Testing and Fit

Objectives

Review basic approaches to hypothesis testing

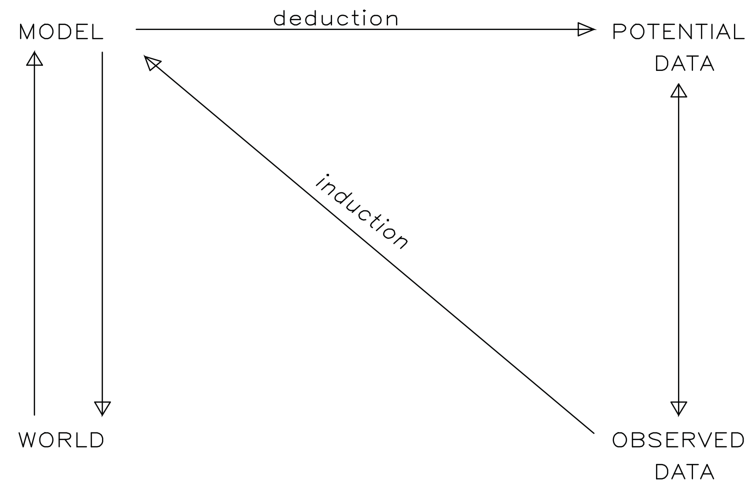
Introduce methods for evaluating the fit and influence of individual observations

Consider various measures of overall fit

Show how the BIC and AIC measures can be used to compare models

Hypothesis Testing

Barnett's Model of Inference



Tests of Individual Coefficients

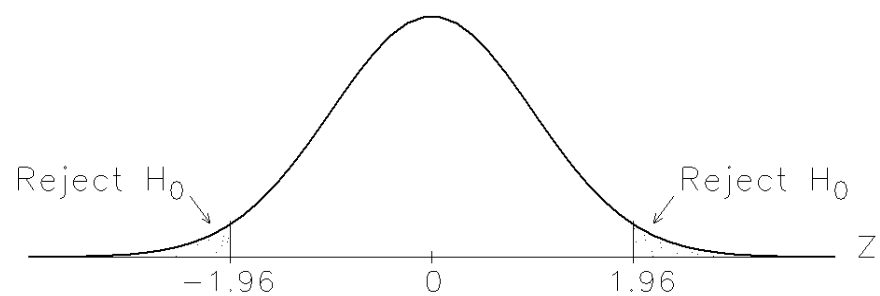
ML estimators are distributed asymptotically normal

$$\widehat{\beta}_k \overset{a}{\sim} \mathbf{N}\left(\beta_k, \mathrm{Var}\left(\widehat{\beta}_k\right)\right)$$

Consider H_0: $\beta_k = 0$, using the ML estimates $\widehat{\beta}_k$ and $\widehat{\sigma}_{\widehat{\beta}_k}$

$$z = \frac{\widehat{\beta}_k - 0}{\widehat{\sigma}_{\widehat{\beta}_k}}$$

Under the assumption justifying ML, if H_0 is true, then:



Types of Errors in Statistical Inference

Type I Error: Rejecting H_0 when it is true

Type II Error: Accepting H_0 when it is false

$H_0: \beta_k = 0$

	Accepts H_0	Rejects H_0
In fact, $\beta = 0$	No error	Type I (reject true)
In fact, $\beta \neq 0$	Type II (accept false)	No error

Question (for you)

When setting the alpha level, with which type of error are we concerned?

Computing a *z*-test

```
. logit lfp k5 k618 age wc hc lwg inc, nolog
------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          k5 |  -1.462913   .1970006    -7.43   0.000    -1.849027   -1.076799
        k618 |  -.0645707   .0680008    -0.95   0.342    -.1978499    .0687085
         age |  -.0628706   .0127831    -4.92   0.000    -.0879249   -.0378162
          wc |   .8072738   .2299799     3.51   0.000     .3565215    1.258026
          hc |   .1117336   .2060397     0.54   0.588    -.2920969     .515564
         lwg |   .6046931   .1508176     4.01   0.000     .3090961    .9002901
         inc |  -.0344464   .0082084    -4.20   0.000    -.0505346   -.0183583
       _cons |    3.18214   .6443751     4.94   0.000     1.919188    4.445092
------------------------------------------------------------------------------
```

Having young children has a significant effect on the probability of working (z=*-7.43*, p<*0.001* for a two-tailed test).

One Tailed Tests

I believe that having children can only have a negative effect on LFP.

Therefore, I divide `P>|z|` by 2 (0.342/2=.171)

Having older children does not significantly decrease a woman's probability of working ($z = -0.95$, *p=.17* for a one-tailed test)

Testing complex hypotheses

$\beta_1 - \beta_2 = 0$

$\beta_1 = \beta_2 = 0$

$\beta_1 = 2 * \beta_2$

Wald and LR tests

Compares estimates obtained after constraints imposed to estimates obtained without constraints

The *unconstrained* estimator $\hat{\beta}$ maximizes the log likelihood function

What constraint does the hypothesis $H_0: \beta = 0$ impose?

Wald Test

Steps

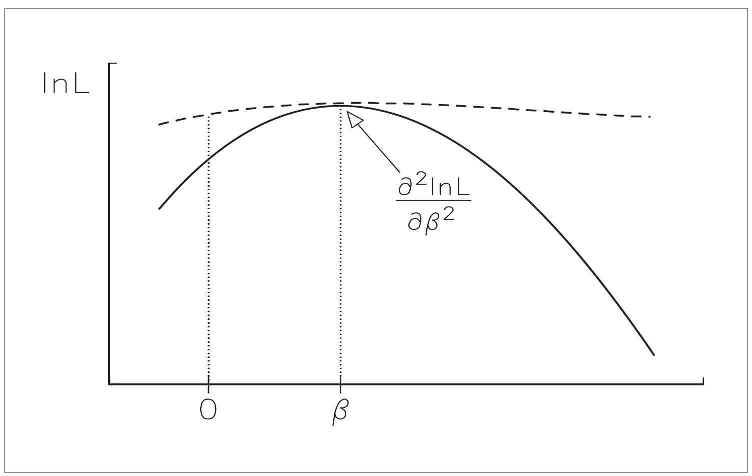
Estimate the model without constraints

Assess the constraint by considering:

The distance: $\hat{\beta} - 0$

The curvature of $\ln L$: $\frac{\partial^2 \ln L}{\partial \beta^2}$

Graphically



Wald Test

A flexible formula for complex hypothesis tests

Consider linear constraints of the form: $\mathbf{Q\beta} = \mathbf{r}$

$\boldsymbol{\beta}$ is the vector of parameters being tested

$\mathbf{Q}$ is a matrix of constants that combine the β 's

$\mathbf{r}$ is a vector of constants

$$W = \left[\mathbf{Q}\hat{\boldsymbol{\beta}} - \mathbf{r}\right]' \left[\mathbf{Q}\widehat{Var}\left(\hat{\boldsymbol{\beta}}\right)\mathbf{Q}'\right]^{-1} \left[\mathbf{Q}\hat{\boldsymbol{\beta}} - \mathbf{r}\right] \sim \chi^2$$, where

$\mathbf{Q}\hat{\boldsymbol{\beta}} - \mathbf{r}$ measures the distance between the estimated and hypothesized values

$\left[\mathbf{Q}\widehat{Var}\left(\hat{\boldsymbol{\beta}}\right)\mathbf{Q}'\right]^{-1}$ reflects the variability in the estimator

Wald Example 1

$H_0 : \beta_{k5} = 0$

```
. test k5

 ( 1)  k5 = 0.0

           chi2(  1) =   55.14
         Prob > chi2 =    0.0000
```

- The effect of having young children on the probability of entering the labor force is significant at the .01 level ($X^2(1)$=55.14, p<.01).

Note: $(-7.426)^2 = 55.145$

Wald Example 2

$H_0 : \beta_{wc} = \beta_{hc} = 0$

```
. test wc hc

 ( 1)  wc = 0.0
 ( 2)  hc = 0.0

           chi2(  2) =   17.66
         Prob > chi2 =    0.0001
```

- The hypothesis that the effects of the husband's and the wife's education are simultaneously equal to zero can be rejected at the .01 level ($X^2(2)$=17.66, p<.01).

Wald Example 3

$H_0 : \beta_{wc} = \beta_{hc}$

```
. test wc=hc

 ( 1)  wc - hc = 0.0

           chi2(  1) =    3.54
         Prob > chi2 =    0.0600
```

- It was hypothesized that the effects of husband's and wife's education are equal. This hypothesis is rejected at the .05 level ($X^2(1)$=3.54, p=.06).

Tests are probabilistic

Consider three sets of hypotheses

Testing two coefficients separately:

$H_a: \beta_1 = 0$

$H_b: \beta_2 = 0$

Test two coefficients are equal

$H_c: \beta_1 = \beta_2$

Test two coefficients are equal to 0

$H_d: \beta_1 = \beta_2 = 0$

Question (for you)

Are there any necessary links among these?

The LR Test

The idea of nested models

Constrained model (M_C) = *Unconstrained* (M_U) + *constraints*.

The *constrained* model is said to be *nested* in the *unconstrained* model.

Consider the following models:

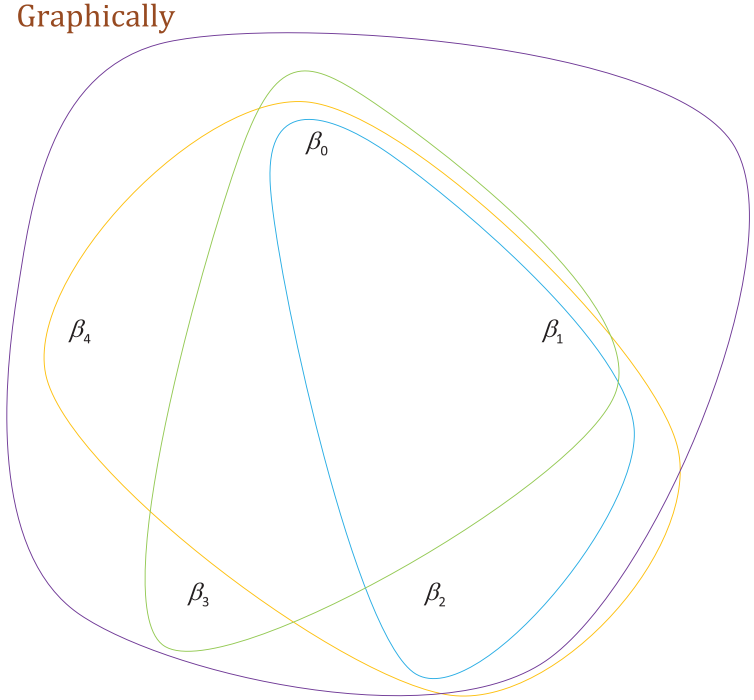
$$M_1 : \Pr(y=1 \mid x) = \Lambda(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$$

$$M_2 : \Pr(y=1 \mid x) = \Lambda(\beta_0 + \beta_1 x_1 + \beta_3 x_3)$$

$$M_3 : \Pr(y=1 \mid x) = \Lambda(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_4)$$

$$M_4 : \Pr(y=1 \mid x) = \Lambda(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4)$$

Graphically



In summary

M_C with L_C is nested in M_U with L_U

β_C can be created from β_U by imposing constraints

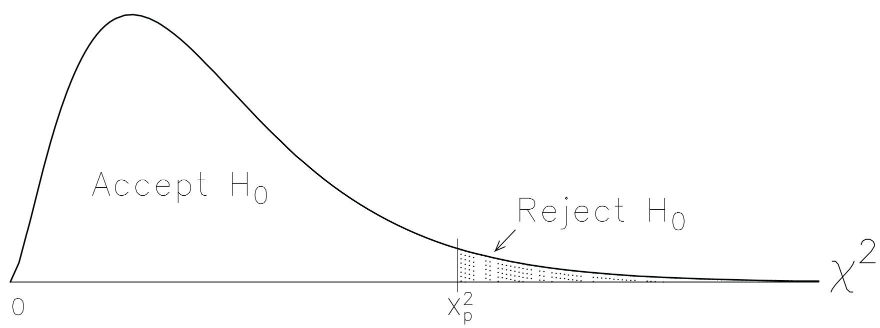
H_0: The constraints imposed on β_U are true

The *likelihood ratio statistic* equals:

$$G^2(M_C \mid M_U) = 2\ln L_U - 2\ln L_C$$

Under very general conditions, if H_0 is true

$G^2 \sim \chi^2$ with degrees of freedom equal to number of constraints



LR test

Steps

Estimate the unconstrained model

Store the results

Estimate the constrained model

Store the results

Calculate LR statistic using the command `lrtest`

LR Example 1

$H_0 : \beta_{k5} = 0$

Estimate the unconstrained model:

```
. quietly logit lfp k5 k618 age wc hc lwg inc
. estimates store full
```


Estimate the constrained model:

```
. quietly logit lfp k618 age wc hc lwg inc
. estimates store dropk5
```

Calculate LR statistic:

```
. lrtest full dropk5

Likelihood-ratio test                                 LR chi2(1)  =     66.48
(Assumption: dropk5 nested in full)                   Prob > chi2 =    0.0000
```

The effect of having young children is significant at the .01 level (LRX^2=66.5,p<.001).

LR Example 2

$H_0: \quad \beta_{WC} = \beta_{HC} = 0$

```
. quietly logit lfp k5 k618 age lwg inc
. estimates store dropwchc
. lrtest full dropwchc

Likelihood-ratio test                                 LR chi2(2)  =     18.50
(Assumption: dropwchc nested in full)                 Prob > chi2 =    0.0001
```

The hypothesis that the effects of the husband's and the wife's education are simultaneously equal to zero can be rejected at the .01 level (LRX^2=18.50, p=.001)

LR Example 3

All coefficients are simultaneously equal to zero

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$

```
. quietly logit lfp
. estimates store constant

. lrtest full constant

Likelihood-ratio test                                 LR chi2(7)  =    124.48
(Assumption: constant nested in full)                 Prob > chi2 =    0.0000
```

We can reject the hypothesis that all coefficients except the intercept are zero at the .01 level (LRX^2 =124.48, p<.001).

Or,

```
. logit lfp k5 k618 age wc hc lwg inc

Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood = -453.03179
Iteration 2:   log likelihood = -452.63343
Iteration 3:   log likelihood = -452.63296
Iteration 4:   log likelihood = -452.63296

Logistic regression                               Number of obs   =        753
                                                  LR chi2(7)      =     124.48
                                                  Prob > chi2     =     0.0000
Log likelihood = -452.63296                       Pseudo R2       =     0.1209

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          k5 |  -1.462913   .1970006    -7.43   0.000    -1.849027   -1.076799
        k618 |  -.0645707   .0680008    -0.95   0.342    -.1978499    .0687085
         age |  -.0628706   .0127831    -4.92   0.000    -.0879249   -.0378162
          wc |   .8072738   .2299799     3.51   0.000     .3565215    1.258026
          hc |   .1117336   .2060397     0.54   0.588    -.2920969     .515564
         lwg |   .6046931   .1508176     4.01   0.000     .3090961    .9002901
         inc |  -.0344464   .0082084    -4.20   0.000    -.0505346   -.0183583
       _cons |    3.18214   .6443751     4.94   0.000     1.919188    4.445092
------------------------------------------------------------------------------
```

Practical Issues in Computing the LR Test

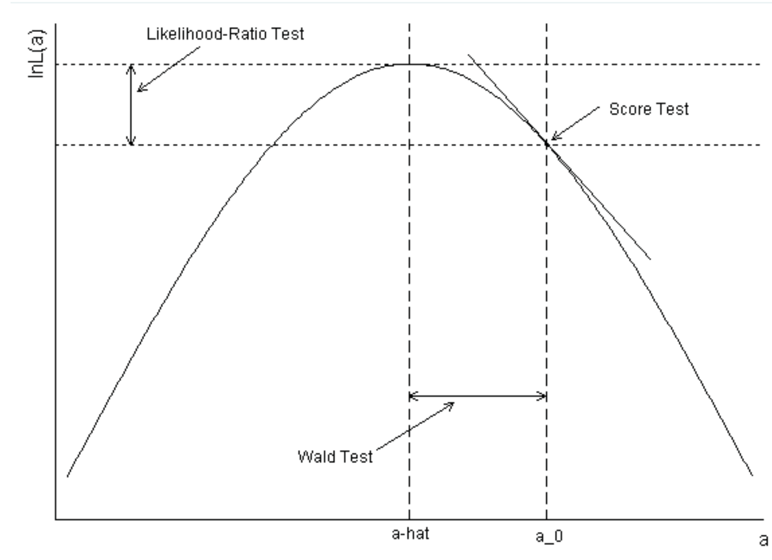
The same sample must be used for all models being compared

Since ML excludes cases with missing data, the sample size can change when the variables in a model change

To insure that the sample size does not change:

- Construct a data set that excludes every observation that has missing values for any of the variables used in any of the models being tested
- Or, impute missing values

Comparing the LR and Wald Tests



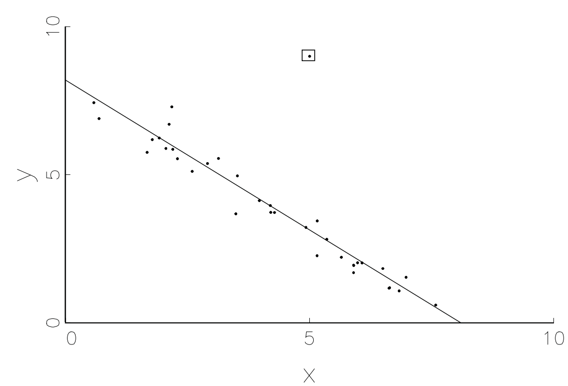
The LR and Wald tests are *asymptotically* equivalent, but in finite samples they will have different values.

Hypothesis	*df*	LR Test G^2	*p*	Wald Test *W*	*p*
$\beta_1 = 0$	1	66.5	<0.01	55.1	<0.01
$\beta_4 = \beta_5 = 0$	2	18.5	<0.01	17.7	<0.01
All slopes $= 0$	7	124.5	<0.01	95.0	<0.01

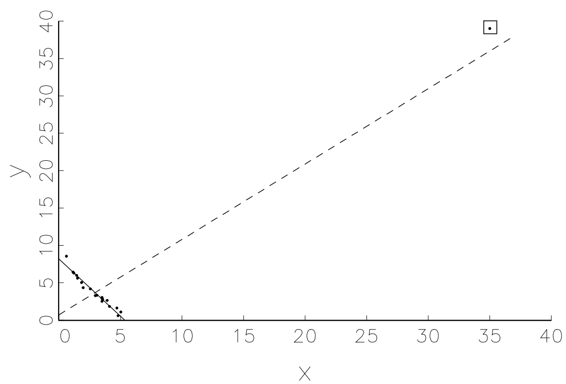
Residuals, Outliers and Influence

Residuals measure the difference between the model's prediction for a given case and the observed value for that case

Outliers are cases that fit poorly



Influential Observations are cases that have strong effects on the estimates obtained



Residuals for Binary Outcomes

$$r_j = \frac{y_j - m_j \widehat{\pi}_j}{\sqrt{m_j \widehat{\pi}_j \left(1 - \widehat{\pi}_j\right)}}$$

Where:

- j is used to index observations
- M_j is the total number of observations sharing j's covariate pattern
- y_j is the total number of positive responses among observations sharing j's covariate pattern.

Covariate patterns

Two observations are said to share the same covariate pattern if the independent variables for the two observations are identical


```
. spex icpsr_nes3
. logit didvote age white male collgrad
. predict prob1
. label var prob1 "Probability of voting."
. predict resid , residual
. label var resid "Pearson residuals."
. sort resid

. list resid prob1 didvote age white male collgrad in 1/10 , clean

           resid      prob1   didvote   age   white    male   collgrad
  1.   -5.645635   .9695801      0_No    86   1_Yes    0_No      1_Yes
  2.   -5.266423   .9651995      0_No    79   1_Yes    0_No      1_Yes
  3.   -3.967376   .9254445      0_No    38   1_Yes   1_Yes      1_Yes
  4.   -3.967376   .9254445     1_Yes    38   1_Yes   1_Yes      1_Yes
  5.   -3.967376   .9254445      0_No    38   1_Yes   1_Yes      1_Yes
  6.   -3.967376   .9254445     1_Yes    38   1_Yes   1_Yes      1_Yes
  7.   -3.967376   .9254445     1_Yes    38   1_Yes   1_Yes      1_Yes
  8.   -3.967376   .9254445      0_No    38   1_Yes   1_Yes      1_Yes
  9.   -3.646626   .9300596      0_No    63    0_No    0_No      1_Yes
 10.   -3.127295   .9072354      0_No    26   1_Yes   1_Yes      1_Yes
```

Standardized Pearson residual

Pearson residuals do not have a standard deviation equal to 1

Pregibon proposed the *standardized Pearson residual* which fully standardizes the residuals

The option `rstandard` (or `rs`) generates Pearson residuals normalized to have an expected standard deviation equal to 1

Data

```
. codebook havesex age female white  , compact

Variable     Obs Unique      Mean      Min       Max  Label
-------------------------------------------------------------------------------
havesex     2091      2  .4002869        0         1  Ever have sex? (1=yes)
age         2091     97  16.06536  11.4167   20.6667  Respondent's age
female      2091      2  .5088474        0         1  Female? (1=yes)
white       2091      2  .5791487        0         1  R is non-hispanic white.
-------------------------------------------------------------------------------

. sum havesex age female white

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     havesex |      2091    .4002869    .4900736          0          1
         age |      2091    16.06536    1.727781    11.4167    20.6667
      female |      2091    .5088474    .5000413          0          1
       white |      2091    .5791487    .4938138          0          1
```


Logit

```
. logit havesex age female white

Iteration 0:   log likelihood = -1407.5103
Iteration 1:   log likelihood = -1223.4792
Iteration 2:   log likelihood = -1221.8798
Iteration 3:   log likelihood = -1221.8787
Iteration 4:   log likelihood = -1221.8787

Logistic regression                               Number of obs   =       2091
                                                  LR chi2(3)      =     371.26
                                                  Prob > chi2     =     0.0000
Log likelihood = -1221.8787                       Pseudo R2       =     0.1319

------------------------------------------------------------------------------
     havesex |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |    .534397   .0318338    16.79   0.000     .4720038    .5967901
      female |  -.1039031   .0978596    -1.06   0.288    -.2957044    .0878981
       white |  -.5570166   .0992522    -5.61   0.000    -.7515473   -.3624859
       _cons |  -8.701246   .5232923   -16.63   0.000     -9.72688   -7.675612
------------------------------------------------------------------------------

. predict brmrstd , rs
```



```
. sort brmrstd , stable

. *list large negative residuals (cases overpredicted by model)
. list brmrstd age female white if  brmrstd<-2 , clean

         brmrstd       age   female   white
  1.   -2.683817   20.1667    1_Yes    0_No
  2.   -2.660666   18.9167     0_No   1_Yes
  3.   -2.660666   18.9167     0_No   1_Yes
  4.   -2.660666   18.9167     0_No   1_Yes
  5.   -2.516644   15.0833    1_Yes    0_No
  6.   -2.516644   15.0833    1_Yes    0_No
  7.   -2.516644   15.0833    1_Yes    0_No
  8.   -2.516644   15.0833    1_Yes    0_No
  9.   -2.516644   15.0833    1_Yes    0_No
 10.   -2.516644   15.0833    1_Yes    0_No
 11.   -2.516644   15.0833    1_Yes    0_No
 12.   -2.516644   15.0833    1_Yes    0_No
 13.   -2.516644   15.0833    1_Yes    0_No
 14.   -2.516644   15.0833    1_Yes    0_No
 15.   -2.516644   15.0833    1_Yes    0_No
 16.   -2.516644   15.0833    1_Yes    0_No
 17.   -2.516644   15.0833    1_Yes    0_No
 18.   -2.455927   17.9167     0_No    0_No
 19.   -2.455927   17.9167     0_No    0_No
 20.   -2.455927   17.9167     0_No    0_No
 21.   -2.455927   17.9167     0_No    0_No
 22.   -2.455927   17.9167     0_No    0_No
 23.   -2.455927   17.9167     0_No    0_No
 24.   -2.455927   17.9167     0_No    0_No
```

```
 25.   -2.339916   17.5833     0_No   1_Yes
 26.   -2.339916   17.5833     0_No   1_Yes
 27.   -2.339916   17.5833     0_No   1_Yes
 28.   -2.339916   17.5833     0_No   1_Yes
 29.   -2.339916   17.5833     0_No   1_Yes
 30.   -2.339916   17.5833     0_No   1_Yes
 31.   -2.339916   17.5833     0_No   1_Yes
 32.   -2.339916   17.5833     0_No   1_Yes
 33.   -2.275481     16.25     0_No   1_Yes
 34.   -2.275481     16.25     0_No   1_Yes
 35.   -2.275481     16.25     0_No   1_Yes
 36.   -2.275481     16.25     0_No   1_Yes
 37.   -2.275481     16.25     0_No   1_Yes
 38.   -2.275481     16.25     0_No   1_Yes
 39.   -2.275481     16.25     0_No   1_Yes
 40.   -2.275481     16.25     0_No   1_Yes
 41.   -2.275481     16.25     0_No   1_Yes
 42.   -2.275481     16.25     0_No   1_Yes
 43.   -2.275481     16.25     0_No   1_Yes
 44.   -2.275481     16.25     0_No   1_Yes
 45.   -2.275481     16.25     0_No   1_Yes
 46.   -2.275481     16.25     0_No   1_Yes
 47.   -2.130246     18.25    1_Yes    0_No
 48.   -2.130246     18.25    1_Yes    0_No
 49.   -2.130246     18.25    1_Yes    0_No
 50.   -2.130246     18.25    1_Yes    0_No
 51.   -2.130246     18.25    1_Yes    0_No
 52.   -2.130246     18.25    1_Yes    0_No
```



```
. *list large positive residuals (cases underpredicted by model)
. list brmrstd age female white if brmrstd>2 , clean

         brmrstd       age   female   white
2020.   2.044268   17.0833    1_Yes   1_Yes
2021.   2.044268   17.0833    1_Yes   1_Yes
2022.   2.044268   17.0833    1_Yes   1_Yes
2023.   2.044268   17.0833    1_Yes   1_Yes
2024.   2.044268   17.0833    1_Yes   1_Yes
2025.   2.044268   17.0833    1_Yes   1_Yes
2026.   2.044268   17.0833    1_Yes   1_Yes
2027.   2.044268   17.0833    1_Yes   1_Yes
2028.   2.044268   17.0833    1_Yes   1_Yes
2029.   2.128586   13.8333     0_No    0_No
2030.   2.128586   13.8333     0_No    0_No
2031.   2.128586   13.8333     0_No    0_No
2032.   2.128586   13.8333     0_No    0_No
2033.   2.128586   13.8333     0_No    0_No
2034.   2.223591   14.3333    1_Yes    0_No
2035.   2.223591   14.3333    1_Yes    0_No
2036.   2.223591   14.3333    1_Yes    0_No
2037.   2.223591   14.3333    1_Yes    0_No
2038.   2.223591   14.3333    1_Yes    0_No
2039.   2.223591   14.3333    1_Yes    0_No
2040.   2.223591   14.3333    1_Yes    0_No
2041.   2.223591   14.3333    1_Yes    0_No
2042.   2.223591   14.3333    1_Yes    0_No
2043.   2.281919      14.5     0_No    0_No
2044.   2.281919      14.5     0_No    0_No
2045.    2.30179      14.5    1_Yes    0_No
2046.    2.30179      14.5    1_Yes    0_No
```

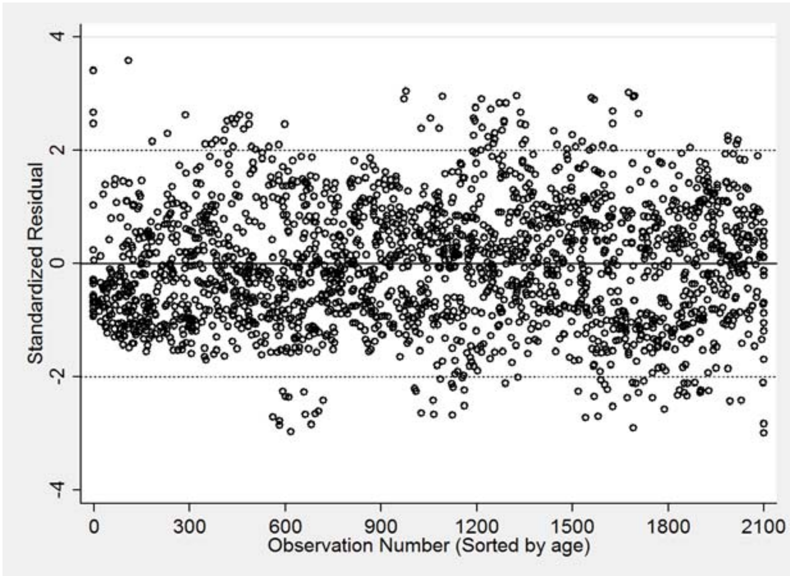


```
2047.    2.30179      14.5    1_Yes    0_No
2048.    2.30179      14.5    1_Yes    0_No
2049.    2.30179      14.5    1_Yes    0_No
2050.    2.30179      14.5    1_Yes    0_No
2051.     2.3142   16.5833    1_Yes   1_Yes
2052.     2.3142   16.5833    1_Yes   1_Yes
2053.     2.3142   16.5833    1_Yes   1_Yes
2054.     2.3142   16.5833    1_Yes   1_Yes
2055.     2.3142   16.5833    1_Yes   1_Yes
2056.     2.3142   16.5833    1_Yes   1_Yes
2057.   2.516161   12.8333     0_No    0_No
2058.   2.533524   16.6667    1_Yes   1_Yes
2059.   2.533524   16.6667    1_Yes   1_Yes
2060.   2.533524   16.6667    1_Yes   1_Yes
2061.   2.533524   16.6667    1_Yes   1_Yes
2062.   2.533524   16.6667    1_Yes   1_Yes
2063.   2.533524   16.6667    1_Yes   1_Yes
2064.   2.533524   16.6667    1_Yes   1_Yes
2065.   2.533524   16.6667    1_Yes   1_Yes
2066.   2.533524   16.6667    1_Yes   1_Yes
2067.   2.533524   16.6667    1_Yes   1_Yes
2068.   2.533524   16.6667    1_Yes   1_Yes
2069.   2.533524   16.6667    1_Yes   1_Yes
2070.   2.533524   16.6667    1_Yes   1_Yes
2071.   2.533524   16.6667    1_Yes   1_Yes
2072.   2.657436        16     0_No    0_No
2073.   2.657436        16     0_No    0_No
2074.   2.657436        16     0_No    0_No
2075.   2.657436        16     0_No    0_No
2076.   2.657436        16     0_No    0_No
2077.   2.657436        16     0_No    0_No
```



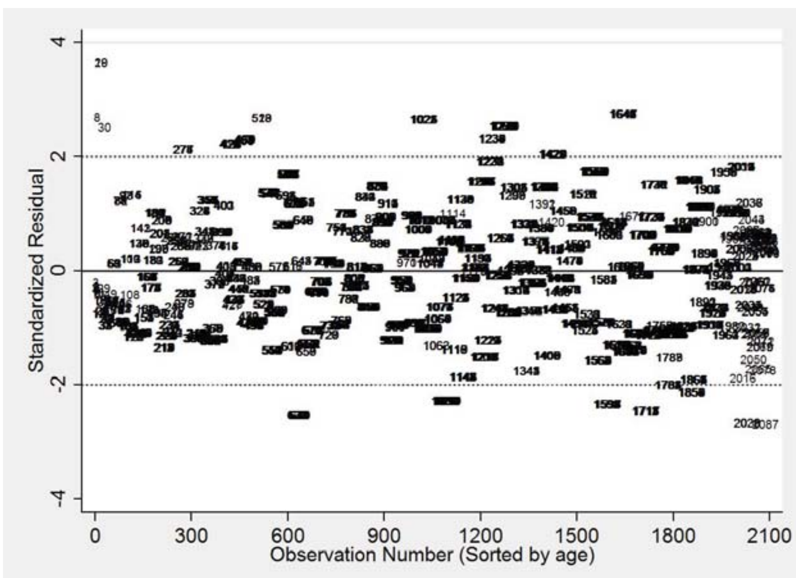
```
2078.   2.675848   14.6667     0_No    0_No
2079.   2.675848   14.6667     0_No    0_No
2080.   2.675848   14.6667     0_No    0_No
2081.   2.689943   12.5833     0_No    0_No
2082.   2.741484   17.6667    1_Yes   1_Yes
2083.   2.741484   17.6667    1_Yes   1_Yes
2084.   2.741484   17.6667    1_Yes   1_Yes
2085.   2.741484   17.6667    1_Yes   1_Yes
2086.   2.741484   17.6667    1_Yes   1_Yes
2087.   2.741484   17.6667    1_Yes   1_Yes
2088.   2.741484   17.6667    1_Yes   1_Yes
2089.   2.741484   17.6667    1_Yes   1_Yes
2090.   3.642644     12.75     0_No    0_No
2091.   3.642644     12.75     0_No    0_No
```

An *index plot* can be used to search for outliers



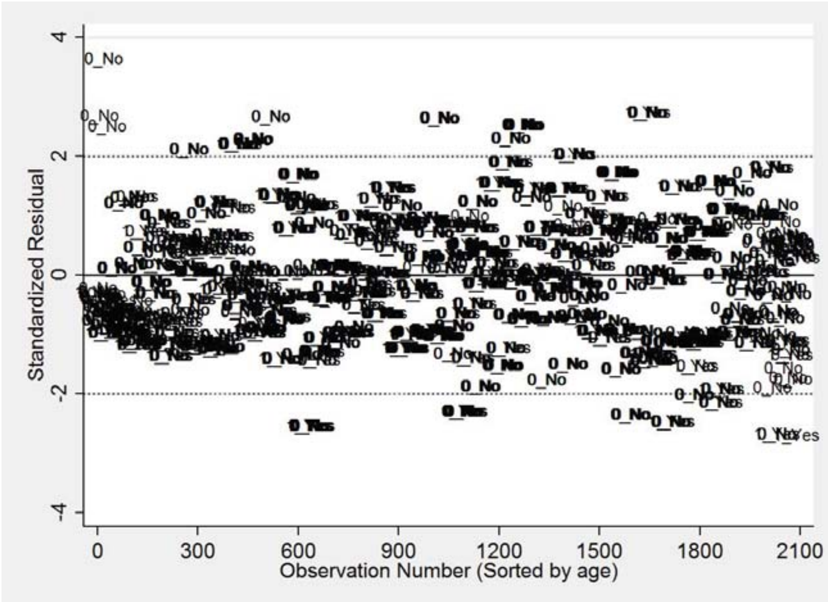

```
. predict brmrstd, rs
. label var brmrstd "Standardized Residual"
. sort age, stable

. generate index = _n
. label var index "Observation Number (Sorted by age)"

. graph twoway scatter brmrstd index, msymbol(Oh) mcolor(black) ///
>         xtitle("Observation Number (Sorted by age)") xlabel(0(300)2100) ///
>         ylabel(-4(2)4) yline(0, lpattern(solid)) ///
>         yline(2 -2, lpattern(dot))
```


Adding id numbers




```
. graph twoway scatter brmrstd index, msymbol(none) ///
>         mlabel(index) mcolor(black) mlabposition(0) ///
>         xtitle("Observation Number (Sorted by age)") xlabel(0(300)2100) ///
>         ylabel(-4(2)4) yline(0, lpattern(solid)) ///
>         yline(2 -2, lpattern(dot))
```

Index plot with abpledge identifiers

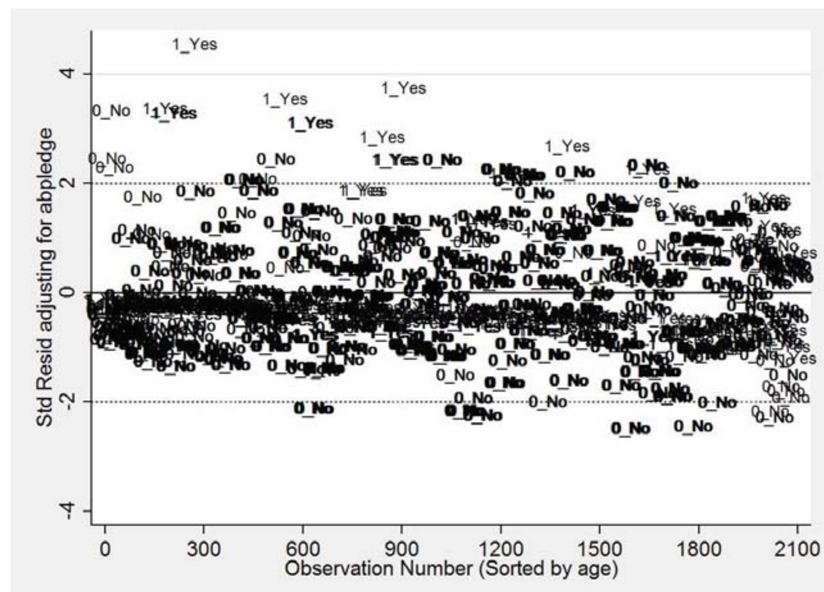



```
. graph twoway scatter brmrstd index, msymbol(none) ///
>         mlabel(abpledge) mcolor(black) mlabposition(0) ///
>         xtitle("Observation Number (Sorted by age)") xlabel(0(300)2100) ///
>         ylabel(-4(2)4) yline(0, lpattern(solid)) ///
>         yline(2 -2, lpattern(dot))
```


Include abpledge in model

```
. logit havesex age female white abpledge, nolog

Logistic regression                               Number of obs   =       2091
                                                  LR chi2(4)      =     478.51
                                                  Prob > chi2     =     0.0000
Log likelihood = -1168.2554                       Pseudo R2       =     0.1700

------------------------------------------------------------------------------
     havesex |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .5339842   .0326036    16.38   0.000     .4700824     .597886
      female |  -.0085701   .1006576    -0.09   0.932    -.2058554    .1887153
       white |  -.6461797   .1025277    -6.30   0.000    -.8471303    -.445229
    abpledge |  -1.888253   .2127139    -8.88   0.000    -2.305164   -1.471341
       _cons |  -8.516833   .5347806   -15.93   0.000    -9.564984   -7.468683
------------------------------------------------------------------------------S
```


Plotting the new residuals:



```
. graph twoway scatter brmrstd2 index, msymbol(none) ///
>         mlabel(abpledge) mcolor(black) mlabposition(0) ///
>         xtitle("Observation Number (Sorted by age)") xlabel(0(300)2100) ///
>         ylabel(-4(2)4) yline(0, lpattern(solid)) ///
>         yline(2 -2, lpattern(dot)) jitter(30)
```


Summary

```
. sum brmrstd brmrstd2

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     brmrstd |      2091   -.0319772    1.047555  -2.683817   3.642644
    brmrstd2 |      2091   -.0248651    1.029938   -2.47515   4.547635
```

Question (for you)

How did the addition of abpledge impact the fit of the model?

Influence for Binary Outcomes

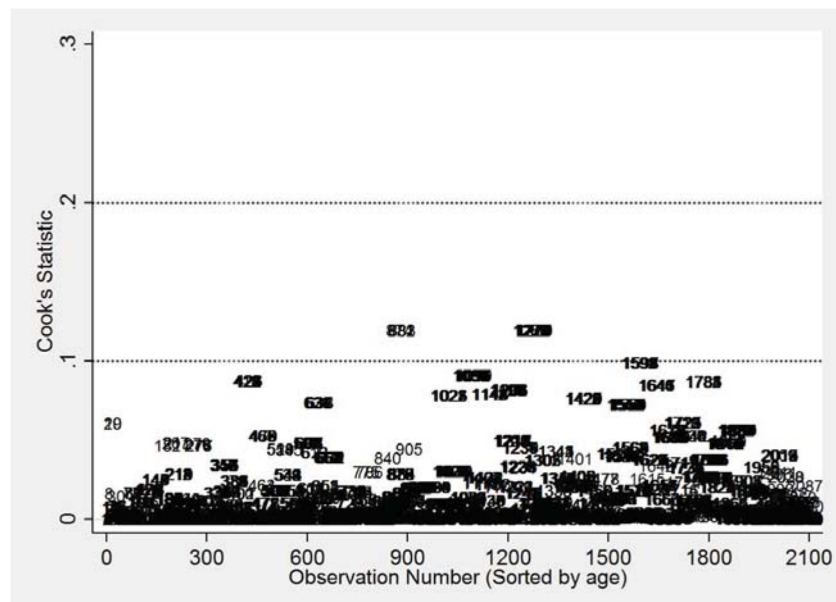
Influence indicates what impact an observation has on the $\hat{\beta}$'s

To determine influential observations (i.e., *high-leverage* points), compute the change in $\hat{\beta}$ when dropping that observation

Since it is computationally impractical to estimate the model N times, use derived approximations that only require estimating the model once

Cook's distance summarizes the effect of removing a single observation: There is not critical value for significantly influential observations

Index Plots of Influential Cases



```
. predict brmcook, dbeta
. label var brmcook "Cook's Statistic"
. graph twoway scatter brmcook index, ///
>         mlabel(index) msymbol(none) mlabposition(0) ///
>         xtitle("Observation Number(Sorted by age)") xlabel(0(300)2100) ///
>         ylabel(0(.1).3) yline(.1 .2, lpattern(dot))
```


Scalar Measures of Fit

A single number to summarize the overall fit of the model

But, Long (1997): I am unaware of convincing evidence that selecting a model that maximizes the value of a given measure of fit results in a model that is optimal in any sense other than the model having a larger value of that measure.

R^2 in the LRM

In the LRM:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3}$$

R^2 can be defined in multiple ways

% Explained Variation $\quad \dfrac{TSS - RSS}{TSS} = 1 - \dfrac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$

Ratio $Var(y)$ to $Var(\hat{y})$ $\quad \dfrac{\widehat{Var}(\hat{y})}{\widehat{Var}(y)} = \dfrac{\widehat{Var}(\hat{y})}{\widehat{Var}(\hat{y}) + \widehat{Var}(\hat{\varepsilon})}$

Transform LR $\quad 1 - \left[\dfrac{L(M_\alpha)}{L(M_\beta)}\right]^{2/N}$

Transform F $\quad \dfrac{FK}{FK + (N - K - 1)}$

Pseudo-R^2 in the BRM

Percent of Explained "Variation"

Define $\hat{y}$ as $\hat{\pi} = \widehat{\Pr}(y \mid \mathbf{x})$:

$$R^2_{\text{Efron}} = 1 - \frac{\sum_{i=1}^{N}\left(y_i - \hat{\pi}_i\right)^2}{\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2}$$

LR Index (aka McFadden's R^2, and Stata's Pseudo R^2)

$$R^2_{\text{McF}} = 1 - \frac{\ln \hat{L}\left(M_\beta\right)}{\ln \hat{L}\left(M_\alpha\right)}$$

ML R^2 (aka Cox-Snell R^2)

$$R^2_{\text{ML}} = 1 - \left[\frac{L\left(M_\alpha\right)}{L\left(M_\beta\right)}\right]^{2/N}$$

If $G^2 = -2\ln\left[L\left(M_\alpha\right)/L\left(M_\beta\right)\right]$,

$R^2_{\text{ML}} = 1 - \exp(-G^2/N)$

Cragg and Uhler's R^2 (aka Nagelkerke R^2)

$$R^2_{\text{C\&U}} = \frac{R^2_{\text{ML}}}{\max R^2_{\text{ML}}} = \frac{1 - \left[L\left(M_\alpha\right)/L\left(M_\beta\right)\right]^{2/N}}{1 - L\left(M_\alpha\right)^{2/N}}$$

McKelvey and Zavoina's R^2:

For models defined in terms of $y^* = \mathbf{x}\boldsymbol{\beta} + \varepsilon$,

$$R^2_{M\&Z} = \frac{\widehat{Var}\left(\hat{y}^*\right)}{\widehat{Var}\left(\hat{y}^*\right)} = \frac{\widehat{Var}\left(\hat{y}^*\right)}{\widehat{Var}\left(\hat{y}^*\right) + Var(\varepsilon)}$$

Where,

$$\widehat{Var}\left(\hat{y}^*\right) = \hat{\boldsymbol{\beta}}'\,\widehat{Var}(\mathbf{x})\hat{\boldsymbol{\beta}}$$

Count-R^2

Let the observed y equal 0 or 1, then:

$\widehat{\pi}_i = \widehat{\Pr}(y=1 \mid \mathbf{x}_i) = F(\mathbf{x}_i \hat{\beta})$

Define the expected outcome $\hat{y}$ as

$$\hat{y}_i = \begin{cases} 0 & \text{if } \widehat{\pi}_i \le 0.5 \\ 1 & \text{if } \widehat{\pi}_i > 0.5 \end{cases}$$

The proportion of correct predictions:

$$R^2_{\text{Count}} = \frac{1}{N} \sum_j n_{jj}$$

Question (for you)

Without knowledge about the independent variables, you can always correctly predict at least 50 percent of the cases. How?

Observed Outcome	Predicted Outcome $\hat{y}=1$	Predicted Outcome $\hat{y}=0$	Row Total
$y=1$	n_{11} :: correct	n_{12} :: incorrect	n_{1+}
$y=0$	n_{21} :: incorrect	n_{22} :: correct	n_{2+}
Column Total	n_{+1}	n_{+2}	N

Adjusted Count R^2

Adjust for simply guessing the largest margin

Subtract out the largest margin:

$$R^2_{\text{AdjCount}} = \frac{\sum_j \#\text{ of correct for row} - \text{largest margin}}{N - \text{largest margin}}$$

Interpretation:

Knowledge of the independent variables, compared to basing our prediction on the marginal distributions alone, reduces the error in prediction by $[100 \times R^2_{\text{AdjCount}}]$ percent

Information measures

A different approach is based on ideas from information theory

These include:

Akaike's information criterion (AIC)

Bayesian information criterion (BIC)

Both information criteria (IC) are based on the notion:

IC	=	Fit	+	Complexity
	=	(log likelihood)	+	(function of N and # of parameters)

Since fit is negative and complexity is positive, a smaller IC indicates a better fit

Formula

AIC = -2lnL + 2*k

BIC = -2lnL + ln(N)*k

A Question (for you)

Which IC has a larger penalty for model complexity?

Varieties of BIC

BIC

BIC'

BIC^{STATA}

What matters is the comparison within variety

For two models the difference is the same for all variations:

$BIC_1 - BIC_2 = BIC_1' - BIC_2' = BIC_1^S - BIC_2^S$

Chose M_2 if it is smaller (e.g., most negative):

$BIC_1 - BIC_2 > 0$

$BIC_1 > BIC_2$

Chose M_1 if it is smaller (e.g., most negative):

$BIC_1 - BIC_2 < 0$

$BIC_1 < BIC_2$

Degrees of evidence for the BIC
From Raftery (1996)

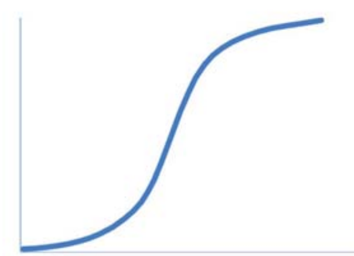
Absolute Difference	Evidence
0-2	Weak
2-6	Positive
6-10	Strong
>10	Very Strong

Example

```
. gen age2 = age*age
. label var age2 "Age-squared"

. quietly logit havesex age female white abpledge, nolog
. quietly fitstat, save
```



```
. quietly logit havesex age age2 female, nolog
. fitstat , dif

Measures of Fit for logit of havesex

                               Current          Saved        Difference
Model:                           logit          logit
N:                                2091           2091                 0
Log-Lik Intercept Only       -1407.510      -1407.510             0.000
Log-Lik Full Model           -1236.600      -1168.255           -68.344
D                         2473.199(2087)  2336.511(2086)     136.688(1)
LR                           341.821(3)     478.510(4)       136.688(1)
Prob > LR                        0.000          0.000             0.000
McFadden's R2                    0.121          0.170            -0.049
McFadden's Adj R2                0.119          0.166            -0.048
ML (Cox-Snell) R2                0.151          0.205            -0.054
Cragg-Uhler(Nagelkerke) R2       0.204          0.276            -0.073
McKelvey & Zavoina's R2          0.209          0.303            -0.094
Efron's R2                       0.152          0.205            -0.054
Variance of y*                   4.159          4.720            -0.561
Variance of error                3.290          3.290             0.000
Count R2                         0.677          0.699            -0.022
Adj Count R2                     0.194          0.247            -0.054
AIC                              1.187          1.122             0.064
AIC*n                         2481.199       2346.511           134.688
BIC                         -13482.746     -13611.789           129.043
BIC'                          -318.885       -447.928           129.043
BIC used by Stata             2503.781       2374.738           129.043
AIC used by Stata             2481.199       2346.511           134.688

Difference of  129.043 in BIC' provides very strong support for saved model.
```

End T&F