e  $V_{k'}$ ,  $Tr$  representa o traço de uma matriz e  $\|V_k\|_{HS} = \sqrt{\langle V_k, V_k \rangle_{HS}} = \sqrt{Tr(V_k^2)}$  é a norma Hilbert-Schmidt de  $V_k$ . O coeficiente  $RV$  varia entre 0 e 1 e quanto mais próximo de 1 for o coeficiente  $RV$  mais semelhantes são as matrizes  $V_k$  e  $V_{k'}$ .

A estrutura de covariâncias de um processo multivariado pode estimar-se através de uma matriz de covariâncias compromisso obtida a partir de um conjunto de amostras de referência retiradas do processo quando está sob controlo. Em seguida, definimos a matriz de covariâncias compromisso como na metodologia STATIS ([9]).

Consideremos:  $K$  amostras de referência de dimensão  $n$  em que cada observação da amostra é descrita por  $p$  variáveis, i.e.,  $K$  quadros de dados  $X_{n \times p}$  recolhidos em  $K$  instantes de tempo diferentes, quando o processo está sob controlo; as matrizes de covariâncias associadas a estes quadros de dados.

A matriz de covariâncias compromisso é definida como uma média

ponderada das  $K$  matrizes de covariâncias  $V_k$ :

$$V_{comp} = \sum_{k=1}^K \alpha_k V_k,$$

onde os pesos  $\alpha_k$  são os elementos do vetor próprio associado ao maior valor próprio da matriz  $Z$  dos coeficientes  $RV$  entre os  $V_k$ 's:

$$Z = \begin{pmatrix} 1 & RV(V_1, V_2) & \cdots & RV(V_1, V_K) \\ RV(V_2, V_1) & 1 & \cdots & RV(V_2, V_K) \\ \vdots & \vdots & \ddots & \vdots \\ RV(V_K, V_1) & RV(V_K, V_2) & \cdots & 1 \end{pmatrix}$$

A carta de controlo, que designamos por carta- $RV$ , é implementada do seguinte modo.

- Para cada nova amostra  $j$  retirada do processo, representamos o valor do coeficiente  $RV$  entre a matriz de covariâncias associada a esta amostra e a matriz de covariâncias compromisso:

$$RV(V_j, V_{comp}) = \frac{Tr(V_j V_{comp})}{\sqrt{Tr(V_j^2) Tr(V_{comp}^2)}}.$$

- O Limite de Controlo da carta,  $LC$ , é calculado a partir da taxa de falsos alarmes:

$$\alpha = P(RV < LC | \text{processo está sob controlo}).$$

Como a distribuição exata do coeficiente  $RV$  é desconhecida, fixamos o limite de controlo  $LC$  num percentil empírico de ordem  $\alpha$  da distribuição por amostragem do coeficiente  $RV$ , quando o processo está sob controlo.

- Se  $RV(V_j, V_{comp}) < LC$ , considera-se que o processo está fora de controlo. Caso contrário, considera-se que o processo está sob controlo.

### 3 Desempenho da carta para um processo normal bivariado

Nesta secção, vamos analisar o desempenho da carta- $RV$  para processos normais bivariados, usando como medida de eficiência o ARL - número esperado de amostras retiradas até à emissão de sinal, para uma taxa de falsos alarmes  $\alpha$  de 0.005. Para processos normais multivariados de dimensão superior ( $p = 3$  e  $p = 4$ ), pode ver-se o desempenho da carta- $RV$  em [6].

Nesta secção vamos também explorar a distribuição empírica do coeficiente  $RV$  quando o processo normal bivariado está sob controlo e estudar a distribuição do RL quando o processo está fora de controlo.

Gerámos processos normais bivariados  $N_2(\boldsymbol{\mu}, \Sigma)$ , com  $\boldsymbol{\mu} = (0,0)'$  e  $\Sigma = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix}$ . Considerámos diferentes estruturas da matriz de covariâncias quando o processo está sob controlo e quando está fora de controlo.

**Observação 3.1** *Note-se que se considerássemos outro vetor de valores esperados, obteríamos os mesmos resultados, visto que o desempenho da carta é independente de tal vetor.*

**Observação 3.2** *Note-se que a covariância entre as variáveis coincide com a correlação dado que as variâncias são unitárias.*

Para uma taxa de falsos alarmes  $\alpha = 0.005$ , estimamos o LC da carta; trata-se do quantil empírico de ordem 0.005 da distribuição sob controlo do coeficiente  $RV$ , obtido por simulação de Monte Carlo a partir de 100000 réplicas, recorrendo ao algoritmo descrito de seguida.

**Algoritmo 3.3** *Para cada  $i=1,100000$ , repetir os passos:*

1. Gerar 4 amostras de referência de dimensão  $n$  da distribuição  $N_2(\boldsymbol{\mu}, \Sigma)$ , supondo o processo sob controlo.

2. Determinar a matriz de covariâncias compromisso,  $V_{comp}$ .
3. Gerar uma nova amostra de dimensão  $n$  da distribuição  $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , supondo o processo sob controlo e determinar a respetiva matriz de covariâncias  $V_i$ .
4. Calcular  $RV(V_i, V_{comp})$ .

Determinar o quantil de ordem 0.005 dos valores  $RV$  obtidos.

A distribuição dos valores do coeficiente  $RV$ , quando o processo está sob controlo, encontra-se nas figuras 1 e 2, para a covariância nula e a covariância igual 0.75, respetivamente.

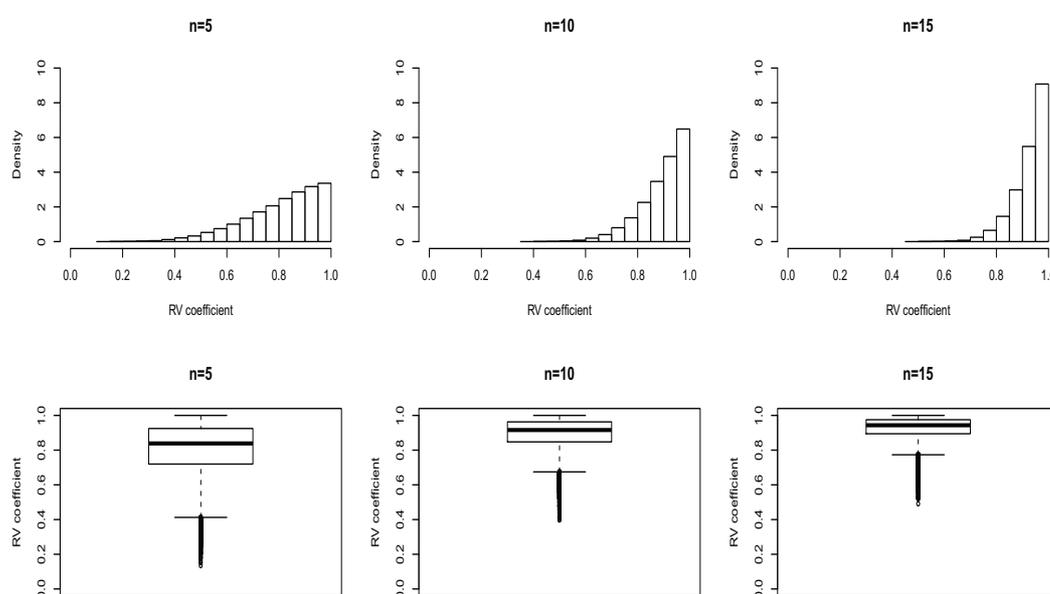


Figura 1: Distribuição do coeficiente  $RV$  quando o processo está sob controlo (covariância nula), para  $n = 5, 10, 15$

Observa-se que a distribuição do coeficiente  $RV$  é assimétrica negativa, predominando os valores mais elevados do coeficiente  $RV$ , isto é, mais próximos de 1. Observa-se ainda, em qualquer dos casos, a existência de bastantes *outliers* inferiores na distribuição do coeficiente  $RV$ .

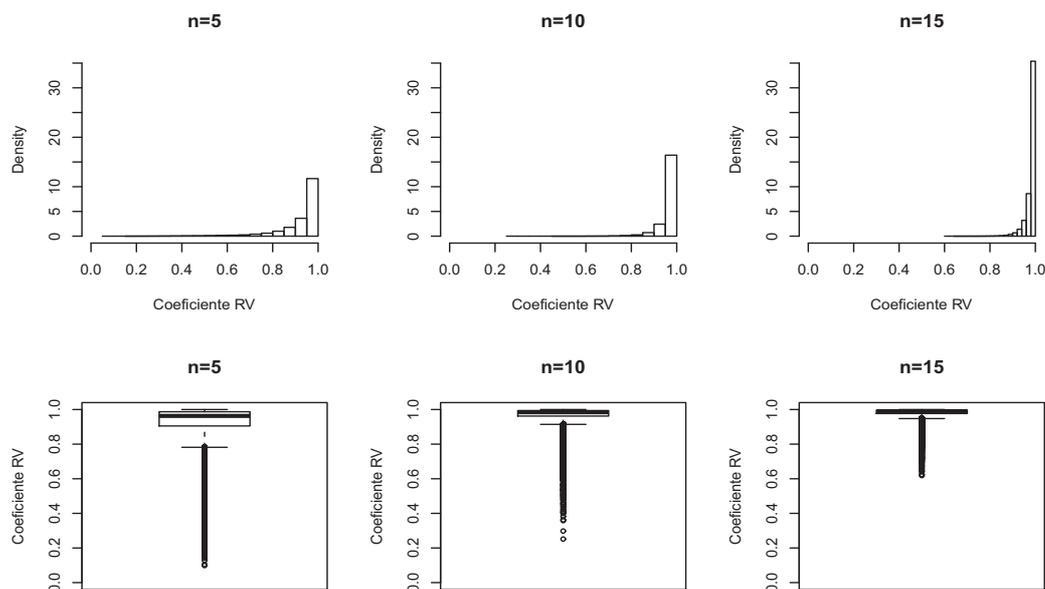


Figura 2: Distribuição do coeficiente  $RV$  quando o processo está sob controlo (covariância igual a 0.75), para  $n = 5, 10, 15$

A distribuição do coeficiente  $RV$  está cada vez mais concentrada à direita (a dispersão diminui) à medida que aumenta a dimensão da amostra, isto é, o coeficiente  $RV$  toma valores cada vez mais próximos de 1 à medida que  $n$  aumenta. Assim, o Limite de Controlo da carta aumenta à medida que aumenta a dimensão da amostra.

Como medida de eficiência da carta, usamos o ARL. Para uma taxa de falsos alarmes  $\alpha = 0.005$ , estimámos o ARL a partir de 10000 réplicas para diferentes estruturas da matriz de covariâncias, recorrendo ao algoritmo descrito a seguir.

**Algoritmo 3.4** Para cada  $i=1, \dots, 10000$  (réplicas)

1. Repetir os passos abaixo até que a carta- $RV$  emita sinal.
  - (a) Gerar 4 amostras de referência de dimensão  $n$  da distribuição  $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  com o processo sob controlo.

- (b) Calcular  $V_{comp}$ .
- (c) Gerar uma amostra da distribuição  $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  com o processo fora de controlo.
- (d) Calcular  $RV(V_i, V_{comp})$  e comparar com  $LC$  previamente estimado, usando o algoritmo 3.3.

2. Registrar o número de amostras até à emissão de sinal,  $RL_i$ .

Calcular a média dos valores  $RL$ , i.e., a estimativa de  $ARL$ .

Na Tabela 1 apresentamos os resultados obtidos para  $n = 5, 10, 15$  quando  $\sigma_{12} = 0$  ou  $\sigma_{12} = 0.75$  sob controlo. As estimativas de  $ARL$  sob controlo estão a negrito.

Tabela 1: Estimativas de  $LC$  e  $ARL$  para  $n = 5, 10, 15$ , sendo  $\sigma_{12} = 0$  ou  $\sigma_{12} = 0.75$  quando o processo está sob controlo

$\sigma_{12}=0$ sob controlo				$\sigma_{12}=0.75$ sob controlo			
$n$	5	10	15	$n$	5	10	15
$LC$	0.360	0.593	0.698	$LC$	0.390	0.747	0.863
$\sigma_{12}$	$ARL$			$\sigma_{12}$	$ARL$		
-0.95	31.2	6.3	2.5	0.75	<b>178.4</b>	<b>204.9</b>	<b>180.3</b>
-0.5	122.3	48.2	22.4	0.5	39.7	16.6	9.4
-0.3	168.7	105.7	66.9	0.3	18.1	5.6	3.1
0	<b>198.6</b>	<b>188.3</b>	<b>193.4</b>	0.1	9.7	2.8	1.6
0.3	161.9	105.7	66.9	0	7.2	2.1	1.8
0.5	121.8	48.8	22.6	-0.3	3.5	1.8	1.1
0.75	68.0	15.8	6.0	-0.5	2.3	1.1	1.0
0.95	31.9	6.3	2.5	-0.75	1.4	1.0	1.0

A partir da Tabela 1, podem tirar-se as seguintes conclusões:

- O limite de controlo da carta e o  $ARL$  dependem da dimensão da amostra e da estrutura da matriz de correlações.

- O ARL quando o processo está sob controle é elevado e aproximadamente igual ao valor esperado  $\alpha^{-1} = 200$ . Quando o processo está fora de controle, o ARL diminui rapidamente à medida que a dimensão da amostra aumenta.
- Se a correlação é nula ou próxima de 0 sob controle, a carta deteta a existência de correlações positivas ou negativas, sendo as maiores correlações em valor absoluto mais facilmente detetáveis.
- Se a correlação é elevada e positiva sob controle, a carta-*RV* deteta diminuições na correlação, correlações negativas e também correlações nulas. No entanto, a carta-*RV* não deteta atempadamente aumentos de  $\sigma_{12}$  quando o seu valor alvo é igual a 0.75. A carta é mais sensível a detetar correlações elevadas negativas.

A distribuição do RL quando o processo está fora de controle (covariância 0.3, 0.5 e 0.95) e a covariância é nula sob controle é apresentada para  $n = 5$  (Figura 3) e  $n = 15$  (Figura 4). Considerámos outros valores da covariância quando o processo está fora de controle, como por exemplo, 0.75, e os resultados foram semelhantes aos apresentados nas figuras 3 e 4. Resultados adicionais para o caso em que a covariância sob controle é 0.75 permitiram tirar conclusões análogas às obtidas a partir das figuras 3 e 4.

Em qualquer um dos casos apresentados, a distribuição do RL é assimétrica positiva, pelo que há predominância de valores baixos do RL. Verifica-se que a mediana do RL é inferior ao valor médio do RL, i.e., ao ARL. Observa-se a existência de *outliers* superiores na distribuição do RL em qualquer uma das situações em que o processo está fora de controle. À medida que nos afastamos da estrutura de correlações sob controle, a dispersão da distribuição do RL diminui.

**Observação 3.5** *Se a matriz de covariâncias compromisso  $V_{comp}$  é conhecida a priori, a distribuição do RL é geométrica, mas se estimarmos os parâmetros, a distribuição do RL não é geométrica.*

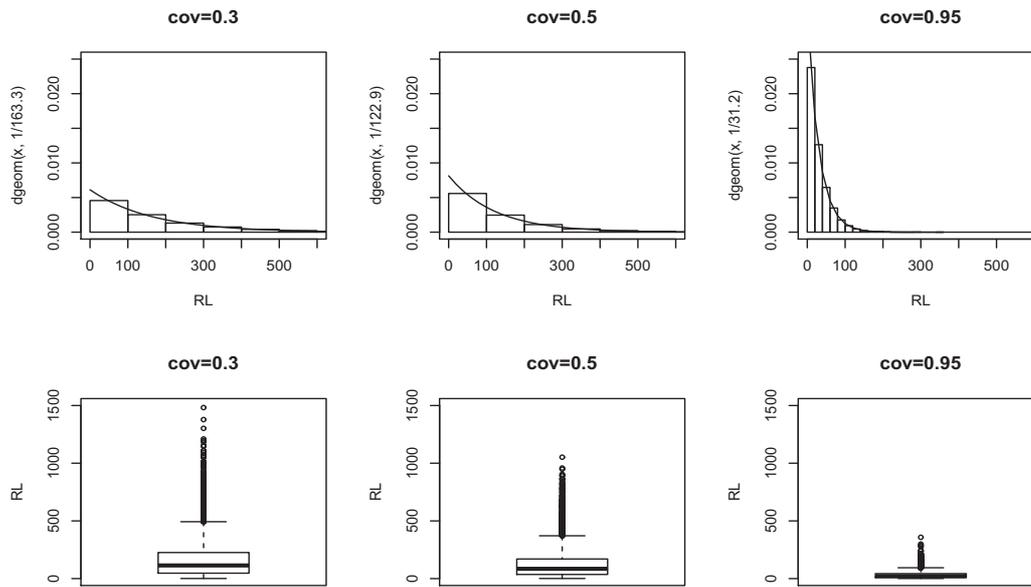


Figura 3: Distribuição do RL quando o processo está fora de controle, para  $n = 5$  e covariância sob controle nula.

*Neste caso, estimamos os parâmetros porque calculamos a matriz de covariâncias compromisso, mas a distribuição geométrica parece ajustar-se bem ao RL (ver figuras 3 e 4), pelo que se pode calcular a mediana e outros quantis da distribuição do RL a partir da distribuição geométrica.*

Os resultados de simulação obtidos para processos normais bivariados sugerem que a carta- $RV$  permite detetar facilmente alterações na correlação entre as variáveis, podendo constituir assim uma técnica de monitorização muito útil numa grande variedade de aplicações industriais.

Neste trabalho considerámos apenas dois valores da covariância sob controle  $\sigma_{12} = 0$  e  $\sigma_{12} = 0.75$ , mas em [6] foram considerados outros valores de  $\sigma_{12}$  e os resultados obtidos foram semelhantes aos apresentados aqui.

Para processos normais com  $p = 3$  e  $p = 4$  (ver [6]), os resultados

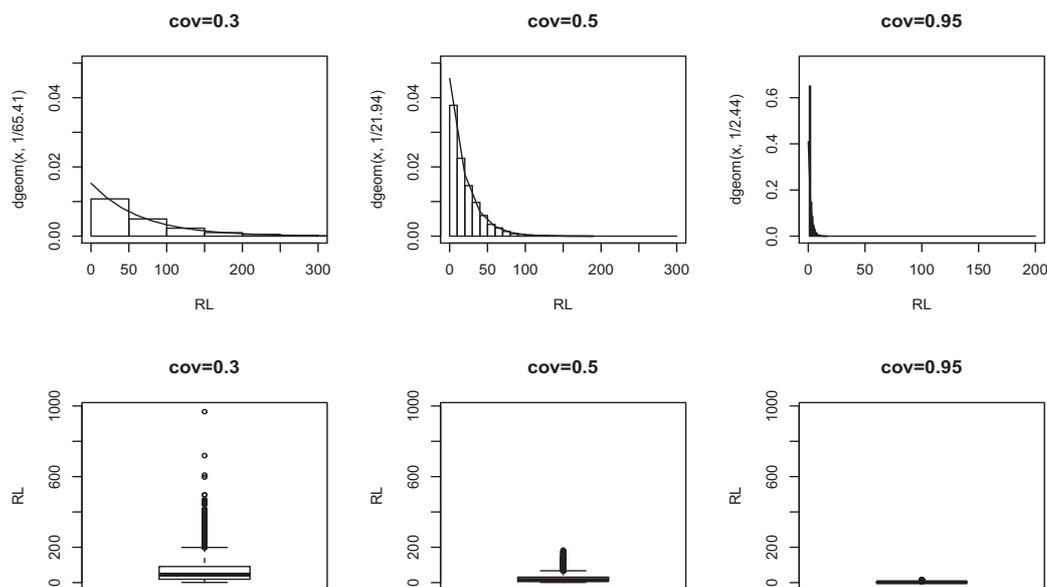


Figura 4: Distribuição do RL quando o processo está fora de controle, para  $n = 15$  e covariância sob controle nula.

são análogos aos obtidos para  $p = 2$  e permitem tirar conclusões que apoiam as referidas neste trabalho.

Referimos ainda que a carta- $RV$  constitui uma contribuição útil para as cartas existentes na literatura para a monitorização da matriz de covariâncias.

## Agradecimentos

Este trabalho é financiado por Fundos Nacionais através da FCT (Fundação para a Ciência e a Tecnologia) no âmbito dos projetos UID/EEA/50014/2013 e PEst-OE/MAT/UI0006/2014.

## Referências

- [1] Alt, F.B. (1985). Multivariate quality control. In Kotz, S., Johnson, N.L. (eds.): *Encyclopedia of Statistical Sciences* (Vol. 6), 111-122. Wiley, New York.
- [2] Chen, G., Cheng, S.W. and Xie, H. (2005). A new multivariate control chart for monitoring both location and dispersion, *Communications in Statistics - Simulation and Computation* 34, 203-217.
- [3] Costa, A.F.B. and Machado, M.A.G. (2008a). A new chart based on sample variances for monitoring the covariance matrix of multivariate processes, *International Journal of Advanced Manufacturing Technology* 41, 770-779.
- [4] Costa, A.F.B. and Machado, M.A.G. (2008b). A new multivariate control chart for monitoring the covariance matrix of bivariate processes, *Communications in Statistics - Simulation and Computation* 37, 1453-1465.
- [5] Escoufier, Y. (1973). Le traitement des variables vectorielles. *Biometrics* 29, 751-760.
- [6] Figueiredo, A. and Figueiredo, F. (2014). Monitoring the variability of a multivariate normal process using STATIS. In Gilli, M., Gonzalez-Rodriguez, G., Neto-Reyes, A. (eds.): *Proceedings of COMPSTAT 2014*, 443-450.
- [7] Hawkins, D. M. and Maboudou-Tchao, E. M. (2008). Multivariate exponentially weighted moving covariance matrix. *Technometrics* 50, 155-166.
- [8] Hotelling, H. (1947). Multivariate quality control, illustrated by the air testing of sample bombsights. In Eisenhart, C., Hastay, M.W., Wallis, W.A. (eds.): *Techniques of Statistical Analysis*, 111-184. McGraw Hill, New York.
- [9] Lavit, C., Escoufier, Y., Sabatier, R. and Traissac, P. (1994). The ACT (Statis method), *Computational Statistics and Data Analysis* 18, 97-119.
- [10] Stoumbos, Z.G., Reynolds, Jr., M., Ryan, T.P. and Woodall, W.H. (2000). The state of statistical process control as we proceed into 21st century, *Journal of the American Statistical Association* 95, 992-998.

- [11] Tang, P.F., Barnett, N.S. (1996a). Dispersion control for multivariate processes. *Australian Journal of Statistics* 38, 235-251.
- [12] Tang, P.F., Barnett, N.S. (1996b). Dispersion control for multivariate processes - some comparisons. *Australian Journal of Statistics* 38, 253-273.
- [13] Woodall, W.H. (2006). The use of control charts in health-care and public-health surveillance, *Journal of Quality Technology* 38, 89-104.
- [14] Yeh, A.B., Lin, D.K.-J., McGrath, R.N. (2006). Multivariate control charts for monitoring covariance matrix: a review. *Quality Technology and Quantitative Management* 3, 415-436.
- [15] Zhang, G. and Chang, S.I. (2008). Multivariate EWMA control charts using individual observations for process mean and variance monitoring and diagnosis, *International Journal of Production Research* 46, 6855-6881.