Logit and Probit Models for Categorical Response Variable

Objective:

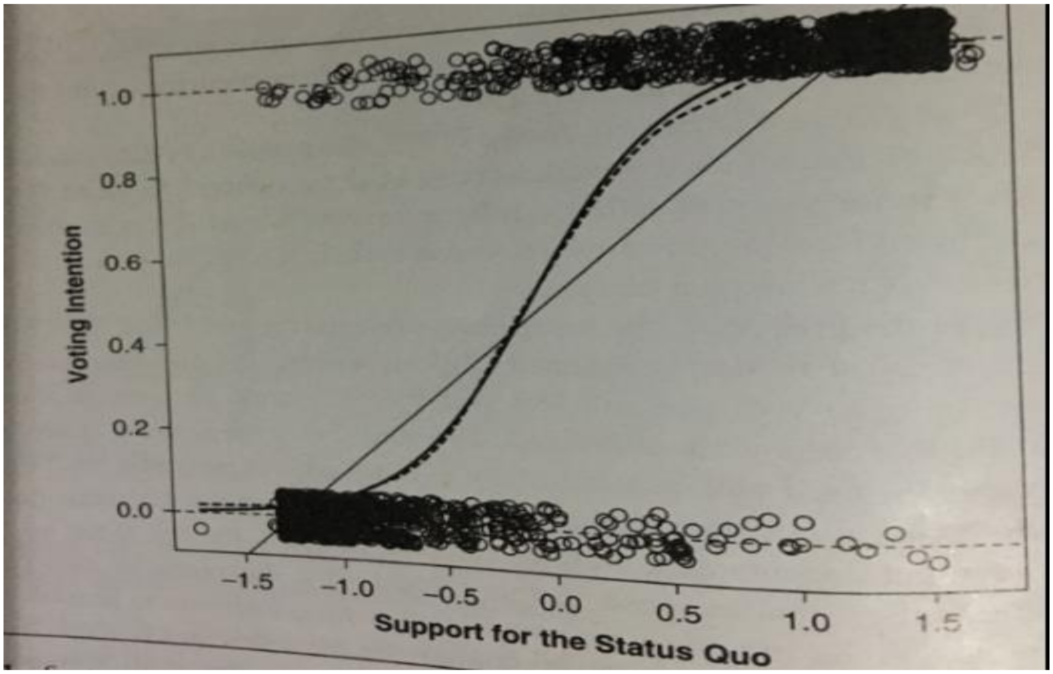
- Introduction to Generalized Linear Models: Logit and Probit are special cases of GLMs.
- Importance and usage of qualitative response variable with qualitative and quantitative regressors.

Dichotomous Data

- An example is for instance the Chilean plebiscite of yes and no vote for the military government to be ousted.
- Conditional average $E(Y \mid x_i)$ is the proportion of 1s ie the conditional probability of sampling a yes .
- $\pi_i = P_r(Y_i) = P(Y = 1 \mid X = x_i)$ ……………………….1

  $E(Y \mid x_i) = \pi_i(1) + (1 - \pi_i)(0) = \pi_i$ ……………………….2

  For discrete X we can evaluate the conditional proportion of Y at each X. The
- Collection of conditional proportion for Y represents sample non parametric regression of dichotomous Y on X.
- For continuous X (here support for status quo) value local averaging can be used.



- Notice at low levels of support for status quo the proportion of yes variables is close to 0 and at high levels of support for status quo the proportion of yes is close to 1.

- In the middle areas the non parametric regression is elongated S curve.

Linear Probability Model

- The non parametric model can be used but we can also formulate as a linear regression model it using the generic regression assumptions:
- $Y_i = \alpha + \beta X_i + \varepsilon_i$ 3

  $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ ε_i and $\varepsilon_{i'}$ are independent

  $E(Y_i) = \alpha + \beta X_i$ 4

  By 2 , 4 transforms to

  $\pi_i = \alpha + \beta X_i$

- Since π_i is a conditional probability therefore the aforementioned model is called the **Linear Probability Model**.

- Since Yi can take values of 0 or 1 the errors ε_i are also dichotomous. The error distribution is not normally distributed but if the sample size is large enough by central limit theorem the assumption of normality is not required.
- When $Y_i=1$

  $\varepsilon_i = 1 - E(Y_i) = 1 - \pi_i$ subtract 3 from 4……………………5
- When $Y_i=0$

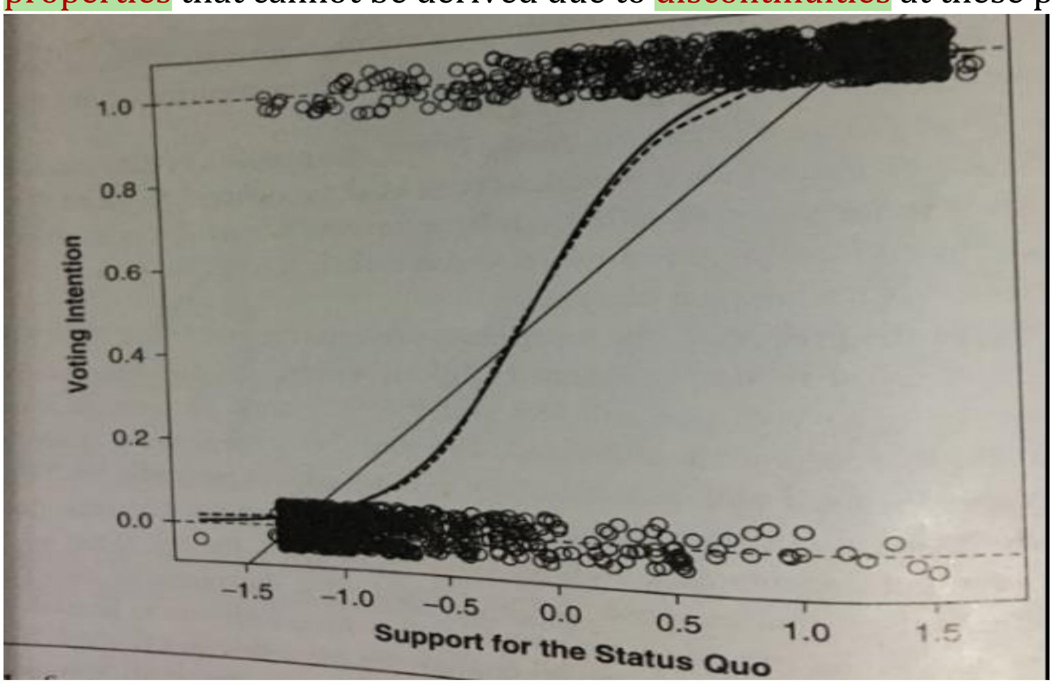
  $\varepsilon_i = 0 - E(Y_i) = 0 - \pi_i = -\pi_i$ subtract 3 from 4………………………6
- The variance of ε is not constant

  $V(\varepsilon_i) = \pi_i(1-\pi_i)^2 + (1-\pi_i)(-\pi_i)^2 = \pi_i(1-\pi_i)$……………………..7

  The heteroscedasticity of the errors creates problems at the boundaries where π_i gets close to 0 or 1

  π_i can only lie in the interval [0,1] as the probabilities cannot be less than 0 or more than 1.

  If we have a broad X range we can confine the X range for probabilities between 0 and 1. This one solution. It is called the Constrained Linear Model. The abrupt change of slope at 0 and 1 provides statistical properties that cannot be derived due to discontinuities at these points.



Transformation of π: Logit and Probit

- By transformation we will ensure that π will be constrained in the interval [0,1].
- We will use a positive monotone (non decreasing) function that maps the linear predictor $\eta = \alpha + \beta X$ into unit interval .
- $\pi_i = P(\eta_i) = P(\alpha + \beta X_i)$ 8
  Where the CDF P(.) is chosen before-hand and then parameters α and β can be estimated.

- $P^{-1}(\pi_i) = \eta_i = \alpha + \beta X_i$ 9
- Where P^{-1} is the inverse CDF.

The transformation P(.) is usually chosen as cdf of normal distribution N(0,1):

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-\frac{1}{2}Z^2)dZ$$ 10

or Logistic distribution:

$$\Lambda(z) = \frac{1}{1+e^{-z}}$$ 11

With the normal distribution $\phi(.)$ creates linear probit model:

$$\pi_i = \phi(\alpha + \beta X_i)$$ 12

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha+\beta X_i} \exp(-\frac{1}{2}Z^2)dZ$$ 13

With the logistic distribution $\phi(.)$ creates linear logit model:

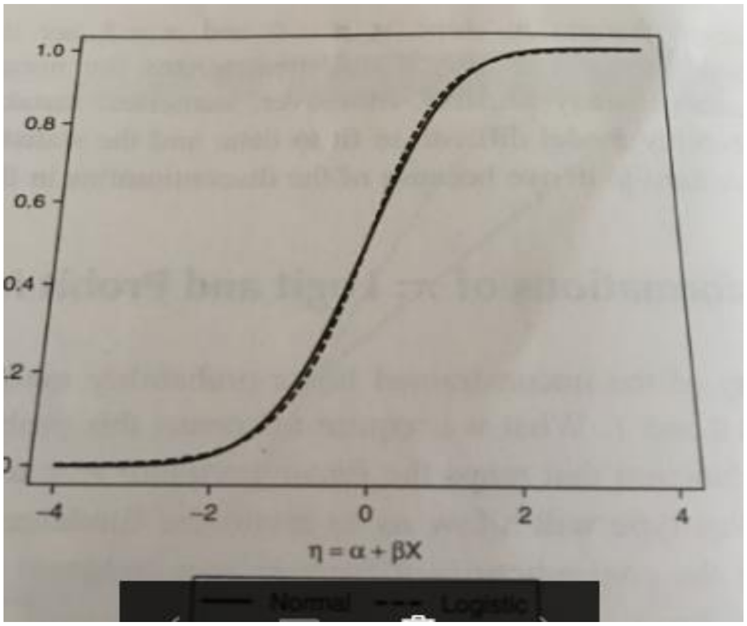
$$\pi_i = \Lambda(\alpha + \beta X_i)$$ 14

$$\pi_i = \frac{1}{1+\exp[-(\alpha + \beta X_i)]}$$ 15

Can be written as

$$\pi_i = \frac{\exp[(\alpha + \beta X_i)]}{1+\exp[(\alpha + \beta X_i)]}$$

Logit and probit are functions give very similar as they are both linear on the range π_i =.2 and π_i = .8. They can only be distinguished by a large data. See how logit and probit are superimposed for the Chile plebiscite data.



Comparing logit and probit:

- The logistic CDF ($\Lambda(z) = \frac{1}{1+e^{-z}}$) is very simple while the normal CDF($\phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-\frac{1}{2} Z^2) dZ$) has a unevaluated integral. For the dichotomous data we get good approximations of normal CDF but for polytomous data logit model is better.
- Transforming equation 15 to obtain:

$$\frac{\pi_i}{1-\pi_i} = \exp(\alpha + \beta X_i) \text{ ……………………..16}$$

$\frac{\pi_i}{1-\pi_i}$ is the odds of Y_i=1 voting to say yes.

Taking log of both sides:

$$\log_e \frac{\pi_i}{1-\pi_i} = \alpha + \beta X_i \text{ ....17}$$

By equation 14 $\pi_i = \Lambda(\alpha + \beta X_i)$ therefore $\Lambda^{-1}(\pi_i) = \alpha + \beta X_i$

Now 17 can be transformed as

$\log_e \frac{\pi_i}{1-\pi_i} = \Lambda^{-1}(\pi_i)$ The inverse transformation is called logit of π

Probability π	Odds $\frac{\pi}{1-\pi}$	Logit $\log_e \frac{\pi}{1-\pi}$
.01	1/99 = 0.0101	−4.60
.05	5/95 = 0.0526	−2.94
.10	1/9 = 0.1111	−2.20
.30	3/7 = 0.4286	−0.85
.50	5/5 = 1	0.00
.70	7/3 = 2.3333	0.85
.90	9/1 = 9	2.20
.95	95/5 = 19	2.94
.99	99/1 = 99	4.60

Here if we have odds that are even ie equal to 1 corresponding to $\pi_i = .5$ then logit =0. The logit is symmetric about 0 .Therefore it is a good candidate for the response variable in a linear model.

By the equation $\frac{\pi_i}{1-\pi_i} = \exp(\alpha + \beta X_i)$ the model obtained is an additive model for log odds.

This is also a multiplicative model since:

$$\frac{\pi_i}{1-\pi_i} = e^{\alpha}(e^{\beta})^{X_i}$$ ……………………………….18

By increasing X by 1 changes the logit by β since by 17

$$\log_e \frac{\pi_i}{1-\pi_i} = \alpha + \beta X_i$$

And the odds increases by a factor of e^{β} using equation 18

Alternatively to interpret β:

Considering the relationship between π and X in equation

$$\pi_i = \frac{1}{1+\exp[-(\alpha+\beta X_i)]}$$

The slope found by $\frac{d\pi}{dX_i} = \beta\pi(1-\pi)$

The slope is maximum when π =1/2

$$\beta\pi(1-\pi) = \beta\frac{1}{2}(1-\frac{1}{2}) = \frac{\beta}{4}$$

π	$\beta\pi(1-\pi)$
.01	$\beta \times .0099$
.05	$\beta \times .0475$
.10	$\beta \times .09$
.20	$\beta \times .16$
.50	$\beta \times .25$
.80	$\beta \times .16$
.90	$\beta \times .09$
.95	$\beta \times .0475$
.99	$\beta \times .0099$

We can see that the relationship between π and $\beta\pi(1-\pi)$ is nearly linear between π=.2 and π=.8

http://web.stanford.edu/class/psych252/tutorials/Tutorial_LogisticRegression.html