Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA. **Rect@**

Volumen 15. Pagina 159 a 171.

EL MODELO POISSON GENERALIZADO INFLADO DE CEROS: UNA APLICACIÓN EN EL ENTORNO EDUCATIVO UNIVERSITARIO

GARCÍA-ARTILES, MARÍA-DOLORES
mariadolores.gartiles@ulpgc.es
Universidad de Las Palmas de Gran Canaria
Departamento de Métodos Cuantitativos en Economía y Gestión
Facultad de Economía, Empresa y Turismo, 35017 Las Palmas de Gran Canaria

GÓMEZ-DÉNIZ, EMILIO
emilio.gomez-deniz@.ulpgc.es
Universidad de Las Palmas de Gran Canaria
Departamento de Métodos Cuantitativos en Economía y Gestión
Facultad de Economía, Empresa y Turismo, 35017 Las Palmas de Gran Canaria

DÁVILA CÁRDENES, NANCY
nancy.davila@ulpgc.es
Universidad de Las Palmas de Gran Canaria
Departamento de Métodos Cuantitativos en Economía y Gestión
Facultad de Economía, Empresa y Turismo, 35017 Las Palmas de Gran Canaria

Recibido (12/09/2014)
Revisado (17/12/2014)
Aceptado (21/12/2014)

RESUMEN: En este trabajo presentamos la distribución de Poisson generalizada inflada de ceros, que se ha mostrado útil en escenarios en los que aparece una amplia presencia de ceros en la muestra. La presentación del modelo se completa con el desarrollo de un programa específico basado en Mathematica que supera algunas limitaciones de otros programas como STATA o EViews que no incluye la distribución de Poisson inflada de ceros en su catálogo. Las ventajas del modelo presentado y del programa propuesto se ilustran con un ejemplo real que se ajusta bien a sus características: el análisis de los factores que influyen en la asistencia de los estudiantes universitarios a las tutorías académicas. Este ejemplo es especialmente adecuado para mostrar la utilidad de la metodología presentada dado que incluye un elevado número de ceros reflejo del alto número de veces que los estudiantes contestan que no han asistido a tutorías. El lugar de residencia, la asistencia a clases de teoría, así como el seguimiento de la evaluación continua se muestran como las variables que parecen explicar la asistencia a la tutoría académica.

Palabras claves: Modelo de Poisson generalizado, modelo de Poisson inflado de ceros, covariable, tutoría académica, educación universitaria.

ABSTRACT: This paper presents the zero-inflated generalised Poisson distribution, which is useful when there is a large presence of zeros in the sample. After presenting the model, we develop a specific program based on Mathematica, overcoming some limitations of alternative approaches such as STATA or EViews, which do not include the zero-inflated Poisson distribution among its routines. The advantages of the model used and the proposed program are illustrated with a real example that is very appropriate to its features, namely an analysis of the factors influencing university students' attendance at tutoring sessions. This example is particularly suitable to show the usefulness of the methodology presented because it includes a large number of zeros, reflecting the many occasions on which the students do not attend these sessions. The students' place of residence, their attendance at lectures and the application of continual assessment are variables that seem to account for attendance at tutoring sessions.

Keywords: Poisson model generalized, zero inflated Poisson model covariate, tutoring sessions, Higher Education.

1. Introducción

La distribución clásica de Poisson es la distribución básica utilizada en los trabajos de naturaleza econométrica cuando se intenta explicar una variable discreta en función de otras (covariables). Sin embargo su uso suele producir ajustes que no superan la hipótesis nula del test Chi-cuadrado que postula una distribución de probabilidad totalmente especificada como el modelo matemático de la población que ha generado la muestra. La razón básicamente consiste en que la mayoría de los datos empíricos que aparecen en la literatura correspondiente presentan el fenómeno de sobre-dispersión (varianza mayor que la media) y la distribución de Poisson se muestra como un modelo no adecuado ya que para ella la varianza y la media coinciden. En este caso resulta aconsejable partir de una distribución que permita recoger dicho fenómeno.

En este trabajo se utiliza como caso alternativo a la distribución de Poisson la distribución generalizada de Poisson (Consul (1989) y Consul y Jain (1973)). Se trata ésta de una distribución discreta dependiente de dos parámetros que se presenta como una extensión natural de la primera y que la incluye como caso particular cuando uno de estos parámetros toma el valor cero. El formato que presenta la misma es sencillo, no incorporando ninguna función especial (entiéndase función gamma, hipergeométrica Gaussiana, función modificada de Bessel, etc.) de manera que los procedimientos de estimación resultan sencillos y asequibles mediante un tratamiento de datos susceptible de ser implementado en cualquier programa estadístico.

En numerosos estudios de recuento, datos dentales, actuariales, biológicos, etc., además del fenómeno de la sobre-dispersión antes mencionado, se ponen de manifiesto otras peculiaridades como la denominada sobrevaloración en cero, observándose una frecuencia del valor cero apreciablemente superior a la frecuencia teórica determinada con la distribución de Poisson y, en nuestro caso también, frente a la distribución de Poisson generalizada. Resulta en este caso conveniente desarrollar y tener en cuenta los denominados modelos inflados de ceros, válidos cuando hay un porcentaje elevado de ceros en la muestra. Se destacan aquí por la conexión que tiene con la modelización utilizada los trabajos de Famoye y Singh (2006) y Gupta et al. (2004), entre los numerosos trabajos que sobre la materia se han publicado en los últimos años.

Se muestra para este caso, así como para el modelo sin inflar de ceros, el procedimiento para estimar los parámetros del modelo mediante el método de máxima verosimilitud. Puesto que los paquetes estándar de estadística existentes en el mercado (STATA, EViews, etc.) no contienen en su catálogo la distribución generalizada de Poisson, se desarrolla, utilizando el software Mathematica (versión 9.0.) un programa que presenta la ventaja, entre otros, de calcular de manera simbólica la matriz de información observada así como las varianzas asintóticas de los estimadores. El programa desarrollado para la aplicación llevada a cabo en este trabajo puede ser solicitado a los autores del mismo. Resulta obvio que otros paquetes estadísticos, como R y Matlab, permiten llevar a cabo también esta tarea. En particular R dispone de códigos y paquetes específicos para trabajar con modelos inflados así como con la distribución generalizada de Poisson. La principal ventaja de Mathematica frente a estos otros paquetes radica en su potencial de cálculo simbólico.

Se ilustra el modelo presentado con una aplicación real para intentar analizar las posibles causas que expliquen el reducido número de estudiantes que utiliza el recurso de las tutorías académicas en la asignatura de Matemáticas Empresariales (en los estudios de Economía y Empresa) en el marco del Espacio Europeo de Educación Superior (EEES) en la Universidad de Las Palmas de Gran Canaria, ULPGC. En este contexto, se ha detectado una reducción del número de estudiantes que utilizan dicho recurso, por tanto, se observa una amplia presencia del valor cero en la variable a estudiar, la asistencia a la tutoría académica.

En definitiva, en este trabajo se plantea analizar las variables que puedan explicar los factores que inciden en este hecho, haciendo uso de un modelo probabilístico paramétrico que tiene en cuenta la sobrevaloración en cero, como ocurre en este caso en el que la mayoría de estudiantes no hace uso de las tutorías.

La estructura del trabajo es como sigue. En la sección 2 se describe el modelo a utilizar. En la sección 3 se presenta un ejemplo real ilustrativo aplicado en el contexto de la tutoría académica universitaria que, sin duda, supone un campo interesante de aplicación de la metodología utilizada. Se muestra con detalle los instrumentos, variables así como el procedimiento seguido. Finalmente se presentan los resultados y la discusión de los mismos.

2. El Modelo Poisson generalizado inflado de ceros

En la literatura estadística la distribución de Poisson es el modelo básico para respuestas tipo conteo. Esto es debido fundamentalmente a su sencillez y a que aparece implementada en la mayoría de los paquetes estadísticos existentes en el mercado. Sin embargo, puesto que la mayoría de los datos empíricos que aparecen en la literatura presentan el fenómeno de sobre-dispersión (varianza mayor que la media) y en menor medida de infra-dispersión (varianza menor que la media) la distribución de Poisson se muestra como un modelo no adecuado ya que para ella la varianza y la media coinciden. De ahí que resulte habitual modelar fenómenos de esta naturaleza utilizando otras distribuciones como la binomial negativa o la distribución de Poisson generalizada, entre otras. En este trabajo hemos optado por trabajar con esta última ya que se presenta como un caso natural de extensión de la distribución clásica de Poisson.

Distribución de Poisson generalizada

La distribución de Poisson generalizada es una distribución discreta que pertenece a la familia Lagrangiana de distribuciones y que incluye como caso particular a la distribución de Poisson. Su función de probabilidad (véase Consul (1989), Consul y Jain (1973) y Famoye y Singh (2006)) viene dada por:

$$f_{\lambda,\alpha}(x)=\frac{1}{x!}\lambda\left(\lambda+x\alpha\right)^{x-1}\exp\{-(\lambda+x\alpha)\},\quad x=0,1,\ldots,\mathrm{K}, \tag{1}$$

donde $\lambda>0$, $0\leq\alpha<1$ ó $\lambda>0$, $\alpha\leq 0$, $\max(-1,-\lambda/m)<\alpha\leq 0$, siendo m el mayor entero tal que $1+m\alpha/\lambda>0$. Cuando $\alpha=0$ la distribución de Poisson generalizada se reduce a la distribución de Poisson con parámetro $\lambda>0$.

La media y la varianza de la distribución de Poisson generalizada vienen dadas por (véase Consul y Jain (1973)):

$$E(X)=\frac{\lambda}{1-\alpha},\quad var(X)=\frac{\lambda}{(1-\alpha)^3},$$

respectivamente. Es sencillo comprobar que la media es menor, igual o mayor que la varianza si α es mayor, igual o menor que cero. Esto garantiza la versatilidad de dicha distribución para ajustar datos que presenten sobre-dispersión, equidispersión o infradispersión.

El cambio de variable $\lambda=\theta(1-\alpha)$ permite escribir la función de probabilidad (1) como:

$$f_{\theta,\alpha}(x)=\frac{\theta(1-\alpha)}{x!}[\theta+\alpha(x-\theta)]^{x-1}\exp\left\{-[\theta+\alpha(x-\theta)]\right\},\quad x=0,1,\ldots,\mathrm{K}. \tag{2}$$

Bajo esta reparametrización de la distribución de Poisson generalizada se tiene, obviamente, que $E(X)=\theta$, facilitando con ello interpretaciones cuando se desee introducir covariables en el modelo.

Sin embargo, a pesar de su enorme popularidad, la distribución de Poisson presenta una serie de limitaciones importantes. Una de ellas, y al mismo tiempo una de sus principales características, es la igualdad de la media y varianza condicionales, fenómeno conocido como equi-dispersión. Dicha distribución no resulta adecuada para modelar fenómenos en el que la varianza supera a la media, fenómeno conocido como sobre-dispersión, que está presente en la mayoría de los datos de recuento habitualmente utilizados, dificultando por tanto la obtención de un buen ajuste y, por ende, de las interpretaciones que se deriven del mismo.

Por otro lado, la distribución de Poisson se presenta como un modelo demasiado sencillo para captar el exceso de ceros que está presente en muchos datos, esto es, subestima la frecuencia real de ceros, tendiendo además a sobreestimar la frecuencia real de valores pequeños y a subestimar la de valores elevados. La abundancia de ceros en los datos ha sido analizada a partir de diferentes aproximaciones. La más destacada es la propuesta de Mullahy (1986), que introdujo el llamado modelo de Poisson inflado de ceros. El inflado de ceros se observa cuando los datos muestrales presentan una frecuencia más elevada para la ocurrencia de cero que el resto de valores observados.

Modelo inflado de ceros

La especificación del modelo en este caso viene dada por:

$$g_{\phi,\theta,\alpha}(y)=\begin{cases}(1-\phi)+\phi f_{\theta,\alpha}(0), & y=0,\\ \phi f_{\theta,\alpha}(y), & y>0,\end{cases} \tag{3}$$

donde $f_{\theta,\alpha}(y)$ es la distribución parental, la distribución de Poisson generalizada dada en (2) y $0<\phi\leq 1$ es un parámetro de inflación. Aplicaciones de modelos de esta naturaleza pueden verse en Famoye y Singh (2006) y Gupta et al. (2004).

El ajuste de las distribuciones se realiza por máxima verosimilitud tal como se indica a continuación. Supóngase que se dispone de una muestra $(y_1,y_2,\ldots,y_n)$ de tamaño n tomada de la función de probabilidad $g_{\phi,\theta,\alpha}(y)$. El logaritmo de la función de verosimilitud del modelo que incluye covariables viene dado por:

$$l(y_i;\phi,\theta_i,\alpha)=n_0\log(1-\phi+\phi f_{\theta_i,\alpha}(0))+\sum_{y_i>0}\log(\phi f_{\theta_i,\alpha}(y_i)),$$

donde n_0 es el número de ceros en la muestra. Las ecuaciones normales de las que se obtendrán los estimadores de ϕ, α y β_j ($j=1,2,\ldots,q$) resultan:

$$\frac{\partial\ell(y_i;\phi,\beta_i,\alpha)}{\partial\phi}=\frac{n-n_0}{\phi}-\frac{n_0\left(1-f_{\theta_i,\alpha}(0)\right)}{1-\phi+\phi f_{\theta_i}(0)}=0,$$

$$\frac{\partial\ell(y_i;\phi,\beta_i,\alpha)}{\partial\alpha}=\frac{n_0\phi}{1-\phi+\phi f_{\theta_i,\alpha}(0)}\frac{\partial f_{\theta_i,\alpha}(0)}{\partial\alpha}+\sum_{y_i>0}\frac{1}{f_{\theta_i,\alpha}(y_i)}\frac{\partial f_{\theta_i,\alpha}(y_i)}{\partial\alpha}=0,$$

$$\frac{\partial\ell(y_i;\phi,\beta_i,\alpha)}{\partial\beta_s}=\frac{n_0\phi}{1-\phi+\phi f_{\theta_i,\alpha}(0)}\frac{\partial f_{\theta_i,\alpha}(0)}{\partial\theta_i}\frac{\partial\theta_i}{\partial\beta_s}+\sum_{y_i>0}\frac{1}{f_{\theta_i,\alpha}(y_i)}\frac{\partial f_{\theta_i,\alpha}(y_i)}{\partial\theta_i}\frac{\partial\theta_i}{\partial\beta_s}=0,\quad s=1,2,\ldots,q,$$

y en la que $\partial\theta_i/\partial\beta_s=x_{is}\theta_i$. Es sencillo obtener que $f_{\theta_i,\alpha}(0)=\exp(-\theta_i(1-\alpha))$, de donde resulta $\partial f_{\theta_i,\alpha}(0)/\partial\alpha=\theta_i\exp(-\theta_i(1-\alpha))$. Por otro lado, se tiene:

$$\frac{\partial f_{\theta_i,\alpha}(y_i)}{\partial\theta_i}=f_{\theta_i,\alpha}(y_i)\left[\frac{1}{\theta_i}-\frac{(1-\alpha)(1+(1-\alpha)(\theta_i-y_i))}{\theta_i+\alpha(y_i-\theta_i)}\right],$$

$$\frac{\partial f_{\theta_i,\alpha}(y_i)}{\partial\alpha}=f_{\theta_i,\alpha}(y_i)\left[\frac{(y_i-1)(y_i-\theta_i)}{\theta_i+\alpha(y_i-\theta_i)}-\frac{1}{1-\alpha}-y_i\right].$$

ç

Obsérvese que si $\phi=1$ y $\alpha=0$, estamos bajo el modelo clásico de Poisson, y el estimador máximo verosímil del parámetro θ coincide, en este caso, con la media muestral, $\bar{y}$, en el modelo sin covariables.

Para el cálculo de la matriz de información de Fisher puede consultarse el trabajo de Famoye y Singh (2006) en el que la misma se calcula para una reparametrización de la distribución de Poisson generalizada ligeramente diferente a la que se va a utilizar en este trabajo.

3. Aplicación del modelo a las tutorías académicas universitarias

En los últimos años se ha teorizado y escrito mucho sobre el paradigma del proceso de Bolonia y su implementación a través del EEES. Bajo el nuevo modelo se ha propuesto un cambio en la orientación de la enseñanza pasando hacia un proceso enseñanza-aprendizaje, reforzando el papel del estudiante, el rol del profesor y la implicación de las instituciones con el objeto de garantizar un cierto éxito.

En este marco se da especial relevancia a la acción tutorial debido a que en el proceso enseñanza-aprendizaje la tarea de enseñar va más allá de la transmisión de conceptos de una disciplina, contempla la necesidad de orientar al estudiante en su desarrollo personal y curricular. Además, como afirman Arbizu, Lobato y Del Castillo (2005) es necesario ligar la formación universitaria a las demandas laborales lo que requiere una orientación profesional así como una formación práctica para desarrollar la capacidad de adaptación a las nuevas circunstancias, necesidades y demandas.

Numerosos informes y trabajos se han elaborado para tratar de orientar al profesor sobre esta nueva tarea, que rebasa las fronteras de la acción académica, algunos de ellos se encuentran en García Nieto (2008), Michavila y García (2003) y Sans Oro (2005).

Junto a la acción tutorial, el nuevo sistema de créditos, ligado al trabajo personal del estudiante implica que éste debe desarrollar un trabajo autónomo de aprendizaje, pero esto no debe suponer que el profesor se desvincule de esta faceta del proceso, sino todo lo contrario, que el estudiante puede y debe contar con el apoyo del profesor que adquiere un nuevo rol en el que además de enseñar, debe asistir, orientar y asesorar al estudiante para que éste desarrolle un aprendizaje activo.

En esta línea Romero, Zurita Ortega y Zurita Molina (2010) afirman que dentro de la acción tutorial la tutoría juega un papel fundamental al objeto de favorecer el trabajo personal del estudiante, a quien la guía del profesor le permitirá alcanzar las competencias para progresar en su aprendizaje. Durante este proceso existe además una retroalimentación que permite al profesorado conocer la visión de los estudiantes sobre la asignatura, sus estrategias de trabajo y sus carencias formativas, ya que durante la tutoría el trato más personalizado genera un ambiente de confianza que favorece el flujo de información entre las partes implicadas. El profesor pasa a ser, como afirma Cano (2009), no sólo el encargado de transmitir conocimientos sino a actuar como tutor, a orientar y a facilitar el aprendizaje de competencias.

Centrando la atención en la tutoría académica, el profesor debe actuar como tutor que resuelve las dudas originadas por los conocimientos que imparte, orienta al estudiante sobre los métodos de trabajo, le ayuda a corregir determinadas carencias y a buscar soluciones que contribuyan al éxito del estudiante en la materia de estudio. En definitiva, se considera que la principal función del profesor, como también afirma Herrera (2007), es posibilitar, facilitar y guiar al estudiante para que pueda acceder intelectualmente a los contenidos y prácticas profesionales de una determinada disciplina.

No obstante, para poder afrontar la tutoría académica y que el profesorado pueda desarrollar esta faceta, conviene no perder de vista la realidad de la configuración de la universidad española. En la misma, como se recoge en Gairín, Feixas, Guillamón y Quinquer (2004), los estudiantes acuden a las tutorías para resolver dudas ligadas a los contenidos de la asignatura, en función de la disponibilidad del profesor y de forma voluntaria sin que ello tenga repercusión en su evaluación.

Por ello, es conveniente contextualizar el marco en el que se desarrolla la docencia en el trabajo que aquí se presenta, para entender por qué principalmente se centra en la tutoría académica. La población que componen los estudiantes de nuevo ingreso se caracteriza por la heterogeneidad, determinada por, las diferentes opciones de acceso desde los bachilleratos a los ciclos formativos de grado superior, pasando por la incorporación de los mayores de 25 y 45 años; una diversidad cultural cada vez mayor; el acceso de estudiantes con diversidad funcional que pueden requerir de una atención diferenciada; los estudiantes de programas de movilidad y aquellos que tienen cargas familiares y/o están incorporados al mundo laboral, entre otros.

Además, los estudiantes de primer curso y en el primer semestre en la Universidad tienen unas características especiales que merecen ser consideradas, pues se enfrentan por primera vez a un nuevo modelo de enseñanza, con nuevos compañeros, nuevas formas de aprender, nuevos profesores, en grupos de enseñanza que duplican, como mínimo, su entorno de aprendizaje habitual en la enseñanza secundaria y bachillerato.

Resulta evidente que en este escenario, con grupos con una media de 80 estudiantes, pues el total de matriculados no asiste regularmente a clase, cabe preguntarse hasta qué punto resulta factible abordar tareas de acción tutorial individualizadas, o por pequeños grupos, por lo que las tutorías más allá de lo académico resultan difíciles de abordar por parte del profesorado.

Presentado el marco en el que se desarrolla la docencia, después de analizar los resultados en los que se constata un bajo rendimiento académico, como se muestra en Dávila-Cárdenes, García-Artiles, Pérez-Sánchez y Gómez-Déniz (2015), a la vez que se observa una baja asistencia a las tutorías por parte de los estudiantes, se trata de analizar las causas por las que no asisten con más regularidad a las tutorías académicas siendo éste un recurso importante que podría mejorar el éxito en la asignatura.

3.1 *Objetivos*

Objetivo general

Analizar qué variables pueden determinar la asistencia de los estudiantes a las tutorías presenciales en el despacho del profesor en la asignatura Matemáticas Empresariales.

Objetivo específico

Estudiar las distintas covariables que pueden influir en la baja asistencia de los estudiantes a las tutorías académicas, haciendo uso de la distribución de Poisson generalizada y la distribución de Poisson generalizada inflada de ceros.

3.2 *Método*

Participantes

En el curso 2012-2013 la asignatura de Matemáticas Empresariales, que corresponde al primer semestre del primer curso del grado en Administración y Dirección de Empresas de la Universidad de Las Palmas de Gran Canaria cuenta con unos 700 estudiantes matriculados, distribuidos en 7 grupos.

Para abordar el estudio se considera una muestra de 244 estudiantes que respondieron a un cuestionario que se distribuyó al principio del segundo semestre del curso durante las horas de Estadística Básica para las Ciencias Sociales, que pertenece al mismo departamento, ya que la asignatura objeto de estudio no tiene continuidad en el segundo semestre. La encuesta fue supervisada por los profesores que impartieron la asignatura de Matemáticas Empresariales, de modo que existió contacto directo con los estudiantes, y se les informó que la encuesta era para un estudio de investigación sobre las tutorías que deseaban llevar a cabo los profesores de Matemáticas.

Instrumentos

Para llevar a cabo el estudio se diseñó un cuestionario de 9 preguntas. Con la primera cuestión se trata de determinar cuántas veces había necesitado el estudiante de la ayuda y orientación del profesor para poder llevar con éxito la materia de estudio, lo que se formula con la pregunta número de veces que ha asistido a tutorías en el despacho del profesor.

Como se sabe que gran parte de los estudiantes no utiliza la tutoría, se pregunta de qué manera resuelven las dudas, si durante las clases de teoría o prácticas con el profesor de la asignatura, o recibiendo algún tipo de ayuda como pueden ser las clases particulares, o bien, utilizando los recursos disponibles en el Campus Virtual de la asignatura, o en general de la Web. Es de destacar que en el mencionado campus virtual los estudiantes disponen de todo el material para el seguimiento de las clases, de videos explicativos de contenidos de la materia, problemas resueltos y exámenes de otros años.

La siguiente cuestión es conocer su opinión sobre las tutorías presenciales, interesa saber si el estudiante considera que son importantes, en el sentido de que les puede ayudar a corregir determinadas carencias que pueden encontrar en el transcurso del estudio de la materia.

El seguimiento constante de la asignatura es un factor primordial en el buen rendimiento de la misma, por lo que la evaluación continua que se efectúa a lo largo del semestre, y que representa el 40% del total de la evaluación, proporciona una gran ayuda para superar con éxito la materia de estudio, es por eso que la información académica proporcionada por la nota obtenida por el estudiante en la evaluación continua es un dato importante para influir en su asistencia a tutorías.

Otros factores que pueden intervenir en la asistencia a las tutorías presenciales en el despacho del profesor se refieren a una serie de datos personales del estudiante. La opción de acceso a la universidad es un dato constatado que en la asignatura que nos ocupa, como son la matemáticas, influye notablemente en el entendimiento de la materia. También el conocer si el estudiante tiene beca o si trabaja puede inferir positiva o negativamente a la asistencia a tutorías.

Por último, se considera si el municipio de residencia de los estudiantes durante el curso podría ser una variable a tener en cuenta para la asistencia a tutorías ya que la Universidad de Las Palmas de Gran Canaria, está situada en la capital de la isla de Gran Canaria, a unos 7 kilómetros del centro de la ciudad. Si bien la mayor parte de sus estudiantes han nacido y residen en la isla hay estudiantes que proceden de otras islas, o son participantes en programas de movilidad que viven durante el curso en la capital o en residencias cercanas a la Universidad, sin embargo, un porcentaje importante de estudiantes reside en otras zonas de la isla.

La Figura 1 muestra el mapa de la isla de Gran Canaria, en el que se destaca la ubicación de las cinco zonas en que se ha dividido la isla para el estudio. Se señala también en la tabla adjunta las distintas zonas con la distancia máxima y mínima, en término medio, a la Universidad.

Zonas consideradas de la isla de Gran Canaria	*Distancia mín.-máx. a la ULPGC en Km*
Las Palmas de G.C.	0-12
Telde	20-30
Zona Centro	30-60
Zona Norte	25-50
Zona Sur	50-90

Figura 1. Mapa de la isla de Gran Canaria con las zonas consideradas

Variables

La muestra con la que se ha elaborado este estudio contiene 244 encuestas, en las cuales tal y como se esperaba, el 70% de los estudiantes indica que nunca ha asistido a una tutoría presencial en el despacho del profesor.

En la Tabla 1 se muestra la distribución de frecuencias del número de asistencia a tutorías para los 244 elementos de la muestra, es una variable discreta que toma el valor 7 si el estudiante ha asistido más de seis veces. Es interesante destacar que la distribución del número de asistencias muestra un patrón muy asimétrico. Sólo el 6.56% de los estudiantes ha asistido más de tres veces a tutorías, mientras que un 93.44% de los mismos han asistido tres veces o menos a las mismas.

Tabla 1. Distribución del número de veces que el estudiante asiste a tutoría presencial

Número de Asistencias	*Frecuencia*	*Porcentaje observado*
0	172	70.49%
1	23	9.42 %
2	20	8.20 %
3	13	5.33 %
4	8	3.28 %
5	2	.82 %
6	0	.00 %
7	6	2.46%

Con el objeto de incorporar al modelo las cuestiones planteadas en la encuesta, las etiquetas utilizadas para cada una de las variables consideradas se detallan en la Tabla 2, así como el porcentaje de respuestas afirmativas a cada una de las cuestiones formuladas.

Tabla 2. Variables utilizadas y definición de las mismas

VARIABLES	DEFINICIÓN	Respuestas afirmativas
ASISTENCIA	Número de veces que ha asistido a tutorías presenciales	30% (ha asistido alguna vez a tutorías)
TEORÍA	Resuelve las dudas en las clases de teoría con el profesor	60%
PARTICULARES	Resuelve las dudas en clases particulares	47%
WEB	Resuelve las dudas utilizando recursos de la web	50%
TAYUDAN	Las tutorías presenciales ayudan a entender la materia	77%
BECA	El estudiante dispone de beca	33%
TRABAJO	El estudiante trabaja	11%
EVCONT	Nota correspondiente a la evaluación continua	43% (aptos)
CSOCIALES	Procede de Bachillerato de Ciencias Sociales	66%
CTECNOLÓGICO	Procede de Bachillerato Científico Tecnológico	23%
OTROS	Procede de otros accesos que no son los dos anteriores	11%
RLPGC	Reside en las Palmas de Gran Canaria	55%
RTELDE	Reside en el Municipio de Telde	11%
RCENTRO	Reside en Municipios que se encuentran en la Zona Centro	7%
RNORTE	Reside en Municipios que se encuentran en la Zona Norte	7%
RSUR	Reside en Municipios que se encuentran en la Zona Sur	20%

En términos medios, aproximadamente la mitad resuelve sus dudas en el transcurso de las clases presenciales, o bien con algún tipo de ayuda externa o utilizando los recursos de la Web.

Sin embargo, es interesante notar el elevado número de estudiantes, 77%, que reconoce que las tutorías académicas son un recurso importante que les permite corregir las carencias en los contenidos matemáticos y les ayudan a la comprensión de la materia.

Con respecto a los datos personales, la mayoría de los estudiantes proceden del Bachillerato de Ciencias Sociales y residen en Las Palmas de Gran Canaria; un 33% de los estudiantes encuestados ha solicitado beca y sólo un 11% afirma que trabaja.

Procedimiento

La variable ASISTENCIA puede tomar valores entre 0 y 7, donde el 7 representa a los estudiantes que asistieron más de 6 veces a tutorías. El resto de variables, excepto la correspondiente a la nota de la evaluación continua, que oscila de 0 a 4 puntos, toman los valores 0 y 1 para las respuestas negativas y afirmativas respectivamente. Los descriptivos de la muestra se recogen en la Tabla 3.

En el modelo que se estudia, la variable aleatoria de interés es el número de veces que un estudiante de Matemáticas Empresariales de la ULPGC acude a tutoría presencial en el despacho del profesor (por tanto una variable discreta), que cuenta con un número elevado de observaciones que toman el valor cero, razón que motiva el presente trabajo, y que pretende averiguar cuáles son las causas que provocan esta inflación de ceros.

Tabla 3. Descriptivos

VARIABLE	Media.	S.D	Mín	Máx
ASISTENCIA	.7622	1.5047	0	7
TEORÍA	.6024	.4903	0	1
PARTICULARES	.4672	.4999	0	1
WEB	.5041	.5010	0	1
TAYUDAN	.7746	.4284	0	1
BECA	.3319	.4718	0	1
TRABAJO	.1060	.3091	0	1
EVCONT	1.6799	1.1744	0	4
CSOCIALES	.6598	.4747	0	1

CTECNOLÓGICO	.2295	.4213	0	1
OTROS	.1106	.3143	0	1
RLPGC	.5451	.4989	0	1
RTELDE	.1106	.3143	0	1
RCENTRO	.0696	.2551	0	1
RNORTE	.0737	.2619	0	1
RSUR	.2008	.4014	0	1

Para ello se asume que la variable ASISTENCIA, número de veces que un estudiante acude a tutorías, obedece a cierto modelo probabilístico, i.e. se trabaja con un modelo lineal generalizado, con distribución de probabilidad paramétrica y, por tanto, dependiente de parámetros, siendo uno de los parámetros igual a la media de la distribución. Esto último facilitará la interpretación cuando se desee introducir covariables en el modelo considerado. Aunque es habitual trabajar con la distribución de Poisson ésta no resulta adecuada en este caso ya que la varianza supera la media. En este sentido los datos muestrales dan una media de 0.76 y una varianza de 2.26; por tanto los datos tienen naturaleza sobre-dispersa. Asimismo, la distribución de Poisson no capta el exceso de ceros presente en los datos, el inflado de ceros se observa cuando los datos muestrales presentan una frecuencia más elevada para la ocurrencia de cero que el resto de valores observados. El índice del inflado de ceros se define como el cociente entre el número de elementos de la muestra con valor cero y el número de los mismos con valor distinto de cero. Si este cociente es mayor que 1, la muestra presenta inflado de ceros. En nuestro caso, y para la muestra que se ha tenido en cuenta, tenemos que este índice viene dado por

$$\text{Índice de inflado de ceros} = \frac{172}{72} = 2.38,$$

mucho mayor que 1. Por tanto, como ya se había anticipado, el porcentaje de ceros en la muestra es mucho mayor que el de no ceros y este hecho se tendrá en cuenta utilizando la distribución de Poisson generalizada descrita anteriormente.

Ajuste de modelos sin covariables. En primer lugar se examina y se comparan los ajustes de la variable dependiente objeto de estudio utilizando la distribución de Poisson generalizada (2). Se ajustan los datos mediante esta distribución y a continuación mediante la distribución de Poisson generalizada inflada de ceros (3) sin covariables, con el objeto de controlar el exceso de ceros. El ajuste de las distribuciones se realiza por máxima verosimilitud tal como se ha descrito en el modelo.

Para ello puede calcularse directamente el máximo del logaritmo de la función de verosimilitud o resolver las ecuaciones normales utilizando algún programa informático. En nuestro caso, como ya se señaló anteriormente, se ha utilizado el software Mathematica en su versión 9.0. Este programa presenta la ventaja de calcular de manera simbólica la matriz de información observada así como las varianzas asintóticas de los estimadores.

El comportamiento del ajuste se discute mediante diferentes tipos de contrastes. Además se tiene en cuenta el valor del máximo del logaritmo de la función de verosimilitud y de los criterios de información de Akaike $AIC = 2k\text{-}2l$ y Bayesiano $BIC = -2l + k \log(n)$, donde k es el número de parámetros del modelo, l el valor del máximo del logaritmo de la función de verosimilitud y n el tamaño muestral para el modelo estimado (véase Akaike (1974) y Leroux (1992)). Como es bien conocido un modelo con menor valor de estos dos últimos estadísticos será siempre preferido.

Como prueba específica de la bondad del ajuste se realiza el test Chi-cuadrado, comparando las frecuencias absolutas observadas empíricamente y las correspondientes frecuencias absolutas teóricas obtenidas con los dos modelos (inflado y sin inflar) considerados. Los cálculos proporcionan unos valores de -277.709 y -270.966 para el máximo del logaritmo de la función de verosimilitud en el modelo sin inflar de ceros e inflado de ceros, respectivamente. Por otro lado, el valor del test Chi-cuadrado resulta 12.41 y 0.08 con 2 y 1 grados de libertad, respectivamente de nuevo para ambos modelos. Estos valores se ven refrendados en el diagrama de barras que se muestra en la Figura 2.

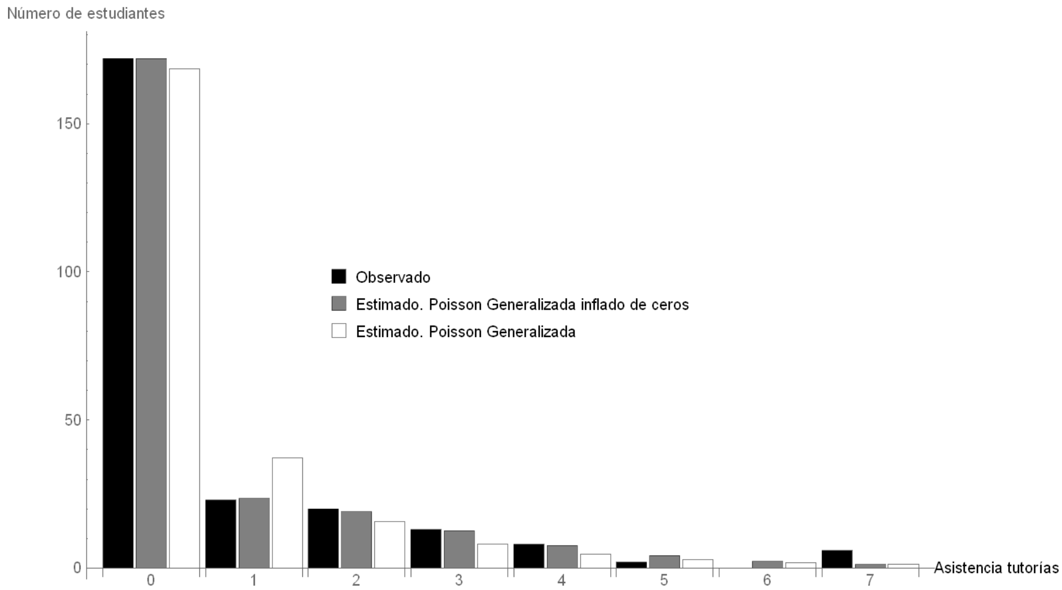

Figura 2. Ajuste sin covariables de la distribución de Poisson generalizada y la inflada de ceros

Finalmente incorporamos el Test Score (Dean y Paul (2000)) que permite comparar ambos ajustes. Según este test se puede contrastar la hipótesis nula $H_0 : \phi = 1$ frente a la hipótesis alternativa $H_1 : \phi \neq 1$, i.e. contrastar el modelo homogéneo frente al modelo inflado de ceros sin más que calcular el estadístico $T = UJ(1, \hat{\theta}, \hat{\alpha})^{-1} U^{\mathrm{T}}$, donde U es el vector que tiene como componentes las ecuaciones normales y en las que se reemplaza ϕ por 1 y los parámetros θ y α por los estimadores máximo verosímiles $\hat{\theta}$ y $\hat{\alpha}$, respectivamente. J es la matriz de información de Fisher. Resulta bien conocido que este estadístico sigue una distribución Chi-cuadrado con un grado de libertad, $\chi^2(1)$. Un valor de T superior a $\chi^2_{0.05,1} = 3.841$ rechazará el modelo homogéneo frente al modelo inflado de ceros. En este trabajo se observa que el valor del estadístico T es 126.212, muy superior a $\chi^2_{0:05;1} = 3.841$ y, por tanto, se rechaza el modelo homogéneo frente al modelo inflado de ceros.

Ajuste de modelos con covariables. Resulta bien conocido que el modelo lineal general constituye una generalización de la regresión por mínimos cuadrados ordinarios, relacionando la distribución aleatoria de la variable dependiente en el experimento con la parte no aleatoria a través de la denominada función de enlace (*link*). En este trabajo se asume que la variable dependiente *Y* está generada por la distribución de Poisson generalizada. La media de la distribución $E(Y) = \theta$ depende de las variables independientes a través de la expresión $\theta = h^{-1}(X\beta)$, donde el predictor lineal $X\beta$ es una combinación lineal de parámetros desconocidos, siendo *h* la función de enlace y en la que los parámetros desconocidos se estiman por máxima verosimilitud. La especificación más utilizada en este caso es considerar el parámetro θ exponencial, asegurando con ello la no negatividad del mismo. Esto es,

$$\theta_i = \exp\left(\sum_{s=1}^{q} x_{is}\beta_s\right), \quad \left(i = 1,2,\ldots,n\right).$$

Obteniéndose, por tanto, el modelo log-lineal de modo que $E(Y_i) = q_i$, $(i = 1, 2, ..., n)$.

3.3 *Resultados*

En este caso los resultados se muestran en las Tablas 4 y 5. Como se observa en la tabla 4, las variables significativas que se obtienen aplicando la distribución de Poisson generalizada son TEORÍA (al 1%) y TAYUDAN y EVCONT (al 5%). En la tabla 5 se reflejan los resultados obtenidos con la distribución de Poisson generalizada inflada de ceros, en este caso las variables TEORÍA y EVCONT mantienen su significatividad apareciendo como nuevas las variables que se refieren al lugar de residencia del estudiante, como son RCENTRO y RNORTE (al 5%).

Tabla 4. Resultados obtenidos con la distribución de Poisson generalizada con covariables

VARIABLE	Parámetro	Estimación	E.E.	$\lvert t \rvert$	Pr > $\lvert t \rvert$
	α	.4365	.0617	3.7944	0.4404
CONSTANTE	β_0	-2.7441	.5458	5.0272	.0000
TEORÍA	β_1	1.4150	.3428	4.1271	.0000
PARTICULARES	β_2	.0876	.2423	.3616	.7179
WEB	β_3	-.2609	.2301	1.1340	.2579
TAYUDAN	β_4	.8762	.4152	2.1103	.0359
BECA	β_5	.0181	.2492	.0727	.9421
TRABAJO	β_6	.2655	.3465	.7663	.4442
EVCONT	β_7	.2618	.1064	2.4594	.0146
TECNOLÓGICO	β_8	-.3770	.2863	1.3168	.1892
OTROS	β_9	.4929	.3403	1.4485	.1488
RTELDE	β_{10}	.4823	.3184	1.5147	.1312
RCENTRO	β_{11}	-.0052	.5555	.0094	.9924
RNORTE	β_{12}	.5467	.4268	1.2808	.2015
RSUR	β_{13}	.4087	.2981	1.3706	.1718

$L_{\max} = -251.681$, $AIC = 533.362$, $BIC = 585.819$

Tabla 5. Resultados obtenidos con la distribución de Poisson generalizada inflada de ceros con covariables

VARIABLE	Parámetro	Estimación	E.E.	$\lvert t \rvert$	Pr >$\lvert t \rvert$
	ϕ	.5502	.0715	7.6910	.0000
	α	.1921	.0617	2.5701	.0828
CONSTANTE	β_0	-2.5271	.7172	3.5233	.0005
TEORÍA	β_1	1.6524	.3666	4.5064	.0000
PARTICULARES	β_2	.1365	.2344	.5824	.5608
WEB	β_3	-.2190	.2328	.9407	.3478
TAYUDAN	β_4	.7799	.5063	1.5404	.1248
BECA	β_5	.2204	.2773	.7947	.4275
TRABAJO	β_6	-.2321	.3685	.6299	.5293
EVCONT	β_7	.2941	.1059	2.7762	.0059

TECNOLÓGICO	β_8	-.2712	.2969	.9134	.3619
OTROS	β_9	.4751	.3079	1.5431	.1241
RTELDE	β_{10}	.1433	.3531	.4060	.6850
RCENTRO	β_{11}	.9947	.4416	2.2524	.0252
RNORTE	β_{12}	1.0519	.4788	2.1969	.0290
RSUR	β_{13}	.5520	.2916	1.8927	.0596

$L_{\max} = -245.402$, $AIC = 522.803$, $BIC = 578.758$

Los resultados obtenidos con el AIC y el BIC corroboran la idea de que la distribución de Poisson generalizada inflada de ceros mejora a las estimaciones realizadas hasta el momento.

4. Conclusiones y discusión

La distribución clásica de Poisson ha sido tradicionalmente utilizada en los trabajos de naturaleza econométrica cuando se intenta explicar una variable discreta en función de otras (discretas o no). Sin embargo, el uso de esta última suele producir ajustes que no superan el test de la Chi-cuadrado. La razón estriba en que la mayoría de los datos empíricos que aparecen en la literatura correspondiente presentan el fenómeno de sobre-dispersión (varianza mayor que la media) y la distribución de Poisson se muestra como un modelo no adecuado ya que para ella la varianza y la media coinciden. En este caso resulta aconsejable partir de una distribución que permita recoger dicho fenómeno.

La distribución generalizada de Poisson se presenta como caso alternativo a la distribución de Poisson como una extensión natural de la primera y que la incluye como caso particular. En este trabajo se utiliza el modelo inflado de ceros, válido cuando hay un porcentaje elevado de ceros en la muestra, utilizándose el método de máxima verosimilitud como procedimiento para estimar los parámetros del modelo. Puesto que los paquetes estándar de estadística existentes en el mercado (STATA, EViews, etc.) no contienen en su catálogo la distribución generalizada de Poisson, utilizando el software Mathematica (versión 9.0.) se desarrolla un programa que presenta la ventaja de calcular de manera simbólica la matriz de información observada así como las varianzas asintóticas de los estimadores.

Se ilustra el modelo presentando como caso de estudio la determinación de factores explicativos del uso de la tutoría universitaria en la asignatura de Matemáticas en los estudios de Economía y Empresa en la ULPGC, en la que se ha detectado una reducción del número de estudiantes que utilizan el recurso de la tutoría académica. Los resultados obtenidos de los modelos aplicados permiten concluir que la distribución de Poisson generalizada inflada de ceros proporciona el mejor ajuste.

Si bien es reconocido que la tutoría académica mejora la calidad del proceso educativo, facilitando la comunicación entre estudiantes y profesores y promoviendo un ambiente de confianza como proponen Arco y Fernández (2011) y Sáiz y Román (2011). El desarrollo del curso permite observar que los estudiantes, en general, infrautilizan este recurso que les puede proporcionar una mejora en su rendimiento académico y que además puede reducir la tasa de abandono. En el análisis llevado a cabo no se pretende redundar en las bondades de las tutorías académicas sino incidir en las variables que pueden explicar la baja asistencia a las mismas en el contexto de la asignatura objeto de estudio.

Del análisis realizado resulta significativo que asisten a tutorías aquéllos que resuelven sus dudas en clase de teoría, lo que se puede explicar por el hecho de que los estudiantes que preguntan en clase son los que están siguiendo la asignatura. Una buena nota en la evaluación continua promueve la asistencia a tutorías, esto podría justificarse por la motivación que produce en el estudiante el éxito en el seguimiento continuado de la materia y lo que ello le repercutirá en el resultado del examen final. También son significativas las variables que agrupan a los estudiantes cuya residencia se localiza en el centro y en la zona norte de la isla, el acceso a la Facultad desde estas zonas requiere un mayor tiempo de desplazamiento por ello se puede entender que éstos aprovechen que las tutorías están programadas a continuación de su horario de clase para acudir a las mismas.

Para concluir, el caso de estudio es especialmente adecuado para mostrar la utilidad de la metodología presentada dado que incluye un elevado número de ceros reflejo del alto número de veces que los estudiantes contestan que no han asistido a tutorías.

Agradecimientos

Los autores agradecen los comentarios y sugerencias realizados por dos evaluadores anónimos así como por el Editor de la revista.

Referencias bibliografícas

1. H. Akaike (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19 (6), 716-723.
2. F. Arbizu, C. Lobato y L. Del Castillo (2005). Algunos modelos de abordaje de la tutoría universitaria. *Revista de Psicodidáctica*, 10, (1), 7-22.
3. J. L. Arco y F. D. Fernández (2011). Eficacia de un programa de tutorías entre iguales para la mejora de los hábitos de estudio del alumnado universitario. R*evista de Psicodidáctica*, 16 (1), 163-180.
4. R. Cano (2009). Tutoría universitaria y aprendizaje por competencias. Cómo lograrlo". *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 12 (1),181-204.
5. P. C. Consul (1989). Generalized Poisson distributions. Properties and applications. Marcel Dekker, Inc. New York and Basel.
6. P. C. Consul y G. C. Jain (1973). A generalization of the Poisson distribution. Technometrics, 15, 4, 791-799.
7. N. Dávila Cárdenes, M. D. García-Artiles, J. M. Pérez-Sánchez y E. Gómez-Déniz (2015). Un modelo logit asimétrico que puede explicar la probabilidad de éxito en el rendimiento académico. Revista de Investigación Educativa, 33, 1, 27-45.
8. D. Dean y S. Paul (2000). Score tests for zero inflation in generalized linear models". The Canadian Journal of Statistics, 27, 3, 563-570.
9. F. Famoye y K. P. Singh (2006). Zero-innflated generalized Poisson regression model with an application to domestic violence data. Journal of Data Science, 4, 117-130.
10. P. L. Gupta, R. C. Gupta y R. C. Tripathi (2004). Score test for zero inflated generalized Poisson regression model. Communications in Statistics: Theory and Methods, 33, 1, 47-64.
11. J. Gairín, M. Freixas, C. Guillamón y D. Quinquer (2004). La tutoría académica en el espacio europeo de la Educación Superior. Revista Interuniversitaria de Formación del Profesorado, 18 (1),.61-77.
12. N. García Nieto (2008). La función tutorial de la Universidad en el actual contexto de la Educación Superior. Revista Interuniversitaria de Formación del Profesorado, 22,1, 21-48.
13. L. Herrera (2007). La acción tutorial en el Espacio Europeo de Educación Superior. Memoria del proyecto de innovación en tutorías para segundo curso de la titulación de maestro. (PIT 034). Universidad de Granada, 4-29.
14. B. G. Leroux (1992). Consistent Estimation of a Mixing Distribution, The Annals of Statistics, 20, 1350-1360.
15. F. Michavila y García Delgado (eds.) (2003). La tutoría y los nuevos modelos de aprendizaje en la Universidad. Madrid: CAM-Cátedra UNESCO.
16. J. Mullahy (1986). Specification and testing of some modified count data models. Journal of Econometrics, 33 (3), 341-365.
17. Romero, C., Zurita Ortega, F. y Zurita Molina, F. (2010). La autonomía y orientación en el Espacio Europeo de Educación Superior mediante el portafolio y la tutoría. Estudios sobre Educación, 19, 261-282.
18. M. C. Sáiz y J. M. Román (2011). Cuatro formas de evaluación en educación superior gestionadas desde la tutoría. Revista de Psicodidáctica, 16 (1), 145-161.
19. R. Sans Oro (2005). Integración del estudiante en el sistema universitario. La tutoría. Cuadernos de Integración Europea, 2, 69-95.
20. S. Wolfram (2003). The Mathematica Book. Wolfram Media, Inc.