

REGRESIÓN LOGÍSTICA PARA DATOS BINARIOS: UNA APLICACIÓN.

Lorenzo Díaz, M^a Carmen; Méndez Naya, Isabel y Porto Vila, Rosalía
Dpto de Econometría y Métodos Cuantitativos
Facultad de Económicas. Universidad de Santiago

RESUMEN.-

En este trabajo se estudia el uso de la regresión logística para estimar y predecir la variable dicotómica "ocurrencia/ no ocurrencia de un día con incendio" en función de una serie de variables independientes.

Por otra parte, también se introduce un análisis de tablas de contingencia con el objeto de discernir posibles grupos diferenciadores en el período temporal analizado.

Una vez seleccionado el mejor modelo de regresión logística y realizadas las estimaciones diarias, se procede a estimar y predecir la variable discreta "número de incendios diarios" mediante la combinación regresión logística-distribución de Poisson.

PALABRAS CLAVE.- Regresión logística, tablas de contingencia, variables categóricas, distribución de Poisson, incendios forestales.

I. INTRODUCCIÓN

La necesidad de predecir adecuadamente la ocurrencia de incendios forestales es de vital importancia para un más eficiente reparto y empleo de recursos de prevención y extinción. Los administradores de los incendios forestales están a menudo envueltos en la toma de decisiones en ambientes de cambio incierto. Decisiones sobre la movilidad diaria de las cuadrillas y aviones contra incendios son algunos ejemplos. Para llevar a cabo esto se necesita información fiable sobre el número de incendios que puedan ocurrir en cada área.

Los primeros modelos de probabilidad utilizados en la predicción del número de incendios provocados por personas estuvieron basados en la distribución binomial y en la distribución de Poisson. En el modelo binomial se utilizaba una estimación del número de personas en la superficie forestal y se operaba bajo la hipótesis de que la probabilidad de que una persona prenda fuego es independiente de que lo haga otra. En seguida se propuso un modelo de probabilidad de Poisson basándose en que la probabilidad de que una persona provoque un incendio (p) era baja y el número de personas en la superficie forestal (n) grande, siendo además la hipótesis bajo la que opera este modelo que el número esperado de incendios por día ($= np$) es constante.

A continuación el objetivo fue desprenderse de la hipótesis tan poco realista de que el número de incendios esperado por día (el parámetro de la distribución de Poisson) permanecía constante y las investigaciones

se dirigieron en la línea de explicar la variabilidad diaria de λ en función de otros factores influyentes en el fenómeno. Aparecen entonces las propuestas de explicar el parámetro a través de la regresión logística.

Hoy en día la regresión logística es un área activa de investigación estadística en la que en los últimos años se suceden los avances metodológicos (Hosmer and Lemeshow (1989)). Esta proliferación en su estudio y análisis es consecuencia fundamental de las limitaciones de la regresión múltiple y del análisis discriminante para predecir el comportamiento de una variable dicotómica.

Así, si nos planteamos el uso de la regresión múltiple no se puede asumir su hipótesis fundamental que es la distribución normal de los errores. Otra dificultad, a la hora de plantearse el uso de la regresión múltiple para predecir una variable dicotómica, es que los valores predichos no pueden ser interpretados como probabilidades ya que no hay ninguna restricción que le imponga a la variable respuesta tomar valores entre 0 y 1.

Por otra parte, el análisis discriminante necesita la hipótesis de normalidad multivariante de las variables explicativas así como la de igualdad de las matrices de varianzas-covarianzas en las dos categorías de la variable respuesta (Press and Wilson (1978)), con lo cual no es posible introducir en el modelo las variables categóricas como explicativas. El modelo de regresión logística requiere supuestos menos restrictivos que el análisis discriminante e incluso cuando las hipótesis de este último se satisfacen la regresión logística también proporciona un buen ajuste.

En este trabajo se hace una aplicación de este tipo de regresión para, posteriormente, estimar el número de incendios diarios mediante una distribución de Poisson (Martell et al.(1987)).

Se estructura en dos secciones. Se comienza, en la primera sección, con una descripción de las principales características de la regresión logística, haciendo mención a la distribución seguida por los errores y a la expresión matemática de la función. Posteriormente, se describe la relación entre la probabilidad estimada por la regresión logística y el parámetro λ de Poisson. En la segunda sección, se hace un análisis de tablas de contingencia de la variable Y "ocurrencia/ no ocurrencia de un día con incendio" respecto a las distintas estaciones (primavera, verano, otoño e invierno) del período a estudiar (1988-1993). Puesto que el resultado obtenido ha sido la clara presencia de diferencias por grupos entre estaciones, se presentan las estimaciones de las correspondientes regresiones así como las bondades de ajuste. Una vez seleccionado el mejor modelo, se estima "el número de incendios diarios" para el período 1988-1994 a través de la estimación diaria del parámetro λ de Poisson.

II. LA REGRESIÓN LOGÍSTICA Y LA DISTRIBUCIÓN DE POISSON.

A. LA REGRESIÓN LOGÍSTICA

La variable respuesta cuyo comportamiento se trata de predecir, es una variable dicotómica Y codificada como 1 ó 0, representando, respectivamente, la presencia o ausencia de un suceso aleatorio. Se asume que la variable Y depende de un conjunto de variables explicativas $X=(X_1,...,X_n)$. Para cada valor observado de X, la variable Y es una variable de Bernoulli con función de distribución de probabilidad:

Y	1	0
P(Y)	$\Pi(X)$	$1-\Pi(X)$

siendo la probabilidad de que ocurra $Y=1$ su valor esperado, $E(Y/X)=\Pi(X)$. Así del modelo de regresión

$$Y(X) = E(Y/X) + \varepsilon = \Pi(X) + \varepsilon$$

se tiene la distribución de probabilidad para los errores

ε	$1-\Pi(X)$	$-\Pi(X)$
P(ε)	$\Pi(X)$	$1-\Pi(X)$

siendo por tanto ε una variable de Bernoulli con media 0 y varianza dependiente de los valores de las variables explicativas

$$V(\varepsilon)=\Pi(X)(1-\Pi(X)).$$

Es evidente entonces que en esta situación no se puede establecer un modelo de regresión basado en la hipótesis de normalidad de los errores.

Con la notación introducida, la hipótesis fundamental en la regresión logística es la linealidad de

$$\ln\left(\frac{p(X)}{1-p(X)}\right)$$

respecto de las variables explicativas $X=(X_1, \dots, X_n)$

$$\ln\left(\frac{p(X)}{1-p(X)}\right) = b_0 + b_1 X_1 + \dots + b_n X_n$$

equivalentemente

$$\frac{p(X)}{1-p(X)} = e^{b_0} e^{b_1 X_1} \dots e^{b_n X_n}$$

Se parte entonces de la hipótesis de una relación exponencial entre los distintos valores de las variables explicativas y el cociente entre la probabilidad de que ocurra el suceso ($Y=1$) y no ocurra ($Y=0$); además este efecto exponencial de cada variable es multiplicativo con el efecto de las demás .

Resolviendo ahora la ecuación (3) para $\Pi(X)$ se obtiene la relación no lineal entre la probabilidad de ($Y=1$) y las variables explicativas:

$$\Pi(X) = \frac{1}{1 + e^{-b_0 - z}}$$

donde $z = \beta_1 X_1 + \dots + \beta_n X_n$, siendo los β_i , $i=1, \dots, n$, los coeficientes a estimar.

La función $\Pi(X)$ es creciente ante una variación positiva de la combinación lineal z de las variables explicativas.

La curva logística estimada

$$P(Z) = \frac{1}{1 + e^{b_0 + e^{-z}}}$$

tiene un punto de inflexión en $z=0$ y asíntotas $y=0$ e $y=1$. Además presenta un comportamiento simétrico con respecto a su punto de inflexión en el sentido de que $P(Z) = 1 - P(-Z)$. De hecho, la curva (5) es solución de la ecuación diferencial $P'(Z) = P(Z)(1 - P(Z))$ y, por tanto, implícitamente se está suponiendo que la variación de la probabilidad de un suceso disminuye a medida que aumenta esta, siendo máxima la variación en el punto de inflexión $z=0$.

B. LA DISTRIBUCIÓN DE POISSON

Puede obtenerse mediante el límite de una secuencia de distribuciones binomiales cuando el número de casos N tiende a infinito, la probabilidad p tiende a cero y el producto Np permanece igual a una constante λ . Es decir, el modelo de Poisson se presenta cuando se estudia el número de veces que tiene lugar un suceso (U), cuya probabilidad es muy pequeña, en un número grande de observaciones N , siendo la frecuencia de ocurrencia del evento constante en el tiempo y en el espacio.

La probabilidad de que una variable aleatoria de Poisson con parámetro $\lambda > 0$ tome el valor u es

$$f(u; \lambda) = \frac{\lambda^u e^{-\lambda}}{u!} \quad u=0,1,2,\dots; \lambda > 0$$

La probabilidad de que U tome valores mayores o iguales que uno se expresa, por tanto, como:

$$P(U \geq 1) = 1 - e^{-\lambda}$$

siendo el parámetro λ la media y la varianza de la variable.

Si se supone ahora que el parámetro λ depende de un conjunto de variables independiente $X = (X_1, \dots, X_n)$, y se asume que la probabilidad de que ocurra uno o más sucesos en presencia de $X = (X_1, \dots, X_n)$ viene estimada a través de la curva logística por $\Pi(X)$ tal como se indica en (4), se tiene que

$$P(U \geq 1) = 1 - e^{-\lambda(X)} = \Pi(X)$$

de donde

$$I(X) = -\ln(1 - \Pi(X))$$

III REGRESIÓN LOGÍSTICA Y PROBABILIDAD DE UN DÍA CON INCENDIO

A. CONSIDERACIONES GENERALES

La ignición del incendio se desencadena a través de mecanismos que afectan a la temperatura del combustible y a su contenido de agua. Los aspectos más importantes de esta relación están indicados en los trabajos de Rothermel(1972), y deben ser considerados a la hora de construir un modelo de predicción de ocurrencia de incendio.

Así, para establecer un modelo matemático para aproximar la probabilidad de ignición se debe tener en cuenta que ésta aumenta con el incremento de la temperatura del combustible y disminuye al aumentar su humedad.

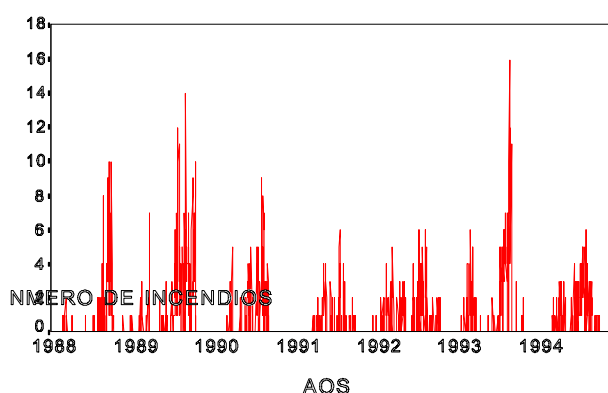
Puesto que no se conocen las relaciones exactas entre la temperatura y la humedad del combustible vegetal que originan la ignición, ni éstas son medidas que figuren en los datos históricos de incendios¹, en los procesos de estimación se consideran como variables explicativas aquéllas cuyos valores son conocidos y que afectan a la temperatura y a la humedad del combustible. Algunas de estas variables son: la temperatura ambiente, humedad climatológica, orientación espacial, velocidad y dirección del viento, etc.

Aún teniendo en cuenta que un modelo de predicción de la probabilidad de ignición se debería establecer a partir de datos en tiempo real sobre la temperatura y la humedad del combustible vegetal, actualmente uno de los modelos más usados es el logístico, utilizando como variables explicativas las citadas anteriormente.

B. ANÁLISIS PRELIMINAR: VARIACIÓN ESTACIONAL EN LA OCURENCIA DIARIA DE LOS INCENDIOS FORESTALES.

Se utilizan como base del análisis los datos históricos (Gráfica 1) del número de incendios por día en el período 1988-1994 en un área de 321 km² del sureste de Pontevedra. Estas estadísticas han sido facilitadas por el ICONA. La causa de la elección de esta área de estudio está estrechamente relacionada con la presencia en este espacio de la estación metereológica de Lourizán, de la cual proceden los datos de las variables explicativas utilizadas en este trabajo (Cuadro2).

Gráfico 1. Número de incendios diarios en el período 1988-1994.



¹ Incluso en los países punteros en la investigación de la ocurrencia de incendios forestales, tales como EEUU y Canadá, no están disponibles las estadísticas que recojan información de estas variables (Martell et al. (1987), Vega (1994)).

Antes de iniciar el proceso de estimación de la dependencia de la variable dicotómica Y igual a:

1 si día con incendio

0 si día sin incendio

con las variables explicativas (Cuadro 2) influyentes, se plantea la siguiente cuestión: ¿Existen diferencias significativas en el comportamiento de la variable a lo largo de las diferentes estaciones incluidas en el período de estudio?

La ocurrencia diaria de incendios -proceso aleatorio discreto- es infrecuente y altamente variable (Gráfica 1). Mediante el uso de la tabla de contingencia correspondiente se contrasta si la variable respuesta "número de incendios por día" se registra independiente de las estaciones anuales. Ahora bien, para evitar el problema de demasiadas celdas con nulos o con poca frecuencia -los tests de las tablas de contingencia no son fiables cuando las frecuencias esperadas son menos de cinco en más del 20% de las celdas- se realiza el análisis de tablas de contingencia de la variable "ocurrencia/no ocurrencia de un día con incendio" por estaciones (verano, otoño, invierno y primavera).

En el cuadro 1 el número de casos esperados (ESP) en las distintas celdas no coinciden con los datos observados bajo el supuesto de independencia. El nivel de significación asociado al estadístico ϕ^2 es menor que 0,01 y por lo tanto se rechaza la hipótesis de que las dos variables son independientes. Es decir, el número de días con incendio no es independiente de las distintas estaciones del año.

CUADRO 1. Tabla de contingencia de días con o sin incendio por estaciones para 1988-1993.

		ESTACIONES DEL AÑO						
		INV	PRI	VER	OTO	%TOT	PHI	SIG
0	OBS	465	404	250	451	75,8	0,36	0,000
	ESP	411	380	388	391			
1	OBS	77	97	262	64	24,2		
	ESP	131	121	124	124			

Dado que las estaciones marcan grupos diferentes para la variable dicotómica "presencia/ausencia de incendio por día" se procede a dos modelos de análisis:

- 1.- Introducir en un sólo modelo de regresión logística (RL) la variable estación como una variable categórica, identificadora de los cuatro grupos (primavera, verano, otoño e invierno).
- 2.- Estimar un modelo de regresión para cada estación (RLV, RLO, RLI, RLP).

²El coeficiente phi modifica el chi cuadrado de Pearson dividiéndolo por el tamaño de la muestra y extrayendo la raíz cuadrada del resultado. Su uso se justifica por ser una medida que minimiza la influencia del tamaño de la muestra en el test χ^2 .

C. ANÁLISIS DE REGRESIÓN LOGÍSTICA POR ESTACIONES.

1. VARIABLES INDEPENDIENTES

En el área objeto de este estudio, se supone un comportamiento estable en el uso de la tierra y en las condiciones de la vegetación, y por tanto se parte de que la ocurrencia del fuego está fuertemente influenciado por las condiciones de temperatura y humedad del combustible forestal y en consecuencia por las condiciones meteorológicas. En España, a diferencia de otros países³, no hay un sistema de evaluación conjunto de tales condiciones climatológicas en su incidencia en la humedad del combustible. Dado que, como se comentó anteriormente, no se dispone de datos de las variables explicativas que serían necesarias para la predicción, hemos tomado como factores influyentes las variables climatológicas citadas en el cuadro 2 proporcionadas por la estación meteorológica de Lourizán. Asumimos que los datos facilitados por la estación son representativos de toda la zona objeto de estudio.

CUADRO 2. Variables independientes consideradas en la regresión logística.

VARIABLES	DESCRIPCIONES
TMAX	Temperatura máxima (°C).
TMME	Temperatura máxima media de los dos días anteriores.
TMIN	Temperatura mínima (°C).
HMAX	Porcentaje de humedad máxima.
HMI	Porcentaje de humedad mínima.
HMME	Porcentaje de humedad mínima media de los dos días anteriores.
P	Precipitación (litros/m ² de agua caída por día).
DV	Dirección del viento.
RV	Velocidad del viento (km/h.).
HRSOL	Horas de sol diarias.
ESCAR	Escarcha existente o no por día.

2. SELECCIÓN DEL MODELO: ESTIMACIÓN Y PREDICCIÓN

Para proceder a la estimación del modelo en la práctica tendremos $m=2.070$ observaciones:

$$(X_{11}, \dots, X_{111}, Y_1), (X_{1i}, \dots, X_{11i}, Y_i), \dots, (X_{1m}, \dots, X_{11m}, Y_m)$$

correspondientes al período de estimación 1988-1993⁴; siendo Y_i el valor de la variable respuesta dicotómica

³Van Wagner, 1987, para Canadá, explica el Canadian Forest Fire Danger Rating System (FFDRS) compuesto por seis índices de medidas de humedad de varios componentes del complejo vegetal.

⁴Dada la falta de datos disponibles para la variable DV para los períodos 1/11/91-30/11/91, 1/5/92-30/6/92 y 1/8/93-31/8/93, y dada la evidencia de la importancia de esta variable en el estudio de la ocurrencia de incendios forestales, se ha procedido a eliminar del análisis los casos correspondientes. De esta forma, en vez de 2.188 casos correspondientes al total de días del período 1988-1993 aparecen tan sólo 2.070.

"presencia/ausencia de días con incendio" cuando los valores observados de las variables explicativas del cuadro 2 son X_{1i} , ..., X_{11i} para un día i determinado.

El objetivo, ahora, es identificar los principales factores del cuadro 2 que deberían incluirse en los cinco modelos de regresión logística propuestos: LR, LRP, LRV, LRO, LRI.

En la regresión logística, la herramienta utilizada para estimar los parámetros del modelo es el método de máxima verosimilitud. Su forma de proceder consiste en construir una función de probabilidad de la muestra en función de los parámetros a estimar; y a continuación determinar los estimadores máximo-verosímiles de los parámetros eligiéndolos como aquellos que maximizan esta función. Esto es, se seleccionan los coeficientes ($\beta_0, \beta_1, \dots, \beta_n$) que hacen que el modelo represente mejor los datos observados.

Las variables de los distintos modelos -cada uno con término constante- del cuadro 3 han sido seleccionadas mediante el método stepwise hacia adelante (Forward Stepwise).

CUADRO 3. Variables seleccionadas

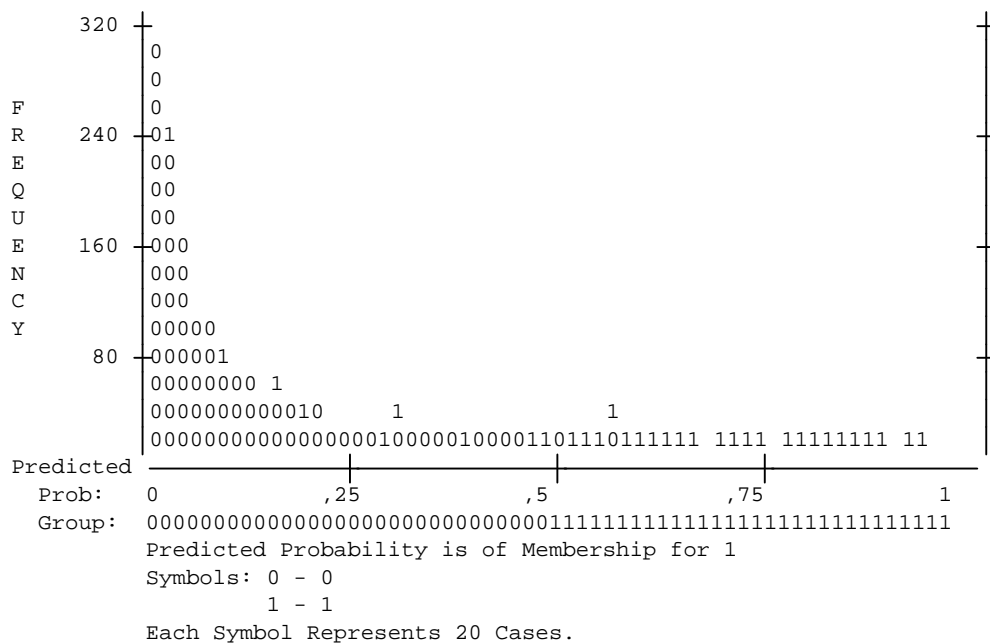
MODELOS	VARIABLES	%T	%1
LR	DV, EST ⁵ , HMME, HMI, RV, TMAX, TMME, TMI	83,12	53,40
LRP	DV, HMI, TMME	83,03	30,93
LRV	DV, HMI, RV, TMME	59,78	40,32
LRO	HMME, HMI, TMME	90,49	48,44
LRI	HMME, HMI, TMAX, TMI	86,46	36,36

El modelo más realista con los datos observados es LR, es decir, aquel que introduce una única regresión logística para todo el período recogiendo la diferenciación por estaciones mediante la variable categórica EST. En el cuadro 3 en la última columna, se observa como es LR la que clasifica mejor los días con incendio, de hecho el 53,4% de los días con incendio han sido correctamente clasificados. A su vez, el 83,12% de los días totales fueron bien clasificados.

En general, si la probabilidad estimada de un evento es menor que 0,5, se predice que el evento no ocurrirá. Si la probabilidad es mayor que 0,5, se predice que el evento ocurrirá. De la figura 2 - histograma de la probabilidad estimada de los días con incendio(1) y sin incendio (0) del modelo LR- se observa como hay ochenta días con incendio con una probabilidad estimada de menos del 0,25. Teniendo en cuenta que si el modelo se ajusta bien, los casos para los cuales el evento ha ocurrido debería situarse a la derecha de 0,5, mientras que los casos que no ocurrieron se situarían a la izquierda, parece conveniente tomar una probabilidad más baja para catalogar a un día sin incendio.

⁵Variable categórica identificativa de las cuatro estaciones del año.

Gráfico 2. Grupos obseavados y la probabilidad predicha por el modelo LR para 1988-1993.



Una vez ajustado el modelo de la regresión logística y contrastado la significación de los coeficientes de las variables independientes (Cuadro 4), el siguiente paso será interpretar los coeficientes estimados del modelo.

Cuadro 4. Resultados de la regresión logística LR.

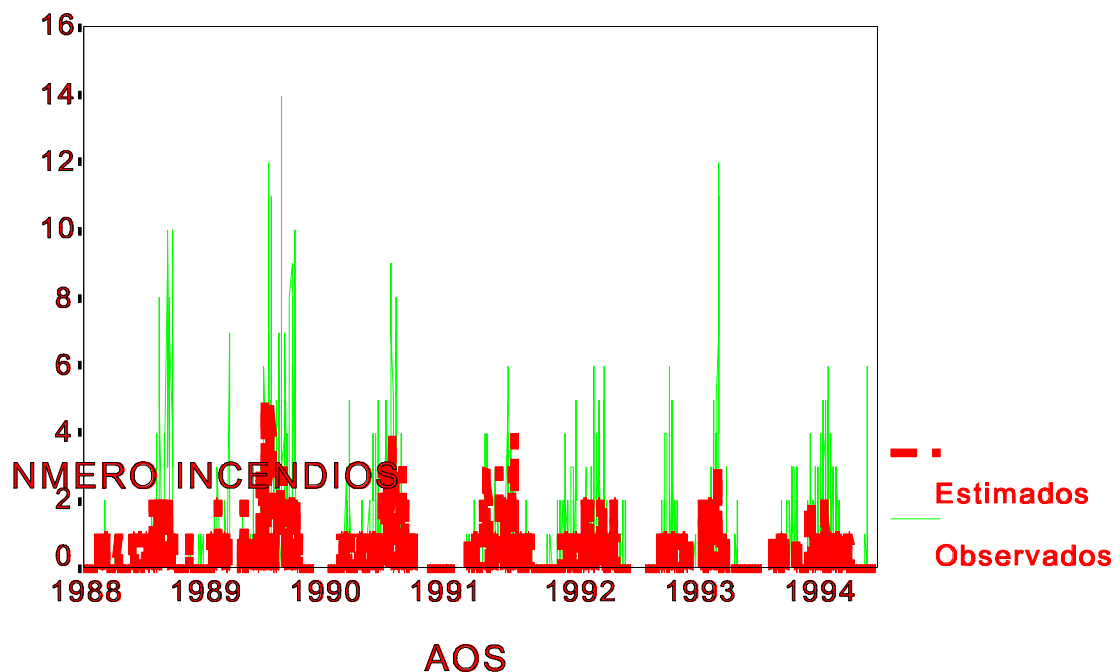
VARIABLES	EXP(β)	R	Significación
Norte (Referencia)	4,0939		
Suroeste	0,4912	0,000	0,5721
Noreste	1,0875	0,0000	0,9466
Sureste	2,3939	0,0000	0,5470
Noroeste	7,3962	0,0000	0,2608
Este	2,1097	0,0000	0,5762
Sur	1,2570	0,0000	0,8619
Oeste	0,0097	0,0000	0,5898
Primavera	0,6307	-0,0714	0,0002
Verano	1,6152	0,0554	0,0027
Otoño(Referencia)	0,8240		
Invierno	1,1912	0,0000	0,2966
HMME	0,9746	-0,0661	0,0005
HMI	0,9669	-0,1011	0,0000
RV	1,0937	0,0296	0,0455
TMAX	1,0734	0,0375	0,0224
TMME	1,2241	0,1135	0,0000
TMI	0,9046	-0,0509	0,0049

El estadístico R, de rango -1 y +1, se usa para calcular la correlación parcial entre la variable dependiente y cada una de las variables independientes. Un valor pequeño nos indica que la variable tiene una contribución parcial al modelo. Un valor positivo nos indica que ante un incremento en el valor de la variable hay un aumento en la probabilidad de que ocurra el evento. Lo contrario ocurre si el valor es negativo.

Hay que admitir desde el punto de vista de la ecuación 3) que el número de veces que es más probable un día con incendio frente a un día sin incendio aumenta exponencialmente con cada valor de cada variable explicativa siempre que presente un $EXP(\beta_i)$ mayor que la unidad, y disminuye exponencialmente si el $EXP(\beta_i)$ es menor que la unidad. Según la columna del cuadro 4 bajo el título $EXP(\beta_i)$, un incremento en las variables que muestran una $EXP(\beta_i)$ mayor que 1 (viento del noroeste, viento del norte, temperatura máxima, verano) provocan un aumento en el ratio $\frac{\Pi(X)}{1-\Pi(X)}$, mientras que aquellas variables con el $EXP(\beta_i)$ menor que la unidad (viento del oeste, viento del suroeste, el invierno, la humedad mínima) hacen que el ratio disminuya.

Finalmente, en el gráfico 3 se expone el número de incendios diarios observados para el período 1988-1994 y el número de incendios diarios predichos mediante la fórmula 7) que integra la estimación de la logística con la distribución de Poisson (para más detalle consultar Lorenzo e Iglesias (1995)). Se observa como la media de la distribución de Poisson se ajusta mejor a los datos observados de valor bajo que a los de valor elevado.

Gráfico 3. Número de incendios forestales diarios observados y predichos.



IV. CONCLUSIONES.

El modelo de regresión logística, como se ha indicado, es ampliamente utilizado dada la relativa sencillez en su manejo y en su interpretación. Sin embargo, uno de sus inconvenientes a la hora de predecir la ocurrencia de incendio es suponer un crecimiento de la probabilidad de ignición simétrico con respecto al punto de inflexión.

En el estudio de la probabilidad de ignición del incendio forestal y, pensando por ejemplo en la temperatura como variable explicativa, sería más adecuado un modelo que reflejase asimetría en el crecimiento, siendo éste más rápido a medida que aumenta la temperatura. En esta línea se propone para trabajos futuros el estudio del posible uso de una modificación de la curva de Gompertz (curva monótona creciente entre 0 y 1 con crecimiento asimétrico) en la predicción de la ocurrencia de un día con incendio.

Además sería interesante en el futuro, y todo con vistas a facilitar al gestor de recursos un instrumento fiable de predicción en ocurrencia de incendio, contrastar el modelo causal presentado en esta comunicación así como el basado en la curva de Gompertz con el enfoque de series de tiempo multivariante para datos diarios. Una primera aproximación al problema, pero con datos mensuales, aparece en Lorenzo y Chas (1996).

V. BIBLIOGRAFÍA

- * Hosmer, D.W. and Stanley Lemeshow. 1989. John Wiley & Sons. New York.
- * Lorenzo D., M.C.; Iglesias P., M.C.; 1995. "Modelos de probabilidad para el estudio de la ocurrencia de incendios forestales". IX Congreso Nacional ASEPELT-España Santiago de Compostela.
- * Lorenzo D., M.C.; Chas A., M.L. 1996. "Forest Fires in Spain: an Univariate Time Series Analysis". LI International Conference Econometrics of Environment and Transdisciplinarity Portugal.
- * Martell, D.L.; S. Otukol and B.J. Stocks. 1987. A logistic model for predicting daily people-caused forest fire occurrence in Ontario. Can. J. For. Res. 17:394-401.
- * Press S. James and Sandra Wilson. 1978. Choosing between logistic regression and discriminant analysis. Journal of the American Statistical Association. 73(364):699-705.
- * Quiñoá L.; Méndez N.; Porto V. y otros. 1992. "Aspectos metodológicos en la planificación de la lucha contra los incendios forestales". VI Congreso Nacional ASEPELT-España Granada.
- * Rothermel, R.C. 1972. A mathematical model for predicting fire spread in wildland fields". USDA Forest Serv. Res. Pap. Int-115.
- * Van Wagner, C.E. 1987. Development and Structure of the Canadian Forest Fire Weather Index System. Canadian Forestry Service. Forestry Technical Report 35. Ottawa.
- * Vega García, C. 1994. Predicting human-caused forest fire occurrence in Whitecourt Forest. Tesis. Alberta (Canadá).