II Jornadas de Enseñanza e Investigación Educativa en el campo de las Ciencias Exactas y Naturales *Actas*, II: 62-68, 2009. La Plata.

UNA APLICACIÓN DEL PROCESO DE POISSON EN LA ENSEÑANZA Y APRENDIZAJE DE LA ESTADÍSTICA.

LEDESMA, A. I.

Facultad de Ingeniería. Facultad de Ciencias Económicas. Universidad Nacional de La Plata.
La Plata, 1900
aledesma@fibertel.com.ar ; alicia.ledesma@ing.unlp.edu.ar

RESUMEN

En los últimos años se ha dado un incremento en la preocupación social por los problemas relacionados con la calidad de los servicios, y en particular, de la enseñanza universitaria. El objetivo de este trabajo es presentar una propuesta que sirva de orientación en el aprendizaje de algunas técnicas y metodologías estadísticas adecuadas para el alumno de grado en distintas carreras universitarias. Se pretende lograr una mejor enseñanza de la asignatura Estadística basándose en la resolución de problemas y de casos prácticos con datos reales de diversos aspectos del ámbito de la tecnología y de las ciencias. Para lograr con los objetivos planteados se presenta, a modo de ejemplo, una aplicación al estudio del proceso de Poisson. En particular se realiza un estudio estadístico del tráfico de automóviles particulares que pasan por un punto fijo de la autopista La Plata-Buenos Aires.

Palabras clave. enseñanza y aprendizaje de estadística, proceso de poisson.

INTRODUCCIÓN

En los últimos años se ha dado un incremento en la preocupación social por los problemas relacionados con la calidad de los servicios, y en particular, de la enseñanza universitaria. El objetivo de este trabajo es presentar una propuesta que sirva de orientación en el aprendizaje de algunas técnicas y metodologías estadísticas adecuadas para el alumno de grado en distintas carreras universitarias. Se pretende lograr una mejor enseñanza de la asignatura Estadística basándose en la resolución de problemas y de casos prácticos con datos reales de diversos aspectos del ámbito de la tecnología y de las ciencias. Para lograr con los objetivos planteados se presenta, a modo de ejemplo, una aplicación al estudio del proceso de Poisson. En particular se realiza un estudio estadístico del tráfico de automóviles particulares que pasan por un punto fijo de la autopista La Plata-Buenos Aires. El punto fijo tomado en esta presentación es el peaje situado en la localidad de Hudson. A partir de datos experimentales se ha de establecer un modelo teórico de comportamiento que se quiere modelar. Se propone introducir el proceso de Poisson en forma sencilla, de manera que el alumno sea capaz de entender sus propiedades principales. El trabajo se centró en la necesidad de profundizar en el análisis de datos reales y el desarrollo de casos prácticos que propicien una mejor implicación del estudiante en el proceso de enseñanza y aprendizaje.

METODOLOGÍA

Para poder tratar con la incertidumbre existente en los procesos de índole estadística es necesario estimar a partir de muestras de la población en estudio un modelo de tipo estadístico que caracterice al proceso que dio origen a los datos. Se tratará el aprendizaje del tema ya mencionado, mediante un caso práctico con datos reales, lo cual constituye una aproximación a la realidad por procedimientos empíricos. Esto implica que el estudiante debe conocer los diversos aspectos teóricos del modelo que se quiere estudiar. A los alumnos se les entregará con suficiente anticipación el material que les permitirá conocer la información teórica y práctica *objetiva y científica* sobre el tema a tratar, para conseguir así los objetivos cognoscitivos que nos proponemos. La lectura de este material por parte del los alumnos, previamente a ser tratado en clase, permitirá orientarlos y motivarlos.
La presentación será planificada y presentada de una manera cuidadosa para que resulte adecuada y clara, de manera que el alumno entienda tanto el objetivo final de la actividad como la manera en que luego deberá presentar los resultados. Es muy importante en la didáctica de la Estadística tener en cuenta que vale la pena realizar un buen trabajo con datos reales para crear en el alumno el hábito del tratamiento estadístico de datos numéricos y además para que desarrolle la habilidad necesaria para el ajuste de modelos teóricos a datos experimentales y luego realice las necesarias pruebas de diagnóstico de validez del modelo propuesto. Los alumnos trabajarán en grupos de tres y tanto el profesor como los auxiliares docentes los irán guiando en sus estudios. Cada grupo presentará luego un informe del trabajo realizado. Así mismo cada comisión expondrá el trabajo en clase, de manera que pueda establecerse un debate entre todos los alumnos de la clase. En este debate se analizarán aquellos conceptos que de acuerdo a la experiencia de los docentes ofrecen mayores dificultades al estudiante, con el propósito de reflexionar acerca de los medios más adecuados para reducir estas dificultades, por ejemplo, bibliografía adecuada, ejercicios ilustrativos, uso de software, etc.

FUNDAMENTACIÓN TEÓRICA

Proceso Puntual o de Conteo

Según establece Papoulis (1991), un proceso estocástico $\{N(t), t \geq 0\}$ es un **proceso puntual o de conteo** si N (t) representa el número total de eventos que han ocurrido hasta el momento t. Así por ejemplo si, N (t) es el número de goles marcado por un futbolista hasta t, $\{N(t), t \geq 0\}$ es un proceso de conteo. En este caso un gol es un **evento.**

Podemos decir que, para que un proceso sea puntual o de conteo, N (t) debe satisfacer:

a. $N(t) \geq 0$
b. N (t) es entero.
c. Si $s \leq t$ entonces $N(s) \leq N(t)$.
d. Para $s < t$, $N(t) - N(s)$ es el número de eventos ocurridos en el intervalo (s, t).

Proceso de Poisson.

Siguiendo lo expuesto por Papoulis (1991), supongamos que se presentan "eventos" en puntos temporales aleatorios, y sea N (t) el número de eventos ocurridos en el intervalo de tiempo [0, t]. Se dice que estos eventos constituyen un proceso de Poisson con tasa λ, λ > 0, si:

(a) $N(0) = 0$
(b) El número de eventos ocurridos en intervalos de tiempo disjuntos son independientes.
(c) La distribución del número de eventos que ocurre en un intervalo dado, depende tan solo de la longitud del intervalo y no de su ubicación.
(d) $\lim_{h \to 0} \dfrac{P\{N(h) = 1\}}{h} = \lambda$
(e) $\lim_{h \to 0} \dfrac{P\{N(h) \geq 2\}}{h} = 0$

De la condición (a) vemos que el proceso se inicia en el tiempo 0. La condición (b), la suposición del *incremento independiente*, dice, por ejemplo, que el número de eventos hasta el momento t [es decir N (t)] es independiente del número de eventos que ocurren entre *t* y *t+s* [que es N (t+s)-N (t)]. La condición (c), la suposición del *incremento estacionario*, señala que la distribución de probabilidad de N (t+s)-N (t) es la misma para todos los valores de *t.* Las condiciones (d) y (e) indican que la probabilidad de que ocurra un evento en un intervalo pequeño de tiempo, de longitud *h*, es de aproximadamente λ.h, mientras que la probabilidad de dos o mas es aproximadamente cero.

Puede demostrarse que bajo las suposiciones anteriores, para un proceso de Poisson con tasa λ, $P\{N(t) = k\} = e^{-\lambda t} \dfrac{(\lambda t)^k}{k!}$, $k = 0,1,2...$ Es decir el número de eventos en un intervalo de longitud t tiene una distribución de Poisson con media λ.t.

APLICACIÓN

Se desarrollará aquí, a modo de ilustración, la propuesta de una aplicación que es posible llevarla a cabo con la participación activa de los alumnos, referente al estudio de un modelo estocástico de Poisson. En esta aplicación se utilizan métodos estadísticos para el estudio de la intensidad de tráfico de automóviles particulares por el peaje de la autopista La Plata-Buenos Aires situado en Hudson. El período que se propone analizar se extiende desde las siete hasta las diez horas de tres días laborables.

Llamaremos $\{N(t),\ \text{con}\ \ t > 0\}$ al proceso de recuentos que representa el número de vehículos particulares que pasan por el punto considerado en el intervalo temporal (0, t] y $\lambda > 0$ a la tasa o intensidad del tráfico de estos vehículos, se llegará a establecer que,

$P\{\, N\,(t) = k\ \} = e^{-\lambda\, t}\,\dfrac{(\lambda\, t)^k}{k\,!}$ con k = 0, 1, 2,…; es decir que, el proceso de recuentos N (t) resulta ser un proceso de Poisson. Se considera entonces que:

* Ha de ser posible dividir el intervalo de tiempo considerado en un número grande de pequeños subintervalos, de manera que la probabilidad de ocurrencia de un evento en cada uno de ellos sea muy pequeña.

* La probabilidad de dos o más ocurrencias en cada subintervalo tiene que ser suficientemente pequeña como para ignorarla.

* La probabilidad de una ocurrencia en cada uno de los subintervalos debe permanecer constante a lo largo del período que se está considerando.

* Una ocurrencia (o no ocurrencia) en un subintervalo no debe afectar a otra ocurrencia (o no ocurrencia) en cualquiera de los otros subintervalos, es decir las ocurrencias deben ser independientes o lo que es lo mismo, en este caso: existencia de independencia estadística entre las variables aleatorias que cuentan el número de vehículos particulares que pasan por el peaje de Hudson en intervalos de tiempo distintos.

Según Campos (2008) los pasos a seguir son:

Primer paso. Establecer el modelo supuesto para el tráfico de vehículos.

Segundo paso. Presentar los datos experimentales obtenidos y compararlos adecuadamente con los valores teóricos correspondientes.

Tercer paso. Realizar el ajuste estadístico del modelo teórico. Si se comprueba que dicho ajuste es satisfactorio, se podrá concluir la validez del modelo propuesto.

Modelo supuesto

Como ya dijimos antes supondremos que el tráfico de vehículos particulares por un el peaje de Hudson de la autopista La Plata-Buenos Aires se comporta como un proceso de Poisson.

Datos experimentales obtenidos

Actualmente de acuerdo a los datos proporcionados por el Órgano de Control de Concesiones Viales (OCCOVI) se observan 2102 automóviles por hora, en la estación de peaje de Hudson en la dirección La Plata hacia Buenos Aires. Considerando intervalos de 15 segundos de duración, que a efectos del análisis del ajuste del modelo se tomaran como unidad de tiempo (1 unidad de tiempo = 15 segundos), se muestran los datos en la Tabla 1. La Tabla 1, debe interpretarse como sigue, X representa el número de automóviles particulares que pasan cada 15 segundos por el peaje considerado. Los valores que toma esta variable van de cero a once. Así puede verse que hay 46 intervalos de 15 segundos cada uno durante los cuales no pasa ningún auto, 137 intervalos en los que pasa un auto particular por cada uno y así sucesivamente.

X	Valores Observados	Frecuencia Relativas Observadas	Probabilidad de Poisson $P(x) = P(X = x) = e^{-2,84} \frac{2,84^x}{x!}$	Valores Esperados
0	46	0,0622	0,0584	43,23
1	137	0,1851	0,1659	122,78
2	166	0,2243	0,2356	174,36
3	153	0,2068	0,2231	165,06
4	108	0,1459	0,1584	117,22
5	79	0,1068	0,0900	68,59
6	28	0,0378	0,0426	31,53
7	11	0,0149	0,0173	12,79
8	8	0,0108	0,0061	4,54
9	3	0,0041	0,0019	1,43
10	0	0	0,0006	0,41
11	1	0,0014	0,0002	0,14

Tabla 1. Valores observados y valores ajustados

La Figura 1 muestra el gráfico de barras de los valores observados, sospechamos que los datos pueden ajustarse a una distribución de Poisson.

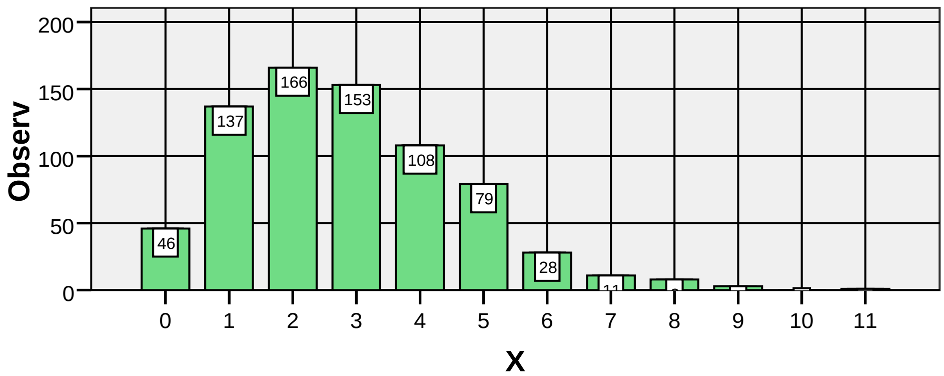

Figura 1. Gráfico de barras de valores observados

Estimamos a la tasa del Proceso de Poisson mediante la media muestral, es decir:

$\hat{\lambda} = \overline{X} = \frac{1}{n} \sum_{i=0}^{k} x_i f_i$ con $f_1 + f_2 + \ldots + f_k = n$.

A partir de los datos que aparecen en las dos primeras columnas de la Tabla 1,

$$\hat{\lambda} = \overline{X} = \frac{0 \times 46 + 1 \times 137 + 2 \times 166 + 3 \times 153 + 4 \times 108 + 5 \times 79 + 6 \times 28 + \ldots + 11 \times 1}{46 + 137 + 166 + 153 + 108 + 79 + 28 + 11 + 8 + 3 + 1} = \frac{2102}{740} = 2,84054$$

Tomamos $\hat{\lambda} = 2,84$ automóviles particulares por unidad de tiempo.

En la Figura 2 pueden compararse los valores observados con los valores teóricos esperados. Como primera aproximación puede decirse que el ajuste es bueno.

Contraste de bondad de ajuste Chi –Cuadrado

Examinamos luego el problema de verificar si el conjunto de datos observados puede ajustarse a una distribución de Poisson o afirmar que proviene de una población con esa distribución. Utilizamos para ello una prueba estadística de bondad de ajuste, en particular la prueba Chi- Cuadrado, que mide, como el nombre lo indica, el grado de ajuste que existe entre la distribución obtenida a partir de la muestra y la distribución teórica que se supone debe seguir esa muestra. Para esta prueba es necesario agrupar o distribuir las observaciones de la muestra en intervalos de clase, preferiblemente del mismo tamaño. Según establece Pérez, C (2001), los pasos a seguir para realizar esta prueba son:

Primero. Formulación de las hipótesis.

H_0: los datos muestrales siguen una distribución de Poisson.

H_1: los datos muestrales no siguen una distribución Poisson.

Segundo. Establecer un nivel de significación.

Se estableció un nivel de significación del 5%, $\alpha = 0.05$.

Tercero. Definir el estadístico de prueba

El estadístico de prueba está definido como:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(f_{oi} - f_{ei}\right)^2}{f_e} \qquad (1)$$

Donde, foi es el total de valores observados que caen en el intervalo i-ésimo, f_{ei} el número esperado de valores en el intervalo i-ésimo y k el número de intervalos de clase en que se distribuyen las observaciones. El estadístico de prueba definido en (1) tiene distribución Chi-Cuadrado con $\nu = 11$ grados de libertad en el caso que estamos considerando.

Cuarto. Comparar el valor crítico de χ^2_{11} obtenido de la tabla con el valor de χ^2 obtenido de (1) para rechazar o aceptar H_o. El valor crítico obtenido para una probabilidad de 0,05 (nivel de significación del 5%) con 11 grados de libertad es: $\chi^2_{11} = 19{,}673$ y el valor del estadístico calculado es $\chi^2 = 16{,}81$. Recordando que la región crítica, al nivel de significación α, se obtiene teniendo en cuenta $P(\chi^2_\nu > \chi^2_{\nu;\alpha}) = \alpha$ el criterio de decisión es:

Si $\chi^2_\nu \le \chi^2_{\nu;\alpha}$ no se rechaza H_o (no se tiene suficiente evidencia para ello)

Si $\chi^2_\nu > \chi^2_{\nu;\alpha}$ se rechaza H_o

Concluimos entonces que con un nivel de significación del 5% no se rechaza la hipótesis nula. El p-valor del contraste encontrado fue 0,1135. Esta aplicación fue simulada mediante el software SPSS. Con lo realizado se acepta el modelo de Poisson como válido, porque las discrepancias no son consideradas como significativas. En definitiva se realiza así un estudio estadístico del tráfico de automóviles particulares que pasan por un punto fijo y determinado de la autopista ya mencionada, en la dirección La Plata-Buenos Aires y se encuentra que en condiciones razonables y bastante generales el número de automóviles que pasan por ese punto considerado en el intervalo temporal (0, t], que indicamos X(t), es un proceso de Poisson con intensidad de tráfico estimada por el valor $\hat{\lambda} = 4 \text{ x } 2.84 = 11.36$ automóviles particulares por minuto.

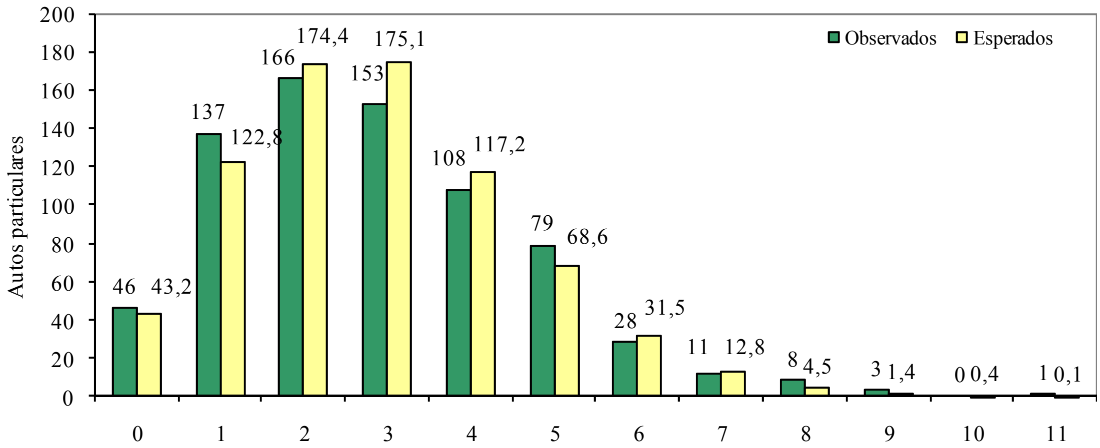

Figura 2. Comparación de los valores observados con los valores teóricos esperados

CONCLUSIONES

Se prueba así que mediante datos muestrales es posible determinar que el tráfico de automóviles, en este caso por un punto determinado de una autopista, es un Proceso de Poisson. En general en cursos donde ya se efectuó esta experiencia se observó interés en el tema y en la manera de enfocarlo, como asimismo buena predisposición para trabajar en grupos.

BIBLIOGRAFÍA

Campos, C.A. (2008); *Aprendizaje de la Estadística a través de casos prácticos*; CPS de Ingenieros (CONSULTING, PROYECTOS Y SISTEMAS, S.L); II Jornadas de innovación docente, tecnologías de la información y de la comunicación e investigación educativa en la universidad de Zaragoza, 2008. http://ice.unizar.es/uzinnova/jornadas/pdf/95.pdf

Papoulis, A (1991). *Variables and Stochastic Processes*. New York, USA. Ed. McGraw-Hill Internacional Editions, Third Edition., pág. 290-298.

Pérez, C. (2001), *Técnicas Estadísticas con SPSS*. Madrid, Prentice Hall, Pearson Educación S.A; pág. 288-289.