**INTRODUCTION**

In the linear regression model, the dependent variable is quantitative. We will be discussing a model wherein the regressand or the dependent variable itself is qualitative in nature.

The fundamental difference between the two is: In a model where Y is quantitative, our objective is to estimate its expected or mean value given the value of the regressors. Whereas when Y is qualitative, our objective is to find the probability of something happening. For example: A student getting admission into IGIDR. Hence qualitative response regression models are also known as **Probability Models**.

Suppose we want to study how variables such as GRE, GPA and prestige of the undergraduate institution affect a student’s admission into a graduate school. The student getting admission is a yes or no decision. Hence the response variable ‘Admit’ can take only two values, say 1 for getting admission and 0 otherwise[[1]](#footnote-2). Therefore, we get a binary choice variable since the regressand can take only two values.

There are 3 ways to develop a probability model for a binary response variable:

1. Linear Probability Model
2. The Logit Model
3. The Probit Model

Now let us first see why we discard the Linear Probability Model. Consider a regression equation,

 Yi = β’ + Ui

Suppose Yi takes only two values with probability and 1- i.e.

 Yi = 1 with probability

 = 0 with probability 1-

Now, E(Yi |xi) = 1.+ 0.(1- =

Therefore, it is the probability that the given event will occur.

We want to model this probability of occurrence as a function of certain characteristics of the individual i.e. we want Pi = F(β’) which is some linear combination of characteristics .

Under Linear Probability Model, F(β’) = β’

The probability distribution of Ui is given as:

|  |  |
| --- | --- |
| Ui  |  |
| -β’ | 1-β’ |
| 1-β’ | β’ |

Here, there is a problem of interpretation since the probability distribution depends on the unknown parameter β and Ui has only a two-point observation and is not normally distributed. Further, heteroskedasticity is present in the model. Also there is no way we can ensure 0 β’Xi 1

So we want a functional form such that:

(1) -β’Xi +

(2) 0 F(β’Xi) 1

where F(β’Xi) = Pi = E(Yi |Xi) = P(Yi = 1|Xi).

The functional form that satisfies these two conditions is a probability distribution function. Mentioned below are the distribution functions of the Logit and the Probit functions:



**What is a latent variable?**

Latent variables are variables that are not directly observed but are rather inferred from other variables that are observed (directly measured). So instead we observe a proxy variable that takes the following form:

 Yi = 0 if Yi\* ≤ 0

 = 1 if Yi\* > 0

In our example, the latent or the unobservable variable is the utility/satisfaction that is derived from getting admission into IGIDR.

**THE PROBIT MODEL**



**THE LOGIT MODEL**







 **ODDS RATIO**

It describes the strength of [association](http://en.wikipedia.org/wiki/Association_%28statistics%29) or non-[independence](http://en.wikipedia.org/wiki/Independence_%28probability_theory%29) between two binary [data](http://en.wikipedia.org/wiki/Data) values. It is used as a descriptive statistic and plays an important role in the [logistic regression](http://en.wikipedia.org/wiki/Logistic_regression).

In our example, if Pi is the probability of getting admission, then 1-Pi is the probability of not getting admission. Both are non-linear in the X’s and the ’s. They can be linearized though, as follows:

1-

Is the Odds Ratio in favor of getting admission

 is the logit (or, log of odds ratio) of the binary variable ‘ADMIT’

Now, the slope coefficient βk gives the change in the logit of Y (admit) for a unit change in GRE, GPA or Prestige of Institution (), which is independent of the value of the regressor. In other words the odds ratio, for a particular variable tells us how the odds of an event change with one unit change in the variable. This ratio is independent of the value of .If βk = 0 then the odds ratio is equal to 1; if βK > 0 then the odds ratio is greater than 1, and if βK < 0 then the odds ratio is less than 1.

**MARGINAL EFFECTS**

In the linear regression model, the slope coefficient measures the change in average values of the dependent variable for a unit change in a particular explanatory variable holding all other variables constant. In the LPM, the slope coefficient measures the change in probability of an event occurring as a result of a one unit change in the value of the explanatory variable. In probit and logit models, the interpretation of the slope is slightly different however.

In the Logit model, shows that the rate of change in probability with respect to an explanatory variable X involves not only but also the level of probability from which the change is measured (but since the Pi is evaluated based on all the variables of the model, even , depends on all the variables of the model). For example, the effect of a unit change in Xi on Pi is greatest when P=0.5 and the least when P=0 or 1. So clearly, the marginal effect of on Pi is not fixed as in the linear regression model.

In the Probit model, suppose we want to find the effect of a unit change in X (say, GPA score) on the probability that Y=1, i.e. that a student gets admission into a graduate school. This needs to be conditioned on a particular value of the GPA score. First the normal density function is ascertained depending on the chosen GPA score. From this value of Z, we find the normal density area and multiply it by the estimated slope coefficient. This gives us the rate of change of the probability with respect to the GPA score.

**ESTIMATION**

**Inconsistency of OLS in Probit and Logit Models**

Rewrite this as,

Now,

Note: This means we cannot use the usual trick of obtaining from OLS estimation and then use it to do a feasible Probit/Logit estimation. This is a general result for all non-linear models:

and,

**Estimation of Logit with WLS**

The logit is represented as

 1-

Which is the odds ratio in favor of possessing a particular attribute such as getting admission into a graduate school. Taking the natural log,

To estimate this, it is necessary to distinguish two types of data:

1. Data at individual level

In this case, OLS estimation becomes infeasible. If a student is admitted into a college,

These expressions are meaningless and thus for such data, we need to use maximum likelihood estimation (MLE) to estimate the parameters.

1. Grouped Data

In this case, there are multiple observations in a particular category. For example, there may be 20 students from a particular undergraduate school with a common GRE score out of which 17 get admission into the graduate programme while the remaining 3 don’t. Thus, we compute

.

The estimated logit then is,

However OLS is still not applicable as the error term is heteroscedastic i.e.,

 each observation with a given GRE class is distributed independently as a binomial variable. Thus, Weighted Least Squares is applicable and we can use

 as an estimator of

1. Estimate for each the probability of getting admission as
2. Obtain the logit as
3. Transform the model to overcome heteroscedasticity as follows,

with

1. Estimate this model with OLS (regression through origin)

**Nonlinear estimation**

Under the estimation of normal distribution of the disturbance term, the OLS estimators are not only BLUE but also BUE i.e., they are the best unbiased estimators in the entire class of estimators, linear or not. But if we drop the assumption of normality then it is possible to obtain nonlinear estimators that perform better than OLS estimators.

Nonlinear least squares estimation involves solving nonlinear normal equations. Usually, an analytical solution with an explicit closed form solution is not possible. An iterative numerical search procedure is used instead and NLLS is done as follows:

The Normal Equations:

Use the Newton-Raphson Method to iterate and obtain the answer:



**Maximum Likelihood Estimation**

In the method of maximum likelihood, the unknown parameters are estimated in such a way that the probability of observing the given Y’s is as high as possible. For this, the function to be optimized is first specified. Then, the function is maximized using calculus techniques. However in practice, the differentiation is easier in log terms. Evoking the assumption of independent samples,

**Probit**

**Logit**

For maximisation,

As with nonlinear estimation, a closed form solution is difficult. Iterative procedures such as Newton Raphson must be used. Simply replace the function Q with For convergence to global maximum we need not worry about starting values, since the log L function is globally concave.



**Choice between Probit and Logit**

In probit, the tails are given less importance than in the logit.[[2]](#footnote-3) Thus, if there is heterogeneity in the data, logit should be chosen. It is for this reason that logit is used more often in empirical studies – it performs well in moderately balanced data as well as data with outliers. The estimates from the Logit and Probit models are related by the approximate rule = 1.6



**EXAMPLE**

The following example attempts to model various factors which influence the admission of a student into a graduate school. The dependent variable – admit/don’t admit – is binary. The explanatory variables used in the model are the GRE (Graduate Record Exam) score, the GPA (Grade Point Average) and the prestige of the undergraduate institution (coded as “rank” in the analysis). The GRE and GPA scores are treated to be continuous variables whereas the variable rank takes values from 1 to 4. Institutions with rank 1 have the highest prestige and those with rank 4 have the lowest prestige.

**THE LOGIT MODEL**

We now run the logistic regression model. To model 1s rather than 0s, we use the **descending** option. We do this because by default, the **proc logistic** command models 0s rather than 1s, in this case that would mean predicting the probability of not getting into graduate school (**admit**=0) versus getting in (**admit**=1). Mathematically, the models are equivalent, but conceptually, it probably makes more sense to model the probability of getting into graduate school versus not getting in. The **class** statement tells SAS that **rank** is a categorical variable (i.e. not continuous). The **param=ref** option after the slash requests dummy coding (1-4) rather than the default effects coding for the levels of **rank**. The name of our dataset is “problog”.

proc logistic data=sasuser.problog descending;

class rank / param=ref ;

model admit = gre gpa rank;

run;

**INTERPRETING THE LOGIT MODEL**

**The LOGISTIC Procedure**

 Analysis of Maximum Likelihood Estimates

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameter | DF | Standard Estimate | Wald Error | Chi-Square | Pr>ChiSq  |
|  |  |  |  |  |  |
| Intercept  | 1 | -5.5414 |  1.1381 |  23.7081 | <.0001 |
| GRE  | 1 | 0.00226 | 0.00109 |  4.2842 | 0.0385 |
| GPA  | 1 | 0.8040 | 0.3318 |  5.8714 |  0.0154 |
| RANK 1  | 1 | 1.5514 | 0.4178 | 13.7870 | 0.0002 |
| RANK 2  | 1 | 0.8760 |  0.3667 | 5.7056 | 0.0169 |
| RANK 3 | 1 |  0.2112 | 0.3929 | 0.2891 | 0.5908 |

* For every one unit change in **GRE**, the log odds of admission (versus non-admission) increases by 0.002.
* For a one unit increase in **GPA**, the log odds of being admitted to graduate school increases by 0.804.
* The coefficients for the categories of rank have a slightly different interpretation because rank is a qualitative variable.

 Odds Ratio Estimates

|  |  |  |  |
| --- | --- | --- | --- |
|  Effect | Point Estimate | 95% Wald Upper Limit | 95% Wald Lower Limit  |
|  |  |  |  |
|  GRE  | 1.002 | 1.000 | 1.004 |
|  GPA  |  2.235 | 1.166 |  4.282 |
|  RANK 1 vs 4  | 4.718 | 2.080 | 10.701 |
|  RANK 2 vs 4  | 2.401 | 1.170 | 4.927 |
|  RANK 3 vs 4  | 1.235 | 0.572 | 2.668 |

 We have taken the rank 4 category as the benchmark category with respect to which we are seeing changes in the log odds ratio. So here, having attended an undergraduate institution with a **rank** of 1, versus an institution with a **rank** of 4, increases the log odds of admission by 1.55.

The table above gives the coefficients as odds ratios. An odds ratio is the exponentiated coefficient, and can be interpreted as the multiplicative change in the odds for a one unit change in the predictor variable. For example, for a one unit increase in **GPA**, the odds of being admitted to graduate school (versus not being admitted) increase by a factor of 2.24.

 **THE PROBIT MODEL**

Below we run the probit regression model using **proc logistic**. The **link=probit** option fits a probit model rather than the default logit model. In this case again, SAS will model the event of getting admission i.e. **admit** takes on the value of 1. (If we omitted the **descending** option, SAS would model **admit** being 0 and our coefficients would be reversed.)

proc logistic data=sasuser.problog descending;

class rank / param=ref ;

model admit = gre gpa rank /link=probit;

run;

**Interpreting the Probit Model**

 Analysis of Maximum Likelihood Estimates

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameter | DF | Standard Estimate | Wald Error | Chi-Square | Pr >ChiSq |
|  |  |  |  |  |  |
| Intercept | 1 | -3.3225 | 0.6633 | 25.0872 | <.0001 |
| GRE | 1 | 0.00138 | 0.000650 | 4.4767 | 0.0344 |
| GPA  | 1 | 0.4777 | 0.1972 | 5.8685 | 0.0154 |
| RANK 1 | 1 | 0.9359 | 0.2453 | 14.5606 | 0.0001 |
| RANK 2 | 1 | 0.5205 | 0.2109 | 6.0904 | 0.0136 |
| RANK 3 | 1 | 0.1237 | 0.2240 | 0.3053 | 0.5806 |

The probit regression coefficients give the change in the probit index, also called a z-score, for a one unit increase in the predictor variable.

* For every one unit change in **GRE**, the z-score increases by 0.001.
* For a one unit increase in **GPA**, the z-score increases by 0.478.
* The coefficients for the categories of rank have a slightly different interpretation. We have taken the rank 4 category as the category with respect to which we are seeing changes in the Z score. So here, having attended an undergraduate institution with a **rank** of 1, versus an institution with a **rank** of 4, increases the z-score by 0.936.

# Bibliography

* Goldberger, A. S. (1991). *A Course in Econometrics* (First ed.). London, England: Harvard University Press.
* Gujarati, D. N., & Sangeetha. (2007). *Basic Econometrics* (4th Edition ed.). New Delhi: Tata McGraw-Hill Publishing Company Limited.
* Sarkar, S. (2011). Jadavpur University Workshop Lecture Notes., (p. 53). Kolkata.
* UCLA: Statistical Consulting Group. (n.d.). *SAS Topics: Logistic (and Categorical) Regression.* Retrieved April 17, 2013, from Introduction to SAS: http://www.ats.ucla.edu/stat/sas/topics/logistic.htm
1. This can also be modeled differently: we can model “rejection from a graduate school” and let the value 1 represent not getting selected and let the value 0 represent the event of getting selected. However, in this report we will model “admission”. [↑](#footnote-ref-2)
2. This is because in probit, approximately 68% of the area under the normal curve lies between the value of $μ\pm σ$, about 95% of the area lies between $μ\pm 2σ$ and about 99.7% of the area lies between $μ\pm 3σ$. So by the time the tails are reached, there is no probability mass left to assign to the extreme values. [↑](#footnote-ref-3)