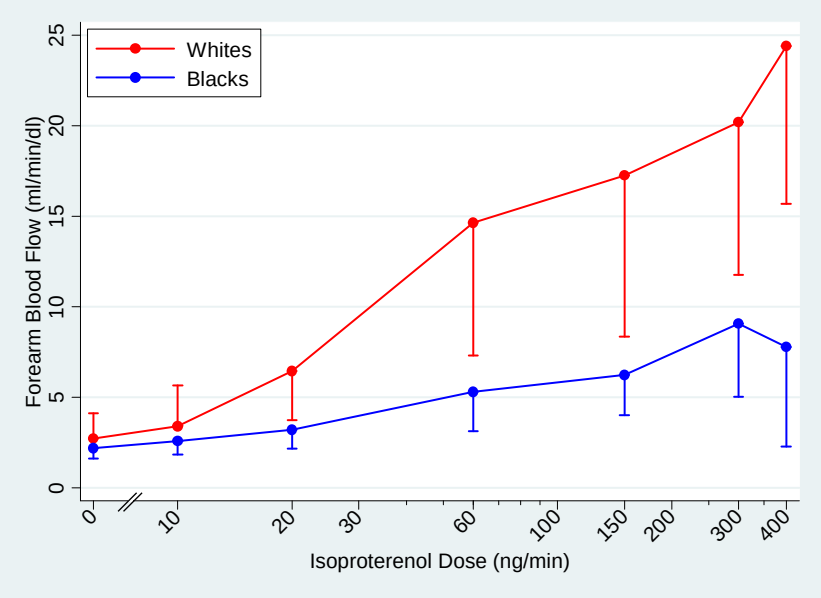
X. Mixed Effects Analysis of Variance

- Analysis of variance with multiple observations per patient
  - These analyses are complicated by the fact that multiple observations on the same patient are correlated with each other
- Response-feature approach to mixed effects analysis of variance
  - Reduce multiple response measures on each patient to a single statistic that captures the most biologically important aspect of the response
  - Perform a fixed effects analysis on this response feature
  - Using a regression slope as a response feature
  - Using an area under the curve as a response feature
- Generalized estimating equations (GEE) approach to mixed effects analysis of variance
  - GEE analysis with logistic or Poisson models

Lang et al. (1995) studied the effect of isoproterenol, a β-adrenergic agonist, on forearm blood flow in a group of 22 normotensive men. Nine of the study subjects were black and 13 were white. Each subject's blood flow was measured at baseline and then at escalating doses of isoproterenol.



The vertical bars indicate estimated standard deviations within each race-dose group.

There are a number of difficulties with analyzing these data.

1. Responses from the same patient are likely to be correlated. If Mr. Smith's response is 2 standard deviations above the mean day-2-treatment response on day 2, it is unlikely that he will be below the mean day-3-treatment response on day 3.

2. There is likely to be inherent variability between patients in how they respond to therapy that must be accounted for in our analysis.

3. We observe 22×7 = 154 responses. However, these observations only come from 22 patients. If we wish to make inferences about patients in general, our effective sample size is 22 rather than 154.

A common error in analyzing data like these is to use a fixed effects model. For example, a model such as

regress response race##dose

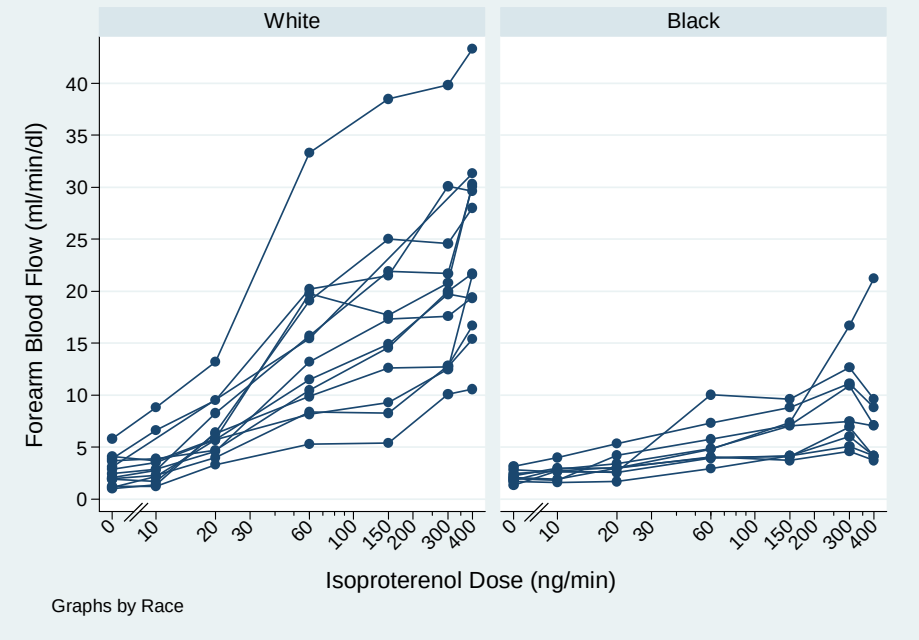
assumes that each patient's response equals
- a constant +
- an effect due to race +
- an effects due to dose +
- dose-race interaction effects +
- an independent error term.

The analysis is exactly the same as if we had had 154 distinct patients with each patient observed at a single-dose. This analysis will have 140 degrees of freedom and will seriously overestimate the significance of the dose-treatment effect.

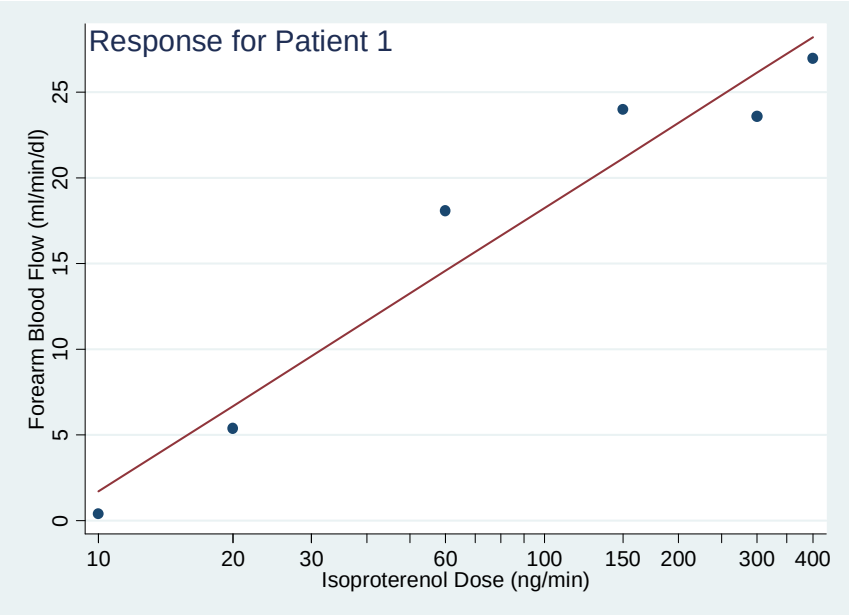
1. The Response-Feature Approach to Mixed Effects Analysis of Variance

The simplest valid way of analyzing mixed effects data is to compress each patient's response values into a single biologically sensible measure and then do an appropriate fixed effects analysis of the condensed response.

Consider the Isoproterenol-race data.

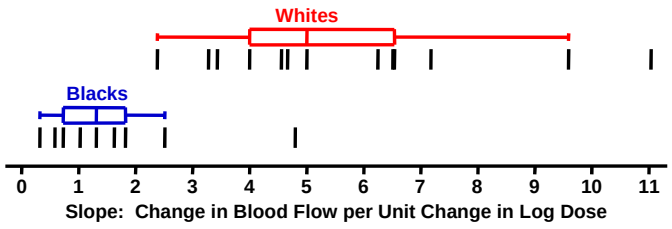


Note that there appears to be a log-linear relationship between dose and blood flow.



This suggests

1. Fit separate linear regression lines to the responses for each patient. This will give us 22 slope estimates - one for each patient.
2. Perform a Wilcoxon rank-sum test on these slopes to determine whether the slopes of black and white patients are different. It is prudent to use a non-parametric test because the individual patient slopes may have a non-normal distribution. However, you could also test these slopes with a *t*-test.



The Wilcoxon-Mann-Whiney rank sum test is significant with P=.0006.

Note that the responses between patients really are independent, so this analysis does not make any silly assumptions.

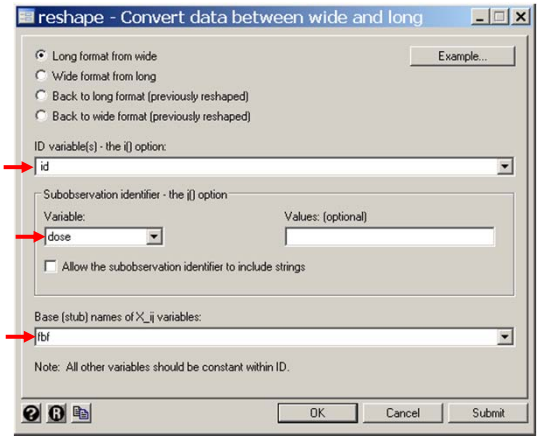
The same idea can be used in many other ways. The key idea is to compress the response data in a way that is biologically sensible. This may involve area under the curve, an average, or a weighted average.

2. Response Feature Analysis Using Stata

```
Exploratory Analysis of Repeated Measures Data Using Stata

. *  11.2.Isoproterenol.log  See Text p.364
. *
. *  Plot mean forearm blood flow by race and log dose of isoproterenol
. *  using the data of Lang et al. (1995).  Show standard deviation for
. *  each race at each drug level.
. *
. use C:\WDDtext\11.2.Isoproterenol.dta, clear
. * Statistics > Summaries... > Tables > Table of summary statistics (table).
. table race, row
 ----------------------
     Race |      Freq.
----------+-----------
    White |         13
    Black |          9
          |
    Total |         22
----------------------
. * Data > Describe data > List data
. list if id == 1| id == 22

     +-------------------------------------------------------------------------+
     | id    race   fbf0   fbf10   fbf20   fbf60   fbf150   fbf300   fbf400 |
     |-------------------------------------------------------------------------|
  1. |  1   White      1     1.4     6.4    19.1       25     24.6       28 |
 22. | 22   Black    2.1     1.9       3     4.8      7.4     16.7     21.2 |
     +-------------------------------------------------------------------------+
```

```
. generate baseline = fbf0

. *
. * Convert data from one record per patient to one record per observation.
. *
. * Data > Create > Other variable-trans... > Convert data between wide...
. reshape long fbf, i(id) j(dose)                                          {1}
(note: j = 0 10 20 60 150 300 400)

Data                               wide   ->   long
-----------------------------------------------------------------------------
Number of obs.                       22   ->     154
Number of variables                  10   ->       5
j variable (7 values)                     ->   dose
xij variables:
                  fbf0 fbf10 ... fbf400   ->   fbf
-----------------------------------------------------------------------------
```

{1} The ***reshape long*** command converts data from one record per **patient** to one record per **observation**. In this command, *i*(*id*) specifies that the *id* variable identifies observations from the same subject. The variable ***fbf*** is the first three letters of variables ***fbf0***, ***fbf10***, …, ***fbf400***; ***j*(*dose*)** defines ***dose*** to be a new variable whose values are the **trailing digits** in the names of the variables ***fbf0***, ***fbf10***, …, ***fbf400***. That is, *dose* will take the values 0, 10, 20, …, 300, 400. One record will be created for each value of *fbf0*, *fbf10*, …,*fbf400*. **Other variables** in the file that are not included in this command (like *race* or *baseline*) are assumed not to vary with *dose* and are replicated in each record for each specific patient.

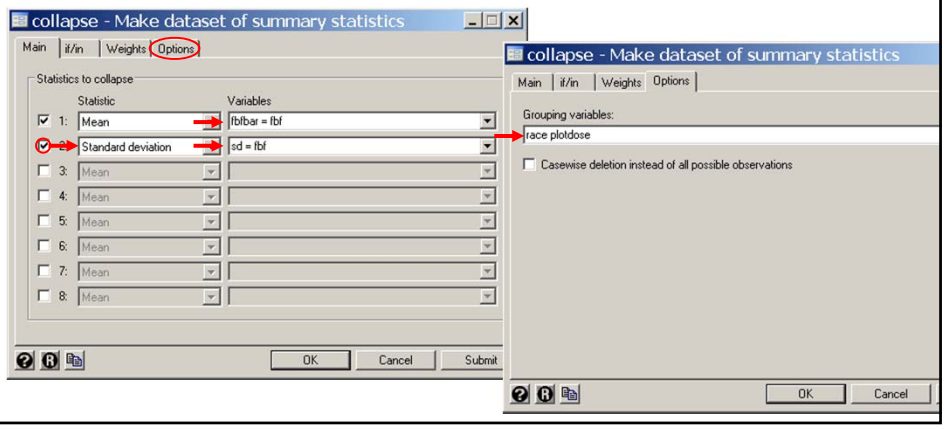



```
. * Data > Describe data > List data
. list if id == 1 | id == 22

     +-------------------------------------+
     | id   dose    race    fbf   baseline |
     |-------------------------------------|
  1. |  1      0   White      1          1 |
  2. |  1     10   White    1.4          1 |
  3. |  1     20   White    6.4          1 |
  4. |  1     60   White   19.1          1 |
  5. |  1    150   White     25          1 |
     |-------------------------------------|
  6. |  1    300   White   24.6          1 |
  7. |  1    400   White     28          1 |
148. | 22      0   Black    2.1        2.1 |
149. | 22     10   Black    1.9        2.1 |
150. | 22     20   Black      3        2.1 |
     |-------------------------------------|
151. | 22     60   Black    4.8        2.1 |
152. | 22    150   Black    7.4        2.1 |
153. | 22    300   Black   16.7        2.1 |
154. | 22    400   Black   21.2        2.1 |
     +-------------------------------------+
```

```
. generate delta_fbf = fbf - baseline
(4 missing values generated)

. label variable delta_fbf "Change in Forearm Blood Flow"

. label variable dose "Isoproterenol Dose (ng/min)"

. generate plotdose = dose

. replace plotdose = 6    if dose == 0                                {2}
(22 real changes made)

. label variable plotdose "Isoproterenol Dose (ng/min)"

. generate logdose = log(dose)
(22 missing values generated)

. label variable logdose "Log Isoproterenol Dose"
```

{2} We want to create Figures 10.1 and 10.2 that plot dose on a logarithmic scale. We also want to include the baseline dose of zero on these figures. Since the logarithm of zero is undefined, we create a new variable called *plotdose* that equals *dose* for all values greater than zero and equals 6 when *dose* = 0. We will use a graphics editor to relabel this value zero with a break in the *x*-axis when we create these figures.

```
. *
. *  Save long format of data for subsequent analyses
. *
. save C:\WDDtext\11.2.Long.Isoproterenol.dta, replace
file C:\WDDtext\11.2.Long.Isoproterenol.dta saved

. *
. *  Generate Figure 11.1
. *
. * Data > Create... > Other variable-trans... > Make dataset of means...
. collapse (mean) fbfbar = fbf (sd) sd = fbf, by(race plotdose)
```



```
. generate blackfbf = .
(14 missing values generated)

. generate whitefbf = .
(14 missing values generated)

. generate whitesd = .
(14 missing values generated)

. generate blacksd = .
(14 missing values generated)

. replace whitefbf = fbfbar if race == 1                                {3}
(7 real changes made)

. replace blackfbf = fbfbar if race == 2
(7 real changes made)
```

{3} The variable ***whitefbf*** equals the mean forearm blood flow for **white** subjects and is missing for black subjects; ***blackfbf*** is similarly defined for **black** subjects. The variables ***blacksd*** and ***whitesd*** give the standard deviations for black and white subjects, respectively.

```
. replace blacksd = sd if race == 2
(7 real changes made)

. replace whitesd = sd if race == 1
(7 real changes made)

. label variable whitefbf "Forearm Blood Flow (ml/min/dl)"

. label variable blackfbf "Forearm Blood Flow (ml/min/dl)"

. generate wsdbar = whitefbf - whitesd                                  {4}
(7 missing values generated)

. generate bsdbar = blackfbf - blacksd
(7 missing values generated)

. replace wsdbar = whitefbf + whitesd if plotdose < 20                  {5}
(2 real changes made)

. twoway connected whitefbf plotdose, color(red)                      /// {6}
>     || rcap whitefbf wsdbar plotdose, color(red)                    ///
>     || connected blackfbf plotdose, color(blue)                     ///
>     || rcap blackfbf bsdbar plotdose, color(blue)                   ///
>     ||, ytitle(Forearm Blood Flow (ml/min/dl))                       ///
>         legend(ring(0) position(11) col(1) order(1 "Whites" 3 "Blacks")) /// {7}
>         xtitle(Isoproterenol Dose (ng/min)) xscale(log)              /// {8}
>         xlabel(6 "0" 10 20 30 60 100 150 200 300 400, angle(45))     /// {9}
>         xmtick(40(10)90 250 350)
```

{4} The **distance** between ***whitefbf*** and ***wsdbar*** equals the standard deviation of the forearm blood flow for **white** subjects at each dose; ***bsdbar*** is similarly defined for black patients.

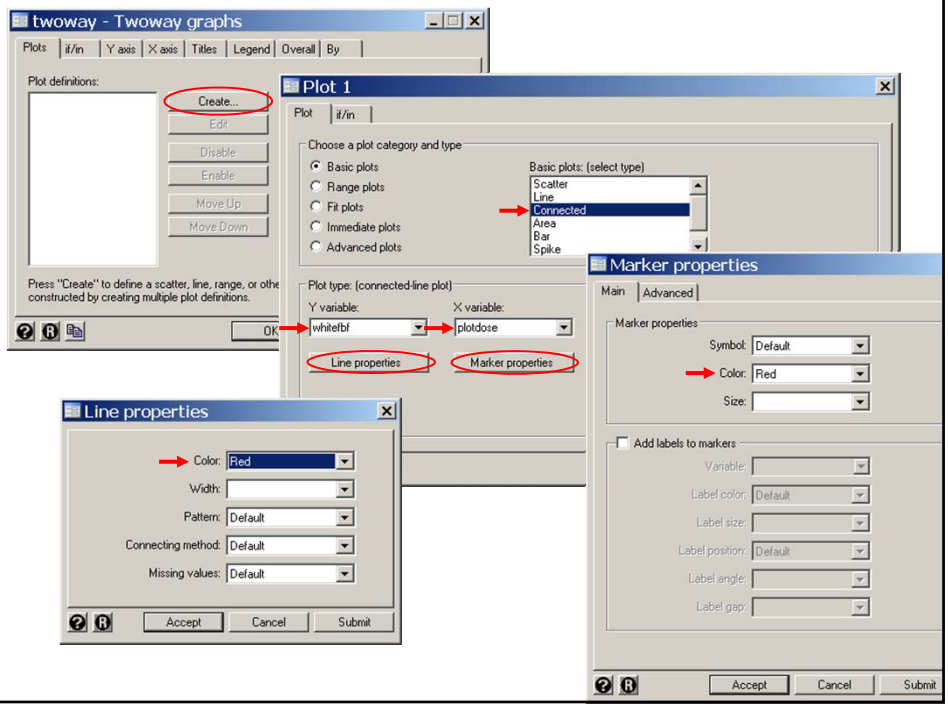
{5} This command will result in drawing the first two error bars for whites above the line to avoid collisions between the standard deviation bars for the two races.

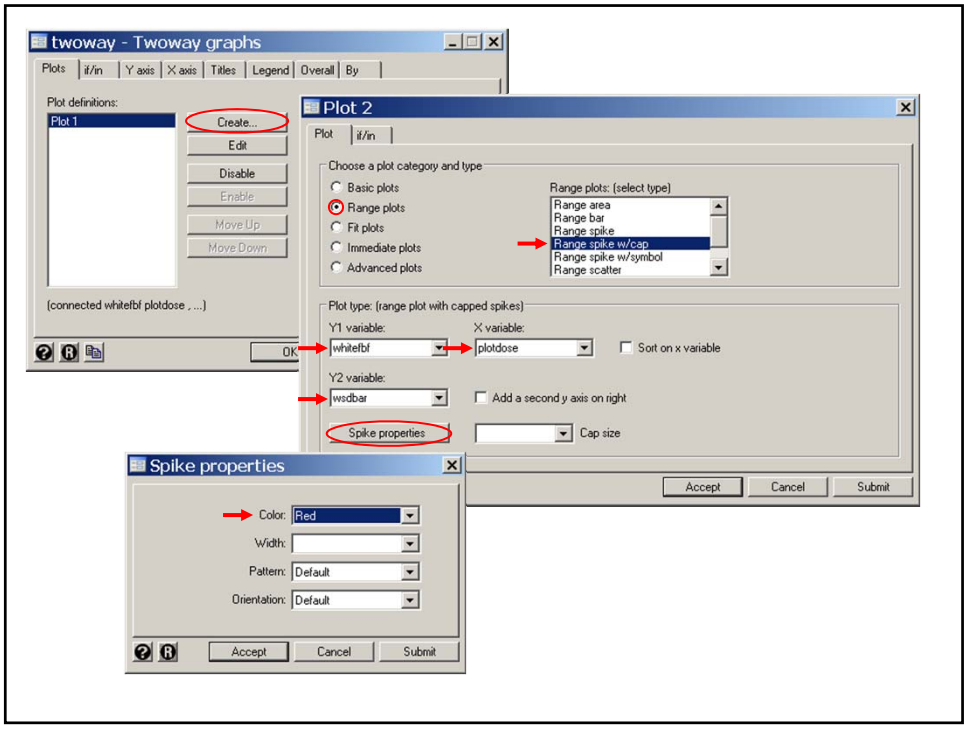
{6} This ***twoway connected*** command draws a scatter-plot of *whitefbf* by *plotdose* and connects the observations with straight lines.

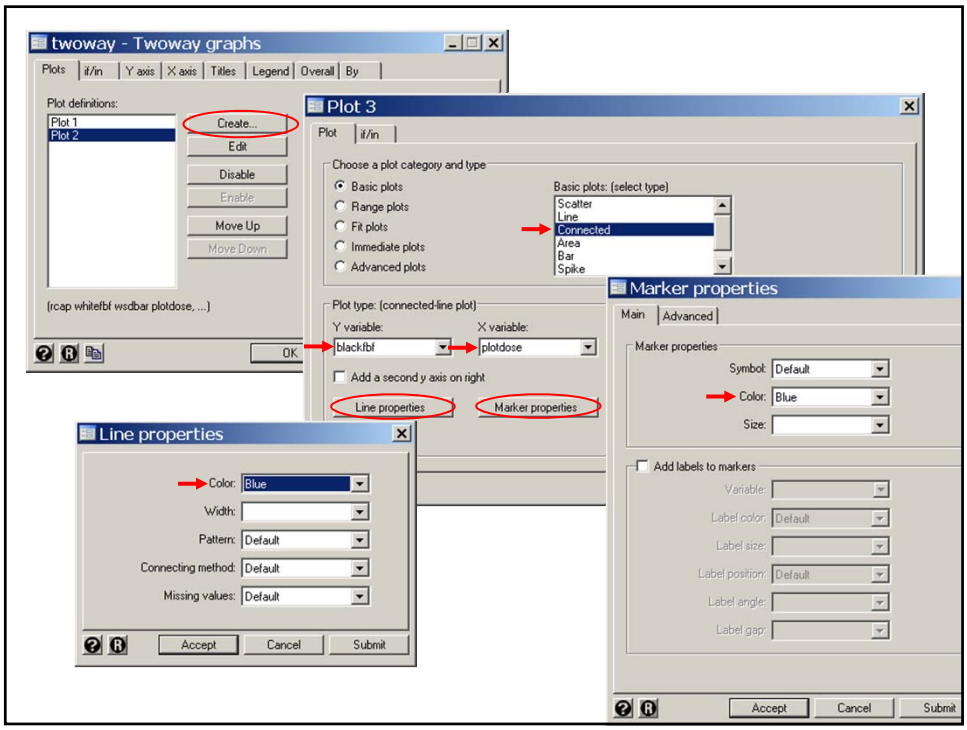
{7} The first and third variables (keys) of this plot are the mean blood flows for whites and blacks, respectively. This order option restricts the legend to these two variables and labels them "Whites" and "Blacks" respectively.

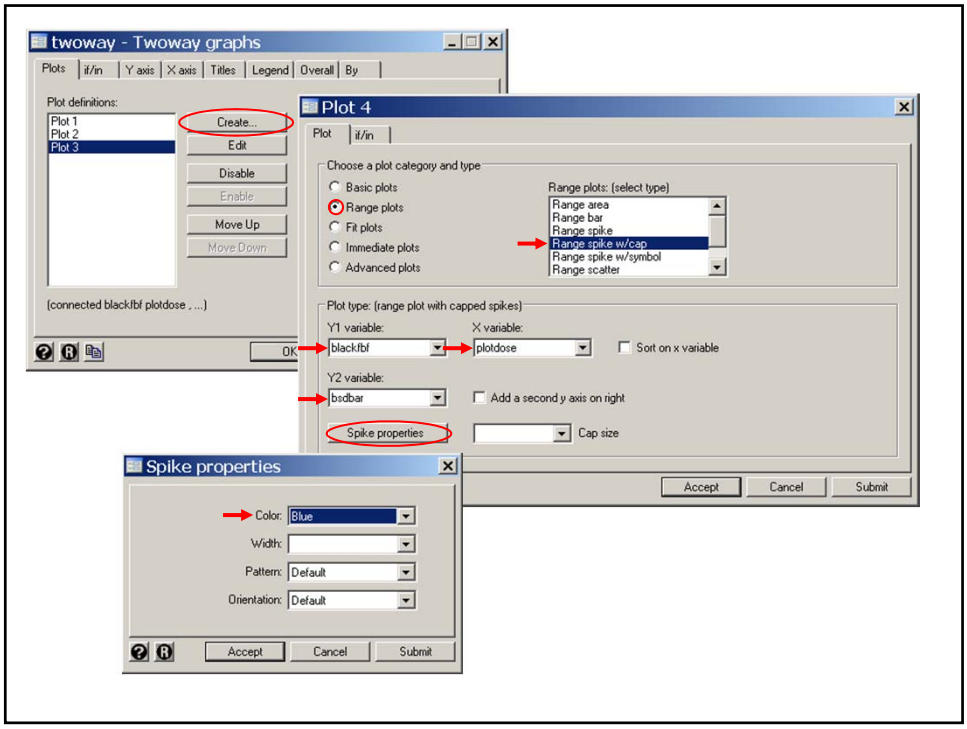
{8} The *xscale(log)* option of the *graph* command causes the *x*-axis to be drawn on a **logarithmic scale**.

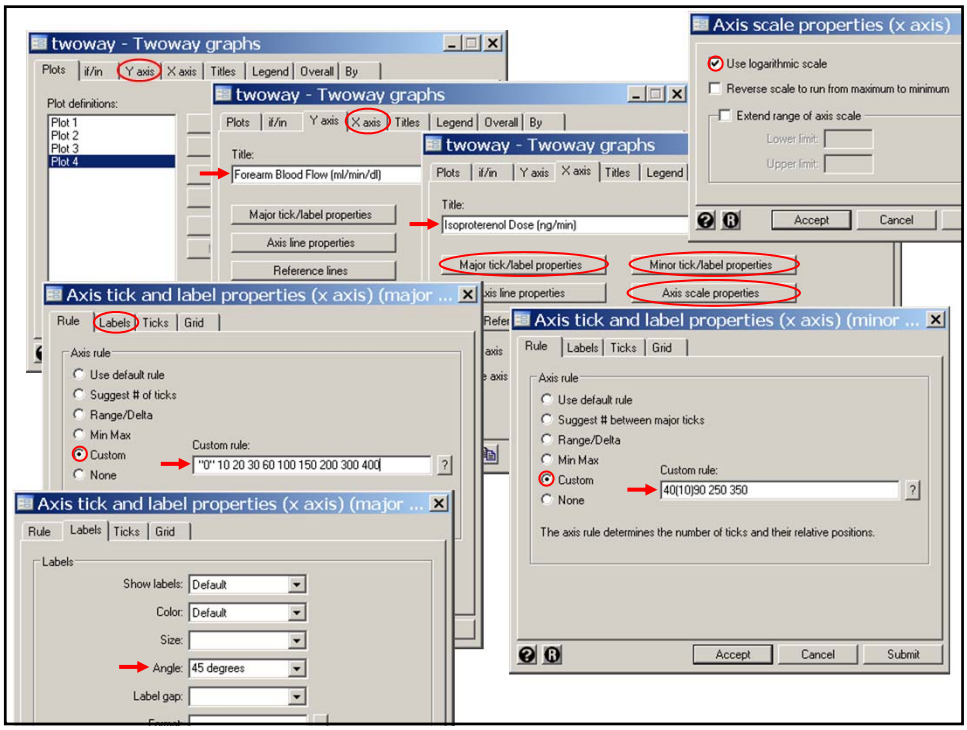
{9} The value 6 is assigned the label "0"

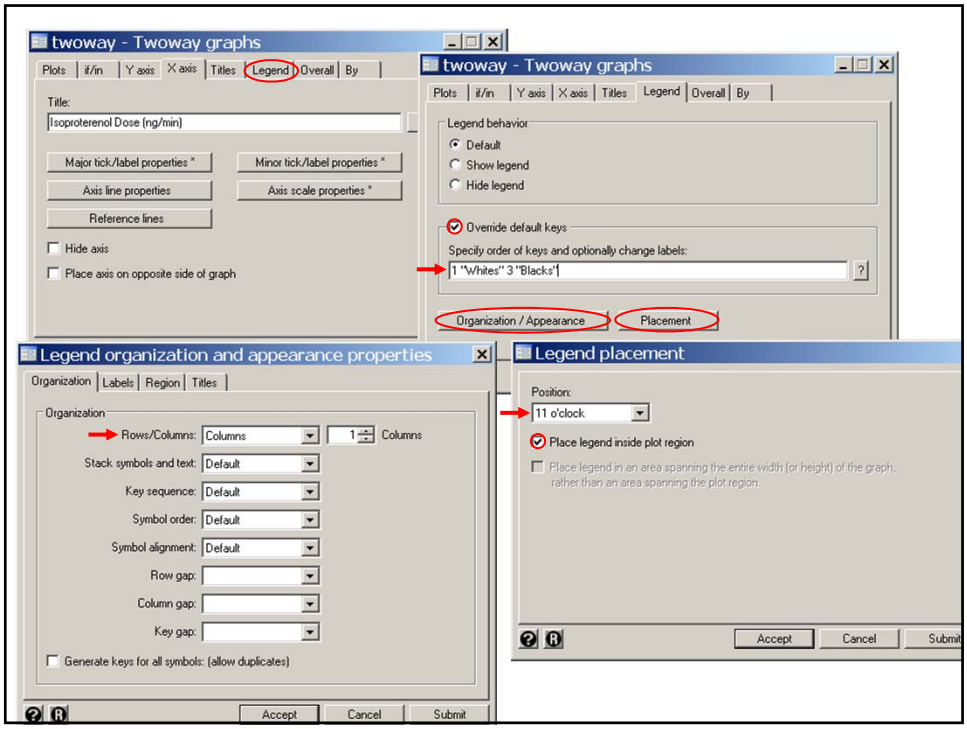


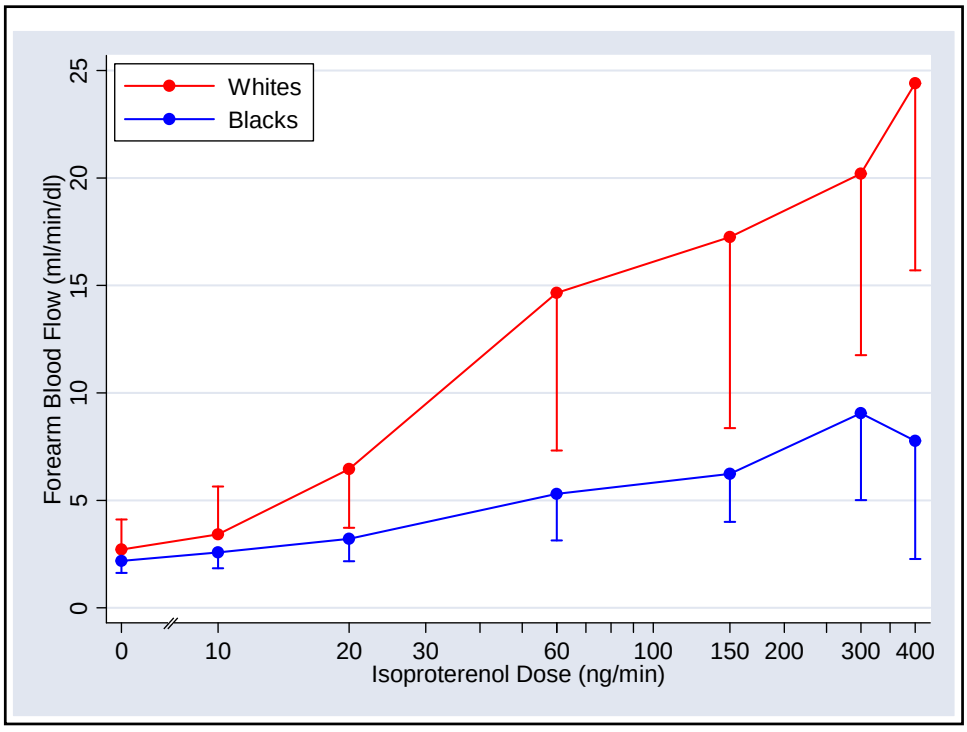

twoway - Twoway graphs
Plots
if/in
Y axis
X axis
Titles
Legend
Overall
By
Plot definitions:
Plot 1
Create...
Edit
Disable
Enable
Move Up
Move Down
(connected whitefbf plotdose , ...)
OK
Plot 2
Plot
if/in
Choose a plot category and type
Basic plots
Range plots
Fit plots
Immediate plots
Advanced plots
Range plots: (select type)
Range area
Range bar
Range spike
Range spike w/cap
Range spike w/symbol
Range scatter
Plot type: (range plot with capped spikes)
Y1 variable:
X variable:
whitefbf
plotdose
Sort on x variable
Y2 variable:
wsdbar
Add a second y axis on right
Spike properties
Cap size
Accept
Cancel
Submit
Spike properties
Color:
Red
Width:
Pattern:
Default
Orientation:
Default
Accept
Cancel
Submit


twoway - Twoway graphs
Plots
if/in
Y axis
X axis
Titles
Legend
Overall
By
Plot definitions:
Plot 1
Plot 2
Create...
Edit
Disable
Enable
Move Up
Move Down
(rcap whitefbf wsdbar plotdose, ...)
OK
Plot 3
Plot
if/in
Choose a plot category and type
Basic plots
Range plots
Fit plots
Immediate plots
Advanced plots
Basic plots: (select type)
Scatter
Line
Connected
Area
Bar
Spike
Plot type: (connected-line plot)
Y variable:
X variable:
blackfbf
plotdose
Add a second y axis on right
Line properties
Marker properties
Line properties
Color:
Blue
Width:
Pattern:
Default
Connecting method:
Default
Missing values:
Default
Accept
Cancel
Submit
Marker properties
Main
Advanced
Marker properties
Symbol:
Default
Color:
Blue
Size:
Add labels to markers
Variable:
Label color:
Default
Label size:
Label position:
Default
Label angle:
Label gap:
Accept
Cancel
Submit


twoway - Twoway graphs
Plots
if/in
Y axis
X axis
Titles
Legend
Overall
By
Plot definitions:
Plot 1
Plot 2
Plot 3
Create...
Edit
Disable
Enable
Move Up
Move Down
(connected blackfbf plotdose , ...)
OK
Plot 4
Plot
if/in
Choose a plot category and type
Basic plots
Range plots
Fit plots
Immediate plots
Advanced plots
Range plots: (select type)
Range area
Range bar
Range spike
Range spike w/cap
Range spike w/symbol
Range scatter
Plot type: (range plot with capped spikes)
Y1 variable:
blackfbf
X variable:
plotdose
Sort on x variable
Y2 variable:
bsdbar
Add a second y axis on right
Spike properties
Cap size
Accept
Cancel
Submit
Spike properties
Color:
Blue
Width:
Pattern:
Default
Orientation:
Default
Accept
Cancel
Submit


twoway - Twoway graphs
Plots
if/in
Y axis
X axis
Titles
Legend
Overall
By
Plot definitions:
Plot 1
Plot 2
Plot 3
Plot 4
Title:
Forearm Blood Flow (ml/min/dl)
Major tick/label properties
Axis line properties
Reference lines
Title:
Isoproterenol Dose (ng/min)
Major tick/label properties
Minor tick/label properties
Axis scale properties
Axis scale properties (x axis)
Use logarithmic scale
Reverse scale to run from maximum to minimum
Extend range of axis scale
Lower limit:
Upper limit:
Accept
Cancel
Axis tick and label properties (x axis) (major ...
Rule
Labels
Ticks
Grid
Axis rule
Use default rule
Suggest # of ticks
Range/Delta
Min Max
Custom
None
Custom rule:
"0" 10 20 30 60 100 150 200 300 400
Axis tick and label properties (x axis) (minor ...
Rule
Labels
Ticks
Grid
Axis rule
Use default rule
Suggest # between major ticks
Range/Delta
Min Max
Custom
None
Custom rule:
40(10)90 250 350
The axis rule determines the number of ticks and their relative positions.
Accept
Cancel
Submit
Axis tick and label properties (x axis) (major ...
Rule
Labels
Ticks
Grid
Labels
Show labels:
Default
Color:
Default
Size:
Angle:
45 degrees
Label gap:

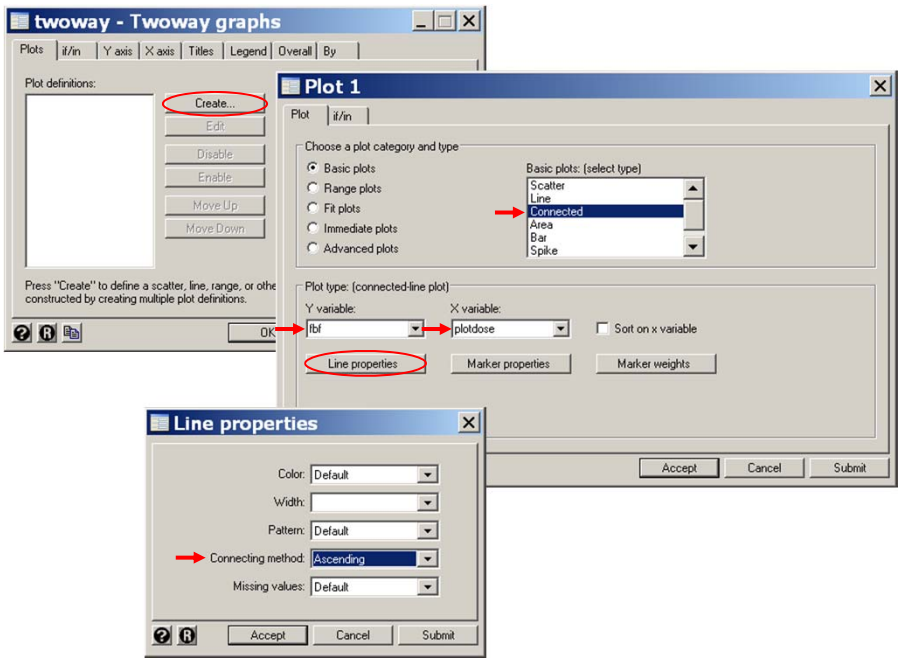

twoway - Twoway graphs
Plots
if/in
Y axis
X axis
Titles
Legend
Overall
By
Title:
Isoproterenol Dose (ng/min)
Major tick/label properties
Minor tick/label properties
Axis line properties
Axis scale properties
Reference lines
Hide axis
Place axis on opposite side of graph
twoway - Twoway graphs
Legend behavior
Default
Show legend
Hide legend
Override default keys
Specify order of keys and optionally change labels:
1 "Whites" 3 "Blacks"
Organization / Appearance
Placement
Legend organization and appearance properties
Organization
Labels
Region
Titles
Rows/Columns:
Columns
Stack symbols and text:
Key sequence:
Symbol order:
Symbol alignment:
Row gap:
Column gap:
Key gap:
Generate keys for all symbols: (allow duplicates)
Accept
Cancel
Submit
Legend placement
Position:
11 o'clock
Place legend inside plot region
Place legend in an area spanning the entire width (or height) of the graph, rather than an area spanning the plot region.
Accept
Cancel
Submit

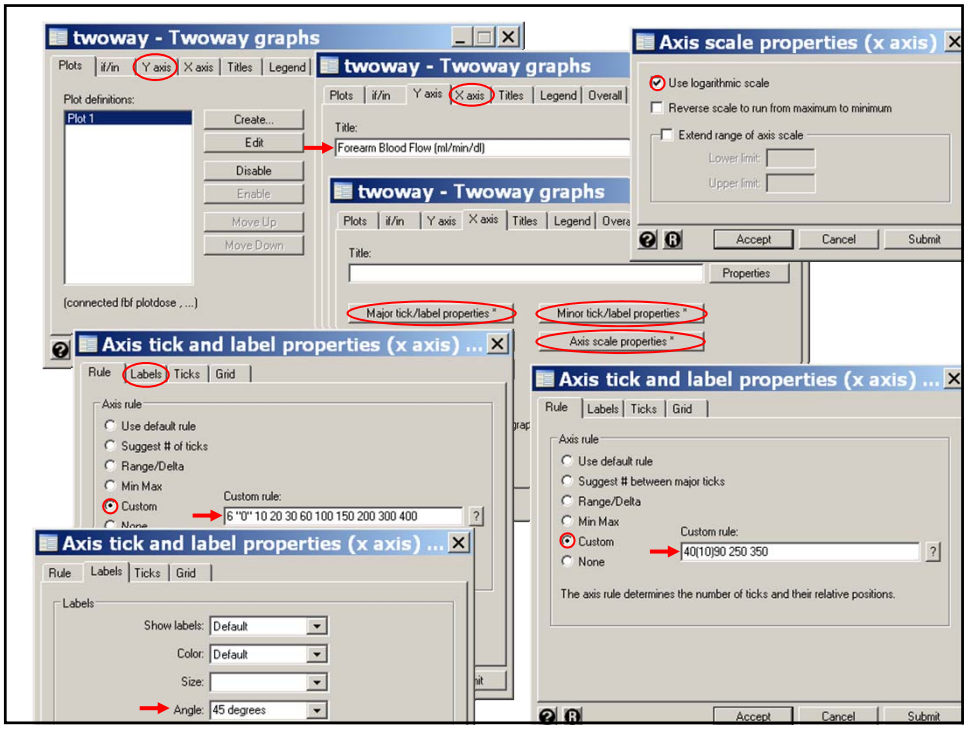

Whites
Blacks
Forearm Blood Flow (ml/min/dl)
0
5
10
15
20
25
0
10
20
30
60
100
150
200
300
400
Isoproterenol Dose (ng/min)

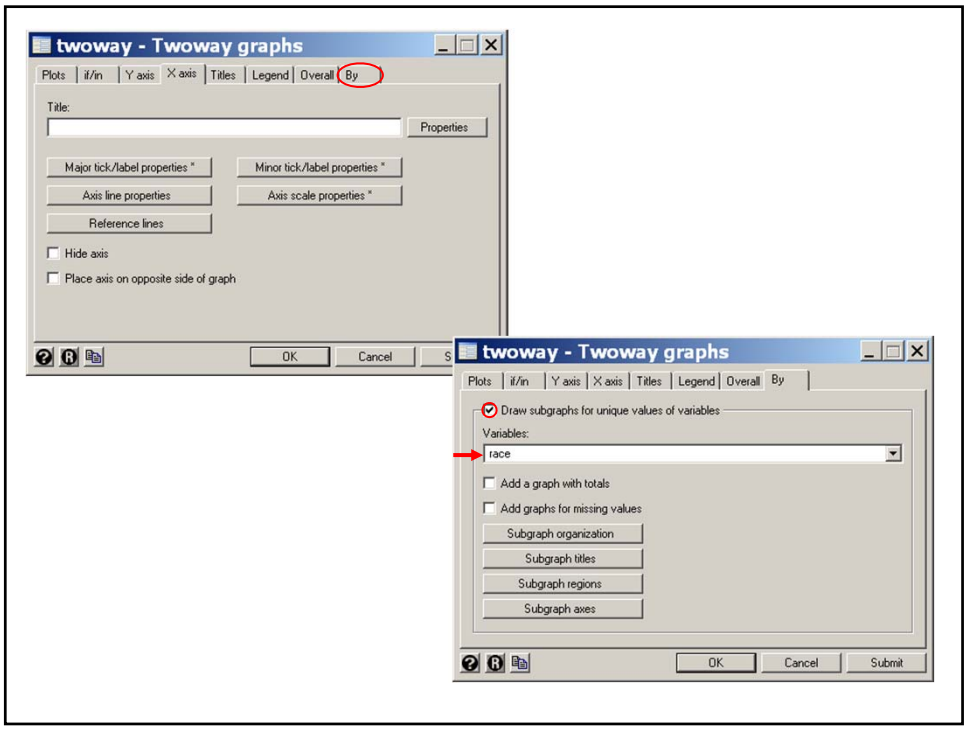
```
. *
. *  Plot individual responses for white and black patients
. *
. use C:\WDDtext\11.2.Long.Isoproterenol.dta, clear                          {10}

. sort id plotdose

. twoway connect fbf plotdose, connect(L)  xscale(log)                  /// {11}
>     xlabel(6 "0" 10 20 30 60 100 150 200 300 400, angle(45))          ///
>     xtick(40(10)90 250 350) ylabel(0(5)40, angle(0))                  ///
>     ytitle(Forearm Blood Flow (ml/min/dl)) by(race)
```

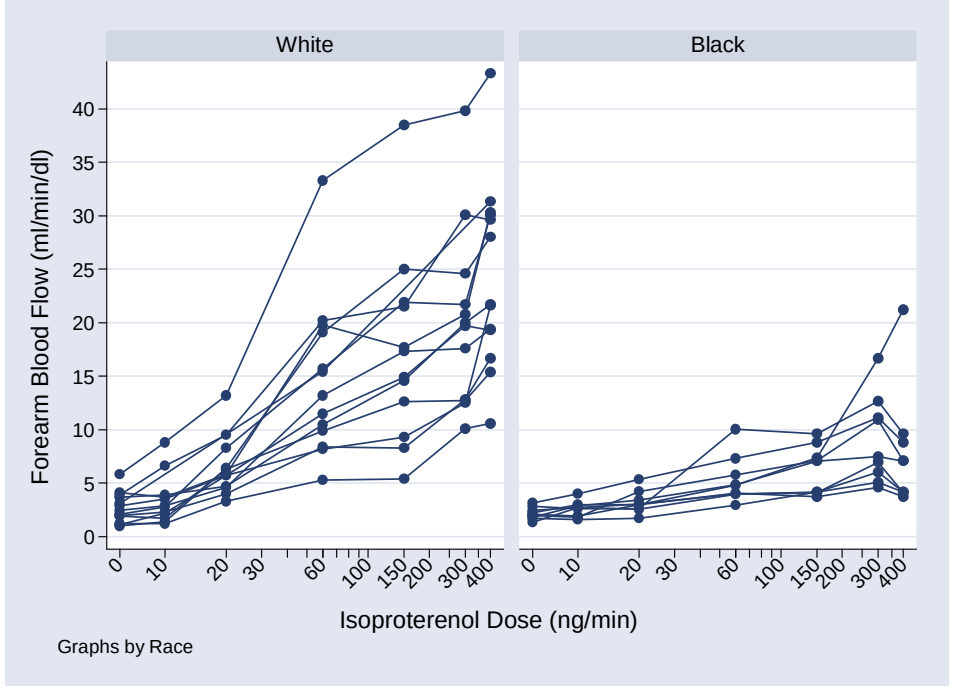
{10} We restore the long form of the data set. Note that this data set was destroyed in memory by the preceding *collapse* command.

{11} The ***connect(L)*** option specifies that straight lines are to **connect** consecutive points as long as the values of the *x*-variable, *plotdose*, are **increasing**. Otherwise the points are not connected. Note that in the preceding command we sorted the data set by *id* and *plotdose*. This has the effect of grouping all observations on the same patient together and of ordering the values on each patient by increasing values of *plotdose*. Hence, ***connect(L)*** will connect the values for **each patient** but will not **connect** the last value of one patient with the first value of the next. ***by(race)*** causes seprate graphs to be made for each race.




twoway - Twoway graphs
Plots
if/in
Y axis
X axis
Titles
Legend
Plot definitions:
Plot 1
Create...
Edit
Disable
Enable
Move Up
Move Down
(connected fbf plotdose , ...)
Title:
Forearm Blood Flow (ml/min/dl)
Overall
Major tick/label properties "
Minor tick/label properties "
Axis scale properties "
Axis scale properties (x axis)
Use logarithmic scale
Reverse scale to run from maximum to minimum
Extend range of axis scale
Lower limit:
Upper limit:
Accept
Cancel
Submit
Properties
Axis tick and label properties (x axis) ...
Rule
Labels
Ticks
Grid
Axis rule
Use default rule
Suggest # of ticks
Range/Delta
Min Max
Custom
None
Custom rule:
6 "0" 10 20 30 60 100 150 200 300 400
Suggest # between major ticks
40(10)90 250 350
The axis rule determines the number of ticks and their relative positions.
Labels
Show labels:
Default
Color:
Size:
Angle:
45 degrees


twoway - Twoway graphs
Plots
if/in
Y axis
X axis
Titles
Legend
Overall
By
Title:
Properties
Major tick/label properties "
Minor tick/label properties "
Axis line properties
Axis scale properties "
Reference lines
Hide axis
Place axis on opposite side of graph
OK
Cancel
Draw subgraphs for unique values of variables
Variables:
race
Add a graph with totals
Add graphs for missing values
Subgraph organization
Subgraph titles
Subgraph regions
Subgraph axes
Submit



The following log file and comments illustrates how to perform the response feature analysis described in the preceding section.

```
. *  11.5.Isoproterenol.log
. *
. *  Perform a response feature analysis of the effect of race and dose of
. *  isoproterenol on blood flow using the data of Lang et al. (1995).
. *  For each patient, we will perform separate linear regressions of change in
. *  blood flow against log dose of isoproterenol.  The response feature that we
. *  will use is the slope of each individual regression curve.
. *
. use C:\WDDtext\11.2.Long.Isoproterenol.dta, clear
. *
. *  Calculate the regression slope for the first patient
. *
```

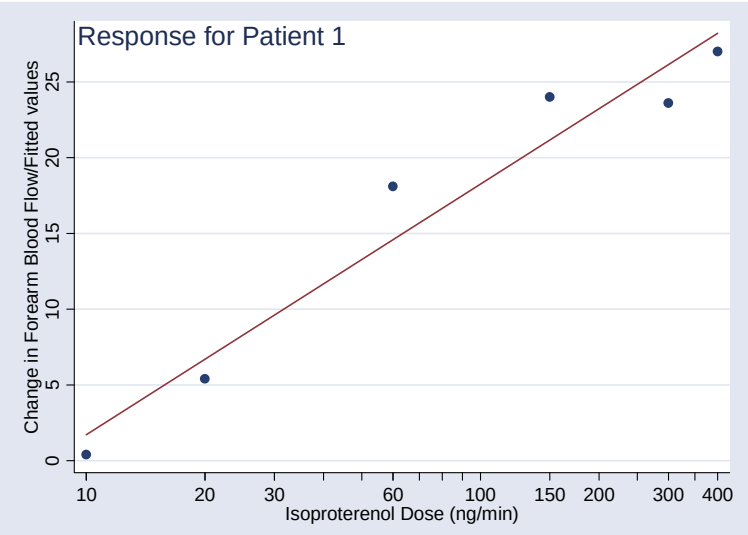
```
. regress delta_fbf logdose if id == 1                                    {1}

   Source | SS                    df         MS         Number of obs =     6
----------+-------------------------------------       F( 1, 4)      = 71.86
    Model |  570.114431   1       570.114431            Prob > F      =0.0011
 Residual |  31.7339077   4       7.93347694            R-squared     =0.9473
----------+-------------------------------------       Adj R-squared =0.9341
    Total |  601.848339   5       120.369668            Root MSE      =2.8166
------------------------------------------------------------------------------
 delta_fbf |  Coef.    Std. Err.   t          P>|t|      [95% Conf. Interval]
-----------+------------------------------------------------------------------
   logdose |  7.181315  .8471392    8.48       0.001     4.82928  9.533351
     _cons |-14.82031  3.860099    -3.84      0.018      -25.53767 -4.10296
------------------------------------------------------------------------------

. predict yhat
(option xb assumed; fitted values)
(22 missing values generated)
```

{1} We **regress** change in blood **flow** against **log dose** of isoproterenol for the observations from the **first** patient. Note that *logdose* is missing when *dose* = 0. Hence, only the six positive doses are included in this analysis. The regression slope for this patient is **7.18**. We could obtain the slopes for all 22 patients with the command

```
by id: regress delta_fbf logdose
```

However, this would require extracting the slope estimates by hand and re-entering them into Stata. This is somewhat tedious to do and is prone to transcription error. Alternately, we can use the ***statsby*** command as explained below.

```
. scatter delta_fbf dose if dose !=0 & id==1                    ///
>     || line yhat  dose if dose !=0 & id==1                    /// {2}
>     ||, ylabel(0 5 10 15 20 25) xscale(log)                   ///
>        xlabel(10 20 30 60 100 150 200 300 400)                ///
>        xtick(10(10)90 250 350) legend(off)                    ///
>        title("Response for Patient 1", position(11) ring(0))      {3}
```



{3} The position of a graph title is controled in the same way as the graph legend.

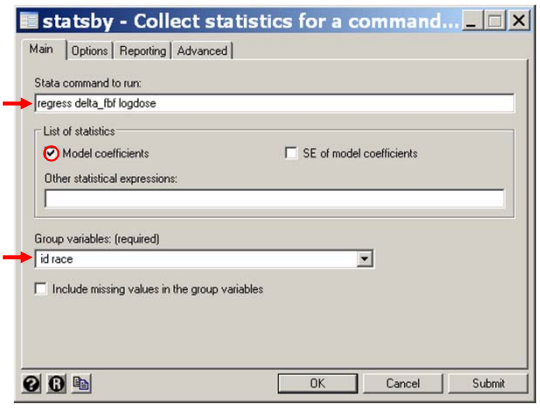
{2} Note that

```
lfit yhat dose
```

will not give the desired results since we are regressing *delta_fbf* against *logdose*.

```
. *
. * Calculate regression slopes for each patient.
. * Reduce data set to one record per patient.
. * The variable slope contains the regression slopes.
. * Race is include in the following by statement to keep this
. * variable in the data file.
. *
. * Statistics > Other > Collect statistics for a command across a by list
. statsby  slope = _b[logdose], by(id race) clear:                  ///     {3}
>     regress delta_fbf logdose
(running regress on estimation sample)

      command:  regress delta_fbf logdose
        slope:  _b[logdose]
           by:  id race

Statsby groups
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
.....................
```

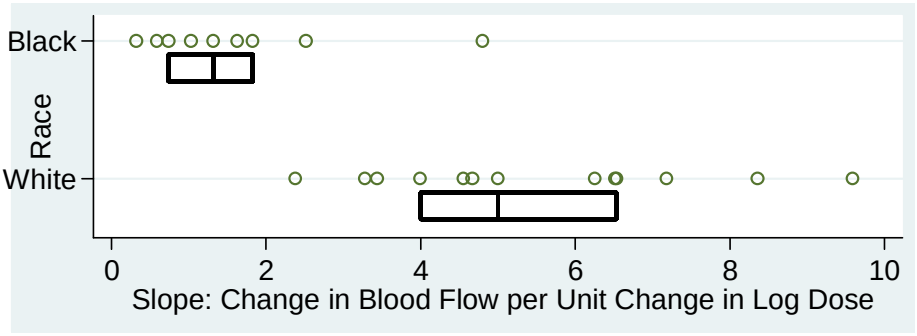
{3} This ***statsby*** command performs a **separate regression** of ***delta_fbf*** against ***logdose*** for each unique combination of values of the variables given by the ***by*** option. In this example, these variables are ***id*** **and** ***race***. The original data set is discarded and is replaced by a new data set with one record per patient. The term ***slope = _b[logdose]*** creates a new variable called ***slope*** that contains the **slope coefficient** of each individual regression. The variables that remain in the data set are ***slope*** and the ***by*** option variables (***id*** and ***race***). Note that, since *id* uniquely specifies each patient, it is not necessary to specify *race* in the *by* option to generate these regressions. However, we include *race* in the *by* option in order to keep this variable in the data set. The *clear* option allows the original data set to be replaced even if it has not been saved.




```
. * Data > Describe data > List data
. list id _b_logdose race                                          {4}
     +-----------------------+
     | id   _b_log~e    race |
     |-----------------------|
  1. |  1   7.181315   White |
  2. |  2   6.539237   White |
  3. |  3   3.999704   White |
  4. |  4   4.665485   White |
  5. |  5   4.557809   White |
     |-----------------------|
  6. |  6   6.252436   White |
  7. |  7   2.385183   White |
  8. |  8   8.354769   White |
  9. |  9   9.590916   White |
 10. | 10   6.515281   White |
     |-----------------------|
 11. | 11   3.280572   White |
 12. | 12   3.434072   White |
 13. | 13   5.004545   White |
 14. | 14   .5887727   Black |
 15. | 15   1.828892   Black |
     |-----------------------|
 16. | 16   .3241574   Black |
 17. | 17    1.31807   Black |
 18. | 18   1.630882   Black |
 19. | 19   .7392463   Black |
 20. | 20   2.513615   Black |
     |-----------------------|
 21. | 21   1.031773   Black |
 22. | 22   4.805952   Black |
     +-----------------------+
```

{4} We **list** the individual **slope estimates** for each patient. Note that the highlighted slope estimate for the first patient is identical to the estimate obtained earlier with the *regress* command.

```
. stripplot slope, over(race)  boffset(-0.1)                          /// {5}
    box(lwidth(medthick)   barwidth(0.1)) xsize(5)
                                                    ///
    xtitle(Slope: Change in Blood Flow per Unit Change in Log Dose)
```



{5} This graph, which is similar to the box plot figure, highlights the difference in the distribution of slope estimates between blacks and whites.

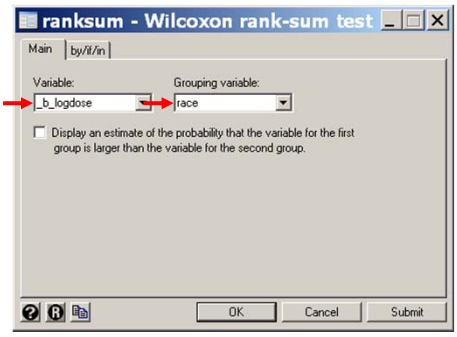
The *stripplot* command is a user-contributed command that must be downloaded using the *findit* command before use. The command

```
. graph7 slope,  by(race) oneway box
```

produces a similar graph.

```
. *
. *  Do ranksum test on slopes.
. *
. * Statistics > Nonparametric... > Tests... > Wilcoxon rank-sum test
. ranksum slope, by(race)                                                {6}
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

        race |      obs    rank sum    expected
-------------+---------------------------------
       White |       13         201       149.5
       Black |        9          52       103.5
-------------+---------------------------------
    combined |       22         253         253

unadjusted variance      224.25
adjustment for ties       -0.00
                        ----------
adjusted variance        224.25

Ho: slope(race==White) = slope(race==Black)
             z =   3.439
    Prob > |z| =   0.0006
```

{6} This ***ranksum*** command performs a Wilcoxon-Mann-Whitney rank sum test of the **null hypothesis** that the **distribution of slopes** is the same for both races. The test is highly significant giving a *P* value of **0.0006**.




```
. *
. *  Do t tests comparing change in blood flow in blacks
. *  and whites at different doses

. *
. use C:\WDDtext\11.2.Long.Isoproterenol.dta, clear
                                                                {7}
 . sort dose
. *  Data > Create or change data > Keep or drop observations
. drop if dose == 0
(22 observations deleted)

. * Statistics > Summaries... > Classical... > Two-group mean-comparison test
. by dose: ttest delta_fbf , by(race) unequal                   {8}
```

{7} The preceding ***statsby*** command **deleted** most of the data. We must read in the **data** set before performing *t* tests at the different doses.

{8} This *ttest* command performs independent ***t* tests** of ***delta_fbf*** in blacks and whites at **each dose** of isoproterenol. The output for doses 60, 150 and 300 have been omitted. The highlighted output from this command is also given in the following Table 10.1.

```
-> dose = 10

Two-sample t test with unequal variances

------------------------------------------------------------------------------
   Group |   Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   White |    12   .7341667    .3088259    1.069804    .0544455    1.413888
   Black |     9   .3966667    .2071634    .6214902   -.081053    .8743863
---------+--------------------------------------------------------------------
combined |    21   .5895238    .1967903    .9018064    .1790265    1.000021
---------+--------------------------------------------------------------------
    diff |          .3375    .3718737               -.4434982    1.118498
------------------------------------------------------------------------------
Satterthwaite's degrees of freedom:  18.0903

                 Ho: mean(White) - mean(Black) = diff = 0

    Ha: diff < 0                 Ha: diff ~= 0                 Ha: diff > 0
      t =   0.9076               t =   0.9076                  t =   0.9076
 P < t =   0.8120           P > |t| =   0.3760          P > t =   0.1880
```

```
-> dose = 20

Two-sample t test with unequal variances

------------------------------------------------------------------------------
   Group |   Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   White |    12   3.775833    .6011875    2.082575    2.452628    5.099038
   Black |     9       1.03    .3130229    .9390686     .308168    1.751832
---------+--------------------------------------------------------------------
combined |    21   2.599048    .4719216    2.162616    1.614636    3.583459
---------+--------------------------------------------------------------------
    diff |          2.745833    .6777977               1.309989    4.181677
------------------------------------------------------------------------------
Satterthwaite's degrees of freedom:  16.1415

                 Ho: mean(White) - mean(Black) = diff = 0

    Ha: diff < 0                 Ha: diff ~= 0                 Ha: diff > 0
      t =   4.0511               t =   4.0511                  t =   4.0511
 P < t =   0.9995           P > |t| =   0.0009          P > t =   0.0005
```

{Output omitted. See Table 10.1}

```
-> dose = 400

Two-sample t test with unequal variances

------------------------------------------------------------------------------
Group     |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   White |      13    21.69308    2.163637    7.80110    16.97892    26.40724
   Black |       9     5.58666      1.803   5.410649    1.427673     9.74566
---------+--------------------------------------------------------------------
combined |      22    15.10409    2.252517   10.56524    10.41972    19.78846
---------+--------------------------------------------------------------------
    diff |            16.10641   2.816756               10.2306    21.98222
------------------------------------------------------------------------------
Satterthwaite's degrees of freedom:  19.9917

                 Ho: mean(White) - mean(Black) = diff = 0

     Ha: diff < 0                 Ha: diff ~= 0                 Ha: diff > 0
     t =   5.7181                 t =   5.7181                 t =   5.7181
 P < t =   1.0000             P > |t| =   0.0000             P > t =   0.0000
```

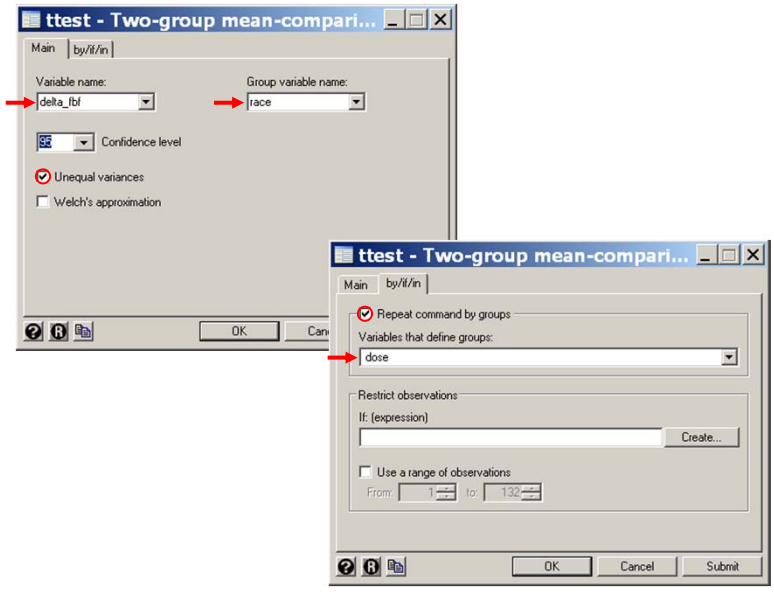


Table 10.1

	Dose of Isoproterenol (ng/min)					
	10	**20**	**60**	**150**	**300**	**400**
White Subjects						
Mean Change from Baseline	0.734	3.78	11.9	14.6	17.5	21.7
Standard Error	0.309	0.601	1.77	2.32	2.13	2.16
95% Confidence Interval	0.054 to 1.4	2.5 to 5.1	8.1 to 16	9.5 to 20	13 to 22	17 to 26
Black Subjects						
Mean Change from Baseline	0.397	1.03	3.12	4.05	6.88	5.59
Standard Error	0.207	0.313	0.607	0.651	1.30	1.80
95% Confidence Interval	-0.081 to 0.87	0.31 to 1.8	1.7 to 4.5	2.6 to 5.6	3.9 to 9.9	1.4 to 9.7
Mean Difference						
White – Black	0.338	2.75	8.82	10.5	10.6	16.1
95% Confidence Interval	-0.44 to 1.1	1.3 to 4.2	4.8 to 13	5.3 to 16	5.4 to 16	10 to 22
P value	0.38	0.0009	0.0003	0.0008	0.0005	<0.0001

3. The Area-Under-the-Curve Response Feature

A response feature that is often useful in response feature analysis is the **area under the curve.**

Let $y_i(t)$ be the response from the i^{th} patient at the time t.
$y_{ij} = y_i(t_j)$ at times $t_1, t_2, \ldots t_j, \ldots, t_n$

We can estimate the area under the curve $y_i(t)$ between t_1 and t_n as follows:

Draw a scatterplot of y_{ij} against t_j for $j = 1,2,\ldots,n$. Then draw straight lines connecting the points (t_1,y_{i1}), (t_2,y_{i2}), …, (t_n,y_{in}).

We estimate the area under the curve to be the area under these lines. Specifically, the area under the line from (t_j,y_{ij}) to $(t_j{+}_1,y_{i,j+1})$ is

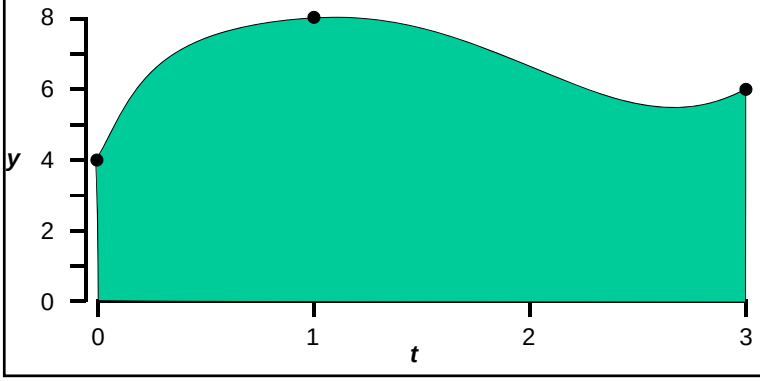
$$\left(\frac{y_{ij} + y_{i,j+1}}{2}\right)\left(t_{j+1} - t_j\right)$$

Hence, the area under the entire curve is estimated by

$$\sum_{j=1}^{n-1}\left(\frac{y_{ij}+y_{i,j+1}}{2}\right)\left(t_{j+1}-t_j\right) \qquad \{10.1\}$$

For example, if $n = 3$, $t_1 = 0$, $t_2 = 1$, $t_3 = 3$, $y_{i1} = 4$, $y_{i2} = 8$, and $y_{i3} = 6$ then equation (10.1) reduces to

$$\left(\frac{4+8}{2}\right)(1-0)+\left(\frac{8+6}{2}\right)(3-1)=20.$$

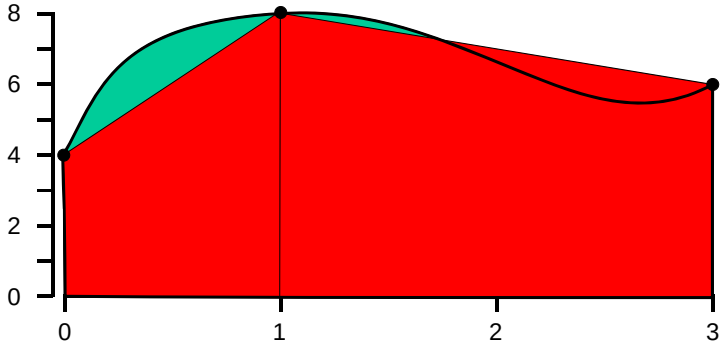


Hence, the area under the entire curve is estimated by

$$\sum_{j=1}^{n-1}\left(\frac{y_{ij}+y_{i,j+1}}{2}\right)\left(t_{j+1}-t_j\right) \qquad \{10.1\}$$

For example, if $n = 3$, $t_1 = 0$, $t_2 = 1$, $t_3 = 3$, $y_{i1} = 4$, $y_{i2} = 8$, and $y_{i3} = 6$ then equation (0.13) reduces to

$$\left(\frac{4+8}{2}\right)(1-0)+\left(\frac{8+6}{2}\right)(3-1)=20.$$

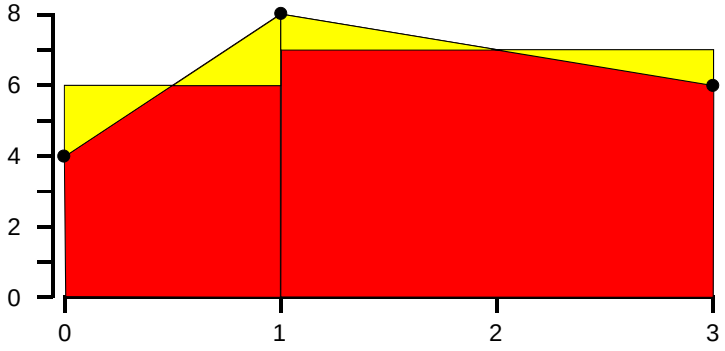


Hence, the area under the entire curve is estimated by

$$\sum_{j=1}^{n-1}\left(\frac{y_{ij}+y_{i,j+1}}{2}\right)\left(t_{j+1}-t_j\right) \qquad \{10.1\}$$

For example, if $n = 3$, $t_1 = 0$, $t_2 = 1$, $t_3 = 3$, $y_{i1} = 4$, $y_{i2} = 8$, and $y_{i3} = 6$ then equation (0.13) reduces to

$$\left(\frac{4+8}{2}\right)(1-0)+\left(\frac{8+6}{2}\right)(3-1)=20.$$



In a response feature analysis based on area under the curve, we use equation {10.1} to calculate this area for each patient and then perform a one-way analysis of variance on these areas.

Equation {10.1} can be implemented in Stata as follows. Let

id be the patient's identification number *i,*
time be the patient's time of observation t_j,
response be the patient's response $y_i\left(t_j\right)$.

Then the area under the response curve for study subjects can be calculated by using the following Stata code

```
sort id time
*
* Delete records with missing values for time or response
*
*  Data > Create or change data > Keep or drop observations
drop if time == . | response == .
generate area=(response+response[_n+1])*(time[_n+1]-time)/2 if id==id[_n+1]
collapse (sum) area = area , by(id)
*
* The variable area is now the area under the curve for
* each patient defined by equation {10.1}.  The data file
* contains one record per patient.
```

4. Generalized Estimating Equations (GEE)

This is a popular and more sophisticated approach to modeling mixed effects response data.

It is basically a generalization of the generalized linear model to allow repeated measures per subject. An appropriate correlation structure for the responses from each patient is built into the model.

Let n be the number of patients studied,
n_i, number of observations on the i^{th} patient,
y_{ij} be the response of the i^{th} patient at her j^{th} observation,
$x_{ij1}, x_{ij2}, \ldots, x_{ijq}$ be q covariates that are measured on her at this time,
$\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ijq})$ denote the values of all of the covariates for the i^{th} patient at her j^{th} observation.

Then the model used by GEE analysis assumes that:

1. The distribution of y_{ij} belongs to the exponential family of distributions.

2. The expected value of y_{ij} given the patient's covariates $x_{ij1}, x_{ij2}, \ldots, x_{ijq}$ is related to the model parameters through an equation of the form

$$g\left[\mathrm{E}\left[y_{ij} \mid \mathbf{x}_{ij}\right]\right] = \alpha + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \ldots + \beta_q x_{ijq} \qquad \{10.2\}$$

g is the link function

$\alpha + \beta_1 x_{ij_1} + \beta_2 x_{ij2} + \ldots + \beta_q x_{ijq}$ is the linear predictor.

3. Responses from different patients are mutually independent.

When there is only one observation per patient (for all i), model {10.2} is, in fact, the generalized linear model. In this case,

when g is the identity function ($g[y] = y$), and y_{ij} is normally distributed, {10.2} reduces to multiple linear regression;

when g is the logit function and y_{ij} has a binomial distribution, {10.2} describes logistic regression;

when g is the logarithmic function and y_{ij} has a Poisson distribution, this model becomes Poisson regression.

Model {10.2} differs from the generalized linear model in that it does not make any assumptions about how observations on the same patient are correlated.

5. Common Correlation Structures

Let ρ_{jk} denote the population correlation coefficient between j^{th} the and k^{th} observations on the same patient. If all patients have n observations, then

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2n} \\ \rho_{31} & \rho_{32} & 1 & \cdots & \rho_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \rho_{n3} & \cdots & 1 \end{bmatrix} \qquad \{10.3\}$$

R is called the **correlation matrix** for repeated observations on study subjects. In this matrix, the coefficient in the j^{th} row and k^{th} column is the correlation coefficient between the j^{th} and k^{th} observations.

{10.3} is called an **unstructured correlation** matrix. It

- makes no assumptions about the correlation structure
- requires $n(n-1) / 2$ correlation parameters.

An **exchangeable correlation** structure assumes that

$$\mathbf{R} = \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{bmatrix} \qquad \{10.4\}$$

Any two distinct observations from the same patient have the same correlation coefficient ρ.

Many data sets have much more complicated correlation structures.

Observations on a patient taken closer in time are often more correlated than observations taken far apart.

Correlation structure may vary among patients.

6. GEE Analysis and the Huber-White Sandwich Estimator

GEE analysis is computationally and methodologically complex. The basic idea of the analysis can be summarize as follows:

1. We select a working correlation matrix $\mathbf{R}_i$ for each patient. $\mathbf{R}_i$, – usually with an exchangeable correlation structure.
2. We estimate the working variance-covariance matrix for the i^{th} patient.
3. Using the working variance-covariance structure we obtain estimates of the model parameters
4. We estimate the variance-covariance matrix of our model parameters using a technique called the **Huber-White sandwich estimator**.
5. We use our parameter estimates and the Huber-White variance-covariance matrix to test hypotheses or construct confidence intervals from relevant weighted sums of the parameter estimates (see Sections 5.14 through 5.16).

7. Example: Analyzing the Isoproterenol Data with GEE

Suppose that in model {10.2} , y_{ij} is a normally distributed random component and $g[y] = y$ is the identity link function. Then model {10.2} reduces to

$$\mathrm{E}\left[y_{ij} \mid \mathbf{x}_{ij} \right] = \alpha + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \ldots + \beta_q x_{ijq} \qquad \{10.5\}$$

Model {10.5} is a special case of the GEE model {10.2}.

Let

y_{ij} be the change from baseline in forearm blood flow for the i^{th} patient at the j^{th} dose of isoproterenol,

$$white_i = \begin{cases} 1: \text{ if the } i^{th} \text{ patient is white} \\ 0: \text{ if he is black} \end{cases}$$

and

$$dose_{jk} = \begin{cases} 1: \text{ if } j = k \\ 0: \text{ otherwise} \end{cases}$$

We will assume that y_{ij} is normally distributed and

$$\mathrm{E}\left[y_{ij} \mid white_i, j \right] = \alpha + \beta \times white_i + \sum_{k=2}^{6} \left(\gamma_k dose_{jk} + \delta_k \times white_i \times dose_{jk} \right) \qquad \{10.6\}$$

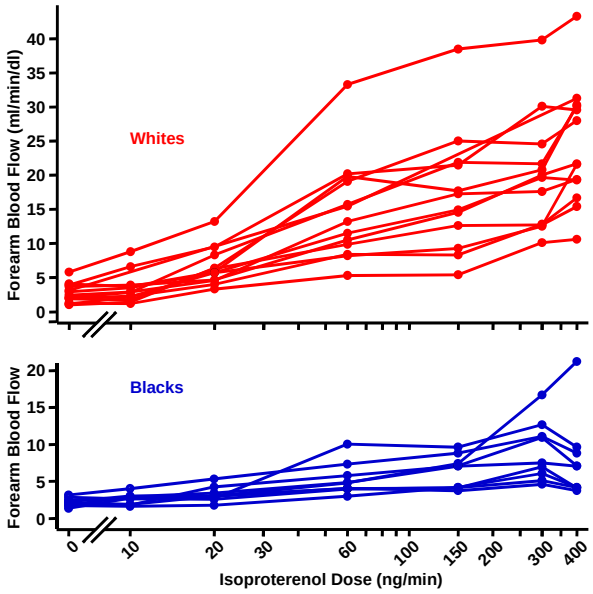
where α, β, $\{\gamma_k,\ \delta_k : k = 2, \ldots, 6\}$ are the model parameters. Model {10.6} is a special case of model {10.5). Note that this model implies that the expected change in blood flow is

α	for a black man on the first dose,	{10.7}
$\alpha + \beta$	for a white man on the first dose,	{10.8}
$\alpha + \gamma_j$	for a black man on the j^{th} dose with $j > 1$, and	{10.9}
$\alpha + \beta + \gamma_j + \delta_j$	for a white man on the j^{th} dose with $j > 1$.	{10.10}

It must be noted that patient 8 in this study has four missing blood flow measurements. This concentration of missing values in one patient causes the choice of the working correlation matrix to have an appreciable effect on our model estimates.

Regardless of the working correlation matrix, the working variance for y_{ij} in model {10.5} is constant.

Figure 10.2 suggests that this variance is greater for whites than blacks and increases with increasing dose.



Hence, it is troubling to have our parameter estimates affected by a working correlation matrix that we know is wrong.

Also, the Huber-White variance-covariance estimate is only valid when the missing values are few and randomly distributed.

For these reasons, we delete patient 8 from our analysis. Without patient 8, the Huber-White variance-covariance matrix is unaffected by the choice of $\mathbf{R}_i$.

Let $\hat{\alpha}$, $\hat{\beta}$, $\left\{\hat{\gamma}_k,\ \hat{\delta}_k : k = 2, ..., 6\right\}$ denote the GEE parameter estimates from the model. Then our estimates of the mean change in blood flow in blacks and whites at the different doses are given by equations {10.7} through {10.10} with the parameter estimates substituting for the true parameter values. Subtracting the estimate of equation {10.7} from that for equation {10.8} gives the estimated mean difference in change in flow between whites and blacks at dose 1, which is

$$\left(\hat{\alpha} + \hat{\beta}\right) - \hat{\alpha} = \hat{\beta} \qquad \{10.11\}$$

Subtracting the estimate of equation {10.9}) from that for equation {10.10} gives the estimated mean difference in change in flow between whites and blacks at dose $j > 1$, which is

$$\left(\hat{\alpha} + \hat{\beta} + \hat{\gamma}_j + \hat{\delta}_j\right) - \left(\hat{\alpha} + \hat{\gamma}_j\right) = \left(\hat{\beta} + \hat{\delta}_j\right) \qquad \{10.12\}$$

Tests of significance and 95% confidence intervals can be calculated for these estimates using the Huber-White variance-covariance matrix.

This is done in the same way as was illustrated in logistic regression. These estimates, standard errors, confidence intervals and *P* values are given in the next table.

Figure 10.2

	Dose of Isoproterenol (ng/min)					
	10	**20**	**60**	**150**	**300**	**400**
White Subjects						
Mean Change from Baseline	0.734	3.78	11.9	14.6	17.5	21.2
Standard Error	0.303	0.590	1.88	2.27	32.09	2.23
95% Confidence Interval	0.14 to 1.3	2.6 to 4.9	8.2 to 16	10 to 19	13 to 22	17 to 26
Black Subjects						
Mean Change from Baseline	0.397	1.03	3.12	4.05	6.88	5.59
Standard Error	0.200	0.302	0.586	0.629	1.26	1.74
95% Confidence Interval	0.0044 to 0.79	0.44 to 1.6	2.0 to 4.3	2.8 to 5.3	4.4 to 9.3	2.2 to 9.0
Mean Difference						
White – Black	0.338	2.75	8.79	10.5	10.6	15.6
95% Confidence Interval	-0.37 to 1.0	1.4 to 4.0	4.9 to 13	5.9 to 15	5.9 to 15	10 to 21
P value	0.35	<0.0005	<0.0005	<0.0005	<0.0005	<0.0001

The null hypothesis that there is no interaction between race and dose on blood flow is

$H_0: \delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_6 = 0$

Under this null hypothesis a chi-squared statistic can be calculated that has as many degrees of freedom as there are interaction parameters (in this case five).

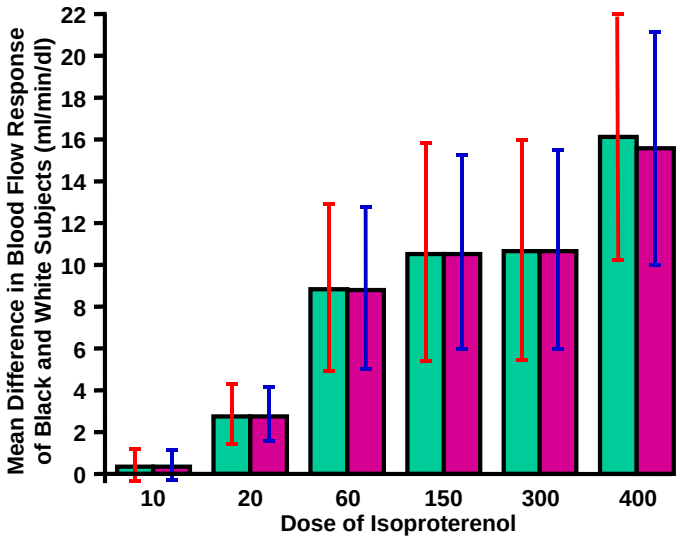
This statistic equals 40.41, which is highly significant ($P < 0.00005$). Hence, we can conclude that the observed interaction is certainly not due to chance.

The GEE and response feature analysis (RFA) in Tables 10.2 and 10.1 should be compared. Note that the mean changes in blood flow in the two races and six dose levels are very similar. They would be identical were if not for the fact that patient 8 is excluded from the GEE analysis but is included in the RFA.

This is a challenging data set to analyze in view of the fact that the standard deviation of the response variable

- increases with dose and
- differs between the races.

The following figure compares the mean difference between blacks and whites at the six different doses. The **green** and **magenta** bars are from the RFA and GEE analyses, respectively.



In this example, response feature analysis and GEE give virtually identical results.

8. Using Stata to Analyze the Isoproterenol Data Set Using GEE

The following log file and comments illustrate how to perform the GEE analysis for the isoproterenol data

```
. *  11.11.Isoproterenol.log
. *
. *  Perform a GEE analyses of the effect of race and dose
. *  of isoproterenol
. *  on blood flow using the data of Lang et al. (1995).
. *
. use C:\WDDtext\11.2.Long.Isoproterenol.dta, clear

. *  Data > Create or change data > Keep or drop observations
. drop if dose == 0 | id == 8                                          {1}
(28 observations deleted)

. generate white = race == 1
```

{1} We **drop** all records with *dose* = **0** or *id* = **8**. When ***dose* = 0**, the change from baseline, ***delta_fbf***, is by definition, **zero**. We eliminate these records as they provide no useful information to our analyses. Patient **8** has **four** missing **values**. These missing values have an adverse effect on our analysis. For this reason we eliminate all observations on this patient (see Section 7).

```
. *
. *  Analyze data using classification variables with
. *  interaction
. *
. * Statistics > Longitudinal... > Generalized est... > Generalized...(GEE)
. xtgee delta_fbf dose##white, i(id) robust                          {2}
Iteration 1: tolerance = 2.061e-13

GEE population-averaged model                   Number of obs      =       126
Group variable:                         id      Number of groups   =        21
Link:                             identity      Obs per group: min =         6
Family:                           Gaussian                     avg =       6.0
Correlation:                  exchangeable                     max =         6
                                                Wald chi2(11)      =    506.86
Scale parameter:                  23.50629      Prob > chi2        =    0.0000

                              (standard errors adjusted for clustering on id)
```

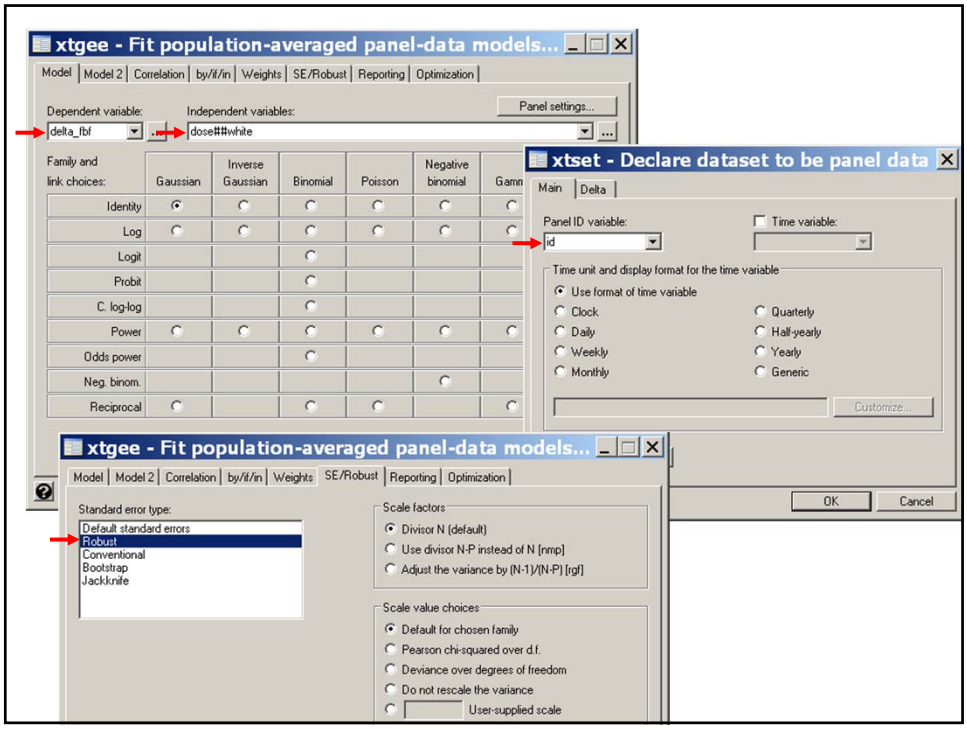
{2} This ***xtgee*** command analyzes model {10.6}. The syntax of *i.dose*white* is analogous to that used for the logistic command in Chapter 4. The **default link function** is the **identity** function. For the **identity link** function the **default random component** is the **normal distribution**. Hence, we do not need to specify either of these aspects of our model explicitly in this command. The ***i(id)*** option specifies ***id*** to be the variable that **identifies** all observations made on the same **patient**. The **exchangeable correlation structure** is the **default** working correlation structure, which we use here. The ***robust*** option specifies that the Huber-White sandwich estimator is to be used. The **table of coefficients** generated by this command is similar to that produced by other **Stata regression commands.**

Note that if we had **not** used the ***robust*** option the model would have assumed that the **exchangeable** correlation structure was true. This would have led to **inaccurate confidence intervals** for our estimates. I strongly recommend that this option always be used in any GEE analysis.

delta_fbf	Coef.	Semi-robust Std. Err.	z	P>|z|	[95% Conf.	Interval]	
dose							
20	.6333333	.2706638	2.34	0.019	.1028421	1.163825	
60	2.724445	.6585882	4.14	0.000	1.433635	4.015254	
150	3.656667	.7054437	5.18	0.000	2.274022	5.039311	
300	6.478889	1.360126	4.76	0.000	3.813091	9.144687	
400	5.19	1.830717	2.83	0.005	1.601861	8.77814	
1.white	.3375	.363115	0.93	0.353	-.3741922	1.049192	{3}
dose#white							
20 1	2.408333	.5090358	4.73	0.000	1.410642	3.406025	
60 1	8.450556	1.823352	4.63	0.000	4.876852	12.02426	
150 1	10.17667	2.20775	4.61	0.000	5.849557	14.50378	
300 1	10.30444	2.305474	4.47	0.000	5.785798	14.82309	
400 1	15.22667	2.748106	5.54	0.000	9.840479	20.61285	
_cons	.3966667	.2001388	1.98	0.047	.0044017	.7889316	{4}

{3} The highlighted term are the estimated **mean**, ***P* value** and 95% **confidence interval** for the difference in response between **white** and **black** men on the **first** dose of isoproterenol (10 ng/min). The parameter estimate associated with the *white* covariate is $\hat{\beta} = 0.3375$ in model {10.6}. The highlighted values in this and in subsequent lines of output are entered into Table 10.2.

{4} The highlighted terms are the estimated **mean**, **standard error** and 95% **confidence interval** for **black** men on the **first** dose of isoproterenol. The parameter estimate associated with *_cons* is $\hat{\alpha} = 0.3967$.




```
. lincom _cons + 1.white                                                    {5}

 ( 1)  1.white + _cons = 0.0

------------------------------------------------------------------------------
   delta_fbf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .7341667    .30298     2.42   0.015     .1403367    1.327997
------------------------------------------------------------------------------

. lincom _cons+ 20.dose                                                     {6}

 ( 1) 20.dose + _cons = 0.0

------------------------------------------------------------------------------
   delta_fbf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |       1.03   .3024088     3.41   0.001     .4372896     1.62271
------------------------------------------------------------------------------

. lincom _cons+ 20.dose +  1.white + 20.dose#1.white                        {7}

 ( 1) 20.dose + 1.white + 20.dose#1.white + _cons = 0.0

------------------------------------------------------------------------------
   delta_fbf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   3.775833   .5898076     6.40   0.000     2.619832    4.931835
------------------------------------------------------------------------------
```

{5} This command calculates $\hat{\alpha}+\hat{\beta}$, the **mean** response for **white** men at the **first** dose of isoproterenol, together with related statistics.

{6} This command calculates $\hat{\alpha}+\hat{\gamma}_2$ the **mean** response for **black** men at the **second** dose of isoproterenol, together with related statistics.

{7} This command calculates $\hat{\alpha}+\hat{\beta}+\hat{\gamma}_2+\hat{\delta}_2$, the **mean** response for **white** men at the **second** dose of isoproterenol, together with related statistics.

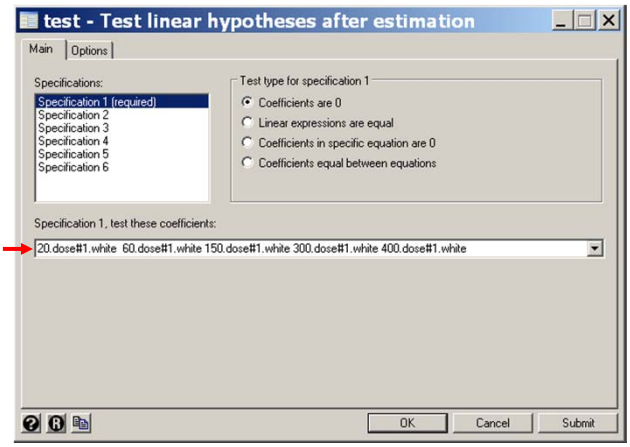
```
. lincom                1.white + 20.dose#1.white                          {8}

 ( 1) 1.white + 20.dose#1.white = 0.0

------------------------------------------------------------------------------
   delta_fbf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   2.745833   .6628153     4.14   0.000     1.446739    4.044927
------------------------------------------------------------------------------

. lincom _cons + 60.dose
                                                {output omitted. See Table 10.2}
. lincom _cons + 60.dose +  1.white +  60.dose#1.white
                                                {output omitted. See Table 10.2}
. lincom                    1.white +  60.dose#1.white
                                                {output omitted. See Table 10.2}
. lincom _cons + 150.dose
                                                {output omitted. See Table 10.2}
. lincom _cons + 150.dose + 1.white + 150.dose#1.white
                                                {output omitted. See Table 10.2}
. lincom                    1.white + 150.dose#1.white
                                                {output omitted. See Table 10.2}
. lincom _cons + 300.dose
                                                {output omitted. See Table 10.2}
. lincom _cons + 300.dose + 1.white + 300.dose#1.white
                                                {output omitted. See Table 10.2}
. lincom                    1.white + 300.dose#1.white
                                                {output omitted. See Table 10.2}
. lincom _cons + 400.dose
```

{8} This calculates $\hat{\beta} + \hat{\delta}_2$, the **mean difference** in response between **white and black** men at the **second** dose of isoproterenol, together with related statistics. Analogous *lincom* commands are also given for dose 3, 4, 5, and 6.

```
( 1) 400.dose + _cons = 0

------------------------------------------------------------------------------
   delta_fbf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   5.586667   1.742395     3.21   0.001     2.171636    9.001698
------------------------------------------------------------------------------

. lincom _cons + 400.dose + 1.white + 400.dose#1.white

 ( 1) 400.dose + 1.white + 400.dose#1.white + _cons = 0.0

------------------------------------------------------------------------------
   delta_fbf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   21.15083   2.233954     9.47   0.000     16.77236    25.5293
------------------------------------------------------------------------------

. lincom                    1.white + 400.dose#1.white

 ( 1) 1.white + 400.dose#1.white = 0.0

------------------------------------------------------------------------------
   delta_fbf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   15.56417   2.833106     5.49   0.000     10.01138    21.11695
------------------------------------------------------------------------------
```

```
. * Statistics > Postestimation > Tests > Test linear hypotheses
. test 20.dose#1.white  60.dose#1.white 150.dose#1.white    ///
>     300.dose#1.white 400.dose#1.white                                    {9}

 ( 1)  20.dose#1.white = 0
 ( 2)  60.dose#1.white = 0
 ( 3)  150.dose#1.white = 0
 ( 4)  300.dose#1.white = 0
 ( 5)  400.dose#1.white = 0

           chi2(  5) =   40.41
         Prob > chi2 =    0.0000
```

{9} This command test the **null hypothesis** that the **interaction** parameters δ_2, δ_3, δ_4, δ_5, and δ_6 are **simultaneously** equal to zero. That is, it tests the null hypothesis that the effects of race and dose on change in blood flow are additive. This test, which has five degrees of freedom, gives $P < 0.00005$, which allows us to reject the null hypothesis with overwhelming statistical significance.



9. GEE Analyses with Logistic or Poisson Models

GEE analyses can be applied to any generalized linear model with repeated measures data.

For logistic regression we use the logit link function and a binomial random component.

For Poisson regression we use the logarithmic link function and a Poisson random component.

In Stata, the syntax for specifying these terms is the same as in the *glm* command.

For logistic regression, we use the *link(logit)* and *family(binomial)* options to specify the link function and random component, respectively.

For Poisson regression, these options are *link(log)* and *family(poisson)*.

10. What we have covered

- Analysis of variance with multiple observations per patient
  - These analyses are complicated by the fact that multiple observations on the same patient are correlated with each other
- Response-feature approach to mixed effects analysis of variance
  - Reduce multiple response measures on each patient to a single statistic that captures the most biologically important aspect of the response: the *statsby* command
  - Perform a fixed effects analysis on this response feature
  - Using a regression slope as a response feature
  - Using an area under the curve as a response feature
- Generalized estimating equations (GEE) approach to mixed effects analysis of variance: the *xtgee* command
  - GEE analysis with logistic or Poisson models

Cited Reference

Lang CC, Stein CM, Brown RM, Deegan R, Nelson R, He HB, Wood M, Wood AJ. Attenuation of isoproterenol-mediated vasodilatation in blacks. N Engl J Med 1995;333:155-60.

For additional references on these notes see.

Dupont WD. *Statistical Modeling for Biomedical Researchers: A Simple Introduction to the Analysis of Complex Data. 2nd ed.* Cambridge, U.K.: Cambridge University Press; 2009.