15.1 TIPOS DE VARIABLES

15.1.1 Predictores y variables predichas

Supongamos que queremos predecir el peso de alguien a partir de su altura. En este caso, el peso es la variable predicha y la altura es el predictor.

La diferencia matemática clave entre las variables predictoras y predichas es que la función de verosimilitud expresa la probabilidad de los valores de la variable predicha como una función de los valores de la variable predictora. La función de verosimilitud no describe las probabilidades de los valores de la variable predictora.

La variable predicha depende de la variable predictora, al menos matemáticamente en la función de verosimilitud, **pero no causalmente** en el mundo real, la variable predicha también se puede llamar variable "dependiente". Las variables predictoras a veces se denominan variables "independientes". La diferencia conceptual clave entre las variables independientes y dependientes es que el valor de la variable dependiente depende del valor de la variable independiente.

15.1.2 Tipos de escala: “metric”, ordinal, nominal y “count”

Los elementos se pueden medir en diferentes escalas. Por ejemplo, los participantes en una carrera a pie se pueden medir por el tiempo que tardaron en correr la carrera o por su posición en la carrera ("primero, segundo, tercero, etc.), o por el nombre del equipo al que representan. Estas tres medidas son ejemplos de escalas métricas, ordinales y nominales, respectivamente.

Ejemplos de escalas métricas incluyen tiempo de respuesta (latencia o duración), temperatura, altura y peso. Esos son casos de un tipo específico de escala métrica denominada relación escala, porque tienen un punto cero natural en la escala. El punto cero en la escala corresponde a una ausencia completa del material que se está midiendo. Por ejemplo, cuando la duración es cero, no ha transcurrido el tiempo, y cuando el peso es cero, no hay fuerza hacia abajo. Debido a que estas escalas tienen un punto cero natural, es significativo hablar de proporciones de cantidades que se miden, y por eso se llaman escalas de proporción.

Un caso especial de datos de escala métrica son los “count data” también denominados datos de frecuencia. Por ejemplo, la cantidad de automóviles que pasan por una intersección durante una hora es un conteo. Los datos de conteo solo pueden tener valores que sean enteros no negativos.

Los ejemplos de escalas ordinales incluyen el lugar en una carrera o la calificación del grado de en un acuerdo (5 = totalmente de acuerdo, 4 = levemente de acuerdo, 3 = ni de acuerdo ni en desacuerdo, 2 = levemente en desacuerdo y 1 = totalmente en desacuerdo). No hay información de distancia o métrica en una escala ordinal.

Los ejemplos de escalas nominales, también conocidas como categóricas, incluyen un partido político o la cara de un dado. Para escalas nominales, no hay distancia entre categorías ni orden entre categorías.

15.2 COMBINACIÓN LINEAL DE PREDICTORES

15.2.1 Función lineal de un solo predictor métrico

Supongamos que hemos identificado una variable para predecir ***(y)*** y una variable para que sea el predictor ***(x)***. Supongamos que hemos determinado que ambas variables son métricas. El siguiente problema que debemos abordar es cómo modelar una relación entre ***x y***. La respuesta habitual a esta pregunta es: una relación lineal

Las funciones lineales conservan la proporcionalidad (en relación con una línea de base adecuada). Si duplicas la entrada, se duplica la salida. Si el costo de una barra de chocolate es una función lineal de su peso, entonces cuando el peso se reduce en un 10%, el costo debe reducirse en un 10%.

La forma matemática general para una función lineal de una sola variable es:

$$y = \beta_0 + \beta_1 x$$

El valor del parámetro ***β0*** se llama intersección porque es donde la línea interseca el eje cuando ***x = 0***

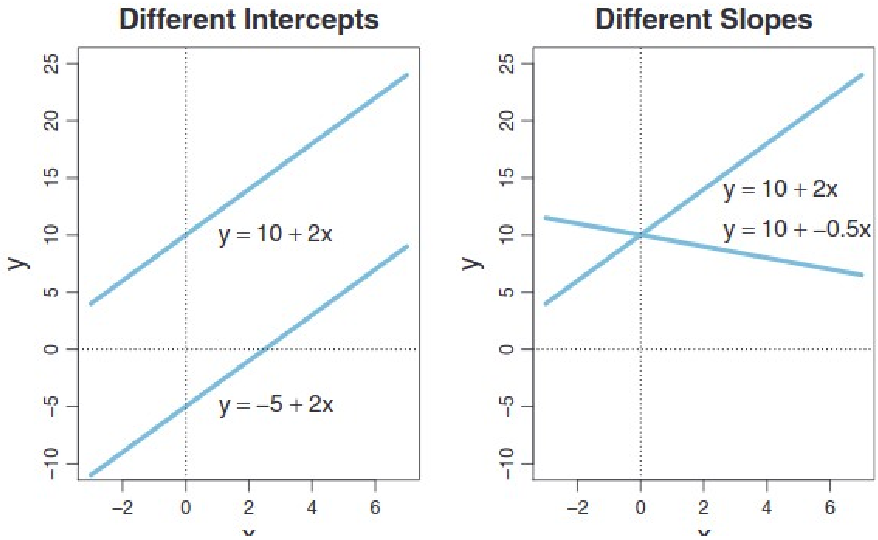
El valor del parámetro ***β1*** se llama pendiente porque indica cuánto aumenta ***y*** *c*uando ***x*** aumenta en 1.

Cuando ***β0*** es diferente a 0, la transformación no conserva la proporcionalidad. Por ejemplo:

y = 10 + 2x

Cuando se duplica de ***x = 1*** a ***x = 2***

y aumenta de ***y = 12*** a ***y = 14***, que no es el doble.



15.2.2 Combinación aditiva de predictores métricos

En general, una combinación lineal de K variables predictoras tiene la forma:

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K \\ &= \beta_0 + \sum_{k=1}^{K} \beta_k x_k \end{aligned}$$

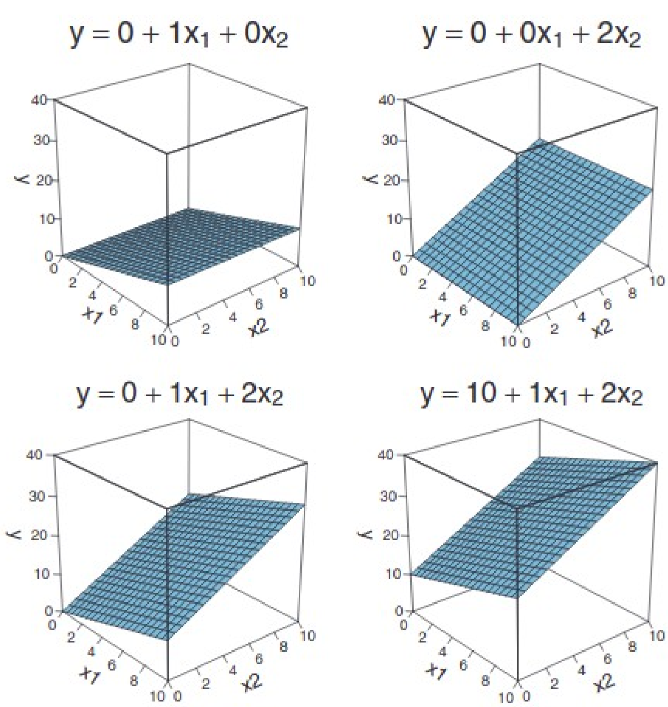
Ejemplos de funciones lineales de dos variables, ***x1*** y ***x2.***

Arriba izquierda: Sólo ***x1*** tiene una relaciòn con ***y***.

Arriba derecha: Sólo ***x2*** tiene influencia.

Inferior izquierda: ***x1*** y ***x2*** tienen un aditivo

Inferior derecha: Se agrega una intersección distinta de cero.



15.2.3. Interacción no aditiva de predictores métricos

La influencia combinada de dos predictores no tiene por qué ser aditiva. Considere, por ejemplo, la autoevaluación de felicidad de una persona, predicha a partir de su salud general y sus ingresos anuales. Es probable que si la salud de una persona es muy mala, la persona no esté feliz, independientemente de sus ingresos. Y si la persona no tiene ingresos, probablemente no sea feliz, independientemente de su salud. Pero si la persona es saludable y rica, entonces la persona tiene una mayor probabilidad de ser feliz.

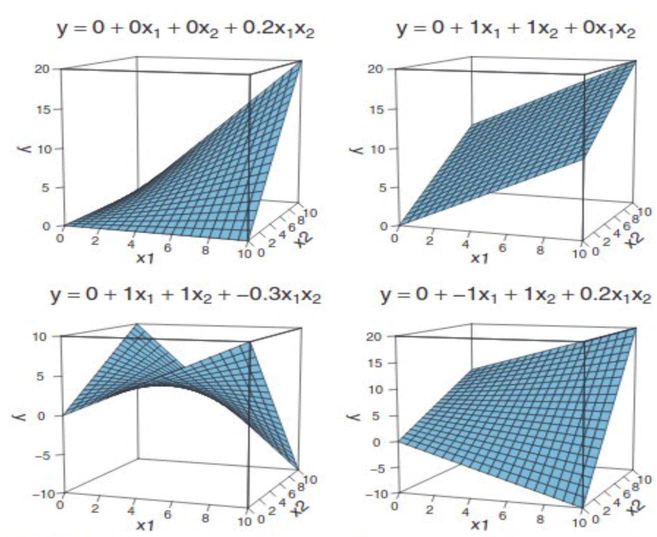
Un gráfico de este tipo de interacción no aditiva entre los predictores aparece en el panel superior izquierdo. El eje vertical, etiquetado como ***y***, es la felicidad. Los lados horizontales, ***x1*** y ***x2***, son salud e ingresos. Observe que ya sea ***x1 = 0*** o **x2 = 0**, entonces ***y = 0***. Pero si tanto ***x1 > 0*** como ***x2> 0***, entonces ***y> 0***. La forma específica de interacción trazada aquí es multiplicativa:

y = 0 + 0x1 + 0x2 + 0.2x1x2.

A modo de comparación, el panel superior derecho de la Figura muestra una combinación no interactiva (es decir, aditiva) de ***x1*** y ***x2***. Observe que el gráfico de la interacción tiene una curvatura en ella, pero el gráfico de la combinación de aditivos es plana.

El panel inferior izquierdo de la Figura muestra una interacción multiplicativa en la que los predictores individuales aumentan el resultado, pero las variables combinadas lo disminuyen. Un ejemplo real de esto ocurre con algunos medicamentos: individualmente, cada uno de los dos medicamentos puede mejorar los síntomas, pero cuando se toman juntos, los dos medicamentos pueden interactuar y causar un deterioro de la salud.

El panel inferior derecho de la Figura muestra una interacción multiplicativa en la que la dirección de la influencia de una variable depende de la magnitud de la otra variable. Nótese que cuando ***x2 = 0,*** un aumento en la variable ***x1*** conduce a una disminución en ***y***. Pero cuando ***x2 = 10***, entonces un aumento en la variable ***x1*** conduce a un aumento en ***y.*** De nuevo, el gráfico de la interacción muestra una curvatura.



15.2.4. Predictores nominales

Las secciones anteriores asumieron que el predictor era métrico. Pero, ¿y si el predictor es nominal, como la afiliación a un partido político o el sexo? Una formulación conveniente hace que cada valor del predictor nominal genere una desviación particular de la diferencia de su nivel de línea base.

Por ejemplo, considere la posibilidad de predecir la altura a partir del sexo (hombre o mujer). Podemos considerar la altura promedio general en ambos sexos como la altura de referencia. Cuando un individuo tiene el valor "hombre", eso agrega una una desviación hacia arriba a la altura prevista. Cuando un individuo tiene el valor "mujer", eso agrega una desviación hacia abajo a la altura pronosticada.

Representaremos el pronosticador nominal por un vector

$$\overrightarrow{x} = \langle x_{[1]}, \ldots, x_{[J]} \rangle$$

donde ***J*** es el número de categorías que tiene el predictor. Se usarà un subíndice entre corchetes para indicar un elemento particular del vector, por analogía con los índices en ***R***. Por lo tanto, el "primer componente del vector se denota ***x[1]*** y el j-ésimo componente se denota ***x[j]*** .Cuando un individuo tiene el nivel ***j*** del predictor nominal, esto se representa estableciendo

$$x_{[j]} = 1 \text{ and } x_{[i \neq j]} = 0.$$

Por ejemplo, sexo es ***x***, siendo el nivel 1 masculino y el nivel 2 femenino. Entonces, el hombre se representa como:

$$\overrightarrow{x} = \langle 1, 0 \rangle$$

y la mujer:

$$\overrightarrow{x} = \langle 0, 1 \rangle$$

Ahora que tenemos una representación formal para la variable predictora nominal, podemos crear una representación formal para el modelo genérico de cómo el predictor influye en la variable predicha. Como se mencionó anteriormente, la idea es que existe un nivel de línea de base de la variable predicha, y cada categoría del predictor indica una desviación por encima o por debajo de ese nivel de línea de base. Denotaremos el valor de la línea de base de la predicción como β0. La definición para el j-ésimo nivel del predictor se denota como β[j].

$$\begin{aligned} y &= \beta_0 + \beta_{[1]} x_{[1]} + \cdots + \beta_{[J]} x_{[J]} \\ &= \beta_0 + \overrightarrow{\beta} \cdot \overrightarrow{x} \end{aligned}$$

15.2.4.2 Combinación aditiva de predictores nominales

Suponga que tenemos dos (o más) predictores nominales de un valor métrico. Por ejemplo, podríamos estar interesados en predecir el ingreso en función de la asociación con el partido político y el sexo. Ahora consideramos la influencia conjunta de esos predictores. Si las dos influencias son aditivas, entonces el modelo se convierte en:

$$
\begin{aligned}
y &= \beta_0 + \overrightarrow{\beta}_1 \cdot \overrightarrow{x}_1 + \overrightarrow{\beta}_2 \cdot \overrightarrow{x}_2 \\
&= \beta_0 + \sum_j \beta_{1[j]} x_{1[j]} + \sum_k \beta_{2[k]} x_{2[k]}
\end{aligned}
$$

con las limitaciones:

$$
\sum_j \beta_{1[j]} = 0 \quad \text{and} \quad \sum_k \beta_{2[k]} = 0
$$

15.2.4.3 Interacción no aditiva de predictores nominales

En muchas aplicaciones, un modelo aditivo no es adecuado para describir la combinación de dos predictores. Por ejemplo, considere predecir los ingresos anuales con relación a la afiliación a un partido político y sexo (en los Estados Unidos contemporáneo). Los hombres, en promedio,tienen ingresos más altos que las mujeres. Los republicanos, en promedio, tienen mayores ingresos que los demócratas. Pero puede ser que las influencias del sexo y los partidos políticos combinen de forma no aditiva. Quizás las personas que son tanto republicanas como hombres tienen un promedio más altos ingresos de lo que se predeciría simplemente sumando los aumentos de ingresos promedio por ser Republicano y por ser hombre.

RESUMEN

Scale type of predictor x					
		Metric		Nominal	
Single group	Two groups	Single predictor	Multiple predictors	Single factor	Multiple factors
β_0	$\beta_{x=1}$ $\beta_{x=2}$	β_0 $+\beta_1 x$	β_0 $+\sum_k \beta_k x_k$ $+\sum_{j,k} \beta_{j\times k} x_j x_k$ $+\left[\text{higher order interactions}\right]$	β_0 $+\overrightarrow{\beta} \cdot \overrightarrow{x}$	β_0 $+\sum_k \overrightarrow{\beta}_k \cdot \overrightarrow{x}_k$ $+\sum_{j,k} \overrightarrow{\beta}_{j\times k} \cdot \overrightarrow{x}_{j\times k}$ $+\left[\text{higher order interactions}\right]$

15.3. VINCULACIÓN DE PREDICTORES COMBINADOS A RUIDO DE DATOS PREDICHOS

Una vez combinadas las variables predictoras, es necesario asignarlas a la variable predicha. Este mapeo matemático se llama función de enlace (inversa).

$$y = f\,(\mathrm{lin}(x))$$

Esta función se llama de enlace inverso, porque tradicionalmente se piensa en la función de enlace como transformar el valor ***y*** en una forma que se pueda vincular al modelo lineal. Es decir, la función de enlace va de ***y*** a los predictores, no de los predictores a ***y***.
La razón de esta terminología es que las flechas en los diagramas jerárquicos de los modelos bayesianos fluirá de la combinación lineal hacia los datos y, por lo tanto, es natural para las funciones mapear hacia los valores predichos.

En segundo lugar, el valor ***y*** que resulta de la función de enlace ***f (lin (x))*** no es un valor de datos per se. En cambio, ***f (lin (x))*** es el valor de un parámetro que expresa una tendencia central de los datos, normalmente su media. Por lo tanto, la función a veces se llama función media (en lugar de función de enlace inverso) y se escribe
μ = f () en lugar de ***y = f ()***.

Hay situaciones en las que es apropiada una función de enlace sin identidad. Por ejemplo, predecir el tiempo de respuesta en función de la cantidad de cafeína consumida. El tiempo de respuesta disminuye a medida que aumenta la dosis de cafeína, aunque la mayor parte de la disminución se produce incluso con una pequeña dosis. Por lo que ***(y)*** a partir de la dosis ***(x)*** tendría una pendiente negativa. Esta pendiente negativa en una función lineal implica que para una dosis muy grandes de cafeína, el tiempo de respuesta se vuelve negativo, lo cual es imposible. Por lo que una función lineal no puede ser utilizada para grandes dosis de cafeína por lo que podríamos querer usar una función enlace que se haga asintótica alrededor de cero como una función exponencial con ***y = exp β 0 + β 1 x***.

15.3.1.1 The logistic function

Una función de enlace frecuente es la logística:

$$y = \mathrm{logistic}(x) = 1 \Big/ \Big(1 + \exp(-x)\Big)$$

Observe el signo negativo delante de la ***x.*** El valor ***y*** de la función logística varía entre 0 y 1. La función logística es casi ***0*** cuando ***x*** es un negativo grande, y es casi ***1*** cuando ***x*** es un positivo grande. En nuestras aplicaciones, ***x*** es una combinación lineal de predictores. Para un solo predictor métrico, la función logística se puede escribir:

$$y = \mathrm{logistic}(x; \beta_0, \beta_1) = 1 \Big/ \Big(1 + \exp\left(-(\beta_0 + \beta_1 x)\right)\Big)$$

La función logística se parametriza de la siguiente forma:

$$y = \mathrm{logistic}\,(x; \gamma, \theta) = 1 \Big/ \Big(1 + \exp\left(-\gamma\,(x - \theta)\right)\Big)$$

$$y = \text{logistic}\left(\gamma\ \left(\sum_k w_k x_k - \theta\right)\right),\ \text{with}\ \left(\sum_k w_k^2\right)^{1/2} = 1.$$

(Este punto no me quedo claro pueden profundizar en la pág 19 del pdf que envío o 437 del texto impreso).

15.3.1.2 La función normal acumulada

Otra función de enlace de uso frecuente es la distribución normal acumulada. Es cualitativamente muy similar a la función logística. Los modeladores utilizarán la logística o la normal acumulada dependiendo de la conveniencia matemática o la facilidad de interpretación.
Por ejemplo, cuando consideramos variables ordinales predichas será natural modelar las respuestas en términos de una variable subyacente continua que tiene una variabilidad normalmente distribuida, lo que lleva a utilizar la normal acumulada como modelo de probabilidades de respuesta.

La función norma acumulada se denota:

$$\Phi(x; \mu, \sigma)$$

donde ***x*** es un número real y ***μ*** y ***σ*** son valores de parámetros, denominados media y desviación estándar de la distribución normal.
El parámetro ***μ*** gobierna el punto en el que la normal acumulada, $\Phi(x)$ es igual a .5

Terminología: La inversa de la normal acumulada se llama función probit. ("Probit" significa "unidad de probabilidad"; Bliss, 1934). La función probit mapea un valor ***p***, para ***0.0 ≤ p ≤ 1.0***, en la línea real infinita, y una gráfica de la función probit se ve muy parecido a la función logit. Puedes ver el enlace expresado de cualquiera de estas formas:

$$y = \Phi(\text{lin}(x))$$
$$\text{probit}(y) = \text{lin}(x)$$

15.3.2. De la tendencia central prevista a los datos ruidosos

En el mundo real, siempre hay una variación en ***y*** que no podemos predecir a partir de ***x***. Por tanto lo que se puede hacer es predecir la probabilidad de que ***y*** tenga un valor particular, dependiendo de ***x***. No predecimos que ***y*** sea exactamente ***f (lin (x))*** porque seguramente estaríamos equivocados. En cambio, predecimos que ***y*** tienda a estar cerca de ***f (lin (x))***.

Para precisar esta noción de tendencia probabilística, necesitamos especificar una probabilidad de distribución para ***y*** que depende de ***f (lin (x))***.
Para mantener la notación manejable, primero definimos ***μ = f (lin (x))***. El valor ***μ*** representa la tendencia central de la predicción de los valores de ***y***, que pueden ser o no la media. Con esta notación, luego denotamos la distribución de probabilidad de ***y*** como alguna función de densidad de probabilidad por especificar, abreviado como "pdf":

$$y \sim \mathrm{pdf}\left(\mu, [\mathrm{scale,shape,etc.}]\right)$$

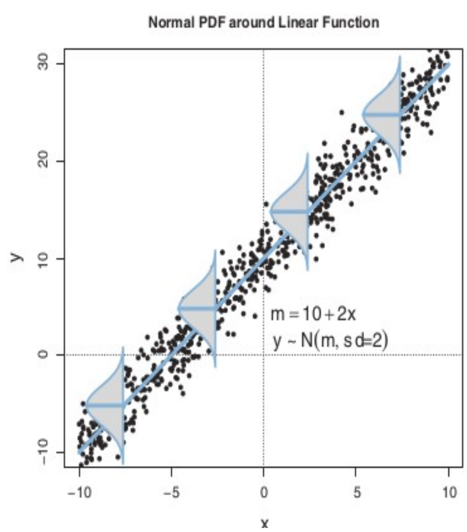
Como lo indican los términos entre corchetes después de ***μ***, el pdf puede tener varios parámetros que controlan la escala de la distribución (la desviación estándar, la forma, etc).
La forma del pdf depende de la escala de medición de la variable predicha.

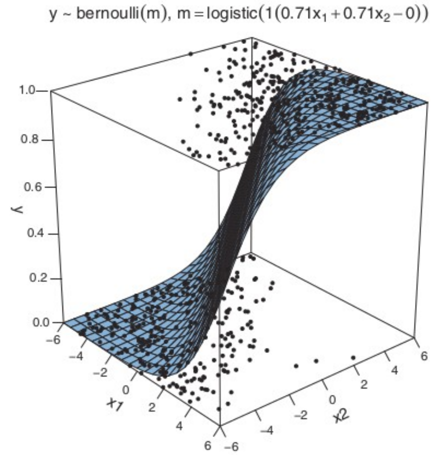
Si la variable predicha es métrica y puede extenderse infinitamente en direcciòn tanto en positiva como en negativa, entonces un PDF típico para describir el ruido en los datos es una distribución normal.

Una distribución normal tiene un parámetro medio ***μ*** y un parámetro de desviación estándar ***σ***,entonces escribiríamos ***y ~ normal (μ, σ)*** con ***μ = f (lin (x))***. En particular, si el enlace función es la identidad, entonces tenemos un caso de regresión lineal convencional.

Si la variable predicha es dicotómica, con ***y*** ∈ ***{0, 1}***, entonces un PDF típico es el Distribución de Bernoulli, ***y ~ Bernoulli (μ)***, donde ***μ = f (lin (x))*** es la probabilidad de que ***y = 1***.

Dado que ***μ*** debe estar entre ***0*** y ***1***, la función de enlace debe convertir ***lin (x)*** a un valor entre ***0*** y ***1***. Una función de enlace típica para este propósito es la función logística.





RESUMEN

Table 15.2 For the generalized linear model: typical noise distributions and inverse link functions for describing various scale types of the predicted variable *y*

Scale type of predicted *y*	Typical noise distribution $y \sim \mathrm{pdf}(\mu, [\text{parameters}])$	Typical inverse link function $\mu = f(\mathrm{lin}(x), [\text{parameters}])$
Metric	$y \sim \mathrm{normal}(\mu, \sigma)$	$\mu = \mathrm{lin}(x)$
Dichotomous	$y \sim \mathrm{bernoulli}(\mu)$	$\mu = \mathrm{logistic}(\mathrm{lin}(x))$
Nominal	$y \sim \mathrm{categorical}(\ldots, \mu_k, \ldots)$	$\mu_k = \dfrac{\exp(\mathrm{lin}_k(x))}{\sum_c \exp(\mathrm{lin}_c(x))}$
Ordinal	$y \sim \mathrm{categorical}(\ldots, \mu_k, \ldots)$	$\mu_k = \Phi((\theta_k - \mathrm{lin}(x))/\sigma) - \Phi((\theta_{k-1} - \mathrm{lin}(x))/\sigma)$
Count	$y \sim \mathrm{poisson}(\mu)$	$\mu = \exp(\mathrm{lin}(x))$

15.4. EXPRESIÓN FORMAL DEL GLM

El modelos lineal generalizado puede ser escrito de la siguiente forma:

$$\mu = f\left(\mathrm{lin}(x), [\text{parameters}]\right)$$
$$y \sim \mathrm{pdf}\left(\mu, [\text{parameters}]\right)$$

Como se ha explicado anteriormente, los predictores ***x*** son combinados en la función lineal ***lin (x).*** La función ***f*** en la primera ecuación se llama función de enlace inverso.
Los datos, ***y***, se distribuyen alrededor de la tendencia central ***μ*** según la función densidad de probabilidad etiquetada como "pdf".