Gráficos de Control para Procesos con Datos Auto correlacionados

Dindo Valdez Blanco

1. Introducción

Los gráficos de control tan utilizados en el control de la calidad de los procesos industriales a menudo presentan el problema de la auto correlación entre sus muestras, como se sabe el Dr. Shewhart fue el primero en usar las cartas de control en 1924, por esta razón se dice que es el pionero del "control estadístico de calidad"

Una de las asunciones básicas de las gráficas de Shewhart radica en la independencia de las muestras tomadas en intervalos periódicos del tiempo. Es por esta razón que los gráficos convencionales pueden dar resultados engañosos al indicar demasiadas falsas alarmas en el proceso cuando los datos se encuentran auto correlacionados[1] a través del tiempo. Lo anterior se puede presentar cuando los procesos de producción son muy lentos y las muestras son tomadas a intervalos de tiempo relativamente cortos.

Recientemente se han realizado estudios de técnicas de monitoreo y control de los procesos alternativos a las técnicas clásicas de control, una de estas es la implementación de las cartas de control con datos auto correlacionados.

2. ¿Cómo determinar si los datos están auto correlacionados?

La auto correlación de los datos se puede detectar de muchas formas, por ejemplo de forma gráfica a través del diagrama de dispersión. Para esto se consideran los datos desfasados en k unidades de tiempo: X_t, y X_{t-k} con k = 1,2, 3,... Luego se procede a determinar visualmente si existe alguna correlación entre ellos.

Otra manera más objetiva para determinar la presencia de auto correlación en las muestras es a través del cálculo de la función de auto correlación

$$\rho_k = \frac{Cov(x_t, x_{t-k})}{V(x_t)} \quad , k = 0,1,2, \ldots$$

Donde $Cov(x_t, x_{t-k})$ es la función de autocovarianza con un rezago k, y se asume que las varianzas del proceso son iguales.

La función de auto correlación estimada a través de las muestras está dada por el estimador:

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T}(x_t - \overline{x})(x_{t-k} - \overline{x})}{\sum_{t=1}^{T}(x_t - \overline{x})^2} \quad , k = 0,1,2, \ldots$$

Este estimador se denomina función de auto correlación muestral. Para determinar si las autocorrelaciones de distinto orden k son significativas, se tiene que la distribución de las auto correlaciones en muestras grandes es: $\hat{\rho}(k) \sim N(0,1/T)$. Del resultado anterior se deduce que la desviación típica es $1/\sqrt{T}$, por tanto al tomar dos desviaciones

[1] La auto correlación se refiere a la dependencia de los datos a través del tiempo

estándar se obtiene in intervalo de confianza aproximado al 95% de confiabilidad $\pm 2/\sqrt{T}$. Es así que cuando se grafican las auto correlaciones para una muestra de datos, se suelen incluir las "bandas de dos desviaciones estándar", las que son utilizadas para determinar si alguna auto correlación muestral es significativamente distinta de cero. En la práctica para determinar si alguna de ellas es significativamente distinta de cero, se toma la regla que dice lo siguiente: si la función de auto correlación muestral en valor absoluto de orden k es más grande que $2/\sqrt{T}$, entonces dicha auto correlación es estadísticamente significativo al 5%.

3. ¿Cómo analizar los datos cuando presentan auto correlación?

Una mala decisión es incrementar los intervalos de las muestras para evitar tropezar con este problema, esto porque aún cuando la presencia de auto correlación complica el procedimiento de análisis de las cartas de control, no se puede desechar información importante acerca de la dinámica del sistema y tratar de no tomar en cuenta este hecho como si todo estuviera bajo control, es decir asumir que las variables del proceso son independientes e idénticamente distribuidos como una variable normal.

Una forma alternativa de tomar en cuenta los datos con auto correlación es utilizar la metodología del análisis de series de tiempo, el cuál se constituye en una herramienta útil para el análisis de la dinámica del proceso que tiene datos con auto correlación en el dominio del tiempo discreto. Esto consiste en aplicar un modelo auto regresivo de medias móviles una vez identificado el grado del auto correlación de los datos.

Una vez que se ha ajustado un modelo Auto regresivo de Medias Móviles adecuado a los datos y que se han verificado los supuestos del modelo ajustado, se realizan las gráficas de control de Shewhart con los residuales ajustados, es decir que el diagnóstico y monitoreo del proceso se lo realiza ahora a través del uso de los residuales del modelo, teniendo en cuenta que dichos residuales ahora no presentan la auto correlación de los datos originales.

Veamos un ejemplo del caso, supongamos que se tienen los registros del peso de 100 muestras de un producto en un proceso de llenado automático.

N° muestra	X_1	N° muestra	X_1	N° muestra	X_1	N° muestra	X_1	N° muestra	X_1
1	9.2	21	9.9	41	9.2	61	9.7	81	7.4
2	10.4	22	9.4	42	10.7	62	9.1	82	9.4
3	9.5	23	9.6	43	8.9	63	10	83	9
4	8.9	24	10	44	10.7	64	9.6	84	8.8
5	11.8	25	13.3	45	11.9	65	11.8	85	10.6
6	10.4	26	10.2	46	10.2	66	9.4	86	11.1
7	11.4	27	10.5	47	9.6	67	10.7	87	10
8	10.3	28	8.6	48	10.2	68	9.8	88	9.3
9	8.1	29	11.5	49	11.8	69	9.9	89	8.2
10	11.9	30	13.2	50	11.8	70	11.5	90	12
11	10.7	31	9.9	51	8.5	71	9.5	91	10.8
12	11.5	32	8.7	52	10	72	9.6	92	9.7
13	10.3	33	8	53	10.3	73	10.4	93	9.3
14	10.5	34	12	54	11.3	74	10.2	94	9.8
15	12.2	35	11.4	55	12.2	75	11.8	95	12.1
16	10.6	36	12.1	56	11.4	76	10.1	96	10.6
17	8.9	37	10.7	57	10.7	77	11.4	97	12
18	8.6	38	10.7	58	9.9	78	9.6	98	10.2
19	11.1	39	11.1	59	11	79	9	99	10.4
20	12.3	40	12.4	60	11.7	80	11.3	100	12.1

Para determinar si los datos están auto correlacionados, se realizan los gráficos de dispersión de los datos, para distinto rezagos de orden k = 1,2,..... En nuestro caso sólo presentamos dos gráficos, en la figura 1 se puede observar que los datos desplazados en un periodo no muestran auto correlación, en cambio en la figura 2 es evidente que existe una auto correlación en los datos desplazados en cinco periodos.

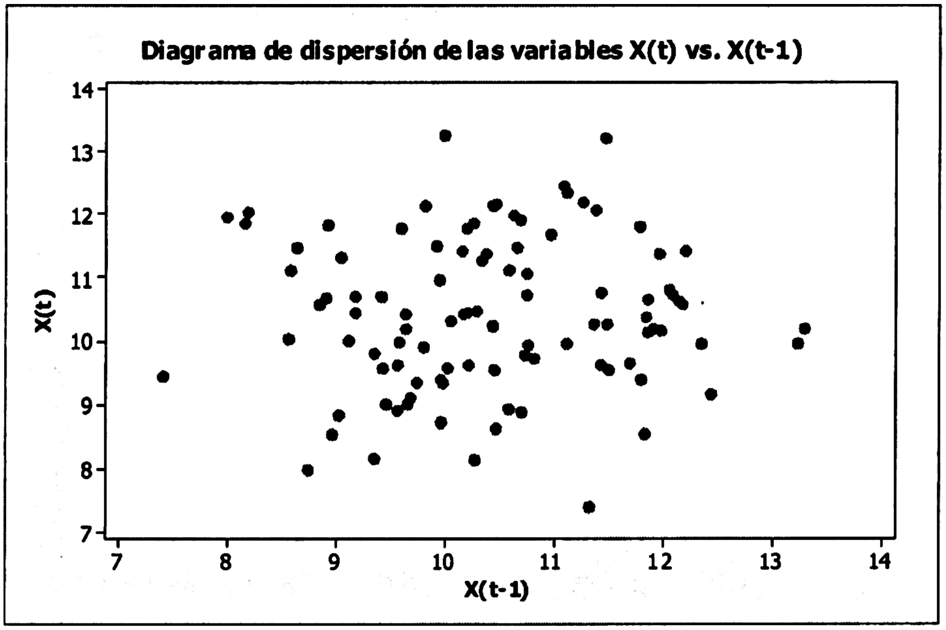

Figura 1. El gráfico no indica la existencia de auto correlación de los datos rezagados un periodo

A continuación se muestran las auto correlaciones hasta un desplazamiento de orden 11.

Rezago k	1	2	3	4	5	6	7	8	9	10	11
Auto correlación $\hat{\rho}_k$	0.02	-0.16	-0.22	0.02	0.47	-0.08	-0.23	-0.17	0.03	0.47	-0.05

Para determinar si alguna de las auto correlaciones es significativamente distinta de cero se procede a determinar si $|\hat{\rho}_k| > 2/\sqrt{T}$,en nuestro caso, *T=100 y* $2/\sqrt{T} = 0.2$ es la línea del intervalo de confianza al 95%. La figura 3 muestra que la auto correlación de orden 5 es importante en este proceso en particular, algo que se evidencia en la figura 2.

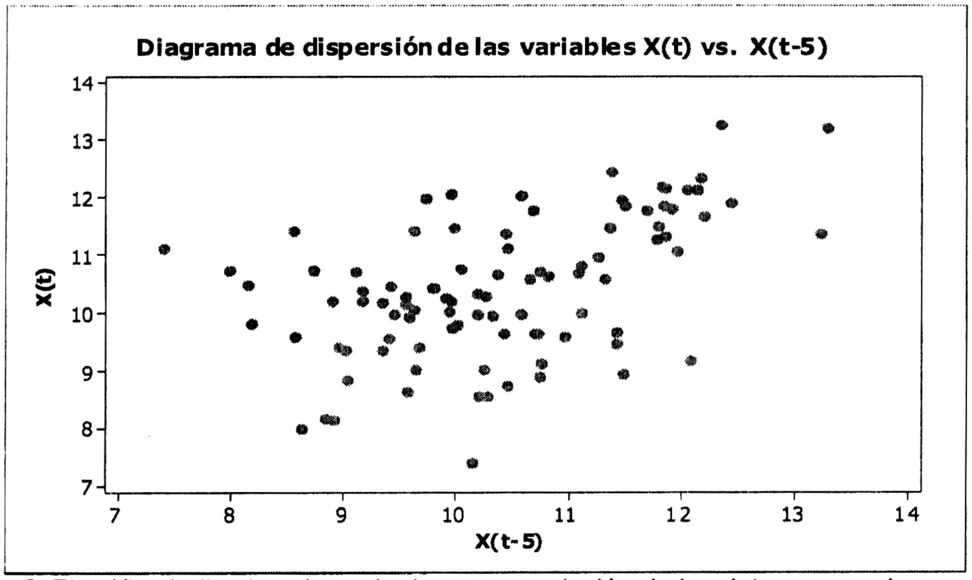

Figura 2. El gráfico indica la existencia de auto correlación de los datos rezagados en cinco periodos

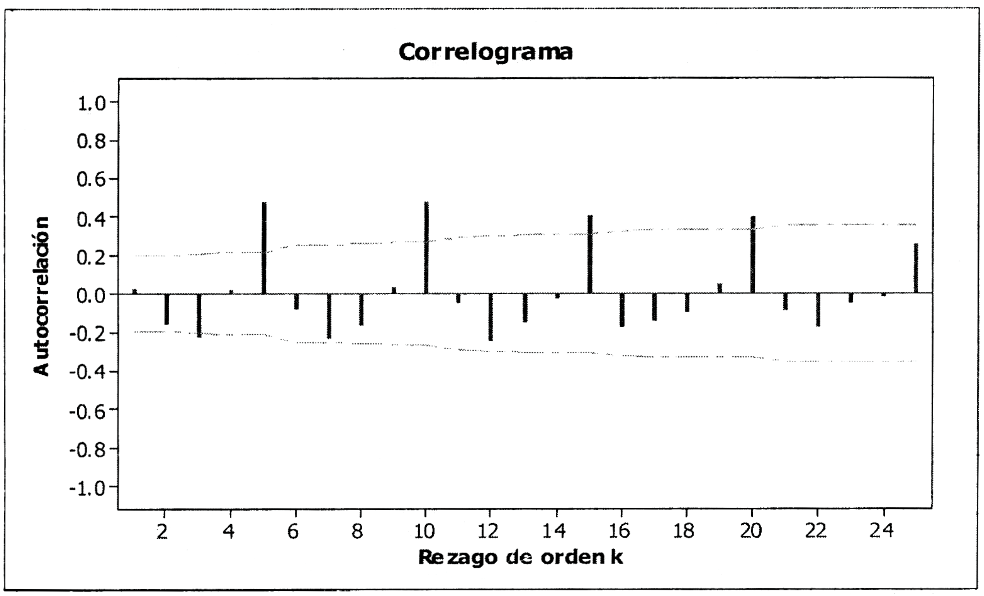

Figura 3. El gráfico de las auto correlaciónes muestra la existencia de auto correlación de orden 5

Por tal razón se plantea un modelo Auto regresivo AR(5), con el cuál se estiman los datos y se calculan los residuales, los siguientes gráficos muestran los resultados obtenidos.

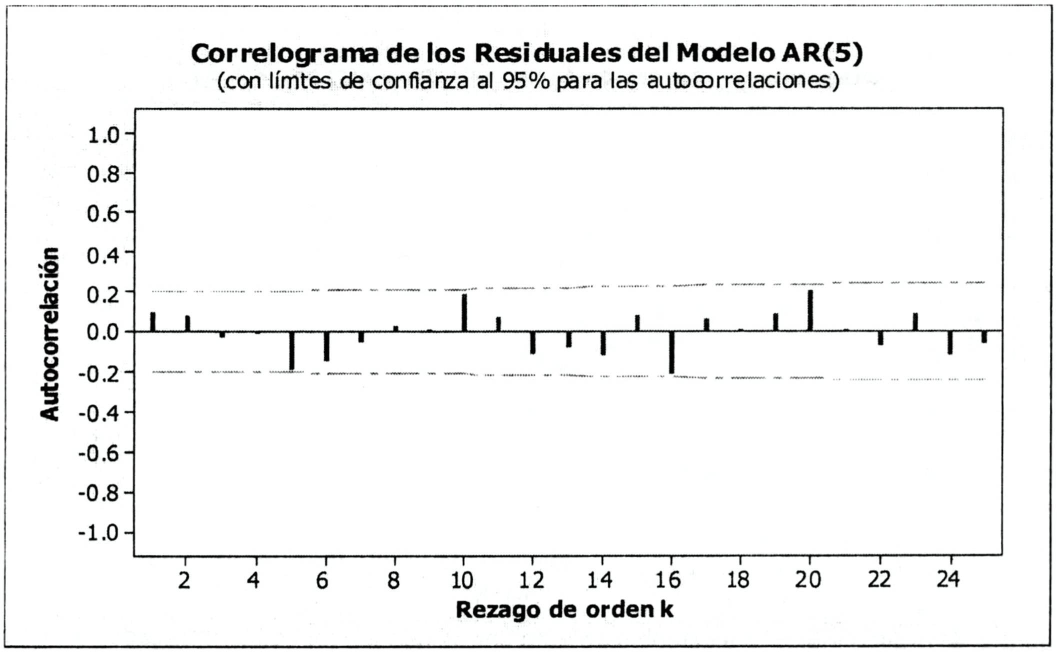

Figura 4. El gráfico de las auto correlaciónes con el modelo ajustado no muestra auto correlaciones significativas en los datos

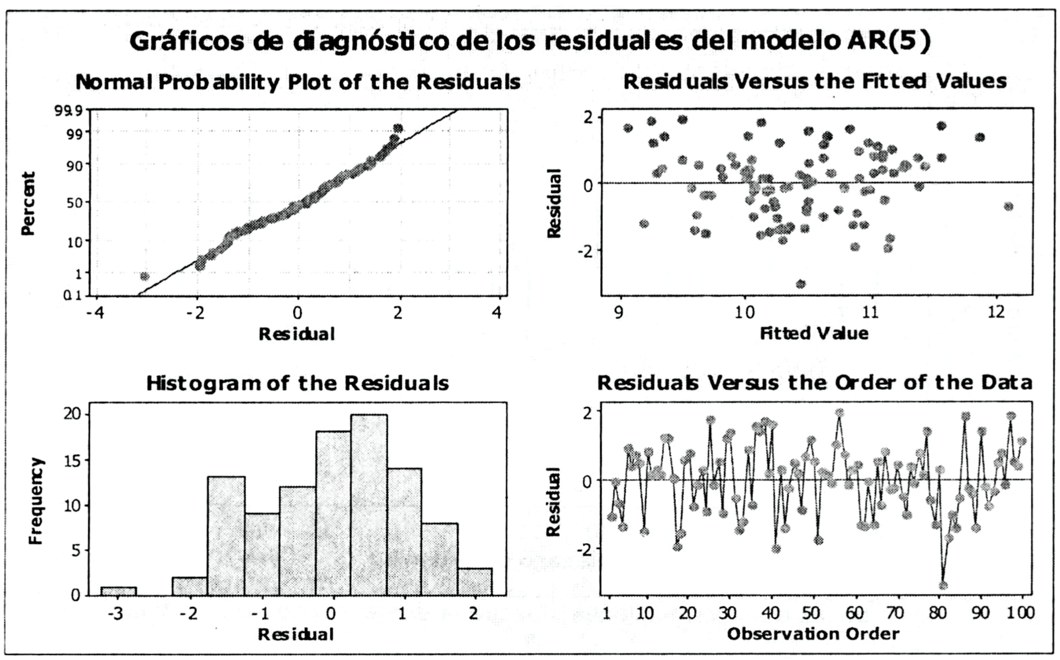

Figura 5. Los gráficos de diagnostico indican que los supuestos se cumplen

Ambas gráficas muestran que el modelo auto regresivo AR (5) ajusta de forma adecuada los datos, y minimiza el problema de la auto correlación. En nuestro caso el modelo ajustado es:

$$\hat{X}_t = 7.5352 - 0.0145X_{t-1} - 0.038X_{t-2} - 0.149X_{t-3} - 0.0039X_{t-4} + 0.4833X_{t-5}$$

Por último, una vez ajustado el modelo, se procederá a graficar las cartas de control de Shewhart utilizando los residuales del modelo, lo que dará una información más adecuada del proceso.

Bibliografía

[1] Castillo E.D. nad Montgomery, *Short Run Statistical Process and Alternative Methods*, Revista Electónica Engineering Intenational, 10(2) 87-97, 1994.
[2] Grant E. L., *Control Estadístico de Calidad*, Ed. Continental, México, 2002.

Índice