

God spiked the integers

MCD

Instituto Tecnológico Autónomo de México

Abril 2021

1. Modelos Lineales Generalizados

La idea general de los Modelos Lineales Generalizados consiste en modelar el valor esperado de Y , a través de una función paramétrica simple de las variables explicativas. Los Modelos Lineales Generalizados se parecen mucho a las primeras computadoras mecánicas; cuyas piezas móviles dentro de ellos, "los parámetros", interactúan para producir predicciones no obvias.

Comúnmente, podemos leer los parámetros directamente para comprender las predicciones; es decir, en los modelos lineales gaussianos, cada parámetro individual puede interpretarse o tiene un significado claro en la escala de predicción. Para los modelos lineales generalizados no podemos leer los parámetros directamente e interpretar de la misma forma.

El más común y usado modelo lineal generalizado es por conteo. Consiste en contabilizar enteros no negativos; por ejemplo: 0, 1, 2 ... etc. El problema radica en su complejidad para modelar; es decir, cuando lo que deseamos predecir es un conteo, la escala de los parámetros nunca es la misma que la escala de la predicción.

Uno de los tipos de modelación por conteo es el modelo de Regresión Binomial.

2. Regresión binomial

En esta sección se presenta uno de los métodos más comunes de conteo relacionados con la clasificación binaria: Vivo - Muerto, Aceptado - Rechazado - Derecha -Izquierda para lo cuál el total de ambas categorías es conocido. La diferencia consistirá en que se podrán agregar variables predictoras.

La distribución binomial esta denotada por:

$$y \sim \text{Binomial}(n, p)$$

donde y es un contador (cero o número entero positivo), p es la probabilidad de éxito del evento o juicio y n el número de eventos o juicios. Además, se cumple que la distribución binomial tiene máxima entropía (lo más probable que suceda) cuando cada prueba debe resultar en uno de los dos eventos y el valor esperado es constante. No hay otro supuesto de probabilidad previo a la observación para una variable de este tipo binomial.

Observaremos un ejemplo de la regresión logística para datos que siguen un comportamiento binomial; pues es el método más común cuando los datos se organizan en casos de un solo ensayo, de modo que la variable de resultado solo puede tomar valores etiquetados en 0 y 1. Asimismo, esta regresión no garantiza que la función de distribución posterior sea una distribución Normal.

2.1. Estudio de los chimpancés

Se trata de un experimento destinado a evaluar las tendencias pro-sociales de los chimpancés. La estructura del experimento imita muchos experimentos comunes realizados en estudiantes humanos por economistas y psicólogos.

Un chimpancé se sienta en un extremo de una mesa larga con dos palancas, una a la izquierda y otra a la derecha. Sobre la mesa hay cuatro platos que pueden contener alimentos deseables; los dos platos en el lado derecho de la mesa están unidos por un mecanismo a la palanca derecha y de la misma manera, los dos platos del lado izquierdo están unidos por un mecanismo a la palanca izquierda, como se muestra en la siguiente Figura 1.

Cuando el animal tira de la palanca izquierda o derecha, los dos platos del mismo lado se deslizan hacia los extremos opuestos de la mesa, esto entrega lo que esté en esos platos a los extremos opuestos.

Ahora bien, en todos los ensayos experimentales, ambos platos del lado del animal que controla las palancas contienen alimentos, pero solo uno de los platos del otro lado de la mesa contiene un alimento. Por lo tanto, mientras que ambas palancas entregan comida al animal que controla las palancas, solo una de las palancas entrega comida al otro lado de la mesa.

Hay dos condiciones experimentales. En la condición de compañero, otro chimpancé está sentado en el extremo opuesto de la mesa, mientras que en la condición de control, el otro lado de la mesa está vacío. Finalmente, dos tratamientos de contrapeso alternan qué lado, izquierdo o derecho, tiene un alimento para el otro lado de la mesa. Esto ayuda a detectar cualquier preferencia individual

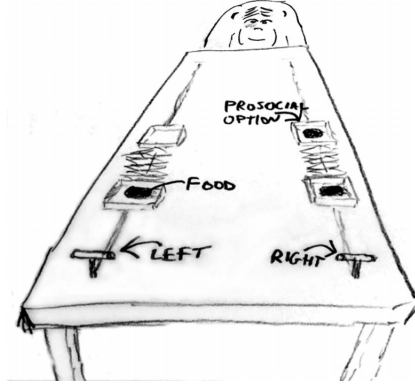


FIGURE 11.2. Chimpanzee prosociality experiment, as seen from the perspective of the focal animal. The left and right levers are indicated in the foreground. Pulling either expands an accordion device in the center, pushing the food trays towards both ends of the table. Both food trays close to the focal animal have food in them. Only one of the food trays on the other side contains food. The partner condition means another animal, as pictured, sits on the other end of the table. Otherwise, the other end was empty.

Figura 1: chimpancés

para los animales que tienen el control.

En términos de modelos lineales, queremos estimar la interacción entre la etiqueta o condición (presencia - ausencia de otro animal) y la opción (qué lado es pro-social).

Queremos inferir lo que sucede en cada combinación de "*prosocleft*" y "*condition*". Hay cuatro combinaciones como se observa:

- (1) *prosoc_left*= 0 and *condition*= 0: Two food items on right and no partner.
- (2) *prosoc_left*= 1 and *condition*= 0: Two food items on left and no partner.
- (3) *prosoc_left*= 0 and *condition*= 1: Two food items on right and partner present.
- (4) *prosoc_left*= 1 and *condition*= 1: Two food items on left and partner present.

Figura 2: combinaciones

Se construyen variables *index* que contengan los valores del 1 al 4. Finalmente el modelo queda como sigue:

$$L_i \sim \text{Binomial}(1, p_i)$$

$$\text{logit}(p_i) = \alpha_{\text{actor}[i]} + \beta_{\text{tratamiento}[i]}$$

donde α_i y β_k se determinarán. Adicionalmente, L indica la etiqueta de la variable "*pulledleft*". Dado que los recuentos de participantes son solo 0 o 1, podría tener el mismo tipo de modelo definido utilizando un Bernoulli. Para el último paso, se convirtió el parámetro a la escala de resultados, lo que significa usar una función logit inversa.

Tendremos que trabajar un poco para interpretarlo. Los primeros 7 parámetros son la intersección única de cada chimpancé.

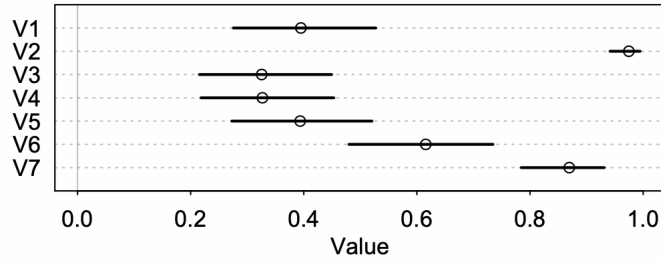


Figura 3: interpretacion

Cada fila es un chimpancé, los números corresponden a los valores de *actor*. Cuatro de los chimpancés, los números 1, 3, 4 y 5, muestran preferencia por la palanca derecha. Dos chimpancés, los números 2 y 7, muestran la preferencia opuesta y se observa que la preferencia del número 2 es muy fuerte. El artículo propone que si inspeccionan los datos, se observará que el actor 2 nunca tiró de la palanca derecha en ningún ensayo o tratamiento. Asimismo, el autor sugiere que hay diferencias sustanciales y que este es exactamente el tipo de efecto que dificulta los experimentos puros en la ciencia del comportamiento. Tener mediciones repetidas, como en este experimento, y medirlas es muy útil.

Ahora consideremos los efectos de los tratamientos, con suerte estimados con mayor precisión porque el modelo podría restar la variación de la mano entre los chimpancés. En la escala logit observamos:

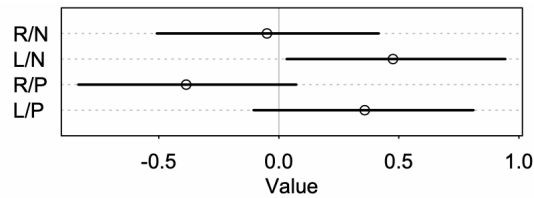


Figura 4: tratamientos

Se agregaron etiquetas de tratamiento en lugar de los nombres de los parámetros. L / N significa "prosocial a la izquierda / sin pareja" R / P significa "prosocial

a la derecha / pareja”. Para comprender estas distribuciones, es útil considerar algunas de nuestras expectativas. Lo que buscamos es evidencia de que los chimpancés eligen más la opción pro-social cuando hay un compañero presente. Esto implica comparar la primera fila con la tercera fila y la segunda fila con la cuarta fila. Probablemente ya pueda ver que no hay mucha evidencia de intención pro-social en estos datos.

En general, no hay ninguna evidencia convincente de elección prosocial en este experimento.

Referencias

McElreath Richard ”Statistical Rethinking - A Bayesian Course with Examples in R and Stan”