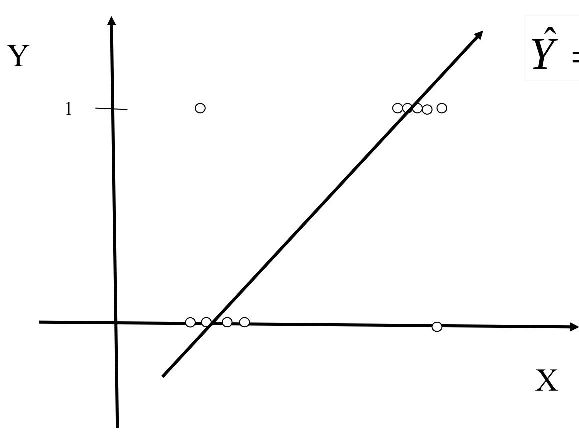
Logistic regression

- Estimates a model relating yes or no outcomes to exogenous factors
- Intuitive coding: 0=no; 1=yes (different programs handle differently)
- Use observed outcomes to tap underlying probabilities

Logistic regression:

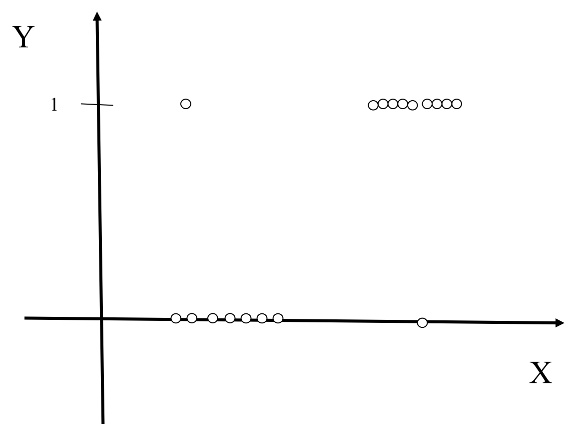
problems with OLS



$$\hat{Y} = a + bX + e$$

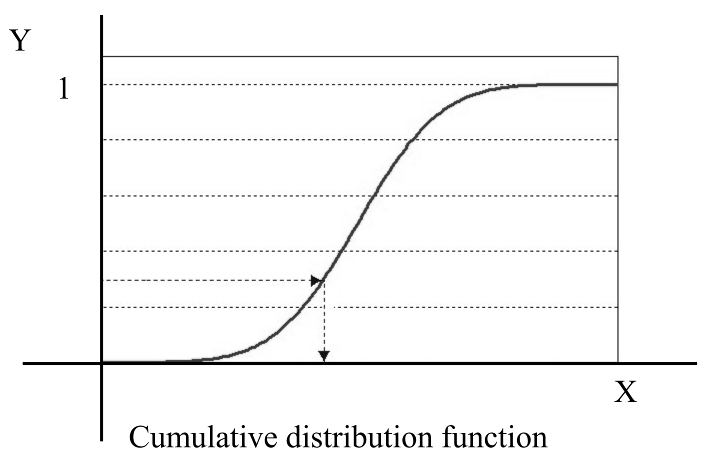
Logistic regression:

reconceptualize: probability of 1’s for each fixed X

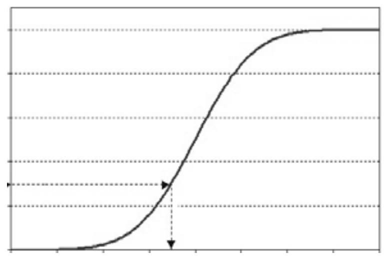


Logistic regression:

estimating the slope

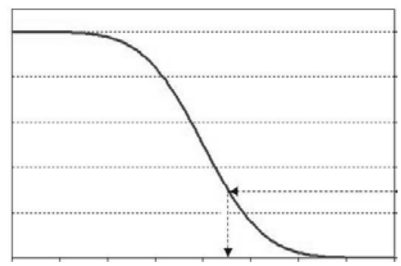


Logistic regression



+ relationship

- relationship



Logistic model

$$P(Y_t = 1) = F(X_t)$$

$F(X_t)$ = cumulative density function of probabilities, expressed as function of X’s

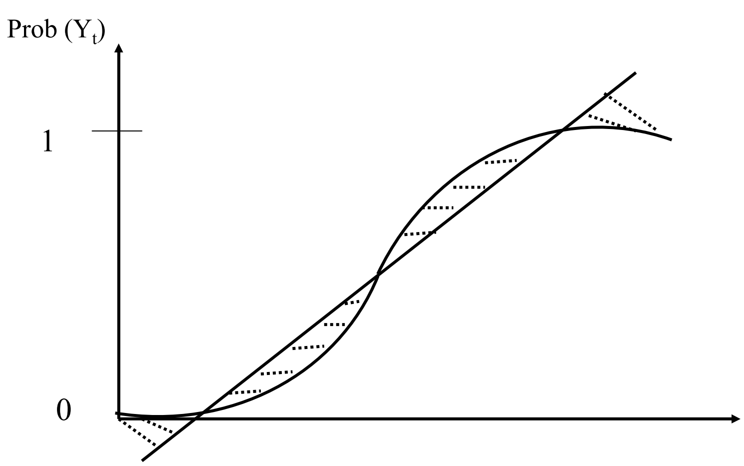
Y_t = binary variable 0 or 1 over *t* individuals

Logistic model

- How do exogenous variables affect the probabilities of choice?
- Expect $F(X_t)$ to be nonlinear because:

1) $F(X_t)$ must fall between 0 and 1 (nonlinear at boundaries)
2) Empirical observation shows cumulative density function follows S-shaped form
3) Additive form particularly inappropriate with several exogenous variables (interaction between variables)

Problems using OLS vs. Logit: errors vary systematically with X



$$\varepsilon_t = F(X_t) - (\beta_1 + \beta_2 X)$$

Problems using OLS vs. Logit: errors from dichotomous dependent variable

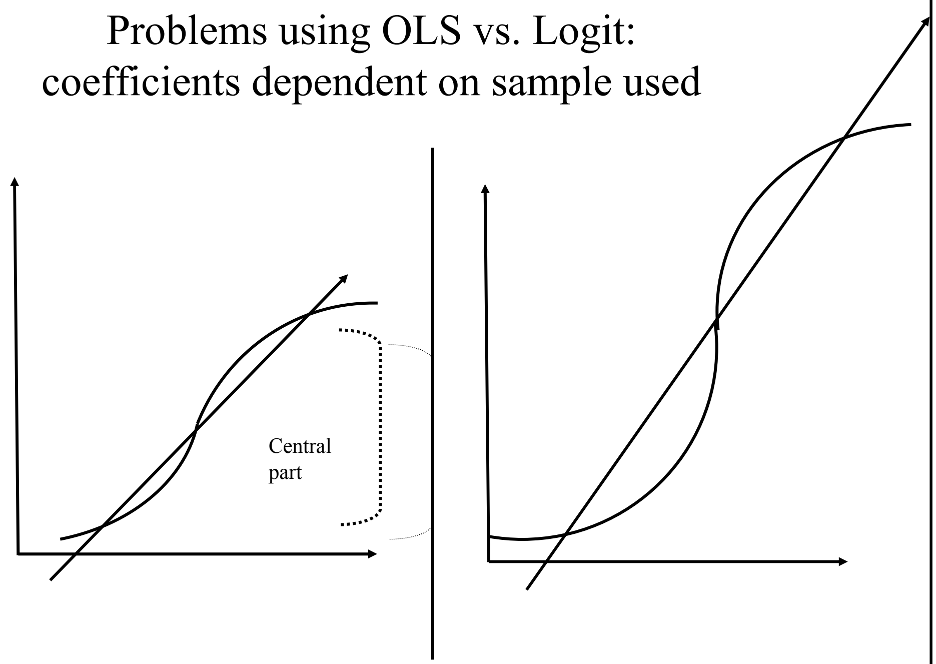
$$1 - F(X_t)$$
$$0 - F(X_t) => -F(X_t)$$

For S-shaped function

$$1 - (a + bX)$$
$$0 - (a + bX) => -(a + bX)$$

For OLS function

Problems using OLS vs. Logit: coefficients dependent on sample used



Functional form of relationship

$$P_t = \text{Prob}(Y_t = 1) = F(X_t\beta), \text{and}$$
$$1 - P_t = \text{Prob}(Y_t = 0) = 1 - F(X_t\beta)$$

$F(X_t\beta)$ = cumulative distribution function describing how probabilities are related to exogenous variables [in matrix notation]

Which cumulative distribution function to use?

$$P = 1/(1 + e^{-X\beta}), \text{where}$$
$$X\beta = a + b_1 X_1 + b_2 X_2 ... b_k X_k$$

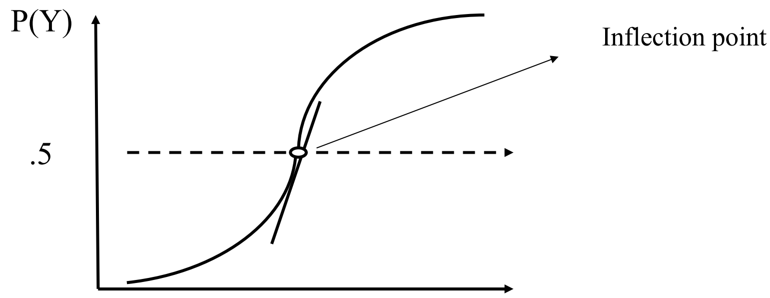
SAS default

Ranges from 0 to 1, as XB ranges from $-\infty$ to $+\infty$

Alternative formula (algebraically equivalent):

$$P = e^{X\beta} / (1 + e^{X\beta})$$

Convenient properties: inflection point



Inflection point (e.g., threshold point): where slope of line shows greatest rate of change

Convenient properties: calculating 1-P

$$\text{If } P = 1/(1+e^{-X\beta}) \text{ then,}$$

$$\begin{aligned}
1-P &= 1-(1/(1+e^{-X\beta})) \\
&= (1+e^{-X\beta}-1)/(1+e^{-X\beta}) \\
&= e^{-X\beta}/(1+e^{-X\beta}) \\
&= e^{-X\beta}/(1+1/e^{X\beta}) \\
&= e^{-X\beta}/((e^{X\beta}+1)/e^{X\beta}) \\
&= e^{-X\beta}*(e^{X\beta}/(e^{X\beta}+1)) \\
&= (1/e^{X\beta})*(e^{X\beta}/(e^{X\beta}+1)) \\
1-P &= 1/(1+e^{X\beta})
\end{aligned}$$

Logging the odds:
producing an OLS model

$$
\begin{aligned}
L &= \log(P/(1-P)) \\
&= \log P - \log(1-P) \\
&= \log(1/(1+e^{-X\beta})) - \log(e^{-X\beta}/(1+e^{-X\beta})) \\
&= [\log 1 - \log(1+e^{-X\beta})] - [\log e^{-X\beta} - \log(1+e^{-X\beta})] \\
&= -\log(1+e^{-X\beta}) - \log e^{-X\beta} + \log(1+e^{-X\beta}) \\
&= -\log e^{-X\beta} \\
&= -(-X\beta) \\
L &= X\beta
\end{aligned}
$$

Log of the odds

Thus:

Logging the odds yields a linear regression model:

$$X\beta = a + b_1 X_1 + b_2 X_2 \ldots b_k X_k$$

Log of the odds

- L=logit, or log of the odds
- As P goes from 0 to 1, L ranges from $-\infty$ to $+\infty$
- When P=.5, then log(P/1-P) = log 1 = 0 (logit=0 when P=.5), hence
- When P<.5, logit is negative, when P>.5, logit is positive

Estimating coefficients: MLE vs. OLS

- Maximum likelihood technique vs. least squares
- OLS: parameter estimates (b's) that minimize SSE
- MLE: parameter estimates that imply the highest likelihood of obtaining observed sample
- Maximizes the log of the likelihood function

Estimating coefficients: MLE

From above:

$$P_t = 1/(1 + e^{-X_t\beta}), \ and$$
$$1 - P_t = 1/(1 + e^{X_t\beta})$$

✓***Logic of MLE***: uses these to derive an expression (***the likelihood function***) for the likelihood of observing the pattern of successes (Y_t=1) and non-successes (Y_t=0) in the data

✓Value of likelihood function depends on unknown parameters (β)

Estimating coefficients: MLE

$$Likelihood \ \ function =$$

$$\text{Prob}(Y_1, Y_2 ... Y_{T1} ..., Y_T) = \prod_{t=1}^{T} P_t^{Y_t}(1 - P_t)^{1-Y_t}$$

✓Likelihood of obtaining the given sample = the ***product of the probabilities*** of the individual observations having the observed outcomes (e.g., multiplication of all the probabilities of P*(1-P) for each case)

Source: Hanushek & Jackson, 1977, section 7.5

Estimating coefficients: MLE

✓Choose b's that maximize the likelihood function

✓What underlying parameters are most likely to produce the observed data?

✓Maximizes the log of the likelihood function, not the likelihood function itself

✓Parameters are linear on the log of the odds

✓Use any kind of independent variables (e.g., dummies, X^2, logX, interval vars)

Interpretation of coefficients

- After MLE, coefficients interpretable as:

"Increment to the log odds of achieving a value of 1 on the dependent variable associated with a unit increase in the independent variable"

- Not readily interpretable, so transform to odds ratios and probabilities

Interpretation of coefficients: senility data

- Coefficients: The log odds of being senile decreases by .3235 for each unit increase in intelligence.
- Odds ratio (antilog of parameter estimate):

  With each unit increase in intelligence, the odds of senility is multiplied by .724 (increasing intelligence decreases senility).

Calculating probablities

$$P = e^{X\beta} / (1 + e^{X\beta})$$

$$\Downarrow$$

$$P = e^{a+bX} / (1 + e^{a+bX})$$

Substitute values of X into equation to get predicted probabilities.