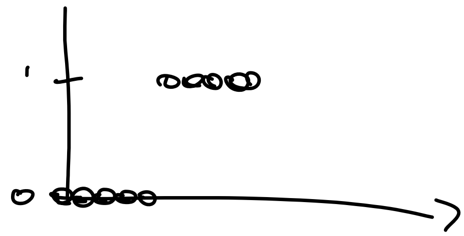
Output is dicotomic

y_i can take value only 0 and 1

linear regression model would not be good



we want to model the probability of $y = 1$, prob of success $P(y_i = 1|\Omega)$

$= E(y_i|\Omega)$ for the Bernoulli r.v.

we model P_i

$$P_i = E(y_i|\Omega) = F(x_i\beta)$$

• $F(-\infty) = 0$ • $F(\infty) = 1$ • $f(x) = \frac{\partial F(x)}{\partial x} > 0$ • the function should also be 1 to 1 (not harmonic)

these are also properties of CDFs

if we use the one of the normal we are in the probit model

$$\frac{\partial P_i}{\partial x_{ij}} = \frac{\partial F(x\beta)}{\partial x_{ij}} = f(x_i\beta)\beta_j$$

marginal effect caused by increase in x_i in a non linear model

$\Phi(x_i\beta)$ is probit model

LOGIT MODEL

$$\Lambda(x) = \frac{1}{1+e^{x}} = \frac{e^{x}}{1+e^{x}}$$

$$\lambda(x) = \frac{e^{x}}{(1+e^{x})^2}$$ is the derivative needed to compute the partial effect

$$\text{Prob}(Y_1 = y_1, \dots Y_n = y_n) = \prod_{\substack{y_i=0 \\ \downarrow}} [1-F(x_i\beta)] \prod_{y_i=1} F(x_i\beta)$$

LIKELIHOOD

the y that are $=0$

prob of not success

prob of success

$$P_r(Y_1 = y_1) = P_r(y_1=1) I(y_1=1) + P(y_1=0) I(y_1=0)$$

for the first obs

$$L(\beta; y, x) = \sum \log F(x_i\beta) + (1-y_i)\log(1-F(x_i\beta))$$

one cannot use the R^2 to assess the quality of the fit

- we can use likelihood ratio index

$$LRI = \frac{\text{loglik(Model 1)}}{\text{loglik(Model 0)}}$$

(likelihood ratio index)

$\to$ model with all the variables

$\hookrightarrow$ model with only the intercept $\nearrow$ $x = $ iota

measure of how much we improve in explanation adding the variables

LRI is not bounded between 0,1

Specific for LOGIT / PROBIT model

HIT AND MISS

OBS. OUTPUTS \ PREDICTED OUT	$\hat{P}(y_i=0)$	$\hat{P}(y_i=1)$
$y_i=0$	N_1	N_3 (miss)
$y_i=1$	N_4 (miss)	$N_2 = \sum_{i=1}^{N} I(P_{y_i}=1 \cap y_i=1)$

$\Sigma I(P_{y_i=0} \cap y_i=0)$

hat on because it is estimated

$\hat{P}(y_i=0) = F(x_i\hat{\beta}) < 0.5$

if the model predicts < 0.5 we expect to see a failure

then a rough measure of quality will be $\frac{N_1+N_2}{N}$

number of times the model correctly predicts the output

Q-ML or PSEUDO-ML

in ML we made the assumptional distribution, that it was correctly specified

$f_{y|x}(y|x)$ were generated from a known distribution for which the parameters were unknown

but what if this is misspecified

$y = X\beta + \varepsilon$ $\quad$ ε is iid $\sim N(0; D)$ errors are heteroskedastic

what happens if we maximize the wrong loglik function

LIKELIHOOD OF

H_1:

$y = X\beta + \varepsilon$ but now we assume the errors to be homoskedastic

what happens to $\varepsilon \sim N(0; I_N\sigma^2)$

$\hat{\beta}$, is $\hat{\beta}$ still consistent and what is its asymptotic distribution

$$\hat{\beta} = \arg\max_{\beta} \lambda(H_1; y_i, x_i; \beta)$$

$$= \arg\max_{\beta} \sum_{i=1}^{N} -\frac{1}{2} \log 2\pi + \log \sigma^2 + \left(\frac{(y_i - x_i'\beta)}{\sigma}\right)^2$$

① $\hat{\beta} \xrightarrow{p} \beta^0$

$\hat{\beta}$ is still a consistent estimator

② $\sqrt{N}(\hat{\beta} - \beta^0) \xrightarrow{d} N(0; H^{-1} \cdot OPG \cdot H^{-1})$

this should be equal to the white variance

$$H^{-1} = E\left[\frac{\partial^2 \lambda}{\partial\theta\partial\theta'}\bigg|_{\theta=\hat{\theta}}\right] \ (?)$$

the product is

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - x_i\beta)^2}$$

$$\log\sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(y_i - x_i\beta)^2$$

$$+ \frac{x_i}{\sigma^2}(y_i - x_i\beta) = X'y - X'X\beta$$

$$(\beta'X'X - y'X)(X'y - X'X\beta)$$

$$\beta'X'Xy \ldots \beta'X'X X'X\beta$$

$$- y'XX'y + y'XX'X\beta$$

we have a variable y that is binary, it can be only 1 or 0, so a regression does not make sense

$y = X\beta + \varepsilon$ regression for y usually continuous

in this case we want to model the probability of Y being $= 1$ or $= 0$

$\downarrow$

we model this as a function of the independent variables $F(x'\beta)$

in OLS $F(x'\beta) = x'\beta$, in this case this is not good because we need $Y \in [0, 1]$

$$F(x'\beta)$$

LOGIT

$$F(x'\beta) = \frac{e^{x'\beta}}{1 + e^{x'\beta}}$$

PROBIT

$$F(x'\beta) = \Phi(x'\beta)$$

these models are estimated through MLE

MARGINAL EFFECTS

Change in the probability that $y=1$
Given a unit change in one indep. var. x

$$\frac{\partial p}{\partial x_j} = \beta_j \quad \text{in OLS}$$

$$\frac{\partial p}{\partial x_j} = F'(\underline{x}'\beta)\beta_j$$

depends on x, typically we compute it at $\bar{x}$

or in case compute the average marginal effect

$$\frac{\sum F'(x'\beta)\beta_j}{n}$$

For logit $\dfrac{e^{x\beta}}{1+e^{x\beta}}$

$$\frac{\partial p}{\partial x_j} = \frac{e^{x\beta}}{(1+e^{x\beta})^2}\cdot\beta$$

ODDS ratio $\rightarrow$ $=2$ means that $P(y=1)$ is twice as likely as $P(y=0)$

$\ln \dfrac{p}{1-p} = x'\beta$ for logit

ODDS RATIO is $\dfrac{p}{1-p}$

$$p = \frac{e^{x\beta}}{1+e^{x\beta}} \qquad \ln p = \ln\left(\frac{e^{x\beta}}{1+e^{x\beta}}\right)$$

$$= x\beta - \ln\left(1 + e^{x\beta}\right)$$

$$p = \frac{e^{x\beta}}{1+e^{x\beta}}$$

$$\frac{p}{1-p} = \frac{\dfrac{e^{x\beta}}{1+e^{x\beta}}}{1 - \dfrac{e^{x\beta}}{1+e^{x\beta}}}$$

$$\ln\left(\frac{p}{1-p}\right) = \frac{x\beta - \ln\left(1+e^{x\beta}\right)}{\ln\left(1 - e^{x\beta}/1+e^{x\beta}\right)}$$

$$p = \frac{e^{x\beta}}{1+e^{x\beta}} \qquad 1 - \frac{e^{x\beta}}{1+e^{x\beta}} = \frac{1+\cancel{e^{x\beta}}-\cancel{e^{x\beta}}}{1+e^{x\beta}} = \frac{1}{1+e^{x\beta}} = 1-p$$

$$\frac{p}{1-p} = \frac{e^{x\beta}}{(1+e^{x\beta})} \cdot (1+e^{x\beta}) = e^{x\beta}$$

$$\boxed{\ln \frac{p}{1-p} = x\beta}$$

i have the yi which are =0 or =1

$$P(Y_1 = y_1, Y_2 = y_2 \ldots Y_N = y_N) =$$

$$\prod_{y=0} [1 - F(x'\beta)] \prod_{y=1} F(x'\beta)$$

$$(1-y_i) \ln(1 - F(x'\beta)) + y_i \ln F(x'\beta)$$

so that when $y_i = 1$ we have $\ln(F(x'\beta))$
and when $y_i = 0$ we have $\ln(1 - F(x'\beta))$

---

if i misspecify the function and use a normal with variance not diagonal but $\sigma^2 \varepsilon I_N$
i know that the sandwich estimator is

$$H^{-1} \, OPG \, H^{-1}$$

$$-\frac{1}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}(y_i - x_i\beta)^2$$

$$H^{-1} = -\sigma^2 \varepsilon (X'X)^{-1}$$

$$G_i = \frac{1}{\sigma^2} x_i (y_i - x_i\beta) = \frac{1}{\sigma^2} x_i' e_i$$

and gradient that is

$$G = \frac{1}{\sigma^2} \varepsilon' X$$

hence $H^{-1} G'G H^{-1}$

$$G'G = \frac{1}{\sigma^4} X' \underbrace{\varepsilon\varepsilon'}_{\text{diag}(\varepsilon_i^2)} X$$

$$= \frac{1}{\cancel{\sigma^4}} \cancel{\sigma^2} \varepsilon (X'X)^{-1} X' \text{diag}(\hat{\varepsilon}_i^2) X \cancel{\sigma^2} \varepsilon (X'X)^{-1}$$

$$= (X'X)^{-1} X' \text{diag}(\varepsilon_i^2) X (X'X)^{-1}$$

$$\frac{1}{\sigma^2} x_i (y_i - x_i \beta)$$

$$\frac{\partial}{\partial \beta} \frac{1}{\sigma^2} (x_i y_i - x_i x_i' \beta) = -\frac{1}{\sigma^2} (X'X)$$

$$H^{-1} = -\sigma^2 (X'X)^{-1}$$