

Regresión Logística



Q Agenda de hoy

- 1 Regresión Logística
- 2 Estimación de la ecuación de regresión
- 3 Prueba de significancia
- 4 Interpretación de la ecuación de regresión
- 5 Transformación Logit

Regresión Logística (1/8)

Hasta ahora hemos utilizado modelos de regresión en los que la variable dependiente es continua: ventas mensuales, número de homicidios, ...



¿Qué pasa en el caso en el que la variable dependiente es discreta?



Ejemplos: Género de una persona, Cliente paga o no paga el próximo mes, Email es spam o no, Banco aprueba tarjeta de crédito o no, Voto de Sara va a votar en segunda vuelta)

Regresión Logística (2/8)

La Regresión Logística nos permite, dado un conjunto particular de valores de las variables independientes, estimar la probabilidad de pertenencia a cada categoría de la variable dependiente

Regresión Lineal

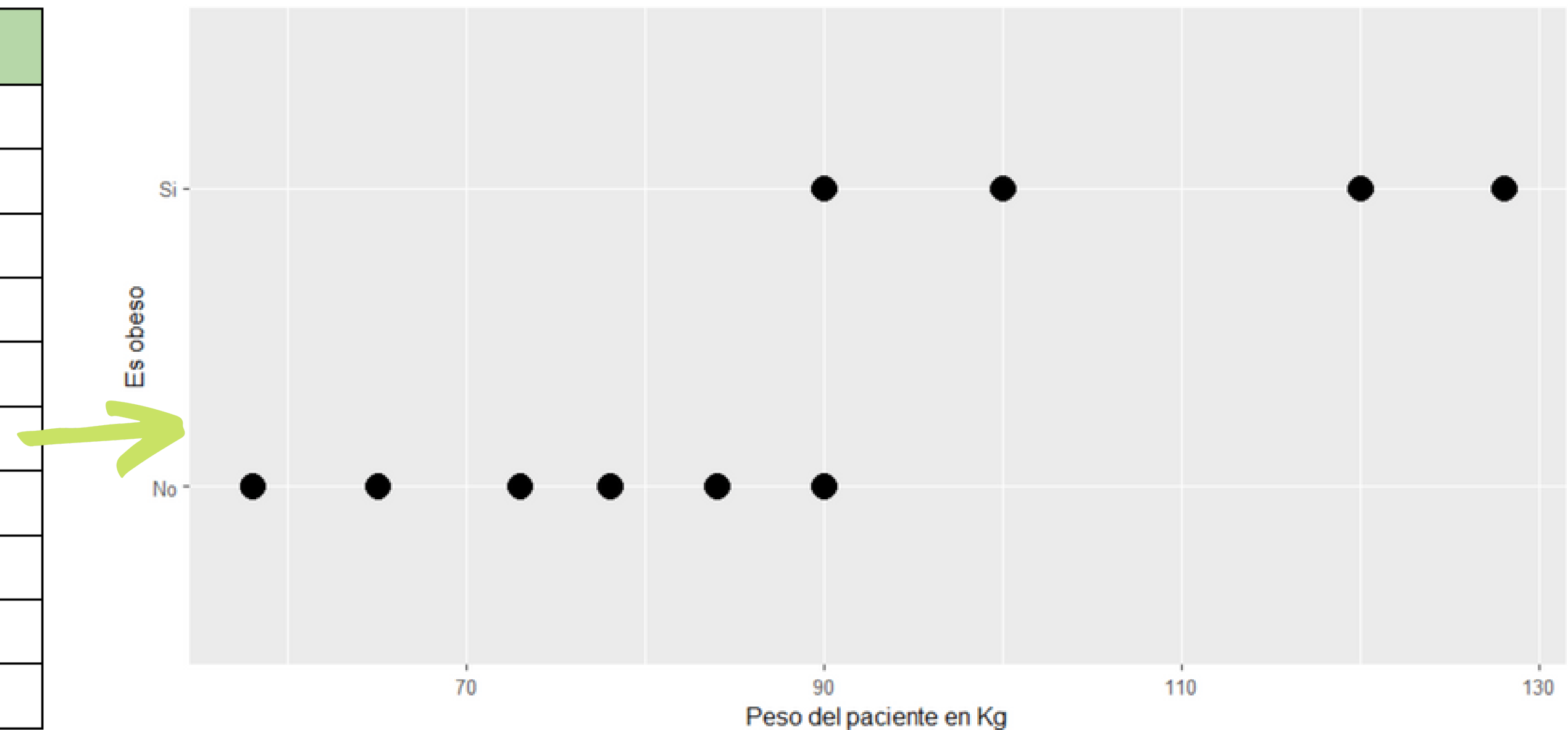
Los valores estimados son el promedio de la variable dependiente dados determinados valores de las variables independientes

Regresión Logística

Los valores estimados son la probabilidad de un nivel particular de la variable dependiente dados determinados valores de las variables independientes

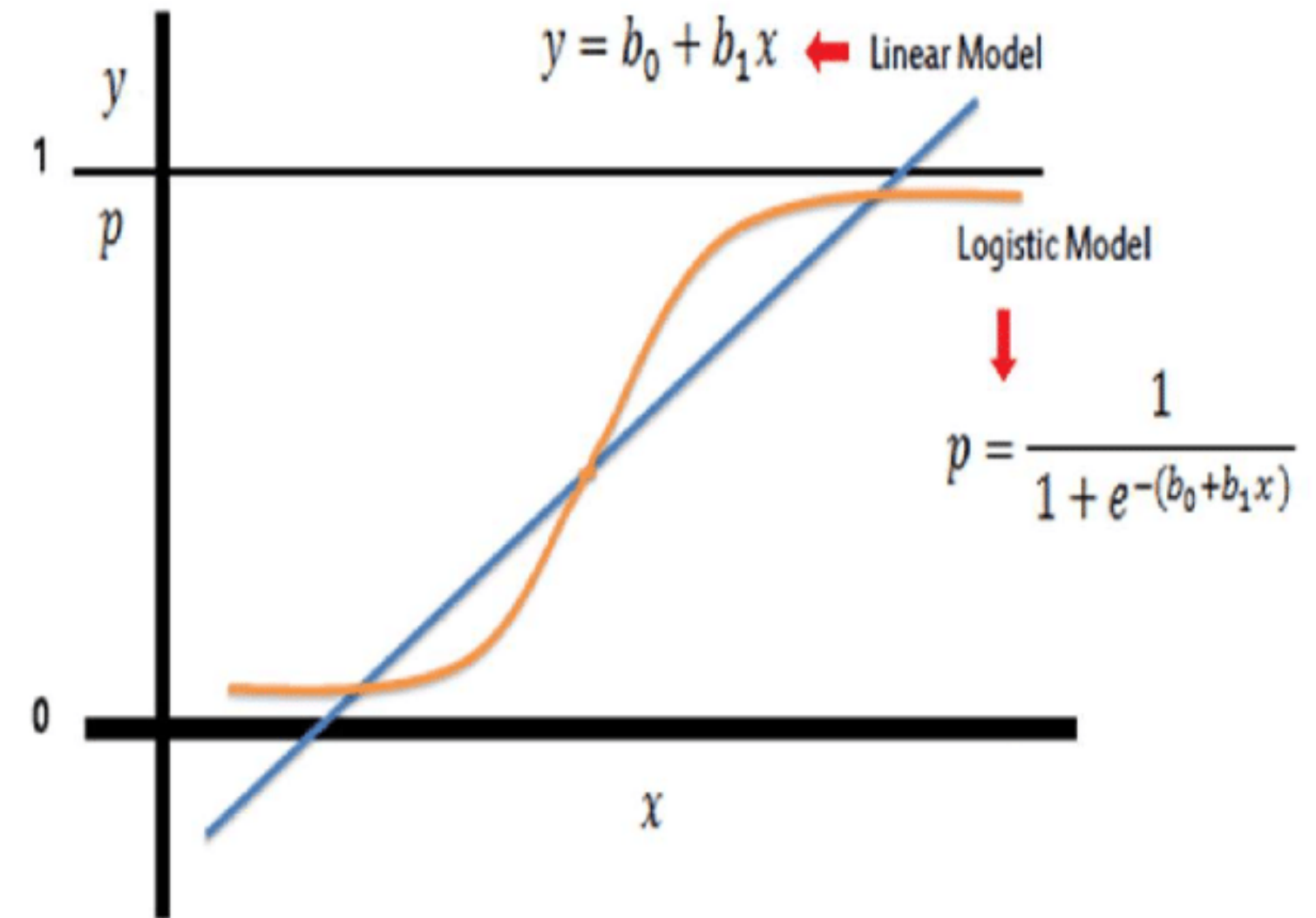
Regresión Logística (3/8)

Paciente i	Peso (Kg)	Es Obeso
1	100	Si
2	73	No
3	90	No
4	90	Si
5	120	Si
6	128	Si
7	65	No
8	58	No
9	78	No
10	84	No



Regresión Logística (4/8)

La Regresión Logística NO ajusta una línea a los datos (como la Regresión Lineal) sino una función con forma de **S** conocida como la **Función Logística**



Regresión Logística (5/8)

Ecuación de Regresión Logística




$$E(y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}} = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}} \quad (1)$$

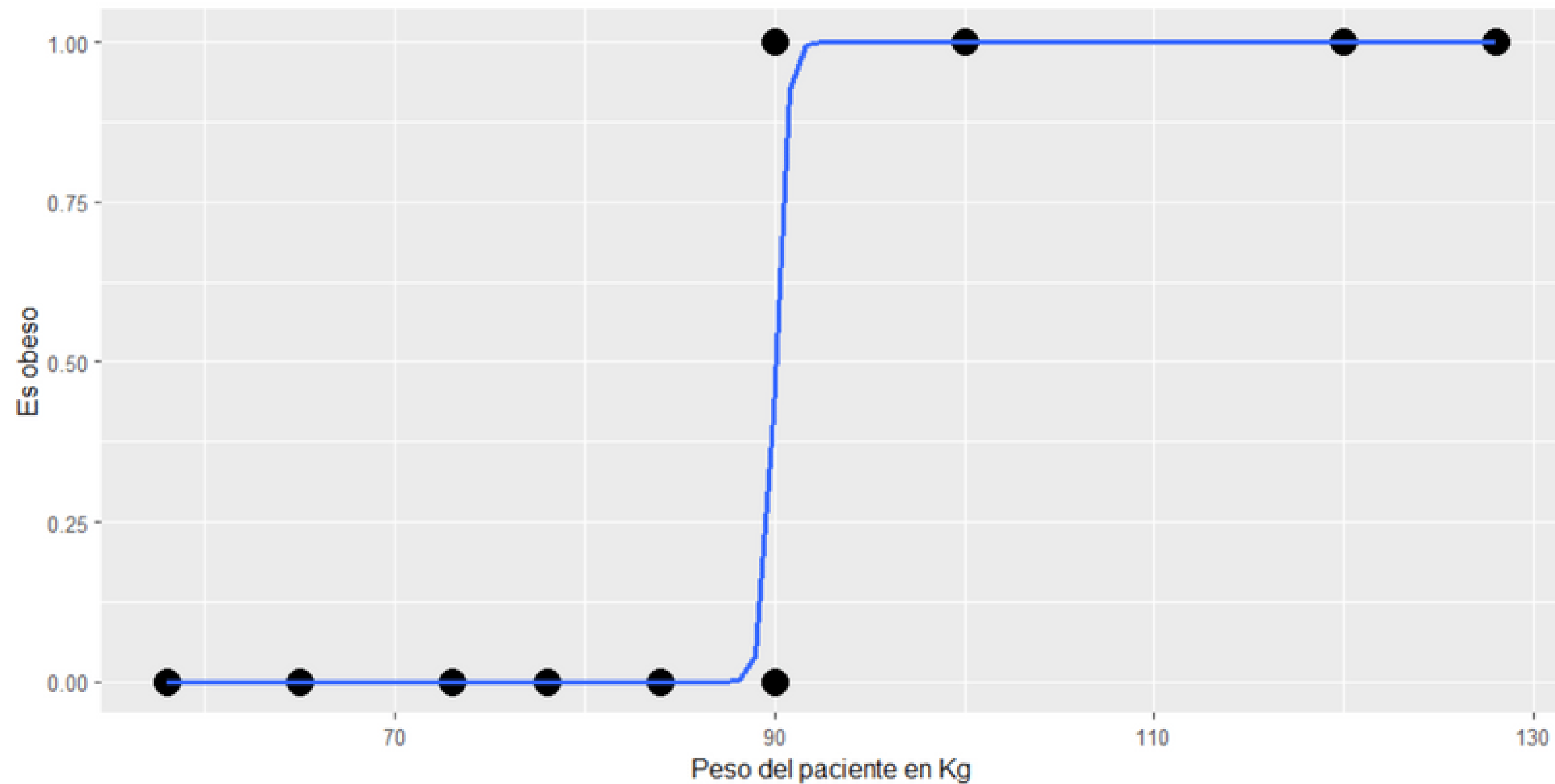
Regresión Logística (6/8)

En la práctica, los valores de los parámetros del modelo no se conocen y es necesario estimarlos usando **datos muestrales**

Ecuación de Regresión Logística Estimada

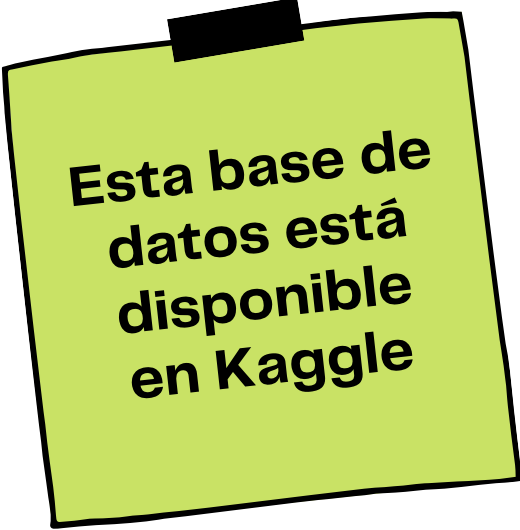

$$\hat{y} = \text{estimacion de } P(y = 1|x_1, x_2, \dots, x_p) = \frac{e^{b_0+b_1x_1+\dots+b_px_p}}{1 + e^{b_0+b_1x_1+\dots+b_px_p}} \quad (2)$$

Regresión Logística (7/8)



Regresión Logística (8/8)

Utilicemos una base de datos que contiene información de empleados de una compañía para predecir cuál es la probabilidad de que estos empleados renuncien (**IBM HR Analytics Employee Attrition & Performance**)



Esta base de
datos está
disponible
en Kaggle