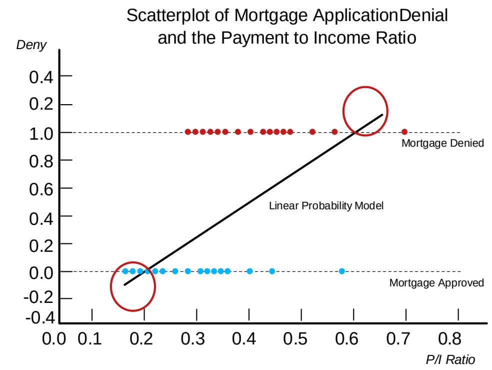
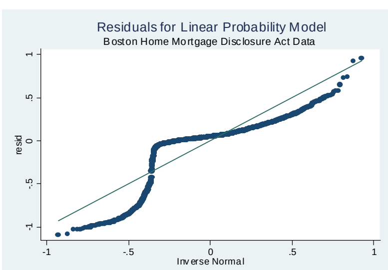
Variable	Definition
P/I ratio	Ratio of total monthly debt payments to total monthly income
housing expense-to-total-monthly-income	Ratio of monthly housing expenses to total monthly income
loan-to-value ratio	Ratio of size of loan to assessed value of property
consumer credit score	
	"1 if no ""slow"" payments or delinquencies"
	2 if one or two slow payments or delinquencies
	3 if more than two slow payments
	4 if insufficient credit history for determination
	5 if delinquent credit history with payments 60 days overdue
	6 if delinquent credit history with payments 90 days overdue
mortgage credit score	
	1 if no late mortgage payments
	2 if no mortgage payment history
	3 if one or two late mortgage payments
	4 if more than two late mortgage payments
public bad credit record	
	"1 if any public record of credit problems (bankruptcy, charge-offs,"
	0 otherwise
Additional Applicant Characteristics	
denied mortgage insurance	
	"1 if applicant applied for mortgage insurance and was denied,"
	0 otherwise
self-employed	
	"1 if self-employed, 0 otherwise

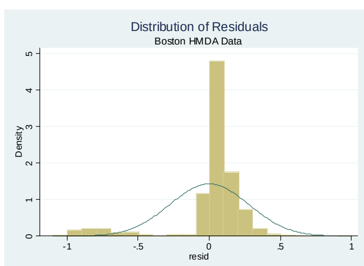
Variable	Definition
single	
	"1 if applicant reported being single,"
	0 otherwise
high school diploma	
	1 if applicant graduated from high school
	0 otherwise
unemployment rate	
1989 Massachusetts unemployment rate in the applicant's industry	3.8
condominium	
	"1 if unit is a condominium,"
	0 otherwise
black	
	1 if applicant is black
	0 if white
deny	
	1 if mortgage application denied
	0 otherwise

Variables included in the Regression Models of Mortgage Decisions

Mortgage applicants with a high ratio of debt payments to income (*P/I ratio*) are more likely to have their application denied (*deny = 1* if denied, *deny = 0* if approved). The linear probability model uses a straight line to model the probability of denial, conditional on the *P/I ratio*.

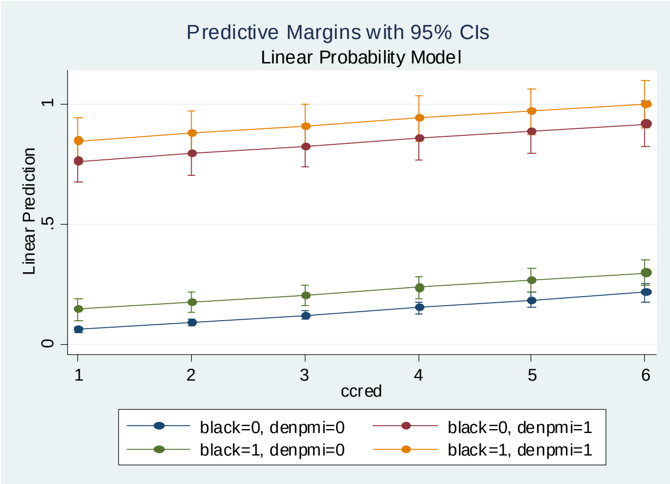







```
regress deny black pi_rat hse_inc ltv_med ltv_high ccred mcred pubrec denpmi
selfemp, r

Linear regression                                      Number of obs     =      2,380
                                                       F(10, 2369)       =      67.22
                                                       Prob > F          =     0.0000
                                                       R-squared         =     0.2663
                                                       Root MSE          =     .27875

------------------------------------------------------------------------------
             |               Robust
        deny |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |   .0836967   .0225623     3.71   0.000     .0394529    .1279406
      pi_rat |   .4487963   .1135962     3.95   0.000     .2260381    .6715545
     hse_inc |  -.0480226    .109559    -0.44   0.661     -.262864    .1668187
     ltv_med |   .0314498   .0127391     2.47   0.014     .0064688    .0564308
    ltv_high |   .1890511   .0501681     3.77   0.000     .0906732     .287429
       ccred |   .0307716   .0045843     6.71   0.000     .0217819    .0397612
       mcred |   .0209104   .0112898     1.85   0.064    -.0012284    .0430493
      pubrec |   .1970876   .0348812     5.65   0.000     .1286867    .2654885
      denpmi |   .7018841   .0451051    15.56   0.000     .6134345    .7903337
     selfemp |   .0598438   .0205233     2.92   0.004     .0195983    .1000894
       _cons |  -.1829933   .0276729    -6.61   0.000    -.2372589   -.1287277
------------------------------------------------------------------------------
```

margins , at(ccred = (1 (1) 6) black = (0 1) denpmi =(0 1))
marginsplot, subtitle("Linear Probability Model")



Theory Behind Logit & Probit Analysis

$$y_i^* = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \mu_i$$

Model explaining the underlying (but unobserved) propensity (y^*) for the bank to deny a loan based on individual applicant characteristics (x_{i1}, x_{i2}, ..., x_{ik})

$$y_i^* = 0 \Rightarrow \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \mu_i$$

If the propensity y* is to equal zero then

$$\Rightarrow -\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right) = \mu_i$$

$$y_i^* > 0 \Rightarrow -\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right) < \mu_i$$

$$\mu_i > -\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right)$$

and, therefore, $y^* > 0$ implies that the **random error term** must be equal to or greater than the constant value determind by the individual's characteristics.

We actually only observe the dummy variable y_i which takes on the values:

$$y_i = \begin{Bmatrix} 1 \text{ if } y^* > 0 \\ 0 \text{ otherwise} \end{Bmatrix}$$

$$P_i = \Pr\left(y_i = 1\right) = \Pr\left[\mu_i > -\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right)\right]$$

$$= 1 - F\left[-\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right)\right]$$

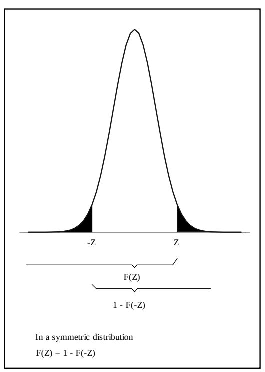
where *F[..]* is the cumulative distribution function of the random error term *mu.*
This expression says that for any particular observation, *i,* the probability that an event will occur equals:

$$P_i = \Pr\left(y_i = 1\right) = \Pr\left[y_i^* > 0\right]$$

If the distribution of *mu* is symmetric, since:

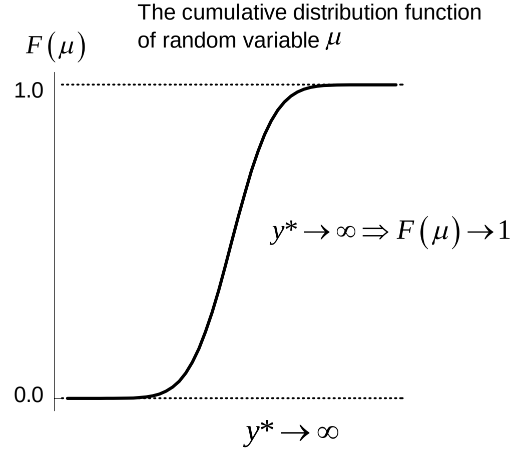
$$1 - F\left(-Z\right) = F\left(Z\right)$$

We can write (see the graph to the left):



$$P_i = 1 - F\left[-\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right)\right] = F\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right)$$

The cumulative distribution function of random variable μ



$$L=\prod_{y_i=1} P_i \prod_{y_i=0}\left(1-P_i\right)$$

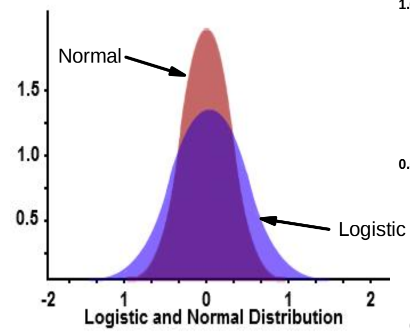
The likelihood function is just the resolution of a series of Bernoulli outcomes with each P_i *equal to:*

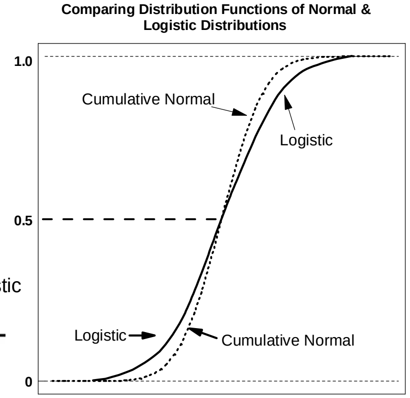
$$P_i = 1-F\left[-\left(\beta_0+\sum_{j=1}^{k}\beta_j x_{ij}\right)\right] = F\left(\beta_0+\sum_{j=1}^{k}\beta_j x_{ij}\right)$$

Choice of probability distribution for the error term:

$$y_i^* = \beta_0+\sum_{j=1}^{k}\beta_j x_{ij} + \mu_i$$

	Normal	Logistic
Density $\left[f\left(x\mid\mu,\sigma^2\right)\right]$	$=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$	$=\frac{\exp\left(-\frac{x-\mu}{s}\right)}{s\left[1+\exp\left(-\frac{x-\mu}{s}\right)\right]^2}$
CDF $\left[F(Z)\right]$	$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{Z}\exp^{-t^2/2}\,dt$	$=\frac{\exp(Z_i)}{1+\exp(Z_i)}$





$$Z_i=\beta_0+\sum_{j=1}^{k}\beta_j x_{ij}$$

Let: $Z_i=\beta_0+\sum_{j=1}^{k}\beta_j x_{ij}$

$$F\left(Z_i\right)=\frac{\exp\left(Z_i\right)}{1+\exp\left(Z_i\right)}$$

Therefore, the odds ratio is:

$$\frac{\Pr\left(y_i=1\right)}{\Pr\left(y_i=0\right)}=\frac{P_i}{1-P_i}=\frac{F\left(Z_i\right)}{1-F\left(Z_i\right)}=\frac{\frac{\exp\left(Z_i\right)}{1+\exp\left(Z_i\right)}}{1-\frac{\exp\left(Z_i\right)}{1+\exp\left(Z_i\right)}}=\exp\left(Z_i\right)=\exp\left(\beta_0+\sum_{j=1}^{k}\beta_j x_{ij}\right)$$

and the ***log odds ratio*** is: $\log\left[\frac{F\left(Z_i\right)}{1-F\left(Z_i\right)}\right]=\log\left[\exp\left(Z_i\right)\right]=Z_i=\beta_0+\sum_{j=1}^{k}\beta_j x_{ij}$

and the ***logit regression model*** is: $\log\left(\frac{P_i}{1-P_i}\right)=Z_i=\beta_0+\sum_{j=1}^{k}\beta_j x_{ij}$

For observation *i* the estimated probability is:

$$\hat{P}_i=\frac{\exp\left(\hat{\beta}_0+\sum_{j=1}^{k}\hat{\beta}_j x_{ij}\right)}{1+\exp\left(\hat{\beta}_0+\sum_{j=1}^{k}\hat{\beta}_j x_{ij}\right)}$$

The marginal impact on the estimated probability of an event's occurring due to a change is a particular x_h:

$$\frac{\partial}{\partial x_h}\hat{P}_i=\hat{\beta}_h\left\{\frac{\exp\left(\hat{\beta}_0+\sum_{j=1}^{k}\hat{\beta}_j x_{ij}\right)}{\left[1+\exp\left(\hat{\beta}_0+\sum_{j=1}^{k}\hat{\beta}_j x_{ij}\right)\right]^2}\right\}$$

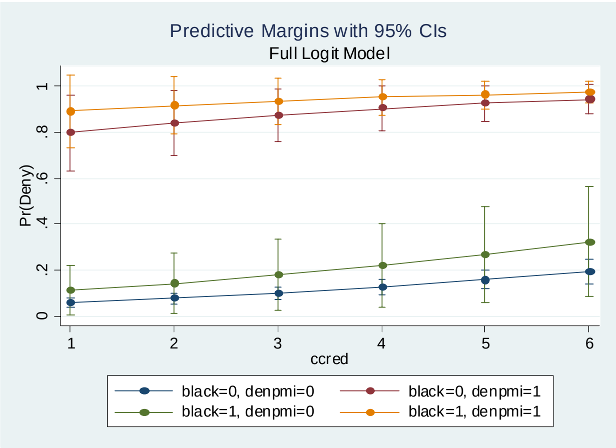
```
logit deny black pi_rat hse_inc ltv_med ltv_high ccred mcred pubrec denpmi
selfemp single hischl probunmp blk_pi blk_hse

Iteration 0:   log likelihood =  -872.0853
Iteration 1:   log likelihood = -667.32451
Iteration 2:   log likelihood = -647.72477
Iteration 3:   log likelihood = -627.64798
Iteration 4:   log likelihood =  -627.1807
Iteration 5:   log likelihood = -627.17867
Iteration 6:   log likelihood = -627.17867

Logistic regression                             Number of obs     =      2,380
                                                LR chi2(15)       =     489.81
                                                Prob > chi2       =     0.0000
Log likelihood = -627.17867                     Pseudo R2         =     0.2808

------------------------------------------------------------------------------
        deny |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |   .7581488   .7390005     1.03   0.305    -.6902657    2.206563
      pi_rat |   5.207461   1.152529     4.52   0.000     2.948545    7.466376
     hse_inc |  -.8454345   1.393284    -0.61   0.544     -3.57622    1.885351
     ltv_med |   .4782353   .1597177     2.99   0.003     .1651943    .7912763
    ltv_high |   1.474876   .3161379     4.67   0.000     .8552567    2.094495
       ccred |   .2978086   .0401532     7.42   0.000     .2191098    .3765074
       mcred |   .2180232   .1443593     1.51   0.131    -.0649158    .5009622
      pubrec |   1.236939   .2066445     5.99   0.000     .8319235    1.641955
      denpmi |   4.594167   .5567479     8.25   0.000     3.502961    5.685373
     selfemp |   .6428772   .2146674     2.99   0.003     .2221368    1.063618
      single |   .4178672   .1545136     2.70   0.007      .115026    .7207083
      hischl |   -1.15973     .41516    -2.79   0.005    -1.973429   -.3460318
    probunmp |    .056445   .0343813     1.64   0.101    -.0109411    .1238312
      blk_pi |  -2.013371   2.504404    -0.80   0.421    -6.921913    2.895171
     blk_hse |   2.285755   2.992161     0.76   0.445    -3.578772    8.150282
       _cons |  -4.853938   .6152655    -7.89   0.000    -6.059836    -3.64804
------------------------------------------------------------------------------
```

$$\frac{\partial}{\partial x_h}\hat{P}_i = \hat{\beta}_h \left\{ \frac{\exp\left(\hat{\beta}_0 + \sum_{j=1}^{k}\hat{\beta}_j x_{ij}\right)}{\left[1+\exp\left(\hat{\beta}_0 + \sum_{j=1}^{k}\hat{\beta}_j x_{ij}\right)\right]^2} \right\}$$

```
. test single hischl probunmp

 ( 1)  [deny]single = 0
 ( 2)  [deny]hischl = 0
 ( 3)  [deny]probunmp = 0

           chi2(  3) =   17.16
         Prob > chi2 =    0.0007
------------------------------------------
. test black blk_pi blk_hse

 ( 1)  [deny]black = 0
 ( 2)  [deny]blk_pi = 0
 ( 3)  [deny]blk_hse = 0

           chi2(  3) =   14.02
         Prob > chi2 =    0.0029
------------------------------------------
```

margins , at(ccred = (1 (1) 6) black = (0 1) denpmi =(0 1))

marginsplot, subtitle("Full Logit Model")



Probit regression as an alternative to Logit

$$y_i^* = \beta_0 + \sum_{j=1}^{k}\beta_j x_{ij} + \mu_i$$

Assume that *mu*$_i$ has a normal distribution

$$\mu_i > -\left(\beta_0 + \sum_{j=1}^{k}\beta_j x_{ij}\right)$$

$$\Pr\left(y_i = 1\right) = P_i = F\left(\beta_0 + \sum_{j=1}^{k}\beta_j x_{ij}\right) = \int_{-\infty}^{\beta_0 + \sum_{j=1}^{k}\beta_j x_{ij}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz_i$$

```
. probit deny black pi_rat hse_inc ltv_med ltv_high ccred mcred pubrec denpmi
selfemp single hischl probunmp blk_pi blk_hse, r

Probit regression                               Number of obs     =      2,380
                                                Wald chi2(15)     =     315.60
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -628.33216               Pseudo R2         =     0.2795

------------------------------------------------------------------------------
             |               Robust
        deny |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |   .2462823   .4480002     0.55   0.582     -.631782    1.124347
      pi_rat |   2.571702   .6623891     3.88   0.000     1.273443    3.869961
     hse_inc |  -.5380259   .7429056    -0.72   0.469    -1.994094    .9180423
     ltv_med |   .2158857   .0819145     2.64   0.008     .0553362    .3764352
    ltv_high |   .7881879   .1814846     4.34   0.000     .4324847    1.143891
       ccred |   .1578707   .0211387     7.47   0.000     .1164396    .1993017
       mcred |   .1114918   .0758759     1.47   0.142    -.0372223    .2602059
      pubrec |   .7047677   .1156898     6.09   0.000     .4780198    .9315156
      denpmi |   2.590097   .2937427     8.82   0.000     2.014371    3.165822
     selfemp |   .3475272   .1145316     3.03   0.002     .1230494     .572005
      single |   .2255941   .0802929     2.81   0.005     .0682229    .3829653
      hischl |  -.6198276   .2308051    -2.69   0.007    -1.072197    -.167458
    probunmp |   .0297126   .0181141     1.64   0.101    -.0057903    .0652156
      blk_pi |  -.5792648   1.472873    -0.39   0.694    -3.466043    2.307514
     blk_hse |   1.231736   1.693551     0.73   0.467    -2.087564    4.551035
       _cons |  -2.543412   .3487455    -7.29   0.000    -3.226941   -1.859884
------------------------------------------------------------------------------

. margins , at(ccred = (1 (1) 6) black = (0 1) denpmi =(0 1))

Predictive margins                              Number of obs     =      2,380
Model VCE    : Robust
```

. marginsplot, subtitle("Full Probit Model")
Variables that uniquely identify margins: ccred black denpmi