Biostatistics 140.624
2011 EXAM

STATA LOG

(*NEEDED TO ANSWER EXAM QUESTIONS*)

1. Multiple Linear Regression

```
. * Part 0.  Data management  MULTIPLE LINEAR REGRESSION
.
.
. * Input data
.
. use  fev.dta, clear

.
.
. codebook

-------------------------------------------------------------------------------------------------------
Id                                                                                          (unlabeled)
-------------------------------------------------------------------------------------------------------

                  type:  numeric (long)

                 range:  [201,90001]                  units:  1
         unique values:  654                      missing .:  0/654

                  mean:   37169.6
             std. dev:   23690.9

           percentiles:        10%       25%       50%       75%       90%
                              6145     15801     36071     53651     73342

-------------------------------------------------------------------------------------------------------
Age                                                                                         (unlabeled)
-------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [3,19]                       units:  1
         unique values:  17                       missing .:  0/654

                  mean:   9.93119
             std. dev:   2.95394

           percentiles:        10%       25%       50%       75%       90%
                                 6         8        10        12        14

-------------------------------------------------------------------------------------------------------
fev                                                                                         (unlabeled)
-------------------------------------------------------------------------------------------------------

                  type:  numeric (double)

                 range:  [.791,5.793]                 units:  .001
         unique values:  575                      missing .:  0/654

                  mean:   2.63678
             std. dev:   .867059

           percentiles:        10%       25%       50%       75%       90%
                             1.612     1.979    2.5475      3.12     3.816
```

1. Multiple Linear Regression (cont'd)

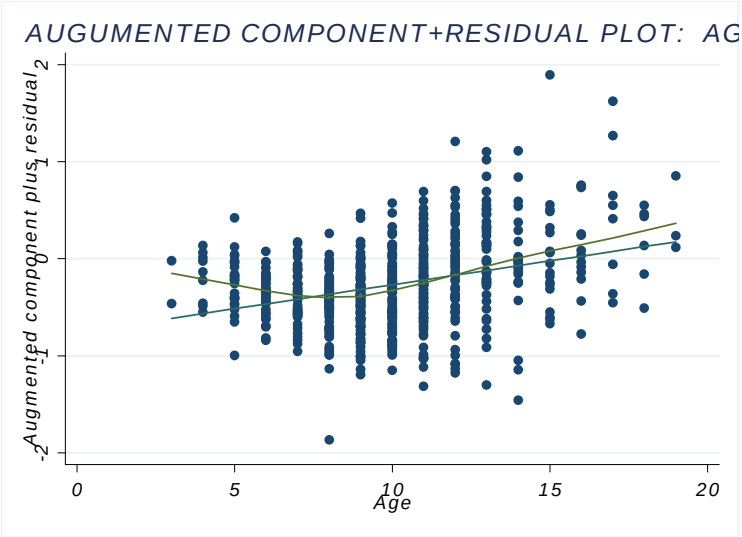
```
-------------------------------------------------------------------------------------------------
Hgt                                                                                  (unlabeled)
-------------------------------------------------------------------------------------------------

                  type:  numeric (double)

                 range:  [46,74]                      units:  .1
         unique values:  56                       missing .:  0/654

                  mean:   61.1436
             std. dev:   5.70351

           percentiles:        10%       25%       50%       75%       90%
                                53        57      61.5      65.5      68.5

-------------------------------------------------------------------------------------------------
Sex                                                                                  (unlabeled)
-------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/654

            tabulation:  Freq.  Value
                           318  0
                           336  1

-------------------------------------------------------------------------------------------------
Smoke                                                                                (unlabeled)
-------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/654

            tabulation:  Freq.  Value
                           589  0
                            65  1

.
.
.
.
```

1. Multiple Linear Regression (cont'd)

```
. * Part a.  MULTIPLE LINEAR REGRESSION
.
.
.
. regress fev Smoke Sex Age Hgt

      Source |       SS       df       MS              Number of obs =     654
-------------+------------------------------           F(  4,   649) =  560.02
       Model |  380.640282     4  95.1600706           Prob > F      =  0.0000
    Residual |  110.279554   649  .169922271           R-squared     =  0.7754
-------------+------------------------------           Adj R-squared =  0.7740
       Total |  490.919836   653   .75179148           Root MSE      =  .41222

------------------------------------------------------------------------------
         fev |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Smoke |  -.0872464   .0592535    -1.47   0.141    -.2035981    .0291053
         Sex |   .1571029   .0332071     4.73   0.000     .0918967    .2223092
         Age |   .0655093   .0094886     6.90   0.000     .0468774    .0841413
         Hgt |   .1041994   .0047577    21.90   0.000     .0948571    .1135418
       _cons |  -4.456974   .2228392   -20.00   0.000    -4.894547   -4.019401
------------------------------------------------------------------------------

.
. * Multicollinearity
.
. vif

    Variable |       VIF       1/VIF
-------------+----------------------
         Age |      3.02    0.331234
         Hgt |      2.83    0.353391
       Smoke |      1.21    0.826744
         Sex |      1.06    0.943194
-------------+----------------------
    Mean VIF |      2.03

.
.
.
```

1. Multiple Linear Regression (cont'd)

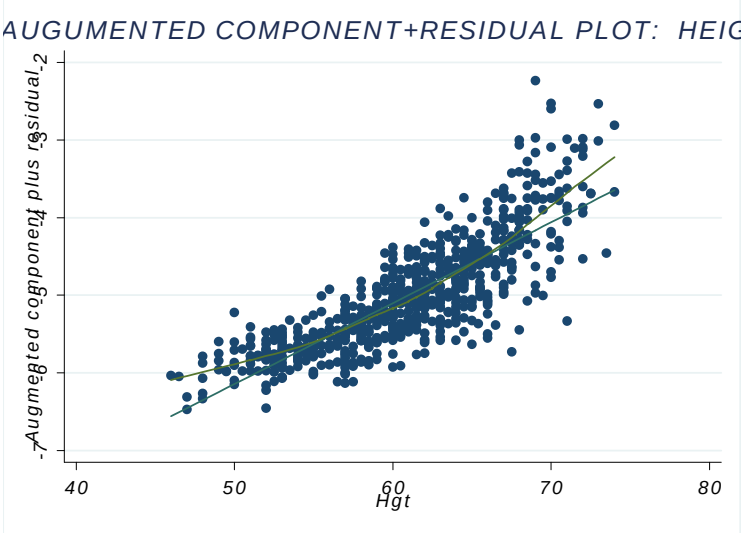
```
. * Scale for age
.
. acprplot Age ,  lowess title("AUGUMENTED COMPONENT+RESIDUAL PLOT:  AGE") graphregion(fcolor(white)
> )

. graph export exam2011_mlr/acprplotAge.emf, replace
(file C:\jt\bio624\2011\Exam 2011\exam2011_mlr/acprplotAge.emf written in Enhanced Metafile format)
```



.
.

1. Multiple Linear Regression (cont'd)

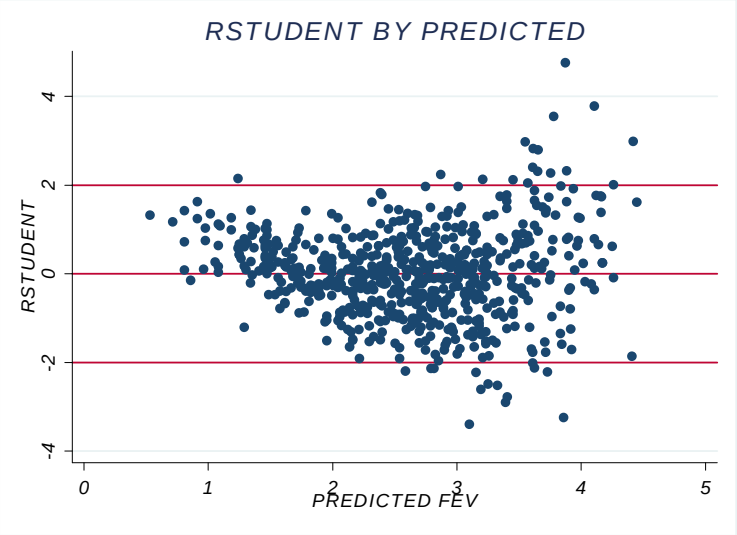
```
. * Scale for Height
.
. acprplot Hgt ,  lowess title("AUGUMENTED COMPONENT+RESIDUAL PLOT:  HEIGHT") graphregion(fcolor(whi
> te))

. graph export exam2011_mlr/acprplotHeight.emf, replace
(file C:\jt\bio624\2011\Exam 2011\exam2011_mlr/acprplotHeight.emf written in Enhanced Metafile forma
> t)
```




```
.
.
. * Predicteds
. predict y_hat
(option xb assumed; fitted values)

. label variable y_hat "PREDICTED FEV"

.
.
.
```

1. Multiple Linear Regression (cont'd)

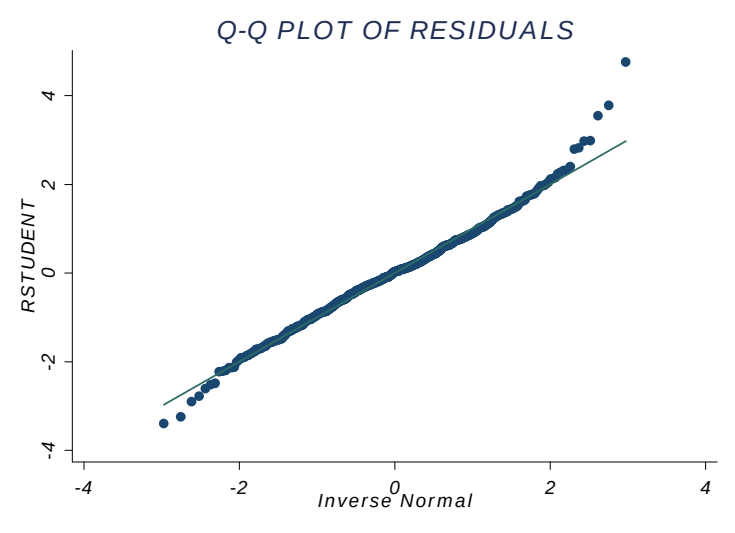
```
. * Studentized residuals
.
. predict estud, rstudent

. label variable estud "RSTUDENT"

.
. graph twoway scatter estud y_hat , yline(-2 0 2) title("RSTUDENT BY PREDICTED")  graphregion(fcolo
> r(white))

. graph export exam2011_mlr/rstudent.wmf,replace
(file C:\jt\bio624\2011\Exam 2011\exam2011_mlr/rstudent.wmf written in Windows Metafile format)
```



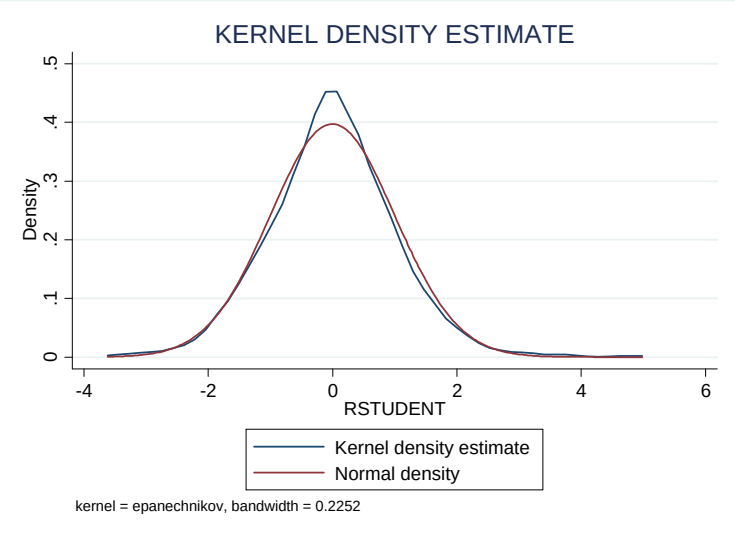
.
.
.

1. Multiple Linear Regression (cont'd)

```
. * Q-Q plot
.
. qnorm estud , title("Q-Q PLOT OF RESIDUALS")   graphregion(fcolor(white))

. graph export exam2011_mlr/q-q.wmf, replace
(file C:\jt\bio624\2011\Exam 2011\exam2011_mlr/q-q.wmf written in Windows Metafile format)
```




```
.
. * Kernel density plot
.
. kdensity estud , normal title("KERNEL DENSITY ESTIMATE")  graphregion(fcolor(white))

. graph export exam2011_mlr/kdensity.wmf, replace
(file C:\jt\bio624\2011\Exam 2011\exam2011_mlr/kdensity.wmf written in Windows Metafile format)
```



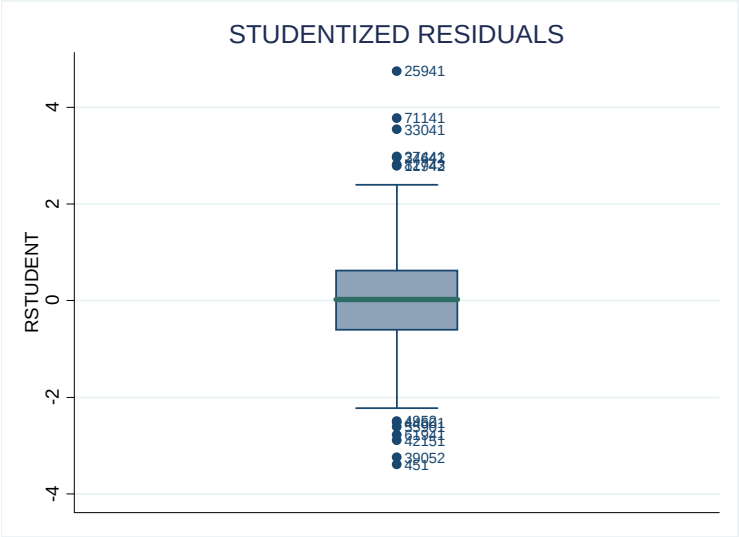

```
.
.
```

1. Multiple Linear Regression (cont'd)

```
.
. * Shapiro-Wilk Test
.
. swilk estud

                  Shapiro-Wilk W test for normal data

    Variable |    Obs       W           V         z       Prob>z
-------------+-------------------------------------------------
       estud |    654    0.98924      4.615     3.720    0.00010

.
.
.
```

1. Multiple Linear Regression (cont'd)

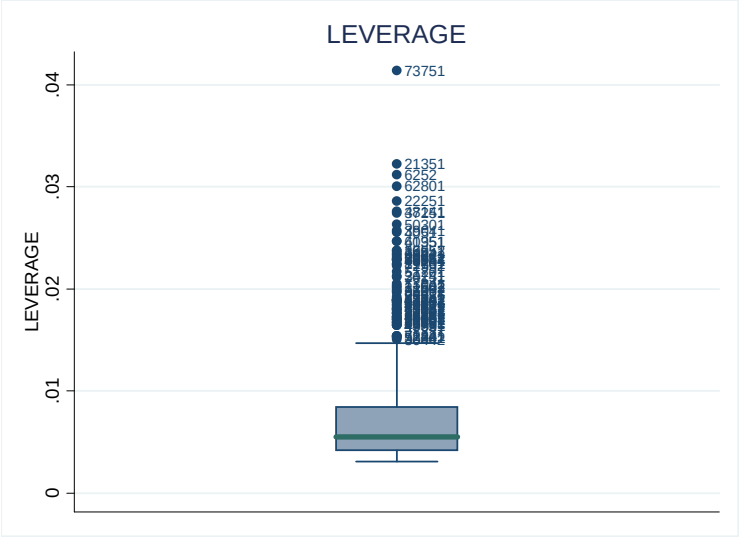
```
. * Boxplot studentized resIduals
.
. graph box estud,   marker(1,mlab(Id)) outergap(200) medtype(cline) medline(lwidth(1.0)) title("STU
> DENTIZED RESIDUALS") graphregion(fcolor(white))

. graph export exam2011_mlr/boxrstudent.wmf, replace
(file C:\jt\bio624\2011\Exam 2011\exam2011_mlr/boxrstudent.wmf written in Windows Metafile format)
```



.

1. Multiple Linear Regression (cont'd)

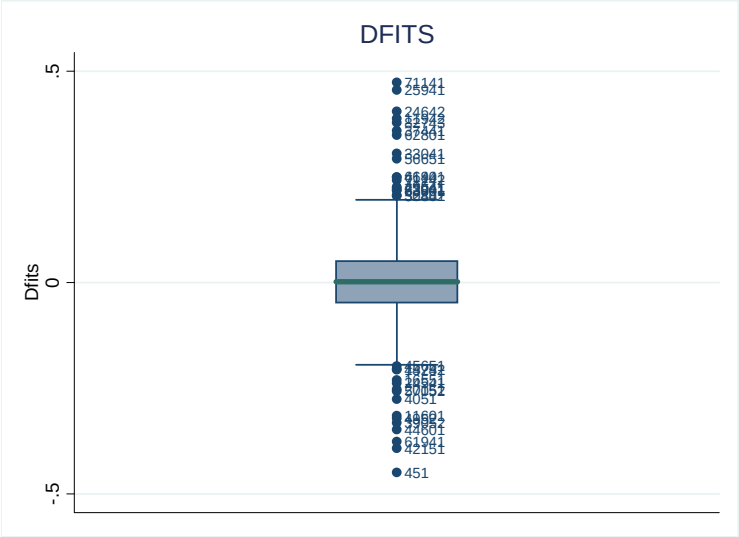
```
. * Boxplot leverages
.
. predict lev, leverage

. label variable lev "LEVERAGE"

.
. graph box lev,     marker(1,mlab(Id)) outergap(200) medtype(cline) medline(lwidth(1.0)) title("LEV
> ERAGE") graphregion(fcolor(white))


. graph export exam2011_mlr/boxleverage.wmf, replace
```




```
(file C:\jt\bio624\2011\Exam 2011\exam2011_mlr/boxleverage.wmf written in Windows Metafile format)

.
.
```

1. Multiple Linear Regression (cont'd)

```
. * Boxplot dfits
.
. predict dfits, dfits

.
. graph box dfits,    marker(1,mlab(Id)) outergap(200) medtype(cline) medline(lwidth(1.0)) title("DF
> ITS") graphregion(fcolor(white))

. graph export exam2011_mlr/boxdfits.wmf, replace
(file C:\jt\bio624\2011\Exam 2011\exam2011_mlr/boxdfits.wmf written in Windows Metafile format)
```



.
.

1. Multiple Linear Regression (cont'd)

```
. * Check linear scale for age
.
. mkspline preage 7 postage= Age, marginal

. regress fev Smoke Sex preage postage  Hgt

      Source |       SS       df       MS              Number of obs =     654
-------------+------------------------------           F(  5,   648) =  474.79
       Model |  385.651411     5  77.1302822           Prob > F      =  0.0000
    Residual |  105.268425   648  .162451274           R-squared     =  0.7856
-------------+------------------------------           Adj R-squared =  0.7839
       Total |  490.919836   653   .75179148           Root MSE      =  .40305

------------------------------------------------------------------------------
         fev |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Smoke |  -.1331518   .0585229    -2.28   0.023    -.2480692   -.0182345
         Sex |   .1438075    .032557     4.42   0.000     .0798776    .2077374
      preage |  -.1043106   .0319527    -3.26   0.001    -.1670538   -.0415673
     postage |   .1817805   .0327296     5.55   0.000     .1175116    .2460495
         Hgt |   .1109448   .0048079    23.08   0.000     .1015039    .1203857
       _cons |  -3.739915   .2532639   -14.77   0.000    -4.237231   -3.242598
------------------------------------------------------------------------------

.
```

2. Longitudinal Data Analysis

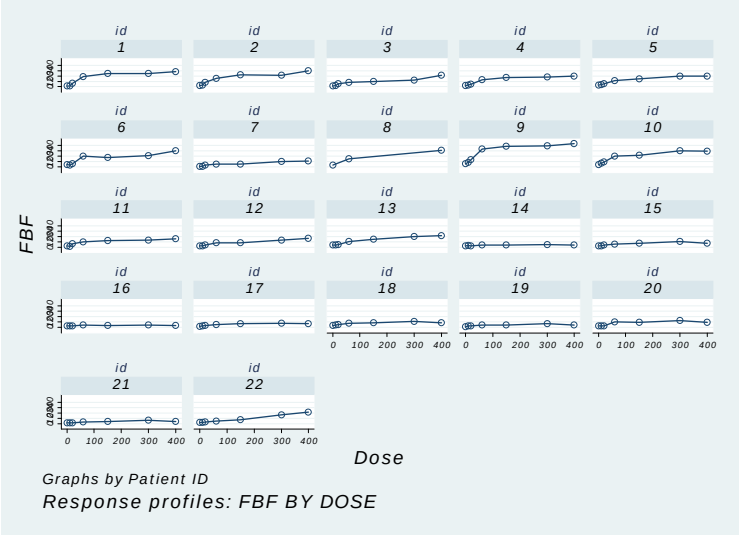
```
. * Part 0.  Data management   LONGITUDINAL DATA ANALYSIS
.
.
. * Input data
.
. use  lda_11.2.Long.Isoproterenol.dta, clear

.
.
.
.
. * Dictionary
.
. codebook

-------------------------------------------------------------------------------------------------------
id
Patient ID
-------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [1,22]                       units:  1
         unique values:  22                       missing .:  0/154

                  mean:      11.5
             std. dev:   6.36499

           percentiles:        10%       25%       50%       75%       90%
                                 3         6      11.5        17        20

-------------------------------------------------------------------------------------------------------
dose                                                                                 Isoproterenol Dose
(ng/min)
-------------------------------------------------------------------------------------------------------

                  type:  numeric (int)

                 range:  [0,400]                      units:  10
         unique values:  7                        missing .:  0/154

            tabulation:  Freq.  Value
                            22  0
                            22  10
                            22  20
                            22  60
                            22  150
                            22  300
                            22  400

-------------------------------------------------------------------------------------------------------
race
   Race
-------------------------------------------------------------------------------------------------------

                  type:  numeric (float)
                 label:  race

                 range:  [1,2]                        units:  1
         unique values:  2                        missing .:  0/154

            tabulation:  Freq.   Numeric  Label
                            91         1  White
                            63         2  Black
```

2. Longitudinal Data Analysis (cont'd)

```
-------------------------------------------------------------------------------------------------
fbf
(unlabeled)
-------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [1,43.3]                     units:  .01
         unique values:  122                      missing .:  4/154

                  mean:    9.5868
             std. dev:   8.92505

           percentiles:        10%       25%       50%       75%       90%
                                 2      3.05       5.8      13.2     21.65
-------------------------------------------------------------------------------------------------
baseline
(unlabeled)-------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [1,5.8]                      units:  .01
         unique values:  20                       missing .:  0/154

                  mean:   2.49636
             std. dev:   1.11815

           percentiles:        10%       25%       50%       75%       90%
                              1.24       1.8      2.18       3.1       3.9

-------------------------------------------------------------------------------------------------

-------------------------------------------------------------------------------------------------
logdose                                                                                    Log
Isoproterenol Dose
-------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [2.3025851,5.9914646]        units:  1.000e-07
         unique values:  6                        missing .:  22/154

            tabulation:  Freq.  Value
                           22  2.3025851
                           22  2.9957323
                           22  4.0943446
                           22  5.0106354
                           22  5.7037826
                           22  5.9914646
                           22  .

.
.
.
.
```

2. Longitudinal Data Analysis (cont'd)

```
. sort id dose

.
.
. * Plot individual profiles
.
. graph twoway (connected fbf dose, sort msymbol(circle_hollow) msize(large) clwidth(medthick) ),
by(id, capt
> ion(Response profiles: FBF BY DOSE))  ytitle(FBF) xtitle(Dose) title(id) subtitle(, size(vlarge))
.
. graph export exam2011_lda\individuals.emf,replace
```



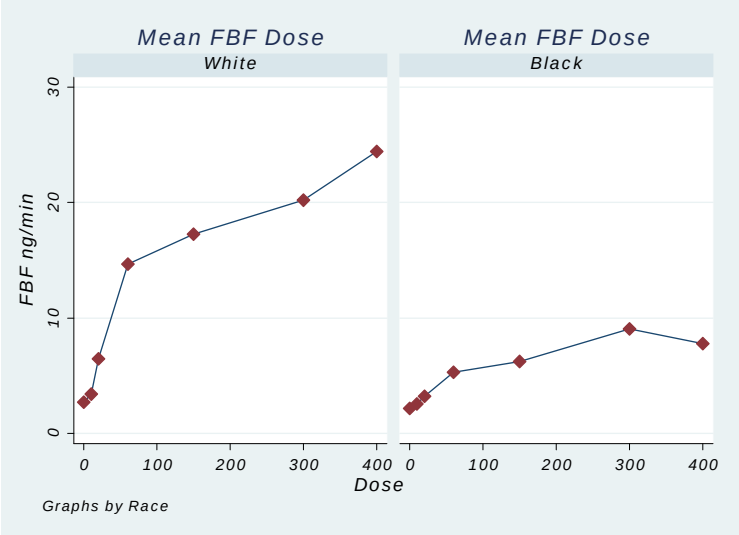
.
.
.

2. Longitudinal Data Analysis (cont'd)

```
. * Plot mean profiles
.
.
. * Save dataset  (dataset in memory destroyed by collapse)
.
. save exam2011_lda.dta, replace
file exam2011_lda.dta saved

.
.
. * Use collapse to create dataset of means
.
. sort id race dose

.
. collapse (mean) fbf,  by (race dose)
.
. list race dose fbf

     +----------------------+
     |  race   dose      fbf |
     |----------------------|
  1. | White      0   2.7146 |
  2. | White     10   3.4167 |
  3. | White     20   6.4583 |
  4. | White     60   14.658 |
  5. | White    150    17.25 |
     |----------------------|
  6. | White    300     20.2 |
  7. | White    400   24.408 |
  8. | Black      0   2.1811 |
  9. | Black     10   2.5778 |
 10. | Black     20   3.2111 |
     |----------------------|
 11. | Black     60   5.3022 |
 12. | Black    150   6.2344 |
 13. | Black    300   9.0567 |
 14. | Black    400   7.7678 |
     +----------------------+

.
```

2. Longitudinal Data Analysis (cont'd)

```
. graph twoway (connected fbf dose, connect(ascending) msymbol(diamond) msize(medlarge)
mcolor(maroon)), by(ra
> ce) ytitle("FBF ng/min") xtitle(Dose) title("Mean FBF Dose")

. graph export exam2011_lda\meanprofiles.emf,replace
(file C:\jt\bio624\2011\Exam 2011\exam2011_lda\meanprofiles.emf written in Enhanced Metafile format)
```

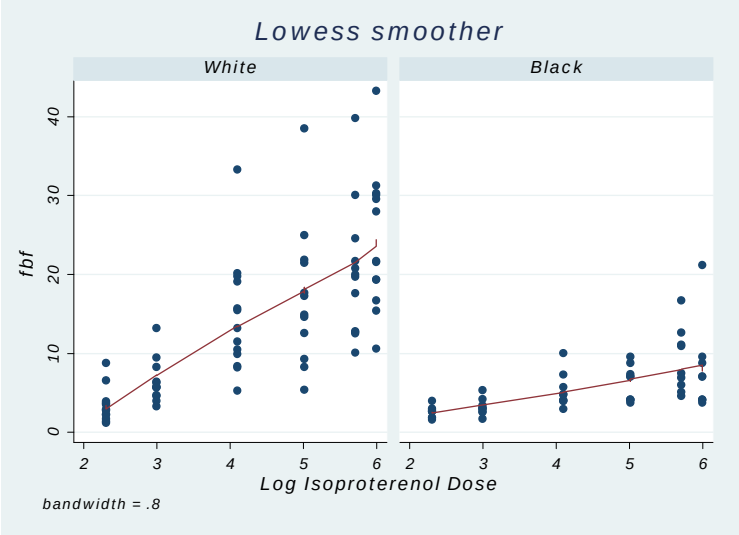



```
.
.
.
. * Restore full dataset
. use exam2011_lda, clear

.
.
.
```

2. Longitudinal Data Analysis (cont'd)

```
. lowess fbf dose, by(race)

. graph export exam2011_lda\lowess_dose.emf,replace
```


```
(file C:\jt\bio624\2011\Exam 2011\exam2011_lda\lowess_dose.emf written in Enhanced Metafile format)

.
.
.
. lowess fbf logdose, by(race)

. graph export exam2011_lda\lowess_logdose.emf,replace
(file C:\jt\bio624\2011\Exam 2011\exam2011_lda\lowess_logdose.emf written in Enhanced Metafile
format)
```




```
.
.
```

2. Longitudinal Data Analysis (cont'd)

```
. * GEE robust variance estimation
.
. regress fbf logdose race, cluster(id) robust

Linear regression                                     Number of obs =     128
                                                      F(  2,    21) =   26.16
                                                      Prob > F      =  0.0000
                                                      R-squared     =  0.5526
                                                      Root MSE      =  6.1436

                                    (Std. Err. adjusted for 22 clusters in id)
------------------------------------------------------------------------------
             |               Robust
         fbf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logdose |   3.796058    .548374     6.92   0.000     2.655652    4.936464
        race |  -8.774381   1.865356    -4.70   0.000     -12.6536    -4.89516
       _cons |   6.728496   3.170691     2.12   0.046      .134684    13.32231
------------------------------------------------------------------------------

.
.
. * Mixed Effects Model
.
. xtmixed fbf logdose race || id: , nolog

Mixed-effects REML regression                 Number of obs      =       128
Group variable: id                            Number of groups   =        22

                                              Obs per group: min =         2
                                                             avg =       5.8
                                                             max =         6


                                              Wald chi2(2)       =    206.17
Log restricted-likelihood = -384.61626        Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         fbf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logdose |   3.769244   .2747792    13.72   0.000     3.230686    4.307801
        race |  -9.007875   2.161515    -4.17   0.000    -13.24437   -4.771382
       _cons |    7.31212   3.455869     2.12   0.034     .5387418     14.0855
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Identity                 |
                   sd(_cons) |   4.663395   .8474464      3.266023    6.658634
-----------------------------+------------------------------------------------
                sd(Residual) |    4.20522   .2902626       3.67312    4.814402
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =    53.11 Prob >= chibar2 = 0.0000

.
.
.
```

2. Longitudinal Data Analysis (cont'd)

```
. * Race x log dose interaction
.
. xi: xtmixed fbf i.race*logdose || id: , nolog
i.race            _Irace_1-2          (naturally coded; _Irace_1 omitted)
i.race*logdose    _IracXlogdo_#       (coded as above)

Mixed-effects REML regression                   Number of obs      =       128
Group variable: id                              Number of groups   =        22

                                                Obs per group: min =         2
                                                               avg =       5.8
                                                               max =         6


                                                Wald chi2(3)       =    413.11
Log restricted-likelihood = -355.94821          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         fbf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    _Irace_2 |   7.094235   2.840116     2.50   0.012      1.52771    12.66076
    logdose |    5.33464   .2777858    19.20   0.000      4.79019     5.87909
_IracXlogd~2 |  -3.692266   .4261185    -8.66   0.000    -4.527443    -2.85709
       _cons |  -8.546494   1.839536    -4.65   0.000    -12.15192   -4.941071
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Identity                 |
                  sd(_cons) |   4.757129   .8145465      3.400922    6.654159
-----------------------------+------------------------------------------------
               sd(Residual) |   3.223097   .2234613      2.813575    3.692226
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =    85.96 Prob >= chibar2 = 0.0000

.
. describe _I*

              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------------------------------------
--
_Irace_2        byte   %8.0g                  race==2
_IracXlogdo_2   float  %9.0g                  (race==2)*logdose

.
. testparm  _IracX*

 ( 1)  [fbf]_IracXlogdo_2 = 0

           chi2(  1) =   75.08
         Prob > chi2 =    0.0000
```

3. Multiple Logistic Regression

```
. * Input data                        MULTIPLE LOGISTIC REGRESSION ANALYSIS
.
. use logistic_1.4.11.Sepsis.dta, clear

.
.
.
. * Data dictionary
.
. codebook

-------------------------------------------------------------------------------------------
id                                                                              Patient ID
-------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [1,455]                      units:  1
         unique values:  455                      missing .:  0/455

                  mean:       228
             std. dev:   131.491

           percentiles:        10%       25%       50%       75%       90%
                                46       114       228       342       410

-------------------------------------------------------------------------------------------
treat                                                                         Treatment
-------------------------------------------------------------------------------------------

                  type:  numeric (float)
                 label:  treatmnt

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/455

            tabulation:  Freq.   Numeric  Label
                           231         0  Placebo
                           224         1  Ibuprofen

-------------------------------------------------------------------------------------------
race                                                                                 Race
-------------------------------------------------------------------------------------------

                  type:  numeric (float)
                 label:  race

                 range:  [0,2]                        units:  1
         unique values:  3                        missing .:  0/455

            tabulation:  Freq.   Numeric  Label
                           293         0  White
                           130         1  Black
                            32         2  Other

-------------------------------------------------------------------------------------------
apache                                                          Baseline APACHE Score
-------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [0,41]                       units:  1
         unique values:  38                       missing .:  1/455

                  mean:   15.3304
             std. dev:   7.08579

           percentiles:        10%       25%       50%       75%       90%
                                 7        10        14        20        25

-------------------------------------------------------------------------------------------
o2del                                          Oxygen Delivery at Baseline (ml/min/m^2)
-------------------------------------------------------------------------------------------

                  type:  numeric (float)
```

3. Multiple Logistic Regression (cont'd)

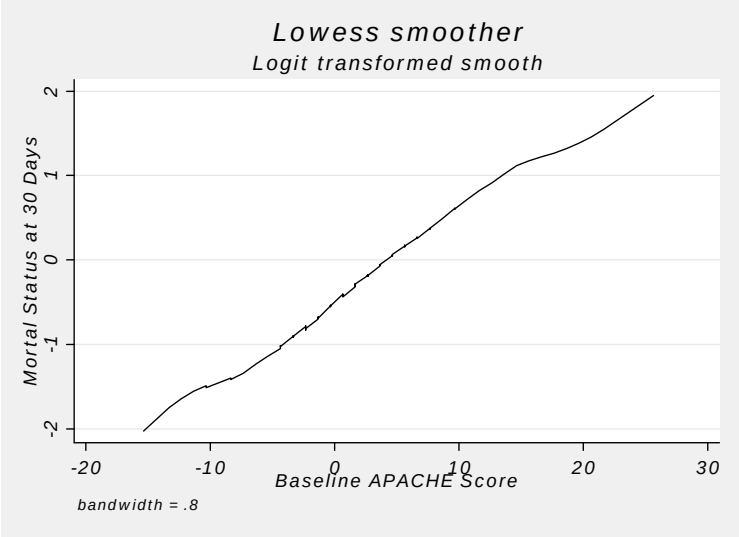
```
                range:  [316.88,2584.3401]          units:  .0001
        unique values:  168                      missing .:  287/455

                 mean:   1023.82
            std. dev:   409.443

          percentiles:        10%       25%       50%       75%       90%
                           567.73   765.195     947.2   1233.08   1525.51

-------------------------------------------------------------------------------
fate                                                      Mortal Status at 30 Days
-------------------------------------------------------------------------------

                 type:  numeric (float)
                label:  fate

                range:  [0,1]                       units:  1
        unique values:  2                        missing .:  0/455

           tabulation:  Freq.   Numeric  Label
                          279         0  Alive
                          176         1  Dead

-------------------------------------------------------------------------------
temp0                                                  Baseline Temperature (deg. F)
-------------------------------------------------------------------------------

                 type:  numeric (float)

                range:  [91.58,107]                 units:  .01
        unique values:  122                      missing .:  0/455

                 mean:   100.427
            std. dev:    2.0261

          percentiles:        10%       25%       50%       75%       90%
                               98      99.5     100.7     101.6    102.38

-------------------------------------------------------------------------------
o2use                                                   Oxygen Use at Baseline 1=yes
-------------------------------------------------------------------------------

                 type:  numeric (float)

                range:  [0,1]                       units:  1
        unique values:  2                        missing .:  0/455

           tabulation:  Freq.  Value
                          287  0
                          168  1

.
.
.
```

3. Multiple Logistic Regression (cont'd)

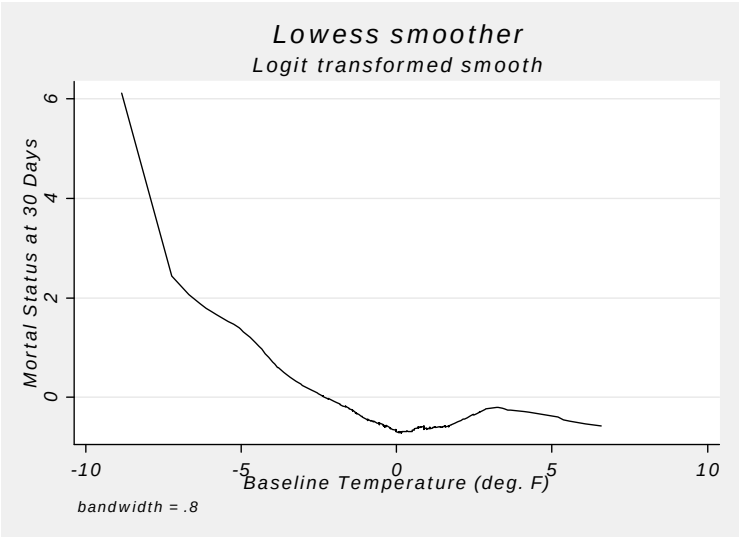
```
. * Center continuous variables - apache score and baseline temperature
.
. egen mean_apache = mean(apache)

. replace apache = apache - mean_apache
(454 real changes made)

.
. egen mean_temp0 = mean(temp0)

. replace temp0 = temp0 - mean_temp0
(455 real changes made)

.
. codebook apache temp0

-------------------------------------------------------------------------------
apache                                                  Baseline APACHE Score
-------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [-15.330397,25.669603]       units:  1.000e-08
         unique values:  38                       missing .:  1/455

                  mean:  -1.8e-07
             std. dev:   7.08579

           percentiles:        10%       25%       50%       75%       90%
                            -8.3304   -5.3304   -1.3304    4.6696    9.6696

-------------------------------------------------------------------------------
temp0                                                Baseline Temperature (deg. F)
-------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [-8.8468552,6.573143]        units:  1.000e-09
         unique values:  122                      missing .:  0/455

                  mean:   3.0e-07
             std. dev:    2.0261

           percentiles:        10%       25%       50%       75%       90%
                           -2.42686  -.926857    .27314   1.17314   1.95314

.
.
.
.
. * Combine Other race with Black
.
. replace race =1 if race == 2
(32 real changes made)
```

3. Multiple Logistic Regression (cont'd)

```
. * Lowess plots
.
. lowess fate apache, logit

. graph export exam2011_logistic\lowess_apache.emf,replace
(file C:\jt\bio624\2011\Exam 2011\exam2011_logistic\lowess_apache.emf written in Enhanced Metafile
format)
```




```
.
.
. lowess fate temp0, logit

. graph export exam2011_logistic\lowess_temp0.emf,replace
(file C:\jt\bio624\2011\Exam 2011\exam2011_logistic\lowess_temp0.emf written in Enhanced
>  Metafile format)
```




```
.
.
.
```

3. Multiple Logistic Regression (cont'd)

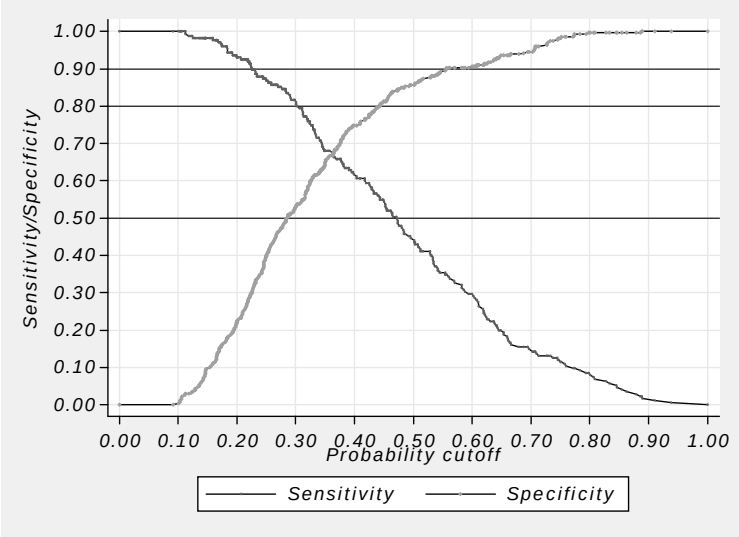
```
. * Collinearity
.
. quietly regress fate treat race apache o2use temp0

. vif

    Variable |       VIF       1/VIF
-------------+----------------------
        race |      1.10    0.912368
       o2use |      1.07    0.938214
      apache |      1.03    0.974095
       treat |      1.01    0.991360
       temp0 |      1.01    0.993284
-------------+----------------------
    Mean VIF |      1.04

.
.
.
. * Fit logistic regression
.
. logistic fate treat race apache o2use temp0

Logistic regression                               Number of obs   =        454
                                                  LR chi2(5)      =      72.67
                                                  Prob > chi2     =     0.0000
Log likelihood = -266.33456                       Pseudo R2       =     0.1201

------------------------------------------------------------------------------
        fate | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |   .8305325   .1754578    -0.88   0.379     .5489493    1.256553
        race |   1.203958   .2745228     0.81   0.416     .7700569    1.882347
      apache |   1.122294   .0183239     7.07   0.000     1.086948    1.158789
       o2use |   1.104204   .2481899     0.44   0.659     .7107653    1.715427
       temp0 |   .8561688   .0450142    -2.95   0.003     .7723361     .949101
------------------------------------------------------------------------------

. logit

Logistic regression                               Number of obs   =        454
                                                  LR chi2(5)      =      72.67
                                                  Prob > chi2     =     0.0000
Log likelihood = -266.33456                       Pseudo R2       =     0.1201

------------------------------------------------------------------------------
        fate |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |  -.1856883   .2112594    -0.88   0.379    -.5997492    .2283726
        race |   .1856143    .228017     0.81   0.416    -.2612909    .6325194
      apache |   .1153746   .0163272     7.07   0.000     .0833738    .1473753
       o2use |   .0991245   .2247682     0.44   0.659     -.341413     .539662
       temp0 |  -.1552877   .0525763    -2.95   0.003    -.2583354     -.05224
       _cons |  -.5320871   .1905595    -2.79   0.005    -.9055768   -.1585973
------------------------------------------------------------------------------

.
. estat ic

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |    454   -302.6705   -266.3346      6     544.6691    569.3777
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note

.
.
.
. * Check fit
.
. * Option number numbers unique covariate patterns
. *   -- observations with the same covariate pattern have the same
. *        number.
.
. predict unique_patterns, number
(1 missing value generated)
```

3. Multiple Logistic Regression (cont'd)

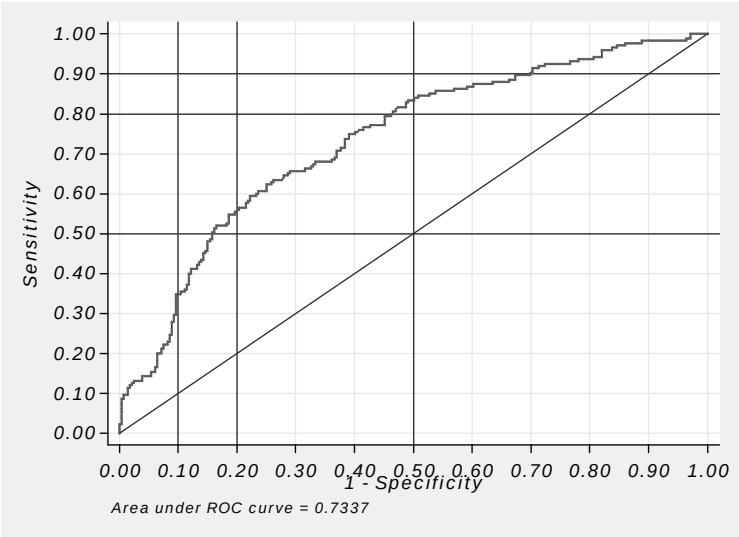
```
. codebook unique_patterns

-------------------------------------------------------------------------------------------------
unique_patterns                                                                 covariate pattern
-------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [1,447]                      units:  1
         unique values:  447                      missing .:  1/455

                  mean:   223.815
             std. dev:   130.036

           percentiles:        10%       25%       50%       75%       90%
                                45       110     223.5       337       404

.
. estat gof

Logistic model for fate, goodness-of-fit test

       number of observations =       454
 number of covariate patterns =       447
            Pearson chi2(441) =       446.49
                  Prob > chi2 =         0.4182

.
. estat gof, group(10)

Logistic model for fate, goodness-of-fit test

  (Table collapsed on quantiles of estimated probabilities)

       number of observations =       454
             number of groups =        10
     Hosmer-Lemeshow chi2(8) =        13.44
                  Prob > chi2 =         0.0976

.
.
.
. * Predicted probabilities for ROC analysis
.
. predict prob_full
(option pr assumed; Pr(fate))
(1 missing values generated)
.
```

3. Multiple Logistic Regression (cont'd)

```
. lsens, clwidth(medium medium) msize(tiny tiny)    ///
>     ylabel(0(.1)1, angle(horizontal)) yline(.5 .8 .9) xlabel(0(.1)1)

. graph export exam2011_logistic\lens.emf,replace
(file C:\jt\bio624\2011\Exam 2011\exam2011_logistic\lens.emf written in Enhanced Metafil
> e format)
```




```
.
.
. lroc,  clwidth(medium medium) msize(tiny tiny)    ///
>     ylabel(0(.1)1, angle(horizontal)) yline(.5 .8 .9) xlabel(0(.1)1) xline(.1 .2 .5)

Logistic model for fate

number of observations =       454
area under ROC curve   =    0.7337

. graph export exam2011_logistic\roc.emf,replace
(file C:\jt\bio624\2011\Exam 2011\exam2011_logistic\roc.emf written in Enhanced Metafile
>  format)
```



3. Multiple Logistic Regression (cont'd)

```
. * Automatic selection of variables
.
. sw, lockterm1 pe(0.05): logistic fate (treat apache) race o2use temp0
                      begin with term 1 model
p = 0.0026 <  0.0500  adding  temp0

Logistic regression                               Number of obs   =        454
                                                  LR chi2(3)      =      71.96
                                                  Prob > chi2     =     0.0000
Log likelihood = -266.69259                       Pseudo R2       =     0.1189

------------------------------------------------------------------------------
        fate | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |   .8436549   .1773148    -0.81   0.419     .5588097    1.273696
      apache |   1.123954   .0182049     7.21   0.000     1.088834    1.160208
       temp0 |   .8541421   .0447455    -3.01   0.003     .7707946     .946502
------------------------------------------------------------------------------

. estat ic

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |    454   -302.6705   -266.6926      4     541.3852    557.8576
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note

.
. predict prob_forw
(option pr assumed; Pr(fate))
(1 missing values generated)

.
.
.
. sw, lockterm1 pr(0.05): logistic fate (treat apache) race o2use temp0
                      begin with full model
p = 0.6592 >= 0.0500  removing o2use
p = 0.4693 >= 0.0500  removing race

Logistic regression                               Number of obs   =        454
                                                  LR chi2(3)      =      71.96
                                                  Prob > chi2     =     0.0000
Log likelihood = -266.69259                       Pseudo R2       =     0.1189

------------------------------------------------------------------------------
        fate | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |   .8436549   .1773148    -0.81   0.419     .5588097    1.273696
      apache |   1.123954   .0182049     7.21   0.000     1.088834    1.160208
       temp0 |   .8541421   .0447455    -3.01   0.003     .7707946     .946502
------------------------------------------------------------------------------

. estat ic

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |    454   -302.6705   -266.6926      4     541.3852    557.8576
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note

.
. predict prob_back
(option pr assumed; Pr(fate))
(1 missing values generated)

.
.
.
. roccomp fate prob_full prob_forw prob_back

                            ROC                    -Asymptotic Normal--
                   Obs       Area     Std. Err.      [95% Conf. Interval]
-------------------------------------------------------------------------
prob_full          454     0.7337       0.0241        0.68642     0.78101
```

3. Multiple Logistic Regression (cont'd)

```
prob_forw          454     0.7308       0.0243        0.68325     0.77838
prob_back          454     0.7308       0.0243        0.68325     0.77838
------------------------------------------------------------------------------
Ho: area(prob_full) = area(prob_forw) = area(prob_back)
    chi2(1) =     0.91       Prob>chi2 =   0.3401

.
.
.
. * Effect modification
.
. xi: logistic fate i.treat*race apache o2use temp0
i.treat           _Itreat_0-1         (naturally coded; _Itreat_0 omitted)
i.treat*race      _ItreXrace_#        (coded as above)

Logistic regression                               Number of obs   =        454
                                                  LR chi2(6)      =      74.04
                                                  Prob > chi2     =     0.0000
Log likelihood = -265.64808                       Pseudo R2       =     0.1223

------------------------------------------------------------------------------
        fate | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   _Itreat_1 |   1.000304   .2639423     0.00   0.999     .5963926    1.677768
        race |   1.556795   .4923647     1.40   0.162     .8375742    2.893606
_ItreXrace_1 |   .5975301   .2628551    -1.17   0.242     .2522981     1.41516
      apache |   1.121895    .018359     7.03   0.000     1.086483    1.158461
       o2use |   1.087076   .2452638     0.37   0.711     .6985757    1.691633
       temp0 |   .8542466   .0449025    -3.00   0.003      .770621    .9469471
------------------------------------------------------------------------------


.
. xi: logistic fate i.treat*apache race o2use temp0
i.treat           _Itreat_0-1         (naturally coded; _Itreat_0 omitted)
i.treat*apache    _ItreXapach_#       (coded as above)

Logistic regression                               Number of obs   =        454
                                                  LR chi2(6)      =      74.42
                                                  Prob > chi2     =     0.0000
Log likelihood = -265.46171                       Pseudo R2       =     0.1229

------------------------------------------------------------------------------
        fate | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   _Itreat_1 |   .8051923   .1718636    -1.02   0.310     .5299264    1.223443
      apache |   1.098477   .0246391     4.19   0.000     1.051231    1.147846
_ItreXapac~1 |   1.044003   .0341096     1.32   0.187      .979245    1.113044
        race |   1.223925   .2797887     0.88   0.377     .7819346    1.915751
       o2use |   1.111336   .2499877     0.47   0.639     .7151107      1.7271
       temp0 |   .8522303     .04508    -3.02   0.003      .768301    .945328
------------------------------------------------------------------------------
```

4. Ordered Logistic Regression

```
. * Part 0.  Data management   ORDERED LOGISTIC ANAD MULTINOMIAL LOGISTIC REGRESSION
.
.
. * Input data
.
. use http://www.stata-press.com/data/r11/fullauto, clear
(Automobile Models)

.
.
.
.
. * Dictionary
.
. codebook rep77 foreign weight gratio
-------------------------------------------------------------------------------------------------------
--
rep77                                                                                          Repair
Record 1977
-------------------------------------------------------------------------------------------------------
--

                  type:  numeric (int)
                 label:  repair

                 range:  [1,5]                        units:  1
         unique values:  5                        missing .:  8/74

            tabulation:  Freq.   Numeric  Label
                            3         1  Poor
                           11         2  Fair
                           27         3  Average
                           20         4  Good
                            5         5  Excellent
                            8         .
-------------------------------------------------------------------------------------------------------
--
foreign
Foreign
-------------------------------------------------------------------------------------------------------
--

                  type:  numeric (int)
                 label:  foreign

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/74

            tabulation:  Freq.   Numeric  Label
                           52         0  Domestic
                           22         1  Foreign

-------------------------------------------------------------------------------------------------------
--
weight
Weight (lbs.)
-------------------------------------------------------------------------------------------------------
--

                  type:  numeric (int)

                 range:  [1760,4840]                  units:  10
         unique values:  64                       missing .:  0/74

                  mean:   3019.46
             std. dev:   777.194

           percentiles:        10%       25%       50%       75%       90%
                              2020      2240      3190      3600      4060
```

4. Ordered Logistic Regression (cont'd)

```
-------------------------------------------------------------------------------------------------------------
--
gratio
Gear Ratio
-------------------------------------------------------------------------------------------------------------
--

                  type:  numeric (float)

                 range:  [2.19,3.89]                 units:  .01
         unique values:  36                       missing .:  0/74

                  mean:   3.01486
             std. dev:   .456287

           percentiles:        10%       25%       50%       75%       90%
                              2.43      2.73     2.955      3.37      3.72

.
.
. * Combine Poor/Fair and Good/Excellent
.
. replace rep77 = 2 if rep77==1
(3 real changes made)

.
. replace rep77 = 4 if rep77==5
(5 real changes made)

.
. codebook rep77

-------------------------------------------------------------------------------------------------------------
--
rep77                                                                                          Repair
Record 1977
-------------------------------------------------------------------------------------------------------------
--

                  type:  numeric (int)
                 label:  repair

                 range:  [2,4]                       units:  1
         unique values:  3                        missing .:  8/74

            tabulation:  Freq.   Numeric  Label
                            14         2  Fair
                            27         3  Average
                            25         4  Good
                             8         .
.
. * Tabulate repair rating by manufacturer (foreign vs. domestic)
.
. tab rep77 foreign, exact

Enumerating sample-space combinations:
stage 3:  enumerations = 1
stage 2:  enumerations = 7
stage 1:  enumerations = 0

    Repair |
    Record |         Foreign
      1977 |  Domestic    Foreign |     Total
-----------+----------------------+----------
      Fair |        12          2 |        14
   Average |        20          7 |        27
      Good |        13         12 |        25
-----------+----------------------+----------
     Total |        45         21 |        66

           Fisher's exact =                 0.084

.
.
.
```

4. Ordered Logistic Regression (cont'd)

```
. * Fit ordered logistic
.
. ologit rep77 foreign weight gratio, or nolog

Ordered logistic regression                     Number of obs   =         66
                                                LR chi2(3)      =       7.41
                                                Prob > chi2     =     0.0600
Log likelihood =  -66.40787                     Pseudo R2       =     0.0528

------------------------------------------------------------------------------
       rep77 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     foreign |   5.234107   4.387806     1.97   0.048     1.012212    27.06535
      weight |   1.000709    .000545     1.30   0.193      .999641    1.001777
      gratio |   1.679785   1.670048     0.52   0.602      .239324     11.7902
-------------+----------------------------------------------------------------
       /cut1 |   2.813564   4.144583                     -5.30967     10.9368
       /cut2 |   4.787285   4.179856                    -3.405081    12.97965
------------------------------------------------------------------------------

.
.
.
. test weight gratio

 ( 1)  [rep77]weight = 0
 ( 2)  [rep77]gratio = 0

           chi2(  2) =    1.80
         Prob > chi2 =    0.4067

.
.
. * Brant test
.
. brant

Brant Test of Parallel Regression Assumption

    Variable |      chi2   p>chi2    df
-------------+--------------------------
         All |      1.93    0.586     3
-------------+--------------------------
     foreign |      0.91    0.341     1
      weight |      0.01    0.935     1
      gratio |      1.12    0.290     1
----------------------------------------

A significant test statistic provides evidence that the parallel
regression assumption has been violated.
```

5. Multinomial Logistic Regression

```
. * Fit multinomial logistic
.
. mlogit rep77 foreign weight gratio, rrr baseoutcome(2) nolog

Multinomial logistic regression                 Number of obs   =         66
                                                LR chi2(6)      =       9.32
                                                Prob > chi2     =     0.1566
Log likelihood = -65.453295                     Pseudo R2       =     0.0664

------------------------------------------------------------------------------
       rep77 |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Fair         |  (base outcome)
-------------+----------------------------------------------------------------
Average      |
     foreign |   1.019474   1.239134     0.02   0.987     .0941395    11.04029
      weight |   1.000467   .0006906     0.68   0.499      .999114    1.001821
      gratio |   5.624233   7.804793     1.24   0.213     .3705478    85.36551
-------------+----------------------------------------------------------------
Good         |
     foreign |   8.062134   10.24878     1.64   0.101     .6673938    97.39077
      weight |   1.001066   .0007833     1.36   0.173     .9995322    1.002603
      gratio |   3.499149   5.108267     0.86   0.391      .200137    61.17831
------------------------------------------------------------------------------

.
.
.
. test foreign

 ( 1)  [Fair]o.foreign = 0
 ( 2)  [Average]foreign = 0
 ( 3)  [Good]foreign = 0
       Constraint 1 dropped

           chi2(  2) =    4.77
         Prob > chi2 =    0.0920

.
. test weight

 ( 1)  [Fair]o.weight = 0
 ( 2)  [Average]weight = 0
 ( 3)  [Good]weight = 0
       Constraint 1 dropped

           chi2(  2) =    1.87
         Prob > chi2 =    0.3923

.
. test gratio

 ( 1)  [Fair]o.gratio = 0
 ( 2)  [Average]gratio = 0
 ( 3)  [Good]gratio = 0
       Constraint 1 dropped

           chi2(  2) =    1.55
         Prob > chi2 =    0.4603

.
```

5. Multinomial Logistic Regression (cont'd)

```
. test weight gratio

 ( 1)  [Fair]o.weight = 0
 ( 2)  [Average]weight = 0
 ( 3)  [Good]weight = 0
 ( 4)  [Fair]o.gratio = 0
 ( 5)  [Average]gratio = 0
 ( 6)  [Good]gratio = 0
       Constraint 1 dropped
       Constraint 4 dropped

           chi2(  4) =    3.54
         Prob > chi2 =    0.4715
```