Gráfico de control para procesos binomiales con exceso de ceros

Silvia Joekes, Marcelo Smrekar, Andrea Righetti

Ponencia presentada en I Congreso Argentino de Estadística - XLIII Coloquio Argentino de Estadística - XX Reunión Científica del GAB realizado en 2015 en Buenos Aires, Argentina

ISSN 2451-8131

1 ° Congreso Argentino de Estadística (CAE I)

XLIII Coloquio Argentino de Estadística

XX Reunión Científica del GAB

Libro de Resúmenes extendidos

6 al 9 Octubre de 2015

Ciudad Autónoma de Buenos Aires

República Argentina

GRÁFICO DE CONTROL PARA PROCESOS BINOMIALES CON EXCESO DE CEROS

SILVIA JOEKES[1], MARCELO SMREKAR[2], ANDREA RIGHETTI[1]

[1] *Instituto de Estadística y Demografía, U.N.C.,*

[2] *Lab. de Ingeniería y Mantenimiento Industrial, U.N.C.*

joekess@eco.unc.edu.ar

RESUMEN

Cuando se analizan datos de procesos en los que se observan muchas muestras conformes, es decir, muchas muestras sin defectos, la utilización del modelo ZIB (Zero Inflated binomial) puede ser una muy buena alternativa. Esto es especialmente cierto para el caso en que los datos muestran una frecuencia más elevada de ceros que la que se debería esperar si la muestra hubiera sido generada mediante una distribución binomial. Tradicionalmente estos procesos fueron monitoreados empleando la distribución binomial pero, bajo estas circunstancias, la distribución binomial tiende a subestimar la variabilidad del proceso. En esta situación, los gráficos de control tienen límites muy estrictos que determinan excesivas señales de falsa alarma, altos costos de inspección y frecuentes paradas del proceso. Cuando no se tiene en cuenta el exceso de ceros, se genera un modelo mal especificado y, en consecuencia, el gráfico de control resultante no cumple con la función para la cual ha sido construido En este trabajo se propone la utilización del modelo lineal generalizado para establecer la bondad de ajuste entre los datos observados y la distribución asumida para la construcción del gráfico. Se muestra además la aplicación del gráfico ZIB a un proceso con datos de una planta de autopartes, con análisis y discusión de los resultados.

Palabras clave: *Procesos de Alta Calidad, Modelo Binomial con Exceso de Ceros, Gráficos de Control ZIB, Modelo Lineal Generalizado.*

Introducción

Debido al avance tecnológico y a la automatización de los procesos de producción, un proceso bien diseñado podría tener un número mayor de conteo de ceros de los que se esperaría encontrar bajo distribución Poisson o binomial. En particular, en el ámbito de la producción, es cada vez más frecuente encontrar muestras conformes, es decir, muestras sin ninguna unidad defectuosa. Estos procesos se conocen actualmente como procesos de alta calidad en los que, gracias al control continuo, se han logrado estándares de calidad realmente muy elevados.

Tradicionalmente estos procesos han sido monitoreados empleando las distribuciones Poisson o binomial, pero ambas distribuciones tienden a subestimar la media y la variabilidad del proceso en presencia de gran cantidad de ceros. Los gráficos de control resultantes tienen límites de control muy estrictos que conduce a una mayor tasa de falsas alarmas en la detección de señales de fuera de control.

Una forma de solucionar este inconveniente consiste en extender las distribuciones básicas de Poisson y binomial a una que pueda ser utilizada para modelar el efecto de una mayor cantidad de ceros. De esta manera, surgieron los procesos “zero-inflated” o procesos con “excesos de ceros”, que son aquellos procesos en los que se observa que los datos muestrales presentan una frecuencia más elevada de unidades conformes (cero defectos) que la que se debería esperar si la muestra hubiera sido generada mediante una distribución Poisson o binomial.

En los últimos años, varios investigadores han proporcionado nuevos modelos que se obtienen mediante la modificación de los modelos tradicionales existentes. Particularmente, los modelos con exceso de ceros, que incluyen a los modelos zero-inflated Poisson (ZIP) y zero-inflated binomial (ZIB), son modelos que se obtienen combinando adecuadamente el modelo tradicional con una singular distribución de ceros.

Sin embargo, los resultados obtenidos de estas investigaciones son muy variables debido a que dependen de la magnitud de la proporción de ceros, de los cambios en la media o en la proporción de no conformidades, del tamaño de las muestras, de la variabilidad y del estado en el que se encuentra el proceso ya sea bajo control o fuera de control. Esta situación conduce a una cierta complejidad a la hora de seleccionar el procedimiento de control adecuado.

En este trabajo se presenta una metodología más simple para determinar si un proceso es binomial con un número elevado de ceros debido a una tasa de no conformidades muy chica, o si se tiene una masa de probabilidad asociada con la presencia de ceros que es superior a la esperada para una distribución binomial estándar.

A efectos de realizar la elección adecuada del modelo, se propone utilizar el modelo lineal generalizado por ser éste un procedimiento muy flexible para tratar con datos de conteos. El modelo lineal generalizado es una extensión del modelo lineal clásico a una familia más general que involucran una variedad de distribuciones seleccionadas de una familia exponencial, como lo son las distribuciones, normal, binomial, Poisson, Gamma y binomial negativa. Fueron introducidos por primera vez por Nelder y Wedderburn (1972) y progresivamente se han aplicado en diversas disciplinas como agricultura, ecología, economía, ingeniería, y medicina entre otras.

Modelos con exceso de ceros o modelos cero-inflados

En una situación ideal, sólo unidades perfectas serían observadas es decir, unidades o muestras de unidades todas conformes. No obstante, una unidad o una muestra pueden estar afectadas por un mecanismo aleatorio que produce no conformidades. Si tal mecanismo, que puede ser controlable o no, ocurre, aparecerán entonces algunas no conformidades.

En el contexto del control estadístico de procesos esta situación indica que, además de la posibilidad de que el número de no conformidades en la muestra tome un valor cero bajo variación aleatoria, puede observarse un número adicional de ceros con probabilidad ϕ.

La función de probabilidad de la variable *Y* cero inflada está dada por;

$$P(Y=y)=\phi\, I_{(y,0)}+(1-\phi)\, g(y;\theta) \qquad y=0,1,2\ldots..$$

donde $I_{(y,0)}=1$ si $y = 0$, y $I_{(y,0)}=0$ en otro caso. $g(y;\theta)$ es la función de probabilidad de la distribución Poisson o de la distribución binomial. Con vector de parámetros θ.

Cuando la variable *X* sigue una distribución Poisson de parámetro λ, es decir, cuando $g(y,\lambda)=e^{-\lambda}\lambda^{y}/\,y!$, se tiene un modelo ZIP mientras que cuando la variable *X* sigue una distribución binomial con parámetros *n* y *p*, es decir, cuando $g(y,n,p)=C_{n}^{y}\,p^{y}(1-p)^{n-y}$, se tiene un modelo ZIB.

Modelo lineal generalizado aplicado a la distribución binomial con exceso de ceros

Los modelos lineales generalizados (Nelder y Wedderburn, 1972), pueden ser aplicados en casos en que la variable de respuesta no necesariamente tiene distribución normal y para cierto grado de no linealidad en la estructura del modelo. El modelo con exceso de ceros puede considerarse como un modelo mixto Bernoulli-binomial (Vieira *et al*., 2000), donde una posible explicación del exceso de ceros en los datos puede provenir de un mejoramiento considerable en alguna característica de calidad bajo monitoreo.

En los modelos binomiales con exceso de ceros la función de probabilidad de Y se puede expresar, evitando los subíndices por simplicidad, de la siguiente manera;

$$P(Y=y)=\begin{cases} \phi+(1-\phi)\,(1-p)^n & y=0 \\ (1-\phi)\binom{n}{y}p^y\,(1-p)^{n-y} & y=1,2,\dots,n \end{cases}$$

El logaritmo de la verosimilitud basado en estos datos está dado por; $l(\beta,\phi,\mathbf{y},\mathbf{z})$. Para un modelo binomial simple, y expandiendo la función logaritmo, es posible demostrar que la estimación de p y de ϕ, se pueden obtener maximizando los dos términos de $l(\beta,\phi,\mathbf{y},\mathbf{z})$ separadamente de la siguiente manera;

$$l(\beta,\phi,\mathbf{y},\mathbf{z})=l(\beta;\mathbf{y},\mathbf{z})+l(\phi;\ \mathbf{z})$$

En R esta función se conoce como “zibinomial” y por default los dos predictores lineales/aditivos son el logit de μ y el logit de ϕ.

Selección del modelo adecuado

La distribución binomial ha sido ampliamente utilizada para modelar datos de conteo, tales como el número de no conformidades por muestra en procesos productivos. Sin embargo, ante la presencia de una gran cantidad de ceros el investigador deberá determinar si el proceso es binomial con un número elevado de ceros debido a una tasa de no conformidades muy chica, o si se está en presencia de una cantidad de ceros que es superior a la esperada para una distribución binomial estándar. En pocas palabras, se debe decidir entre una distribución binomial estándar y una distribución binomial con exceso de ceros.

Esta decisión es de gran importancia, teniendo en cuenta que la habilidad de un gráfico de control está determinada por la bondad de ajuste entre los datos observados y la distribución asumida para la construcción del gráfico. Cuando se fracasa en tener en cuenta el exceso de ceros, se genera un modelo mal especificado que puede resultar en estimadores sesgados o inconsistentes.

La bondad de ajuste entre modelos competitivos puede establecerse utilizando varios criterios de información estadística. En particular, en los MLG estas estadísticas son de vital importancia para determinar la distribución de probabilidad que mejor ajusta a los datos. En este trabajo, se empleará el Criterio de Akaike que permite evaluar tanto el ajuste del modelo a los datos como la complejidad del modelo. Cuanto más pequeño es el valor del AIC mejor es el ajuste. El AIC es útil para comparar modelos similares con distintos grados de complejidad.

Determinación de los límites de control

Una vez que se ha determinado el modelo adecuado a los datos, es posible construir los límites de control. Si la distribución elegida es la binomial, es posible emplear cualquiera de los gráficos tradicionales. Aquí será considerado el gráfico adecuado para el caso en el que los datos presentan un exceso de ceros es decir, cuando el modelo seleccionado es un modelo ZIB.

Generalmente el límite inferior de control para el modelo ZIB no se considera, lo que es común en los gráficos de atributos. Dado que la distribución ZIB es altamente asimétrica, para definir los límites de control se sugiere emplear límites de probabilidad. Luego, con una señal de falsa alarma predeterminada α, el límite superior *k* de control del gráfico basado en el número de no conformidades se obtiene como el menor valor entero que satisface;

P (de *k* o más no conformidades en la muestra) ≤ α

Análisis de sensibilidad del gráfico de control

Es importante estudiar la sensibilidad del gráfico de control basado en el modelo ZIB. Para ello se emplea la longitud promedio de corrida, ARL. Como el modelo ZIB contiene dos parámetros, cada uno puede tener un impacto diferente sobre la probabilidad de alarma. Aquí será considerado sólo el límite superior del gráfico y se analizará el efecto de los cambios de cada uno de los parámetros del modelo suponiendo un proceso estable.

El ARL se obtiene mediante la relación $\mathrm{ARL} = 1/(1-\beta)$,, es decir, la inversa de la potencia estadística de la prueba siendo β una medida de la habilidad del gráfico de control para detectar desplazamientos en los parámetros del proceso. Cuando *p* toma el valor p_0, que indica el nivel aceptable de calidad o el nivel medio de calidad estimado del proceso, el ARL = 1/α, donde α es la probabilidad de falsa alarma.

Para el modelo ZIB, este valor está dado por;

$$\mathrm{ARL} = \frac{1}{P(\text{un punto fuera de control})} = \frac{1}{\sum_{\mathrm{y=LSC}}^{\infty} (1-\phi)\binom{n}{y} p^{y} (1-p)^{n-y}}$$

El análisis de sensibilidad indicó que este gráfico es relativamente sensible a aumentos en el valor de ϕ. Esto es importante en el control estadístico de procesos dado que ϕ está relacionado con la presencia de unidades conformes. Cuando ϕ aumenta, hay más unidades sin defectos, y luego la cantidad de unidades que deben ser inspeccionadas antes que se dé una señal de alarma aumenta rápidamente para cualquier tamaño de muestra. Por otro lado, a medida que la fracción de no conformes aumenta, alejándose del valor de la hipótesis nula, aumenta la potencia y disminuye la cantidad de muestras a ser inspeccionadas antes de dar la alarma.

Aplicación del modelo ZIB a un proceso industrial

Se presenta la aplicación de la metodología propuesta para el caso de una empresa autopartista que fabrica un tornillo de acero en forma de U, conocido técnicamente como tornillo U-bolt. Este tornillo tiene rosca en los extremos y se utiliza para la sujeción de elásticos en ciertos vehículos. Por ser un componente de seguridad de los automóviles el nivel de defecto aceptado debe ser muy bajo. La característica crítica a controlar está relacionada con las posibles fisuras en la estructura del tornillo.

Las frecuencias observadas indican que la cantidad de muestras en las que no se encontraron tornillos con fisuras tiene una frecuencia más elevada que la que se debería esperar si la muestra hubiera sido generada mediante una distribución binomial. Para comprobarlo, el paso siguiente consistió en emplear el modelo lineal generalizado a efectos determinar si el proceso es binomial con un número elevado de ceros o si se está en presencia de una cantidad de ceros que es superior a la esperada para una distribución binomial estándar, generando un modelo ZIB. Se estableció el modelo que mejor ajusta los datos y el valor de los parámetros estimados.

Con esta información se construyó el límite superior del gráfico de control y se obtuvieron las conclusiones pertinentes.

Resultados y Conclusiones

Cuando se analizan datos de procesos en los que se observan muchas muestras conformes, es decir, muchas muestras con cero defectos, la utilización del modelo ZIB puede ser una muy buena alternativa. Esto es especialmente cierto para el caso en que los datos muestran una frecuencia más elevada de unidades conformes (cero defectos) que la que se debería esperar si la muestra hubiera sido generada mediante una distribución binomial. Tradicionalmente estos procesos fueron monitoreados empleando la distribución binomial pero, bajo estas circunstancias, la distribución binomial tiende a subestimar la variabilidad del proceso. En esta situación, los gráficos de control tienen límites muy estrictos que determinan excesivas señales de falsa alarma, altos costos de inspección y frecuentes paradas del proceso. La utilización del modelo lineal generalizado ha mostrado ser un procedimiento adecuado para establecer la bondad de ajuste entre los datos observados y la distribución asumida para la construcción del gráfico. Cuando no se tiene en cuenta el exceso de ceros, se genera un modelo mal especificado y, en consecuencia, el gráfico de control resultante no cumple con la función para la cual ha sido construido.

Bibliografía

- ✓ Fatahi, A.A; Noorosana, R; Dokouhaki, P. y Babakhani, M. (2010). Truncated Zero Inflated Binomial Control Charts for monitoring rare health events. *International Journal of Research and Reviews in Applied Sciences* (IJRRAS). 4. 4, 380-387.
- ✓ Katemee, N y Mayureesawan, T. (2012a). “Nonconforming Control Charts for Zero Inflated Poisson distribution”. *World Academy of Science, Engineering and Technology*. 69, 149-155.
- ✓ Katemee, N y Mayureesawan, T. (2012b). “Control Charts for Zero-Inflated Poisson Models”. *Applied Mathematical Sciences*. 6-56, 2791-2803.
- ✓ McCullagh & Nelder (1989) *Generalized Linear Models*. Chapman and Hall.
- ✓ Nelder J. A. and Wedderburn R. W. M (1972). ”Generalized Lineal Model”, *Journal of the Royal Statistical Society*. Series A (General), Vol. 135, No. 3, 370-384.
- ✓ Peerajit V., y Mayureesawan T. (2010). “Nonconforming Control Charts for Zero-inflated processes”. *Proceedings of the National Conference on Statistics and Applied Statistics*, Thailand 76, 61-73.
- ✓ Sim, C.H. y Lim, M.H., (2008). “Attribute Charts for Zero-Inflated Processes”. *Communications in Statistics-Simulation and Computation*, 37: 1440-1452.
- ✓ Vieira, A. M. C., Hinde, J. P., Demetrio, C. G. B. (2000). “Zero-inflated proportion data models applied to a biological control assay”. *Journal of Applied Statistics* 27:373–389.
- ✓ Xie, M., He, B., Goh, T. N. (2001). “Zero-inflated Poisson model in statistical process control”. *Computational Statistics & Data Analysis* 38:191–201.
- ✓ Yawsaeng, B. y Mayureesawan, T. (2012).”Control Charts for Zero-Inflated Binomial Models”. *Thailand Statistician*. 10(1). 107-120.