ESTUDIO DE LA CARTA EWMA EN PRESENCIA DE DATOS AUTO CORRELACIONADOS[1]

Joaquín González Borja
Maestría en Ciencias Estadísticas
Profesional en Matemáticas con énfasis en Estadística
Docente Universidad Católica Popular del Risaralda
jgb@ucpr.edu.co

Javier Páez Páez
Profesional en Matemáticas con énfasis en Estadística
Docente Universidad del Tolima
jppaez@gmail.com

Recibido Mayo 9 de 2008 – Aceptado Julio 08 de 2008

RESUMEN

Las cartas de control son aplicadas tradicionalmente a procesos industriales, asumiendo que la secuencia de las observaciones no presenta auto correlación, pero en la práctica esta condición es violada con frecuencia. La presencia de auto correlación tiene un serio impacto sobre el funcionamiento de la carta, causando un incremento sustancial en la frecuencia de falsas alarmas. El presente artículo introduce la forma de construcción e implementación de la carta de promedios móviles ponderados exponencialmente (EWMA), en presencia de datos auto correlacionados propuesta por Montgomery, D.C. y Mastrangelo, C.M. (1991), mediante una rutina de programación y su aplicación a un modelo de promedios móviles auto regresivos ARMA (p, q), obtenidos a través de simulación.

---

[1] Producto derivado del trabajo de grado "Estudio de la Carta Ewma en presencia de datos auto Correlacionados" presentado por Javier Páez Páez, bajo la dirección del Msc. Joaquín González Borja, para optar al título de profesional en Matemáticas con énfasis en Estadística otorgado por la Universidad del Tolima.

Palabras Clave: Cartas de control, promedios móviles ponderados exponencialmente, datos auto correlacionados, modelo de promedios móviles auto regresivos.

ABSTRACT

Control charts are traditionally applied to industrial processes, assuming that the observations sequence does not have any autocorrelation, but this assumption is frequently violated in practice. The presence of autocorrelation has a serious impact on the performance of control charts, causing a dramatic increase in the frequency of false alarms. This paper presents the construction and implementation form of the chart of exponentially weighted moving average EWMA, in presence of auto correlated data proposed by Montgomery, D.C. y Mastrangelo, C.M. (1991), through a programming routine and its application to a model of autoregressive-moving average model ARMA(p, q), obtained by simulation.

Key words: Control charts, exponentially weighted moving average, auto correlated data, autoregressive-moving average model.

1. INTRODUCCIÓN

Las cartas de control son una herramienta muy usada dentro del control estadístico de procesos. Son un instrumento útil para el análisis y el monitoreo de la estabilidad de un proceso a través de una característica de calidad, que sirve para detectar patrones anormales en cualquier lapso de tiempo. En la práctica se desea controlar el valor promedio de la característica de calidad así como su variabilidad. El control del valor promedio se hace a través de cartas donde se utilizan medidas de tendencia central, la más usual: la media. Cartas tales como $\bar{X}$, CUSUM (sumas acumuladas) y EWMA (promedios móviles ponderados exponencialmente) son usadas para tal propósito. Las cartas CUSUM y EWMA, son más eficientes en detectar pequeños cambios en la media que la carta $\bar{X}$. Para controlar la variabilidad, frecuentemente se usan medidas como la desviación estándar, como es el caso de la carta S y rangos para la carta R. Estos tipos de cartas de control estadístico son las más usadas en procesos en línea.

Es de nuestro interés la carta EWMA, sugerida por Roberts, S.W. (1959), la cual está basada en la estadística EWMA y se puede utilizar con subgrupos u observaciones individuales, donde el valor objetivo (límite central, LC) puede ser el estimador del valor promedio de la característica de calidad en estudio. Si los valores de la estadística EWMA, están dentro de la región conformada por los límites de control superior (LCS) e inferior (LCI), el proceso está bajo control, en caso contrario se dice que la variación del proceso se debe a causas asignables, lo cual exige una inmediata intervención a fin de determinar la causa y tomar medidas correctivas, llevando de nuevo el proceso a control.

La carta EWMA tiene como supuestos que las observaciones provienen de una distribución normal y no están auto correlacionadas. El segundo supuesto, frecuentemente es violado en la práctica, esto ocurre cuando las observaciones son tomadas en intervalos cortos de tiempo, muy comunes en procesos industriales químicos. La presencia de auto correlación tiene un profundo impacto en el funcionamiento de las cartas de control usuales, causando un incremento sustancial en la frecuencia de falsas alarmas.

Montgomery, D.C. y Mastrangelo, C.M. (1991), proponen una carta que tiene en cuenta la presencia de datos auto correlacionados en un proceso de control estadístico. La carta es construida, asumiendo que los datos tienen una estructura de auto correlación del tipo ARIMA (0, 1 ,1), para luego aplicarlo a la estadística EWMA, con el fin de obtener LC, LCI, LCS apropiados donde se tenga en cuenta la auto correlación de los datos.

Los autores mencionados, aseguran que si las observaciones del proceso son modelados por un miembro de la familia ARIMA (p, d, q) y además están auto correlacionadas positivamente y el proceso no cambia tan rápidamente en media, se puede usar la carta para datos auto correlacionados con una apropiada escogencia de lambda (constante de la estadística EWMA).

El presente artículo, está dirigido a realizar un estudio respecto a la construcción e ilustración con un ejemplo del uso de la carta propuesta por Montgomery, D.C. y Mastrangelo, C.M. (1991).

El artículo está organizado así: en la sección 2 se presentan algunas consideraciones teóricas básicas sobre las características con que trabaja la carta EWMA ante observaciones no auto-correlacionadas. Se incluye la teoría sobre el ajuste de los datos a un modelo ARMA (p, q) y las características de la carta EWMA usada para datos auto-correlacionados. En la sección 3 se realiza la construcción de la carta EWMA para datos no auto correlacionados y la carta EWMA para datos auto correlacionados, mediante una rutina de programación en el paquete R versión 2.2.0 [7], seguido de una ilustración del funcionamiento de las cartas a través de un modelo ARMA (p, q) simulado y finalmente en la sección 4 se dan las conclusiones y posibles temas de investigación derivados.

2. CONSIDERACIONES TEÓRICAS

2.1 CARTA EWMA PARA DATOS NO AUTO CORRELACIONADOS

La esencia de la carta EWMA es la de monitorear la media de un proceso, teniendo como característica primordial detectar cambios pequeños en la media del proceso. Se puede trabajar con subgrupos u observaciones individuales. Se asume que los datos

X_t provienen de una distribución normal (X_t □ $N(0,\sigma^2)$) y no están auto correlacionados. La estadística EWMA para observaciones individuales esta definida como:

$$Z_t = \lambda X_t + (1-\lambda)Z_{t-1},\ t = 1,2,3,...,n, \qquad (1)\ \ [6]$$

donde $Z_0 = \bar{X}$ (media muestral) y $0 < \lambda \le 1$.

Se tiene que la varianza de la estadística EWMA es: $Var(Z_t) = \frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2t}]\sigma^2$ y cuando $t \to \infty$, $Var(Z_t) = \frac{\lambda}{2-\lambda}\sigma^2$.

En [3] y [4], muestran que si las observaciones X_t están no correlacionadas, los límites de control de la carta EWMA son:

$$LCS = \bar{X} + K\sigma\sqrt{\lambda/[(2-\lambda)n]} \qquad (2)$$

y

$$LCI = \bar{X} - K\sigma\sqrt{\lambda/[(2-\lambda)n]} \qquad (3)$$

Usualmente λ=0.20 tomando K el valor de 3.

2.2 MODELOS ARMA (p, q) ESTACIONARIOS

Cuando los datos presentan estructura de auto correlación se pueden estructurar mediante modelos de promedios móviles auto regresivos:

$$ARMA\ (p,\ q) = \varphi_p(B)X_t = \theta_q(B)Z_t \qquad (4)$$

Donde $\varphi_p(B) = 1-\varphi_1 B - \varphi_2 B^2 - L - \varphi_p B^p$ es un polinomio auto regresivo de orden p, $\theta_q(B) = 1-\theta_1 B - \theta_2 B^2 - L - \theta_q B^q$ es el polinomio de promedios móviles de orden q, B es el operador de rezago y Z_t es una sucesión de “choques” aleatorios independientes e idénticamente distribuidos con media cero y varianza constante σ^2, conocido como ruido blanco [1].

Para ajustar la serie a un modelo ARMA adecuado debemos tener en cuenta que si la serie de tiempo de longitud n presenta tendencia debemos transformarla mediante una diferenciación de orden d. Posteriormente, se deben estimar los parámetros del modelo y los residuales son entonces definidos como la diferencia entre un valor observado y su pronóstico un paso adelante, re escalado por la raíz cuadrada del error cuadrático medio (ECM) de la predicción hecha [2]; esto, es: $\hat{W}_t = \dfrac{X_t - \hat{P}_{t-1}X_t}{\sqrt{ECM(\hat{P}_{t-1}X_t)}} \quad , \quad t = 1,2,3,...,n$

Donde $\hat{P}_{t-1}X$ es el mejor predictor lineal un paso adelante (bajo mínimo ECM) de X_t sobre $X_1, X_2, \cdots, X_{t-1}$. Si el modelo ajustado es el apropiado, $\hat{W}_t$ debe comportarse aproximadamente como un proceso ruido blanco.

2.3 CARTA EWMA PARA DATOS AUTO CORRELACIONADOS

La carta EWMA también puede ser usada en situaciones donde las observaciones están auto correlacionadas. Supóngase que las observaciones del proceso pueden ser modeladas por un ARIMA (0, 1, 1), cuyo modelo es $X_t = X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}$. La estadística EWMA con $\lambda = (1-\theta)$ es el pronóstico un paso adelante óptimo para este proceso. Es decir, si $\hat{X}_{t+1}(t)$ es el pronóstico para la observación en el periodo t+1 hecha al final del periodo t, entonces:

$$\hat{X}_{t+1}(t) = Z_t. \qquad (5)$$

La secuencia de errores de predicción un paso adelante es:

$$e_t = X_t - \hat{X}_t(t-1) \qquad (6)$$

Que se comportan aproximadamente como un ruido blanco. Ahora , si el proceso no es exactamente un ARIMA (0, 1, 1) pero en lugar es modelado por un miembro de la familia ARIMA y las observaciones del proceso están auto correlacionadas en forma positiva y el proceso en media no tiene cambios tan rápidamente, el EWMA con un apropiado valor de λ (el λ que minimice la suma de cuadrados de los errores de predicción un paso adelante SCE_λ) provee un excelente predictor un paso adelante [5]. Los límites de control de la carta sobre los errores satisfacen la siguiente condición de probabilidad: $P_r[-U_{\alpha/2}\sigma_p \le X_t - \hat{X}_t(t-1) \le U_{\alpha/2}\sigma_p] = (1-\alpha)$, donde $U_{\alpha/2}\sigma_p$ es el cuantil $\alpha/2$ superior de la distribución norma estándar. Pero, $P_r[\hat{X}_t(t-1) - U_{\alpha/2}\sigma_p \le X_t \le \hat{X}_t(t-1) + U_{\alpha/2}\sigma_p] = (1-\alpha)$.

Lo cual sugiere que si el EWMA es un conveniente predictor un paso adelante entonces se puede usar Z_t como la línea central sobre una carta de control para el periodo t+1 con:

$$LCS_{t+1} = Z_t + U_{\alpha/2}\sigma_p \qquad (7)$$

Y

$$LCI_{t+1} = Z_t - U_{\alpha/2}\sigma_p \qquad (8)$$

Y las observaciones X_{t+1} podrían ser comparadas con estos límites, determinando si el proceso está o no bajo control. La estimación de σ_p, es frecuentemente:

$$\hat{\sigma}_p = \sqrt{\frac{SCE_\lambda}{n}} \qquad (9)$$

Lo anterior descrito, es la forma de construcción de la carta EWMA para el caso de datos auto correlacionados.

3. CÓDIGO DE LA RUTINA DE PROGRAMACIÓN Y APLICACIÓN

Teniendo en cuenta el desarrollo teórico de la sección 2, procedemos a desarrollar una rutina de programación, con el fin de construir las cartas de control estadístico EWMA tanto para datos no auto correlacionados (programa 2), como para datos auto correlacionados (programa 3). De igual forma, se ilustra el uso de los programas en mención con 100 observaciones obtenidas por simulación estadística, considerando un modelo AR (0.8), con errores distribuidos normalmente con media cero y varianza uno, que se presenta en el Programa 1. La programación se realizó en el paquete de uso libre R, versión 2.2.0 [7].

3.1. PROGRAMA 1: SIMULACIÓN DE n DATOS AJUSTADOS A UN MODELO ARMA (p, q)

```
n<-100
Xt<-arima.sim(n=n, list(ar=c(0.8),ma=c(0)),sd=sqrt(1))                    ecuación (4)
Xt
```

3.2. PROGRAMA 2: CONSTRUCCIÓN DE LA CARTA EWMA PARA DATOS NO AUTO CORRELACIONADOS

```
landa<-0.20
Z0<-mean(Xt)
 l<-0
Zt<-matrix(0,n,1)
for(i in landa){
l<-l+1
Zt[1,l]<-i*Xt[1]+(1-i)*Z0
```

```
for(k in 2:n){
Zt[k,l]<-i*Xt[k]+(1-i)*Zt[k-1,l]}}                                   ecuación (1)
k<-3
sigma<-sd(Xt)
lcs<-Z0+(k*sigma*sqrt((landa/(2-landa))*(1-(1-landa)^2)))
lci<-Z0-(k*sigma*sqrt((landa/(2-landa))*(1-(1-landa)^2)))
for(i in 2:n){
lcs[i]<-Z0+(k*sigma*sqrt((landa/(2-landa))*(1-(1-landa)^(2*i))))    ecuación (2)
lci[i]<-Z0-(k*sigma*sqrt((landa/(2-landa))*(1-(1-landa)^(2*i))))}   ecuación (3)
TIEMPO=1:length(Xt)
plot(TIEMPO,Xt,xlim=c(0,(length(Xt)+2)),ylim=c(min(lci,min(Xt)-
2),max(lcs,max(Xt)+2)),type='l',xaxt='n')
lines(TIEMPO,Xt,lty=1)
lines(TIEMPO,lci,lty=2)
lines(TIEMPO,lcs,lty=3)
axis(1,1:(length(Xt)+2),cex.axis=0.5,las=0)
legend(1,min(Xt),c("Xt","LCI","LCS"),lty=1:3,lwd=1,cex=0.95)
abline(h=Z0,lty=3)
for(i in 1:length(Xt)){
if(Xt[i]>lcs[i])  temp<-1
else if(Xt[i]<lci[i])   temp<-1
else   temp<-20
points(TIEMPO[i],Xt[i],pch=temp)}
```

3.3. PROGRAMA 3: CONSTRUCCIÓN DE LA CARTA EWMA PARA DATOS AUTO CORRELACIONADOS

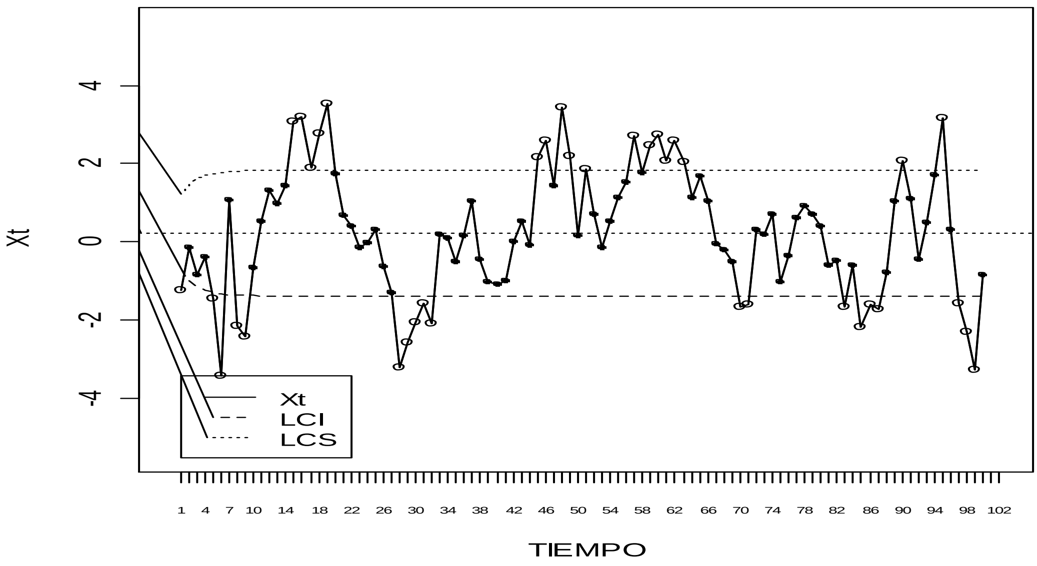
```
l<-0
s<-seq(0.001,1,0.001)
Z<-matrix(0,n,1000)
for(i in s){
l<-l+1
Z[1,l]<-i*Xt[1]+(1-i)*Z0
for(k in 2:n){
Z[k,l]<-i*Xt[k]+(1-i)*Z[k-1,l]}}
m<-0
error<-matrix(0,n,1000)
for(h in 1:1000){
m<-m+1
error[1,m]<-(Xt[1]-Z0)^2
for(j in 2:n){
error[j,m]<-(Xt[j]-Z[j-1,h])^2}}                    ecuación (5) y (6)
t<-0
sc<-matrix(0,1000,1)
for(i in 1:1000){
t<-t+1
sc[t,1]<-sum(error[1:n,i])}
plot (s,sc,"l",xlab="Lambda",ylab="SCE")
min2=function(X){
a<-min(X)
n<-length(X)
i<-1
```

```
while(X[i]!=a){i<-i+1}
I<-i
min2<-c(a,I)}
O<-min2(sc)
lambda<-0.001*O[2]
sigmap<-sqrt(O[1]/n)                                ecuación (9)
h<-0
zt<-matrix(0,n,1)
for(v in 1:n){
h<-h+1
zt[h,1]<-Z[v,O[2]]}
LCS<-rep(0,n)
LCI<-rep(0,n)
LCS[1]<-Z0+sigmap*1.96
LCI[1]<-Z0-sigmap*1.96
for(j in 2:n){
LCS[j]<-zt[j-1]+sigmap*1.96                         ecuación (7)
LCI[j]<-zt[j-1]-sigmap*1.96}                        ecuación (8)
TIEMPO<-1:length(zt)
plot(TIEMPO,zt,xlim=c(0,(length(zt)+2)),ylim=c(min(LCI,min(zt)-
2),max(LCS,max(zt)+2)),type='l',xaxt='n')
lines(TIEMPO,zt,lty=1)
lines(TIEMPO,LCI,lty=2)
lines(TIEMPO,LCS,lty=2)
axis(1,1:(length(zt)+2),cexaxis=0.5,las=0)
legend(1,min(LCI)-0.5,c("zt","LCI","LCS"),lty=1.3,lwd=1,cex=0.95)
```

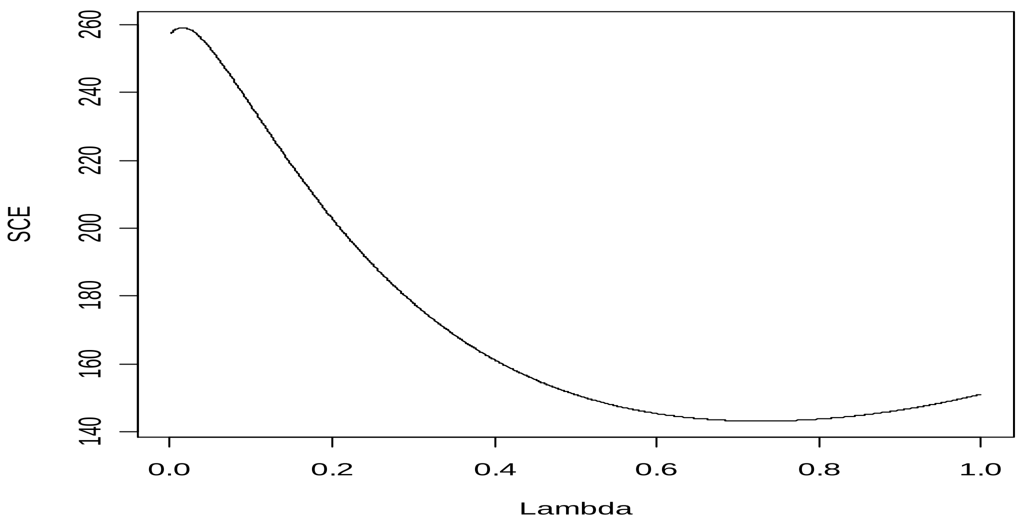
```
for(i in 1:length(Xt)){ if(Xt[i]>LCS[i])   temp<-8
else if(Xt[i]<LCI[i])   temp<-8
else    temp<-19
points(TIEMPO[i],Xt[i],pch=temp)}
for(i in 1:length(zt)){ if(zt[i]>LCS[i])   temp<-4
else if(zt[i]<LCI[i])   temp<-4
else   temp<-19
points(TIEMPO[i],zt[i],pch=temp)}
```

Ejecutando el programa 1, obtenemos una serie temporal de tamaño 100, que tiene una estructura de modelo auto regresivo AR (p), con p=0.8. Es decir, se logran datos altamente auto correlacionados en el tiempo en forma positiva. Con esta serie, se ejecuta el programa 2 que genera la gráfica de la carta de control estadístico EWMA expuesta en 2.1 y es dada en la Figura 1, se puede analizar que esta gráfica presenta un drástico incremento en señales fuera de control, seguidamente se presenta en la Figura 2, la obtención del valor de λ que minimiza la suma de cuadrados de los errores de predicción, λ=0.73, SCE_{λ}=146, requisito para implementar la carta de control para el caso de datos auto correlacionados, expuesta su teoría en 2.3, su rutina de programación en el programa 3 y la gráfica de la carta en la Figura 3, donde podemos ver que las señales fuera de control son realmente las de las posiciones 6, 7, 96 y 98 debido a causas asignables . Indicando que las señales fuera de control de la Figura 1, corresponden a falsas alarmas, al haberse utilizado para datos auto correlacionados.

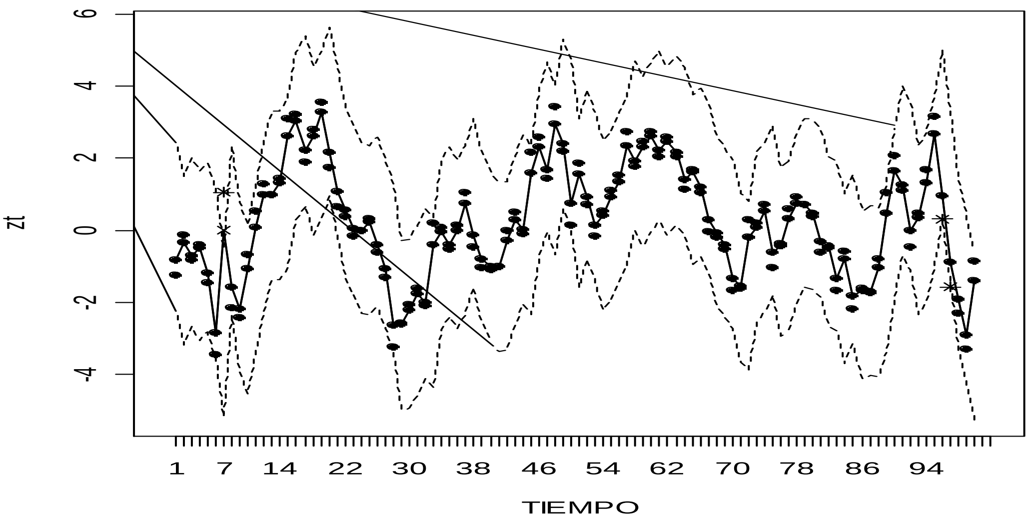
3.4. FIGURA 1: CARTA EWMA PARA DATOS NO AUTO CORRELACIONADOS, CON LOS DATOS SIMULADOS



3.5. FIGURA 2: OBTENCIÓN EN FORMA GRÁFICA DEL VALOR ÓPTIMO DE λ



3.6. FIGURA 3: CARTA EWMA PARA DATOS AUTO CORRELACIONADOS, CON LOS DATOS SIMULADOS



4. CONCLUSIONES

El trabajo anterior permite concluir que el uso de la carta de control EWMA propuesta por Roberts, S.W. (1959), en presencia de observaciones auto correlacionadas no tiene un buen funcionamiento debido a que genera un incremento sustancial en falsas alarmas, precisamente porque incumple uno de los supuestos de la carta. Por el contrario, la carta de control EWMA, propuesta por Montgomery, D.C. y Mastrangelo, C.M. (1991), tiene un buen funcionamiento ante observaciones auto correlacionadas, detectando las señales fuera de control, a las cuales se les pueden asignar la causa que lo produjo y entrar a corregir la situación. La carta con un buen comportamiento para el caso de observaciones auto correlacionadas, considera que si las observaciones del proceso están auto correlacionadas en forma positiva y el proceso en media no tiene cambios tan rápidamente, se pueden modelar las observaciones como un ARIMA (0, 1, 1). Desafortunadamente, si llega a incumplirse una de las dos condiciones la carta no funciona correctamente. Por tal razón, temas de investigaciones posteriores podrían ser:

- Buscar cartas de control alternativas donde su funcionamiento no sea afectado por el incumplimiento de una o las dos condiciones enunciadas anteriormente.
- Usar el modelo al que mejor se ajustan las observaciones auto correlacionadas, para la construcción de la carta EWMA, en lugar de utilizar el modelo ARIMA(0,1,1), analizando su funcionamiento.
- Realizar un estudio sobre las cartas de control multi variadas EWMA, en presencia de datos auto correlacionados en este caso series de tiempo multi variados.

5. BIBLIOGRAFÍA

[1] Box, G.E.P. y Jenkins, G.M. (1976). *Time Series Analysis, Forecasting, and Control*. Holden Day, Oakland, CA.

[2] Brockwell, P.J. y Davis, R.A. (1996). *Introduction to Time Series and Forecasting*. Springer-Verlag, New York.

[3] Hunter, J.S. (1986). The Exponentially Weighted Moving Average. *Journal of Quality Technology* **18**, pp. 203-209.

[4] Montgomery, D.C. (1991). *Introduction to Statistical Quality Control.* 2nd ed., John Wiley Sons, New York, NY.

[5] Montgomery, D.C. y Mastrangelo, C.M. (1991). *Some Statistical Process Control Methods for Auto correlated Data. Journal of Quality Technology* **23**, Nº 3. pp 179-204.

[6] Roberts, S.W. (1959). *Control Chart Test Based on Geometric Moving Averages*. *Technometrics* **1**, pp. 239-251.

[7] The R Development Core Team, (2005). R: *A Language and Environment for Statistical Computing*, version 2.2.0. R Foundation for Statistical Computing.