ECON 4130, Statistics 2
Tore Schweder, 17 November 2003

Lecture note on logistic regression

Chapter 18 in R. Carter Hill, William E. Griffiths, George G. Judge: *Undergraduate Econometrics* (second edition). John Wiley & Sons, Inc. gives a brief introduction to regression analysis with qualitative response. In the case of binary response, probit regression and logistic regression are considered. These two methods are very similar, and outside economics, logistic regression is by far the most popular method.

To briefly explain logistic regression, consider the data in Table 18.1 of Hill et al. For 21 independent respondents were faced with the choice of either taking bus or car for a given trip. The response is denoted y. It is recorded how long time (minutes) each trip would take by either car or bus. This is assumed known to the respondent before he or she made the choice. The data are as follows.

Auto Time	Bus Time	*y*
52,90	4,4	0
4,10	28,5	0
4,10	86,9	1
56,20	31,6	0
51,80	20,2	0
0,20	91,2	1
27,60	79,7	1
89,90	2,2	0
41,50	24,5	0
95,00	43,5	0
99,10	8,4	0
18,50	84	1
82,00	38	1
8,60	1,6	0
22,50	74,1	1
51,40	83,8	1
81,00	19,2	0
51,00	85	1
62,20	90,1	1
95,10	22,2	0
41,60	91,5	1

The response variable *y* is binary – it can only take one of two possible values, 0: the respondent did not take car, or 1: the respondent took car rather than bus.

The question asked is what the probability, *p*, is of choosing automobile transportation given the times it would take by either car or bus.

Assume first that the choice probability only depends on the difference

x = Auto Time – Bus Time

The linear logistic regression model is

$$p = P(Y = 1) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

Since the respondents are assumed independent, the likelihood function is obtained by multiplying the likelihood component generated by each respondent. The first respondent took bus. This event has probability $1 - p$ for the value of *p* determined by the corresponding $x = -48.5$, and the two parameters. Since $y = 0$, we can write $1 - p = p^y (1-p)^{1-y}$. This formula does also work for respondents with $y = 1$, and we have the likelihood function written as the following product.

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} p_i^{y_i} \left(1 - p_i\right)^{1-y_i}$$

Here, $p_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$ is regardede as a function of the two regression parameters (the intercept β_0, and the regression parameter for *x*, β_1).

Stata, or any other statistical system, is capable of maximizing the likelihood function to obtain maximum likelihood estimates. In the same run, the information matrix is calculated, and hence the standard errors of the maximum likelihood estimates. The result is as follows

```
               Value Std. Error t value
(Intercept) -0.23757     0.7496 -0.3169
          x  0.05311     0.0206  2.5776

    Null Deviance(-2log(Likelihood)): 29.06 on 20 degrees of freedom

Residual Deviance(-2log(Likelihood)): 12.33 on 19 degrees of freedom
```

The null model is $H_0 : \beta_1 = 0$, which would be the case if people are insensitive to the value of *x* in their choice of bus versus auto. We obtain a difference in -2log(likelihood) of 16.73, which compared to the chi-square distribution with one df gives a p-value of 0.00004.

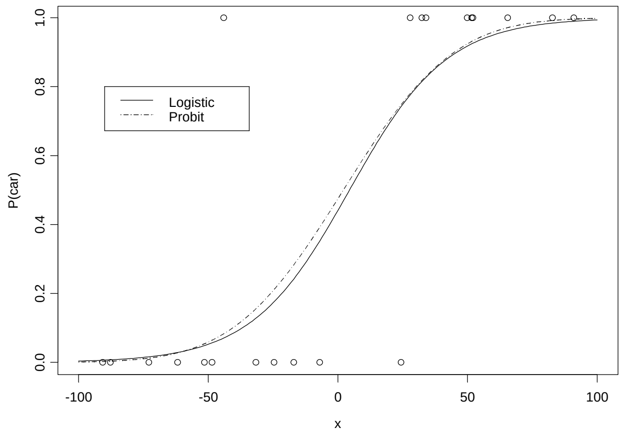
Hill et al. carries out a probit regression. The result is (with the intercept estimate slightly different from what Hill et al. found),

```
               Value Std. Error t value
(Intercept) -0.06462    0.40034 -0.1614
          x  0.02999    0.01027  2.9205

    Null Deviance: 29.06 on 20 degrees of freedom

Residual Deviance: 12.33 on 19 degrees of freedom.
```

We find the residual deviance to be exactly the same in the probit model as in the logistic model. The two models fit the data equally well. Note that the estimates are different. This is the case since the link functions are different. When comparing the two fitted models, we find, however, that the fitted probability curves are very similar. In the following graph, the fitted probability curves are given for both models, together with the observed data points given as small circles. Note that the probit model has the form

$$p = F(\beta_0 + \beta_1 x) = P(Z \le \beta_0 + \beta_1 x)$$

where $Z \sim N(0,1)$.



1.0
0.8
0.6
0.4
0.2
0.0
P(car)
Logistic
Probit
-100
-50
0
50
100
x