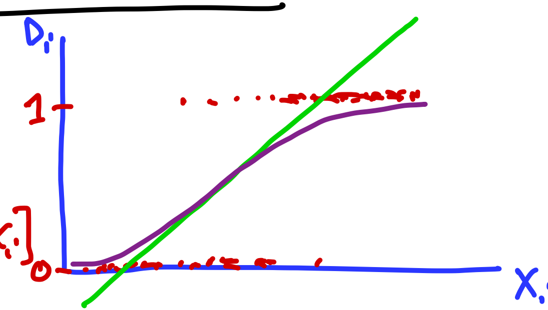
Lecture Note 7: Binary Dependent Variables

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

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



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

How to interpret β's?

→ $\beta_k = \dfrac{\partial Pr[D_i = 1 | X_{ki}]}{\partial X_{ki}}$ if X_{ki} continuous

→ β_k = diff in conditional probabilities if X_{ki} binary

Problems: ① $\hat{Pr}[D_i = 1 | X_i]$ may lie outside [0,1]
② Heteroskedasticity

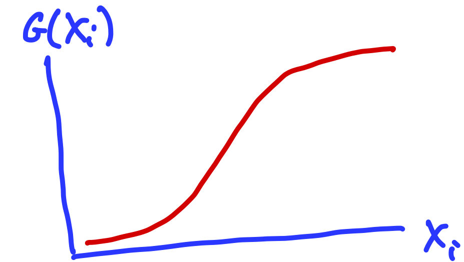
Linear probability model: OLS for binary dep. vars.

Probit and logit models

Start with estimating p for Bernoulli RV:

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

$$p_i = G(X_i\beta) = G(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_K X_{Ki})$$



Choices of $G(\cdot)$:

1. std. normal CDF $\rightarrow$ probit
2. logistic CDF $\rightarrow$ logit

So...

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

Probit:

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

In R, feglm( ) with family='probit'

Logit:

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

In R, feglm( ) with family='logit'

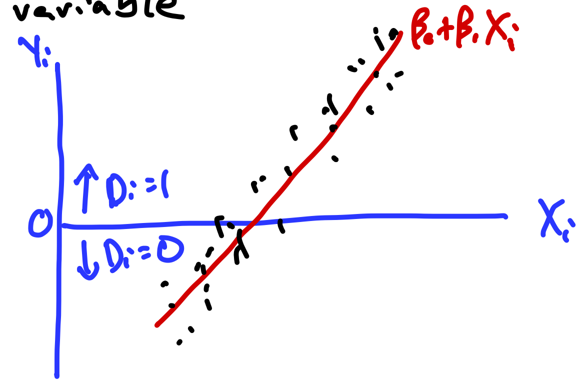
Latent variables

Suppose Y_i is a latent continuous variable

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

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

$\varepsilon_i \sim$ logistic



But we observe:

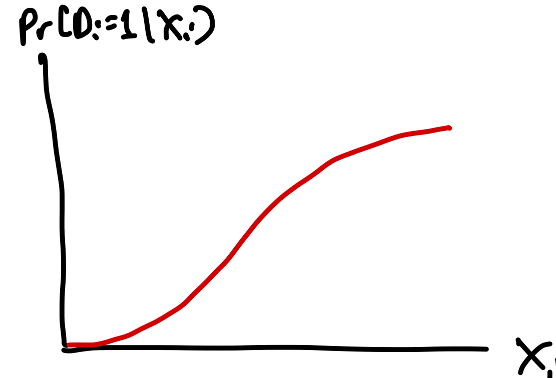
$$D_i = \begin{cases} 1 & \text{if } Y_i \geq 0 \\ 0 & \text{if } Y_i < 0 \end{cases}$$

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

$$\Pr[D_i = 1 \mid X_i] = \Pr[Y_i > 0 \mid X_i]$$
$$= G[X_i'\beta]$$

Φ $\quad$ Λ

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



Logits/probits:

$$\frac{\partial \Pr[D_i = 1 \mid X_i]}{\partial X_{ki}} = \frac{\partial G(X_i'\beta)}{\partial X_{ki}} = g(X_i'\beta)\beta$$

pdf

Two approaches:

1. marginal effects at $\bar{X}$
2. average of marginal effects across i

marginaleffects::avg_slope( )

percentage points

Odds ratio (logit only)

Odds of an event $= \dfrac{p}{1-p}$

In the logit model: $p = \dfrac{e^{x_i'\beta}}{1+e^{x_i'\beta}}$

So... $\text{odds} = \dfrac{p}{1-p} = \dfrac{\frac{e^{x_i'\beta}}{\cancel{1+e^{x_i'\beta}}}}{\cancel{\frac{1+e^{x_i'\beta}}{1+e^{x_i'\beta}} - \frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}}} = e^{x_i'\beta}$

And...

$$\text{odds ratio} = \frac{\text{odds when } X_1 = 1, \text{ controlling for others}}{\text{odds when } X_1 = 0, \text{ controlling " "}} = e^{\beta_1}$$

$\dfrac{\text{men}}{\text{women}} = 1.25 \Rightarrow 25\% \uparrow$ odds for men