View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by IUScholarWorks

Regression Models for Binary Dependent Variables Using Stata, SAS, R, LIMDEP, and SPSS*

Hun Myoung Park, Ph.D.
kucc625@indiana.edu

© 2003-2010
Last modified on October 2010

University Information Technology Services
Center for Statistical and Mathematical Computing
Indiana University
410 North Park Avenue Bloomington, IN 47408
(812) 855-4724 (317) 278-4740
http://www.indiana.edu/~statmath

* The citation of this document should read: "Park, Hun Myoung. 2009. *Regression Models for Binary Dependent Variables Using Stata, SAS, R, LIMDEP, and SPSS*. Working Paper. The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University." http://www.indiana.edu/~statmath/stat/all/cdvm/index.html

This document summarizes logit and probit regression models for binary dependent variables and illustrates how to estimate individual models using Stata 11, SAS 9.2, R 2.11, LIMDEP 9, and SPSS 18.

1. Introduction

A categorical variable here refers to a variable that is binary, ordinal, or nominal. Event count data are discrete (categorical) but often treated as continuous variables. When a dependent variable is categorical, the ordinary least squares (OLS) method can no longer produce the best linear unbiased estimator (BLUE); that is, OLS is biased and inefficient. Consequently, researchers have developed various regression models for categorical dependent variables. The nonlinearity of categorical dependent variable models makes it difficult to fit the models and interpret their results.

1.1 Regression Models for Categorical Dependent Variables

In categorical dependent variable models, the left-hand side (LHS) variable or dependent variable is neither interval nor ratio, but rather categorical. The level of measurement and data generation process (DGP) of a dependent variable determine a proper model for data analysis. Binary responses (0 or 1) are modeled with binary logit and probit regressions, ordinal responses (1^{st}, 2^{nd}, 3^{rd}, …) are formulated into (generalized) ordinal logit/probit regressions, and nominal responses are analyzed by the multinomial logit (probit), conditional logit, or nested logit model depending on specific circumstances. Independent variables on the right-hand side (RHS) are interval, ratio, and/or binary (dummy).

Table 1.1 Ordinary Least Squares and Categorical Dependent Variable Models

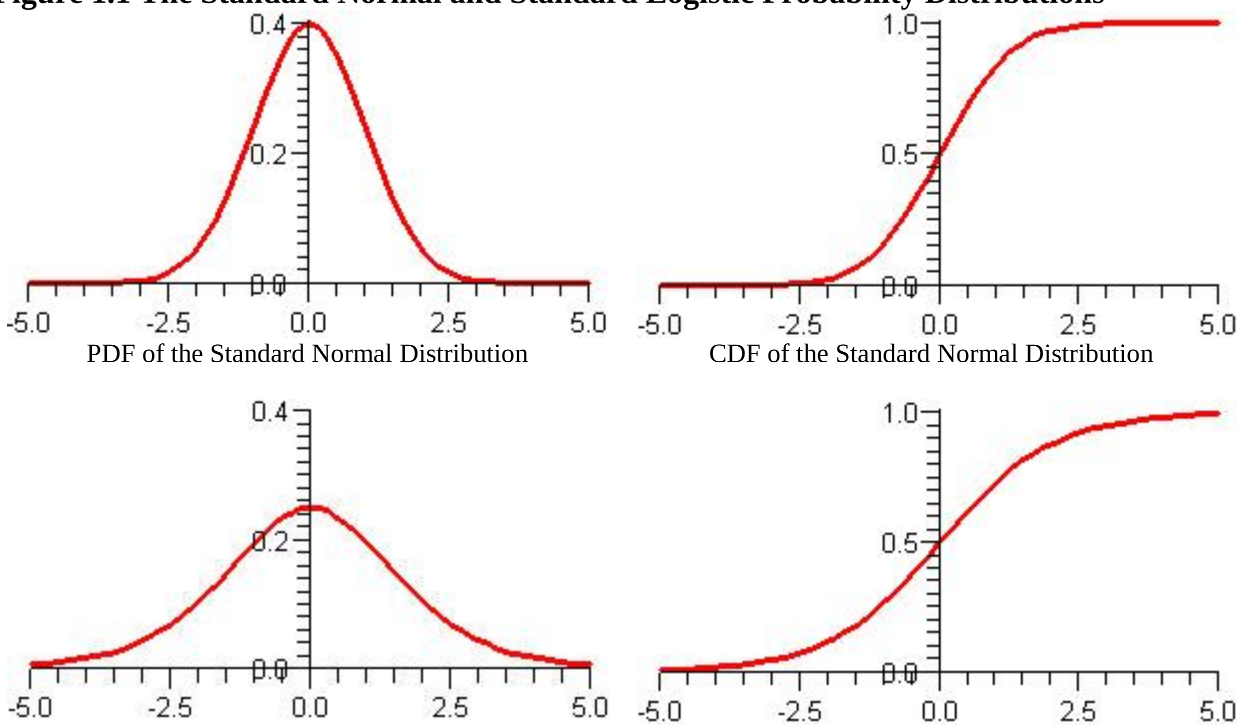
	Model	Dependent (LHS)	Estimation	Independent (RHS)
OLS	Ordinary least squares	Interval or ratio	Moment based method	A linear function of interval/ratio or binary variables $\beta_0+\beta_1X_1+\beta_2X_2...$
Categorical DV Models	Binary response	Binary (0 or 1)	Maximum likelihood method	
	Ordinal response	Ordinal (1^{st}, 2^{nd}, 3^{rd}…)		
	Nominal response	Nominal (A, B, C …)		
	Event count data	Count (0, 1, 2, 3…)		

Categorical dependent variable models adopt the maximum likelihood (ML) estimation method, whereas OLS uses the moment based method. The ML method requires an assumption about probability distribution functions, such as the logistic function and the complementary log-log function. Logit models use the standard logistic probability distribution, while probit models assume the standard normal distribution. This document focuses on logit and probit models only, excluding regression models for event count data (e.g., negative binomial regression model and zero-inflated or zero-truncated regression models). Table 1.1 summarizes categorical dependent variable models in comparison with OLS.

1.2 Logit Models versus Probit Models

How do logit models differ from probit models? The core difference lies in the distribution of errors (disturbances). In the logit model, errors are assumed to follow the standard logistic distribution with mean 0 and variance $\frac{\pi^2}{3}$, $\lambda(\varepsilon) = \frac{e^{\varepsilon}}{(1+e^{\varepsilon})^2}$. The errors of the probit model are assumed to follow the standard normal distribution, $\phi(\varepsilon) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\varepsilon^2}{2}}$ with variance 1.

Figure 1.1 The Standard Normal and Standard Logistic Probability Distributions



PDF of the Standard Normal Distribution

CDF of the Standard Normal Distribution

PDF of the Standard Logistic Distribution

CDF of the Standard Logistic Distribution

The probability density function (PDF) of the standard normal probability distribution has a higher peak and thinner tails than the standard logistic probability distribution (Figure 1.1). The standard logistic distribution looks as if someone has weighed down the peak of the standard normal distribution and strained its tails. As a result, the cumulative density function (CDF) of the standard normal distribution is steeper in the middle than the CDF of the standard logistic distribution and quickly approaches zero on the left and one on the right.

The two models, of course, produce different parameter estimates. In binary response models, the estimates of a logit model are roughly $\pi/\sqrt{3}$ times larger than those of the probit model. These estimators, however, end up with almost the same standardized impacts of independent variables (Long 1997).

The choice between logit and probit models is more closely related to estimation and familiarity than to theoretical or interpretive aspects. In general, logit models reach convergence fairly well. Although some (multinomial) probit models may take a long time to reach convergence, a probit model works well for bivariate models. As computing power improves and new algorithms are developed, importance of this issue is diminishing. For discussion of selecting logit or probit models, see Cameron and Trivedi (2009: 471-474).

1.3 Estimation in SAS, Stata, LIMDEP, R, and SPSS

Table 1.2 summarizes the procedures and commands used for categorical dependent variable models. Note that Stata and R are case-sensitive, but SAS, LIMDEP, and SPSS are not.

Table 1.2 Procedures and Commands for Categorical Dependent Variable Models

	Model	Stata 11	SAS 9.2	R	LIMDEP 9	SPSS17
OLS		`.regress`	REG	`lme()`	Regress$	Regression
Binary	Binary logit	`.logit`, `.logistic`	QLIM, LOGISTIC, GENMOD, PROBIT	`glm()`	Logit$	Logistic regression
	Binary probit	`.probit`	QLIM, LOGISTIC, GENMOD, PROBIT	`glm()`	Probit$	Probit
Bivariate	Bivariate probit	`.biprobit`	QLIM	`bprobit()`	Bivariateprobit$	-
Ordinal	Ordinal logit	`.ologit`	QLIM, LOGISTIC, GENMOD, PROBIT	`lrm()`	Ordered$, Logit$	Plum
	Generalized logit	`.gologit2`*	-	`logit()`	-	-
	Ordinal probit	`.oprobit`	QLIM, LOGISTIC, GENMOD, PROBIT	`polr()`	Ordered$	Plum
Nominal	Multinomial logit	`.mlogit`	LOGISTIC, CATMOD	`multinom()`, `mlogit()`	Mlogit$, Logit$	Nomreg
	Conditional logit	`.clogit`	LOGISTIC, MDC, PHREG	`clogit()`	Clogit$, Logit$	Coxreg
	Nested logit	`.nlogit`	MDC	-	Nlogit$**	-
	Multinomial probit	`.mprobit`	-	`mnp()`	-	-

* A user-written command written by Williams (2005)

** The `Nlogit$` command is supported by NLOGIT, a stand-alone package, which is sold separately.

Stata offers multiple commands for categorical dependent variable models. For example, the `.logit` and `.probit` commands respectively fit the binary logit and probit models, while `.mlogit` and `.nlogit` estimate the mulitinomial logit and nested logit models. Stata enables users to perform post-hoc analyses such as marginal effects and discrete changes in an easy manner.

SAS provides several procedures for categorical dependent variable models, such as PROC LOGISTIC, PROBIT, GENMOD, QLIM, MDC, PHREG, and CATMOD. Since these procedures support various models, a categorical dependent variable model can be estimated by multiple procedures. For example, you may run a binary logit model using PROC LOGISTIC, QLIM, GENMOD, and PROBIT. PROC LOGISTIC and PROC PROBIT of SAS/STAT have been commonly used, but PROC QLIM and PROC MDC of SAS/ETS have advantages over other procedures. PROC LOGISTIC reports factor changes in the odds and tests key hypotheses of a model. The QLIM (Qualitative and LImited dependent variable Model) procedure in SAS analyzes various categorical and limited dependent variable regression models such as censored, truncated, and sample-selection models. PROC QLIM also handles Box-Cox regression and the bivariate probit model. The MDC (Multinomial Discrete Choice) procedure can estimate conditional logit and nested logit models.[1]

In R, `glm()` fits binary logit and probit models in the object- oriented programming concept. Multiple other functions have been developed to fit other categorical dependent variable models. The LIMDEP `Logit$` and `Probit$` commands support a variety of categorical dependent variable models that are addressed in Greene's *Econometric Analysis* (2003). The output format of LIMDEP 9 is slightly different from that of previous version, but key statistics remain unchanged. The nested logit model and multinomial probit model in LIMDEP are estimated by NLOGIT, a separate package. SPSS also supports some categorical dependent variable models and its output is often messy and hard to read.

1.4 Long and Freese's SPost

Stata users may benefit from user-written commands such as J. Scott Long and Jeremy Freese's SPost. This collection of user-written commands conducts many follow-up analyses of various categorical dependent variable models including event count data models. See section 2.2 for the most common SPost commands.

In order to install SPost, execute the following commands consecutively. Visit J. Scott Long's Web site at http://www.indiana.edu/~jslsoc/ to get further information.

```
. net from http://www.indiana.edu/~jslsoc/stata/
. net install spost9_ado, replace
. net get spost9_do, replace
```

[1] An advantage of using SAS is the Output Delivery System (ODS), which makes it easy to manage SAS output. ODS enables users to redirect the output to HTML (Hypertext Markup Language) and RTF (Rich Text Format) formats. Once SAS output is generated in an HTML document, users can easily handle tables and graphics especially when copying and pasting them into a wordprocessor document.

If a Stata command, function, or user-written command does not work in version 11, run the `.version` command to switch the interpreter to old one and execute that command again. For example, `normal()` was `norm()` in old versions.

```
. version 9
```

Also you may update Stata or reinstall user-written commands to get their latest version installed.

```
. update all
```

2. Binary Logit Regression Model

The binary logit model is represented as $\text{Prob}(y=1|x)=\Lambda(x\beta)=\dfrac{\exp(x\beta)}{1+\exp(x\beta)}$, where Λ indicates a link function, the cumulative standard logistic distribution function. This chapter illustrates how to fit the binary logit model. The sample model considered here explores how social trust is affected by education, family income, age, gender, and Internet use (www).

2.1 Binary Logit Model in Stata (.logit)

Stata provides two equivalent commands for the binary logit model that present the same result in different ways. The `.logit` command produces coefficients with respect to logit (log of odds), while `.logistic` reports odd ratios.

```
. logistic trust educate income age male www

Logistic regression                               Number of obs   =       1174
                                                  LR chi2(5)      =     128.68
                                                  Prob > chi2     =     0.0000
Log likelihood = -733.97164                       Pseudo R2       =     0.0806

------------------------------------------------------------------------------
       trust | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     educate |   1.163673   .0304619     5.79   0.000     1.105474    1.224935
      income |   1.030814   .0118919     2.63   0.009     1.007768    1.054387
         age |   1.028411   .0050091     5.75   0.000      1.01864    1.038276
        male |   1.292781    .162669     2.04   0.041     1.010228    1.654362
         www |   1.739745   .2885914     3.34   0.001      1.25686    2.408153
------------------------------------------------------------------------------
```

This model fits the data very well (p<.0000) and all independent variables except for gender are statistically significant at the .01 level. Interpretation of the odds ratio will be discussed in Section 2.2. In order to get the coefficients (log of odds), simply run `.logit` without any argument right after the `.logistic` command.

```
. logit
(output is skipped)
```

Or you may run a separate `.logit` command with all arguments. Both commands report the same goodness-of-fit measures such as likelihood ratio and McFadden's pseudo R^2.

```
. logit trust educate income age male www

Iteration 0:   log likelihood = -798.31217
Iteration 1:   log likelihood = -734.25733
Iteration 2:   log likelihood = -733.97169
Iteration 3:   log likelihood = -733.97164

Logistic regression                               Number of obs   =       1174
                                                  LR chi2(5)      =     128.68
                                                  Prob > chi2     =     0.0000
Log likelihood = -733.97164                       Pseudo R2       =     0.0806

------------------------------------------------------------------------------
       trust |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     educate |   .1515812   .0261774     5.79   0.000     .1002745    .2028879
      income |   .0303485   .0115364     2.63   0.009     .0077376    .0529595
         age |   .0280152   .0048707     5.75   0.000     .0184688    .0375616
        male |    .256796   .1258287     2.04   0.041     .0101762    .5034157
         www |   .5537383   .1658815     3.34   0.001     .2286165    .8788601
       _cons |  -4.983007    .478359   -10.42   0.000    -5.920574   -4.045441
------------------------------------------------------------------------------
```

A coefficient of `.logit` is the corresponding logarithmic transformed odds ratio of `.logistic`. For example, the coefficient of education is .1516= log(1.1637) or 1.1637=exp(.1516).

Stata has post-estimation commands that conduct follow-up analyses. The following `.predict` command with the `residual` option computes residuals and then stores them into a new variable `resid`.

```
. predict resid, residual
```

The `.test` and `.lrtest` commands respectively conduct the Wald test and likelihood ratio test. A large chi-squared rejects the null hypothesis that the parameter of education is zero. Education has a significant positive impact on social trust.

```
. test educate

 ( 1)  [trust]educate = 0

           chi2(  1) =   33.53
         Prob > chi2 =    0.0000
```

Marginal effects and discrete changes are very useful when interpreting the result of a binary logit or probit model. The marginal effect of a continuous independent variable x_c is the partial derivative with respect to that variable. The discrete change of a binary independent variable (dummy variable) x_b is the difference in predicted probabilities of $x_b = 1$ and $x_b = 0$, holding all other independent variables constant at their reference points. x_{-b} denotes all independent variables other than x_b Marginal effects and discrete changes look similar but are not equal in conceptual and numerical senses.

$$\frac{\partial P(y=1|x)}{\partial x_c} = \frac{\exp(x\beta)}{[1+\exp(x\beta)]^2} = \Lambda(x\beta)(1-\Lambda(x\beta)\beta_c \text{ (marginal effect of } x_c\text{)}$$

$$\frac{\Delta P(y=1|x)}{\Delta x_b} = P(y=1|x_{-b}, x_b=1) - P(y=1|x_{-b}, x_b=0) \text{ (discrete change of } x_b\text{)}$$

The `.mfx` command with `dydx` (partial derivatives), the default option, computes marginal effects for continuous covariates and discrete changes for binary variables at the reference points after the estimation of a linear or nonlinear regression model. You may change reference points using the `at()` option; If this option is not specified, Stata by default uses means of independent variables as reference points. `mean` in the `at()` option below says that if a covariate is not listed in `at()`, its mean is used as its reference point.

```
. mfx, dydx at(mean educate=16 male=0 www=1)

Marginal effects after logit
      y  = Pr(trust) (predict)
         =  .47534926
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
 educate |   .0378032       .0066    5.73   0.000   .024873  .050734        16
  income |   .0075687      .00287    2.63   0.008   .001934  .013203   24.6486
     age |   .0069868      .00121    5.75   0.000   .004606  .009367   41.3075
   male*|   .0640968      .03132    2.05   0.041   .002718  .125475         0
    www*|   .1329051      .03797    3.50   0.000   .058487  .207323         1
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

The predicted probability of trusting most people is .4753 for female WWW users at the average age of 41 who graduated a college (16 years of education) and have average family income of 25 thousands dollars. Marginal effects and discrete changes are listed under `dy/dx`. For a year increase in education after college graduation, the predicted probability of trusting people will increase by 3.78 percent, holding other independent variables constant at the reference points (see the list of values under the label `x`). WWW users are 13.29 percent more likely than non-users to trust people, holding other covariates at the reference points.

2.2 Using SPost Commands in Stata

SPost commands provide useful follow-up analysis commands (ado files) for categorical dependent variable models (Long and Freese 2003). The `.fitstat` command reports various goodness-of-fit measures such as log likelihood, McFadden's R^2 (or Pseudo R^2), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). 1467.943 labeled as `D(1168)` is -2*Log-likelihood (=-2*-733.972) and 1,168=N-K=1,174-6, where K denotes the number of parameters including the intercept.

```
. net install spost9_ado, replace from(http://www.indiana.edu/~jslsoc/stata/)
checking spost9_ado consistency and verifying not already installed...

. fitstat

Log-Lik Intercept Only:       -798.312   Log-Lik Full Model:            -733.972
D(1168):                      1467.943   LR(5):                          128.681
                                         Prob > LR:                        0.000
McFadden's R2:                   0.081   McFadden's Adj R2:                0.073
ML (Cox-Snell) R2:               0.104   Cragg-Uhler(Nagelkerke) R2:       0.140
McKelvey & Zavoina's R2:         0.140   Efron's R2:                       0.105
Variance of y*:                  3.826   Variance of error:                3.290
Count R2:                        0.654   Adj Count R2:                     0.175
```

```
AIC:                              1.261   AIC*n:                        1479.943
BIC:                          -6787.682   BIC':                          -93.340
BIC used by Stata:             1510.352   AIC used by Stata:            1479.943
```

The likelihood ratio statistic is based on the difference of log likelihoods between the null model and the full model. 128.68=-2*[(-798.312)-(-733.972)].

The binary logit (log of the odds) model can be expressed in a log-linear form of $\ln\Omega(x) = x\beta$, where $\Omega(x)$ is the odds of the success (y=1) given *x* (Long 1997: 79). The odds ratio is used to examine the change in the odds when an independent variable x_{odds} increases by δ; a odds ratio greater than 1 means that the odds increase as that variable increase by δ (pp. 80-82).

The odds: $\Omega(x) = \dfrac{P(y=1|x)}{P(y=0|x)} = \dfrac{P(y=1|x)}{1-P(y=1|x)} = \dfrac{\Lambda(x\beta)}{1-\Lambda(x\beta)}$

Odds ratio: $\dfrac{\Omega(x_{-odds}, x_{odds}+\delta)}{\Omega(x_{-odds}, x_{odds})} = \exp(\beta_{odds}\delta)$

The `.listcoef` command produces a table of unstandardized coefficients (parameter estimates), factor (percent) changes in odds, and standardized coefficients. The `help` option helps read the output of `.listcoef`. Find factor changes in odds under the labels `e^b` and `e^bStdX`. Factor changes in odds are, in fact, the odds ratios that `.logistic` produced on page 6.

Long (1997) discusses interpretation of binary response models using factor changes in odds and predicted probabilities. For a unit increase in education, for example, the odds are expected to increase by a factor of 1.1637=exp(.1516). Alternatively, for a standard deviation change in education, the odds will change by a factor of 1.4763=exp(.1516*2.5697). Notice that the last column under `SDofX` lists standard deviations of covariates. The odds of trusting people are 1.2928=exp(.2568) times larger for men than for women, holding all other variables constant.

```
. listcoef, help

logit (N=1174): Factor Change in Odds

  Odds of: 1 vs 0

----------------------------------------------------------------------
       trust |      b         z     P>|z|    e^b    e^bStdX      SDofX
-------------+--------------------------------------------------------
     educate |   0.15158    5.791   0.000   1.1637   1.4763     2.5697
      income |   0.03035    2.631   0.009   1.0308   1.2068     6.1943
         age |   0.02802    5.752   0.000   1.0284   1.4559    13.4071
        male |   0.25680    2.041   0.041   1.2928   1.1364     0.4978
         www |   0.55374    3.338   0.001   1.7397   1.2554     0.4108
----------------------------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
     e^b = exp(b) = factor change in odds for unit increase in X
 e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
   SDofX = standard deviation of X
```

You may interpret factor change in odds in a reverse way. Pay attention to `reverse` of the `.listcoef` command. For a standard deviation change in education, the odds of having NO social trust are expected to decrease by a factor of .6774=exp(-.1516*2.5697). The odds of NOT trusting people are .7735=exp(-.2568) times smaller for men than for women. The labels `e^b` and `e^bStdX` below should be `e^(-b)` and `e^(-bStdX)`, respectively.

```
. listcoef, reverse

logit (N=1174): Factor Change in Odds

  Odds of: 0 vs 1

----------------------------------------------------------------------
       trust |      b         z     P>|z|    e^b    e^bStdX      SDofX
-------------+--------------------------------------------------------
     educate |   0.15158    5.791   0.000   0.8593   0.6774     2.5697
      income |   0.03035    2.631   0.009   0.9701   0.8286     6.1943
         age |   0.02802    5.752   0.000   0.9724   0.6869    13.4071
        male |   0.25680    2.041   0.041   0.7735   0.8800     0.4978
         www |   0.55374    3.338   0.001   0.5748   0.7966     0.4108
----------------------------------------------------------------------
```

Alternatively, you may use percent changes in the odds by adding the `percent` option. For example, the odds of trusting people are 29.3 percent larger for men than for women, holding all other covariates constant.

```
. listcoef, percent help

logit (N=1174): Percentage Change in Odds

  Odds of: 1 vs 0

----------------------------------------------------------------------
       trust |      b         z     P>|z|      %      %StdX      SDofX
-------------+--------------------------------------------------------
     educate |   0.15158    5.791   0.000     16.4     47.6     2.5697
      income |   0.03035    2.631   0.009      3.1     20.7     6.1943
         age |   0.02802    5.752   0.000      2.8     45.6    13.4071
        male |   0.25680    2.041   0.041     29.3     13.6     0.4978
         www |   0.55374    3.338   0.001     74.0     25.5     0.4108
----------------------------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
       % = percent change in odds for unit increase in X
   %StdX = percent change in odds for SD increase in X
   SDofX = standard deviation of X
```

The `.prvalue` command lists predicted probabilities of positive and negative outcomes for a given set of values for the independent variables. The following example predicts, as shown in `.mfx` above, that 47.53 percent of female WWW users will trust most people at the reference points (educate=16, income=24.65, age=41.31), while 52.47 percent will not.

```
. prvalue, x(educate=16 male=0 www=1) rest(mean)

logit: Predictions for trust

Confidence intervals by delta method

                                95% Conf. Interval
  Pr(y=1|x):          0.4753   [ 0.4277,    0.5230]
  Pr(y=0|x):          0.5247   [ 0.4770,    0.5723]

     educate     income        age       male        www
x=        16  24.648637  41.307496          0          1
```

The `.prtab` command constructs a table of predicted values (probabilities) for all combinations of categorical variables listed. Both `.prtab` and `.prvalue` report the same predicted probability of .4753 that female WWW users trust most people. The table below suggests that male WWW users are more likely to trust than their counterparts (53.94 percent versus 34.24 percent, respectively). The `x()` option specifies particular values of covariates other than their means as reference points. The `rest()` option sets the reference points of independent variables that are not specified in `x()`.

```
. prtab male www, x(educate=16 male=0 www=1) rest(mean)

logit: Predicted probabilities of positive outcome for trust

--------------------------------
          |       WWW Use
   Gender | Non-users      Users
----------+---------------------
   Female |    0.3424     0.4753
     Male |    0.4024     0.5394
--------------------------------

      educate     income        age       male        www
x=         16  24.648637  41.307496          0          1
```

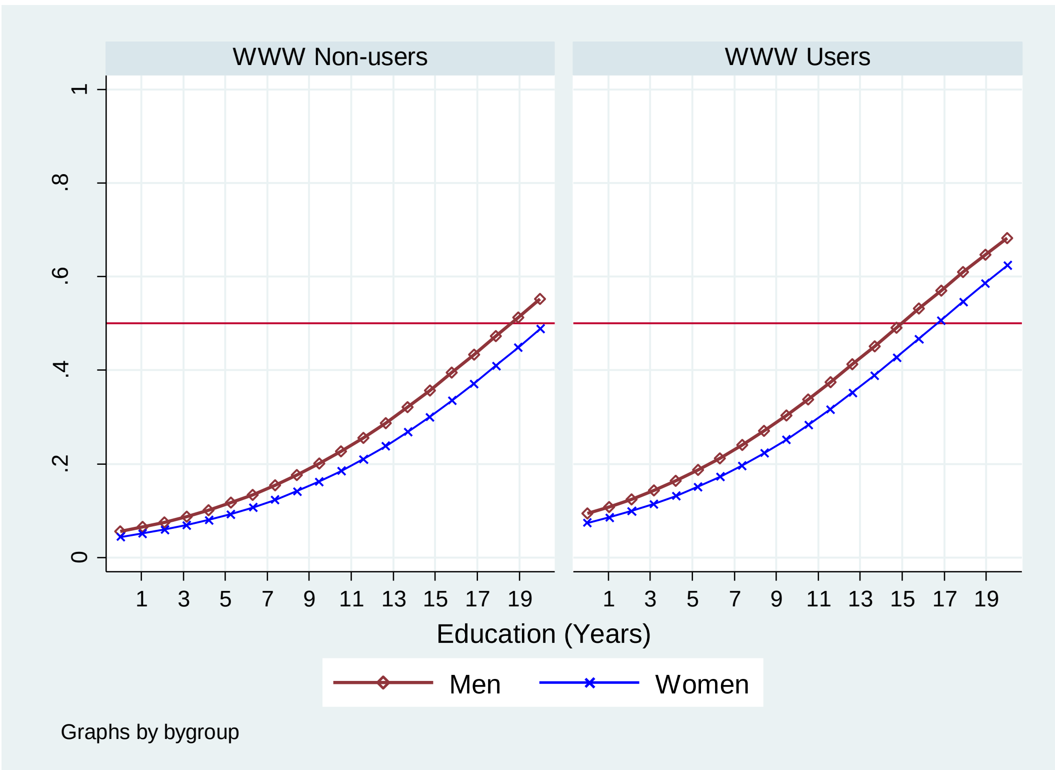
The most useful command for binary response models is `.prchange`, which calculates marginal effects and discrete changes at a given set of values of independent variables. The predicted probability of .4753 and the marginal effects (discrete changes) are the same as what `.mfx` produced above. Read marginal effects under the last `MargEfct` (or `-+1/2`) column and discrete changes under `0->1` (when changing the value from 0 to 1). For an additional year of education after college, the predicted probability of trusting people is expected to increase by 3.78 percent (marginal effect) when holding all other covariates constant at their reference points. WWW users are 13.29 percent (discrete change) more likely than non-users to trust people, holding other variable at their reference points.

```
. prchange, x(educate=16 male=0 www=1) rest(mean)

logit: Changes in Probabilities for trust

         min->max      0->1     -+1/2    -+sd/2  MargEfct
educate    0.5264    0.0111    0.0378    0.0968    0.0378
 income    0.1936    0.0064    0.0076    0.0468    0.0076
    age    0.4397    0.0049    0.0070    0.0934    0.0070
   male    0.0641    0.0641    0.0640    0.0319    0.0640
    www    0.1329    0.1329    0.1372    0.0567    0.1381

              0       1
Pr(y|x)  0.5247  0.4753

       educate   income      age     male      www
   x=       16  24.6486  41.3075        0        1
sd_x=  2.56971  6.19427  13.4071  .497765  .410755
```

SPost `.prgen` computes a series of predictions (predicted probabilities in this case) by holding all variables but one interval variable constant and allowing that variable to vary (Long and Freese 2003). The first command below computes predicted probabilities that male WWW users (`male=1` and `www=1`) trust most people when education changes from 0 through 20 years, holding other independent variables at the reference points, and then stores them into new variables, whose names begin with `Logit_ed11`.

```
. prgen educate, from(0) to(20) ncases(20) x(male=1 www=1) rest(mean) gen(Logit_ed11)

logit: Predicted values as educate varies from 0 to 20.

      educate      income         age        male         www
x=   14.24276   24.648637   41.307496           1           1

. prgen educate, from(0) to(20) ncases(20) x(male=1 www=0) rest(mean) gen(Logit_ed10)

logistic: Predicted values as educate varies from 0 to 20.

      educate      income         age        male         www
x=   14.24276   24.648637   41.307496           1           0

. prgen educate, from(0) to(20) ncases(20) x(male=0 www=1) rest(mean) gen(Logit_ed01)

logistic: Predicted values as educate varies from 0 to 20.

      educate      income         age        male         www
x=   14.24276   24.648637   41.307496           0           1

. prgen educate, from(0) to(20) ncases(20) x(male=0 www=0) rest(mean) gen(Logit_ed00)

logistic: Predicted values as educate varies from 0 to 20.

      educate      income         age        male         www
x=   14.24276   24.648637   41.307496           0           0
```

Figure 2.1 Predicted Probabilities of Trusting Most People (Binary Logit Model)



After generating predicted probabilities of other groups (male WWW non-users, female users, and female non-users), you can draw Figure 2.1. See the Stata script in Appendix for necessary data manipulation. Figure 2.1 suggests that education and WWW use influence social trust significantly but gender does not.

2.3 Binary Logit Model in SAS: PROC LOGISTIC and PROC PROBIT

SAS has several procedures for the binary logit model such as LOGISTIC, PROBIT, GENMOD, and QLIM procedures. PROC LOGISTIC is commonly used for the binary logit model, but PROC PROBIT is also able to estimate the binary logit model.

Unlike PROC QLIM, LOGISTIC, PROBIT, and GENMOD procedures by default use a smaller value in the dependent variable as success (positive event). As a consequence, magnitudes of the coefficients remain the same, but their signs are opposite to those of PROC QLIM, Stata, and LIMDEP. The DESCENDING (DESC) option in PROC LOGISTIC and PROC GENMOD forces SAS to use a larger value as success. Notice that a SAS procedure is comprised of a series of statements, each of which ends with a semi-colon.

```
PROC LOGISTIC DESCENDING DATA = masil.gss_cdvm;
     MODEL trust = educate income age male www;
RUN;
```

Alternatively, you may explicitly specify the category of successful event using the EVENT option. EVENT=LAST (or EVENT='1') use the last ordered category (1) as a successful event. Both approaches produce the same results.

```
PROC LOGISTIC DATA = masil.gss_cdvm;
    MODEL trust(EVENT=LAST) = educate income age male www;
RUN;
```

```
                                     The LOGISTIC Procedure

                                       Model Information

                    Data Set                      MASIL.GSS_CDVM
                    Response Variable             trust                   trust
                    Number of Response Levels     2
                    Model                         binary logit
                    Optimization Technique        Fisher's scoring


                             Number of Observations Read        1174
                             Number of Observations Used        1174


                                        Response Profile

                                Ordered                       Total
                                  Value          trust     Frequency

                                      1              1           492
                                      2              0           682
```

```
                        Probability modeled is trust=1.


                           Model Convergence Status

                  Convergence criterion (GCONV=1E-8) satisfied.


                             Model Fit Statistics

                                                 Intercept
                                   Intercept           and
                     Criterion          Only    Covariates

                     AIC            1598.624      1479.943
                     SC             1603.693      1510.352
                     -2 Log L       1596.624      1467.943

                     Testing Global Null Hypothesis: BETA=0

             Test                 Chi-Square       DF     Pr > ChiSq

             Likelihood Ratio       128.6811        5         <.0001
             Score                  121.5344        5         <.0001
             Wald                   109.6453        5         <.0001


                    Analysis of Maximum Likelihood Estimates

                                          Standard          Wald
       Parameter    DF    Estimate           Error    Chi-Square    Pr > ChiSq

       Intercept     1     -4.9830          0.4784      108.5101        <.0001
       educate       1      0.1516          0.0262       33.5302        <.0001
       income        1      0.0303          0.0115        6.9200        0.0085
       age           1      0.0280         0.00487       33.0824        <.0001
       male          1      0.2568          0.1258        4.1650        0.0413
       www           1      0.5537          0.1659       11.1431        0.0008


                             Odds Ratio Estimates

                                 Point          95% Wald
                   Effect     Estimate      Confidence Limits

                   educate       1.164       1.105       1.225
                   income        1.031       1.008       1.054
                   age           1.028       1.019       1.038
                   male          1.293       1.010       1.654
                   www           1.740       1.257       2.408


         Association of Predicted Probabilities and Observed Responses

               Percent Concordant      68.4    Somers' D    0.371
               Percent Discordant      31.3    Gamma        0.373
```

```
                    Percent Tied            0.4    Tau-a          0.181
                    Pairs                335544    c              0.686
```

Stata and SAS produce the same results. Log likelihood is -733.9716 =(1467.943/-2); SAS report -2*log likelihood 1467.943. Likelihood ratio is 128.681=1596.624-1467.943. McFadden's pseudo R^2 is .0806=1-(1467.943/1596.624). AIC and BIC (or Schwarz information criterion) are 1479.943 and 1510.352, respectively, in both outputs. Parameter estimates and their standard errors are the same. However, Stata and SAS respectively conduct z test and Wald test to examine the effects of individual independent variables but produce the same p-values, except for rounding errors. For example, Stata's z score 5.79 for education is the square root of the Wald statistic 33.53.

If you want to get the output in the HTML format, use ODS statements before and after a SAS procedure. `ODS HTML` redirects SAS output to the HTML format. The output is skipped.

```
ODS HTML;
PROC LOGISTIC . . .
. . .
ODS HTML CLOSE;
```

PROC LOGISTIC by default reports odds changes when independent variables increase by a unit. The odds changes (ratios) under `Odds Ratio Estimates` are the same as what Stata `.listcoef` produced in Section 2.2. For a unit ($1,000) increase in family income, the odds of having social trust are expected to change by a factor of 1.031=exp(.0303), holding all other covariates constant. The odds of having social trust are 1.293=exp(.2568) times larger for men than for women; conversely, the odds of having no social trust are .7734=exp(-.2568) times smaller for men than for women.

The UNITS statement specifies a unit other than means of covariates. The SD in UNITS indicates a standard deviation increase in covariates listed (`educate`, `income`, and `age` in this example). UNITS adds factor changes in odds to the end of the LOGISTIC output. Read numbers under `Odds Ratios` (other output is skipped below). For a standard deviation increase in family income, the odds are expected to increase by a factor of 1.207=exp(.0303*6.1943). You may find the same number under `e^bStdX` of `.listcoef` in Section 2.2.

```
PROC LOGISTIC DATA = masil.gss_cdvm;
    MODEL trust(EVENT='1') = educate income age male www;
    UNITS educate=SD income=SD age=SD;
RUN;
```

```
                                    Odds Ratios

                         Effect          Unit     Estimate

                         educate       2.5697        1.476
                         income        6.1943        1.207
                         age          13.4071        1.456
```

Let us compute marginal effects manually. See Park (2004) for computation in detail. If you are not familiar with SAS, you may skip this part. The first step is to get parameter estimates and reference points. In PROC LOGISTIC, add OUTEST=masil.blm to store parameter estimates into a SAS data set *masil.blm*. PROC MEANS with MEAN and STD computes means and standard deviations of variables listed in the VAR statement and then store them into *masil.meanX*. Notice that SAS, unlike Stata and R, is not case-insensitive.

```
PROC LOGISTIC DESCENDING DATA = masil.gss_cdvm OUTEST=masil.blm;
     MODEL trust = educate income age male www;

PROC MEANS MEAN STD DATA = masil.gss_cdvm;
     VAR educate income age male www;
     OUTPUT OUT=masil.meanX;
RUN;
(output is skipped)
```

Next, convert two SAS data sets into matrices, *bHat* and *X* in PROC IML. Then, compute predicted probability and marginal effects. Pay attention to comments enclosed by /* and */.

```
PROC IML;
USE masil.blm; /* get a row vector of parameter estimates */
READ ALL VAR{Intercept educate income age male www} INTO bHat;
K=NCOL(bHat); /* get the number of regressors */

USE masil.meanX;
READ ALL VAR{educate income age male www} INTO X;
meanX = {1} || X[4,]; /* a row vector of means of independent variables */
sdX = {0} || X[5,]; /* a row vector of standard deviations of independent variables */

referX = meanX; /* set reference points */
referX[1,2]=16; referX[1,5]=0; referX[1,6]=1; /* education=16, male=0, www=1 */

xb = bHat * T(referX);
prob = exp(xb)/(1+exp(xb)); /* compute a predicted probability */

PRINT referX prob;

margin = prob * (1-prob) * T(bHat); /* compute marginal effects */
marginSD = prob * (1-prob) * T(bHat # sdX);

result = T(bHat) || T(exp(bHat))||T(exp(bHat # sdX)) || margin||marginSD || T(meanX)||T(sdX);
result = result[2:K,];

PRINT result[ROWNAME={"educate", "income", "age", "male", "www"}
      COLNAME={"b" "exp(b)" "exp(b*sdX)" "MargEffect" "MargEffect(SD)" "Mean of X" "SD of X"}];

QUIT; /* terminate PROC IML */
```

The following is the output of the PROC IML above. Compare marginal effects with what `.prchange` reported in Section 2.2. Notice that .0640 and .1381 are not correct discrete changes of gender and WWW use, respectively. Factor changes in the odds are also listed under labels `exp(b)` and `exp(b*sdX)`.

```
                 referX                                                   prob

                      1        16 24.648637 41.307496         0         1 0.4753497
```

```
                                         result
                 b     exp(b) exp(b*sdX) MargEffect MargEffect(SD) Mean of X   SD of X

educate 0.1515807 1.1636722  1.4762701  0.0378031       0.097143  14.24276 2.5697123
income  0.0303475 1.0308127  1.2068103  0.0075684       0.046881 24.648637 6.1942699
age     0.0280151 1.0284112  1.4558671  0.0069867      0.0936722 41.307496 13.407127
male    0.2567949   1.29278  1.1363525  0.0640427      0.0318782 0.4505963 0.4977653
www     0.5537335 1.7397362   1.255393  0.1380969       0.056724 0.7853492 0.4107548
```

PROC PROBIT is primarily designed for the binary probit model but can estimate the same binary logit model as well. The /DIST=LOGISTIC option indicates the link function (probability distribution) to be used in maximum likelihood estimation.

```
PROC PROBIT DATA = masil.gss_cdvm;
     MODEL trust = educate income age male www /DIST=LOGISTIC;
RUN;
```

```
                                    The Probit Procedure

                                      Model Information

              Data Set                    MASIL.GSS_CDVM
              Dependent Variable                   trust    trust
              Number of Observations                1174
              Name of Distribution              Logistic
              Log Likelihood                  -733.97164


                         Number of Observations Read        1174
                         Number of Observations Used        1174


                                   Class Level Information

                                  Name        Levels    Values

                                  trust            2    0 1


                                       Response Profile

                                 Ordered                   Total
                                   Value     trust     Frequency

                                       1     0               682
                                       2     1               492

PROC PROBIT is modeling the probabilities of levels of trust having LOWER Ordered Values in the
response profile table.

           Algorithm converged.


                                 Type III Analysis of Effects
```

```
                          Wald
Effect        DF    Chi-Square    Pr > ChiSq

educate        1       33.5304        <.0001
income         1        6.9204        0.0085
age            1       33.0827        <.0001
male           1        4.1650        0.0413
www            1       11.1433        0.0008


        Analysis of Maximum Likelihood Parameter Estimates

                        Standard   95% Confidence     Chi-
Parameter  DF Estimate     Error        Limits        Square Pr > ChiSq

Intercept   1   4.9830    0.4784   4.0454   5.9206   108.51      <.0001
educate     1  -0.1516    0.0262  -0.2029  -0.1003    33.53      <.0001
income      1  -0.0303    0.0115  -0.0530  -0.0077     6.92      0.0085
age         1  -0.0280    0.0049  -0.0376  -0.0185    33.08      <.0001
male        1  -0.2568    0.1258  -0.5034  -0.0102     4.17      0.0413
www         1  -0.5537    0.1659  -0.8789  -0.2286    11.14      0.0008
```

Unlike PROC LOGISTIC, PROC PROBIT does not have the DESCENDING (or DESC) option. Therefore, you have to switch the signs of coefficients when comparing with PROC LOGISTIC, Stata, and LIMDEP. PROC PROBIT does not have the UNITS statement to compute factor changes in the odds.

2.4 Binary Logit Model in SAS: PROC QLIM and PROC GENMOD

PROC QLIM estimates not only logit and probit models, but also censored, truncated, and sample-selected models. You may provide the probability distribution of a dependent variable in the ENDOGENOUS statement or in the DISCRETE option of the MODEL statement.

```
PROC QLIM DATA=masil.gss_cdvm;
     MODEL trust = educate income age male www;
     ENDOGENOUS trust ~ DISCRETE(DIST=LOGIT);

PROC QLIM DATA=masil.gss_cdvm;
     MODEL trust = educate income age male www /DISCRETE(DIST=LOGIT);
RUN;
```

```
                        The QLIM Procedure

                 Discrete Response Profile of trust

        Index          Value           Frequency     Percent

          1              0                   682       58.09
          2              1                   492       41.91


                        Model Fit Summary

          Number of Endogenous Variables                1
```

```
                    Endogenous Variable                      trust
                    Number of Observations                    1174
                    Log Likelihood                      -733.97164
                    Maximum Absolute Gradient            0.0000275
                    Number of Iterations                        13
                    Optimization Method               Quasi-Newton
                    AIC                                       1480
                    Schwarz Criterion                         1510


                             Goodness-of-Fit Measures

       Measure                        Value    Formula

       Likelihood Ratio (R)          128.68    2 * (LogL - LogL0)
       Upper Bound of R (U)          1596.6    - 2 * LogL0
       Aldrich-Nelson                0.0988    R / (R+N)
       Cragg-Uhler 1                 0.1038    1 - exp(-R/N)
       Cragg-Uhler 2                 0.1397    (1-exp(-R/N)) / (1-exp(-U/N))
       Estrella                       0.108    1 - (1-R/U)^(U/N)
       Adjusted Estrella             0.0981    1 - ((LogL-K)/LogL0)^(-2/N*LogL0)
       McFadden's LRI                0.0806    R / U
       Veall-Zimmermann              0.1714    (R * (U+N)) / (U * (R+N))
       McKelvey-Zavoina              0.3489

       N = # of observations, K = # of regressors

Algorithm converged.


                              Parameter Estimates

                                                Standard                 Approx
        Parameter     DF        Estimate           Error    t Value    Pr > |t|

        Intercept      1       -4.983009        0.478382     -10.42      <.0001
        educate        1        0.151581        0.026178       5.79      <.0001
        income         1        0.030349        0.011536       2.63      0.0085
        age            1        0.028015        0.004871       5.75      <.0001
        male           1        0.256796        0.125829       2.04      0.0413
        www            1        0.553738        0.165881       3.34      0.0008
```

PROC QLIM produces various goodness-of-fit measures and, unlike other procedures, reports *t* scores, which are the same as *z* score in Stata (see Section 2.1). Therefore, PROC QLIM is more comparable to Stata and LIMDEP than other alternative procedures in SAS.

PROC GENMOD provides flexible methods to estimate generalized linear and nonlinear models. The DISTRIBUTION (DIST) and the LINK=LOGIT options respectively specify a probability distribution and a link function.

```
PROC GENMOD DATA = masil.gss_cdvm DESC;
     MODEL trust = educate income age male www /DIST=BINOMIAL LINK=LOGIT;
RUN;
```

```
                              The GENMOD Procedure

                                Model Information

                Data Set              MASIL.GSS_CDVM
                Distribution                Binomial
                Link Function                  Logit
                Dependent Variable             trust    trust


                      Number of Observations Read        1174
                      Number of Observations Used        1174
                      Number of Events                    492
                      Number of Trials                   1174


                                Response Profile

                           Ordered                 Total
                             Value    trust    Frequency

                                 1    1              492
                                 2    0              682

PROC GENMOD is modeling the probability that trust='1'.


                      Criteria For Assessing Goodness Of Fit

                Criterion                    DF           Value        Value/DF

                Log Likelihood                        -733.9716
                Full Log Likelihood                   -733.9716
                AIC (smaller is better)               1479.9433
                AICC (smaller is better)              1480.0153
                BIC (smaller is better)               1510.3523


           Algorithm converged.


                   Analysis Of Maximum Likelihood Parameter Estimates

                                Standard     Wald 95% Confidence          Wald
   Parameter    DF    Estimate       Error           Limits            Chi-Square    Pr > ChiSq

   Intercept     1     -4.9830      0.4784     -5.9206     -4.0454         108.51        <.0001
   educate       1      0.1516      0.0262      0.1003      0.2029          33.53        <.0001
   income        1      0.0303      0.0115      0.0077      0.0530           6.92        0.0085
   age           1      0.0280      0.0049      0.0185      0.0376          33.08        <.0001
   male          1      0.2568      0.1258      0.0102      0.5034           4.17        0.0413
   www           1      0.5537      0.1659      0.2286      0.8789          11.14        0.0008
   Scale         0      1.0000      0.0000      1.0000      1.0000

NOTE: The scale parameter was held fixed.
```

Instead of the LINK=LOGIT option, you may provide a corresponding link function manually using the FWDLINK and INVLINK statements. The following is an example.

```
PROC GENMOD DATA = masil.gss_cdvm DESC;
     FWDLINK link=LOG(_MEAN_/(1-_MEAN_));
     INVLINK invlink=1/(1+EXP(-1*_XBETA_));
     MODEL trust = educate income age male www /DIST=BINOMIAL;
RUN;
(output is skipped)
```

2.5 Binary Logit Model in R

In R, `glm()` fits binary logit and probit models. This function returns associated statistics and functions such as `coef()` and `vcov()` in an object. Unlike Stata and SAS, R does not give you all answers with a single function. Accordingly, you need to get specific answers using statistics and functions that `glm()` returns.

Let us read a data set first using `read.table()`. The following example reads a CSV file and saves into a data frame *df*. A delimiter is specified in `sep=''` and `header=T` reads variable names from the first row. The `attach()` function adds the data frame to R search path so that variables in the data frame are accessed by their names alone (without their data frame name).

```
> df<-read.table('http://www.indiana.edu/~statmath/stat/all/cdvm/gss_cdvm.csv',
+                sep=',', header=T)
> attach(df)
```

In the `glm()` below, a dependent variable is followed by a tilde (~) and a list of independent variables separated by a plus (+) sign. The `family=` option specifies a link function. The `glm()` returns associated statistics and functions in an object *blm*. `summary(blm)` reports the summary of the estimated binary logit model.

```
> blm<-glm(trust~educate+income+age+male+www, data=df, family=binomial(link="logit"))

> summary(blm)

Call:
glm(formula = trust ~ educate + income + age + male + www, family = binomial(link = "logit"),
    data = df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8263  -0.9987  -0.6752   1.1494   2.1516

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.983009   0.478359 -10.417  < 2e-16 ***
educate      0.151581   0.026177   5.791 7.02e-09 ***
income       0.030349   0.011536   2.631 0.008522 **
age          0.028015   0.004871   5.752 8.83e-09 ***
male         0.256796   0.125829   2.041 0.041267 *
www          0.553738   0.165881   3.338 0.000843 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1596.6  on 1173  degrees of freedom
Residual deviance: 1467.9  on 1168  degrees of freedom
```

```
AIC: 1479.9

Number of Fisher Scoring iterations: 4
```

R reports the same parameter estimates, standard errors, and z scores that Stata produced. R does not, however, display goodness-of-fit measures except for AIC and, like SAS PROC LOGISTIC, returns -2*log likelihood of null and full models (see Section 2.3) instead. For instance, 1,467.9 of `Residual deviance:` is -2*log likelihood of the full model. `df.null` (=1,173) and `df.residual` (=1,168) are degrees of freedom of null and full models, respectively. Therefore, the likelihood ratio and its p-value are computed as,

```
> blm$deviance/-2
[1] -733.9716

> AIC(blm)
[1] 1479.943

> LRtest<-blm$null.deviance - blm$deviance
> LRtest
[1] 128.6811

> dchisq(LRtest, blm$df.null - blm$df.residual)
[1] 2.214737e-26
```

The likelihood ratio is 128.6811, which is large enough to reject the null hypothesis of poor fit (no difference between null and full models). McFadden's pseudo R^2 is computed on the basis of the two deviances (log likelihoods of null and full models): .0806=1-(1467.9/1596.6). Notice that a comment begins with the pound sign (#).

```
> 1-bpm$deviance/bpm$null.deviance # McFadden's pseudo R square
[1] 0.08056336
```

Now, let us compute factor changes in the odds of having success. Create vectors of means and standard deviations of covariates using `c()`, `mean()`, and `sd()`. Notice that 1 is for the intercept. *bHat* and *K* are a vector of parameter estimates and a scalar for the length of *bHat* (number of parameters).

```
> meanX<-c(1, mean(educate), mean(income), mean(age), mean(male), mean(www))
> sdX<-c(1, sd(educate), sd(income), sd(age), sd(male), sd(www))
> bHat<-coef(blm) # vector of parameter estimates
> K<-length(bHat) # the number of parameters
```

Next, compute factor changes of the odds. The following `cbind()` combines individual vectors into a matrix. `Exp(bHat*sdX)` is factor changes when covariates increase by their standard deviations. `colnames(fcOdds)` puts column names to the data frame *fcOdds*.

```
> fcOdds<-cbind(bHat, exp(bHat), exp(bHat*sdX), meanX, sdX)
> fcOdds<-fcOdds[2:K,]
> colnames(fcOdds)<-c("b", "e^b", "e^(b*sd)", "Mean of X", "SD of X")
```

The following output is very similar to what `.listcoef` produced in Section 2.2.

```
> fcOdds

                 b   exp(b) exp(b*sd)  Mean of X    SD of X
educate 0.15158121 1.163673  1.476272 14.2427598  2.5697123
income  0.03034856 1.030814  1.206818 24.6486371  6.1942699
age     0.02801520 1.028411  1.455869 41.3074957 13.4071272
```

```
male     0.25679598 1.292781  1.136353  0.4505963  0.4977653
www      0.55373840 1.739745  1.255396  0.7853492  0.4107548
```

Finally, compute marginal effects at the same reference points. `%*%` below obtains the element by element product, a scalar of *xb* in this case. The scalar *prob* contains the predicted probability of 47.53 percent that female WWW users with 16 years of education (educate=16, male=0, and www=1) trust most people, holding other covariates at their means.

```
> referX<-c(1, 16, mean(income), mean(age), 0, 1) # set reference points
> xb<-bHat %*% referX # element by element product

> prob<-exp(xb)/(1+exp(xb)) # compute a pridicted probability
> prob
          [,1]
[1,] 0.4753492
```

Marginal effects are $\Lambda(x\beta)(1-\Lambda(x\beta)\beta_c$ in the binary logit model. When covariates increase by their standard deviations from the reference points, the marginal effects are `prob*(1-prob)*bHat*sdX`. Compare the following result with what `.prchange` computed in Section 2.2 and the PROC IML output in Section 2.3. Notice that .0640 and .1381 below are not discrete changes of gender and WWW use. See Section 3.4 for computing discrete changes.

```
> margEffect<-cbind(bHat, prob*(1-prob)*bHat, prob*(1-prob)*bHat*sdX, meanX,sdX)
> margEffect<-margEffect[2:K,]
> colnames(margEffect)<-c("b", "MargEffect", "MargEffect(SD)", "Mean of X", "SD of X")
> margEffect

                 b  MargEffect MargEffect(SD)  Mean of X    SD of X
educate 0.15158121 0.037803193     0.09714333 14.2427598  2.5697123
income  0.03034856 0.007568699     0.04688256 24.6486371  6.1942699
age     0.02801520 0.006986775     0.09367259 41.3074957 13.4071272
male    0.25679598 0.064042951     0.03187836  0.4505963  0.4977653
www     0.55373840 0.138098116     0.05672447  0.7853492  0.4107548
```

2.6 Binary Logit Model in LIMDEP (Logit$)

LIMDEP can read data in the ASCII text (CSV) and Excel format. The following script clears the worksheet (`RESET$`), defines data size (`ROWS;999999$`), and then reads an Excel file `gss_cdvm.xls`. Notice that each command ends with $ and subcommands are separated by a semi-colon.

```
RESET$
ROWS;999999$
READ;FILE="C:\Temp\Limdep\gss_cdvm.xls"$
```

The `Logit$` command estimates various logit models in LIMDEP. A dependent variable is specified in the `Lhs=` (left-hand side) subcommand and a list of independent variables in the `Rhs=` (right-hand side). You have to explicitly specify `ONE` for the intercept.

```
LOGIT;Lhs=TRUST;
      Rhs=ONE,EDUCATE,INCOME,AGE,MALE,WWW$

Normal exit from iterations. Exit status=0.

+---------------------------------------------+
| Binary Logit Model for Binary Choice        |
| Maximum Likelihood Estimates                |
```

```
| Model estimated: Sep 09, 2009 at 04:25:56PM.|
| Dependent variable               TRUST     |
| Weighting variable                None     |
| Number of observations            1174     |
| Iterations completed                 5     |
| Log likelihood function      -733.9716     |
| Number of parameters                 6     |
| Info. Criterion: AIC =         1.26060     |
|   Finite Sample: AIC =         1.26066     |
| Info. Criterion: BIC =         1.28650     |
| Info. Criterion:HQIC =         1.27037     |
| Restricted log likelihood    -798.3122     |
| McFadden Pseudo R-squared     .0805957     |
| Chi squared                   128.6811     |
| Degrees of freedom                   5     |
| Prob[ChiSqd > value] =        .0000000     |
| Hosmer-Lemeshow chi-squared =   3.64573    |
| P-value=  .88759 with deg.fr. =      8     |
+----------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
---------+Characteristics in numerator of Prob[Y = 1]
 Constant|   -4.98300913       .47835906   -10.417   .0000
 EDUCATE |     .15158121       .02617738     5.791   .0000   14.2427598
 INCOME  |     .03034856       .01153642     2.631   .0085   24.6486371
 AGE     |     .02801520       .00487072     5.752   .0000   41.3074957
 MALE    |     .25679598       .12582872     2.041   .0413    .45059625
 WWW     |     .55373840       .16588151     3.338   .0008    .78534923

+----------------------------------------------------------------------+
| Information Statistics for Discrete Choice Model.                    |
|                            M=Model MC=Constants Only   M0=No Model |
| Criterion F (log L)     -733.97164       -798.31217    -813.75479 |
| LR Statistic vs. MC      128.68107           .00000        .00000 |
| Degrees of Freedom         5.00000           .00000        .00000 |
| Prob. Value for LR          .00000           .00000        .00000 |
| Entropy for probs.       733.97164        798.31217     813.75479 |
| Normalized Entropy          .90196           .98102       1.00000 |
| Entropy Ratio Stat.      159.56630         30.88523        .00000 |
| Bayes Info Criterion       1.28048          1.39009       1.41640 |
| BIC(no model) - BIC         .13592           .02631        .00000 |
| Pseudo R-squared            .08060           .00000        .00000 |
| Pct. Correct Pred.        65.41738           .00000      50.00000 |
| Means:       y=0    y=1    y=2    y=3    y=4    y=5     y=6   y>=7 |
| Outcome     .5809  .4191  .0000  .0000  .0000  .0000  .0000  .0000 |
| Pred.Pr     .5809  .4191  .0000  .0000  .0000  .0000  .0000  .0000 |
| Notes: Entropy computed as Sum(i)Sum(j)Pfit(i,j)*logPfit(i,j).      |
|        Normalized entropy is computed against M0.                   |
|        Entropy ratio statistic is computed against M0.              |
|        BIC = 2*criterion - log(N)*degrees of freedom.               |
|        If the model has only constants or if it has no constants,  |
|        the statistics reported here are not useable.               |
+----------------------------------------------------------------------+
+------------------------------------------+
| Fit Measures for Binomial Choice Model |
| Logit    model for variable TRUST      |
+------------------------------------------+
| Proportions P0= .580920   P1= .419080  |
| N =    1174 N0=     682   N1=     492  |
| LogL=     -733.972 LogL0=    -798.312  |
| Estrella = 1-(L/L0)^(-2L0/n) = .10799  |
+------------------------------------------+
|     Efron |  McFadden  |  Ben./Lerman  |
|    .10474 |    .08060  |       .56407  |
|    Cramer | Veall/Zim. |     Rsqrd_ML  |
|    .10469 |    .17142  |       .10382  |
+------------------------------------------+
| Information  Akaike I.C. Schwarz I.C.  |
| Criteria        1.26060       1.28650  |
+------------------------------------------+
```

```
+-----------------------------------------------------------+
|Predictions for Binary Choice Model.  Predicted value is   |
|1 when probability is greater than  .500000, 0 otherwise.|
|Note, column or row total percentages may not sum to       |
|100% because of rounding. Percentages are of full sample.|
+------+---------------------------------+----------------+
|Actual|         Predicted Value         |                |
|Value |       0                1        | Total Actual   |
+------+----------------+----------------+----------------+
|  0   |    538 ( 45.8%)|    144 ( 12.3%)|    682 ( 58.1%)|
|  1   |    262 ( 22.3%)|    230 ( 19.6%)|    492 ( 41.9%)|
+------+----------------+----------------+----------------+
|Total |    800 ( 68.1%)|    374 ( 31.9%)|   1174 (100.0%)|
+------+----------------+----------------+----------------+

=======================================================================
Analysis of Binary Choice Model Predictions Based on Threshold =  .5000
-----------------------------------------------------------------------
Prediction Success
-----------------------------------------------------------------------
Sensitivity = actual 1s correctly predicted                      46.748%
Specificity = actual 0s correctly predicted                      78.886%
Positive predictive value = predicted 1s that were actual 1s    61.497%
Negative predictive value = predicted 0s that were actual 0s    67.250%
Correct prediction = actual 1s and 0s correctly predicted       65.417%
-----------------------------------------------------------------------
Prediction Failure
-----------------------------------------------------------------------
False pos. for true neg. = actual 0s predicted as 1s            21.114%
False neg. for true pos. = actual 1s predicted as 0s            53.252%
False pos. for predicted pos. = predicted 1s actual 0s          38.503%
False neg. for predicted neg. = predicted 0s actual 1s          32.750%
False predictions = actual 1s and 0s incorrectly predicted      34.583%
=======================================================================
```

Stata, SAS, and LIMDEP produce the same result. The likelihood ratio is 128.6811=-2*[(-798.3122)-(-733.9716)]. While SAS reports AIC*N=1,479.9433, LIMDEP returns an AIC of 1.2606 (=1,479.943/1,174). BIC (Schwarz IC) is 1510.351=1.2865*1174. In order to compute marginal effects, add the `Marginal Effects` and `Means` subcommands to `Logit$`. The following script computes marginal effects at the mean values of independent variables. Other parts in the output are skipped.

```
LOGIT;Lhs=TRUST;
      Rhs=ONE,EDUCATE,INCOME,AGE,MALE,WWW;
      Marginal Effects; Means$
```

```
+---------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|Elasticity|
+---------+--------------+----------------+--------+--------+----------+
---------+Marginal effect for variable in probability
 Constant|   -1.20446697       .11302276   -10.657   .0000
 EDUCATE |     .03663942       .00632491     5.793   .0000   1.27598047
 INCOME  |     .00733570       .00278319     2.636   .0084    .44211529
 AGE     |     .00677169       .00117650     5.756   .0000    .68395424
---------+Marginal effect for dummy variable is P|1 - P|0.
 MALE    |     .06213506       .03043408     2.042   .0412    .06845822
---------+Marginal effect for dummy variable is P|1 - P|0.
 WWW     |     .12861867       .03653176     3.521   .0004    .24698361

+---------------------+
| Marginal Effects for|
+----------+----------+
| Variable | All Obs. |
+----------+----------+
| ONE      | -1.20447 |
| EDUCATE  |   .03664 |
| INCOME   |   .00734 |
```

```
| AGE      |   .00677 |
| MALE     |   .06214 |
| WWW      |   .12862 |
+----------+----------+
```

In order to compare marginal effects computed in Stata and LIMDEP, let us run `.prchange` in Stata without reference points specified. `quietly` before a command run the command but suppresses the output. Stata and LIMDEP produce the same marginal effects (e.g., .0366 for education) and discrete changes (e.g., .1286 for WWW use). Notice that marginal effects and discrete changes vary depending on reference points used (compare with marginal effects in Section 2.2).

```
. quietly logit trust educate income age male www

. prchange

logit: Changes in Probabilities for trust

         min->max      0->1     -+1/2    -+sd/2  MargEfct
educate    0.5259    0.0111    0.0366    0.0939    0.0366
 income    0.1805    0.0057    0.0073    0.0454    0.0073
    age    0.4428    0.0041    0.0068    0.0905    0.0068
   male    0.0621    0.0621    0.0620    0.0309    0.0621
    www    0.1286    0.1286    0.1331    0.0549    0.1338

              0       1
Pr(y|x)  0.5910  0.4090

       educate   income      age     male      www
   x=  14.2428  24.6486  41.3075  .450596  .785349
sd_x=  2.56971  6.19427  13.4071  .497765  .410755
```

2.7 Binary Logit Model in SPSS

In SPSS, the `Logistic Regression` command fits the binary logit model. SPSS generates messy tables, which are often overwhelming for beginners. The tables below are selected from the entire output.

```
LOGISTIC REGRESSION VARIABLES trust
   /METHOD=ENTER educate income age male www
   /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1467.943[a]	.104	.140

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Variables in the Equation

	B	S.E.	Wald	Df	Sig.	Exp(B)

Step 1[a]	educate	.152	.026	33.530	1	.000	1.164
	income	.030	.012	6.920	1	.009	1.031
	age	.028	.005	33.083	1	.000	1.028
	male	.257	.126	4.165	1	.041	1.293
	www	.554	.166	11.143	1	.001	1.740
	Constant	-4.983	.478	108.511	1	.000	.007

a. Variable(s) entered on step 1: educate, income, age, male, www.

SPSS returns the same parameter estimates and their standard errors. Like SAS PROC LOGISTIC, SPSS reports -2*Log-likelihood (1,467.943=-2*733.9716) and Wald statistics. P-values are listed under the label `Sig.` and factor changes in odds under `Exp(B)`. SPSS does not produce Pseudo R^2, AIC, Schwarz, and BIC.

Table 2.1 summarizes parameter estimates and goodness-of-fit measures of the binary logit model produced in Stata, SAS, R. and LIMDEP, excluding the output of PROC PROBIT and SPSS. Parameter estimates, their standard errors, and goodness-of-fit measures are identical except for some rounding errors. Stata, R, and LIMDEP report *z* scores for hypothesis test, while PROC QLIM returns *t* scores and LOGISTIC, GENMOD, and PROBIT procedures conduct chi-square tests. PROC LOGISTIC and Stata `.logit` with SPost are general recommended.

Table 2.1. Parameter Estimates and Goodness-of-fit of the Binary Logit Model

	SAS			Stata	R	LIMDEP
	LOGISTIC	QLIM	GENMOD	`.logit`	`glm()`	`Logit$`
Education	.1516 (.0262)	.1516 (.0262)	.1516 (.0262)	.1516 (.0262)	.1516 (.0262)	.1516 (.0262)
Family income	.0303 (.0115)	.0303 (.0115)	.0303 (.0115)	.0303 (.0115)	.0303 (.0115)	.0303 (.0115)
Age	.0280 (.0049)	.0280 (.0049)	.0280 (.0049)	.0280 (.0049)	.0280 (.0049)	.0280 (.0049)
Gender (male)	.2568 (.1258)	.2568 (.1258)	.2568 (.1258)	.2568 (.1258)	.2568 (.1258)	.2568 (.1258)
WWW use	.5537 (.1659)	.5537 (.1659)	.5537 (.1659)	.5537 (.1659)	.5537 (.1659)	.5537 (.1659)
Intercept	-4.9830 (.4784)	-4.9830 (.4784)	-4.9830 (.4784)	-4.9830 (.4784)	-4.9830 (.4784)	-4.9830 (.4784)
Log likelihood	-733.9716	-733.9716	-733.9716	-733.9716	-733.9716	-733.9716
Likelihood test	128.6811	128.68		128.68	128.6811	128.6811
Pseudo R^2	.0806	.0806		.0806	.0806	.0806
AIC	1479.943	1480.	1479.9433	*1479.943*	1479.943	1479.944
BIC (Schwarz)	1510.352	1510.	1510.3523	*1510.352*		1510.352
H_0 test	Chi-square	t	Chi-square	z	z	z

* PROC LOGISTIC and R report (-2*Log-likelihood).
** AIC*N and BIC*N in Stata and LIMDEP

3. Binary Probit Regression Model

The probit model is represented as $\text{Prob}(y=1|x)=\Phi(x\beta)$, where Φ indicates the cumulative standard normal probability distribution function. Let us fit the binary probit model to see if there is substantial difference between binary logit and probit models.

3.1 Binary Probit Model in Stata (.probit)

Stata `.probit` estimates the binary probit regression model. If you want to get robust standard errors, add the `robust` option to `.logit` and `.probit`. The logit and probit models produce almost similar goodness-of-fit measures but their parameter estimates differ.

```
. probit trust educate income age male www

Iteration 0:   log likelihood = -798.31217
Iteration 1:   log likelihood = -734.10951
Iteration 2:   log likelihood = -733.99746
Iteration 3:   log likelihood = -733.99746

Probit regression                                 Number of obs   =       1174
                                                  LR chi2(5)      =     128.63
                                                  Prob > chi2     =     0.0000
Log likelihood = -733.99746                       Pseudo R2       =     0.0806

------------------------------------------------------------------------------
       trust |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     educate |   .0907207   .0154349     5.88   0.000     .0604689    .1209725
      income |   .0185906   .0068681     2.71   0.007     .0051293    .0320519
         age |   .0173105   .0029496     5.87   0.000     .0115293    .0230916
        male |   .1593935   .0768819     2.07   0.038     .0087077    .3100793
         www |   .3417645   .0992156     3.44   0.001     .1473055    .5362235
       _cons |  -3.030053   .2786062   -10.88   0.000    -3.576111   -2.483995
------------------------------------------------------------------------------
```

The standard normal probability distribution and standard logistic distribution respectively have a unit variance and a variance of $\pi^2/3$. Therefore, a parameter estimate in a binary logit model is about 1.8138 $(=\pi/\sqrt{3})$ larger than its corresponding coefficient in its probit counterpart. Long's suggestion is 1.7 (Long 1997: 48). For instance, the coefficient of education in the binary logit model is .1516, which is similar to .1542 (1.7*.0907). See Cameron and Trivedi (2009: 451-452) for discussion on parameter estimates across models (OLS, binary logit, and binary probit model).

```
. di _pi/sqrt(3)*.0907207
.16454915

. di 1.7*.0907207
.15422519
```

Goodness-of-fit measures are very similar to those of the logit model. Log likelihoods are -733.972 and -733.997 and likelihood ratios are 128.681 and 128.629 in binary logit and probit models, respectively. They produce the same pseudo R^2 of .0806.

```
. fitstat

Measures of Fit for probit of trust
```

```
Log-Lik Intercept Only:       -798.312   Log-Lik Full Model:           -733.997
D(1168):                      1467.995   LR(5):                         128.629
                                         Prob > LR:                       0.000
McFadden's R2:                   0.081   McFadden's Adj R2:               0.073
ML (Cox-Snell) R2:               0.104   Cragg-Uhler(Nagelkerke) R2:      0.140
McKelvey & Zavoina's R2:         0.166   Efron's R2:                      0.105
Variance of y*:                  1.199   Variance of error:               1.000
Count R2:                        0.652   Adj Count R2:                    0.171
AIC:                             1.261   AIC*n:                        1479.995
BIC:                         -6787.630   BIC':                          -93.289
BIC used by Stata:            1510.404   AIC used by Stata:            1479.995
```

In order to get standardized estimates, run SPost's `.listcoef` command. A coefficient is the impact of an independent variable for a unit increase in that variable, while the corresponding number under `bStdX` is the impact of the covariate for a standard deviation increase in that variable. For example, the x-standardized coefficient of education is .2331 (=.0907*2.5697). Notice that factor changes in odds by definition are not available in a probit model.

```
. listcoef, help

probit (N=1174): Unstandardized and Standardized Estimates

 Observed SD: .49361879
   Latent SD: 1.0952088

-------------------------------------------------------------------------------
       trust |      b         z     P>|z|    bStdX    bStdY   bStdXY      SDofX
-------------+-----------------------------------------------------------------
     educate |   0.09072    5.878   0.000   0.2331   0.0828   0.2129     2.5697
      income |   0.01859    2.707   0.007   0.1152   0.0170   0.1051     6.1943
         age |   0.01731    5.869   0.000   0.2321   0.0158   0.2119    13.4071
        male |   0.15939    2.073   0.038   0.0793   0.1455   0.0724     0.4978
         www |   0.34176    3.445   0.001   0.1404   0.3121   0.1282     0.4108
-------------------------------------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
   bStdX = x-standardized coefficient
   bStdY = y-standardized coefficient
  bStdXY = fully standardized coefficient
   SDofX = standard deviation of X
```

The discrete change of a binary variable remains unchanged in the binary probit model, but the marginal effect of a continuous independent variable in the binary probit model is defined as,

$$\frac{\partial P(y=1|x)}{\partial x_c} = \phi(x\beta)\beta_c$$

where ϕ denotes the standard normal probability density function.

You may compute marginal effects and discrete changes using either `.mfx` or SPost's `.prchange`. Marginal effects and discrete changes in the logit and probit models, despite different parameter estimates, are very similar (.0378 versus .0361 for education and .1329 versus .1320 for WWW use). Also two models return the similar predicted probability at the same reference points (.4753 versus .4747).

```
. mfx, at(mean educate=16 male=0 www=1)

Marginal effects after probit
```

```
      y  = Pr(trust) (predict)
         =  .47469509
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
 educate |   .0361195      .00681    5.30   0.000   .022774  .049465         16
  income |   .0074017      .00264    2.81   0.005   .002234  .012569   24.6486
     age |    .006892      .00118    5.83   0.000   .004574   .00921   41.3075
   male*|   .0635132      .03058    2.08   0.038   .003573  .123453          0
    www*|   .1320435       .0374    3.53   0.000   .058748  .205339          1
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

```
. prchange, x(educate=16 male=0 www=1) rest(mean)

probit: Changes in Probabilities for trust

         min->max      0->1     -+1/2    -+sd/2  MargEfct
educate    0.5265    0.0123    0.0361    0.0926    0.0361
 income    0.1916    0.0065    0.0074    0.0458    0.0074
    age    0.4409    0.0051    0.0069    0.0922    0.0069
   male    0.0635    0.0635    0.0634    0.0316    0.0635
    www    0.1320    0.1320    0.1354    0.0558    0.1361

             0       1
Pr(y|x)  0.5253  0.4747

       educate   income      age     male      www
   x=       16  24.6486  41.3075        0        1
sd_x=  2.56971  6.19427  13.4071  .497765  .410755
```

Similarly, `.prtab` and `.prvalue` report same predicted probabilities at the same reference points. Compare the following result with the output presented in Section 2.2.

```
. prtab male www, x(educate=16 male=0 www=1) rest(mean)

probit: Predicted probabilities of positive outcome for trust

--------------------------------
          |       WWW Use
   Gender | Non-users      Users
----------+---------------------
   Female |    0.3427     0.4747
     Male |    0.4029     0.5382
--------------------------------

      educate     income        age       male        www
x=         16  24.648637  41.307496          0          1
```

```
. prvalue, x(educate=16 male=0 www=1) rest(mean)

probit: Predictions for trust

Confidence intervals by delta method

                                95% Conf. Interval
  Pr(y=1|x):          0.4747   [ 0.4281,    0.5213]
  Pr(y=0|x):          0.5253   [ 0.4787,    0.5719]

      educate     income        age       male        www
x=         16  24.648637  41.307496          0          1
```

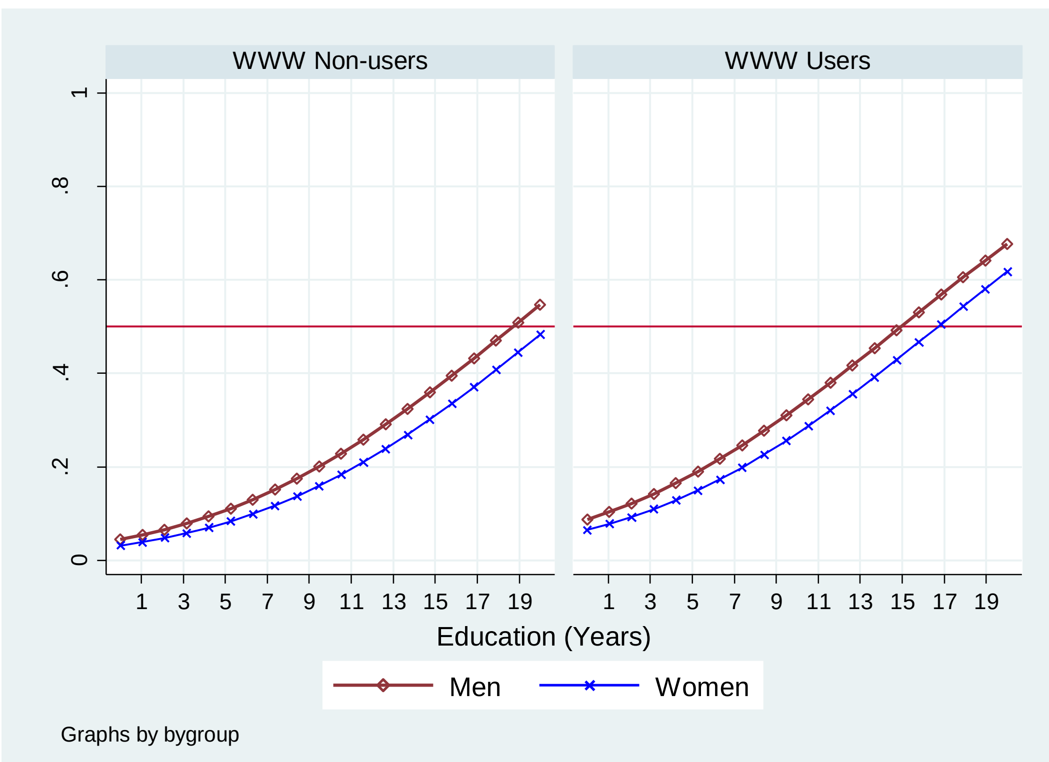
Finally, let us draw a plot of predicted probabilities using `.prgen`. We are using the same reference points and same range of education (0 to 20) to get Figure 3.1. See Appendix for the Stata script used.

```
. quietly probit trust educate income age male www
```

```
. prgen educate, from(0) to(20) ncases(20) x(male=1 www=1) rest(mean) gen(Probit_age11)

probit: Predicted values as educate varies from 0 to 20.

      educate      income         age        male         www
x=   14.24276   24.648637   41.307496           1           1

. prgen educate, from(0) to(20) ncases(20) x(male=1 www=0) rest(mean) gen(Probit_age10)

probit: Predicted values as educate varies from 0 to 20.

      educate      income         age        male         www
x=   14.24276   24.648637   41.307496           1           0

. prgen educate, from(0) to(20) ncases(20) x(male=0 www=1) rest(mean) gen(Probit_age01)

probit: Predicted values as educate varies from 0 to 20.

      educate      income         age        male         www
x=   14.24276   24.648637   41.307496           0           1

. prgen educate, from(0) to(20) ncases(20) x(male=0 www=0) rest(mean) gen(Probit_age00)

probit: Predicted values as educate varies from 0 to 20.

      educate      income         age        male         www
x=   14.24276   24.648637   41.307496           0           0
```

Compare Figure 2.1 and 3.1 to find they are almost identical. This finding is not surprising at all because predicted probabilities, marginal effects, and discrete changes are very similar in binary logit and probit models, although two models produce different parameter estimates and standard errors.

Figure 3.1 Predicted Probabilities of Trusting Most People (Binary Probit Model)



3.2 Binary Probit Model in SAS: PROC PROBIT and PROC LOGISTIC

PROBIT and LOGISTIC procedures estimate the binary probit model. Keep in mind that the coefficients of PROC PROBIT have opposite signs. Stata and SAS produce the same result.

```
PROC PROBIT DATA = masil.gss_cdvm;
     MODEL trust = educate income age male www;
RUN;
```

```
                                      The Probit Procedure

                                        Model Information

              Data Set                    MASIL.GSS_CDVM
              Dependent Variable                   trust    trust
              Number of Observations                1174
              Name of Distribution                Normal
              Log Likelihood                -733.9974633


                             Number of Observations Read        1174
                             Number of Observations Used        1174


                                     Class Level Information

                                   Name       Levels    Values

                                   trust           2    0 1


                                         Response Profile

                                  Ordered                 Total
                                    Value     trust    Frequency

                                        1     0              682
                                        2     1              492

PROC PROBIT is modeling the probabilities of levels of trust having LOWER Ordered Values in the
response profile table.


           Algorithm converged.


                                  Type III Analysis of Effects

                                                  Wald
                           Effect        DF    Chi-Square    Pr > ChiSq

                           educate        1       34.5467        <.0001
                           income         1        7.3266        0.0068
                           age            1       34.4417        <.0001
                           male           1        4.2983        0.0382
```

```
                      www             1        11.8657        0.0006


                 Analysis of Maximum Likelihood Parameter Estimates

                               Standard    95% Confidence      Chi-
        Parameter  DF Estimate    Error        Limits        Square Pr > ChiSq

        Intercept   1   3.0300   0.2786   2.4840   3.5761   118.28     <.0001
        educate     1  -0.0907   0.0154  -0.1210  -0.0605    34.55     <.0001
        income      1  -0.0186   0.0069  -0.0321  -0.0051     7.33     0.0068
        age         1  -0.0173   0.0029  -0.0231  -0.0115    34.44     <.0001
        male        1  -0.1594   0.0769  -0.3101  -0.0087     4.30     0.0382
        www         1  -0.3418   0.0992  -0.5362  -0.1473    11.87     0.0006
```

PROC LOGISTIC requires a normal probability distribution as a link function (/LINK=PROBIT or /LINK=NORMIT) to fit a binary probit model. McFadden's pseudo R^2 is .0806=1-(.1467.995/1596.624). OUTEST stores parameter estimates into a SAS data set *masil.bpm*, which will be used when computing marginal effects later.

```
PROC LOGISTIC DATA = masil.gss_cdvm DESC OUTEST=masil.bpm;
     MODEL trust = educate income age male www /LINK=PROBIT;
RUN;
```

```
                            The LOGISTIC Procedure

                              Model Information

              Data Set                      MASIL.GSS_CDVM
              Response Variable             trust                 trust
              Number of Response Levels     2
              Model                         binary probit
              Optimization Technique        Fisher's scoring


                       Number of Observations Read        1174
                       Number of Observations Used        1174


                               Response Profile

                          Ordered                      Total
                            Value         trust      Frequency

                                1             1            492
                                2             0            682

                         Probability modeled is trust=1.


                           Model Convergence Status

                 Convergence criterion (GCONV=1E-8) satisfied.


                             Model Fit Statistics
```

Criterion	Intercept Only	Intercept and Covariates
AIC	1598.624	1479.995
SC	1603.693	1510.404
-2 Log L	1596.624	1467.995

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	128.6294	5	<.0001
Score	121.5344	5	<.0001
Wald	118.2980	5	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.0298	0.2796	117.4048	<.0001
educate	1	0.0907	0.0158	32.9144	<.0001
income	1	0.0186	0.00682	7.4273	0.0064
age	1	0.0173	0.00295	34.3163	<.0001
male	1	0.1594	0.0769	4.2979	0.0382
www	1	0.3418	0.0995	11.7914	0.0006

Association of Predicted Probabilities and Observed Responses

Percent Concordant	68.4	Somers' D	0.371
Percent Discordant	31.3	Gamma	0.372
Percent Tied	0.4	Tau-a	0.181
Pairs	335544	c	0.686

Stata, PROC LOGISTIC, and PROC PROBIT share the same parameter estimates, but PROC LOGISTIC reports slightly different standard errors (e.g., .0158 versus .0154 for education). The following script fits the same model using /LINK=NORMIT and stores the SAS output in an HTML file `c:\temp\sas\logit.html` using ODS.

```
ODS HTML FILE='c:\temp\sas\probit.html';
PROC LOGISTIC DATA = masil.gss_cdvm DESC;
   MODEL trust(EVENT='1') = educate income age male www /LINK=NORMIT;
RUN;
ODS HTML CLOSE;
```

Let us compute marginal effects using SAS/IML. We stored parameter estimates in *masil.bpm*. The following SAS script highlights the only parts different from the PROC IML in Section 2.3. PROBNORM()=CDF('NORMAL') and PDF('NORMAL') are respectively CDF and PDF of the standard normal distribution.

```
PROC IML;
USE masil.bpm; /* get a row vector of parameter estimates */
READ ALL VAR{Intercept educate income age male www} INTO bHat;
K=NCOL(bHat); /* get the number of regressors */
...

prob = PROBNORM(xb); /* compute a predicted probability */
...

margin = PDF('NORMAL', xb, 0, 1) * T(bHat); /* compute marginal effects */
marginSD = PDF('NORMAL', xb, 0, 1) * T(bHat # sdX);
...

QUIT; /* terminate PROC IML */
```

The predicted probability that female Internet users will trust people is 47.47 percent, holding other covariates at their means. Calculated marginal effects are the same as what `.prchange` returned in Section 3.1.

```
          referX                                                      prob

               1         16 24.648637 41.307496         0         1 0.4746975


                                      result
                         b MargEffect MargEffect(SD) Mean of X   SD of X

          educate 0.0907156  0.0361175      0.0928116  14.24276 2.5697123
          income  0.0185849  0.0073994      0.0458338 24.648637 6.1942699
          age     0.0173094  0.0068915      0.0923958 41.307496 13.407127
          male    0.1593898  0.0634594      0.0315879 0.4505963 0.4977653
          www     0.3417757  0.1360745      0.0558932 0.7853492 0.4107548
```

3.3 Binary Probit Model in SAS: PROC QLIM and PROC GENMOD

PROC QLIM provides various goodness-of-fit statistics. The DIST=NORMAL option below indicates the normal probability distribution to be used in estimation. Compared to PROC LOGISTIC, PROC QLIM reports same parameter estimates and goodness-of-fit statistics but slightly different standard errors.

```
PROC QLIM DATA=masil.gss_cdvm;
     MODEL trust = educate income age male www /DISCRETE (DIST=NORMAL);
RUN;
```

```
                                The QLIM Procedure

                        Discrete Response Profile of trust

                Index          Value            Frequency     Percent

                  1              0                    682       58.09
                  2              1                    492       41.91


                                Model Fit Summary
```

```
                    Number of Endogenous Variables                1
                    Endogenous Variable                       trust
                    Number of Observations                     1174
                    Log Likelihood                       -733.99746
                    Maximum Absolute Gradient               0.00200
                    Number of Iterations                         11
                    Optimization Method                Quasi-Newton
                    AIC                                        1480
                    Schwarz Criterion                          1510


                               Goodness-of-Fit Measures

             Measure                      Value    Formula

             Likelihood Ratio (R)        128.63    2 * (LogL - LogL0)
             Upper Bound of R (U)        1596.6    - 2 * LogL0
             Aldrich-Nelson              0.0987    R / (R+N)
             Cragg-Uhler 1               0.1038    1 - exp(-R/N)
             Cragg-Uhler 2               0.1396    (1-exp(-R/N)) / (1-exp(-U/N))
             Estrella                    0.1079    1 - (1-R/U)^(U/N)
             Adjusted Estrella            0.098    1 - ((LogL-K)/LogL0)^(-2/N*LogL0)
             McFadden's LRI              0.0806    R / U
             Veall-Zimmermann            0.1714    (R * (U+N)) / (U * (R+N))
             McKelvey-Zavoina            0.1662

             N = # of observations, K = # of regressors

Algorithm converged.


                                Parameter Estimates

                                               Standard                 Approx
              Parameter    DF       Estimate      Error    t Value    Pr > |t|

              Intercept     1      -3.030053   0.278616     -10.88      <.0001
              educate       1       0.090721   0.015435       5.88      <.0001
              income        1       0.018591   0.006868       2.71      0.0068
              age           1       0.017310   0.002950       5.87      <.0001
              male          1       0.159393   0.076882       2.07      0.0382
              www           1       0.341764   0.099215       3.44      0.0006
```

PROC GENMOD estimates the binary probit model using the /DIST=BINOMIAL and /LINK=PROBIT options in the MODEL statement. Again, DESC uses a larger value as a positive event (success). PROC QLIM and PROC GENMOD return the same parameter estimates, standard errors, and goodness-of-fit measures.

```
PROC GENMOD DATA = masil.gss_cdvm DESC;
     MODEL trust = educate income age male www /DIST=BINOMIAL LINK=PROBIT;
RUN;
```

```
                                 The GENMOD Procedure

                                  Model Information
```

```
              Data Set                MASIL.GSS_CDVM
              Distribution                  Binomial
              Link Function                   Probit
              Dependent Variable               trust    trust


                           Number of Observations Read        1174
                           Number of Observations Used        1174
                           Number of Events                    492
                           Number of Trials                   1174


                                      Response Profile

                                Ordered                 Total
                                  Value    trust    Frequency

                                      1    1              492
                                      2    0              682

PROC GENMOD is modeling the probability that trust='1'.


                           Criteria For Assessing Goodness Of Fit

              Criterion                    DF           Value        Value/DF

              Log Likelihood                        -733.9975
              Full Log Likelihood                   -733.9975
              AIC (smaller is better)               1479.9949
              AICC (smaller is better)              1480.0669
              BIC (smaller is better)               1510.4040


          Algorithm converged.


                      Analysis Of Maximum Likelihood Parameter Estimates

                                    Standard     Wald 95% Confidence          Wald
   Parameter     DF     Estimate       Error           Limits             Chi-Square    Pr > ChiSq

   Intercept      1      -3.0301      0.2786     -3.5761     -2.4840          118.28        <.0001
   educate        1       0.0907      0.0154      0.0605      0.1210           34.55        <.0001
   income         1       0.0186      0.0069      0.0051      0.0321            7.33        0.0068
   age            1       0.0173      0.0029      0.0115      0.0231           34.44        <.0001
   male           1       0.1594      0.0769      0.0087      0.3101            4.30        0.0382
   www            1       0.3418      0.0992      0.1473      0.5362           11.87        0.0006
   Scale          0       1.0000      0.0000      1.0000      1.0000

NOTE: The scale parameter was held fixed.
```

3.4 Binary Probit Model in R

The `glm()` function fits the binary probit model with `family=binomial(link="probit")`.

```
> bpm<-glm(trust~educate+income+age+male+www, data=df, family=binomial(link="probit"))

> summary(bpm)

Call:
glm(formula = trust ~ educate + income + age + male + www, family = binomial(link = "probit"),
    data = df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8299  -1.0033  -0.6756   1.1496   2.1831

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.030037   0.279632 -10.836  < 2e-16 ***
educate      0.090719   0.015812   5.737 9.63e-09 ***
income       0.018591   0.006820   2.726 0.006410 **
age          0.017311   0.002955   5.858 4.68e-09 ***
male         0.159394   0.076884   2.073 0.038157 *
www          0.341768   0.099532   3.434 0.000595 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1596.6  on 1173  degrees of freedom
Residual deviance: 1468.0  on 1168  degrees of freedom
AIC: 1480

Number of Fisher Scoring iterations: 4
```

Parameter estimates are the same across Stata, PROC LOGISTIC, and PROC QLIM. R and PROC LOGISTIC have the same standard errors, which are slightly different from those of Stata, PROC QLIM, PROC GENMOD, and PROC PROBIT. Let us conduct the likelihood ratio test using deviances of the null and full models. The pseudo R^2 .0806 is also computed from the two deviances.

```
> bpm$deviance/-2
[1] -733.9975

> AIC(bpm)
[1] 1479.995

> LRtest<-bpm$null.deviance-blm$deviance
> LRtest
[1] 128.6811

> dchisq(LRtest, bpm$df.null - bpm$df.residual)
[1] 2.214737e-26

> 1-bpm$deviance/bpm$null.deviance # McFadden's pseudo R square
[1] 0.08056336
```

In order to get the predicted probability, use the same script except for the cumulative standard normal distribution function (CDF) `pnorm()`. The predicted probability is 47.47 percent at the same reference points.

```
> bHat<-coef(bpm) # vector of parameter estimates
> K<-length(bHat) # the number of regressors

> referX<-c(1, 16, mean(income), mean(age), 0, 1)
> xb<-bHat %*% referX # element by element product
> prob<-pnorm(xb)
> prob
          [,1]
```

```
[1,] 0.4746947
```

When calculating marginal effects in the binary probit model, use the standard normal probability density function (PDF) `dnorm()`. The following `for()` loop sets two reference points of 0 and 1 and computes the difference of the two predicted probabilities.

```
> margin<-cbind(bHat, dnorm(xb)*bHat, dnorm(xb)*bHat*sdX, meanX, sdX)

> for (i in c(5, 6)) { # locations of binary variables
+     referX0<-matrix(referX)
+     referX1<-matrix(referX)
+     referX0[i,1]<-0
+     referX1[i,1]<-1
+
+     xb0<-bHat %*% referX0
+     xb1<-bHat %*% referX1
+
+     dChange<-pnorm(xb1)-pnorm(xb0)
+     margEffect[i,2]<-dChange # replace the marginal effect with the discrete change
+ }
>
> margEffect<-margEffect[2:K,]
> colnames(margEffect)<-c("b", "MargEffect", "MargEffect(SD)", "Mean of X", "SD of X")
> margEffect
                 b  MargEffect MargEffect(SD)  Mean of X    SD of X
educate 0.09071919 0.036118888     0.09281515 14.2427598  2.5697123
income  0.01859065 0.007401671     0.04584795 24.6486371  6.1942699
age     0.01731051 0.006891997     0.09240188 41.3074957 13.4071272
male    0.15939356 0.063513240     0.03158862  0.4505963  0.4977653
www     0.34176814 0.132044777     0.05589197  0.7853492  0.4107548
```

Compare above marginal effects with the results of `.prchange` in Section 3.1 and PROC IML in Section 3.2.

3.5 Binary Probit Model in LIMDEP (Probit$)

In LIMDEP, the `Probit$` command estimates various probit models. Do not forget to include the ONE for the intercept. LIMDEP produces the same result as the other software packages.

```
PROBIT;Lhs=TRUST;
    Rhs=ONE,EDUCATE,INCOME,AGE,MALE,WWW;
    Marginal Effects; Means$

Normal exit from iterations. Exit status=0.

| Binomial Probit Model                        |
| Maximum Likelihood Estimates                 |
| Model estimated: Sep 09, 2009 at 11:41:52PM.|
| Dependent variable                TRUST     |
| Weighting variable                 None     |
| Number of observations             1174     |
| Iterations completed                  5     |
| Log likelihood function       -733.9975     |
| Number of parameters                  6     |
| Info. Criterion: AIC =          1.26064     |
|   Finite Sample: AIC =          1.26070     |
| Info. Criterion: BIC =          1.28655     |
| Info. Criterion:HQIC =          1.27041     |
| Restricted log likelihood     -798.3122     |
| McFadden Pseudo R-squared      .0805634     |
| Chi squared                    128.6294     |
| Degrees of freedom                    5     |
| Prob[ChiSqd > value] =         .0000000     |
| Hosmer-Lemeshow chi-squared =   4.81557     |
```

```
| P-value=  .77709 with deg.fr. =       8     |
+---iable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-ndex function for probability
 Constant|   -3.03005313       .27860620   -10.876   .0000
 EDUCATE |     .09072070       .01543488     5.878   .0000   14.2427598
 INCOME  |     .01859061       .00686814     2.707   .0068   24.6486371
 AGE     |     .01731045       .00294962     5.869   .0000   41.3074957
 MALE    |     .15939348       .07688194     2.073   .0382    .45059625
 WWW     |     .34176450       .09921561     3.445   .0006    .78534923

+-----------------ves of E[y] = F[*]   with |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Observations used for means are All Obs.  |
+---------------------  | Standard Error |b/St.Er.|P[|Z|>z]|Elasticity|
+--------+--------------+----numerator of Prob[Y = 4]
 Constant|    -.58627711       .01519985   -38.571   .0000
 EDUCATE |     .03529223       .00600827     5.874   .0000   1.22238383
 INCOME  |     .00723213       .00266928     2.709   .0067    .43350460
 AGE     |     .00673413       .00114709     5.871   .0000    .67646354
---------+Marginal effect for dummy variable is P|1 - P|0.
 MALE    |     .06205251     - .02991770     2.074   .0381    .06799567
---------+Marginal effect for dummy variable is P|1 - P|0.
 WWW     |     .12889554     - .03589934     3.590   .0003    .24616994

+-----------------------------------odel |
| Probit   model for variable TRUST      |
+-------------------------------------0  |
| N =    1174 N0=     682   N1=     492  |
| LogL=     -733.997 LogL0=    -798.312  |
| Estrella = 1-(L/L0)^(-2L0/n) = .10795  |
+--------------------------------------- |
|    .10456 |    .08056  |       .56389  |
|    Cramer | Veall/Zim. |     Rsqrd_ML  |
|    .10440 |    .17135  |       .10378  |
+----------------------------------------+
| Criteria        1.26064       1.28655  |
+----------------------------------------+
+--ed value is |
|1 when probability is greater than  .500000, 0 otherwise.|
|Note, column or row total percentages may not sum to     |
|100% because of rounding. Percentages are of full sample.|
+------+---------------------------------+------          |

|Value |       0               1        | Total Actual   |
+------+----------------+---------------+----------58.1%)|
|  1   |    263 ( 22.4%)|    229 ( 19.5%)|    492 ( 41.9%)|
+------+----------------+---------------+---------------)|
+------+----------------+---------------+----------------+

=======5000
------------------------------------------------------------------------------------------
Sensitivity = actual 1s correctly predicted                         46.545%
Specificity = actual 0s correctly predicted                         78.739%
Positive predictive value = predicted 1s that were actual 1s        61.230%
Negative predictive value = predicted 0s that were actual 0s        67.125%
Correct prediction = actual 1s and 0s correctly predicted           65.247%
-----------------------------------------------------------------------
Prediction Failure
-----------------------------------------------------------------------
False pos. for true neg. = actual 0s predicted as 1s                21.261%
False neg. for true pos. = actual 1s predicted as 0s                53.455%
False pos. for predicted pos. = predicted 1s actual 0s              38.770%
False neg. for predicted neg. = predicted 0s actual 1s              32.875%
False predictions = actual 1s and 0s incorrectly predicted          34.753%
=======================================================================
```

Compare marginal effects above with the following that `.prchange` computed at the means of all independent variables.

```
. prchange

probit: Changes in Probabilities for trust

         min->max      0->1     -+1/2    -+sd/2  MargEfct
educate    0.5262    0.0123    0.0353    0.0905    0.0353
 income    0.1816    0.0059    0.0072    0.0448    0.0072
    age    0.4435    0.0045    0.0067    0.0901    0.0067
   male    0.0621    0.0621    0.0619    0.0309    0.0620
    www    0.1289    0.1289    0.1323    0.0546    0.1330

              0       1
Pr(y|x)  0.5888  0.4112

       educate   income      age     male      www
    x=  14.2428  24.6486  41.3075  .450596  .785349
sd_x=   2.56971  6.19427  13.4071  .497765  .410755
```

3.6 Binary Probit Model in SPSS

SPSS has the `Probit` command to fit the binary probit model. This command requires an additional variable (e.g., *n* in the following example) with constant 1. If you want to use GUI menu (point-and-click), include *n* in `Total Observed:` and independent variables in `Covariate(s)` of a dialog box `Probit Analysis`.

```
COMPUTE n=1.
PROBIT trust OF n WITH educate income age male www
   /LOG NONE
   /MODEL PROBIT
   /PRINT FREQ
   /CRITERIA ITERATE(20) STEPLIMIT(.1).
```

The following tables are selected from messy SPSS output. Stata, SAS, LIMDEP, SPSS and R produce the same parameter estimates and goodness-of-fit measures.

Parameter Estimates

	Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
PROBITa	educate	.091	.015	5.878	.000	.060	.121
	income	.019	.007	2.707	.007	.005	.032
	age	.017	.003	5.869	.000	.012	.023
	male	.159	.077	2.073	.038	.009	.310
	www	.342	.099	3.445	.001	.147	.536
	Intercept	-3.030	.279	-10.876	.000	-3.309	-2.751

a. PROBIT model: PROBIT(p) = Intercept + BX

Chi-Square Tests

		Chi-Square	Dfa	Sig.
PROBIT	Pearson Goodness-of-Fit Test	1174.457	1168	.442

Chi-Square Tests

		Chi-Square	Dfa	Sig.
PROBIT	Pearson Goodness-of-Fit Test	1174.457	1168	.442

a. Statistics based on individual cases differ from statistics based on aggregated cases.

The `Probit` command also fits the binary logit model. The following command reports z scores instead of Wald statistics and does not report factor changes of the odds. The output is skipped.

```
PROBIT trust OF n WITH educate income age male www
   /LOG NONE
   /MODEL LOGIT
   /PRINT FREQ
   /CRITERIA ITERATE(20) STEPLIMIT(.1).
```

Table 3.1 summarizes parameter estimates and goodness-of-fit statistics produced in SAS, Stata, R, and LIMDEP. Parameter estimates are the same across software packages, but standard errors in PROC LOGISTIC and R are slightly different from those computed in other software packages (i.e., PROC QLIM, PROC GENMOD, PROC PROBIT, Stata, LIMDEP, and SPSS). I would recommend PROC LOGISTIC and Stata for the binary probit model.

Table 3.1 Parameter Estimates and Goodness-of-fit of the Binary Probit Model

	SAS			Stata	R	LIMDEP
	LOGISTIC	QLIM	GENMOD	.probit	glm()	Probit$
Education	.0907	.0907	.0907	.0907	.0907	.0907
	(.0158)	(.0154)	(.0154)	(.0154)	(.0158)	(.0154)
Family income	.0186	.0186	.0186	.0186	.0186	.0186
	(.0068)	(.0069)	(.0069)	(.0069)	(.0068)	(.0069)
Age	.0173	.0173	.0173	.0173	.0173	.0173
	(.0030)	(.0030)	(.0029)	(.0029)	(.0030)	(.0029)
Gender (male)	.1594	.1594	.1594	.1594	.1594	.1594
	(.0769)	(.0769)	(.0769)	(.0769)	(.0769)	(.0769)
WWW use	.3418	.3418	.3418	.3418	.3418	.3418
	(.0995)	(.0992)	(.0992)	(.0992)	(.0995)	(.0992)
Intercept	-3.0298	-3.0301	-3.0301	-3.0301	-3.0300	-3.0301
	(.2796)	(.2786)	(.2786)	(.2786)	(.2796)	(.2786)
Log likelihood	-733.9975	-733.9975	-733.9975	-733.9975	-733.9975	-733.9975
Likelihood test	128.629	128.63		128.63	128.6811	128.6294
Pseudo R^2	.0806	.0806		.0806	.0806	.0806
AIC	1479.995	1480.	1479.9949	*1479.995*	1749.995	1749.9914
BIC (Schwarz)	1510.404	1510.	1510.4040	*1510.404*		1510.4097
H_0 test	Chi-square	t	Chi-square	z	z	z

* PROC LOGISTIC and R reports (-2*Log-likelihood).
** AIC*N and BIC*N in Stata and LIMDEP

4. Bivariate Probit Regression Models

Bivariate probit regression models have two equations for two binary dependent variables. This chapter explains how to fit the bivariate probit model and the recursive bivariate regression model with an endogenous variable. The recursive bivariate probit model is formulated as (Maddala 1983:122-123; Greene 2003:715-716),

$$y_1^* = x_1'\beta_1 + y_2\gamma + \varepsilon_1, \quad y_1 = 1 \text{ if } y_1^* > 0, \text{ 0 otherwise,}$$
$$y_2^* = x_2'\beta_2 + \varepsilon_2, \quad y_2 = 1 \text{ if } y_2^* > 0, \text{ 0 otherwise,}$$

where y_1 is a binary dependent variable of interest in equation 1, y_2 is a binary dependent variable of equation 2 that is included in the first equation as an endogenous variable, and x_1 and x_2 are the regressor vectors of two regression equations. A typical bivariate probit model does not include $y_2\gamma$ in the first equation. Disturbances of two equations are assumed to be independent, identically distributed and follow the bivariate standard normal probability distribution with their correlation coefficient ρ:

$$\phi_2(\varepsilon_1, \varepsilon_2, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)}\left(\varepsilon_1^2 + \varepsilon_2^2 - 2\rho\varepsilon_1\varepsilon_2\right)\right]$$

Here we consider a model, where social trust and Internet use are jointly determined. Stata, SAS, and LIMDEP can fit bivariate probit models.

4.1 Bivariate Probit Model in Stata (.biprobit)

In Stata, `.biprobit` estimates bivariate probit models. If both equations have the same specification, you may list two dependent variables followed by covariates. If not, you need to specify equations individually, in each of which a binary variable and independent variables separated by an equal sign. The following two commands fit exactly the same model.

```
. quietly biprobit trust www educate income age male // or

. biprobit (trust = educate income age male) (www = educate income age male)

Fitting comparison equation 1:

Iteration 0:   log likelihood = -798.31217
Iteration 1:   log likelihood = -740.16976
Iteration 2:   log likelihood = -740.02303
Iteration 3:   log likelihood = -740.02303

Fitting comparison equation 2:

Iteration 0:   log likelihood =  -610.5431
Iteration 1:   log likelihood = -564.86129
Iteration 2:   log likelihood = -564.36806
Iteration 3:   log likelihood = -564.36805

Comparison:    log likelihood = -1304.3911

Fitting full model:

Iteration 0:   log likelihood = -1304.3911
```

```
Iteration 1:   log likelihood = -1297.8302
Iteration 2:   log likelihood = -1297.8205
Iteration 3:   log likelihood = -1297.8205

Bivariate probit regression                       Number of obs   =       1174
                                                  Wald chi2(8)    =     185.87
Log likelihood = -1297.8205                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
trust        |
     educate |   .1028598   .0150584     6.83   0.000      .073346    .1323737
      income |   .0202876   .0068117     2.98   0.003     .0069369    .0336384
         age |   .0161267   .0029175     5.53   0.000     .0104085     .021845
        male |    .165699   .0766088     2.16   0.031     .0155486    .3158495
       _cons |  -2.926968   .2750501   -10.64   0.000    -3.466056    -2.38788
-------------+----------------------------------------------------------------
www          |
     educate |   .1478252   .0180092     8.21   0.000     .1125278    .1831225
      income |   .0188763   .0065797     2.87   0.004     .0059803    .0317723
         age |  -.0103983   .0031951    -3.25   0.001    -.0166606   -.0041361
        male |   .0776235   .0864866     0.90   0.369     -.091887     .247134
       _cons |  -1.317766    .289774    -4.55   0.000    -1.885713   -.7498197
-------------+----------------------------------------------------------------
     /athrho |   .2035694   .0565478     3.60   0.000     .0927378     .314401
-------------+----------------------------------------------------------------
         rho |   .2008033   .0542676                      .0924729    .3044355
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0:     chi2(1) =  13.1412    Prob > chi2 = 0.0003
```

This model fits the data well (χ^2=185.87, p<.0000). `.fitstat` and other SPost commands do not work with this model. Instead, `.estat` returns AIC 2,618 and BIC 2,673, respectively.

```
. estat ic

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |   1174           .    -1297.82     11     2617.641    2673.391
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

We can compute marginal effects and conditional marginal effects using `predict(pmarg1)` and `predict(pcond1)`, respectively. If the correlation of disturbances of two equations is zero, they should be identical. Since the likelihood ratio test above rejects the null hypothesis of zero correlation (χ^2=13.1412, p<.0003), marginal effects and conditional marginal effects here are different even at the same reference points.

```
. mfx, predict(pcond1) at(mean educate=16 male=0)

Marginal effects after biprobit
      y  = Pr(trust=1|www=1) (predict, pcond1)
         =   .4744549
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
 educate |   .0371474      .00613    6.06   0.000   .025124  .049171        16
  income |   .0076112      .00272    2.80   0.005   .002278  .012944   24.6486
     age |    .006753      .00117    5.79   0.000   .004467  .009039   41.3075
    male*|   .0643811      .03051    2.11   0.035   .004592  .124171         0
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. mfx, predict(pmarg1) at(mean educate=16 male=0)
```

```
Marginal effects after biprobit
      y  = Pr(trust=1) (predict, pmarg1)
         =  .45422459
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
 educate |   .0407647      .00609    6.69   0.000   .028822  .052708        16
  income |   .0080402       .0027    2.98   0.003   .002752  .013329   24.6486
     age |   .0063912      .00116    5.53   0.000   .004127  .008655   41.3075
   male*|   .0659948      .03045    2.17   0.030   .006316  .125674         0
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

4.2 Recursive Bivariate Probit Model in Stata (.biprobit)

What if Internet use influences social trust directly? In order words, WWW use is the dependent variable in the second equation and is also included in the first equation as an endogenous variable. This is a recursive bivariate probit model, which is explained in Maddala (1983) and Greene (1996, 2003). Since the two equations have different specifications, they should be provided separately in parentheses after the `.biprobit` command. Check the model name `Seemingly unrelated bivariate probit` in the following output.

```
. biprobit (trust = educate income age male www) (www = educate income age male)

Fitting comparison equation 1:

Iteration 0:   log likelihood = -798.31217
Iteration 1:   log likelihood = -734.10951
Iteration 2:   log likelihood = -733.99746
Iteration 3:   log likelihood = -733.99746

Fitting comparison equation 2:

Iteration 0:   log likelihood =  -610.5431
Iteration 1:   log likelihood = -564.86129
Iteration 2:   log likelihood = -564.36806
Iteration 3:   log likelihood = -564.36805

Comparison:    log likelihood = -1298.3655

Fitting full model:

Iteration 0:   log likelihood = -1298.3655
Iteration 1:   log likelihood = -1298.2982
Iteration 2:   log likelihood = -1297.3043
Iteration 3:   log likelihood = -1297.3008
Iteration 4:   log likelihood = -1297.3007

Seemingly unrelated bivariate probit             Number of obs   =       1174
                                                 Wald chi2(9)    =     194.40
Log likelihood = -1297.3007                      Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
trust        |
     educate |   .1228844   .0197756     6.21   0.000      .084125    .1616437
      income |   .0225769   .0066392     3.40   0.001     .0095643    .0355894
         age |   .0126723    .004382     2.89   0.004     .0040837     .021261
        male |   .1682476   .0743747     2.26   0.024     .0224759    .3140193
         www |  -.7178395   .5729155    -1.25   0.210    -1.840733    .4050543
       _cons |  -2.531195   .4938755    -5.13   0.000    -3.499174   -1.563217
-------------+----------------------------------------------------------------
www          |
     educate |   .1510947   .0182167     8.29   0.000     .1153906    .1867988
      income |   .0188034   .0065301     2.88   0.004     .0060047    .0316021
```

```
         age |  -.0101814   .0031937    -3.19   0.001    -.0164409   -.0039219
        male |   .0663948    .086608     0.77   0.443    -.1033538    .2361435
       _cons |  -1.365747   .2928927    -4.66   0.000    -1.939807   -.7916883
-------------+----------------------------------------------------------------
     /athrho |   .6719729   .4621132     1.45   0.146    -.2337523    1.577698
-------------+----------------------------------------------------------------
         rho |   .5862762   .3032758                     -.2295859    .9182416
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0:     chi2(1) =  2.12962    Prob > chi2 = 0.1445
```

This model also fits the data well (χ^2=194.40, p<.0000) and most individual parameters are statistically significant at the .05 level. AIC and BIC are 2,619 and 2,679, respectively.

```
. estat ic

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |   1174          .   -1297.301     12     2618.601    2679.419
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

However, the LR test (χ^2=2.1296) suggests that the two disturbances are not significantly correlated. The estimated correlation .5863 is far away from zero but is not statistically discernable (p<.1445). Therefore, social trust and WWW use may not be jointly determined; each equation may need to be estimated separately or may be analyzed in the bivariate probit model. The binary probit model for WWW use is as follows.

```
. probit www educate income age male

Iteration 0:   log likelihood =  -610.5431
Iteration 1:   log likelihood = -564.86129
Iteration 2:   log likelihood = -564.36806
Iteration 3:   log likelihood = -564.36805

Probit regression                                 Number of obs   =       1174
                                                  LR chi2(4)      =      92.35
                                                  Prob > chi2     =     0.0000
Log likelihood = -564.36805                       Pseudo R2       =     0.0756

------------------------------------------------------------------------------
         www |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     educate |   .1454532   .0178746     8.14   0.000     .1104197    .1804868
      income |   .0189197   .0065902     2.87   0.004     .0060031    .0318362
         age |  -.0103946   .0032009    -3.25   0.001    -.0166682    -.004121
        male |     .08164   .0865442     0.94   0.346    -.0879834    .2512635
       _cons |  -1.288283   .2885836    -4.46   0.000    -1.853896   -.7226694
------------------------------------------------------------------------------
```

In the recursive bivariate probit model, conditional marginal effects make more sense than the typical marginal effects. The predicted probability that citizens trust most people is 47.21 percent at the reference points, given they use the Internet: `pr(trust=1|www=1)`=.4721.

```
. quietly biprobit (trust = educate income age male www) (www = educate income age male)

. mfx, predict(pcond1) at(mean educate=16 male=0 www=1)

Marginal effects after biprobit
      y  = Pr(trust=1|www=1) (predict, pcond1)
         =  .47208977
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
```

```
---------+--------------------------------------------------------------------
 educate |   .0394964      .00635    6.22   0.000   .027053   .05194         16
  income |   .0079921      .00266    3.01   0.003   .002786  .013198   24.6486
     age |   .0061891      .00132    4.67   0.000   .003592  .008786   41.3075
   male*|     .065738      .02987    2.20   0.028   .007193  .124284          0
    www*|   -.2858939      .21383   -1.34   0.181  -.704984  .133196          1
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Stata `.mfx` does not report direct and indirect effects but returns the sum of the two effects. When combining direct and indirect effects, for an additional increase in education from the 16 years, the conditional predicted probability of trusting people will increase by 3.95 percent, holding all other variables constant at their reference points.

The following Stata script illustrates how to compute manually direct and indirect effects of covariates. See the Stata script in Appendix for entire steps of computation. Beginners may skip this part and take a look at the result table only. Find the predicted probability of .4721 in the middle of the output. See Greene (1996, 2007) for related formulas.

```
. quietly biprobit (trust = educate income age male www) (www = educate income age male)
. global rho=e(rho)       // correlation coefficient of disturbances
. global n1 = 6           // the number of parameters in equation 1
. global n2 = 5           // the number of parameters in equation 2

. tabstat educate income age male www, stat(mean) col(variable) save

   stats |   educate    income       age      male       www
---------+--------------------------------------------------
    mean |  14.24276  24.64864   41.3075  .4505963  .7853492
------------------------------------------------------------

. matrix ref1 = r(StatTotal),I(1)     // reference points for equation 1
. matrix ref1[1,1]=16                // education (college graduation)
. matrix ref1[1,4]=0                 // female
. matrix ref1[1,5]=1                 // WWW use

. matrix ref2 = ref1[1,1..$n2]        // reference points for equation 2
. matrix ref2[1,$n2]=1

. // get parameter estimates
. matrix b0=e(b)
. matrix b1=b0[1,1..$n1]              // parameter estimates for equation 1
. matrix b2=b0[1,$n1+1..$n1+$n2]      // parameter estimates for equation 2

. matrix xb1=b1*ref1'     // compute xb1 of equation 1
. matrix xb2=b2*ref2'     // compute xb2 of equation 2

. global xb1=xb1[1,1]     // put xb1 into a global macro for computation
. global xb2=xb2[1,1]     // put xb1 into a global macro for computation

. // compute the predicted probability at the reference points
. di binormal($xb1, $xb2, $rho)/normal($xb2)
.47208977

. // compute direct effects
. global g1=normalden($xb1)*normal(($xb2-($rho)*$xb1)/sqrt(1-($rho)^2))
. matrix directE=$g1/normal($xb2)*b1
. matrix directE=directE[1,1..$n2]

. // compute indirect effects
. global g2=normalden($xb2)*normal(($xb1-($rho)*$xb2)/sqrt(1-($rho)^2))
. matrix indirectE=($g2/normal($xb2)- ///
                  (binormal($xb1,$xb2,$rho)*normalden($xb2))/(normal($xb2)^2))*b2
. matrix indirectE[1,$n2]=0
```

```
. // compute overall effects
. matrix Overall=directE+indirectE
```

...
(the procedure for computing discrete change is skipped)
...

```
. matrix list Marginal

Marginal[4,5]
             Education       Income          Age         Male          WWW
Reference          16    24.648637    41.307496            0            1
   Direct   .05190699     .0095366   .00535285    .07106867   -.3032191
 Indirect   -.0124106   -.00154447   .00083628   -.00545353            0
  Overall   .03949639    .00799213   .00618913    .06573803  -.28589388
```

Read the last line for overall marginal effects and discrete changes and compare with the output of the `.mfx` above. The overall impact of education on social trust is the sum of direct (.0519) and indirect effects (-.0124). Family income also has negative indirect effect -.0015, but age has both positive direct and indirect effects (.0054 and .0008, respectively).

The following two commands compute marginal effects of equation 1 and 2 (`pmarg1` and `pmarg2`). The predicted probability of trusting people is .4196 at the reference points, while the predicted probability of using WWW in the second equation is .8632.

```
. mfx, predict(pmarg1) at(mean educate=16 male=0 www=1)

Marginal effects after biprobit
      y  = Pr(trust=1) (predict, pmarg1)
         =  .41959352
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
 educate |   .0480246      .00759    6.33   0.000   .033147  .062903        16
  income |   .0088233      .00258    3.42   0.001    .00377  .013876   24.6486
     age |   .0049525      .00175    2.82   0.005   .001515   .00839   41.3075
   male*|   .0665716      .02941    2.26   0.024   .008926  .124217         0
    www*|  -.2770971      .20246   -1.37   0.171  -.673911  .119717         1
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. mfx, predict(pmarg2) at(mean educate=16 male=0 www=1)

Marginal effects after biprobit
      y  = Pr(www=1) (predict, pmarg2)
         =  .86317073
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
 educate |   .0331092      .00319   10.37   0.000   .026852  .039366        16
  income |   .0041204      .00145    2.84   0.005   .001277  .006963   24.6486
     age |   -.002231      .00071   -3.13   0.002  -.003628 -.000834   41.3075
   male*|   .0140228      .01825    0.77   0.442  -.021756  .049801         0
    www*|          0           0       .       .         0        0         1
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

4.3 Bivariate Probit Models in SAS: PROC QLIM

In SAS, PROC QLIM is able to estimate both bivariate probit models. Like Stata, SAS allows specifying two equations in a line if they share the same specification. ENDOGENOUS describes characteristics of dependent variables; in this example, they are discrete variables whose disturbances are normally distributed. Stata and SAS report the same correlation of disturbances (ρ=.2008), parameter estimates, and standard errors.

```
PROC QLIM DATA=masil.gss_cdvm;
     MODEL trust www = educate income age male;
     ENDOGENOUS trust www ~ DISCRETE(DIST=NORMAL);
RUN;
```

```
                                    The QLIM Procedure

                            Discrete Response Profile of trust

                    Index          Value            Frequency     Percent

                      1              0                   682       58.09
                      2              1                   492       41.91


                             Discrete Response Profile of www

                    Index          Value            Frequency     Percent

                      1              0                   252       21.47
                      2              1                   922       78.53


                                    Model Fit Summary

                      Number of Endogenous Variables              2
                      Endogenous Variable                 trust www
                      Number of Observations                   1174
                      Log Likelihood                          -1298
                      Maximum Absolute Gradient           0.0004068
                      Number of Iterations                       55
                      Optimization Method               Quasi-Newton
                      AIC                                      2618
                      Schwarz Criterion                        2673

Algorithm converged.

                                   Parameter Estimates

                                                    Standard                 Approx
          Parameter          DF        Estimate        Error    t Value    Pr > |t|

          trust.Intercept     1       -2.926969     0.275060     -10.64      <.0001
          trust.educate       1        0.102860     0.015059       6.83      <.0001
          trust.income        1        0.020288     0.006812       2.98      0.0029
          trust.age           1        0.016127     0.002918       5.53      <.0001
          trust.male          1        0.165699     0.076609       2.16      0.0305
          www.Intercept       1       -1.317767     0.289789      -4.55      <.0001
          www.educate         1        0.147825     0.018010       8.21      <.0001
          www.income          1        0.018876     0.006580       2.87      0.0041
          www.age             1       -0.010398     0.003195      -3.25      0.0011
          www.male            1        0.077624     0.086487       0.90      0.3694
          _Rho                1        0.200803     0.054268       3.70      0.0002
```

Now, let us fit the recursive bivariate probit model. Notice that the two equations are provided in two separate MODEL statements. The ENDOGENOUS statement is needed to indicate the probability distribution of disturbances in the two equations.

```
PROC QLIM DATA=masil.gss_cdvm;
     MODEL trust = educate income age male www;
     MODEL www = educate income age male;
     ENDOGENOUS trust www ~ DISCRETE(DIST=NORMAL);
RUN;
```

```
                                The QLIM Procedure

                        Discrete Response Profile of trust

                  Index          Value             Frequency     Percent

                    1              0                     682       58.09
                    2              1                     492       41.91


                         Discrete Response Profile of www

                  Index          Value             Frequency     Percent

                    1              0                     252       21.47
                    2              1                     922       78.53


                                Model Fit Summary

                    Number of Endogenous Variables              2
                    Endogenous Variable                 trust www
                    Number of Observations                   1174
                    Log Likelihood                          -1297
                    Maximum Absolute Gradient             0.00327
                    Number of Iterations                       52
                    Optimization Method              Quasi-Newton
                    AIC                                      2619
                    Schwarz Criterion                        2679

Algorithm converged.

                               Parameter Estimates

                                                 Standard                 Approx
          Parameter          DF        Estimate        Error    t Value    Pr > |t|

          trust.Intercept     1       -2.532266     0.494644      -5.12      <.0001
          trust.educate       1        0.122857     0.019796       6.21      <.0001
          trust.income        1        0.022575     0.006640       3.40      0.0007
          trust.age           1        0.012681     0.004389       2.89      0.0039
          trust.male          1        0.168258     0.074380       2.26      0.0237
          trust.www           1       -0.716498     0.574098      -1.25      0.2120
          www.Intercept       1       -1.365669     0.292877      -4.66      <.0001
          www.educate         1        0.151091     0.018218       8.29      <.0001
```

```
www.income          1        0.018804        0.006530       2.88     0.0040
www.age             1       -0.010182        0.003193      -3.19     0.0014
www.male            1        0.066424        0.086610       0.77     0.4431
_Rho                1        0.585570        0.303930       1.93     0.0540
```

Stata and PROC QLIM produce the same result except for the correlation of disturbances and parameter estimates of WWW use, which are slightly different (e.g., .5863 versus .5856 in ρ and -.7178 versus -.7165 for WWW use).

4.4 Bivariate Probit Models in LIMDEP (Bivariateprobit$)

`Bivariateprobit$` estimates bivariate probit models in LIMDEP. The `Lhs=` subcommand lists the two binary dependent variables, whereas `Rh1=` and `Rh2=` respectively specify the independent variables for the two equations.

```
BIVARIATEPROBIT;Lhs=TRUST,WWW;
    Rh1=ONE,EDUCATE,INCOME,AGE,MALE;
    Rh2=ONE,EDUCATE,INCOME,AGE,MALE$

Normal exit from iterations. Exit status=0.

+---------------------------------------------+
| FIML Estimates of Bivariate Probit Model    |
| Maximum Likelihood Estimates                |
| Model estimated: Sep 15, 2009 at 03:16:00PM.|
| Dependent variable              TRUWWW      |
| Weighting variable                None      |
| Number of observations            1174      |
| Iterations completed                17      |
| Log likelihood function      -1297.820      |
| Number of parameters                11      |
| Info. Criterion: AIC =         2.22968      |
|   Finite Sample: AIC =         2.22987      |
| Info. Criterion: BIC =         2.27716      |
| Info. Criterion:HQIC =         2.24758      |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
---------+Index    equation for TRUST
 Constant|   -2.92696771       .27487860   -10.648   .0000
 EDUCATE |     .10285982       .01414096     7.274   .0000   14.2427598
 INCOME  |     .02028760       .00707111     2.869   .0041   24.6486371
 AGE     |     .01612671       .00293070     5.503   .0000   41.3074957
 MALE    |     .16569900       .07696720     2.153   .0313    .45059625
---------+Index    equation for WWW
 Constant|   -1.31776621       .29250724    -4.505   .0000
 EDUCATE |     .14782515       .01763456     8.383   .0000   14.2427598
 INCOME  |     .01887630       .00643465     2.934   .0034   24.6486371
 AGE     |    -.01039833       .00328982    -3.161   .0016   41.3074957
 MALE    |     .07762348       .08744329      .888   .3747    .45059625
---------+Disturbance correlation
 RHO(1,2)|     .20080326       .05431808     3.697   .0002

+-----------------------------------------------------+
| Joint Frequency Table for Bivariate Probit Model    |
| Predicted cell is the one with highest probability  |
+-----------------------------------------------------+
|                          WWW                        |
+-------------+---------------------------------------+
|  TRUST      |       0            1        Total     |
|-------------+-------------+------------+------------+
|         0   |       180   |       502  |       682  |
|    Fitted   |   (    36)  |  (   730)  |  (   766)  |
```

```
|-------------+-------------+------------+------------+
|       1     |       72    |      420   |      492   |
|    Fitted   |   (     0)  |  (   408)  |  (   408)  |
|-------------+-------------+------------+------------+
|     Total   |      252    |      922   |     1174   |
|    Fitted   |   (    36)  |  (  1138)  |  (  1174)  |
|-------------+-------------+------------+------------+
+----------------------------------------------------------+
| Bivariate Probit Predictions for TRUST    and WWW      |
| Predicted cell (i,j) is cell with largest probability  |
| Neither TRUST    nor WWW      predicted correctly      |
|                            82 of    1174 observations  |
| Only    TRUST    correctly predicted                   |
|         TRUST    = 0:     143 of     682 observations  |
|         TRUST    = 1:      25 of     492 observations  |
| Only    WWW      correctly predicted                   |
|         WWW      = 0:       4 of     252 observations  |
|         WWW      = 1:      25 of     922 observations  |
| Both    TRUST    and WWW      correctly predicted      |
|         TRUST    = 0 WWW      = 0:      15 of     180  |
|         TRUST    = 1 WWW      = 0:       0 of      72  |
|         TRUST    = 0 WWW      = 1:     359 of     502  |
|         TRUST    = 1 WWW      = 1:     218 of     420  |
+----------------------------------------------------------+
```

The above output suggests that Stata, SAS, and LIMDEP produce same correlation coefficient of errors, parameter estimates, and standard errors with some rounding errors. AIC and BIC are 2617=2.2297*1,174 and 2,673=2.2772*1,174, respectively.

Now, fit the recursive bivariate probit model by adding WWW use to the first equation as an endogenous variable. `Marginal Effect` (or `Margin`) in the following command computes marginal effects and discrete changes at the means of the independent variables.

```
BIVARIATEPROBIT;Lhs=TRUST,WWW;
    Rh1=ONE,EDUCATE,INCOME,AGE,MALE,WWW;
    Rh2=ONE,EDUCATE,INCOME,AGE,MALE;
    Marginal Effect$
```

```
Normal exit from iterations. Exit status=0.

+---------------------------------------------+
| FIML Estimates of Bivariate Probit Model    |
| Maximum Likelihood Estimates                |
| Model estimated: Sep 15, 2009 at 00:21:09PM.|
| Dependent variable               TRUWWW     |
| Weighting variable                 None     |
| Number of observations             1174     |
| Iterations completed                 24     |
| Log likelihood function       -1297.301     |
| Number of parameters                 12     |
| Info. Criterion: AIC =          2.23050     |
|   Finite Sample: AIC =          2.23072     |
| Info. Criterion: BIC =          2.28230     |
| Info. Criterion:HQIC =          2.25003     |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
---------+Index    equation for TRUST
 Constant|   -2.53127459       .62810574    -4.030   .0001
 EDUCATE |     .12288180       .02325478     5.284   .0000   14.2427598
 INCOME  |     .02257666       .00691464     3.265   .0011   24.6486371
 AGE     |     .01267296       .00549849     2.305   .0212   41.3074957
 MALE    |     .16824823       .07532931     2.234   .0255    .45059625
 WWW     |    -.71772906       .79960562     -.898   .3694    .78534923
---------+Index    equation for WWW
```

```
 Constant|   -1.36574036       .29541029    -4.623   .0000
 EDUCATE |     .15109435       .01790608     8.438   .0000   14.2427598
 INCOME  |     .01880339       .00644213     2.919   .0035   24.6486371
 AGE     |    -.01018150       .00326806    -3.115   .0018   41.3074957
 MALE    |     .06639735       .08750730      .759   .4480    .45059625
---------+Disturbance correlation
 RHO(1,2)|     .58621974       .42476829     1.380   .1676

+-------------------------------------------------------+
| Marginal Effects for Ey1|y2=1                          |
+----------+----------+----------+----------+----------+
| Variable | Efct  x1 | Efct  x2 | Efct  h1 | Efct  h2 |
+----------+----------+----------+----------+----------+
| ONE      |   .00000 |   .00000 |   .00000 |   .00000 |
| EDUCATE  |   .05291 |  -.01572 |   .00000 |   .00000 |
| INCOME   |   .00972 |  -.00196 |   .00000 |   .00000 |
| AGE      |   .00546 |   .00106 |   .00000 |   .00000 |
| MALE     |   .07245 |  -.00691 |   .00000 |   .00000 |
| WWW      |  -.30905 |   .00000 |   .00000 |   .00000 |
+----------+----------+----------+----------+----------+
+-------------------------------------------+
| Partial derivatives of E[y1|y2=1] with    |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Effect shown is total of 4 parts above.   |
| Estimate of E[y1|y2=1] = .499957          |
| Observations used for means are All Obs.  |
| Total effects reported = direct+indirect. |
+-------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|       .000000    ......(Fixed Parameter).......
 EDUCATE |     .03718914       .00584175     6.366   .0000   14.2427598
 INCOME  |     .00776473       .00279401     2.779   .0055   24.6486371
 AGE     |     .00651654       .00123352     5.283   .0000   41.3074957
 MALE    |     .06553806       .03045594     2.152   .0314    .45059625
 WWW     |    -.30905460       .38237776     -.808   .4190    .78534923

+-------------------------------------------+
| Partial derivatives of E[y1|y2=1] with    |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Effect shown is total of 4 parts above.   |
| Estimate of E[y1|y2=1] = .499957          |
| Observations used for means are All Obs.  |
| These are the direct marginal effects.    |
+-------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|       .000000    ......(Fixed Parameter).......
 EDUCATE |     .05291298       .01587199     3.334   .0009   14.2427598
 INCOME  |     .00972153       .00344429     2.823   .0048   24.6486371
 AGE     |     .00545698       .00182639     2.988   .0028   41.3074957
 MALE    |     .07244780       .03248863     2.230   .0258    .45059625
 WWW     |    -.30905460       .38237776     -.808   .4190    .78534923

+-------------------------------------------+
| Partial derivatives of E[y1|y2=1] with    |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Effect shown is total of 4 parts above.   |
| Estimate of E[y1|y2=1] = .499957          |
| Observations used for means are All Obs.  |
| These are the indirect marginal effects.  |
+-------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|       .000000    ......(Fixed Parameter).......
```

```
 EDUCATE |    -.01572384       .01418159    -1.109   .2675   14.2427598
 INCOME  |    -.00195680       .00186681    -1.048   .2945   24.6486371
 AGE     |     .00105955       .00097193     1.090   .2756   41.3074957
 MALE    |    -.00690973       .01021978     -.676   .4990    .45059625
 WWW     |       .000000    ......(Fixed Parameter).......

+-----------------------------------------------------------+
| Analysis of dummy variables in the model. The effects are |
| computed using E[y1|y2=1,d=1] - E[y1|y2=1,d=0] where d is |
| the variable. Variances use the delta method.  The effect |
| accounts for all appearances of the variable in the model.|
+-----------------------------------------------------------+
|Variable      Effect   Standard error     t ratio          |
+-----------------------------------------------------------+
 MALE         .065467     .030353            2.157
 WWW         -.296117     .325843            -.909

+-----------------------------------------------------+
| Joint Frequency Table for Bivariate Probit Model    |
| Predicted cell is the one with highest probability  |
+-----------------------------------------------------+
|                         WWW                         |
+-------------+---------------------------------------+
|  TRUST      |       0            1         Total    |
|-------------+------------+------------+------------+
|         0   |       180   |       502   |       682   |
|    Fitted   |   (    54)  |  (   560)  |  (   614)  |
|-------------+------------+------------+------------+
|         1   |        72   |       420   |       492   |
|    Fitted   |   (     0)  |  (   560)  |  (   560)  |
|-------------+------------+------------+------------+
|     Total   |       252   |       922   |      1174   |
|    Fitted   |   (    54)  |  (  1120)  |  (  1174)  |
|-------------+------------+------------+------------+
+--------------------------------------------------------+
| Bivariate Probit Predictions for TRUST    and WWW      |
| Predicted cell (i,j) is cell with largest probability  |
| Neither TRUST    nor WWW      predicted correctly      |
|                          166 of    1174 observations  |
| Only    TRUST    correctly predicted                   |
|         TRUST    = 0:      25 of     682 observations  |
|         TRUST    = 1:      67 of     492 observations  |
| Only    WWW      correctly predicted                   |
|         WWW      = 0:       3 of     252 observations  |
|         WWW      = 1:      67 of     922 observations  |
| Both    TRUST    and WWW      correctly predicted      |
|         TRUST    = 0 WWW      = 0:      21 of     180  |
|         TRUST    = 1 WWW      = 0:       0 of      72  |
|         TRUST    = 0 WWW      = 1:     356 of     502  |
|         TRUST    = 1 WWW      = 1:     213 of     420  |
+--------------------------------------------------------+
```

SAS, Stata, and LIMDEP produce almost the same parameter estimates and log likelihood, but LIMDEP produces slightly different standard errors. The correlation of disturbances is .5862 in Stata and LIMDEP but is slightly different in SAS (ρ=.5856). LIMDEP and Stata report the same conditional predicted probability of 49.9968 percent and conditional marginal effects at the means of covariates. Let us compare the LIMDEP output (direct and indirect effects combined) with the following output computed in Stata:

```
. mfx, predict(pcond1) at(mean male=.450596 www=.785349)

Marginal effects after biprobit
      y  = Pr(trust=1|www=1) (predict, pcond1)
         =  .49996773
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
```

```
 educate |   .0371892      .00611    6.09   0.000   .025213  .049165   14.2428
  income |   .0077648      .00269    2.89   0.004   .002498  .013031   24.6486
     age |   .0065165       .0012    5.43   0.000   .004164  .008869   41.3075
   male*|   .0654669      .03028    2.16   0.031   .006124   .12481   .450596
    www*|  -.2961619      .23328   -1.27   0.204  -.753376  .161052   .785349
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

LIMDEP reports direct and indirect effects separately in addition to direct and indirect effect combined. The first table under the label `Marginal Effects for Ey1|y2=1` right after the parameter estimates summarizes direct and indirect effects. For example, education has a direct effect of .05291 and an indirect effect -.01572, so its overall impact on social trust is the sum of the two effects, which is .0372=.0529-.0157. Stata reports this combined marginal effect. Find the equivalent overall effect in the table under `Total effects reported = direct+indirect` of the above LIMDEP output. LIMDEP produces other two tables for direct (see under `These are the direct marginal effects`) and indirect effects (see under `These are the indirect marginal effects`).

Discrete changes .0655 of male and -.3091 of WWW use under `direct+indirect` in the LIMDEP output are different from those of Stata since LIMDEP computes at the means of all covariates including binary variables; in fact, they are not, by definition, discrete changes (differences in predicted probabilities between `trust=0` and `trust=1`). LIMDEP reports discrete changes (`E[y1|y2=1,d=1]-E[y1|y2=1,d=0]`) separately at the bottom of the output. Find -6.5467 percent for gender and -29.6117 for WWW use.

The following table reports direct, indirect, and overall effects computed manually at the means of covariates in Stata. See the attached Stata script for computation. Notice that the last two numbers (.0655 and -.2962) on row `Overall` are discrete changes of gender and WWW use, respectively.

	Education	Income	Age	Male	WWW
Reference	14.24276	24.648637	41.307496	.45059625	.78534923
Direct	.05291496	.00972179	.0054568	.07244873	-.30910722
Indirect	-.01572574	-.00195703	.00105967	-.00691029	0
Overall	.03718922	.00776475	.00651647	.06546686	-.29616189

Analysis of direct and indirect effects is very useful especially when two effects have opposite signs. For instance, education influences positively social trust in the first equation but has a negative impact (indirect effect) on WWW use in the second equation. Therefore, its overall effect is determined by magnitudes of two effects; the large direct impact dominates in this case, .0372=.0529-.0157. If this specification is correct, a single equation for social trust may mistakenly report an overestimated impact of education. See Greene (1996, 2003) for discussion of computing and interpreting marginal effects in the recursive bivariate probit model.

Table 4.1 compares the results of bivariate probit models across Stata, SAS, and LIMDEP. In the bivariate probit model, all three software packages report the same goodness-of-fit measures, parameter estimates, and the correlation coefficient of disturbance (ρ=.2008), but LIMDEP produces slightly different standard errors. In the recursive bivariate probit model, similarly, Stata, SAS, and LIMDEP produce the same parameter estimates and goodness-of-fit measures, but LIMDEP produce different standard errors. SAS reports a bit different parameter estimate of the endogenous variable (-.7165 versus -.7178) and correlation coefficient (ρ=.5856 versus .5863).

Table 4.1 Parameter Estimates and Goodness-of-fit of Bivariate Probit Models

	Bivariate Probit Model			Recursive Bivariate Probit Model		
	Stata	SAS	LIMDEP	Stata	SAS	LIMDEP
Education	.1029 (.0151)	.1029 (.0151)	.1029 (.0141)	.1229 (.0198)	.1229 (.0198)	.1229 (.0233)
Family income	.0203 (.0068)	.0203 (.0068)	.0203 (.0071)	.0226 (.0066)	.0226 (.0066)	.0226 (.0069)
Age	.0161 (.0029)	.0161 (.0029)	.0161 (.0029)	.0127 (.0044)	.0127 (.0044)	.0127 (.0055)
Gender (male)	.1657 (.0766)	.1657 (.0766)	.1657 (.0770)	.1682 (.0744)	.1682 (.0744)	.1682 (.0753)
WWW use				-.7178 (.5729)	-.7165 (.5741)	-.7177 (.7996)
Intercept	-2.9270 (.2751)	-2.9270 (.2751)	-2.9270 (.2749)	-2.5312 (.4939)	-2.5323 (.4946)	-2.5313 (.6281)
Education	.1478 (.0180)	.1478 (.0180)	.1478 (.0176)	.1511 (.0182)	.1511 (.0182)	.1511 (.0179)
Family income	.0189 (.0066)	.0189 (.0066)	.0189 (.0063)	.0188 (.0065)	.0188 (.0065)	.0188 (.0064)
Age	-.0104 (.0032)	-.0104 (.0032)	-.0104 (.0033)	-.0102 (.0032)	-.0102 (.0032)	-.0102 (.0033)
Gender (male)	.0776 (.0865)	.0776 (.0865)	.0776 (.0874)	.0664 (.0866)	.0664 (.0866)	.0664 (.0875)
Intercept	-1.3178 (.2898)	-1.3178 (.2898)	-1.3178 (.2925)	-1.3657 (.2929)	-1.3657 (.2929)	-1.3657 (.2954)
Log likelihood	-1297.8205	-1298	-1297.820	-1297.3007	-1297	1297.301
Likelihood test	185.87			194.40		
Rho (ρ)	.2008 (.0543)	.2008 (.0543)	.2008 (.0543)	.5863 (.3033)	.5856 (.3039)	.5862 (.4248)
χ^2 to test ρ=0	13.1412			2.1296		
AIC	2617.641	2618	2617.644	2618.601	2619	2618.607
BIC (Schwarz)	2673.391	2673	2673.386	2679.419	2679	2679.420

[*] AIC*N and BIC*N in LIMDEP

5. Conclusion

The regression models discussed so far are of categorical dependent variables (binary, ordinal, and nominal responses). An appropriate regression model is determined largely by the measurement level of the categorical dependent variable of interest. The level of measurement should be considered in conjunction with theory and research questions (Long 1997). You must also examine the data generation process (DGP) of a dependent variable to understand its "behavior." Experienced researchers pay special attention to censoring, truncation, sample selection, and other particular patterns of the DGP. These issues are not addressed in this brief technical note.

Generally speaking, if the dependent variable is binary, you may use the binary logit or probit regression model. For ordinal responses, try to fit either ordered logit or probit regression model. If you have a nominal response variable, investigate the DGP carefully and then choose one of the multinomial logit, conditional logit, and nested logit models. In order to use the conditional logit and nested logit, you need to reshape the data set in advance.

You should check key assumptions of a model before fitting the model. Examples are the parallel regression assumption in ordered logit and probit models and the independence of irrelevant alternatives (IIA) assumption in the multinomial logit model. You may respectively conduct the Brant test and Hausman test for these assumptions. If an assumption of an ordered or nominal response model is violated, find alternative models or consider if a dependent variable can be explored in a binary response model by dichotomizing the variable.

Since logit and probit models are nonlinear, their parameter estimates are difficult to interpret intuitively. The situation becomes even worse in generalized ordered logit and multinomial logit models, where many parameter estimates and related statistics are produced. Consequently, researchers need to spend more time and effort interpreting the results substantively. Simply reporting parameter estimates and goodness-of-fit statistics is not sufficient. J. Scott Long (1997) and Long and Freese (2003) provide good examples of meaningful interpretations using predicted probabilities, factor changes in odds, and marginal effects (discrete changes) of predicted probabilities. It is highly recommended to visualize marginal effects and discrete changes using a plot of predicted probabilities.

In general, logit and probit models require larger N than do linear regression models. Like the Bayesian estimation method, the maximum likelihood estimation method depends on data. You need to check if you have sufficient valid observations especially when your data contain many missing values. Scott Long's rule of thumb says 500 observations and at least additional 10 per independent variable are required in ML estimation. If you have small N, DO NOT include a large number of independent variables. This is the so called "small N and large parameter" problem; you may not be able to reach convergence in estimation and/or may not get reliable results with desirable asymptotic ML properties. In contrast, an extremely large N, say millions to estimate only two parameters, is not always a virtue since it absurdly boosts the statistical power of a test without adding new information. Even a tiny effect, which should have been negligible in a normal situation, may be mistakenly reported as statistically significant.

Regarding statistical software packages, I would recommend the SAS LOGISTIC, QLIM, and MDC procedures of SAS/ETS (see Table 2.1 and 3.1). SAS also has PROC GENMOD and PROC PROBIT, but PROC LOGISTIC and PROC QLIM appear to be best for binary and ordinal response models, and PROC MDC is good for nominal dependent variable models. ODS is another advantage of using SAS. I also strongly recommend using Stata since it provides handy ways to fit various models and also can be assisted by SPost, which has various useful commands such as `.fitstat`, `.prchange`, `.listcoef`, `.prtab`, and `.prgen`. I encourage the SAS Institute to develop additional statements similar to, in particular, `.prchange` and `.prgen`.

LIMDEP supports various regression models for categorical dependent variables addressed in Greene (2003) but does not seem as user-friendly and stable as SAS and Stata. However, LIMDEP computes direct and indirect effects in the recursive bivariate probit model and helps researchers interpret the result in more detail. You may benefits from R's object-oriented programming concept and analyze data flexibly in your own way. SPSS is least recommended mainly due to its limited support for categorical dependent variable models and messy syntax and output.

For logit and probit models for ordinal and nominal outcome variables, see Park, Hun Myoung. 2009. *Regression Models for Ordinal and Nominal Dependent Variables Using SAS, Stata, LIMDEP, and SPSS*. Working Paper. The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University."
http://www.indiana.edu/~statmath/stat/all/cdvm/index_nominal.html

Appendix: Data Sets

The sample data set is a subset of the 2000 and 2002 General Social Survey of NORC (http://www.norc.org).

http://www.indiana.edu/~statmath/stat/all/cdvm/gss_cdvm.csv
http://www.indiana.edu/~statmath/stat/all/cdvm/gss_cdvm.sas7bdat
http://www.indiana.edu/~statmath/stat/all/cdvm/gss_cdvm.dta

http://www.indiana.edu/~statmath/stat/all/cdvm/cdvm_binary.do (Stata script)
http://www.indiana.edu/~statmath/stat/all/cdvm/cdvm_binary.R (R script)

- `trust`: 1 if a respondent trust most people
- `belief`: Religious intensity: no religion (0) through strong (3)
- `educate`: respondent's education (years)
- `income`: family income ($1,000.00)
- `age`: respondent's age
- `male`: 1 for male and 0 for female
- `www`: 1 if a respondent have used WWW

```
. sum trust belief educate income age male www, sep(20)

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       trust |      1174    .4190801    .4936188          0          1
      belief |      1174    1.892675    1.044809          0          3
     educate |      1174    14.24276    2.569712          2         20
      income |      1174    24.64864     6.19427         .5       27.5
         age |      1174     41.3075    13.40713         18         86
        male |      1174    .4505963    .4977653          0          1
         www |      1174    .7853492    .4107548          0          1


. tab trust male, miss

    Social |        Gender
     Trust |    Female       Male |     Total
-----------+----------------------+----------
         0 |       397        285 |       682
         1 |       248        244 |       492
-----------+----------------------+----------
     Total |       645        529 |     1,174


. tab trust www, miss

    Social |        WWW Use
     Trust | Non-users      Users |     Total
-----------+----------------------+----------
         0 |       180        502 |       682
         1 |        72        420 |       492
-----------+----------------------+----------
     Total |       252        922 |     1,174


. tab male www, miss

           |        WWW Use
    Gender | Non-users      Users |     Total
```

```
-----------+----------------------+----------
    Female |       149        496 |       645
      Male |       103        426 |       529
-----------+----------------------+----------
     Total |       252        922 |     1,174
```

. tab belief male, miss

```
      Religious |        Gender
      Intensity |    Female       Male |     Total
----------------+----------------------+----------
    No religion |        80        112 |       192
Somewhat strong |        79         55 |       134
Not very strong |       239        217 |       456
         Strong |       247        145 |       392
----------------+----------------------+----------
          Total |       645        529 |     1,174
```

. tab belief www, miss

```
      Religious |        WWW Use
      Intensity | Non-users      Users |     Total
----------------+----------------------+----------
    No religion |        38        154 |       192
Somewhat strong |        37         97 |       134
Not very strong |        95        361 |       456
         Strong |        82        310 |       392
----------------+----------------------+----------
          Total |       252        922 |     1,174
```

References

Allison, Paul D. 1991. *Logistic Regression Using the SAS System: Theory and Application*. Cary, NC: SAS Institute.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Cameron, A. Colin, and Pravin K. Trivedi. 2009. *Microeconometrics Using Stata*. TX: Stata Press.

Greene, William H. 1996. Marginal Effects in the Bivariate Probit Model. Stern School of Business, New York University.

Greene, William H. 2003. *Econometric Analysis*, 5th ed. Upper Saddle River, NJ: Prentice Hall.

Greene, William H. 2007. *LIMDEP Version 9.0 Econometric Modeling Guide*. Plainview, New York: Econometric Software.

Long, J. Scott, and Jeremy Freese. 2003. *Regression Models for Categorical Dependent Variables Using Stata*, 2nd ed. College Station, TX: Stata Press.

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables: Advanced Quantitative Techniques in the Social Sciences*. Sage Publications.

Maddala, G. S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.

Park, Hun Myoung. 2004. "Presenting the Binary Logit/Probit Models Using the SAS/IML." Proceedings of the 15th Midwest SAS Users Group Conference in Chicago, IL (September 26-28, 2004).

SAS Institute. 2004. *SAS/STAT 9.1 User's Guide*. Cary, NC: SAS Institute.

SPSS Inc. 2007. *SPSS 16.0 Command Syntax Reference*. Chicago, IL: SPSS Inc.

Stata Press. 2007. *Stata Base Reference Manual, Release 10*. College Station, TX: Stata Press.

Stokes, Maura E., Charles S. Davis, and Gary G. Koch. 2000. *Categorical Data Analysis Using the SAS System*, 2nd ed. Cary, NC: SAS Institute.

Acknowledgements

I am grateful to Jeremy Albright and Kevin Wilhite at the UITS Center for Statistical and Mathematical Computing for comments and suggestions. I also thank J. Scott Long in Sociology and David H. Good in the School of Public and Environmental Affairs, Indiana University. A special thanks to many readers around the world who have eagerly provided constructive feedback and encouraged me to keep improving this document.

Revision History

- 2003. 04 First draft
- 2004. 07 Second draft
- 2005. 09 Third draft (Added bivariate logit/probit and nested logit models)
- 2008. 10 Fourth draft (Added SAS ODS and SPSS output)
- 2009. 09 Fifth draft (Estimated models using different data and rewrote chapter 2-4)

- 2010. Edited by Dani Marinova.