

1. Modelos con variable respuesta discreta

En este tema estudiaremos los modelos de regresión en los que la variable respuesta no es continua, sino que es o bien cualitativa o cuantitativa discreta. Como consecuencia, no tendrá sentido suponer una **distribución normal** para la variable respuesta.

1.1. Regresión logística

Consideremos una variable de interés Y dicotómica, y queremos estudiar su relación con otras variables explicativas X mediante un análisis de regresión.

- En este tema estudiaremos los modelos de regresión en los que la variable respuesta no es continua.
- Primero nos centraremos en el caso en que la variable respuesta es binaria y toma sólo dos valores $Y = 0$ e $Y = 1$.

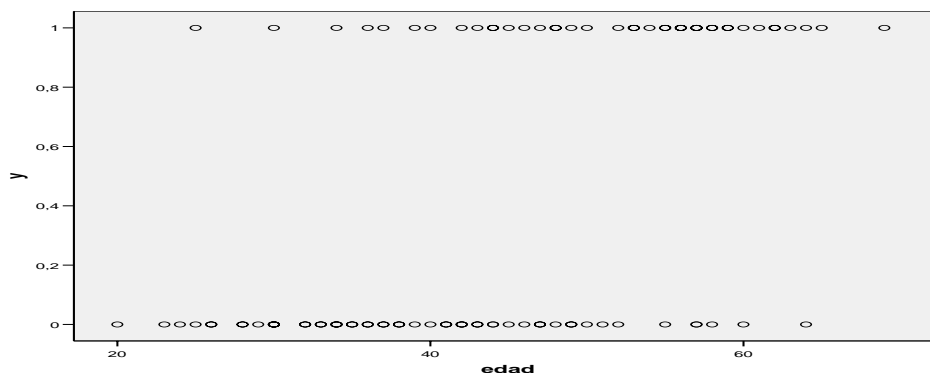


Figura 1: Ejemplo

En la Figura 1 aparece el ejemplo descrito en clase. Vemos que el valor medio de Y dado $X = x$,

$$\pi(x) = \mathbb{E}(Y|X = x),$$

es (en este caso) un número entre 0 y 1. Por tanto el modelo lineal

$$\pi(x) = \beta_0 + \beta_1 x,$$

es claramente inadecuado. Al contrario de lo que se suponía en la regresión lineal, los errores en el modelo

$$y = \pi(x) + \varepsilon \tag{1}$$

no siguen, dado x , una distribución normal con varianza constante (independiente de x). De hecho, para x fijo, en (1) se puede ver que la varianza de ε vale $\pi(x)(1 - \pi(x))$. Si agrupamos los datos del ejemplo anterior obtenemos la siguiente tabla:

Grupo de Edad	n	Presencia	Ausencia	Media (Proporción)
20-29	10	9	1	0.1
30-34	15	13	2	0.13