MODELO LOGIT O REGRESIÓN LOGÍSTICA

Isadore Nabi

Como se señala en (Aldrich & Nelson, 1984, págs. 30-31), la inferencia estadística comienza por asumir que el modelo que se va a estimar y utilizar para hacer inferencias está correctamente especificado. La presunción, *i.e.*, el supuesto de partida, es que la teoría estadística-matemática correspondiente a tal o cual modelo estadístico es la que justifica el uso del mismo. Sin embargo, a lo planteado por los autores hay que agregar que es aún más importante que las propiedades reales del fenómeno a estudiar (establecidas por el marco científico mediante el cual se estudia) deben corresponderse en una magnitud mínima necesaria y suficiente con las propiedades matemáticas de tal o cual modelo estadístico. Los autores señalan que es bastante fácil demostrar que la especificación incorrecta del modelo tiene implicaciones realmente sustanciales, ya que todas las propiedades estadísticas de las estimaciones pueden destruirse. Para decirlo sin rodeos, la especificación incorrecta del modelo conduce a respuestas incorrectas.

Los autores también elaboran una maravilla gnoseológica en su argumentación, relativa a la justificación del difundido uso del supuesto de linealidad, estableciendo una versión modificada de la navaja de Occam, una que no implica reduccionismo filosófico, como sí lo suele ser la que utilizan, por ejemplo, los bayesianos subjetivos en los modelos parsimoniosos (y fue en ese sentido en el que la criticó también Albert Einstein):

"¿Por qué es tan popular la especificación lineal? Hay dos razones básicas (y relacionadas). En la práctica, los modelos lineales son matemáticamente simples, por lo que los estadísticos han podido aprender mucho sobre ellos, y se han escrito programas de computadora para hacer la estimación. Sobre bases teóricas, la simplicidad conduce a su adopción, justificada por una versión de la navaja de Occam: en ausencia de una guía teórica en sentido contrario, comience asumiendo el caso más simple. Así, la Navaja de Occam, por implicación, diría: Con alguna orientación teórica en sentido contrario, no asuma el caso más simple." (Aldrich & Nelson, 1984, pág. 31).

Como señalan los autores en (Aldrich & Nelson, 1984, pág. 32), el problema (desde la perspectiva matemática) con la especificación del modelo de probabilidad lineal es que $\sum b_k X_{ik}$ se usa para aproximar un número de probabilidad, P_i [$P_i \equiv P\ (Y_i = 1)$], restringido a ser de 0 a 1, mientras que $\sum b_k X_{ik}$ no está tan restringido. Una forma de abordar este problema es transformar P_i para eliminar una o ambas restricciones. Para el caso dicotómico, podemos eliminar el límite superior, $P_i = 1$, observando la relación $\frac{P_i}{1-P_i}$. Esta relación debe ser positiva (ya que $0 < P_i < 1$, que efectivamente es una restricción), pero no hay límite superior explícito en el cociente antes expuesto. El resultado de lo cual puede ser cualquier número real desde el infinito negativo hasta el positivo.

Luego se asume que la variable transformada es una función lineal de X, como se señala en (Aldrich & Nelson, 1984, pág. 32) y se presenta a continuación:

$$\log\left[\frac{P_i}{1-P_i}\right] = \sum b_k X_{ik} \equiv Z_i$$

Finalmente, puesto que el monomio que se desea conocer de la expresión anterior es P_i, se pueden aplicar antilogaritmos y manipulación algebraica para aislar (despejar P_i) para obtener la expresión que se presenta a continuación:

$$P_i = \frac{e^{Z_i}}{1+e^{Z_i}}$$

La expresión anterior es conocida como la *función logística*, la cual es continua y puede tomar valores únicamente entre 0 y 1. Es cercana a 0 cuando Z_i tiende a infinito negativo, incrementa monotónicamente con Z_i y es cercana a 1 cuando Z_i tiende a infinito positivo. Forma así una curva suave en forma de S, la cual es simétrica alrededor del punto Z_i tal como se muestra a continuación.

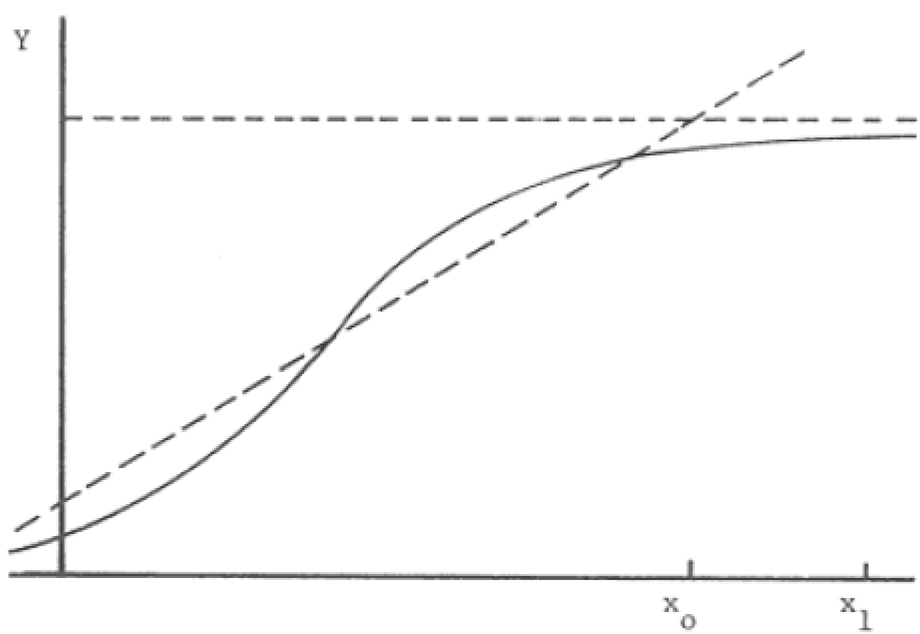


Figure 1
Sigmoid Versus Linear Specifications

Fuente: (Aldrich & Nelson, 1984, pág. 27).

Así, a diferencia del caso de especificación lineal, esto satisface la restricción de encontrarse entre 0 y 1 sin necesidad de también restringir Z_i, es decir, sin restringir $\sum b_k X_{ik}$.

Como señalan los autores, "Estas características de la función descrita en la ecuación 2.4 la convierten en una alternativa atractiva al modelo de probabilidad lineal para variables dependientes dicotómicas. Puede ser bastante razonable, pero ¿por qué este? ¿Por qué no otros? Después de todo, es tan arbitrario como escoger la linealidad. De hecho, hay un número infinito de alternativas a la ecuación 2.4 y, al igual que con la ecuación 2.4, algunas de ellas se han desarrollado como modelos alternativos para la estimación. Siete de estos se muestran en la Figura 3, y los describiremos brevemente para proporcionar una indicación del amplio "menú" de opciones disponibles." (Aldrich & Nelson, 1984, pág. 32).

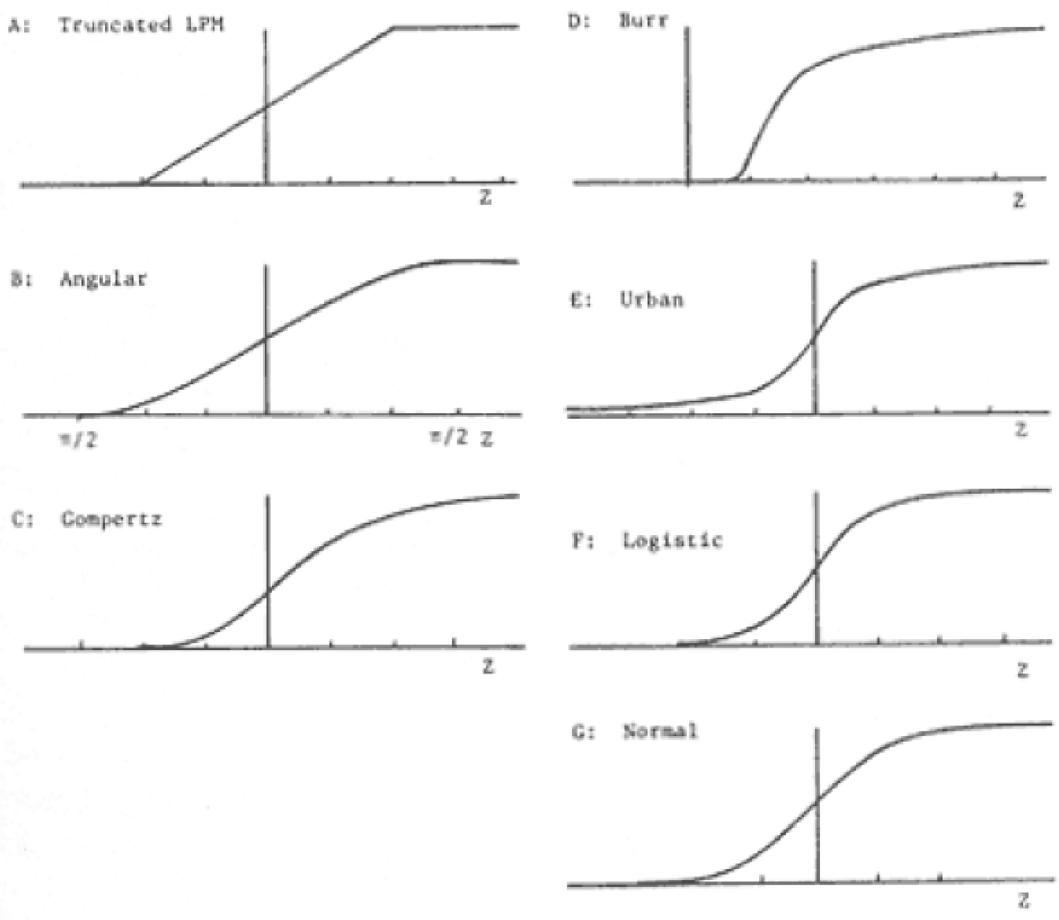


Figure 3
Graphs of Alternative Specifications

Fuente: (Aldrich & Nelson, 1984, pág. 33).

“Las curvas logísticas y normales son tan similares que producen resultados esencialmente idénticos. En la práctica, arrojan probabilidades de elección estimadas que difieren en menos de 0,02 y que pueden distinguirse, en el sentido de significancia estadística, solo con muestras muy grandes. La elección entre ellos, por lo tanto, gira en torno a preocupaciones prácticas como la disponibilidad y flexibilidad de los programas de computadora y las preferencias y experiencias personales. Estos dos han recibido la mayor atención por parte de investigadores (y programadores de computadoras). Es decir, están mejor desarrollados que los otros ejemplos, se usan mucho más ampliamente y están disponibles en programas de computadora con mucha más frecuencia. Su importancia es tal que centraremos la mayor parte de nuestra atención en ellos. Las curvas logísticas y normales son tan similares que producen resultados esencialmente idénticos. En la práctica, arrojan probabilidades de elección estimadas que difieren en menos de 0,02 y que pueden distinguirse, en el sentido de significancia estadística, solo con muestras muy grandes. La elección entre ellos, por lo tanto, gira en torno a preocupaciones prácticas como la disponibilidad y flexibilidad de los programas de computadora y las preferencias y experiencias personales. Estos dos han recibido la mayor atención por parte de investigadores (y programadores de computadoras). Es decir, están mejor desarrollados que los otros ejemplos, se usan mucho más ampliamente y están disponibles en programas de computadora con mucha más frecuencia. Su importancia es tal que centraremos la mayor parte de nuestra atención en ellos.” (Aldrich & Nelson, 1984, pág. 34).

Señala además (Liao, 1994, pág. 11), el *modelo logit* o *regresión logística*[1] asume que el conjunto de datos sigue una distribución binomial. Agrega también que “Para interpretar los resultados de un modelo logit de manera significativa, el modelo en sí debe ser capaz de explicar la variable de respuesta significativamente mejor que el modelo con la intersección solamente. Esto es cierto para todos los modelos lineales generalizados. En un modelo de regresión clásico, se utiliza una prueba F; En un modelo logit (y otros modelos de probabilidad), la prueba más comúnmente utilizada es el estadístico de razón de verosimilitud, que sigue aproximadamente la distribución chi-cuadrado (ver Aldrich y Nelson, 1984; Greene, 1990; McCullagh y Nelder, 1989; entre otros).” (Liao, 1994, págs. 12-13).

Como señalan (McCullagah & Nelder, 1989, pág. 31), la función enlace η empleada por el modelo logit es:

$$\eta = \log\left(\frac{\mu}{1-\mu}\right)$$

[1] Se le conoce como modelo logit porque la regresión logística emplea como función enlace

Como señalan (McCullagah & Nelder, 1989, pág. 31), es importante conocer la forma funcional de la familia de potencias de las funciones enlace, en el contexto de los modelos lineales generalizados, al menos para observaciones cuya media es positiva. Esta familia puede especificarse mediante $\eta = \frac{(\mu^{\lambda}-1)}{\lambda}$ con la característica de que $\lim\limits_{\lambda \to 0} \eta = \log(\mu)$, aunque también puede especificarse mediante el truncamiento $\eta = \{\mu^{\lambda}, cuando\ \lambda \neq 0;\ \log(\mu)\ cuando\ \lambda = 0\}$. La primera forma de especificación tiene la ventaja de experimentar una transición suave (del inglés "smooth"[2]), aunque con ambas hay que realizar acciones complementarias especiales en el caso de $\lambda = 0$.

La función logística posee la siguiente forma general, como se señala en (Wikipedia, 2021):

$$f(x) = \frac{L}{1 - e^{-k(x-x_0)}}$$

Sin embargo, como se señala en (Aldrich & Nelson, 1984, pág. 34), la función logística de uso estadístico estándar toma la siguiente forma[3]: y toma la forma que se presenta a continuación:

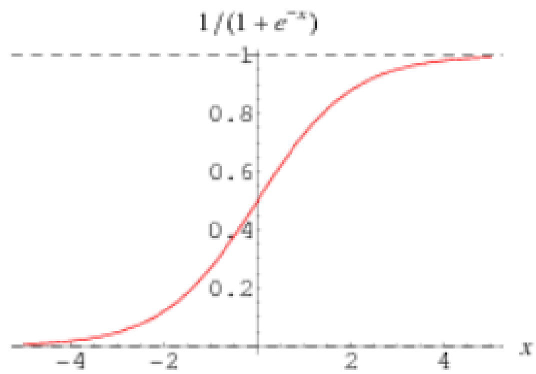
$$f(x) = \frac{e^x}{1 + e^x}$$

Esta distribución adopta empíricamente formas como las que se presentan a continuación:

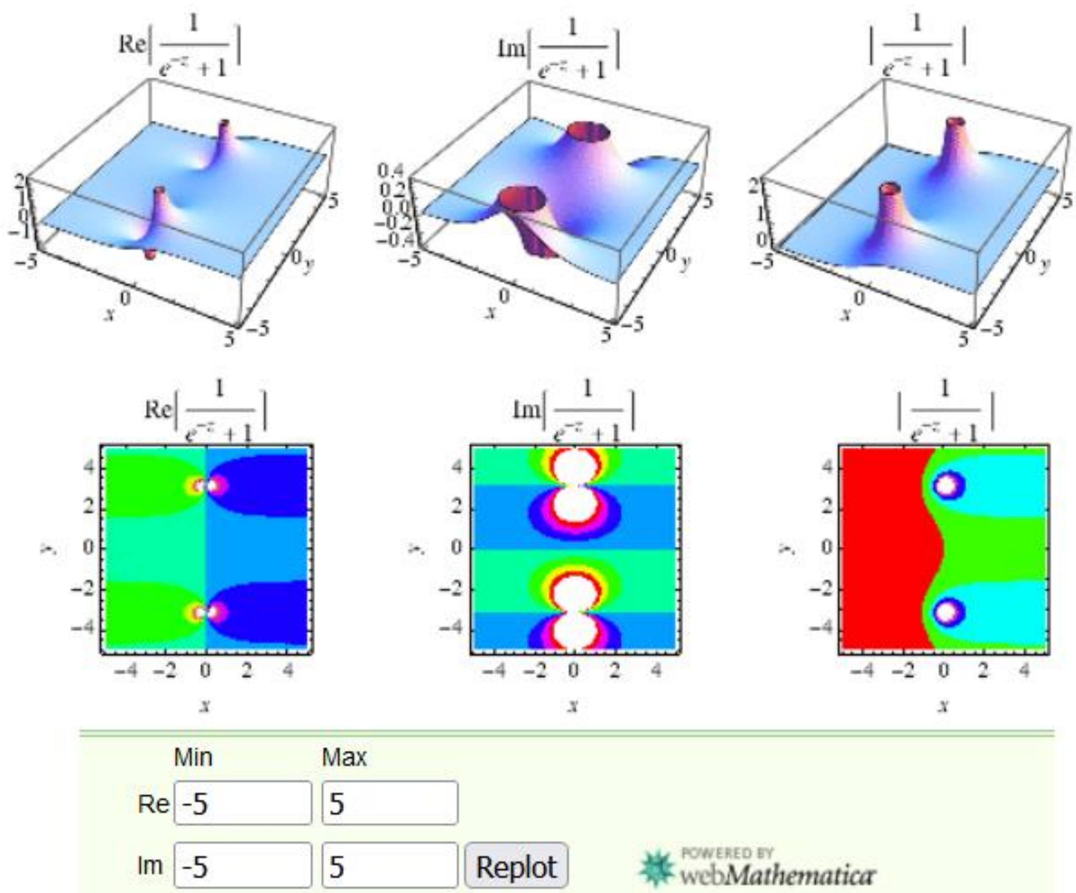
---

[2] Como se señala en (Weisstein, Smooth Function, 2021), una función suave es aquella para cuyas derivadas de orden superior (primera, segunda, tercera, …) derivadas son continuas sobre algún dominio (*i.e.*, existen) en el cual se estudia la función. Se puede decir entonces que una función es suave sobre algún intervalo restringido como (a, b) o $[a, b]$. El número de derivadas que deben existir para que la función sea considerada suave depende del problema que se estudie, sin embargo, puede variar de dos a infinito. Una función para la cual existen derivadas en todos los órdenes y en la que además estas son todas continuas, se conoce como *función-C-infinito*.

[3] Para la transición de una forma (la general) a otra (la estándar) los valores de los parámetros de la forma general deben tomar los siguientes valores: $k = 1$, $x_o = 0$, $L = 1$. Evidentemente, esto tiene implicaciones gnoseológicas importantes, sin embargo, escapa de los objetivos de esta investigación abordarlas, por lo que serán objetivo de un trabajo posterior.



Fuente: (Weisstein, 2021).



Fuente: (Weisstein, 2021).

REFERENCIAS

Aldrich, J. H., & Nelson, F. D. (1984). *Linear Probability, Logit, and Probit Models.* Beverly Hills: Sage University Papers Series. Quantitative Applications in the Social Sciences.

Liao, T. F. (1994). *INTERPRETING PROBABILITY MODELS. Logit, Probit, and Other Generalized Linear Models.* Iowa: Sage University Papers Series. Quantitative Applications in the Social Sciences.

McCullagah, P., & Nelder, J. A. (1989). *Generalized Linear Models* (Segunda ed.). London: Chapman and Hall.

Weisstein, E. W. (21 de Mayo de 2021). *Sigmoid Function*. Obtenido de MathWorld - A Wolfram Web Resource: https://mathworld.wolfram.com/SigmoidFunction.html

Weisstein, E. W. (18 de Mayo de 2021). *Smooth Function*. Obtenido de Wolfram MathWorld - A Wolfram Web Resource: https://mathworld.wolfram.com/SmoothFunction.html

Wikipedia. (7 de Julio de 2021). *Logistic function*. Obtenido de Logistic regression: https://en.wikipedia.org/wiki/Logistic_function