Single-level Models for Binary Responses

Distribution of Binary Data

y_i response for individual i ($i = 1, \ldots, n$), coded 0 or 1

Denote by r the number in the sample with $y = 1$

Mean and variance

$\text{E}(y) = \pi = \Pr(y = 1)$ estimated by $\hat{\pi} = \frac{r}{n}$

$\text{var}(y) = \pi(1 - \pi)$ estimated by $\hat{\pi}(1 - \hat{\pi})$

So the mean and variance of y are determined by a single parameter π

y follows a Bernoulli distribution (special case of binomial distribution)

Linear Probability Model

Model for the mean of y

$$E(y_i) = \pi_i = \beta_0 + \beta_1 x_i$$

or expressed as a model for y_i

$$\begin{aligned} y_i &= \pi_i + e_i \\ &= \beta_0 + \beta_1 x_i + e_i \end{aligned}$$

where $e_i \sim N(0, \sigma_e^2)$

Estimate using ordinary least squares (as for continuous y).

Problems with the Linear Probability Model

- Residuals $e_i = y_i - (\beta_0 + \beta_1 x_i)$ can only take two possible values for a given x_i, so not normally distributed
- $\text{var}(y_i) = \pi(1-\pi) = (\beta_0 + \beta_1 x_i)[1 - (\beta_0 + \beta_1 x_i)]$ which depends on x_i so not homoskedastic
- Relationship between π and x may be nonlinear, although linearity assumption often reasonable for π between 0.2 and 0.8
- Possible to get predicted probabilities outside [0,1]. Again, this is unlikely if π lies between 0.2 and 0.8 for all x (or combinations of values on a set of x variables)

The Generalised Linear Model

Work with a nonlinear transformation of $\beta_0 + \beta_1 x_i$ that ensures predicted probabilities will lie between 0 and 1

General model

$$\pi_i = F(\beta_0 + \beta_1 x_i)$$

where F usually chosen to be the cumulative distribution function (cdf) of a logistic or normal distribution

usually written as

$$F^{-1}(\pi_i) = \beta_0 + \beta_1 x_i$$

where F^{-1} is called the link function

The Logit/Logistic Model

Write $z = \beta_0 + \beta_1 x$

Logistic transformation of z

$$\pi = F(z) = \frac{\exp(z)}{1 + \exp(z)} = \frac{e^z}{1 + e^z}$$

π always lies between 0 and 1

Logit model

$$\log\left(\frac{\pi}{1-\pi}\right) = z = \beta_0 + \beta_1 x$$

where $\pi/(1-\pi)$ is the odds that $y = 1$ and $\log[\pi/(1-\pi)]$ is the log-odds or logit

Interpretation of Logit Coefficients

Take exponentials of each side of logit model:

$$\frac{\pi}{1-\pi} = \exp(\beta_0 + \beta_1 x) = \exp(\beta_0).\exp(\beta_1 x) \quad (1)$$

Now increase x by 1 unit:

$$\frac{\pi}{1-\pi} = \exp(\beta_0 + \beta_1(x+1)) = \exp(\beta_0).\exp(\beta_1 x).\exp(\beta_1) \quad (2)$$

Comparing (1) and (2) we see that a 1-unit increase in x has multiplied the odds that $y = 1$ by $\exp(\beta_1)$, or increased the odds by a factor of $\exp(\beta_1)$.

Interpret $\exp(\beta_1)$ as an odds ratio, comparing the odds for two individuals with x-values spaced 1 unit apart.

Example: State Differences in US Voting

$y = 1$ if intends to vote Bush in 2004 election, 0 otherwise

Variable	$\hat{\beta}$	$se(\hat{\beta})$	$\exp(\hat{\beta})$
Constant	-0.34	0.05	0.71
State (ref=California)			
New York	-0.19	0.08	0.83
Texas	0.69	0.08	2.00

- Odds of voting Bush in California = 0.71
- Odds of voting Bush in New York are 0.83 times odds in California (17% lower)
- Odds of voting Bush in Texas are twice the odds in California (100% higher)

Other Link Functions

Probit model

$$\Phi^{-1}(\pi) = \beta_0 + \beta_1 x$$

where Φ is the cdf of the standard normal distribution

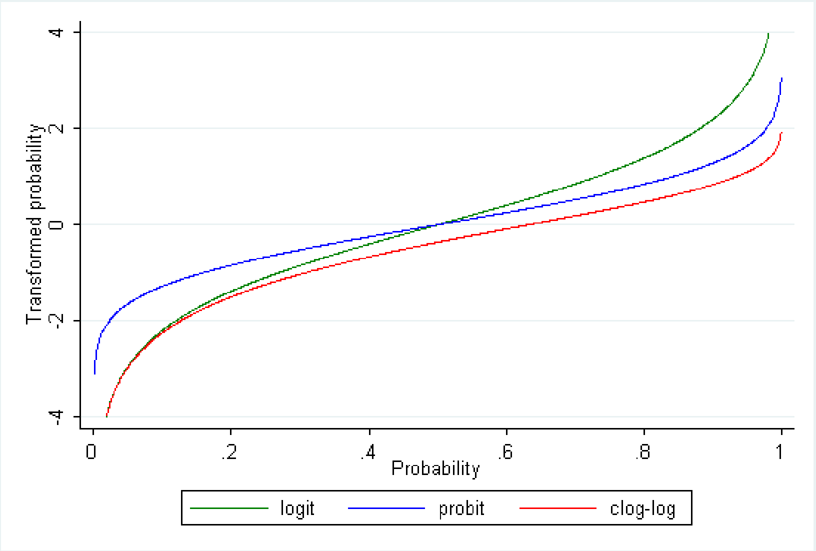
Complementary log-log (clog-log) model

$$F^{-1}(\pi) = \log[-\log(1-\pi)] = \beta_0 + \beta_1 x$$

Choice of link function

- In general, get very similar $\hat{\pi}$ whatever link is used
- Logit(0.5) = probit(0.5) but move further apart as π gets close to 0 or 1
- Logit and clog-log almost indistinguishable for small π

Logit, probit and clog-log transformations of π



Latent Variable Representation

A GLM expresses the mean of a binary y, $\mathrm{E}(y) = \pi$, as a function of covariates x.

Another way to represent a GLM is in terms of a latent (unobserved) continuous variable y^* that underlies y such that

$$y_i = \begin{cases} 1 & \text{if } y_i^* \geq 0 \\ 0 & \text{if } y_i^* < 0 \end{cases}$$

Threshold model

$$y_i^* = \beta_0 + \beta_1 x_i + e_i^*$$

- $e_i^* \sim N(0, 1) \rightarrow$ probit model
- $e_i^* \sim$ standard logistic (with variance $\simeq 3.29$) $\rightarrow$ logit model

Relationship between probit and logit

The residual in the threshold model has fixed variance, but the value it is fixed at depends on the chosen distribution:

Distribution	Variance	Link
Standard normal	1	probit
Standard logistic	3.29	logit

Increasing the variance (and therefore scale of y^*) increases the magnitude of the coefficients.

$$\hat{\beta}_{\text{logit}} \simeq \sqrt{3.29}\ \hat{\beta}_{\text{probit}} = 1.8\ \hat{\beta}_{\text{probit}}$$

US Voting: Logit and Probit Coefficients

Compare coefficients from logit and probit models of voting Bush.

Variable	$\hat{\beta}_{\text{probit}}$	$\hat{\beta}_{\text{logit}}$	$\hat{\beta}_{\text{logit}}/\hat{\beta}_{\text{probit}}$
Constant	−0.21	−0.34	1.62
State (ref=California)			
New York	−0.12	−0.19	1.58
Texas	0.43	0.69	1.60

Ratio of logit:probit is approximately 1.6

But, for this simple model, predicted probabilities of voting Bush will be exactly the same for logit and probit models.

Predicted Probabilities of Voting Bush

Logit model

$$\hat{\pi} = \frac{\exp(-0.34 - 0.19\ NY + 0.69\ TEX)}{1 + \exp(-0.34 - 0.19\ NY + 0.69\ TEX)}$$

California	(NY = TEX = 0)	$\hat{\pi} = 0.71/(1 + 0.71) = 0.42$
New York	(NY = 1, TEX = 0)	$\hat{\pi} = 0.59/(1 + 0.59) = 0.37$
Texas	(NY = 0, TEX = 1)	$\hat{\pi} = 1.42/(1 + 1.42) = 0.59$

Probit model

$$\hat{\pi} = \Phi(-0.21 - 0.12\ NY + 0.43\ TEX)$$

California	(NY = TEX = 0)	$\hat{\pi} = \Phi(-0.21) = 0.42$
New York	(NY = 1, TEX = 0)	$\hat{\pi} = \Phi(-0.33) = 0.37$
Texas	(NY = 0, TEX = 1)	$\hat{\pi} = \Phi(0.22) = 0.59$

Calculate using NORMDIST in Excel or CDFNORM in SPSS

Significance Testing

Suppose we wish to test $\mathrm{H}_0 : \beta_1 = 0$ versus $\mathrm{H}_0 : \beta_1 \neq 0$

Z-ratios

Compare $Z = \hat{\beta}/se(\hat{\beta})$ with N(0,1)

Wald test

Compare Wald $= Z^2$ with χ^2_1

Also used to test more general hypotheses, e.g. $\mathrm{H}_0 : \beta_1 = \beta_2 = 0$ or $\mathrm{H}_0 : \beta_1 = \beta_2$

Likelihood ratio test

Compare difference in deviance (-2 log-likelihood) between two nested models with χ^2_q, where $q =$ difference in number of parameters. Not available in MLwiN for discrete response models.

Confidence Intervals for β and Odds Ratios

95% CI for β

$$(\hat{\beta} - 1.96\ se(\hat{\beta}),\ \hat{\beta} + 1.96\ se(\hat{\beta}))$$

E.g. 95% CI for NY-California difference in log-odds of voting Bush is $-0.19 \pm (1.96 \times 0.08) = -0.35$ to -0.03

95% CI for odds ratio, $\exp(\beta)$

Take exponent of lower and upper limits of 95% interval for β.

E.g. 95% CI for NY vs California odds ratio is $\exp(-0.35)$ to $\exp(-0.03) = 0.70$ to 0.97

Adding Further Predictors to Logit Model

Variable	$\hat{\beta}$	$se(\hat{\beta})$	Wald	$\exp(\hat{\beta})$
Constant	-0.42	0.11	-	0.66
State (ref=California)				
New York	-0.18	0.08	4.4	0.84
Texas	0.71	0.08	76.7	2.03
Female	-0.27	0.07	15.9	0.76
Age (years)	0.005	0.002	5.3	1.005

Note: Wald for joint test of state effects is 111.6 on 2 df ($\chi^2_{2;0.05} = 5.99$). All other tests on 1 df ($\chi^2_{1;0.05} = 3.84$).

- Odds of voting Bush are 25% lower for women than for men (controlling for state and age)
- Odds of voting Bush increased by 0.5% for each 1-year increase in age (controlling for state and gender)

Predicted Probabilities from Logit Model

Variable	Value	$\hat{\pi}$
State	California	0.42
	New York	0.37
	Texas	0.59
Sex	Male	0.49
	Female	0.42
Age (years)	20	0.42
	30	0.43
	40	0.45

- Values of each variable varied in turn, holding others at sample mean (proportions for categorical variables)
- E.g. $\hat{\pi}$ for State calculated for each possible set of values for the State dummies, with Female=0.54 and Age=46.7 years

Interaction Effects

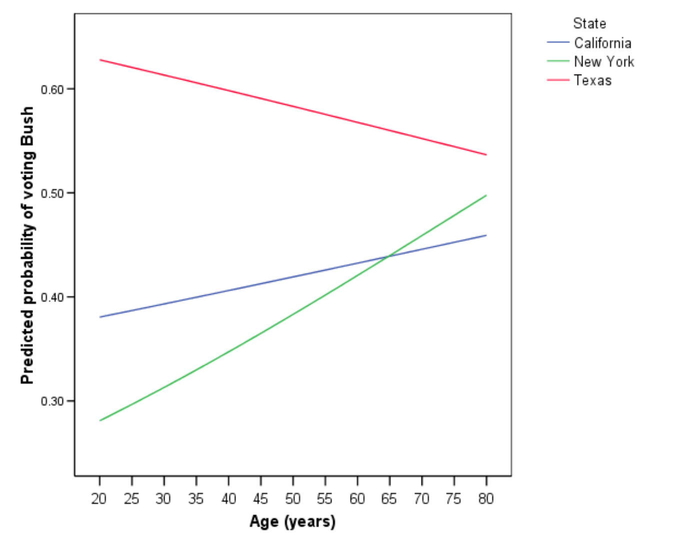
Suppose we believe the effect of age on voting intentions differs across states, then fit:

$$\begin{aligned}\log\left(\frac{\pi}{1-\pi}\right) &= \beta_0 + \beta_1\ \mathsf{NY} + \beta_2\ \mathsf{TEX} + \beta_3\ \mathsf{FEMALE} + \beta_4\ \mathsf{AGE} \\ &+ \beta_5\ \mathsf{NY}\cdot\mathsf{AGE} + \beta_6\ \mathsf{TEX}\cdot\mathsf{AGE}\end{aligned}$$

State	Age effect	Estimated age effect
California	β_4	0.005
New York	$\beta_4 + \beta_5$	0.005+0.010
Texas	$\beta_4 + \beta_6$	0.005−0.012

Test of $H_0 : \beta_5 = \beta_6 = 0$ gives Wald$= 15.3$ on 2 df, so strong evidence of interaction effect

$\hat{\pi}$ by Age and State



Note: Female fixed at 0.54 (sample proportion)