

# Modelos de Clasificación

Herramientas Econométricas

Ricardo Pasquini

# Modelos de clasificación

- Hasta aquí hemos trabajado con variables continuas como variables a explicar/proyectar, pero en una gran cantidad de aplicaciones las predicciones son discretas.
- Estos modelos también son conocidos como **modelos de clasificación**.

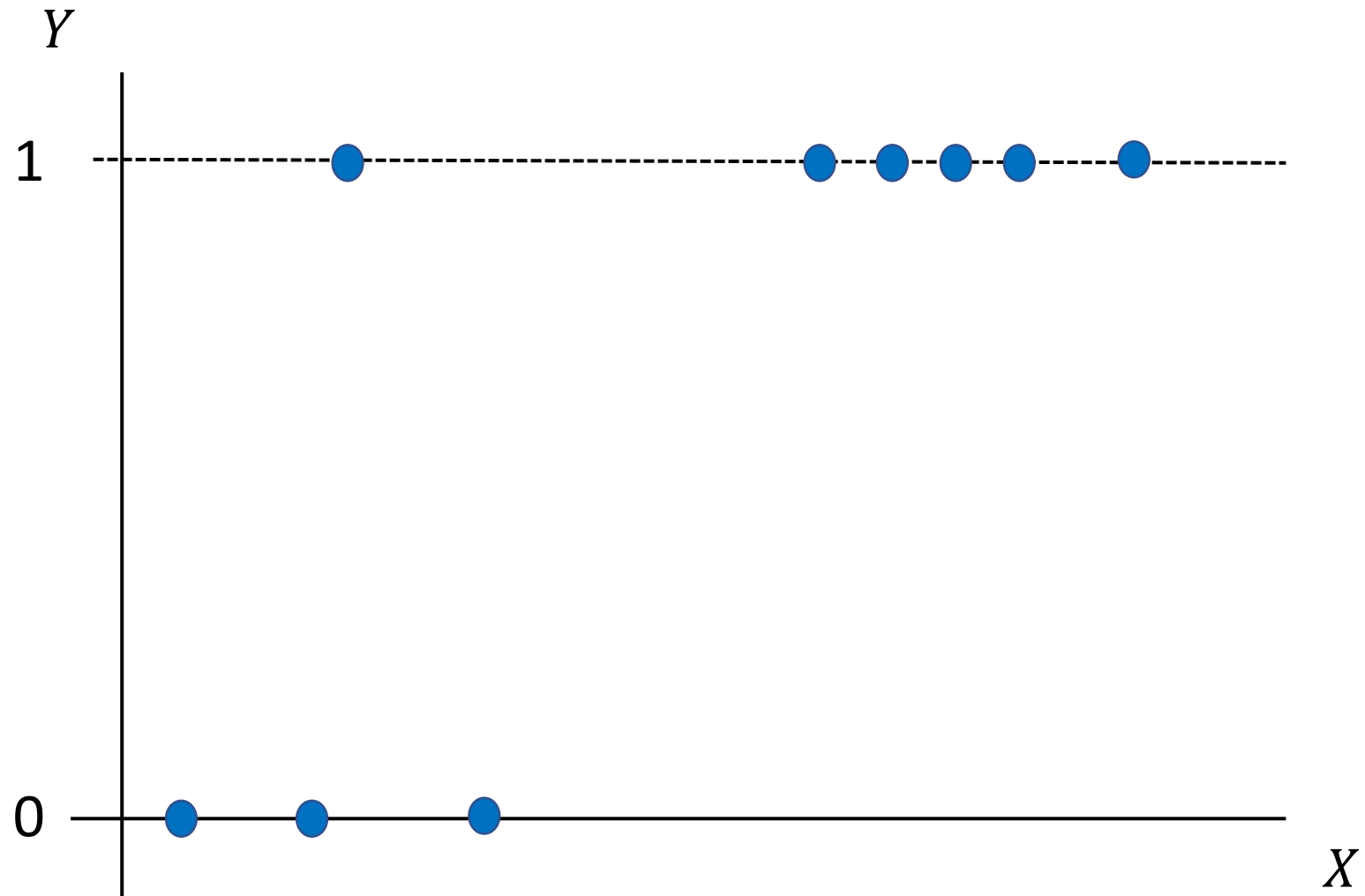
# Aplicación: Probabilidad de Default

- Entidad crediticia otorga préstamos personales.
- Variable objetivo: 1 si impago, 0 de otro modo

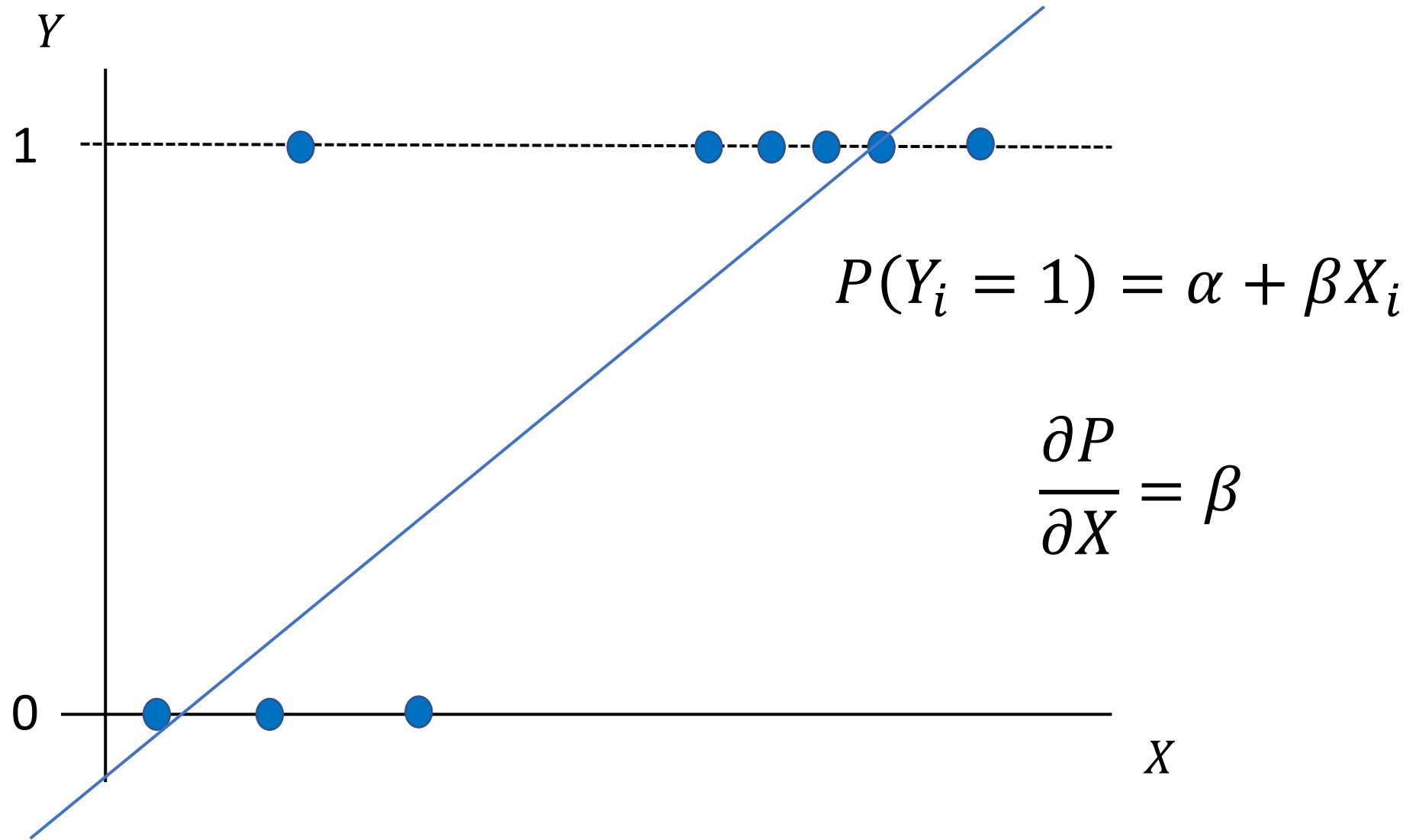
$$\begin{array}{ccccccc} \boxed{P(Y_i = 1)} & = & \alpha & + & \boxed{\text{Monto}} & + & \boxed{\text{Género}} & + \dots + \varepsilon_i \\ & & & & X_i & & Z_i & \\ & & & & \underbrace{\hspace{15em}} & & & \\ & & & & & & \text{Variables} & \\ & & & & & & \text{explicativas} & \end{array}$$

Probabilidad de default

# Modelo lineal para clasificación

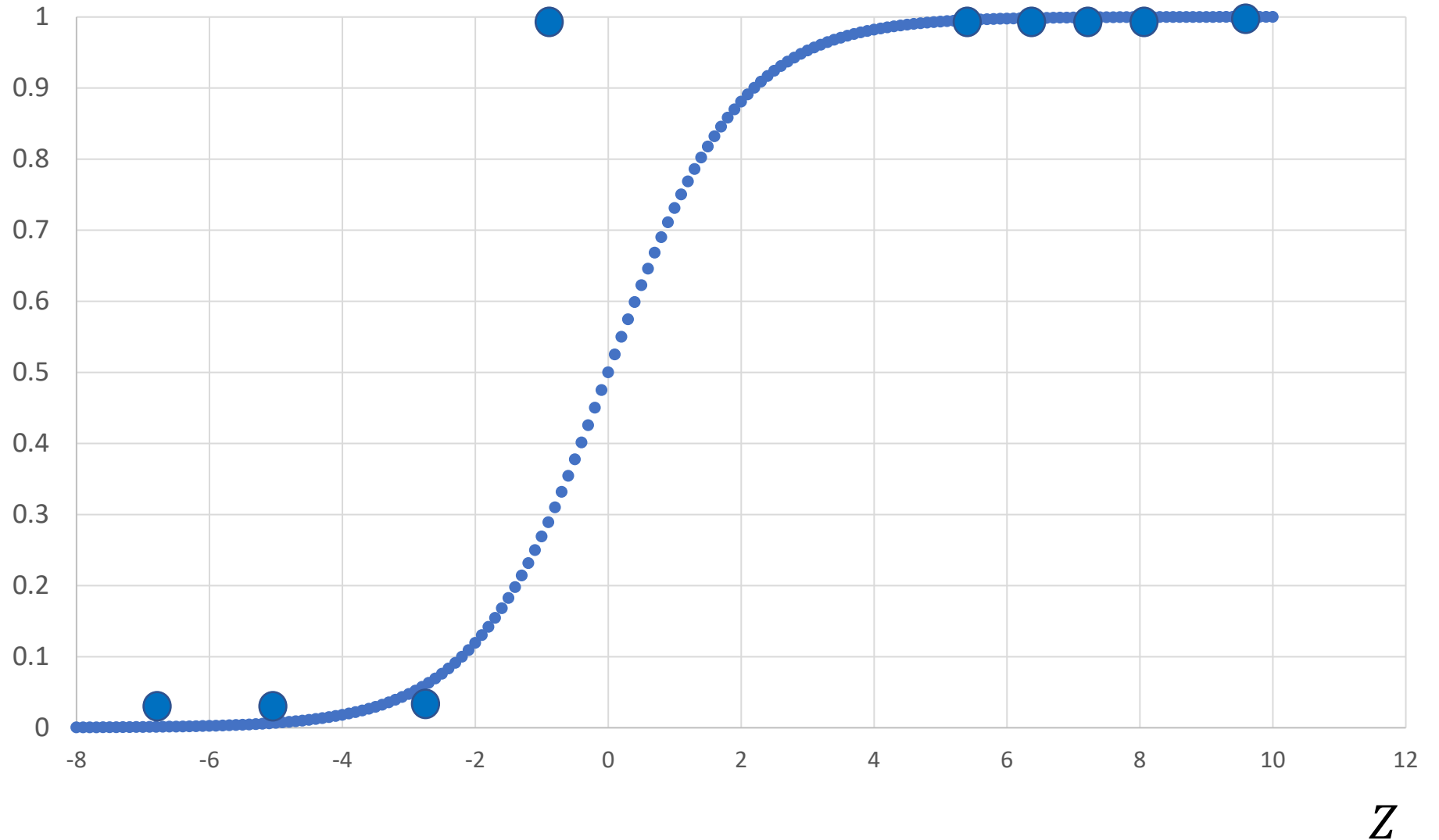


# Modelo lineal para clasificación



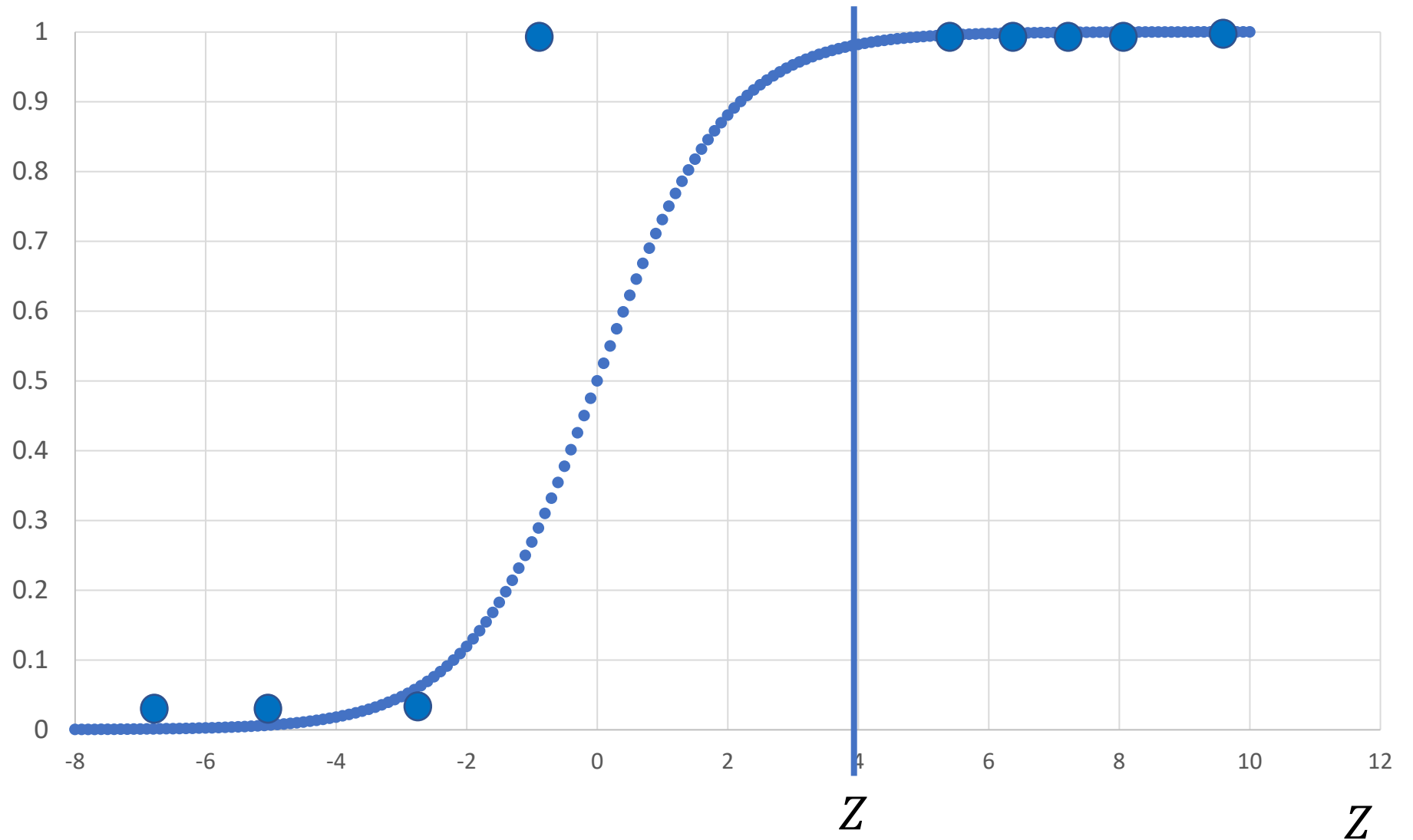
# Modelos de clasificación: Logit

$$\left\{ \begin{array}{l} P(Z) = \frac{1}{1 + e^{-Z}} \\ Z = \alpha + \beta X_i \end{array} \right.$$



# Efecto marginal

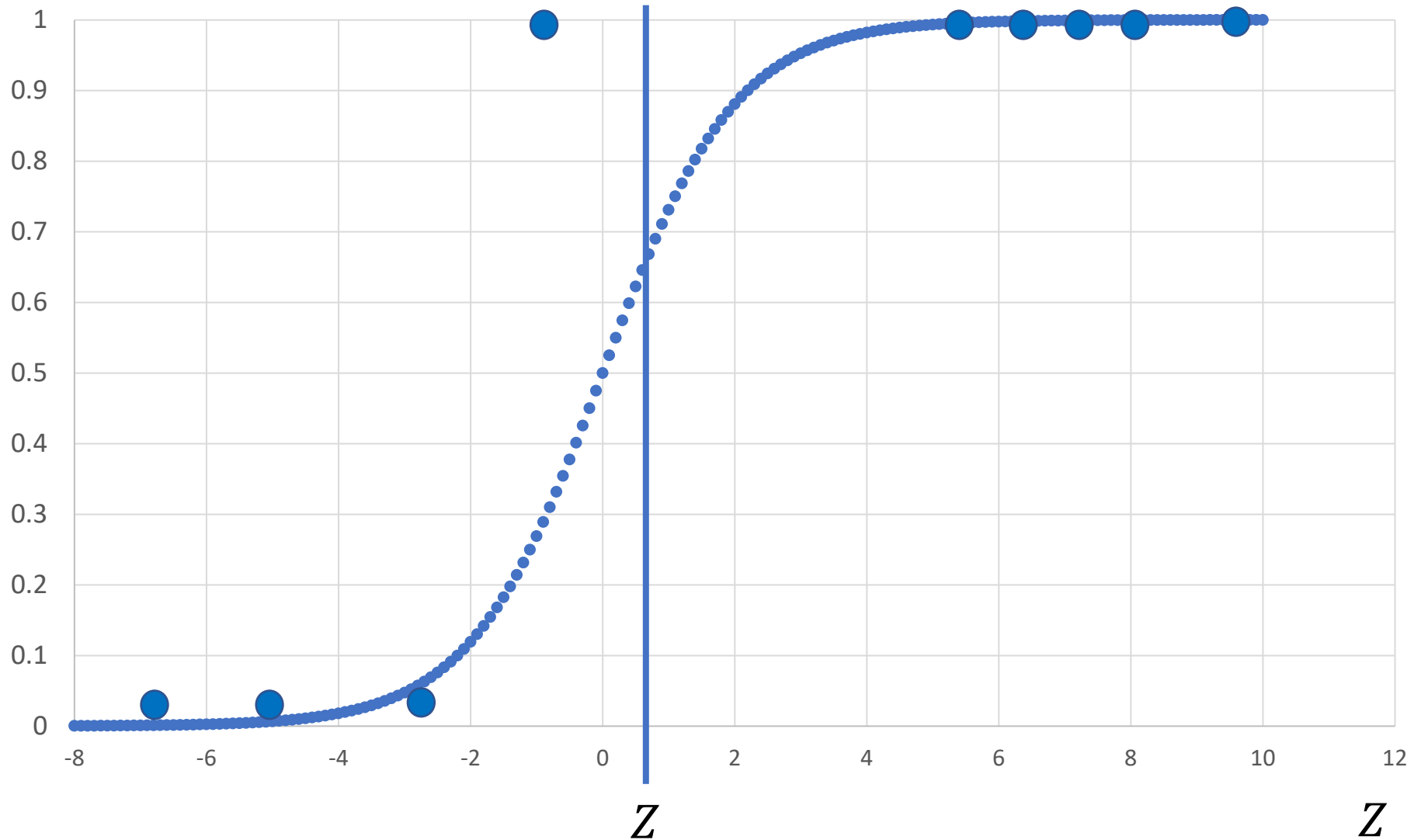
$$P(Z) = \frac{1}{1 + e^{-Z}}$$
$$Z = \alpha + \beta X_i$$



# Efecto marginal

$$P(Z) = \frac{1}{1 + e^{-Z}}$$
$$Z = \alpha + \beta X_i$$

$$\frac{\partial p}{\partial X} = \frac{e^{-Z}}{(1 + e^{-Z})^2} \beta$$

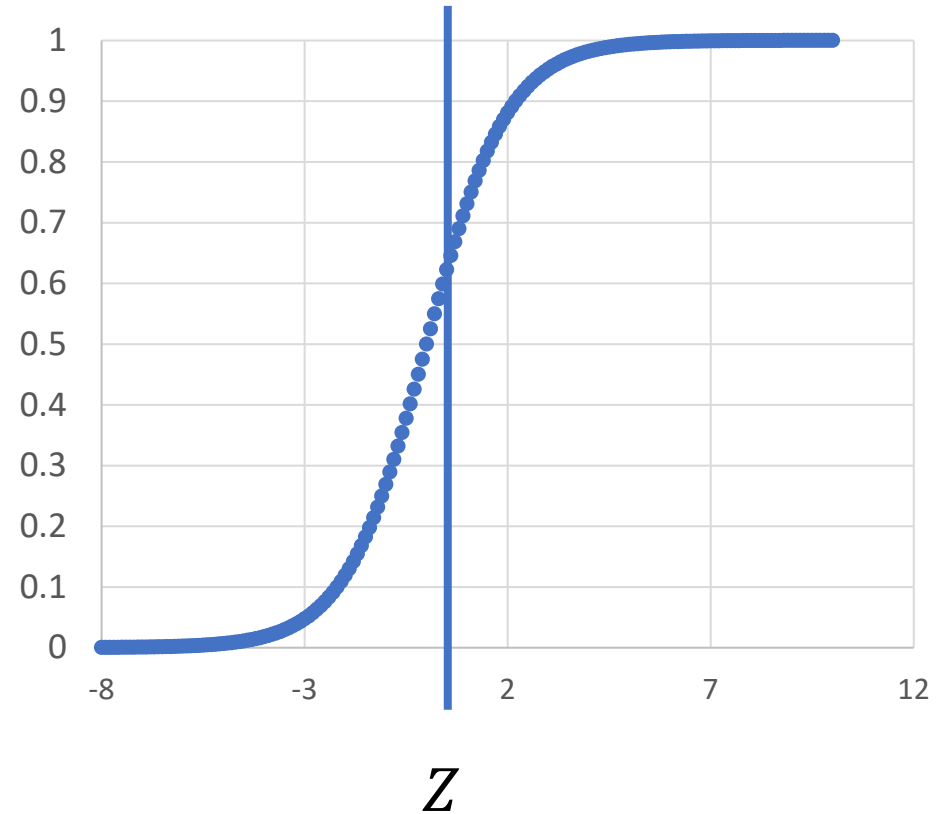




# Efecto marginal

$$\left\{ \begin{array}{l} Z = \alpha + \beta X_i \\ P(Z) = \frac{1}{1 + e^{-Z}} \end{array} \right.$$

$$\frac{\partial p}{\partial X} = \frac{e^{-Z}}{(1 + e^{-Z})^2} \beta$$



- Un valor de interés podría ser conocer el efecto en el valor promedio de Z en los datos de la muestra

# Resumen

- Modelos de clasificación binaria implican una transformación no-lineal que va a garantizar que las predicciones sean apropiadas.
- Recordar que el efecto marginal de un factor explicativo será específico a un nivel deseado.

# Interpretando coeficientes en regresión múltiple

- Al incorporar variables al modelo, los coeficientes asociados capturan el efecto adicional de las mismas, teniendo en cuenta el resto de las variables ya incorporadas.
- Esto es importante porque los factores explicativos pueden estar relacionados entre ellos.

# Modelo: Probabilidad de Default de Crédito

$$P(Y_i = 1) = \frac{1}{1 + e^{-Z}}$$

$Z = \beta_1 + \beta_2 \text{Monto}_i + \beta_3 \text{Sexo}_i + \beta_4 \text{Universidad}_i + \varepsilon_i$

The diagram illustrates the components of a logistic regression model for credit default probability. A vertical line connects the probability term  $P(Y_i = 1)$  to a box labeled "Probabilidad de default". A horizontal line connects the linear combination  $Z$  to a box labeled "Monto en miles de \$". Below  $Z$ , three terms are grouped by brackets and connected to boxes:  $\beta_2 \text{Monto}_i$  is connected to "Monto en miles de \$",  $\beta_3 \text{Sexo}_i$  is connected to "Sexo Femenino", and  $\beta_4 \text{Universidad}_i$  is connected to "Universidad Completa".

Probabilidad de default

Monto en miles de \$

Sexo Femenino

Universidad Completa

# Estimación de Logit por Máxima Verosimilitud

- Eligiendo los parámetros, maximizar la probabilidad de ocurrencia de la muestra de datos en cuestión.

$$\max_{\beta} \prod_{i=1}^n \text{Pr}(Y_i, \beta)$$

$$\max_{\beta} \sum_{i=1}^n \ln(\text{Pr}(Y_i, \beta))$$