
12. Aplicaciones de los modelos lineales generalizados para predecir la distribución de mustélidos nativos e introducidos

Jaime R. Rau, Ramón C. Soriguer, Juan F. Beltrán, David R. Martínez

Introducción

El objetivo de este estudio fue elaborar un modelo matemático que permita predecir con cierta robustez y fiabilidad la probabilidad de encontrar un mustélido nativo (huillín, *Lontra provocax*) o introducido (visón americano, *Mustela vison*) en un curso determinado de agua si se dan o se satisfacen una serie de requerimientos ecológicos.

Métodos

Diseño y tipo de muestreo. Se realizó un muestreo sistemático de los cursos de agua de la IX, X y XI Regiones del sur de Chile en base a una serie de transectos E-W entre las Latitudes 38°S-49°N. En cada uno de estos transectos se identificaron los cursos de agua más próximos y en ellos se registró la presencia/ausencia de mustélidos recorriendo y detectando la presencia de signos (e.g., fecas, huellas, madrigueras) en parcelas que variaron entre 200 a 600 m hasta poder ser consideradas positivas o negativas. En todas ellas se registraron variables y categorías fisiográficas, estructurales, del entorno y estado de conservación del cauce y de presuntos competidores y presuntas presas potenciales. La

información recogida ha sido principalmente de tipo categórica y permitió la construcción de una matriz de datos correspondiente a 85 parcelas.

Proceso analítico. Mediante una aproximación por modelos lineales generalizados, GLM (e.g., McCullough y Nelder 1989), se elaboró un modelo predictivo de la capacidad de selección de hábitat de huillines y visones. GLM son un amplio espectro de modelos en los cuales la regresión lineal es sólo un caso particular. De esta forma GLM facilita una gama más amplia de relaciones entre la variable dependiente (o variable respuesta) y la(s) variable(s) independiente(s) o explicatoria(s). GLM además permite el empleo de funciones de error aplicables en casos en los que no lo es la distribución normal.

En síntesis, un modelo matemático GLM se define por tres componentes: un predictor lineal, una función de error y una función de enlace. Un predictor lineal (PL) se define como la suma de los efectos de las variables explicatorias:

$$PL = a + bx + cy + \dots$$

en donde, a, b, c, ... son parámetros o constantes estimados a partir de los datos observados y x, y, ... son las variables explicatorias o independientes. Estas últimas pueden ser de tipo continuo o categóricas. La función de error depende de la naturaleza de las variables respuesta. Para datos categóricos, del tipo de presencia/ausencia (+/-), la función de error binomial es la más adecuada y esta ha sido la aplicada en nuestro caso. La función de enlace, el tercer componente de un modelo GLM, es la encargada de trasladar los cambios del PL en la variable respuesta (dependiente). Entre las alternativas posibles (logit, probit, etc), la función más aplicada es la logística, que tiene la particularidad de acotar los valores predictivos entre 0 y 1. Con este tipo de función de enlace, la probabilidad de encontrar un visón (respuesta +) es una función logística cuando el predictor lineal

es un polinomio de primer orden. En su forma más simple la función logística se puede expresar como:

$$p = (e^{PL}) / (1 + e^{PL}) \text{ y } PL = (\ln p)/(1 - p), \text{ en donde}$$

p es la probabilidad de obtener una respuesta positiva y e es la base de los logaritmos naturales (\ln).

Proceso de selección de variables, ajuste y evaluación del modelo. En primer lugar, ajustamos cada variable explicatoria al modelo. Para ello utilizamos los programas de Regresión no Lineal (SYSTAT 7.0) y Logistic Regression (BMDP:LR, Dixon 1984). Cada variable probada y con contribución significativa (al menos del 5%) fue incluida en el modelo. Procedimientos tradicionales como el análisis escalonado paso a paso facilitaron la preselección de variables predictoras y permitió reducir su número inicial a sólo 10. No se probaron las contribuciones de los términos cuadráticos o cúbicos.

Una de las suposiciones básicas de un análisis de regresión es la independencia de las observaciones. La distribución de las observadas para mustélidos difícilmente puede soportar esta independencia, por lo que se han incluido la Latitud y la Longitud en el modelo (y/o sus interacciones) con el fin de analizar la autocorrelación espacial. A continuación, se evaluó si las restantes variables explicatorias (excluidas Latitud y Longitud) permanecían significativas. Este procedimiento es equivalente a extraer el efecto del componente espacial (Bustamante 1997).

Para el modelo seleccionado se calculó el porcentaje de clasificaciones correctas y se analizó si esta clasificación fue significativamente mejor que una al azar (Titus et al 1984). Además, en el ajuste definitivo se llevó a cabo un análisis de los residuos: se midió la influencia de una observación sobre el ajuste del modelo y el estadístico \underline{h} (Hosmer y Lemeshow 1989).

Resultados y discusión

En el caso de los visones (presentes en 23 de las 85 parcelas muestreadas) de todas las 70 categorías medidas una primera selección, en función de su capacidad explicatoria y contribución al modelo de regresión logística, redujo su número a sólo 10.

El bajo valor del estadístico “Log Likelihood” ($=-3,37$) y la bondad de ajuste del modelo ($\chi^2 = 0,72$; g.l.=42; $p=1,00$) demostró una excelente capacidad predictiva del modelo. El elevado valor del coeficiente de McFadden ($Rho=0,93$) así también lo confirma. Esta capacidad predictiva fue superior al 97,2 % de todos los casos para un punto de corte de probabilidad de 0,51 (94,8% de las parcelas con visones y el 98,1% de las parcelas sin visones). Para conseguir un 100% de clasificación correcta debíamos de reducir la probabilidad de corte a 0.21.

El modelo tuvo una sensibilidad de 0,95 y una especificidad de 0,98, reduciendo a probabilidades muy bajas las clasificaciones incorrectas (falsos positivos: 0,05; y falsos negativos: 0,02).

En el caso de los huillines (que sólo estuvieron presentes en 4 de las 85 parcelas muestreadas), un modelo preliminar explicó su presencia en base a sólo tres variables independientes: (a) Ausencia de contaminación; (b) Presencia de camarones de río (*Sammastacus spinifrons*) y (c) Presencia de choritos de agua dulce (*Diplodon chilensis*). Interesantemente, una de las 10 variables que explicó la presencia de visones fue la ausencia de huillines, caracoles de río (*Chilina* spp.) y puyes (*Galaxias* spp.).

Bibliografía

Bustamante J (1997) Predictive model for Lesser Kestrel, *Falco naumanni* distribution, abundance and extinction in Southern Spain. *Biological Conservation* 80:153-160.

Hosmer DW y Lemeshow S (1989) Applied Logistic Regression. J. Wiley y Son Inc. New York.

Mc Cullough P y Nelder JA (1989) Generalized Linear Models. Chapman y Hall. New York.

Titus K, Mosher JA y Williams BK (1984) Chance-correlated classification for use in discriminat analysis: ecological applications. American Midl Naturalist 111:1-7.