

Econometric Methods for Political Analysis

Volume II: Causal Estimation for Observational Studies

Kevin Lingfeng Li

Table of contents

I	Maximum Likelihood Estimator	2
1	Binomial Logistic Regression	3
1.1	Binary Outcome Variables	3
1.2	Logistic Regression Model	4
1.3	Odds Ratios and Interpretation	5
1.4	Confidence Intervals and Hypothesis Tests	7
1.5	Likelihood Ratio Test of Models	7
	Implementation in R	7
	Implementation in STATA	7
2	Maximum Likelihood Estimation	8
2.1	Likelihood Functions	8
2.2	Simple Maximum Likelihood Estimator	8
2.3	Gradient Descent Estimation Approaches	8
2.4	Properties of the ML Estimator	8
2.5	Probit Model	8
3	Multinomial and Ordinal Logistic Regression	9
4	Regression for Counts	10
II	Selection on Observables	11
5	Selection on Observables Design	12
6	Matching, Weighting, and Regression	13
7	Partial Identification and Sensitivity Analysis	14

Part I

Maximum Likelihood Estimator

Chapter 1

Binomial Logistic Regression

1.1 Binary Outcome Variables

Some outcome variables Y are limited in the possible values that they can take on. This can include binary outcome variables that only take values 0 or 1, or categorical variables with 3 or more possible values.

Key Definition: Binary Outcome Variables

Binary Outcome Variables have two possible values/categories: $Y = 1$ and $Y = 0$. Examples of binary outcome variables include yes/no, true/false, vote/abstain, treatment/control, and correct/incorrect.

For binary response variables, we are interested in the **proportion** of the units that are in category $Y = 1$ (and knowing that, we know the rest of the units are in category $Y = 0$). More formally, we define $\pi = Pr(Y = 1)$, and $1 - \pi = Pr(Y = 0)$, where $\pi \in [0, 1]$.

- If $\pi = 0$, that means no units are in category $Y = 1$ (and thus all are in category $Y = 0$).
- If $\pi = 1$ that means all units are in category $Y = 1$ (and thus none are in category $Y = 0$).

We can also think of the proportion π as a **probability** of getting an observation in category $Y = 1$ if we were randomly selecting from all the observations.

More formally, the binary Y variable takes a Bernoulli distribution, which has only has one parameter, the probability $\pi = Pr(Y = 1)$. The probability π is also the expectation (mean) of the distribution: $\mathbb{E}[Y] = \pi$. The variance of a Bernoulli distribution is $Var(Y) = \pi(1 - \pi)$. Thus, the variance is not considered a separate parameter (unlike normal regression), since it is just a function of the main parameter π .

Sometimes, we use linear regression to model binary Y variables. We previously discussed the **linear probability model**, which is exactly this. However, the linear probability model has a few key issues:

1. A linear regression model always goes towards $\pm\infty$ on both sides as long as the slope is not 0 (since it is a line). However, probability π must be between 0 and 1. Thus, linear regression may give absurd predictions outside of $\pi \in [0, 1]$.
2. Assumptions of heteroscedasticity are also not possible with a binary Y , but that is more minor than the first issue.

Thus, we need some new model, that restricts π estimates to within 0 and 1.

1.2 Logistic Regression Model

Let us say we have data $(Y_i, X_{1i}, \dots, X_{ki})$ for n observations $i = 1, \dots, n$, where Y is a binary response variable with values 0 and 1. The logistic regression takes the form:

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \alpha + \beta_1 X_{1i} + \dots + \beta_k X_{ki}$$

Where $\pi_i = Pr(Y_i = 1)$, α and β_1, \dots, β_k are regression coefficients that need to be estimated.

However, this form of the model is not very useful - after all, we are interested in the probability an observation i is in category $Y = 1$, or in other words, we are interested in π_i . So, let us isolate π and get the model in that form.

💡 Key Definition: Logistic Regression Model

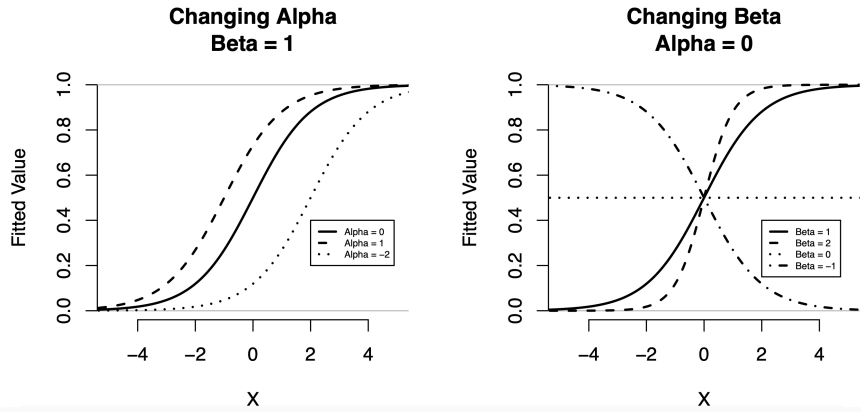
The logistic regression is a model of the probability of an observation i being in category $Y = 1$, given the values of \vec{X} . The model takes the form:

$$\pi_i = \frac{e^{\alpha + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}}{1 + e^{\alpha + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}}$$

Where $\pi_i = Pr(Y_i = 1)$, α and β_1, \dots, β_k are regression coefficients that need to be estimated.

The parameters of the logistic model are estimated using maximum likelihood estimation, which we will cover in detail in the next chapter.

The logistic regression function results, when plotted with one X variable, results in a curve as seen in the figure below:



As we can see, the curve will never exceed 0 or 1 on the Y level. This allows us to have reasonable interpretations of the probability π_i . As we see from above, the parameters change the shape of the logistic curve:

- When α is higher, fitted probabilities shift upward, and when α is lower, fitted probabilities shift downward.
- When $\beta > 0$, increasing X increases π . When $\beta < 0$, increasing X decreases π . The larger the absolute value of β is, the steeper the logistic curve (the stronger the relationship).

However, the actual magnitudes of the coefficients are not directly interpretable (unlike linear regression). To do this, we need to introduce another concept: Odds Ratios.

1.3 Odds Ratios and Interpretation

Odds of an Event

The **odds** are a ratio of the probabilities of an event occurring and an event not occurring. In terms of π , odds are defined as:

$$\text{Odds} = \frac{\pi}{1 - \pi}$$

Since π is the probability of $Y = 1$ occurring, and $1 - \pi$ is the probability of $Y = 1$ not occurring.

You might notice this looks very similar to the first logistic regression I presented in the last section. The logistic regression is a linear model of the log-odds: $\log\left(\frac{\pi}{1-\pi}\right)$.

Conditional Odds are odds of an event occurring, given a specific explanatory variable value.

- For example, the odds of an event occurring, given some explanatory variable $X = x$, is $\pi_x / (1 - \pi_x)$.

Odds Ratios

Odds ratios are a way to compare the conditional odds of an event.

Key Definition: Odds Ratio

An odds ratio is the ratio of two conditional odds, given two different explanatory variables.

Let us say we have the odds of something occurring when $X = 1$, which we label as odds_1 . Then, we have another odds of something occurring when $X = 0$, which we label odds_0 . The odds ratio of these two events are:

$$OR = \frac{\text{odds}_1}{\text{odds}_0} = \frac{\pi_1 / (1 - \pi_1)}{\pi_0 / (1 - \pi_0)}$$

The odds ratio is thus a ratio measure of the relationship between the odds of $Y = 1$ and X between two different values of X .

More intuitively, let us say we start with the odds of $Y = 1$ when $X = 0$ (odds_0). The odds ratio is what we would multiply odds_0 by in order to get the odds of $Y = 1$ when $X = 1$ (odds_1):

$$\text{odds}_0 \times \frac{\text{odds}_1}{\text{odds}_0} = \text{odds}_1$$

This tells us that odds ratios are the rate of multiplicative increase when X increases by 1. This allows us to provide an interpretation of logistic regression.

- For example, if the odds ratio is 1.2, we would multiply the odds when $X = 0$ by 1.2 to get the odds when $X = 1$.
- Since multiplying something by 1.2 means increasing it by 20%, that means that when we increase X from 0 to 1, the odds of $Y = 1$ occurring increase by 20%.
- Similarly, if the odds ratio was 0.9, it would mean when we increase X from 0 to 1, the odds of $y = 1$ occurring would decrease by 10%.

Calculating Odds Ratios

How can we actually calculate the odds ratios? Recall the logistic regression in terms of log-odds (but this time, with only one explanatory variable for simplicity):

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \beta X_i$$

We can rewrite $\pi_i/(1-\pi_i)$ as $\log(\text{odds})$. Now, consider the conditional logs of $X = x$ and $X = x + 1$ (an increase in 1 in X)..

$$\begin{aligned}\log(\text{odds}_x) &= \alpha + \beta(X) \\ \log(\text{odds}_{x+1}) &= \alpha + \beta(X + 1)\end{aligned}$$

The change in log-odds would thus be $\log(\text{odds}_{x+1}) - \log(\text{odds}_x)$, which through log properties, we can simplify as well..

$$\begin{aligned}\log(\text{odds}_{x+1}) - \log(\text{odds}_x) &= \alpha + \beta(X + 1) - [\alpha + \beta X] \\ &= \alpha + \beta X + \beta - \alpha - \beta X \\ &= \beta\end{aligned}$$

We also know through property of logs, that:

$$\log(\text{odds}_{x+1}) - \log(\text{odds}_x) = \log\left(\frac{\text{odds}_{x+1}}{\text{odds}_x}\right) = \beta$$

Now we have a log of an odds ratio. We can un-log and isolate the odds ratio by taking both sides to the power of e , which cancels out the logs to get the final calculation of the odds ratio:

$$\begin{aligned}e^{\log\left(\frac{\text{odds}_{x+1}}{\text{odds}_x}\right)} &= e^\beta \\ \frac{\text{odds}_{x+1}}{\text{odds}_x} &= OR = e^\beta\end{aligned}$$

Or simply, after we find the coefficients of our logistic model, we can then calculate e^β to find the odds ratio. Thus, our interpretations are as follows:

i Interpretation of Odds Ratios

We know that odds ratios are a multiplicative increase. Thus, for every one unit increase in X , the odds of $Y = 1$ are multiplied by e^β . Or more intuitively:

- When $e^\beta > 1$, then a one unit increase in X is associated with a $(e^\beta - 1) \times 100$ percent increase in the odds of being in category $Y = 1$.
- When $e^\beta < 1$, then a one unit increase in X is associated with a $(1 - e^\beta) \times 100$ percent increase in the odds of being in category $Y = 1$.
- When $e^\beta = 1$, then a one unit increase in X has no effect on the odds of a unit being in category $Y = 1$.

1.4 Confidence Intervals and Hypothesis Tests

1.5 Likelihood Ratio Test of Models

Implementation in R

Implementation in STATA

Chapter 2

Maximum Likelihood Estimation

2.1 Likelihood Functions

See lecture 9 econometrics II slides 30-31, MY452A 10-11

2.2 Simple Maximum Likelihood Estimator

MY452A page 12-13

2.3 Gradient Descent Estimation Approaches

MY452A pg 14-18

2.4 Properties of the ML Estimator

MY452A Pg 19, Metrics II pg. 37- 38.

2.5 Probit Model

Chapter 3

Multinomial and Ordinal Logistic Regression

Chapter 4

Regression for Counts

Part II

Selection on Observables

Chapter 5

Selection on Observables Design

Chapter 6

Matching, Weighting, and Regression

Chapter 7

Partial Identification and Sensitivity Analysis