Probit

$Y_i = 1$ if yes

$Y_i = 0$ if no.

For example : random utility U_i of individual i

$$U_i = \beta_0^* + w_i \beta^* + \mu_i$$

where β_0^*, β^* are intercept and slope, w is the characteristic

$\mu_i \sim N(0, \sigma_\mu^2)$

Let S denote a threshold of satisfaction.

$$Y_i = 1 \Leftrightarrow U_i = \beta_0^* + w_i \beta^* + \mu_i \geq S + \varepsilon_i$$

$$Y_i = 0 \Leftrightarrow U_i < S + \varepsilon_i$$

model:

let $\mu_i^* = \varepsilon_i - \mu_i$ and $\mu_i^* \sim N(0, \sigma^2)$

$$\frac{\mu_i^*}{\sigma} \leq \frac{(\beta_0^* - S)}{\sigma} + \frac{w_i \beta^*}{\sigma}$$

$$e_i \leq X_i \beta$$

where

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

$X_i = [1 \quad w_i]$ is a row vector

with one or more characteristics for a total of $K-1$

β is a column vector of intercept and slopes for all characteristics

dimensions:

$$\underset{1\times K}{X_i}, \quad \underset{K\times 1}{\beta}$$

Then:

$$Y_i = 1 \quad \text{if} \quad e_i \leq X_i\beta$$

$$Y_i = 0 \quad \text{if} \quad e_i > X_i\beta$$

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

$$P(Y_i = 1 \mid X_i) = \Phi(X_i\beta) = \int_{-\infty}^{X_i\beta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

$$= \int_{-\infty}^{X_i\beta} \phi(t)\, dt = F(X_i\beta)$$

$$P(Y_i = 0 \mid X_i) = 1 - \Phi(X_i \beta)$$

$$l(y_i; \beta) = \prod_{i: Y_i = 1} \Phi(x_i' \beta) \prod_{i: Y_i = 0} \left[1 - \Phi(x_i \beta) \right]$$

let's assume N_1 persons say 1, N_2 say 0; total $= N$

$$\log l(y_i \cdot) = \sum_{i: Y_i = 1} \ln \Phi(x_i' \beta) + \sum_{i: Y_i = 0} \ln \left[1 - \Phi(x_i' \beta) \right]$$

$$= \sum_{i=1}^{N} Y_i \ln \Phi(x_i \beta) + (1 - Y_i) \ln \left(1 - \Phi(x_i' \beta) \right)$$

MLE estimator:

$$\frac{\partial \log l(.)}{\partial \beta} = \sum_{i: Y_i = 1} \frac{\partial \Phi(x_i' \beta) / \partial \beta}{\Phi(x_i' \beta)} - \sum_{i: Y_i = 0} \frac{\partial \Phi(x_i' \beta) / \partial \beta}{\left[1 - \Phi(x_i' \beta) \right]}$$

note that: $\dfrac{\partial \Phi(x_i' \beta)}{\partial \beta} = \dfrac{\partial \int_{-\infty}^{x_i' \beta} \phi(t)\,dt}{\partial x_i \beta} \cdot \dfrac{\partial x_i \beta}{\partial \beta} = \phi(x_i' \beta) x_i'$

where

$$\phi(x_i\beta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_i\beta^2}{2}\right)$$

also:

- $$\frac{\partial \phi(x_i\beta)}{\partial \beta'} = \frac{\partial \phi(x_i\beta)}{\partial x_i\beta} \cdot \frac{\partial x_i\beta}{\partial \beta'} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_i\beta^2}{2}\right) \cdot -\frac{2x_i'\beta}{2} x_i$$

$$= -\phi(x_i\beta)\, x_i'\beta\, x_i$$

$$\frac{\partial \log l(\cdot)}{\partial \beta} = \sum_{i:y_i=1} \frac{\phi(x_i\beta)}{\Phi(x_i\beta)} x_i' - \sum_{i:y_i=0} \frac{\phi(x_i\beta)}{1-\Phi(x_i\beta)} x_i'$$

At the maximum $\frac{\partial L}{\partial \beta} = 0$

The function is concave and has a max if $\frac{\partial^2 \log l}{\partial \beta \partial \beta'}$ is negative definite.

$$\frac{\partial^2 \log l(\cdot)}{\partial \beta \partial \beta'} = \sum_{i:y_i=1} \frac{-x_i\beta \cdot \phi(x_i\beta) \cdot \Phi(x_i\beta) - [\phi(x_i\beta)]^2}{[\Phi(x_i\beta)]^2} x_i' x_i$$

$$- \sum_{i:y_i=0} \frac{-x_i\beta \cdot \phi(x_i\beta)[1-\Phi(x_i\beta)] + [\phi(x_i\beta)]^2}{[1-\Phi(x_i\beta)]^2} x_i' x_i$$

the solution to the maximization is found by iterations.

Consider 2 parameter values β^1 and β^0. The Taylor expansion about β^0 is:

$$L(\beta^1) = L(\beta^0) + \left[\frac{\partial L(\beta)}{\partial \beta}\right]'_{\beta^0} (\beta^1 - \beta_0)$$

where $L(\,)$ is the log-likelihood function. Let us denote the $\left[\frac{\partial L(\beta)}{\partial \beta}\right]_{\beta^0}$ gradient evaluated at β^0 : $D(\beta^0)$

$$L(\beta^1) = L(\beta^0) + D'(\beta_0)\,(\beta^1 - \beta^0)$$

and assume: $$\beta^1 - \beta^0 = \lambda\, M(\beta^0)\, D(\beta^0)$$

$$L(\beta^1) = L(\beta^0) + \lambda\, D'(\beta^0)\,[M(\beta^0)]^{-1}\, D(\beta^0) \geq L(\beta^0)$$

where $M(\beta^0)$ is a negative definite matrix, such as the HESSIAN $\frac{\partial^2 L}{\partial \beta \partial \beta'}$ evaluated at β_0.

pre multiply $M(\cdot)$ by $-$ to make it positive definite and obtain:

$$\beta^1 = \beta^0 - \lambda \left[M(\beta^0)\right]^{-1} D(\beta_0)$$

let's see how it works for a deterministic model:

$$Y = X' A X + B X + C$$

$$\frac{\partial Y}{\partial x} = 2AX + B'$$

$$\frac{\partial Y}{\partial x} = 0 \Rightarrow X = -\frac{1}{2} A^{-1} B'$$

$$D(X^0) = \left[\frac{\partial Y}{\partial x}\right]_{x_0}$$

$$M(X^0) = 2A$$

$$X^1 = X^0 - \lambda \left[M(X^0)\right]^{-1} D(X^0)$$

let $\lambda = 1$

$$x^1 = x^0 - \lambda \frac{1}{2} A^{-1} \left[2AX^0 + B'\right]$$

$$x^1 = x^0 - x^0 - \frac{1}{2} A^{-1} B$$

$$x^1 = -\frac{1}{2} A^{-1} B$$

max is attained
in one step
of length 1 ($\lambda = 1$)

Variance of $\hat{\beta}_{MLE}$:

- is equal to $E\left[\frac{\partial^2 \log L(\,)}{\partial\beta\,\partial\beta'}\right]^{-1} = \left[I(\beta)\right]^{-1}$

that is the inverse of the information matrix.

Technically it is the inverse of the Hessian evaluated at $\beta = \hat{\beta}_{MLE}$ with a $(-)$ sign. that is the estimated $\hat{V}($

$$\hat{V}(\hat{\beta}_{MLE}) = -M(\beta)^{-1}_{\beta=\hat{\beta}_{MLE}}$$

the estimated variances of $\tilde{\beta}_1, \tilde{\beta}_2 \ldots$ are on the main diagonal.

we know that

$$-E\,\frac{\partial^2 \log L(\,)}{\partial\theta\,\partial\theta'} = E\left[\sum_{i=1}^{N} \frac{\partial \log L_t}{\partial\theta}\left(\frac{\partial \log L_t}{\partial\theta}\right)'\right]$$

therefore, an alternative variance estimator is:
the outer product of scores inverse:

$$\sum_{i=1}^{N}\left[\left(\frac{\partial L_t}{\partial\beta}\right)\left(\frac{\partial L_t}{\partial\beta}\right)'\right]^{-1}_{\beta=\hat{\beta}_{MLE}}$$

advantage: is ALWAYS positive definite.

Inference based on the probit model.

1. Predicted probability.

Given the estimated $\hat{\beta}$ and a vector of characteristics x_i, we can predict the probability that individual i chooses the public transportation:

- $\hat{\Phi}(x_i\hat{\beta}) = \hat{\Phi}$
- $$\text{Asy Var}(\hat{\Phi}) = \frac{\partial \hat{\Phi}'}{\partial \hat{\beta}} \left[\text{Asy Var}[\hat{\beta}]\right] \frac{\partial \hat{\Phi}}{\partial \hat{\beta}}$$

by the delta method, where

$$\frac{\partial \hat{\Phi}}{\partial \hat{\beta}} = \hat{\phi}(x_i\hat{\beta}) \cdot x_i'$$

- next, the asymptotically valid CI for $\hat{\Phi}$ is:

$$\hat{\Phi} \pm 1.96 \cdot \sqrt{\text{Asy Var}(\hat{\Phi})}$$

2. MARGINAL EFFECT

the estimated partial effect tells us how the probability of choosing the public transport would change if the time ~~dist~~ increased by 1 unit (or another individual characteristic increases by 1 unit)

$ME = \phi(x_i \beta) \cdot \beta$

the estimated ME:

- $\hat{ME} = \hat{\phi}(x_i' \hat{\beta}) \cdot \hat{\beta}$

- $\text{Asy var}(\hat{ME}) = \frac{\partial \hat{ME}}{\partial \hat{\beta}}' \left[\text{Asy Var}[\hat{\beta}] \right] \frac{\partial ME}{\partial \hat{\beta}}$

by the delta method, where

$$\frac{\partial ME}{\partial \hat{\beta}} = \hat{\phi}(x_i'\hat{\beta}) \cdot 1 + \frac{\partial \hat{\phi}(x_i \hat{\beta})}{\partial \hat{\beta}} \cdot \hat{\beta}$$

$$= \hat{\phi}(x_i\hat{\beta}) - \hat{\phi}(x_i\hat{\beta})\, x_i'\hat{\beta} x_i'\hat{\beta}$$

$$= \hat{\phi}(x_i\hat{\beta}) \left[1 - (x_i'\hat{\beta})(x_i'\hat{\beta}) \right]$$

- the asymptoticaly valid **CI** for $\hat{ME}$ is:

$$\hat{ME} \pm 1.96 \cdot \sqrt{\text{Asy var}(\hat{ME})}$$

Logit

instead of normal density $\phi(t)$ use

$$f(u_i) = \frac{\exp(-u_i)}{(1+\exp(-u_i))^2}$$

$$\text{Proba}(Y_i=1) = P(u_i \leq X_i'\lambda) = F(X_i'\lambda)$$

$$= \frac{1}{1+\exp(-X_i'\lambda)} = \frac{\exp(X_i'\lambda)}{1+\exp(X_i'\lambda)}$$

$$P(Y_i=0) = 1 - F(X_i'\lambda)$$

$$\ell = \prod_{i:Y_i=1} F(X_i'\lambda) \cdot \prod_{i:Y_i=0} [1 - F(X_i'\lambda)]$$

$$= \prod_{i:Y_i=1} \frac{\exp(X_i'\lambda)}{1+\exp(X_i'\lambda)} \prod_{i:Y_i=0} \frac{1}{1+\exp(X_i'\lambda)}$$

$$L = -\sum_{i=1}^{N} \ln[1+\exp(X_i'\lambda)] + \sum_{i:Y_i=1} X_i'\lambda$$