Intermediate Social Statistics

Lecture 4: Multinomial Discrete choice

Raymond Duch

Nuffield College Oxford University

www.raymondduch.com

1 Measures of Fit

A likelihood ratio index:

$$LRI = 1 - \frac{lnL}{lnL_0} \tag{1}$$

where L_0 is the log-likelihood computed only with a constant, and ln L is the maximized log-likelihood.

2 Probit

What we are modeling here is actually a latent quantity that gives rise to the observed discrete outcomes. We can think of this underlying latent quantity as a "probability" or a "random utility". We model the unobserved net utilities y^* of the choices y via the model

$$y^* = \mathbf{x_i}\beta + \varepsilon_i \tag{2}$$

With either iid N(0,1) for probit. We don't observe the net utility of the choice , just whether it was made or not. So we observe whether a person voted for Labour or for the Opposition; whether an individual made a campaign contribution or did not; whether a country went to war or did not; whether someone died or did not. We posit that

$$\begin{aligned} y_i &= 0 \text{ if } y_i^* \leq 0 \\ &= 1 \text{ if } y_i^* > 0 \end{aligned}$$

Some simple algebra gives us our estimator.

$$\begin{aligned} Prob(y_i = 1|\mathbf{x_i}) &= Prob(y_i^* > 0) \\ &= Prob(\mathbf{x_i}\beta + \varepsilon_i) > 0 \\ &= Prob(\varepsilon_i > -\mathbf{x_i}\beta) \\ &= Prob(\frac{\varepsilon_i}{\sigma} > \frac{-\mathbf{x_i}\beta}{\sigma}) \end{aligned} \tag{3}$$

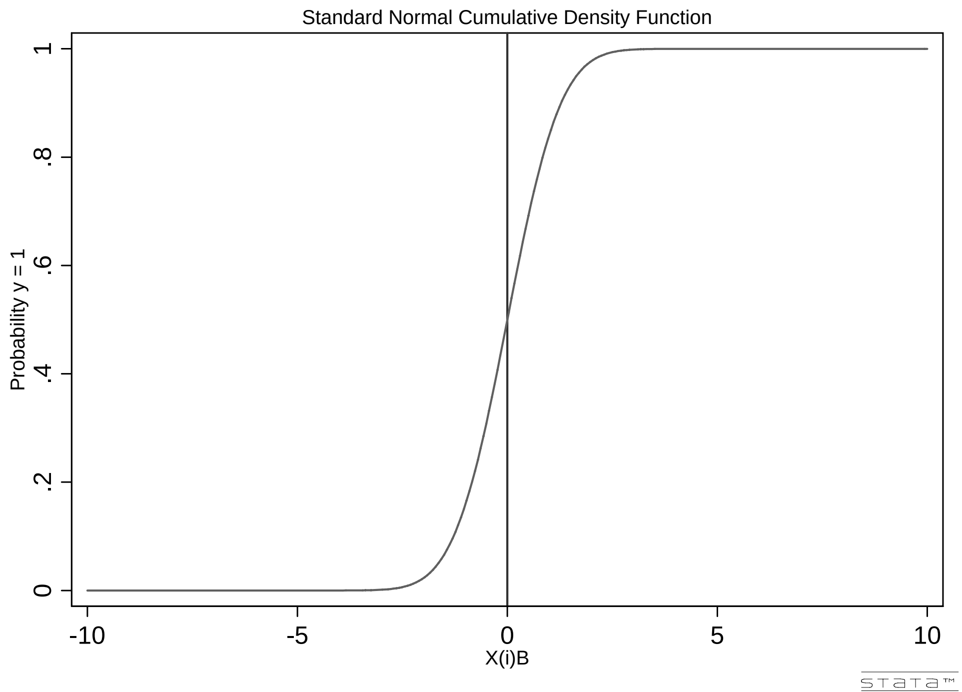
Because the error has a normal distribution this becomes

$$Prob(y_i = 1|\mathbf{x_i}) = 1 - \Phi(\frac{-\mathbf{x_i}\beta}{\sigma}) = \Phi(\frac{\mathbf{x_i}\beta}{\sigma}) \tag{4}$$

Similarly,

$$Prob(y_i = 0|\mathbf{x_i}) = 1 - \Phi(\frac{\mathbf{x_i}\beta}{\sigma}) \tag{5}$$

This is simply the standard normal cumulative density function and as you can see it closely resembles the logistic CDF that we examined last week.


Standard Normal Cumulative Density Function
Probability y = 1
0
.2
.4
.6
.8
1
-10
-5
0
5
10
X(i)B

2.1 Probit Estimation

Maximum likelihood provides a convenient and powerful method for estimating the parameters of this probit model. A key assumption is that the data are identically and independently distributed, which allows us to form a likelihood function for the whole data from the product of the likelihoods for each observation:

$$\begin{aligned}(y_1, y_2, ..., y_n) &= P(y_1)P(y_2)...P(y_n) \\ &= \prod_{y_{i=1}} \Phi(\mathbf{x_i}\beta) \prod_{y_{i=0}} [1 - \Phi(\mathbf{x_i}\beta)]\end{aligned} \tag{6}$$

In Likelihood notation:

$$L = \prod_{y_{i=1}}^{N} \Phi(\mathbf{x_i}\beta)^{y_i} \prod_{y_{i=0}}^{N} [1 - \Phi(\mathbf{x_i}\beta)]^{1-y_i} \tag{7}$$

Each observation thus contributes something to the likelihood, either in the first part when $y_i = 1$, or in the second part when $y_i = 0$ (so 1 - $y_i = 1$). As is typical with MLE, it is easier to work with the log-likelihood:

$$lnL = \sum_{y_{i=1}}^{N} y_i ln\Phi(\mathbf{x_i}\beta) + \sum_{y_{i=0}}^{N} (1 - y_i) ln[1 - \Phi(\mathbf{x_i}\beta)] \tag{8}$$

The only unknowns here are is the vector of β

Can use calculus to optimize to solve for the unknown vector of

But there is no simple analytic solution so this is typically accomplished iteratively. Lets do a couple of iterations by hand

2.2 Probit Estimation Example

Lets return to our example from the UK 2004 Election study.

In this example the dependent variable is Labour incumbent vote and it takes on a value of 1 or 0. The independent variable, is income category that ranges in value from 10 (10,000) to 100 (100,000 or greater).

$$lnL = \sum_{y_{i=1}}^{N} y_i ln\Phi(\alpha + \beta_1 * (Income)) + \sum_{y_{i=0}}^{N} (1 - y_i) ln[1 - \Phi(\alpha + \beta_1 * (Income))] \tag{9}$$

The two unknowns in this likelihood function are α and β_1. For this example, lets pick as our starting values α=1.2 and β_1=-.02. And lets calculate the likelihood for a respondent with an income of 10 (10,000 GBP) and who expressed a Labour preference (vote=1).

$$lnL = 1ln\Phi(1.2 - .02 * (Income))$$
$$= -.173$$

And for the person who expressed a preference for an opposition candidate and who had an income of 50 (50,000 GBP)...

$$lnL = (1 - 0)ln[1 - \Phi(1.2 - .02 * (Income))]$$
$$= -.866$$

	vote	income	alpha	beta	logindi~1	loglike	alpha2	beta2	logindi~2	loglike2
1.	1	10	1.2	-.02	-.1727538	-4.882811	2.38	-.052	-.0319477	-3.484201
2.	1	20	1.2	-.02	-.2380737	-4.882811	2.38	-.052	-.0944455	-3.484201
3.	1	30	1.2	-.02	-.3205539	-4.882811	2.38	-.052	-.2308079	-3.484201
4.	0	40	1.2	-.02	-1.065434	-4.882811	2.38	-.052	-.9621029	-3.484201
5.	0	50	1.2	-.02	-.8657396	-4.882811	2.38	-.052	-.5326208	-3.484201
6.	1	60	1.2	-.02	-.6931472	-4.882811	2.38	-.052	-1.471199	-3.484201
7.	0	70	1.2	-.02	-.5460044	-4.882811	2.38	-.052	-.1096304	-3.484201
8.	0	80	1.2	-.02	-.4224764	-4.882811	2.38	-.052	-.0382607	-3.484201
9.	0	90	1.2	-.02	-.320554	-4.882811	2.38	-.052	-.010782	-3.484201
10.	0	100	1.2	-.02	-.2380737	-4.882811	2.38	-.052	-.0024041	-3.484201

As it turns out α=2.38 and β_1=-.052 maximize the log likelihood.

```
. probit vote income

Iteration 0:   log likelihood = -6.7301167
Iteration 1:   log likelihood =  -3.846122
Iteration 2:   log likelihood = -3.5128548
Iteration 3:   log likelihood = -3.4841193
Iteration 4:   log likelihood = -3.4837079
Iteration 5:   log likelihood = -3.4837078

Probit regression                                 Number of obs   =         10
                                                  LR chi2(1)      =       6.49
                                                  Prob > chi2     =     0.0108
Log likelihood = -3.4837078                       Pseudo R2       =     0.4824

------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |  -.0524063   .0292291    -1.79   0.073    -.1096943    .0048817
       _cons |   2.384189   1.496149     1.59   0.111    -.5482103    5.316588
------------------------------------------------------------------------------
```

2.3 Interpreting the Probit Coefficients

Here are the results of a probit model with Labour as the dependent variable and the same set of independent variables we have used in previous examples...

```
. probit  incumvote retnat class union southwest urban lrself own

Iteration 0:   log likelihood = -507.77976
Iteration 1:   log likelihood = -444.89944
Iteration 2:   log likelihood = -443.61601
Iteration 3:   log likelihood = -443.61387

Probit regression                                 Number of obs   =        785
                                                  LR chi2(7)      =     128.33
                                                  Prob > chi2     =     0.0000
Log likelihood = -443.61387                       Pseudo R2       =     0.1264

------------------------------------------------------------------------------
   incumvote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      retnat |  -.4202345   .0610461    -6.88   0.000    -.5398826   -.3005864
       class |   -.232117   .0510478    -4.55   0.000    -.3321689   -.1320651
       union |   .2976013   .1156959     2.57   0.010     .0708414    .5243611
   southwest |   -.510433   .1970207    -2.59   0.010    -.8965865   -.1242795
       urban |    .189083   .0632399     2.99   0.003      .065135     .313031
      lrself |  -.0795816   .0216375    -3.68   0.000    -.1219903   -.0371729
         own |  -.3172383   .1135865    -2.79   0.005    -.5398637   -.0946129
       _cons |    1.16181   .2511572     4.63   0.000     .6695505    1.654069
------------------------------------------------------------------------------
```

So if we're going to find the marginal effects of a change in some x on the probability of y = 1 (for example, the marginal effect of changing left-right self identification on Labour vote probabilities), we have to note that the parameters of the model, β, are not the marginal effect of x on y. In general:

$$\frac{\partial E[y|\mathbf{x}]}{\partial \mathbf{x_k}} = \frac{d\Phi(\mathbf{x}\beta)}{d(\mathbf{x_k})} = \phi(\mathbf{x}\beta)\beta_k \tag{10}$$

where ϕ denotes the standard Normal Probability Density Function (PDF).

And it's important to note that the values of the marginal effects (which is what the partials of y wrt x are) will change with x. That's what makes them nonlinear. (Think of a line. The marginal effect of x on y does not change anywhere over the course of that line. It is always β. Not here.)

One of the most meaningful strategies for calculating the impact of x on y is to evaluate the impact of incrementing x_k by δ is to increment x_k by δ for each observation in the data and then calculate the average change in y.

$$\begin{aligned} &\frac{\Delta Pr(y_i = 1|\mathbf{x_i})}{\Delta x_{i,k}} \\ &\quad = Pr(y_i = 1|\mathbf{x_i}, x_{i,k} + \delta) - Pr(y_i = 1|\mathbf{x_i}, x_{i,k}) \end{aligned} \tag{11}$$

For each individual, i, a change in the variable x_k to $x_k + \delta$, the predicted probability of an event changes by $\frac{\Delta Pr(y_i=1|\mathbf{x_i})}{\Delta x_{i,k}}$, holding all other variables at their actual values for that individual i.

The impact of this change evaluated over the entire sample of respondents is simply the mean $\Delta Pr(y_i = 1|\mathbf{x_i})$

Lets work through an example from the probit model of Labour vote preference. Here we want to determine the impact on the probability of voting Labour of a unit change in the economic evaluation variable (retnat), i.e., where $x_k = retnat$

$$\begin{aligned}\frac{\Delta Pr(y_i = 1|\mathbf{x_i})}{\Delta x_{i,k}} &= \\ \mathrm{Pr}(Labour_i &= 1|\mathbf{x_i}, retnat_i + 1) - Pr(Labour_i = 1|\mathbf{x_i}, retnat_i)\end{aligned} \tag{12}$$

For each individual in data set the estimate of the $Pr(Labour_i = 1|\mathbf{x_i}, retnat_i)$ is simply

$$\begin{aligned}Pr(Labour_i &= 1|\mathbf{x_i}, retnat_i) \\ &= 1 - \Phi(\alpha + \beta_1 * (retnat) + \beta_2 * (class) + \beta_3 * (union) \\ &+ \beta_4 * (southwest) + \beta_5 * (urban) + \beta_6 * (lrself) + \beta_7 * (own)) \\ &= 1 - \Phi(1.16 - 0.42 * (retnat) - 0.23 * (class) + 0.297 * (union) \\ &- 0.51 * (southwest) + 0.189 * (urban) - 0.079 * (lrself) - 0.317 * (own))\end{aligned} \tag{13}$$

And for each individual the estimate of the $Pr(Labour_i = 1|\mathbf{x_i}, retnat_i + 1)$ is simply

$$\begin{aligned}Pr(Labour_i &= 1|\mathbf{x_i}, retnat_i + 1) \\ &= 1 - \Phi(\alpha + \beta_1 * (retnat) + \beta_2 * (class) + \beta_3 * (union) + \beta_4 * (southwest) \\ &+ \beta_5 * (urban) + \beta_6 * (lrself) + \beta_7 * (own)) \\ &= 1 - \Phi(1.16 - 0.42 * (retnat) - 0.23 * (class) + 0.297 * (union) - 0.51 * (southwest) \\ &+ 0.189 * (urban) - 0.079 * (lrself) - 0.317 * (own))\end{aligned} \tag{14}$$

```
. list retnat class union southwest urban lrself own incumvote XB_1 yhat_1 XB_2 yhat_2 delta in 40/60
```

	retnat	class	union	southw~t	urban	lrself	own	incumvote	XB_1	yhat_1	XB_2	yhat_2	delta
40.	worse	middle	non-union	0	rural	5	own	Opposition	-1.321308	.0931993	-1.321308	.0931993	0
41.	worse	middle	union	0	large town	4	rent	Labour	-.2487211	.4017883	-.2487211	.4017883	0
42.	worse	lower middle	non-union	0	small-medium town	4	own	Labour	-.8205267	.205958	-.8205267	.205958	0
43.	worse	lower middle	union	1	large town	5	own	Opposition	-.923857	.1777804	-.923857	.1777804	0
44.	better	lower middle	non-union	1	rural	left	own	Opposition	-.4408289	.3296684	-.8610634	.1946016	-.1350669
45.	same	working	non-union	0	large town	5	own	Opposition	-.0586738	.476606	-.4789082	.316002	-.160604
46.	same	lower middle	non-union	0	small-medium town	5	own	Opposition	-.4798737	.3156586	-.9001082	.1840313	-.1316273
47.	same	middle	non-union	0	large town	7	own	Opposition	-.6820709	.2475971	-1.102305	.1351645	-.1124326
48.	same	upper middle	non-union	0	rural	6	own	Opposition	-1.212772	.1126084	-1.633007	.0512338	-.0613747
49.	worse	middle	non-union	0	small-medium town	7	own	Opposition	-1.291388	.0982845	-1.291388	.0982845	0
50.	same	working	union	0	large town	6	own	.	.1593459	.5633019	-.2608886	.3970892	-.1662126
51.	worse	middle	union	0	small-medium town	6	own	Opposition	-.9142056	.1803044	-.9142056	.1803044	0
52.	worse	working	non-union	0	small-medium town	3	own	Labour	-.508828	.3054364	-.508828	.3054364	0
53.	same	.	non-union	0	small-medium town	5	own	.	.	.	.	.	.
54.	better	middle	union	0	rural	4	own	Labour	-.1036564	.458721	-.5238909	.3001772	-.1585438
55.	better	middle	non-union	0	small-medium town	5	own	Opposition	-.2917562	.3852365	-.7119907	.2382353	-.1470013
56.	worse	working	non-union	1	small-medium town	4	own	.	-1.098843	.1359183	-1.098843	.1359183	0
57.	same	middle	union	0	large town	5	own	Opposition	-.2253065	.4108704	-.645541	.2592883	-.1515821
58.	better	middle	non-union	0	large town	3	own	Labour	.0564899	.5225242	-.3637446	.3580244	-.1644999
59.	better	middle	non-union	0	small-medium town	left	own	Opposition	.0265701	.5105987	-.3936644	.3469144	-.1636842
60.	better	middle	non-union	0	small-medium town	6	own	Opposition	-.3713378	.355193	-.7915723	.214305	-.1408879

3 Ordered probit

Consider the 4-point democratic satisfaction measure from the 2004 UK Election study. If we use this as a dependent variable and estimate using OLS, all differences will be treated as equal. OLS will also generate predicted y that will be "below" the "not at all satisfied" option, and "above" the "very satisfied" option.

We assume that individual opinion is continuous, but unobserved. So just as we did with probit, we treat this is a latent regression model, where:

$$y^* = \mathbf{x_i}\beta + \varepsilon_i \tag{15}$$

Again the y^* is unobserved, but we do observe discrete outcomes. Here I work out the example of the democratic satisfaction dependent variable with four discrete ordered outcomes: 1, 2, 3 and 4.

$$\begin{aligned} y_i &= 1 \text{ if } y_i^* < \gamma_1 \\ &= 2 \text{ if } \gamma_1 \geq y_i^* < \gamma_2 \\ &= 3 \text{ if } \gamma_2 \geq y_i^* < \gamma_3 \end{aligned} \tag{16}$$

$$= 4 \text{ if } \gamma_3 \leq y_i^*$$

So, thinking of this as a probit, we have:

$$\begin{aligned} Prob(y_i = 1|\mathbf{x_i}) &= Prob(y_i^* < \gamma_1) \\ &= Prob(\mathbf{x_i}\beta + \varepsilon_i < \gamma_1) \\ &= Prob(\varepsilon_i < \gamma_1 - \mathbf{x_i}\beta) \\ &= \Phi(\gamma_1 - \mathbf{x_i}\beta) \end{aligned} \tag{17}$$

$$\begin{aligned} Prob(y_i = 2|\mathbf{x_i}) &= Prob(\gamma_1 \leq y_i^* < \gamma_2) \\ &= Prob(\gamma_1 \leq \mathbf{x_i}\beta + \varepsilon_i < \gamma_2) \\ &= Prob(\varepsilon_i < \gamma_2 - \mathbf{x_i}\beta) - Prob(\varepsilon_i < \gamma_1 - \mathbf{x_i}\beta) \\ &= \Phi(\gamma_2 - \mathbf{x_i}\beta) - \Phi(\gamma_1 - \mathbf{x_i}\beta) \end{aligned} \tag{18}$$

$$\begin{aligned} Prob(y_i = 3|\mathbf{x_i}) &= Prob(\gamma_2 \leq y_i^* < \gamma_3) \\ &= Prob(\gamma_2 \leq \mathbf{x_i}\beta + \varepsilon_i < \gamma_3) \\ &= Prob(\varepsilon_i < \gamma_3 - \mathbf{x_i}\beta) - Prob(\varepsilon_i < \gamma_2 - \mathbf{x_i}\beta) \\ &= \Phi(\gamma_3 - \mathbf{x_i}\beta) - \Phi(\gamma_2 - \mathbf{x_i}\beta) \end{aligned} \tag{19}$$

$$\begin{aligned} Prob(y_i = 4|\mathbf{x_i}) &= Prob(y_i^* \geq \gamma_3) \\ &= Prob(\mathbf{x_i}\beta + \varepsilon_i \geq \gamma_3) \\ &= Prob(\varepsilon_i \geq \gamma_3 - \mathbf{x_i}\beta) \\ &= 1 - \Phi(\gamma_3 - \mathbf{x_i}\beta) \end{aligned} \tag{20}$$

Here are the Stata MLE estimates of a model with the dependent variables, democratic satisfaction, treated it as an ordered discrete variable. Note that here have estimates of the systematic component of the model, $(\mathbf{x_i}\beta)$ and the cut points, (γ_i)

```
. oprobit demsat retnat class union southwest urban lrself own

Iteration 0:   log likelihood = -1298.5771
Iteration 1:   log likelihood = -1264.2758
Iteration 2:   log likelihood = -1264.2328
```

```
Iteration 3:   log likelihood = -1264.2328

Ordered probit regression                         Number of obs   =       1150
                                                  LR chi2(7)      =      68.69
                                                  Prob > chi2     =     0.0000
Log likelihood = -1264.2328                       Pseudo R2       =     0.0264

------------------------------------------------------------------------------
      demsat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      retnat |   -.333087   .0410278    -8.12   0.000    -.4135001    -.252674
       class |   .0081889   .0332542     0.25   0.805    -.0569882    .0733659
       union |   .0963001   .0784095     1.23   0.219    -.0573797    .2499798
   southwest |   .0036102   .1170997     0.03   0.975     -.225901    .2331214
       urban |   .0180474   .0420636     0.43   0.668    -.0643957    .1004905
      lrself |   .0159106   .0149907     1.06   0.289    -.0134706    .0452918
         own |  -.0306966   .0801961    -0.38   0.702    -.1878781    .1264849
-------------+----------------------------------------------------------------
       /cut1 |  -1.886656   .1788768                     -2.237248   -1.536064
       /cut2 |   -1.00096   .1740084                      -1.34201   -.6599098
       /cut3 |   .7507777   .1732733                      .4111683    1.090387
------------------------------------------------------------------------------


. summarize retnat class union southwest urban lrself own

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      retnat |      1454    2.102476    .8110265          1          3
       class |      1389     2.00432    .9942102          1          5
       union |      1490    .2234899    .4167238          0          1
   southwest |      1500    .0866667    .2814398          0          1
       urban |      1489    1.950974    .7817737          1          3
-------------+--------------------------------------------------------
      lrself |      1308    5.308869    2.189775          1         10
         own |      1470    .7496599    .4333563          0          1
```

What is the probability that the average person in our 2004 UK sample was very satisfied with British democracy, i.e., had a democratic satisfaction score of 4? Stata has estimates all of the relevant parameters necessary to answer this question.

$$\begin{aligned} Prob(y_i = 4|\mathbf{x_i}) &= Prob(y_i^* \geq .75) \quad (21) \\ = 1 - \Phi(.75 - (\beta_1 * (retnat) &+ \beta_2 * (class) + \beta_3 * (union) + \beta_4 * (southwest) \\ &+ \beta_5 * (urban) + \beta_6 * (lrself) + \beta_7 * (own))) \\ = 1 - \Phi(.75 - (-0.33 * (2) &+ 0.008 * (2) + 0.096 * (0) + 0.003 * (0) \\ &+ 0.018 * (2) + 0.016 * (5) - 0.031 * (1))) \\ &= 1 - \Phi(1.31)) \\ &= 1 - (.90) \\ &= .10 \end{aligned}$$

And this is pretty consistent with the actual data...

```
             demsat |      Freq.     Percent        Cum.
--------------------+-----------------------------------
not at all satisfied |        169       11.52       11.52
  not very satisfied |        363       24.74       36.26
    fairly satisfied |        792       53.99       90.25
      very satisfied |        143        9.75      100.00
--------------------+-----------------------------------
              Total |      1,467      100.00
```

We can then generate predicted probabilities for each of the values of democratic satisfaction along with standard errors for each point prediction – you will learn more about doing this in the class on Thursday. In this example all of the independent variables are set to their mean or modal values.

```
     Quantity of Interest |      Mean       Std. Err.    [95% Conf. Interval]
--------------------------+-----------------------------------------------
      Pr(demsat=not at a) |   .0926244     .0087204     .0763543     .1102703
      Pr(demsat=not very) |   .2376394     .0130862     .2124331     .2624363
      Pr(demsat=fairly s) |   .5746827     .0145541     .5467807      .602373
      Pr(demsat=very sat) |   .0950535     .0084244     .0799727     .1126458
```

4 Homework Questions

Note: Question 4 is hard and is a bonus question – try to get as much of it answered as you can but you are not required to answer it.

Question 1

I have estimated a logistic model to explore how judicial selection procedures impact court rulings for or against the governor when he or she is a party in the case (governors are named in many cases). My main independent variable is whether the judge is elected or appointed. I also know whether the judge is the same party as the governor named in the case.

Data: 1000 rulings on cases in which governors are named as a party in the case. Drawn from 42 states over 10 years.

Dependent Variable: coded 1 for a ruling for the governor and 0 for a ruling against.

Independent variables: Elected = 1 if judge was elected, 0 if appointed Party = 1 if judge is same party as governor 0 if otherwise

Table 1: Results

Result	Coefficient
Elected	.07
Party	1 .61
Party*Elected	.21
Constant	.01

1. What is the stochastic component of this model?

2. What is the systematic component of this model?

3. What is the difference in the expected probability of ruling in the governors' favor between two judges who are both from the governor's party but one of whom is elected and the other appointed?

4. What is the difference in the expected probability of ruling in the governor's favor between two judges, both of whom are elected, but one who is from the governor's party and one who is not?

5. Comment on the strengths and weaknesses of this model in this situation

Question 2

A five member committee votes 3-2 in favor of a proposal. Assume voting is independent. Let p be the probability that a committee member votes for the proposal.

1. We have no information with which to distinguish committee members (in the language of Bayesian statistics, wed say that the committee members are exchangeable, but I digress). What is the maximum likelihood estimate (MLE) of p, the probability that any particular committee member votes for the proposal?

2. What is the log-likelihood of $p = .5$? Compare this value of the log-likelihood function with that attained at the MLE with a likelihood ratio test. What does this say about the plausibility of H0 : $p = .5$?

3. How would your conclusion about the plausibility of H0 : $p = .5$ change if we observed

   (a) i. a 10 person committee splitting 6-4 in favor of the proposal?

   (b) ii. a 50 person assembly splitting 30-20 in favor of the proposal? i.e., what is happening to the likelihood function and/or the log-likelihood function in these cases relative to the case of a five person committee? In particular, what is happening the 2nd derivative of the log-likelihood function in the neighborhood of the MLE?

Question 3

The data `q3data.dta` has a subset of the data from a study done in the UK in 2004. The following variables are in the dataset:

`trustgov`	Trust in government (1=none, 10=complete)
`econ`	Retrospective economic evaluation (1=much better, 5=much worse)
`demsat`	Satisfaction with Democracy (1=not at all, 4=very)
`labvote`	Labour voter (1= labour voter, 0=other)
`lrself`	Left-right self-placement (1=extreme left, 10=extreme right)

- Estimate an ordered logit model with `demsat` as the dependent variable and the remaining variables as independent variables. Present your results in a proper table.
- Using the coefficient estimates from the previous step, write out the equation to calculate the probability of being in each category. How many equations should there be?
- Calculate the probability of being in each of the categories for someone who thinks the economy is the same as it was a year ago, trusts government completely, was a labour voter and places himself at 3 on the left-right scale. You may use SPost to do this.
- Generate a graph of the predicted probability of being in each category as the values of trust in government change holding the remaining variables constant at their median values. Hint: Use `prgen` from the SPost suite of commands for this.
- Do a likelihood ratio test for the hypothesis that $\beta_{\text{labvote}} = \beta_{\text{lrself}} = 0$. Do this "by hand" using the `chi2tail()` function to generate the p-value. What do you conclude?

Question 4

Download the file nagler.asc.dta from my web site (www.raymondduch.com). This file contains 98,857 cases (welcome to large n research!) from the 1984 Current Population Survey, analyzed by Jonathan Nagler in two articles: The Effects of Registration Laws and Education on Voter Turnout American Political Science Review, 1991, 85:1393–1405; Scobit: an alternative estimator to logit and probit American Journal of Political Science, 1994, 38:230–255. The data in the file comprise the following variables (in column order): turnout 1 if the respondent reports turning out to vote in the 1984 presidential election, 0 otherwise. educ 1 for 0-4 yrs education; 2 for 5-7 yrs; 3 for 8 yrs; 4 for 9-11 yrs; 5 for 12 yrs; 6 for 1-3 yrs college; 7 for 4 yrs college; 8 for 5+ yrs college age age of respondent, in years south 1 if respondent line in the South, 0 otherwise. govelec 1 if a gubernatorial election coincided with the presidential election closing number of days before election day that voter registration closes in the respondents state The following questions ask to you to estimate a series of logistic regression models. Construct a publication-quality table with the parameter estimates and standard errors for each the models, along with some summary information (e.g., goodness-of-fit, deviance, etc).

1. Estimate a logit model predicting turnout with the predictors educ and age and the square of each of these predictors. Provide a brief write-up of the parameter estimates (i.e., assess statistical significance and substantive implications) and the goodness-of-fit of the logistic regression model.
2. How many unique predicted probabilities are produced by this model? Explain how you derived your answer.

3. Compare the predicted probababilites from the logit model with the corresponding predicted probabilities from a probit model. How and why do they differ, if at all? Is there any statistical basis for preferring logit over probit or vice-versa?

4. Augment your logit model from the first part of this question with the following additional contextual predictors: south, govelec, and closing, and interactions between the two education variables (educ and educ2) and the closing date variable (i.e., make the effects of closing date quadratically conditional on the categorical education measure). Discuss the estimates and goodness-of-fit of this model in contrast with those obtained from the model for the previous question. Report a likelihood ratio test of the joint significance of the new predictors.

5. Using the estimates from the second model, plot the implied coefficient for closing as a function of education, given the interaction effects estimated above. Overlay 95 percent confidence intervals around the point estimates. Offer a substantive interpretation of what this plot reveals.

6. Using the estimates from the second model, consider a hypothetical nonsoutherner, in a state without a gubernatorial election, who has 12 years of education and has the median age of a non-southerner with 12 years of education. Plot the predicted probability of turnout for this person, as the closing date requirement varies over the range of closing date requirements observed in non-southern states. Overlay 95 percent confidence intervals around the point estimates.

7. Using the estimates from the second model, consider a hypothetical nonsoutherner, in a state without a gubernatorial election, who has 5+ years of college and has the median age of a non-southerner with 5+ years of college. Plot the predicted probability of turnout for this person, as the closing date requirement varies over the range of closing date requirements observed in non-southern states. Overlay 95 percent confidence intervals around the point estimates. Briefly compare the answers from this question with those from the previous question.

Due in class Wednesday, Week 5.