

# ALGUNAS CARTAS DE CONTROL BIVARIADAS PARA ATRIBUTOS

## SOME BIVARIATE CONTROL CHARTS FOR ATRIBUTES

CAROLINA OSPINA HINCAPIE

*Escuela de Estadística, Universidad Nacional de Colombia, Medellín, cospinah@unal.edu.co*

SERGIO YAÑEZ CANAL

*Escuela de Estadística, Universidad Nacional de Colombia, Medellín, syanez@unal.edu.co*

CARLOS MARIO LOPERA GÓMEZ

*Escuela de Estadística, Universidad Nacional de Colombia, Medellín, cmlopera@unal.edu.co*

Recibido para revisar Junio 18 de 2010, aceptado Octubre 28 de 2009, versión final Octubre 30 de 2009

**RESUMEN:** Muchos procesos industriales son de naturaleza multivariada dado que la calidad de un producto depende de más de una variable. El control multivariado de procesos captura la relación en las variables asociadas al proceso, si se ignora esta correlación y se utilizan gráficos de control univariados para cada variable por separado se puede concluir erróneamente acerca del estado del proceso. En variables continuas correlacionadas se han realizado muchas investigaciones, sin embargo se encuentran pocos trabajos que traten sobre atributos correlacionados. En este trabajo se comparan tres cartas de control para variables aleatorias binomiales bivariadas, correlacionadas entre sí, las cuales miden atributos. Las cartas son: La carta  $T^2$  de Hotelling basada en la aproximación de la distribución binomial multivariada a la distribución normal multivariada. La carta  $MNP$  la cual es una extensión de las cartas  $np$  univariadas, y la carta  $r$  que es una metodología no paramétrica basada en el índice de profundidad de Mahalanobis. La comparación se hace vía simulación utilizando como medida de comparación, la longitud promedio de racha (ARL). Dentro del trabajo se presenta un ejemplo aplicado de las metodologías para construir cartas de control para variables binomiales bivariadas en una empresa de telecomunicaciones. De los resultados se aprecia, en términos generales, que la carta  $MNP$  es la mejor tanto en control como fuera de control.

**PALABRAS CLAVE:** Cartas de Control Bivariadas,  $T^2$  de Hotelling, Variables Binomiales Bivariadas, Carta de Control  $MNP$ , Carta de Control  $r$ , Profundidad de Mahalanobis, ARL.

**ABSTRACT:** Many industrial processes are multivariate in nature since the quality of a product depends on more than one variable. The multivariate control of processes captures the relation between the variables associated with the process, if this correlation is ignored and univariate control charts are used for every variable separately is possible to conclude erroneously over the process status. In the continuous case, many researches have been done, however there are few works that aim to correlated attributes. In this work we compare three charts of control for correlated bivariate binomial random variables, which are associated with attributes. The charts are: Hotelling's  $T^2$  chart based on the approximation of the distribution binomial multivariate to the normal multivariate distribution.  $MNP$  chart which is an extension of univariate  $np$  chart, and  $r$  chart that is a non-parametric methodology based on the Mahalanobis's depth. The comparison is made through of simulation study using as a comparison measure the average run length (ARL). In this work we present an example of the used methodologies to construct control charts for bivariate binomial variables in a telecommunications company. The results shown in general terms that the  $MNP$  chart is the best in both control and out of control.

**KEYWORDS:** Control Bivariate Charts, Hotelling's  $T^2$ , Bivariate Binomial Variables,  $MNP$  Control Chart,  $r$  Control Chart, Mahalanobis's Depth, ARL

## 1. INTRODUCCIÓN

En el mundo de las organizaciones no todos los procesos que se desean medir se pueden describir mediante variables continuas [1], los atributos asociados a los procesos de servicio al cliente, como son: la satisfacción con la venta, la satisfacción con la facturación, entre otros; son variables discretas que deben ser controladas para la administración efectiva de los clientes. Por lo tanto se requiere de metodologías que permitan establecer estándares de cumplimiento también para este tipo de problemas [2]. Existen diversas situaciones en las cuales más de una variable se debe considerar de manera simultánea, por ejemplo la satisfacción con el servicio de internet de banda ancha y la disponibilidad de la red de acceso. El control multivariado de procesos es ideal para capturar la relación que existe entre las variables que determinan la calidad de un producto o servicio, por lo tanto si se ignora esta correlación y se utilizan cartas de control univariadas para cada variable por separado, se puede perder información importante [3]. Entre los métodos más usados para construir cartas de control multivariadas se encuentra la carta de control  $T^2$  de Hotelling, sin embargo este método se restringe, usualmente, al caso de distribuciones normales [4]. En variables continuas correlacionadas se han realizado muchas investigaciones, sin embargo se encuentran pocos trabajos que traten sobre atributos correlacionados [2]. En este trabajo se estudian variables aleatorias binomiales bivariadas correlacionadas, que miden atributos y se indaga sobre cuál metodología es adecuada para establecer un control efectivo sobre procesos que sean medidos con dichas variables. Se comparan tres métodos para construir cartas de control para variables binomiales bivariadas correlacionadas.

El primer método es la carta  $T^2$  de Hotelling o carta  $AN$ , basada en la aproximación de la distribución binomial multivariada a la distribución normal multivariada [5]. El segundo método es la carta  $MNP$  que consiste en la extensión multivariada de las cartas  $np$

univariadas [2], y la carta  $r$ , la cual emplea un método no paramétrico basado en el índice de profundidad de Mahalanobis [4]. En la sección 2 se revisan los conceptos básicos de la construcción de las cartas de control bivariadas.

La sección 3 presenta la metodología de simulación, se detallan los algoritmos empleados y se delimitan los escenarios a ser considerados. Los resultados de la comparación de las tres metodologías, mediante el comportamiento del ARL en control y fuera de control se presentan en la sección 4. La sección 5 presenta un ejemplo aplicado a la industria de telecomunicaciones, el cual ilustra el uso de las tres metodologías. Por último, la sección 6 muestra las conclusiones del trabajo.

## 2. CONCEPTOS BÁSICOS EN CARTAS DE CONTROL BIVARIADAS

### 2.1 Construcción de cartas de control

La construcción de cartas de control se puede dividir en las siguientes fases [6]:

- Fase I: se estiman los parámetros del proceso (estándares de calidad) y se hallan los límites de control.
- Fase II: para cada observación se calcula el estadístico asociado a la carta y se compara con los límites de control definidos en la fase I, con el fin de determinar si el proceso está en control o fuera de control.

### 2.2 Distribución Binomial Bivariada

Si se consideran dos variables aleatorias  $I_1$  e  $I_2$ , las cuales pueden tomar dos posibles valores: 1, si el ensayo resulta un éxito en la variable  $i$  ó 0, en otro caso,  $i = 1, 2$ ; La distribución conjunta de la variable  $(I_1, I_2)$  es Bernoulli bivariada, con probabilidad conjunta  $p_{rs} = \Pr [I_1 = r, I_2 = s]$ , para  $r, s = 0, 1$ ; cuya función generadora de probabilidad (fgp) conjunta de  $(I_1, I_2)$  es [7]:

$$\Pi(t_1, t_2) = E [t_1^{I_1}, t_2^{I_2}] = (t_1 t_2 p_{11} + t_1 p_{10} + t_2 p_{01} + p_{00}) \quad (1)$$

Usando la ecuación anterior es posible determinar los momentos de la distribución Bernoulli bivariada, en particular:  $E [I_1, I_2] = p_{11}$  y  $Cov [I_1, I_2] = p_{11} - p_{+1}p_{1+}$ , así el coeficiente de correlación es:

$$\rho [I_1, I_2] = \frac{p_{11} - p_{+1}p_{1+}}{\sqrt{[p_{+1}(1-p_{+1})][p_{1+}(1-p_{1+})]}} \quad (2)$$

Una propiedad interesante proveniente de fgp conjunta (1) es que las fgp's marginales son:

$$\Pi_1(t) = p_{0+} + p_{1+}t,$$

$$\Pi_2(t) = p_{+0} + p_{+1}t.$$

Donde, cada distribución marginal es una distribución Bernoulli [7].

Considere ahora una secuencia de  $m$  ensayos Bernoulli bivariados. Defina  $(X_1, X_2)$  como:

$$X_1 = \sum_{i=1}^m I_{1i} \text{ y } X_2 = \sum_{i=1}^m I_{2i}.$$

La distribución conjunta de  $(X_1, X_2)$  es binomial bivariada [8]. Luego, partiendo de (1), la fgp conjunta de  $(X_1, X_2)$ , es:

$$\begin{aligned} \Pi_{X_1, X_2}(t_1, t_2) &= E [t_1^{X_1}, t_2^{X_2}] = \left\{ E [t_1^{I_1}, t_2^{I_2}] \right\}^m \\ &= (t_1 t_2 p_{11} + t_1 p_{10} + t_2 p_{01} + p_{00})^m. \end{aligned}$$

Las fgp's de las marginales para  $X_1$  y  $X_2$  son:

$$\Pi_{X_1}(t) = [p_{0+} + p_{1+}t]^m,$$

$$\Pi_{X_2}(t) = [p_{+0} + p_{+1}t]^m.$$

Cada una de cuales corresponde a una distribución binomial con el mismo número de ensayos  $m$ , pero con probabilidades de éxito  $p_{1+}$  y  $p_{+1}$ , para  $X_1$  y  $X_2$ , respectivamente. En adelante  $p_{1+} = p_1$  y  $p_{+1} = p_2$ .

La correlación entre  $X_1$  y  $X_2$  es la misma que en el caso Bernoulli bivariado, que se mostró en (2).

Sea  $\underline{X}_t = \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix}$  un vector aleatorio binomial bivariado con parámetros  $m, p_1, p_2$  y  $\rho_{12}$ , tomado en el tiempo  $t$ . Si se tiene una muestra de vectores de las características de  $\underline{X}_t$ , recolectada a través de  $n$  puntos diferentes en el tiempo, entonces una matriz de datos binomiales bivariados tiene la siguiente representación:

$$\mathbf{X} = [\underline{X}_1 \ \underline{X}_2 \ \cdots \ \underline{X}_n] = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \end{bmatrix},$$

donde,  $X_{it}$  es el conteo de unidades no conformes en el atributo  $i$  ( $i = 1, 2$ ), tomado en el tiempo  $t$  ( $t = 1, 2, \dots, n$ ). En el contexto del artículo, la matriz  $\mathbf{X}$  constituye el conjunto de datos históricos en control.

En adelante se asumirá que  $m$  hace referencia al tamaño de la submuestra y  $n$  indica el número de observaciones tomadas en el tiempo o tamaño de muestra. Para más detalles sobre la distribución binomial bivariada [9].

### 2.3 Carta de control AN

Basados en un conjunto de datos histórico  $\mathbf{X}$ , y asumiendo independencia de las observaciones sucesivas en el tiempo, el vector de medias  $\mu$  y la matriz de covarianzas  $\Sigma$ , asociados a  $\mathbf{X}$ , se pueden estimar respectivamente mediante:

$$\bar{\mathbf{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \end{bmatrix} \text{ y } S = \begin{bmatrix} s_1^2 & s_{12} \\ s_{12} & s_2^2 \end{bmatrix}, \quad (3)$$

donde, para  $i = 1, 2$  se tiene que:

$$\bar{X}_i = \frac{\sum_{t=1}^n X_{it}}{n}, \quad s_i^2 = \frac{\sum_{t=1}^n (X_{it} - \bar{X}_i)^2}{n-1},$$

$$s_{12} = \frac{\sum_{t=1}^n (X_{1t} - \bar{X}_1)(X_{2t} - \bar{X}_2)}{n-1}.$$

La media muestral  $\bar{\mathbf{X}}$  será tomada como un estándar deseable para que el proceso esté en control.

El estadístico para la carta de control AN se define como:

$$G = (\underline{X}_t - \bar{\mathbf{X}})' S^{-1} (\underline{X}_t - \bar{\mathbf{X}}).$$

Para valores suficientemente grandes de  $m$  y  $n$ ,  $G$  tiene aproximadamente una distribución  $\chi^2_2$ , de manera que el límite superior de control (UCL<sub>AN</sub>) de esta carta de control es el valor  $\chi^2_{\alpha,2}$ , donde  $\alpha$  es el nivel de significancia especificado, mientras que el límite inferior de control (LCL<sub>AN</sub>) es cero.

El estimador  $G$  es equivalente al estadístico  $T^2$  de Hotelling tradicional en cartas de control multivariadas. En adelante el estadístico  $G$  se denotará como  $T^2$ .

## 2.4 Carta de control $MNP$

En la referencia [2] se presenta una metodología para desarrollar control estadístico de calidad multivariado para variables tipo atributos, donde se define el estadístico como la suma ponderada de los conteos de unidades no conformes de cada característica de calidad en una muestra.

En el caso bivariado, si  $P = (p_1, p_2)$  es el vector de fracciones de no conformidad, y  $\rho_{12}$  la correlación entre los dos atributos bajo estudio, entonces, para un periodo de tiempo fijo  $t$ , el estadístico para la carta  $MNP$  [2] es:

$$W_t = \frac{X_{1t}}{\sqrt{p_1}} + \frac{X_{2t}}{\sqrt{p_2}}, \quad (4)$$

con las siguientes propiedades:

$$E[W_t] = m(\sqrt{p_1} + \sqrt{p_2}),$$

$$V[W_t] = m \left[ 2 - p_1 - p_2 + 2\rho_{12}\sqrt{(1-p_1)(1-p_2)} \right]$$

Cuando el vector de fracciones de no conformidad del proceso  $P$  y la correlación entre atributos  $\rho_{12}$  no son conocidos, deben ser estimados mediante los datos históricos en control, contenidos en la matriz  $\mathbf{X}$ . Luego, el vector de fracciones de no conformidad se estima como [2]:

$$\bar{P} = \frac{\bar{X}}{m} = \begin{bmatrix} \bar{p}_1 \\ \bar{p}_2 \end{bmatrix}, \quad (5)$$

Siendo  $\bar{X}$  estimado como en (3).

La correlación entre los atributos  $\rho_{12}$  se estima también a partir de la matriz  $\mathbf{X}$ , como:

$$\hat{\rho}_{12} = \frac{\sum_{t=1}^n (X_{1t} - \bar{X}_1)(X_{2t} - \bar{X}_2)}{\sqrt{\sum_{t=1}^n (X_{1t} - \bar{X}_1)^2 \sum_{t=1}^n (X_{2t} - \bar{X}_2)^2}}. \quad (6)$$

Conociendo las estimaciones del vector  $P$  y de la correlación  $\rho_{12}$ , la línea central (CL<sub>MNP</sub>) y los límites de control (UCL<sub>MNP</sub>: límite superior y LCL<sub>MNP</sub>: límite inferior) de las cartas  $MNP$  son:

$$UCL_{MNP} = m(\sqrt{\bar{p}_1} + \sqrt{\bar{p}_2}) + 3\sqrt{\hat{V}(W_t)},$$

$$CL_{MNP} = m(\sqrt{\bar{p}_1} + \sqrt{\bar{p}_2}),$$

$$LCL_{MNP} = m(\sqrt{\bar{p}_1} + \sqrt{\bar{p}_2}) - 3\sqrt{\hat{V}(W_t)},$$

donde  $\hat{V}(W_t)$  se obtiene reemplazando las estimaciones muestrales de  $p_1, p_2$  y  $\rho_{12}$ . Si el estadístico  $W_t$  cae dentro de los límites de control, se puede concluir que el proceso bivariado está en control [2].

## 2.5 Carta de control $r$

La carta de control  $r$  se fundamenta en la noción de profundidad de los datos, la cual se basa en el hecho de que cualquier densidad de probabilidad distingue entre puntos *centrales* y *periféricos*. Las profundidades más grandes corresponden al centro de la distribución, mientras que las más pequeñas corresponden a regiones externas [3]. Una medida de profundidad muy utilizada es la profundidad de Mahalanobis, la cual se define para un vector aleatorio  $\underline{X}_t$ , que proviene de una distribución de referencia  $F$  en  $\mathbb{R}^2$ , como [4]:

$$MD_F(\underline{X}_t) = \frac{1}{[1 + (\underline{X}_t - \mu_F)' \Sigma_F^{-1} (\underline{X}_t - \mu_F)]}.$$

Donde  $\mu_F$  y  $\Sigma_F$  son respectivamente el vector de medias y la matriz de covarianzas de la distribución de referencia  $F$ . Si los parámetros de la distribución de referencia no son conocidos, entonces la versión muestral de  $MD_F(\underline{X}_t)$  es:

$$MD_{F_n}(\underline{X}_t) = \frac{1}{[1+(\underline{X}_t-\bar{X})'S^{-1}(\underline{X}_t-\bar{X})]}.$$

Donde  $\bar{X}$  y  $S$  son respectivamente el vector de medias y la matriz de covarianzas muestrales de los datos históricos en control  $\underline{X}$ , asociados a la **distribución empírica**  $F_n$  [10], y que son estimados mediante la expresión (3).

Sean  $F$  y  $G$  las funciones de distribución de dos poblaciones independientes bidimensionales. Suponga que la matriz  $\underline{X}$  es una muestra tomada de la distribución de referencia  $F$ , la cual está en control. Sea  $\underline{Y} = [\underline{Y}_1 \ \underline{Y}_2 \ \dots]$  una muestra aleatoria tomada desde una población que tiene distribución  $G$  (una nueva medida tomada del proceso). Para determinar si la muestra  $\underline{Y}$  está en control, se deben comparar las distribuciones  $F$  y  $G$ , mediante la siguiente prueba de hipótesis:

$$H_0 : G = F \quad \text{vs.}$$

$H_1$  : Hay un cambio en la posición y/o un incremento en la escala de  $G$  a  $F$ .

Para probar esta hipótesis, se usa el estadístico:

$$R_F(\underline{Y}_j) = \text{Pr}_F (MD_F(\underline{X}_t) \leq MD_F(\underline{Y}_j) | \underline{X}_t \sim F) \\ j = 1, 2, \dots$$

que caracteriza la distancia entre  $F$  y  $G$  con respecto a la profundidad de Mahalanobis. Bajo la hipótesis  $H_0$ , si  $MD_F(\underline{Y}_j)$  es continua, la distribución de  $R_F(\underline{Y}_j)$  definida anteriormente es uniforme en  $[0, 1]$  [11]. Cuando la distribución  $F$  es desconocida la versión muestral del estadístico  $R_F(\underline{Y}_j)$  es:

$$R_{F_n}(\underline{Y}_j) = \frac{\#\{MD_{F_n}(\underline{X}_t) \leq MD_{F_n}(\underline{Y}_j); t=1, 2, \dots, n\}}{n}.$$

Los límites de la carta  $r$  se basan en que bajo  $H_0$ ,  $R_{F_n}(\underline{Y}_j)$  es aproximadamente una distribución uniforme en  $[0, 1]$ , cuyo valor esperado es 0.5 ( $CL_r$ ). Como el valor de  $R_{F_n}(\underline{Y}_j)$  indica la probabilidad de que  $\underline{Y}_j$  sea un punto dentro del conjunto de datos histórico en control  $\underline{X}$ , entonces el límite de control inferior ( $LCL_r$ ) es  $\alpha$ . Donde,  $\alpha$  es la proporción de falsa alarma [3]. Por diseño el límite de control superior ( $UCL_r$ ) se asume con valor de uno.

En una carta de control  $r$  cuando los valores de  $R_{F_n}(\underline{Y}_j)$  están en la región  $\alpha$  (entre 0 y  $\alpha$ ) el proceso está fuera de control estadístico [10].

### 3. ESTUDIO DE SIMULACIÓN

Para analizar las tres metodologías para construir cartas de control para variables binomiales bivariadas (Cartas  $AN$ ,  $MNP$  y  $r$ ), se controlan factores que afectan su desempeño: a) Nivel de correlación entre los dos atributos  $\rho_{12}$ , b) Las proporciones de no conformidad ( $p_1, p_2$ ) y c) El tamaño de la submuestra  $m$ . Esto con el fin de analizar la rapidez de cada uno de los métodos en la detección de un punto fuera de control.

#### 3.1 Factores de simulación y sus niveles

Se consideraron tres niveles de correlación  $\rho_{12} = 0.8, 0.5$  y  $0.2$ , que corresponden a niveles de correlación fuerte, moderado y débil, respectivamente. Se seleccionaron tres niveles para las proporciones de no conformidad, así:  $p_i = 0.01, 0.1$  y  $0.3$ ,  $i = 1, 2$ ; que representan a proporciones de no conformidad baja, media y alta [1]. Los tamaños de submuestra  $m$  fueron calculados a partir de las siguientes situaciones:

- *Caso 1.* Tamaño de submuestra que favorece la aproximación de la distribución binomial bivariedad a la distribución normal bivariada.
- *Caso 2.* Tamaño recomendado por la referencia [2] para la construcción de cartas  $MNP$ .
- *Caso 3.* Tamaño basado en la condición de que al menos se encuentre una observación no conforme en la muestra [1].

#### 3.2 Generación de los datos

Los datos para todo el estudio de simulación son construidos a partir del algoritmo de Ong [12]. Este algoritmo genera variables binomiales bivariadas a partir de las proporciones de no conformidad de las marginales ( $p_1, p_2$ ) y la correlación entre las variables  $\rho_{12}$ .

La siguiente tabla resume los tamaños de submuestra calculados en los tres *casos* definidos en la sección 3.1, para las combinaciones de los factores  $p_1, p_2$  y  $\rho_{12}$ .

**Tabla 1.** Tamaños de submuestra para los escenarios de simulación.

**Table 1.** Subsample sizes for the simulation scenarios

$p_1$	$p_2$	$\rho_{12}$	Caso 1	Caso 2	Caso 3
0.3	0.3	0.8	50	10	10
0.3	0.1	0.8			
0.3	0.01	0.8			
0.1	0.1	0.8	100	30	30
0.1	0.01	0.8			
0.01	0.01	0.8	910	810	300
0.3	0.3	0.5	50	10	10
0.3	0.1	0.5			
0.3	0.01	0.5			
0.1	0.1	0.5	100	30	30
0.1	0.01	0.5			
0.01	0.01	0.5	910	670	300
0.3	0.3	0.2	50	10	10
0.3	0.1	0.2	100	15	
0.3	0.01	0.2			
0.1	0.1	0.2	100	30	30
0.1	0.01	0.2			
0.01	0.01	0.2	910	540	300

En algunas combinaciones de  $p_1, p_2$  y  $\rho_{12}$  (marcadas en la Tabla 1, con sombreado) donde los atributos tienen valores para las proporciones de no conformidad considerablemente diferentes, el algoritmo de Ong no converge. Adicionalmente, algunos tamaños de submuestra correspondientes a los Casos 2 y 3 resultaron iguales. Lo anterior lleva a considerar 23 posibles combinaciones diferentes de  $p_1, p_2$  y  $\rho_{12}$  con su tamaño de submuestra calculado.

En adelante se empleará la siguiente notación para los subíndices,  $i = 1, 2$ ; se refiere a los atributos,  $t = 1, 2, \dots, n$ ; indica el número de observaciones en el tiempo, y  $q = 1, 2, \dots, N$ ; es el número de simulaciones. En todos los escenarios de simulación se emplea la misma combinación de subíndices  $itq$ .

### 3.3 Metodología de simulación

Para cada escenario de simulación se generaron  $N = 5000$  simulaciones que representan los datos históricos que denotaremos  $X_{itq}$  con  $i = 1, 2$ ;  $t = 1, 2, \dots, n$ ; y  $q = 1, 2, \dots, N$ . Sobre los datos históricos se estiman (en Fase I) los parámetros del proceso (estándares de calidad) y los límites de control para cada una de las cartas descritas en la sección 2.

Con el objetivo de examinar la rapidez con que cada uno de los métodos detecta, en Fase II, desviaciones de los datos con respecto a la media. Para cada escenario de simulación se generaron  $N = 5000$  simulaciones en control denotadas  $Y_{itq}$ , las cuales se contaminan mediante la desviación estándar de cada atributo de la siguiente manera:

$$Y_{itqz} = Y_{itq} + C_{iz}$$

Donde, para la simulación  $q$  ( $q = 1, 2, \dots, N$ ),  $Y_{itqz}$  se refiere a la  $t$ -ésima observación contaminada del  $i$ -ésimo atributo,  $C_{iz}$  es la contaminación  $Z$ , para el atributo  $i$ , donde  $Z = 0, \pm 1, \pm 2, \dots, \pm 6$  son los niveles de contaminación tanto en dirección positiva (signo "+") como en dirección negativa (signo "-") a partir de la media del proceso en control. Las contaminaciones se resumen en la siguiente tabla.

**Tabla 2.** Niveles de contaminación  
**Table 2.** Contamination Level

Nivel de Contaminación $Z$	Atributo 1 $C_{1z}$	Atributo 2 $C_{2z}$
0	$0s_1$	$0s_2$
$\pm 1$	$\pm 1s_1$	$\pm 1s_2$
$\pm 2$	$\pm 1s_1$	$\pm 2s_2$
$\pm 3$	$\pm 1s_1$	$\pm 3s_2$
$\pm 4$	$\pm 2s_1$	$\pm 2s_2$
$\pm 5$	$\pm 2s_1$	$\pm 3s_2$
$\pm 6$	$\pm 3s_1$	$\pm 3s_2$

De la Tabla 2 se observa que  $Z = -2$ , representa contaminaciones para los atributos 1 y 2, en la dirección negativa de magnitud  $1s_1$  y  $2s_2$

respectivamente.  $s_1 = \sqrt{m p_1(1 - p_1)}$  y  $s_2 = \sqrt{m p_2(1 - p_2)}$  son las respectivas desviaciones estándar de cada atributo. Las contaminaciones se realizan en la misma dirección, pues la correlación se trabaja en valores positivos. Por último,  $Z = 0$ , hace referencia a los datos en control.

### 3.4 Construcción de las cartas de control

Para la construcción de las cartas de control primero se estiman los parámetros, luego se calculan los límites de control y por último se establece la medida de comparación.

El criterio de comparación utilizado fue la longitud promedio de racha (ARL) tanto en control como fuera de control, es decir, el número promedio de mediciones en la carta hasta encontrar una señal fuera de control. Lo ideal es encontrar una carta con un ARL en control grande y un ARL fuera de control pequeño.

Una forma general de calcular el ARL se obtiene al invertir la proporción de observaciones fuera de control, así:

$$ARL = \frac{1}{\hat{p}^*} \text{ con } \hat{p}^* = \frac{\sum_{q=1}^N \sum_{t=1}^n I_{tqZ}}{Nn}, \quad (7)$$

donde,  $I_{tqZ}$  se define apropiadamente para cada carta de control.

#### 3.4.1 Carta AN

*Estimadores de los parámetros:* Para cada simulación se estiman los parámetros mediante la expresión (3). Luego se promedian los valores obtenidos en las  $N$  simulaciones, así:

$$\bar{X} = \frac{\sum_{q=1}^N \bar{X}_q}{N}, \quad \bar{S} = \frac{\sum_{q=1}^N \bar{S}_q}{N}. \quad (8)$$

*Límite de control:* El límite de control depende de los factores  $m, p_1, p_2, \rho_{12}$  y  $n$ ; por lo tanto es necesario calcularlo de manera empírica, se

utiliza en este trabajo la metodología propuesta en [6], que se describe a continuación:

- Para cada uno de los datos históricos generados  $X_{tq}$ , se calcula el estadístico  $T_{tq}^2$ .
- En cada simulación  $q$  se calcula el máximo  $T_{tq}^2$ .
- El  $UCL_{AN}$  se obtienen como el percentil 95 de la distribución empírica del estadístico  $T^2$ , generada en el punto anterior.

*Medida de comparación:* Se utiliza el ARL, el cual se calcula como en la expresión (7), teniendo en cuenta que  $I_{tqZ}$  se define como:

$$I_{tqZ} = \begin{cases} 1, & \text{si } T_{tqZ}^2 > UCL_{AN} \\ 0, & \text{si } T_{tqZ}^2 \leq UCL_{AN} \end{cases}$$

#### 3.4.2 Carta MNP

*Estimadores de los parámetros:* en cada simulación se estiman las proporciones de no conformidad y la correlación entre atributos como en las expresiones (5) y (6). Luego se promedian estos valores para las  $N$  simulaciones, así:

$$\bar{P} = \left[ \frac{\sum_{q=1}^N \bar{p}_{1,q}}{\frac{\sum_{q=1}^N N}{N}} \right], \quad \hat{\rho}_{12} = \frac{\sum_{q=1}^N \hat{\rho}_{12,q}}{N}.$$

*Límites de control:* Empleando el mismo concepto de la carta AN para cada simulación  $q$  se calculan los límites de control  $UCL_{MNP,q}$  y  $LCL_{MNP,q}$ , obteniendo así una distribución empírica para los límites de control de la carta MNP. Luego, se estiman el  $UCL_{MNP}$  y el  $LCL_{MNP}$  como los percentiles 97.5 en cada caso.

*Medida de comparación:* En este caso, el cálculo del ARL utiliza en la expresión (7) la siguiente definición para  $I_{tqZ}$ :

$$I_{tqZ} = \begin{cases} 1, & \text{si } W_{tqZ} < LCL_{MNP} \text{ ó } W_{tqZ} > UCL_{MNP}, \\ 0, & \text{si } W_{tqZ} \geq LCL_{MNP} \text{ y } W_{tqZ} \leq UCL_{MNP}, \end{cases}$$

donde,  $W_{tqZ}$  es el valor del estadístico MNP resultante de aplicar la ecuación (4) en la simulación  $q$ , bajo un nivel de contaminación  $Z$ .

### 3.4.3 Carta *r*

*Estimadores de los parámetros:* Igual que para la carta *AN* en cada simulación se estiman los parámetros mediante la expresión (3) y se promedian los valores obtenidos en las *N* simulaciones, llegando a la expresión (8).

*Límite de control:* En cada simulación *q* se toma la mínima profundidad  $MD_{F_n}(\underline{X}_{tq})$ , obteniendo una distribución empírica de las mínimas profundidades de Mahalanobis. Luego el  $LCL_r$  se calcula como el percentil 5 de esta distribución empírica.

*Medida de comparación:* Para la carta *r* el cálculo del ARL utiliza en la expresión (7) la siguiente definición para  $I_{tqz}$ :

$$I_{tqz} = \begin{cases} 1, & \text{si } MD_{F_n}(\underline{Y}_{tqz}) < LCL_r, \\ 0, & \text{si } MD_{F_n}(\underline{Y}_{tqz}) \geq LCL_r \end{cases}$$

donde,  $MD_{F_n}(\underline{Y}_{tqz})$  es la profundidad de Mahalanobis para la *t*-ésima observación contaminada a nivel *Z*.

## 4. RESULTADOS

En la Tabla 3, que consta de tres filas y tres columnas, se resumen de manera gráfica los resultados del estudio de simulación cuando las proporciones de no conformidad son iguales. Las filas de la tabla hacen referencia a los casos de simulación y las columnas corresponden a las proporciones de no conformidad estudiadas. Cada celda en la tabla consta de tres figuras (cada una de ellas asociada a un nivel de correlación distinto) donde se muestran los ARL's obtenidos en función del nivel de contaminación *Z*. Por ejemplo, la celda ubicada en la fila 1, columna 3 presenta los resultados para proporciones de no conformidad iguales a 0.3, con tamaños de submuestra calculados para el caso 1 y en las tres correlaciones. De la Tabla 3 se puede concluir que la carta *MNP* presenta un comportamiento similar bajo cualquiera de las situaciones de simulación consideradas. En tal comportamiento se destaca que la carta *MNP* tiene valores de ARL menores o muy cercanos a

los de las otras cartas excepto cuando se tiene un nivel de contaminación  $Z = \pm 3$  y bajo una correlación alta de  $\rho_{12} = 0.8$  donde las cartas *AN* y *r* tienen un ARL apreciablemente menor. Cuando la contaminación se da en la dirección negativa, los ARL's obtenidos en las cartas *AN* y *r* presentan en la mayoría de los casos valores significativamente superiores a los de la carta *MNP*, lo cual se agrava para los tamaños de submuestra correspondientes a los casos 2 y 3. En este sentido el nivel de correlación actúa como un factor atenuante ya que cuando éste disminuye los ARL's de las cartas se acercan. Ahora, si la contaminación se da en la dirección positiva las cartas de control estudiadas en general presentan ARL's muy similares, lo cual se acentúa a medida que disminuye el nivel de correlación. En términos generales se puede concluir que al aumentar las proporciones de no conformidad en los atributos se logran atenuar las diferencias entre los ARL's obtenidos en las tres cartas estudiadas. A continuación se muestran los resultados obtenidos en el caso de proporciones de no conformidad diferentes en los atributos.

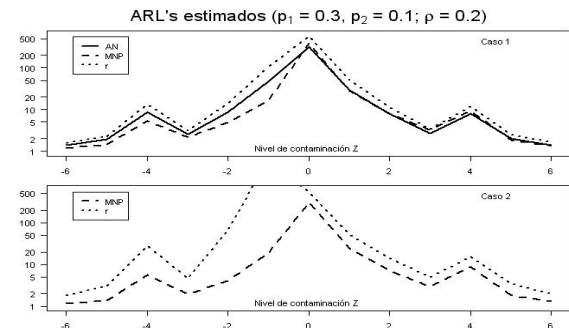
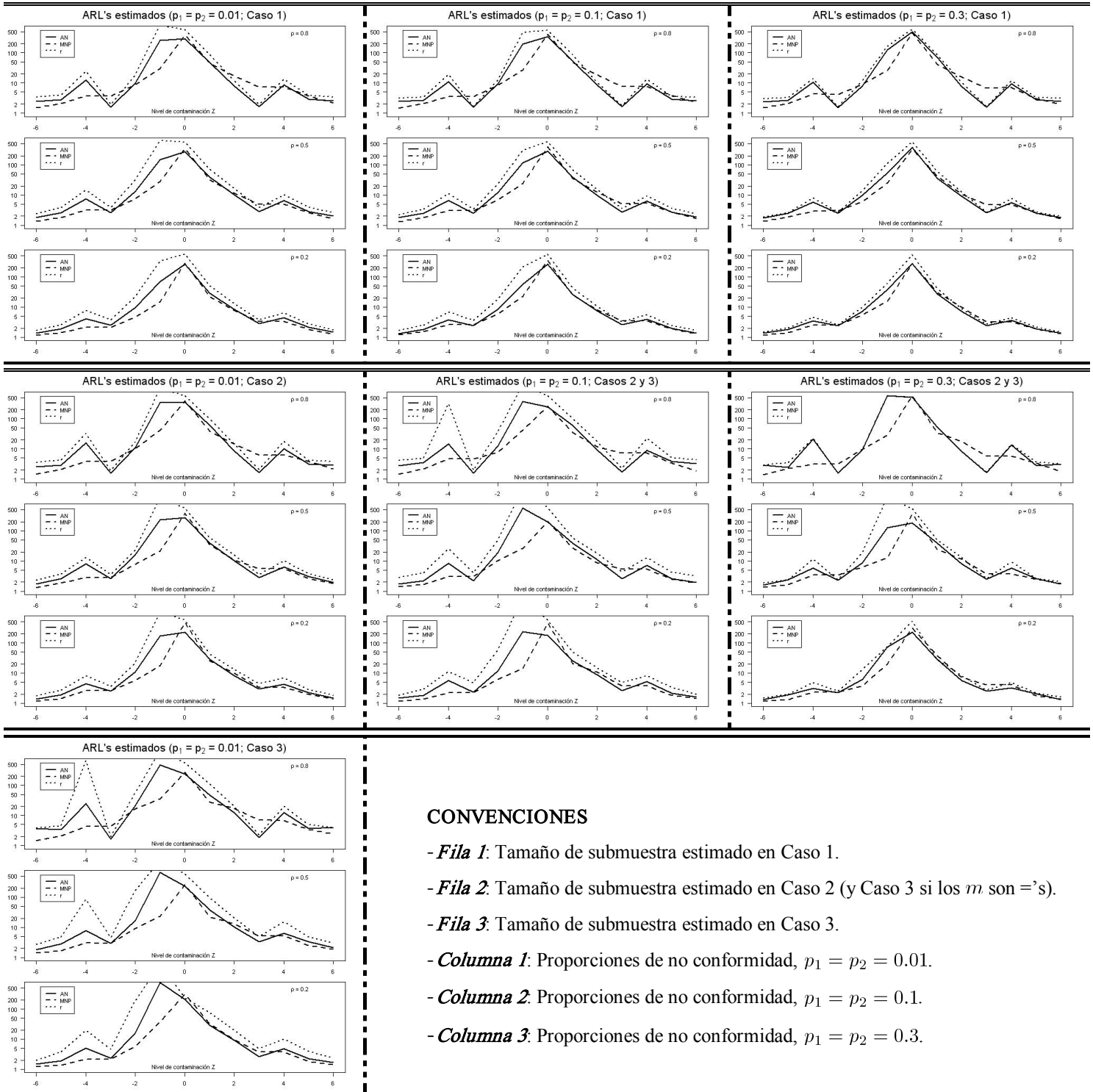


Figura 1. Resultados de simulación para proporciones de no conformidad diferentes.

Figure 1. Simulation Results for different nonconforming proportions.

Cuando las proporciones de no conformidad en los atributos son diferentes en el Caso 1 no se observan conclusiones diferentes a las reportadas en el análisis de la matriz de gráficos (Tabla 3), en el Caso 2 cabe notar que la carta *AN*, no arrojó ningún dato fuera de control, en este caso el tamaño muestral es pequeño ( $n = 15$ ), mientras que las cartas *MNP* y *r* si detectaron puntos fuera de control, sin embargo el mejor comportamiento en esta situación es para la carta *MNP*.





**CONVENCIONES**

- **Fila 1:** Tamaño de submuestra estimado en Caso 1.
- **Fila 2:** Tamaño de submuestra estimado en Caso 2 (y Caso 3 si los  $m$  son ='s).
- **Fila 3:** Tamaño de submuestra estimado en Caso 3.
- **Columna 1:** Proporciones de no conformidad,  $p_1 = p_2 = 0.01$ .
- **Columna 2:** Proporciones de no conformidad,  $p_1 = p_2 = 0.1$ .
- **Columna 3:** Proporciones de no conformidad,  $p_1 = p_2 = 0.3$ .

**Tabla 3.** Matriz de gráficos para los resultados de simulación cuando las proporciones de no conformidad son iguales.  
**Table 3.** Matrix of graphs for the simulation results when of the nonconforming proportions are equal.

## 5. APLICACIÓN CON DATOS REALES

En el área de servicio al cliente de una importante Empresa de Servicios de Telecomunicaciones se quiere establecer un sistema de control que les permita evaluar la percepción del cliente con respecto a las premisas estratégicas básicas de servicio: El *cumplimiento* del tiempo acordado en suministrar el servicio solicitado por el cliente y el *funcionamiento* adecuado de los servicios prestados, apuntado a un factor diferenciador “La Oportunidad”, este atributo es altamente valorado por el cliente de la compañía. Por lo tanto se puede representar lo anterior mediante dos variables Bernoulli, así:

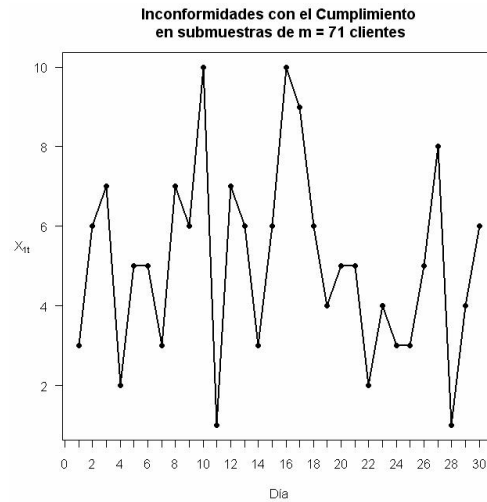
$$I_1 = \begin{cases} 1, & \text{si no se cumplió con el tiempo acordado.} \\ 0, & \text{en otro caso} \end{cases}$$

$$I_2 = \begin{cases} 1, & \text{si el servicio no quedó funcionando} \\ & \text{de manera adecuada.} \\ 0, & \text{en otro caso} \end{cases}$$

Observe que  $I_1$  representa la no conformidad de los clientes con el cumplimiento en el tiempo acordado para suministrar el servicio, mientras que  $I_2$  es la no conformidad del cliente con el funcionamiento adecuado de los servicios prestados en la actividad solicitada.

La empresa desea determinar cuándo se presenta un deterioro en la calidad del servicio a través de de los dos atributos definidos. Para lograr definir la calidad del servicio prestado se comenzó a medir diariamente los dos atributos mediante una encuesta telefónica, aplicada a clientes a los que se les instaló un servicio de telecomunicaciones. La encuesta se genera automáticamente cuando se cierra el servicio prestado. La medición se empezó a realizar a partir del mes de octubre de 2008, durante este mes se observaron los resultados y los datos históricos fueron extraídos en el mes de noviembre de 2008. Se tomaron aleatoriamente  $m = 71$  clientes, agrupados por fecha de solicitud de servicio y se les preguntó acerca de su percepción del cumplimiento y el adecuado funcionamiento del servicio solicitado. A continuación se muestran las gráficas del número de clientes inconformes en estos dos

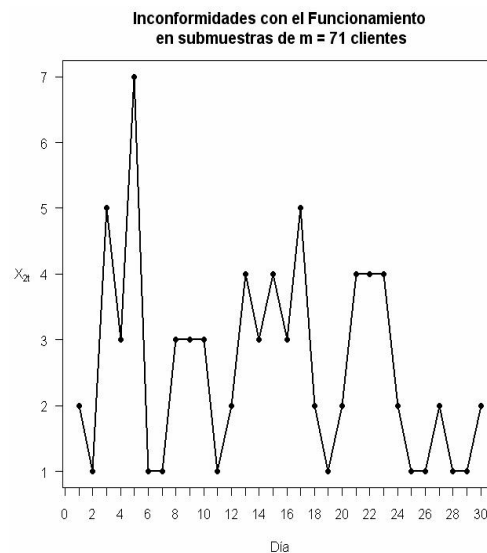
atributos durante un seguimiento de  $n = 30$  días (Fase I).



**Figura 2.** Inconformidades en el Cumplimiento (Fase I).

**Figure 2.** Unconformities in the compliance (Phase I).

De la gráfica anterior se puede observar que en promedio a 5 clientes por día en la muestra no se les cumplió con el tiempo acordado en suministrar el servicio, no hay ningún tipo de tendencia en el gráfico de control del atributo, lo que indica que el atributo cumplimiento en el periodo de evaluación se puede considerar un estándar de servicio.



**Figura 3.** Inconformidades en el Funcionamiento (Fase I).

**Figure 3.** Unconformities in the operation (Phase I)

Con respecto al atributo funcionamiento en los primeros seis días de evaluación de los datos históricos se presentaron comportamientos atípicos, debido a problemas en los sistemas de información que fueron corregidos, sin embargo se determinó que el atributo se podría considerar controlado corrigiendo las causas de los puntos fuera de control.

Partiendo del conjunto de datos históricos se determinan los parámetros del proceso en control como se describió en la sección 2, obteniendo los siguientes resultados.

$$\bar{X} = \begin{bmatrix} 5.0667 \\ 2.6000 \end{bmatrix} \text{ y } S = \begin{bmatrix} 5.7195 & 1.1655 \\ 1.1655 & 2.3172 \end{bmatrix}$$

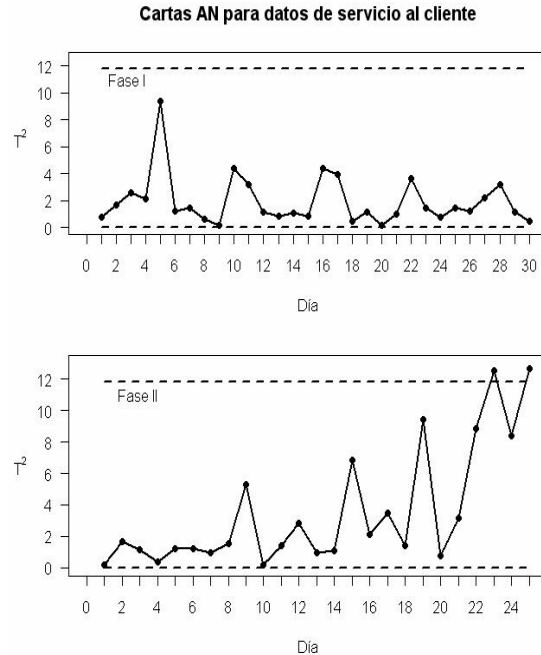
$$\bar{P} = \begin{bmatrix} 0.0714 \\ 0.0366 \end{bmatrix}, \hat{\rho}_{12} = 0.3201$$

Ahora, con base en las estimaciones anteriores se calculan los límites de control, para cada una de las metodologías estudiadas en la sección 2, cuyos resultados se resumen en la Tabla 4.

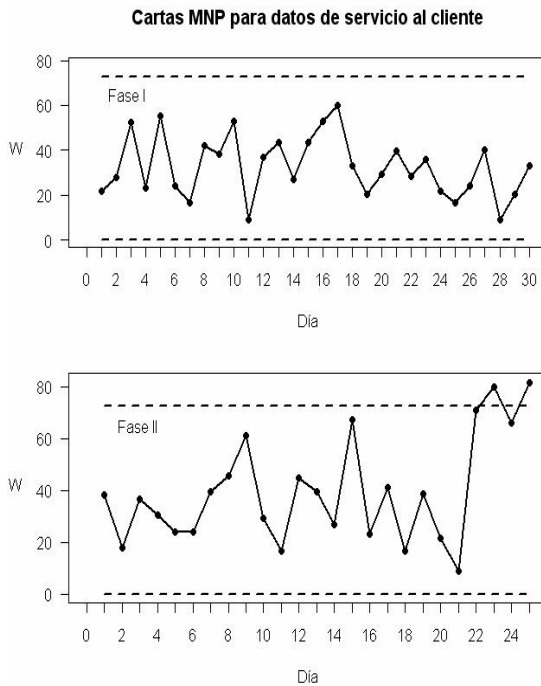
**Tabla 4.** Límites de control para cartas bivariadas  
**Table 4.** Control's limits for bivariate charts

Metodología	LCL	UCL
Carta AN	0	11.83
Carta MNP	0	72.50
Carta r	0.0027	1.00

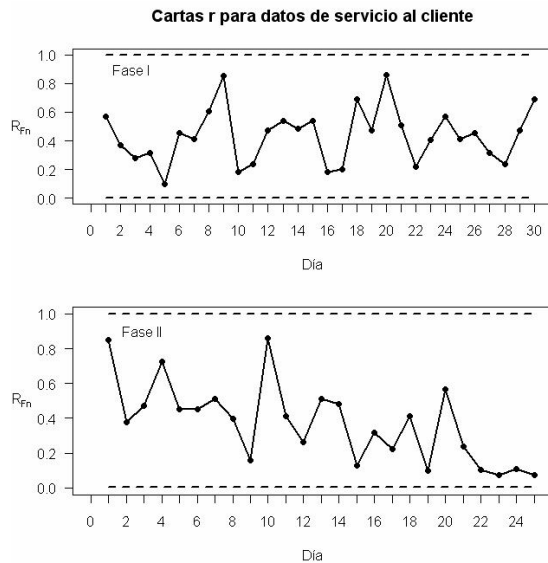
A continuación se muestran las cartas de control resultantes de aplicar las tres metodologías a los datos en Fase I y a un monitoreo posterior de clientes por 25 días en dic/2008 (Fase II).



**Figura 4.** Carta AN para datos reales.  
**Figure 4.** AN chart for real data.



**Figura 5.** Carta MNP para datos reales.  
**Figure 5.** MNP chart for real data.



**Figura 6.** Carta  $r$  para datos reales.

**Figure 6.**  $r$  chart for real data.

A partir de los resultados anteriores se determina que el proceso está fuera de control mediante las cartas de control  $AN$  y  $MNP$ , las cuales ubican dos puntos fuera de control, aunque se puede remarcar que en la carta  $MNP$  los puntos fuera de control están más alejados del límite de control superior. Por otra parte, aún cuando la carta  $r$  no detecta ningún punto fuera de control, se observa una tendencia decreciente en el estadístico  $r$ . Haciendo una revisión detallada del proceso se logró establecer que las causas para que éste se encontrara fuera de control correspondían a la temporada de vacaciones de fin de año, lo cual afectó la calidad de respuesta a las solicitudes de los clientes, tanto en el funcionamiento como en el cumplimiento.

## 6. CONCLUSIONES

Se propone la carta  $MNP$  para vigilar el comportamiento de un proceso para variables tipo atributos correlacionados, pues posee un ARL adecuado tanto en control como fuera de control. Particularmente fuera de control, cuando las contaminaciones son positivas o negativas en la mayoría de las simulaciones detecta primero una desviación pequeña de la media del proceso. Esta carta también presenta un desempeño similar en todos los escenarios de simulación, lo

que la hace robusta para medir si un proceso de esta naturaleza está o no en control estadístico.

Las cartas  $AN$  y  $r$  presentan variaciones significativas en cuanto a su rendimiento cuando las condiciones de simulación cambian, lo cual las hace poco confiables a la hora de determinar de manera temprana si un proceso se encuentra fuera de control. Si se compara el desempeño de las técnicas  $AN$  y  $r$  se puede concluir que son similares cuando la contaminación se da en dirección positiva y el tamaño de submuestra favorece a la aproximación de la distribución binomial a la normal (Caso 1). Cuando se observan los resultados del ejemplo aplicado, se ratifican las conclusiones del estudio de simulación, pues a pesar de que tanto la carta  $AN$  como la  $MNP$  detectan los puntos fuera de control, la técnica  $MNP$  presenta los puntos fuera de control más alejados de los límites de control.

## REFERENCIAS

- [1] LARPKIATTAWORN, S. A Neural Network Approach for Multi-Attribute Process Control with Comparison of Two Current Techniques and Guidelines for Practical Use, Doctoral Dissertation, University of Pittsburgh, 2003.
- [2] LU, X. S., XIE, M., GOH, T. N. AND LAI, C. D. Control Chart for Multivariate Attribute Processes, International Journal of Production Research, 36(12), 3477-3489, 1998.
- [3] ZERTUCHE, F., CANTU, M. AND PIÑA, M. Sistema de Control Multivariado no Paramétrico de Procesos". Cultura Ciencia y Tecnología, 21(4), 11-18, 2007.
- [4] LIU, R. Control Chart for Multivariate Processes. Journal of the American Statistical Association, 90(432), 1380-1387, 1995.
- [5] PATEL, H. I. Quality Control Methods for Multivariate Binomial and Poisson Distributions, Technometrics, 15(1), 103-112, 1973.
- [6] SULLIVAN, J. H. AND WOODALL, W. H. A Comparison of Multivariate Control Charts

for Individual Observations, *Journal of Quality Technology*, 28(4), 398-408, 1996.

[7] KOCHERLAKOTA, S., AND KOCHERLAKOTA, K. *Bivariate Discrete Distributions*, Marcel Dekker, New York, 1992.

[8] MARSHALL, A. W. AND OLKIN, I. A Family of Bivariate Distributions by the Bivariate Bernoulli Distribution, *Journal of the American Statistical Association*, 80(390), 332-338, 1985.

[9] OSPINA, C. *Control Multivariado de Procesos con Variables Binomiales Bivariadas*, Tesis de Maestría, Posgrado en Estadística. Universidad Nacional, Medellín, 2009.

[10] LIU, R. AND SINGH, K. A Quality Index Based On Data Depth and Multivariate Rank Tests, *Journal of the American Statistical Association*, 88(421), 252-260, 1993.

[11] HAMURKAROGLU, C., MERT, M. AND SAYKAN, Y. Nonparametric Control Charts Based on Mahalanobis Depth, *Hacettepe Journal of Mathematics and Statistics*, 33, 57-67, 2004.

[12] ONG, S. H. The Computer Generation of Bivariate Binomial Variables with Given Marginals and Correlation, *Communications in Statistics - Simulation and Computation*, 21(2), 285-299, 1992.