RCAN

Revista Cubana de Alimentación y Nutrición
RNPS: 2221. ISSN: 1561-2929
Volumen 34. Número 1 (Enero – Junio del 2024): 190-196

Notas estadísticas

Escuela de Medicina de La Habana. La Habana

DE VUELTA SOBRE EL CHALLENGER: LA DISTRIBUCIÓN DE POISSON COMO MODELO DE INTERPRETACIÓN DE LA REALIDAD

Sergio Santana Porbén[1].

INTRODUCCIÓN

En una exposición anterior el análisis forense de la tragedia del trasbordador espacial "Challenger" fue el pretexto para demostrar la importancia de los valores no positivos (léase también negativos[*]) en el análisis estadístico-matemático de los resultados de una investigación.[1] En aquella exposición, se recurrió a un modelo binomial logístico para predecir el número de anillos O dañados ante una temperatura ambiental especificada. La explosión del "Challenger" sacudió al mundo entero, máxime cuando fue un evento que recibía una atención mediática sin precedentes. Los análisis forenses posteriores señalaron la deformidad de los anillos "O" ante las bajas temperaturas ambientes como la causa, primero, de la fuga de los gases combustibles, y la posterior explosión después. El modelo logístico binomial aplicado expuso la influencia de los valores extremos de la temperatura ambiente sobre la resistencia de los anillos "O" a la deformidad.

El modelo logístico se ha revelado útil para predecir el riesgo de que ocurra un evento de entre varias alternativas preexistentes. Sin embargo, y en el caso particular del "Challenger", la "materia prima" del análisis de datos lo representan el número de anillos "O" dañados al término de cada misión, y que podría corresponder con un anillo dañado, dos anillos dañados, o incluso, todos los anillos dañados[†]. Por otro lado, si bien tal modelo binomial logístico ha sido santificado por la práctica para el análisis de datos de similar porte, la adopción del mismo obvia un hecho incontrovertible: en condiciones naturales NO caben esperar anillos O dañados[‡]. Si bien tal presunción es loable, también no es menos cierto que, en la realidad de todos los días, siempre ocurrirán anillos O dañados. En consecuencia, se (re)define y se establece que, en condiciones naturales, la ocurrencia de tales anillos dañados será tan "pequeña" que podría

[*] A los fines del presente trabajo, se denomina como un resultado negativo aquel que apoya la hipótesis nula de la ausencia de tratamiento.

[†] En condiciones naturales, tal parece que la seguridad de la misión no es puesta en riesgo si no se rebasa un número crítico de anillos daños que, en este caso, parece ser 2. Entonces, la ocurrencia de 3 ó 4 ó 5 ó 6 anillos dañados no solo es improbable, sino también catastrófico, como parece que ocurrió con la fatídica misión #23.

[‡] El enfatizado es responsabilidad del autor.

[1] Médico. Especialista de Segundo Grado en Bioquímica Clínica. Profesor Asistente Máster en Nutrición en Salud Pública

Recibido: 23 de Marzo del 2024.
Aceptado: 20 de Abril del 2024.

Sergio Santana Porbén. Servicio de Laboratorio Clínico. Hospital Pediátrico Docente "Juan Manuel Márquez". Avenida 31 esquina a calle 76. Marianao. La Habana
Correo electrónico: ssergito@infomed.sld.cu.

asumirse como "despreciable" (o mejor dicho: "insignificante"). En virtud de lo anterior, es inmediato que el modelo binomial no sería la mejor solución del análisis forense de la tragedia del "Challenger", a pesar incluso de la alternativa elaborada por Papke y Wooldridge (1996).[2]

El fin último del procesamiento de los datos es la modelación de la realidad mediante la observación del fenómeno en condiciones controladas. De esta manera, se puede construir un modelo con el cual interpretar correctamente (e intervenir en consecuencia) la realidad. La distribución de Poisson ha sido propuesta en numerosas instancias para el análisis de eventos de incidencia infrecuente.[3] La distribución de Poisson puede incorporar el efecto de *offsets* (tenidos como el número de "exposiciones" acumulados durante la unidad de tiempo dentro de la cual ocurren (o no) los eventos de interés). Un atractivo de la distribución de Poisson sería, además, la existencia de modelos que asumen naturalmente la sobreabundancia de ceros como el modo esperado de operación (o como también se podría decir: en condiciones controladas) de un proceso especificado, como lo sería el de las misiones de los trasbordadores espaciales.[4] En virtud de todo lo anteriormente expuesto, se ha completado el análisis forense de la tragedia del "Challenger" empleando un modelo de regresión basado en la distribución Poisson. En un trabajo acompañante se expondrá la pertinencia de un modelo de regresión basado en una distribución Poisson "inflada" por ceros[§].

Presentación de la distribución Poisson

La distribución Poisson es una de las que están incluidas en la familia de las distribuciones exponenciales. El parámetro μ distingue la distribución Poisson, y se corresponde con el valor esperado de eventos independientemente del tiempo transcurrido de observación. La distribución Poisson puede extenderse para incorporar el efecto del tiempo de observación, o el tamaño de las series temporales asociadas con la ocurrencia del k-ésimo (k = 0, 1, 2, …) evento de interés[**]. El parámetro μ se modificaría en correspondencia:

$$\boxed{\begin{array}{c} P(K = k) = e^{-\mu}\mu^{k}\,\frac{1}{k!}\,, \\ \text{para } k = 0, 1, 2, \ldots, K \end{array}} \qquad [1]$$

$$\boxed{\begin{array}{c} P(K = k) = e^{-rt}(rt)^{k}\,\frac{1}{k!}\,, \\ \text{para } k = 0, 1, 2, \ldots, K \end{array}} \qquad [2]$$

De la expresión anterior, es inmediato que μ = rt: número esperado del k-ésimo evento de interés que se expresa como una fracción del tamaño de la serie temporal; t: unidad del tiempo de observación; r: tasa esperada de ocurrencia del k-ésimo evento en la unidad t de observación del proceso Poisson. De esta exposición se desprende que la influencia de los fallos de los anillos O en la seguridad de las misiones del transbordador *Challenger* podría ser modelada mediante la distribución Poisson t = 6: número de anillos O presentes en cada transbordador; y k = 0 anillo O dañado; k = 1 anillo dañado; k = 2

[§] ZIP (del inglés *Zero-Inflated-Poisson*).

[**] Cada misión del transbordador *Challenger* puede ser tratada como una serie temporal finita comprendida entre el momento del lanzamiento y el momento del aterrizaje. Esta serie temporal se podría extender a la vida útil del transbordador, desde el diseño y la construcción, hasta la retirada del servicio activo y la puesta de baja, y contemplaría las varias misiones de servicio del mismo.

anillos dañados; …; k = 6 anillos dañados; respectivamente[††].

El objetivo ahora del análisis forense de la tragedia del *Challenger* sería la estimación del parámetro r. Para ello, se han propuesto varias estrategias, entre ellas, la estimación máximo-verosímil basada en la maximización del logaritmo de la función de máxima verosimilitud de la distribución del parámetro r según los k eventos observados;[5] la estrategia mínimo-cuadráticos basada en la minimización del cuadrado de los residuales del modelo[‡‡];[6] y la estrategia lineal generalizada desarrollada para distribuciones de la familia exponencial, y donde el parámetro r se hace dependiente de los valores observados de los k eventos mediante una función vinculante[§§].[7] La estrategia máximo-verosímil (ML) ha sido la empleada históricamente en la solución de aplicaciones estadísticas que involucran a la distribución Poisson, es muy flexible en su implementación, y asegura rápida convergencia hacia un máximo global que se corresponde con el vector solución de los parámetros de interés. Sin embargo, la estrategia ML es dependiente de los estimados iniciales del vector de los parámetros, y puede fallar en converger, o converger en un máximo local alejado del valor verdadero del vector de los parámetros.

La estrategia mínimo-cuadráticos (MC) es una visión alternativa (y a la vez complementaria) del problema de la búsqueda del máximo global que se corresponde con el vector solución de los parámetros del modelo propuesto, y no requiere del conocimiento de las derivadas parciales del logaritmo de la función ML para la implementación. Asimismo, la estrategia MC ofrece soluciones del vector de los parámetros del modelo que son también máximo-verosímiles, lo que garantiza que, no importa la estrategia empleada en el problema, todas han de converger en un único (y global) máximo.

La estrategia lineal generalizada (GLM)[***] puede ser vista como una síntesis (y culminación) de las estrategias anteriormente expuestas, y una solución costo-efectiva del problema de la búsqueda del vector ML de los parámetros de un modelo estadístico debido a la sencillez e inmediatez en la implementación de los algoritmos de cálculo. La estrategia GLM también tiene como ventaja adicional que no requiere de estimados iniciales del vector de los parámetros para la puesta en marcha de los algoritmos matemáticos.

De forma similar a lo expuesto en el trabajo precedente,[1] se hipotetizó que la relación entre las realizaciones de los k eventos y la temperatura exterior en el día del lanzamiento del transbordador fuera lineal o parabólica. En la primera de la hipótesis, la relación se describiría mediante 2 parámetros; mientras que, en la segunda, serían 3 parámetros:

$$\boxed{\mu = te^{(\beta_o X_o + \beta_1 X_1)}} \qquad [3]$$

$$\boxed{\mu = te^{(\beta_o X_o + \beta_1 X_1 + \beta_2 X_2^2)}} \qquad [4]$$

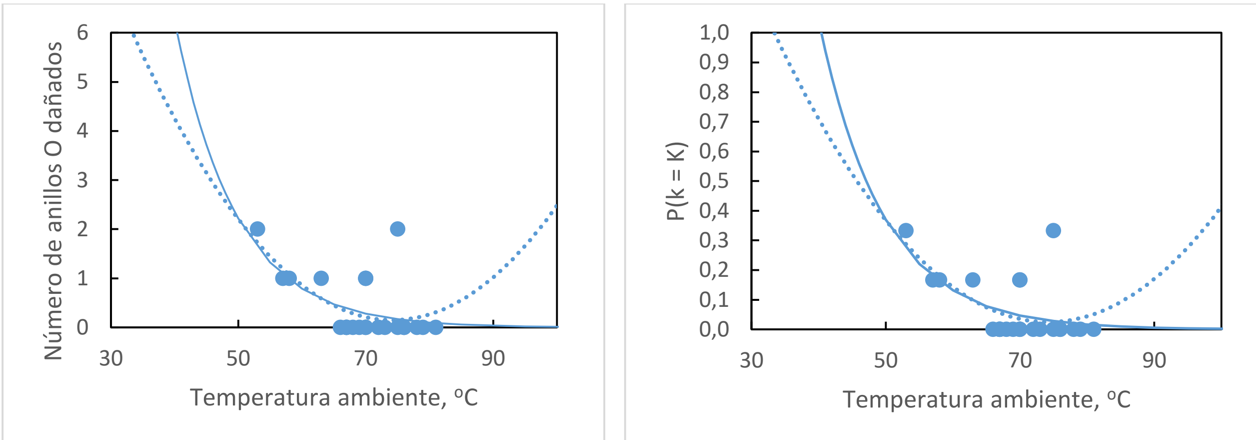
[††] Hasta la fatídica misión #23 de los transbordadores *Challenger*, se habían registrado valores de k = {0, 1, 2}, indicando con ello que la seguridad de la misión no se comprometía siempre y cuando k ≤ 2 anillos O dañados.

[‡‡] Los residuales del modelo se corresponden con las diferencias entre los valores observado y esperado del k-ésimo evento: $\Delta = (y - \mu)$.

[§§] Ninguna estrategia de estimación del parámetro r es superior a la otra, y se espera que todas ellas converjan en un único máximo global, y que coincide con el vector solución de los parámetros del modelo hipotetizado.

[***] Del inglés *Generalized Linear Models*.

Figura 1. Comportamiento de un modelo lineal para la predicción del número de anillos O dañados en el análisis forense de la tragedia del transbordador *Challenger*. El modelo lineal está delineado en la ecuación [**3**] del presente texto. *Izquierda*: Predicción del número de anillos O dañados al término de cada misión. *Derecha*: Predicción de la probabilidad de ocurrencia de un número especificado de anillos O dañados al término de cada misión. Para más detalles: Consulte el texto del presente ensayo.



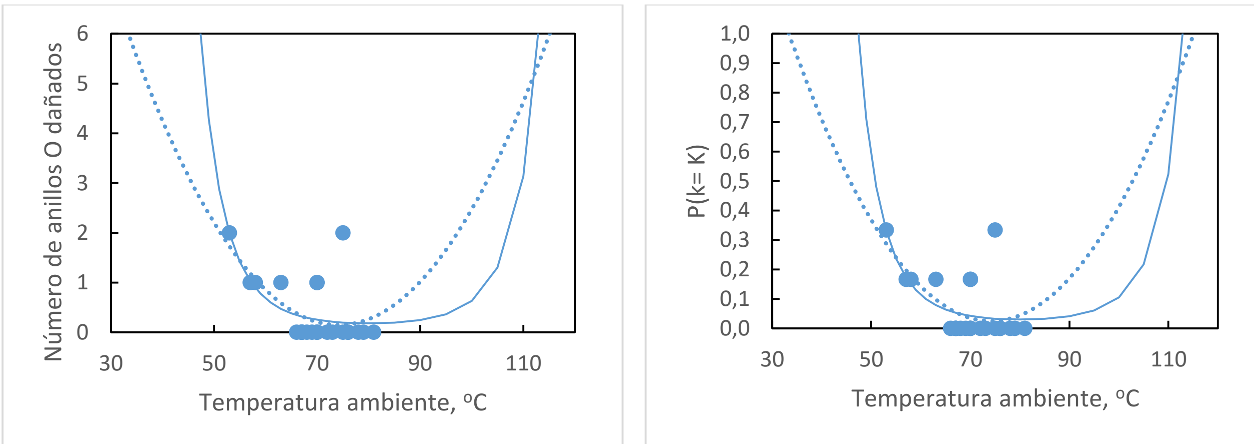
La Figura 1 muestra los resultados del modelo lineal expuesto en la ecuación [**3**]. El modelo predice que el número de anillos O dañados guarda una relación exponencial con la temperatura exterior, y que se anticiparía que, a temperaturas < 40ºC, todos los anillos O estarían dañados, lo que pondría en grave peligro la seguridad de la misión correspondiente. Sin embargo, el modelo no identificaría a las temperaturas exteriores extremadamente cálidas (> 100ºC) como igualmente lesivas para la integridad del transbordador. Igualmente, el modelo lineal predeciría como afectados por las temperaturas ambientales extremadamente frías a un número de anillos O mayor que el propio del diseño de la aeronave (que es de 6).

Por su parte, la Figura 2 muestra el comportamiento del modelo parabólico expuesto en la ecuación [**4**] en la predicción del número de anillos O dañados al término de cada misión de la lanzadera. Se constata ahora que el número de anillos O dañados es ≤ 2 para un rango de temperaturas exteriores entre 50 – 100ºC, reafirmando lo anotado empíricamente que las temperaturas exteriores extremas son equivalentes en su influencia deletérea sobre la estructura, y por ende, la integridad del transbordador. Aun así, el modelo parabólico (al igual que el modelo lineal expuesto más arriba) señala como dañados un número mayor de anillos O que los que fija el diseño.

DISCUSIÓN DE LOS RESULTADOS

Este trabajo ha explorado la pertinencia de la distribución Poisson en la modelación de la influencia de la temperatura ambiental en el número de anillos O dañados como causa de la tragedia del transbordador *Challenger*. Hasta donde alcanza la memoria, es la primera vez que la distribución Poisson se emplea para este tipo de análisis forense.

Figura 2. Comportamiento de un modelo parabólico para la predicción del número de anillos O dañados en el análisis forense de la tragedia del transbordador *Challenger*. El modelo parabólico está delineado en la ecuación [**4**] del presente texto. *Izquierda*: Predicción del número de anillos O dañados al término de cada misión. *Derecha*: Predicción de la probabilidad de ocurrencia de un número especificado de anillos O dañados al término de cada misión. Para más detalles: Consulte el texto del presente ensayo.



Las evidencias empíricas siempre apuntaron hacia una relación parabólica entre el número de anillos O dañados y la temperatura exterior en el día del lanzamiento: una señal de alerta de que las variaciones extremas de la temperatura ambiente son equivalentes en el riesgo para la integridad de la lanzadera. La inclusión de una relación parabólica como la mostrada en la ecuación [**4**] dentro de la estructura del proceso Poisson sirvió para confirmar tales evidencias, y establece sinergias con los resultados de la aplicación anterior de un modelo binomial logístico.[1]

No obstante, se hubiera deseado un modelo Poisson que prescribiera un número máximo de anillos O dañados en los extremos del rango de temperaturas permisibles. Esto es: el modelo Poisson aplicado debería anticipar que el número esperado de anillos O dañados fuera de 6 (coincidente con el número de anillos O presentes en la estructura del *Challenger*) para temperaturas exteriores extremas de 50 y 110ºC, en correspondencia con la relación parabólica hipotetizada. Este, claramente, no es el caso, y el modelo Poisson descrito en esta comunicación predice números muy superiores: un ejemplo de *overshooting*. Sin embargo, este comportamiento no debería sorprender en virtud de la pertenencia de la distribución Poisson a la familia de las distribuciones exponenciales.

Lo anteriormente dicho abre la oportunidad para el examen de la efectividad de otros modelos alternativos, en particular, aquellos que tienen en cuenta la sobreexpresión de las observaciones 0 en el vector de la variable Y dependiente, como lo serían la distribución Poisson inflada por ceros (ZIP)[4,8] y la distribución binomial negativa.[9] En este sentido, se llama la atención sobre los trabajos de Frome en los 1980s, quien propuso el análisis de datos como los propios de las misiones del *Challenger* mediante estrategias mínimo-cuadráticas de solución de modelos lineales embebidos dentro de la estructura de las tasas (léase también cocientes) del número observado de eventos respecto del tamaño de la serie temporal de pertenencia.[10-11] En futuras extensiones de esta investigación se explorarán comparativamente tales estrategias, y las ventajas y limitaciones del empleo de las mismas.

CONCLUSIONES

El modelo Poisson aplicado en el análisis forense de la tragedia del transbordador *Challenger* ha sinergizado con modelos logísticos binomiales preexistentes en cuanto a la demostración de una relación parabólica subyacente entre la temperatura exterior en el día del lanzamiento y el número de anillos O dañados como indicador de la integridad de la lanzadera. Quedan asuntos pendientes sobre la manipulación de la sobrerrepresentación de las observaciones 0 en el vector de la variable dependiente.

Futuras extensiones

En el análisis forense de la tragedia del *Challenger* se ha asumido la dependencia del número de anillos O dañados respecto de la temperatura ambiental. En los modelos presentados, la temperatura en el día del lanzamiento se ha asumido como un factor fijo (esto es: como una variable independiente | explicativa | predictora | "fijada" de antemano). Sin embargo, no se puede soslayar que el número de anillos O dañados es propio de cada misión de lanzamiento, e integra numerosas variables más allá de la temperatura exterior. En futuras extensiones se podría evaluar cómo la misión propiamente dicha (como un factor no controlado | aleatorio) influiría en el número de anillos O dañados mediante modelos mixtos.[12]

NOTAS ADICIONALES

El autor coloca a disposición de los lectores interesados los datos primarios de las misiones del transbordador *Challenger*, y los algoritmos empleados en el tratamiento estadístico-matemático y el análisis de los resultados. Se han cuidado la captación de los datos primarios de las fuentes anotadas, y la exactitud de las maquinarias de estimación de los parámetros de los modelos logísticos expuestos. Cualquier error presente que afecte la interpretación de los resultados expuestos en este ensayo es solo responsabilidad del autor.

AGRADECIMIENTOS

Dr. James Hardin, por sus invaluables sugerencias y comentarios.

REFERENCIAS BIBLIOGRÁFICAS

1. Santana Porbén S. La importancia de los valores no positivos y negativos: A propósito de Poisson, Fisher y el *Challenger*. RCAN Rev Cubana Aliment Nutr 2023;33:445-53.
2. Papke LE, Wooldridge JM. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. J Appl Econometr 1996;11:619-32.
3. Milner Jr DA. Count data and the Poisson distribution. En: Statistics for pathologists [Editores: Milner Jr DA, Meserve E, Soong TR, Mata DA]. Springer Publishing. New York: 2016.
4. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 1992; 34:1-14.
5. Silva JS, Tenreyro S. On the existence of the maximum likelihood estimates in Poisson regression. Econom Lett 2010; 107:310-2.

6. Kukush A, Schneeweis H, Wolf R. Three estimators for the Poisson regression model with measurement errors. Stat Papers 2004;45:351-68.
7. Abdulkabir M, Edem UA, Tunde RS, Kemi BL. An empirical study of generalized linear model for count data. J Appl Comput Math 2015;4(5):1000253. Disponible en: http://doi:10.4172/2168-9679.1000253. Fecha de última visita: 17 de Abril del 2024.
8. Lee AH, Wang K, Yau KK. Analysis of zero-inflated Poisson data incorporating extent of exposure. Biometrical J 2001; 43:963-75.
9. White GC, Bennetts RE. Analysis of frequency count data using the negative binomial distribution. Ecology 1996;77: 2549-57.
10. Frome EL. The analysis of rates using Poisson regression models Biometrics 1983;39:665-74.
11. Frome EL, Checkoway H. Use of Poisson regression models in estimating incidence rates and ratios. Am J Epidemiol 1985; 121:309-23.
12. McCulloch CE, Searle SR. Generalized, linear, and mixed models. John Wiley & Sons. New York: 2004.