Lecture Note 7: Binary dep. vars.

$D_i = \begin{cases} 1 \\ 0 \end{cases}$

$$D_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \varepsilon_i$$
$$= X_i' \beta + \varepsilon_i$$

Interpretation: $\hat{D}_i = X_i'\beta$ expected prob. $Pr[D_i = 1 | X_i]$

Linear prob model

How to interpret β's?

$\rightarrow \beta_k = \dfrac{\partial Pr[D_i = 1 | X_i]}{\partial X_{ki}}$ if X_k is continuous

$\rightarrow \beta_k$ = diff. in conditional prob if X_k is binary

Problems: ① $\hat{Pr}[D_i = 1]$ may lie outside $[0,1]$

② Heterostredasticity

Probit and logit models

- Start with estimating p for Bernoulli RV:

$$L = p^{S}(1-p)^{N-S} = \prod_{i=1}^{N} p^{D_i}(1-p)^{1-D_i}$$

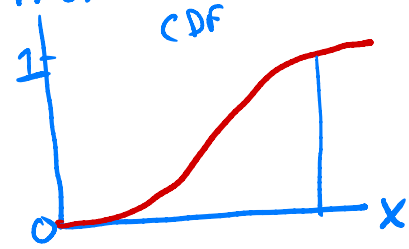
$$G(X_i'\beta) = G(\beta_0 + \beta_1 X_1 \cdots)$$

- So...

$$L = \prod_{i=1}^{N} G(X_i'\beta)^{D_i}\left[1 - G(X_i'\beta)\right]^{1-D_i}$$

- Can use a CDF as $G(\cdot)$:
  1. std. normal CDF → probit
  2. logistic CDF → logit

$Pr[D_i = 1 \mid X]$



vs.

Probit

- $\Pr[D_i = 1 \mid X_i] = \Phi[X_i'\beta] = \int_{-\infty}^{X_i'\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x'\beta)} \, dx$

- In R: feglm( ) with family = 'probit'

Logit

- $\Pr[D_i = 1 \mid X_i] = \Lambda[X_i'\beta] = \frac{e^{X_i'\beta}}{1 + e^{X_i'\beta}} = \frac{1}{1 + e^{-X_i'\beta}}$

$\exp\{X_i'\beta\}$

- In R: feglm( ) with family = 'logit'

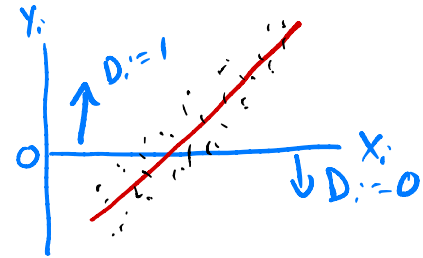
Latent variables

Suppose Y_i is a latent (unobserved) continuous variable:

$$Y_i = X_i'\beta + \varepsilon_i$$

$\varepsilon_i \sim N(0,1)$

$\varepsilon_i \sim$ logistic



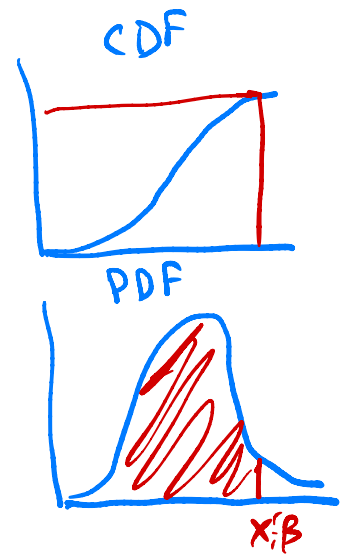
But we observe:

$$D_i = \begin{cases} 1 & \text{if } Y_i > 0 \\ 0 & \text{if } Y_i \le 0 \end{cases}$$

What is $\Pr[D_i = 1 \mid X_i]$?

$$\begin{aligned}
\Pr[D_i = 1 \mid X_i] &= \Pr[Y_i > 0 \mid X_i] \\
&= \Pr[X_i'\beta + \varepsilon_i > 0 \mid X_i] \\
&= \Pr[\varepsilon_i > -X_i'\beta \mid X_i] \\
&= 1 - F[-X_i'\beta] \\
&= F[X_i'\beta]
\end{aligned}$$

G



Marginal effect: $\dfrac{\partial \Pr[D_i=1 \mid X_i]}{\partial X_{ki}}$ ← OLS (LPM) directly gives us this → β_k = derivative

Logit/probit more complex:

- $\dfrac{\partial \Pr[D_i=1 \mid X_i]}{\partial X_{ki}} = \dfrac{\partial F[X_i'\beta]}{\partial X_{ki}} = f(X_i'\beta)\beta_k$

- Two common approaches to summarize marginal effects:
  1. marginal effect at $\bar{X}$
  2. average of the marginal effects across i
- mfx package → logitmfx() and probitmfx()

Odds ratio

Odds of an event: $\frac{p}{1-p}$

In logit model: $p = \frac{e^{X_i'\beta}}{1+e^{X_i'\beta}}$

So...

$$\ln(\text{odds}_i) = \ln\left(\frac{\frac{e^{X_i'\beta}}{1+e^{X_i'\beta}}}{1-\frac{e^{X_i'\beta}}{1+e^{X_i'\beta}}}\right) = \ln\left(e^{X_i'\beta}\right) = X_i'\beta = \beta_0 + \beta_1 X_{i1} \cdots$$

$$\text{odds} = e^{\ln(\text{odds}_i)} = e^{X_i'\beta}$$

And...

$$\text{odds ratio} = \frac{\frac{\Pr[D_i=1 \mid X_{i1}=1, X_2, X_3]}{1-\Pr[D_i=1 \mid X_{i1}=1, \cdots]}}{\frac{\Pr[D_i=1 \mid X_{i1}=0, \cdots]}{1-\Pr[D_i=1 \mid X_{i1}=0 \cdots]}} = e^{\beta_1}$$