Quant II - Problem Set VI
Models for Binary Data
Kenneth Benoit

Assigned: Wednesday, February 25, 2009
Due: Wednesday, March 4, 2009

1. Using R, create a table of probabilities from 0 to 1 (in steps of .10) with the associated odds, log-odds/logits as well as quantiles of the cumulative normal distribution. Explain in one sentence why the latter two transformations are important in the analysis of binary data.

2. You have a dataset with five observations as follows:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>x</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>0</td>
<td>8</td>
</tr>
</tbody>
</table>

(a) Manually calculate the likelihood for \( Y \mid \pi = .2 \).

(b) Manually calculate the log-likelihood of a logit model with

\[
Pr(Y = 1) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}},
\]

where \( \beta_0 = -8 \) and \( \beta_1 = 1 \).

3. Load the Titanic dataset (STATA-format), which provides actual data on the survival of passengers from the 1912 disaster.

(a) Run a logit model regressing \textit{survived} on \textit{age}, \textit{sex}, and \textit{class}. Showing command and output is enough at this stage.

(b) Is there evidence for a “women and children first”-policy? How did class affect survival? Note: For this sub-task, it is enough to interpret the signs of the coefficients.
(c) Briefly interpret all findings with the help of odds ratios (neglect the issue of inference/significance tests). Try to make precise numerical statements.

(d) For all fans of the 1997 film: What is the odds ratio of survival of a woman travelling in first class and a man travelling in third class? (Note: This is one odds ratio, the ratio of the two odds.) What are the two associated expected probabilities of survival?

4. Load the STATA dataset economic_bills.dta. This is a sample of legislative bills dealing with economic matters, which were initiated by governments in Belgium, France, Germany, and the UK during the 1990s. Variables are status (equals 1 if bill was passed), xland (the country), cabinet (0 = initiated from among the government parties on the floor, 1 = initiated by the cabinet), and vpdi_LH92economic (the ideological/veto player distance between the most rightist and most leftist government party on the 1-20 taxes vs. spending scale from the expert survey by Laver and Hunt (1992)).

(a) Run a logit model regressing status on the other variables. Briefly interpret the findings with the help of odds ratios. Try to make precise numerical statements.

(b) What is the expected probability that a bill initiated by the government parties in the UK is passed if the veto player distance equals 0?

(c) What is the expected probability that a French bill initiated by the cabinet is passed if the veto player distance equals 5?

(d) Plot the expected outcome of a German bill initiated by the government parties against ideological distance. Use the sample range as range for the distance variable. Make sure the range of the y-axis is from 0 to 1. Why does the plotted curve not look S-shaped?

(e) Calculate “manually” the share of outcomes correctly predicted by the above model (using \( Pr = .5 \) as cut-off). Compare the result to the success rate obtained when always predicting the sample mode of status.