**THE LOGIT MODEL**

Both the logit and probit model approaches are able to overcome the limitation of the LPM that it can produce estimated probabilities that are negative or greater than one. They do this by using a function that effectively transforms the regression model so that the fitted values are bounded within the (0,1) interval. Visually, the fitted regression model will appear as an **S-shape** rather than a straight line, as was the case for the LPM. This is shown in figure 1.

Probability of *y*

1 …………………………………….

0 *x*

Figure 1: The logit model

The logistic function *F*, which is a function of any random variable, *z*, would be:

 (1)

where *e* is the exponential under the logit approach. The model is so called because the function *F* is in fact the cumulative logistic distribution. So the logistic model estimated would be

 (2)

where again is the probability that = 1.

With the logistic model, 0 and 1 are asymptotes to the function and thus the probabilities will never actually fall to exactly zero or rise to one, although they may come infinitesimally close. In equation (1), as tends to infinity, tends to zero and tends to 1; as tends to minus infinity, tends to infinity and  tends to 0.

Clearly, this model is not linear (and cannot be made linear by a transformation) and thus is not estimable using OLS. Instead, maximum likelihood is usually used.

**The Probit Model**

Instead of using the cumulative logistic function to transform the model, the cumulative normal distribution is sometimes used instead. This gives rise to the probit model. The function *F* in equation (1) is replaced by:

 (3)

This function is the cumulative distribution function for a standard normally distributed random variable. As for the logistic approach, this function provides a transformation to ensure that the fitted probabilities will lie between zero and one. Also as for the logit model, the marginal impact of a unit change in an explanatory variable,  say, will be given by , where is the parameter attached to and 

**Choosing Between the Logit and Probit Models**

For the majority of the applications, the logit and probit models will give very similar characterisations of the data because the densities are very similar. That is, the fitted regression plots (such as figure 1) will be virtually indistinguishable and the implied relationships between the explanatory variables and the probability that  will also be very similar. Both approaches are much preferred to the linear probability model. The only instance where the models may give non-negligibility different results occurs when the split of the between 0 and 1 is very unbalanced for example, when occurs only 10% of the time.

Stock and Watson (2006) suggest that the logistic approach was traditionally preferred since the function does not require the evaluation of an integral and thus the model parameters could be estimated faster. However, this argument is no longer relevant given the computational speeds

now achievable and the choice of one specification rather than the other is now usually arbitrary.

**Estimation of Limited Dependent Variable Models**

Given that both logit and probit are non-linear models, they cannot be estimated by OLS. While the parameters could, in principle, be estimated using non-linear least squares (NLS), maximum likelihood (ML) is simpler and is invariably used in practice. The principle is that the parameters are chosen to jointly maximise a log-likelihood function (LLF). The form of this LLF will depend upon whether the logit or probit model is used, but the general principles for parameter estimation will still apply. That is, we form the appropriate log-likelihood function and then the software package will find the values of the parameters that jointly maximise it using an iterative search procedure.

Once the model parameters have been estimated, standard errors can be calculated and hypothesis tests conducted. While *t*-test statistics are constructed in the usual way, the standard error formulae used following the ML estimation are valid asymptotically only. Consequently, it is common to use the critical values from a normal distribution rather than a *t* distribution with the implicit assumption that the sample size is sufficiently large.

**Parameter Interpretation for Logit and Probit Models**

Standard errors and *t*-ratios will automatically be calculated by the statistical or econometric software package used, and hypothesis tests can be conducted in the usual fashion. However, interpretation of the coefficients needs slight care. It is tempting, but incorrect, to state that a 1-unit increase in , for example, causes a increase in the probability that the outcome corresponding to will be realised. This would have been the correct interpretation for the linear probability model.

However, for logit models, this interpretation would be incorrect because the form of the function is not , for example, but rather , where *F* represents the (non-linear) logistic function. To obtain the required relationship between changes in and , we would need to differentiate *F* with respect to and it turns out that this derivative is . So in fact, a 1-unit increase in will cause a increase in probability. Usually, these impacts of incremental changes in an explanatory variable are evaluated by setting each of them to their mean values. For example, suppose we have estimated the following logit model with 3 explanatory variables using maximum likelihood



Thus we have  , , . We now need to calculate , for which we need the means of the explanatory variables, where is defined as before. Suppose that these are  , , , then the estimate of *F*  will be given by





Thus a 1-unit increase in  will cause an increase in the probability that the outcome corresponding to  will occur by 0.4 × 0.679 = 0.272. The corresponding changes in probability for variables  and are −0.7 × 0.679 = −0.475 and 0.8 × 0.679 = 0.543, respectively. These estimates are sometimes known as the *marginal effects*.

There is also another way of interpreting discrete choice models, known as the random utility model. The idea is that we can view the value of *y* that is chosen by individual *i* (either 0 or 1) as giving that person a particular level of utility, and the choice that is made will obviously be the one that generates the highest level of utility. This interpretation is particularly useful in the situation where the person faces a choice between more than 2 possibilities.

**CHI-SQUARE TEST OF GOODNESS OF FIT AND INDEPENDENCE**

The *X2* (chi-square) distribution is used to test whether (1) the observed frequencies differ "significantly" from expected frequencies when more than two outcomes are possible; (2) two variables are independent and (3) the sampled distribution is binomial, normal, or other. The *X2* statistic calculated from the sample data is given by



where denotes the frequencies and I,, the expected frequencies. If the calculated is greater than the tabular value of at the specified level of significance and degrees of freedom, the null hypothesis Ho is rejected in favor of the alternative hypothesis .

The degrees of freedom for tests of goodness of fit (1 and 3) are given by

d f = c - m - 1 (2)

where c represents the categories and m, the number of population parameters estimated from sample statistics.

The degrees of freedom for tests of independence, or contingency-table tests (2), are given by

 (3)

where r indicates the number of rows of the contingency table and c, the number of columns.

The expected frequency for each cell of a contingency table is



Where *,* and indicate sum over row and column, respectively, of the observed cell and n represents the overall sample size.

**Example**

In the past, 30% of the TVs sold by a store were small-screen, 40% were medium, and 30% were

large. In order to determine the inventory to maintain of each type of TV set, the manager takes a random sample of 100 recent purchases and finds that 20 were small-screen, 40 were medium, and 40 were large. At test *5%* level of significance test the hypothesis that the past pattern of sales *Ho* still prevails.

**Solution**

**Observed and Expected Frequency**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ***Screen Size*** | | | ***Total*** |
| ***Small*** | ***Medium*** | ***Large*** |
| ***Observed Pattern*** | ***20*** | ***40*** | ***40*** | ***100*** |
| ***Past pattern*** | ***30*** | ***40*** | ***30*** | ***100*** |



Find the chi-square critical. **.** Because no population parameter was estimated, *m* = 0. Therefore the degree of freedom is given by**,** df = 3-0-1= 2 . Hence the critical is 5.99. Since the calculated value of  = 5.83 is smaller than the tabular value of

= 5.99 with = 0.05 and df = 2 (see App. 6), we cannot reject *Ho,* that the past sales pattern still prevails.

**Example 2**

A car dealer has collected the data shown in Table 1 on the number of foreign and domestic cars purchased by customers under 30 years old and 30 and above.

**Table 1:Table for Car Buyers**

|  |  |  |  |
| --- | --- | --- | --- |
| **Age** | **Type of Car** | | **Total** |
| **Foreign** | **Domestic** |
| < 30 | 30 | 40 | 70 |
|  | 20 | 80 | 100 |
| Total | 50 | 120 | 170 |

Test at the 1% level of significance if the type of car bought (foreign or domestic) is independent of the age of the buyer.

**Solution**

Constructs a table of expected frequencies (Table 2). For the first cell in row 1 and column 1, we obtain



**Table 2: Table of Expected Frequencies for the Observed Frequencies**

|  |  |  |  |
| --- | --- | --- | --- |
| **Age** | Type of Car | | Total |
| Foreign | Domestic |
|  | 21  29 | 49  71 | 70  100 |
| Total | 50 | 120 | 170 |



Chi-Square Critical: At and 1% level of significance is 6.63

Decision: We reject *Ho,* that age is not a factor in the type of car bought (and conclude that younger people seem more likely to buy foreigncars).

**Non-Parametric Tests**

So far we have assumed our sample is large or come from normally distributed populations. However, it may be inappropriate to treat the populations to be always normal. There are situations in which the use of normal curve may not be appropriate. In such situations techniques that do not make restrictive assumptions about the shape of the population are to be used. These techniques are classified as distribution free or more commonly, non-parametric tests. Such test include:

**Sign Tests**

This test is based on the direction or sign (plus or minus) of a pair of observations and does not depend upon the numerical magnitude of the observations. The null hypothesis for sign test is that there is no real difference between the two pairs of data. We assign positive sign if ***x* >*y* ,** zero for *x = y* and negative sign for *x<y*. Thus, *x* can be rating for brand *A* and *y* can be rating for brand *B* of an item. If a person gives higher score to *A,* we assign positive sign to *x-y,* if same score is given we assign 0, and if higher score is given to *B*, we assign negative sign. As we are testing the differences, we exclude tie evaluation (i.e. zeros). Now, we count the number of positive signs and find out the proportion of the signs (*p*). The null hypothesis is that this proportion should be 0.5, i.e., *H0: p=0.5* and alternative hypothesis Ha *p*. We now use the binomial distribution for the construction of the rejection region. For large values of *n*, using the normal distribution as an approximation to the binomial, the standard deviation is given by:



So

the normal variate is 

**Example**

A market researcher was interested in testing the preference of consumers for a brand of detergent soap. Consumers were asked to rate the brands on cleanliness it provides, by giving them numbers from 1 to 4. The data of consumers’ response and the corresponding signs is given below:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Consumer No. Score for brand *X* Score for brand *Y* Sign *(X – Y)*

1 4 2 +

2 2 3 -

3 1 2 -

4 4 3 +

5 3 2 +

6 3 4 -

7 4 2 +

8 2 1 +

9 1 1 0

10 4 3 +

11 1 2 -

12 1 2 -

13 2 2 0

14 2 3 -

15 4 3 +

16 3 2 +

17 3 3 0

18 4 2 +

19 4 1 +

20 4 2 +

21 4 4 0

22 3 2 +

23 3 3 0

24 3 1 +

25 4 1 +

26 4 3 +

27 4 1 +

28 1 2 -

29 3 2 +

30 3 4 -

31 2 3 -

32 2 1 +

33 2 3 -

34 3 4 -

35 3 2 +

Will you conclude that the effectiveness of the two brands is the same?

**Solution**

The total number of respondents are 35. We find the positive signs are 19 and the negative are 11 and the numbers of zeros are 5. We exclude the ties this gives us a total 30 useable responses. The hypothesis test to be tested 

The sample information gives us the sample proportion



And therefore, the test statistic is given by



At 5% level of significance the acceptance region is . Since the calculated value falls in the acceptance region, the hypothesis is accepted i.e. the consumers perceive no difference between the two brands of the detergent soaps.