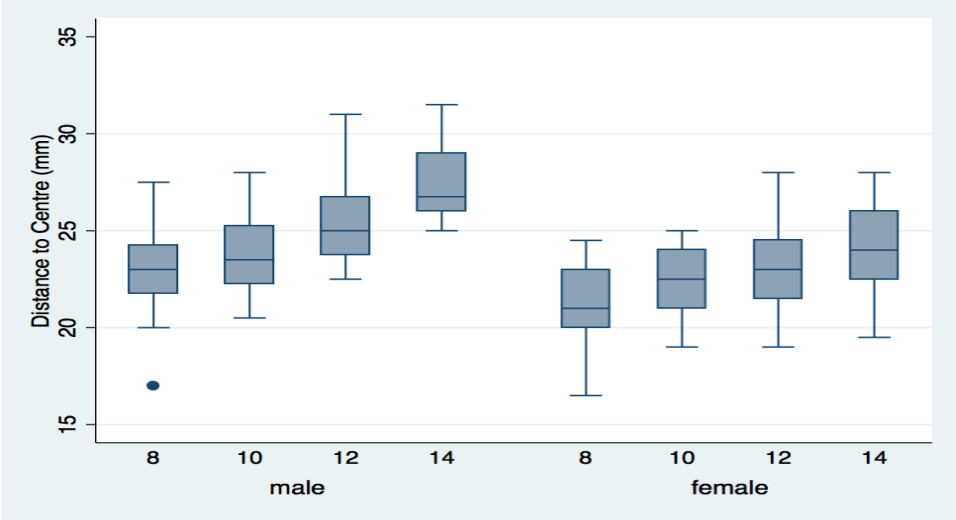
Models In Epidemiology And Biostatistics
Gordon Hilton Fick

Classic Methods With Repeated Measures

Let us now consider a growth study. A measure of growth [based the distance [in mm] between the centre of 2 teeth] was considered for a group of boys and girls. There were 16 boys and 11 girls in the study. This measure [dist] was recorded for each child at ages 8, 10, 12 and 14 years of age. The primary objective was to determine if the boy growth "curves" were different from the girl growth curves. The data is in pott.dta. One might be tempted to use the graph and analyses below.

```
graph box dist,over(age) over(sex)
```




```
table age sex,c(mean dist sd dist)
------------------------------
           |       Gender
 Age (yr)  |     male    female
-----------+------------------
         8 |   22.875  21.18182
           | 2.452889  2.124532
           |
        10 |  23.8125  22.22727
           | 2.136001  1.902152
           |
        12 | 25.71875  23.09091
           | 2.651847   2.36451
           |
        14 | 27.46875  24.09091
           | 2.085416  2.437398
------------------------------
. regr dist age sex as

      Source |       SS       df       MS              Number of obs =     108
-------------+------------------------------           F(  3,   104) =   25.39
       Model |  387.935027     3  129.311676            Prob > F      =  0.0000
    Residual |  529.757102   104  5.09381829           R-squared     =  0.4227
-------------+------------------------------           Adj R-squared =  0.4061
       Total |   917.69213   107  8.57656196           Root MSE      =  2.2569

------------------------------------------------------------------------------
        dist |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |    .784375   .1261673     6.22   0.000     .5341806    1.034569
         sex |   1.032102   2.218797     0.47   0.643    -3.367855     5.43206
          as |  -.3048295   .1976661    -1.54   0.126    -.6968089    .0871498
       _cons |   16.34062   1.416224    11.54   0.000      13.5322    19.14905
------------------------------------------------------------------------------
```

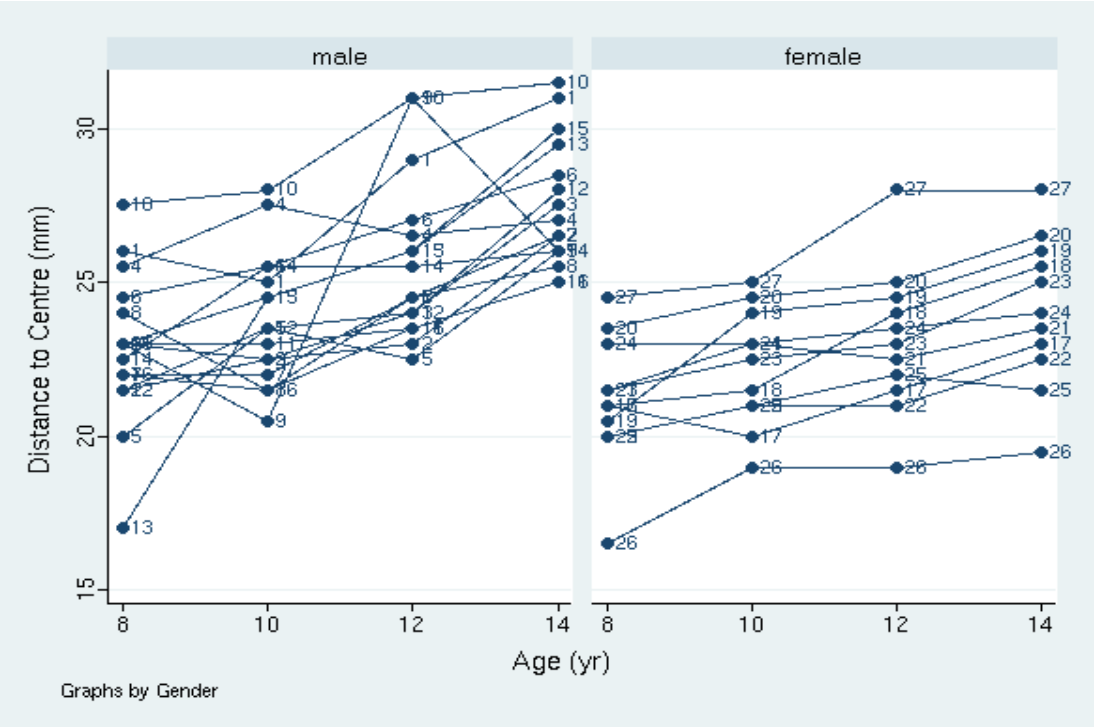
One can see that the averages [or the medians] grow faster for the boys than the girls. The computed standard deviations for each age/sex group are very similar. The Root MSE from the regression is 2.2569 which provides an estimate of the 'assumed constant' standard deviation. The coefficient -0.3048295 provides an estimate of the difference between the mean slope for the girls and the mean slope for the boys. The standard error of this estimate [0.1976661] is based on the Root MSE.
But the Root MSE is an estimate of the cross sectional standard deviation. Cross sectionally, each computed standard deviation 'contains' differences among the children at any given age for each gender. This is inevitable in cross sectional studies. This analysis above [the visuals, the tables and the regression analysis] is incorrect but it may not be immediately clear why it is incorrect.

This type of study can be called a longitudinal study in that each child was followed up over a course of 6 years. Each child contributes 4 measurements once every 2 years. Clearly the measurements from the same child are not independent while any 2 measurements from 2 different children are independent. Further, any comparison between any 2 measurements from the same child can be viewed as a part of the intra-child variability while any comparison between any 2 measurements from 2 different children can be viewed as a part of the inter-child variability.

This type of study can also be called a split unit design in that the comparison between male and female children must be a comparison between children [the whole unit comparisons] while a comparison between 2 different years for a given child is a comparison within children [the split unit comparisons] It can also be noted that such studies are sometimes also called repeated measures studies in that the measure on a given child is repeated 4 times here.

Here is a more appropriate visualization of the data:

```
sort subject age
scatter dist age,connect(ascending) by(sex) mlabel(subject)
```



The lines are drawn from two year period to two year period for a given child to aid the eye in following the course of measurements for a given child. There were no measurements taken between each 2 year period.

For the moment, let us assume that the growth curves are lines. Rather than considering distance as the primary outcome, perhaps we should consider a different outcome. If we were to determine the rate of change of distance per year for each child, then we would have one value [one slope] for each child. Then we could just compare expected slope conditional on gender. If normality of slopes is plausible then a simple t test for the difference in mean slope (and confidence interval for this difference) would do the trick. This is called a response feature analysis. It is certainly the simplest approach and has much to recommend it. It is easy to explain and interpret. In many investigations, we get a clear and, perhaps, definitive result. In a way, we have changed our definition of the primary outcome and formulated our objectives and hypotheses in terms of this [computed] outcome.

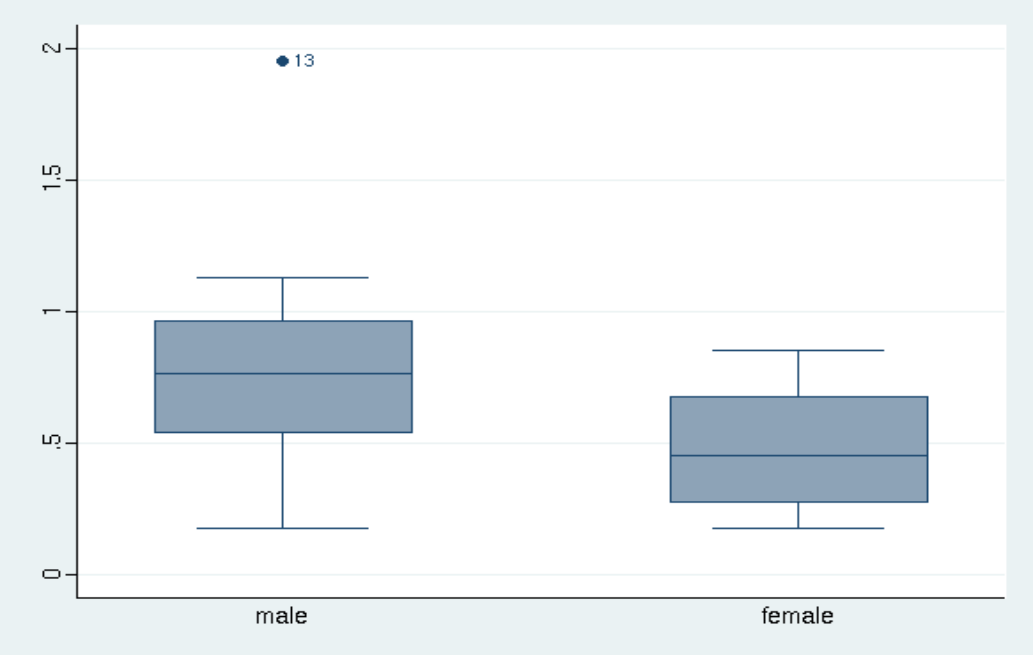
```
gen blist=.
quietly forval num = 1/27 {
regr dist age if subject==`num'
replace blist=_b[age] if subject==`num'
}
by subject: replace blist=. if _n!=1
ttest blist,by(sex)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    male |      16     .784375    .1015729    .4062917    .5678775    1.000873
  female |      11    .4795455     .066214    .2196071    .3320114    .6270795
---------+--------------------------------------------------------------------
combined |      27    .6601852    .0712533    .3702429     .513722    .8066484
---------+--------------------------------------------------------------------
    diff |            .3048295    .1347353                .0273369    .5823222
------------------------------------------------------------------------------
    diff = mean(male) - mean(female)                              t =   2.2624
Ho: diff = 0                                     degrees of freedom =       25

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9837         Pr(|T| > |t|) = 0.0326          Pr(T > t) = 0.0163
```

This bit of Stata code and output provides evidence (p=0.0326) that the slopes are steeper for the boys than the slopes for the girls. So growth appears to be faster for the boys compared to the girls. The boxplot of slopes reveals a troublesome lad.

```
graph box blist, over(sex) marker(1,mlabel(subject))
```



Analysis of the slopes could consider other 'between subjects' characteristics using regression. For example, to assess whether growth depends on initial distance [at age 8] as either a modifier or a confounder:

```
gen sd = sex*dist
regr blist sex dist sd

      Source |       SS       df       MS              Number of obs =      27
-------------+------------------------------           F(  3,    23) =    5.48
       Model |  1.48603663     3  .495345542           Prob > F      =  0.0054
    Residual |  2.07803761    23  .090349461           R-squared     =  0.4169
-------------+------------------------------           Adj R-squared =  0.3409
       Total |  3.56407424    26  .137079778           Root MSE      =  .30058

------------------------------------------------------------------------------
       blist |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |  -2.681152   1.198251    -2.24   0.035    -5.159922    -.202382
        dist |  -.0986842   .0316402    -3.12   0.005    -.1641369   -.0332315
          sd |   .1042985   .0547978     1.90   0.070    -.0090594    .2176565
       _cons |   3.041776     .72766     4.18   0.000     1.536497    4.547056
------------------------------------------------------------------------------

regr blist sex dist

      Source |       SS       df       MS              Number of obs =      27
-------------+------------------------------           F(  2,    24) =    5.78
       Model |  1.15872976     2  .579364882           Prob > F      =  0.0089
    Residual |  2.40534447    24  .100222686           R-squared     =  0.3251
-------------+------------------------------           Adj R-squared =  0.2689
       Total |  3.56407424    26  .137079778           Root MSE      =  .31658

------------------------------------------------------------------------------
       blist |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |  -.4130445   .1322775    -3.12   0.005    -.6860518   -.1400372
        dist |  -.0639122   .0272079    -2.35   0.027    -.1200666   -.0077578
       _cons |   2.246367   .6273935     3.58   0.002     .9514901    3.541243
------------------------------------------------------------------------------
```

So there is no evidence that age8 distance modifies [p=0.070] but, perhaps, some indication that age8 confounds [-0.413 vs -0.305]

There are, inevitably, many aspects of a response curve that could have been selected for study. Perhaps, it would be best if your literature provided some guides to such features [before analysis begins].

Another candidate would be change in distance [final – baseline]. Change scores are widely seen in health research. One way to carry out this analysis involves 'reshaping' the data:

```
by subject:gen ct=_n
drop blist sd
reshape wide dist age alin aquad acub gal gaq gac as,i(subject) j(ct)
gen cs=dist4-dist1
graph box cs,over(sex) marker(1,mlabel(subject))
ttest cs,by(sex)
regr cs sex
regr cs sex dist1
regr dist4 sex dist1
```

The last two analyses provide the same β_1 .

$$y_4 = \beta_0 + \beta_1 * S + \beta_2 * y_1 + \epsilon$$

$$y_4 - y_1 = \beta_0 + \beta_1 * S + [\beta_2 - 1] * y_1 + \epsilon$$

The analysis of variance provides a definitive 'classical' assessment here since each child was measured at all 4 time points. The time points are [assumed] the same for each child and there are no missing values.
First separate out whole unit [between subject] differences from split unit [within subject] differences, by:

```
anova dist subject
```

```
                          Number of obs =     108     R-squared     =  0.5649
                          Root MSE      = 2.22031     Adj R-squared =  0.4252

                  Source |  Partial SS    df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |   518.37963    26  19.9376781       4.04     0.0000
                         |
                 subject |   518.37963    26  19.9376781       4.04     0.0000
                         |
                Residual |    399.3125    81  4.92978395
              -----------+----------------------------------------------------
                   Total |   917.69213   107  8.57656196
```

Then we identify the age, gender and interaction components:

```
anova dist c.age sex c.age#sex,sequential
```

```
                          Number of obs =     108     R-squared     =  0.4227
                          Root MSE      = 2.25695     Adj R-squared =  0.4061

                  Source |    Seq. SS     df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |  387.935027     3  129.311676      25.39     0.0000
                         |
                     age |  235.356019     1  235.356019      46.20     0.0000
                     sex |  140.464857     1  140.464857      27.58     0.0000
                 age*sex |  12.1141519     1  12.1141519       2.38     0.1261
                         |
                Residual |  529.757102   104  5.09381829
              -----------+----------------------------------------------------
                   Total |   917.69213   107  8.57656196
```

The age and interaction comparisons are within subject comparisons but the gender comparison is a between subjects comparison. Comparisons that are between subjects but within gender provide a between subjects error sum of squares [`518.37963 - 140.464857 = 377.91477`] with 26-1 = 25 degrees of freedom while comparisons that are within subjects but not a part of age or interaction comparisons provide a within subjects error sum of squares [`399.3125 - 235.356019 - 12.1141519 = 151.84233`] with 81-1-1 =79 degrees of freedom. It is instructive to carry out this part of the analysis of variance 'by hand'. Such calculation makes it clearer what is going on in this rather complicated situation. Nevertheless, Stata will do the analysis in one step. Figuring out the syntax of the command takes just about as long as doing the 2 simple analyses and then doing the hand calculation :-)

Here is the analysis of variance separating out the linear components of age and age*sex.

```
anova dist sex / subject|sex c.age c.age#sex,sequential

                          Number of obs =     108     R-squared     =  0.8345
                          Root MSE      = 1.38638     Adj R-squared =  0.7759

      Source |    Seq. SS     df       MS           F     Prob > F
-------------+------------------------------------------------------
       Model |    765.8498    28  27.3517786      14.23     0.0000
             |
         sex |  140.464857     1  140.464857       9.29     0.0054
 subject|sex |  377.914773    25  15.1165909
-------------+------------------------------------------------------
         age |  235.356019     1  235.356019     122.45     0.0000
     sex#age |  12.1141519     1  12.1141519       6.30     0.0141
             |
    Residual |   151.84233    79   1.9220548
-------------+------------------------------------------------------
       Total |   917.69213   107  8.57656196
```

This approach provides a very similar statement to the response feature analysis based on mean slope comparison. (p=0.0141)

Split unit studies are often a strong choice at the design stage in part because the key comparison (in this case, the age*sex interaction term) is estimated with the higher precision (being a within subject comparison). The 'main effect' of gender is estimated with lower precision (since it is a between subject comparison) but such a comparison would rarely be of interest in any case.

The above analyses have assumed that the distance/age relationship is linear. An assessment of this assumption proceeds as follows:

```
gen a2 = age*age
gen a3 = a2*age
anova dist c.age c.a2 c.a3 sex c.age#sex c.a2#sex c.a3#sex,sequential

                          Number of obs =     108     R-squared     =  0.4268
                          Root MSE      = 2.29356     Adj R-squared =  0.3867

      Source |    Seq. SS     df       MS           F     Prob > F
-------------+------------------------------------------------------
       Model |  391.649516     7  55.9499309      10.64     0.0000
             |
         age |  235.356019     1  235.356019      44.74     0.0000
          a2 |  1.44675926     1  1.44675926       0.28     0.6011
          a3 |  .389351852     1  .389351852       0.07     0.7861
         sex |  140.464857     1  140.464857      26.70     0.0000
     sex#age |  12.1141519     1  12.1141519       2.30     0.1323
      sex#a2 |  1.19954756     1  1.19954756       0.23     0.6340
      sex#a3 |  .678829966     1  .678829966       0.13     0.7202
             |
    Residual |  526.042614   100  5.26042614
-------------+------------------------------------------------------
       Total |   917.69213   107  8.57656196
```

Then you can do the subtractions as before.
Tests for nonlinearity can be done with single degree of freedom tests or by pooling the quadratic and cubic components for 2 degree of freedom tests. You can separate all the pieces out and possibly pool:

```
anova dist sex / subject|sex c.age c.a2 c.a3 c.age#sex c.a2#sex c.a3#sex,sequential

                          Number of obs =     108     R-squared     =  0.8386
                          Root MSE      = 1.40536     Adj R-squared =  0.7697

                Source |    Seq. SS     df       MS           F     Prob > F
            -----------+----------------------------------------------------
                 Model |  769.564289    32   24.048884      12.18     0.0000
                       |
                   sex |  140.464857     1  140.464857       9.29     0.0054
            subject|sex |  377.914773    25  15.1165909
            -----------+----------------------------------------------------
                   age |  235.356019     1  235.356019     119.17     0.0000
                    a2 |  1.44675926     1  1.44675926       0.73     0.3948
                    a3 |  .389351852     1  .389351852       0.20     0.6583
               sex#age |  12.1141519     1  12.1141519       6.13     0.0155
                sex#a2 |  1.19954756     1  1.19954756       0.61     0.4382
                sex#a3 |  .678829966     1  .678829966       0.34     0.5595
                       |
              Residual |  148.127841    75  1.97503788
            -----------+----------------------------------------------------
                 Total |   917.69213   107  8.57656196
```

Or you can 'ask' for an incomplete subdivision [below] and subtract off the linear parts to get the 2 degree of freedom components.

```
anova dist sex / subject|sex age age#sex,sequential

                          Number of obs =     108     R-squared     =  0.8386
                          Root MSE      = 1.40536     Adj R-squared =  0.7697

                Source |    Seq. SS     df       MS           F     Prob > F
            -----------+----------------------------------------------------
                 Model |  769.564289    32   24.048884      12.18     0.0000
                       |
                   sex |  140.464857     1  140.464857       9.29     0.0054
            subject|sex |  377.914773    25  15.1165909
            -----------+----------------------------------------------------
                   age |   237.19213     3  79.0640432      40.03     0.0000
               age#sex |  13.9925295     3  4.66417649       2.36     0.0781
                       |
              Residual |  148.127841    75  1.97503788
            -----------+----------------------------------------------------
                 Total |   917.69213   107  8.57656196
```

We can see that `13.9925295 - 12.1141519 = 1.19954756 +.678829966 = 1.8783775`
The test for nonlinearity is based on F = `1.8783775/2 / 1.97503788 = 0.47552949`
which would be compared with an $F_{2,75}$ distribution.

```
disp 1 - F(2,75,0.47552949)
.62341697

disp invF(2,75,0.95)
3.1186421
```

We can construct a regression model that removes all the between subject differences and then includes age and age*sex since they are both within subject comparisons:

$E(y)=\beta_0+\sum_{j=2}^{27}\beta_j\delta_j+\beta_{28}\text{age}+\beta_{29}\text{age*sex}$ where δ_j is the indicator for the jth subject. The first 16 indicators $\delta_1\ \delta_2 \ldots \delta_{16}$ are for the boys and $\delta_{17}\ \delta_{18} \ldots \delta_{27}$ are for the girls.

For example:
For id==3 [boy] : $E(y)=\beta_0+\beta_3+\beta_{28}\,\text{age}$
For id==18 [girl]: $E(y)=\beta_0+\beta_{18}+(\beta_{28}+\beta_{29})\,\text{age}$

...so that β_{29} is the difference between the mean girl slope and the mean boy slope

The [rather unwieldy] results look like:

```
gen as=age*sex

regr dist i.subject age as
i.subject         _Isubject_1-27      (naturally coded; _Isubject_1 omitted)

      Source |       SS       df       MS              Number of obs =     108
-------------+------------------------------           F( 28,    79) =   14.23
       Model |    765.8498    28  27.3517786           Prob > F      =  0.0000
    Residual |   151.84233    79   1.9220548           R-squared     =  0.8345
-------------+------------------------------           Adj R-squared =  0.7759
       Total |   917.69213   107  8.57656196           Root MSE      =  1.3864

------------------------------------------------------------------------------
        dist |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 _Isubject_2 |     -4.375   .9803201    -4.46   0.000    -6.326278   -2.423722
 _Isubject_3 |       -3.5   .9803201    -3.57   0.001    -5.451278   -1.548722
 _Isubject_4 |     -1.125   .9803201    -1.15   0.255    -3.076278    .8262783

[snip]

_Isubject_25 |  -3.271875   1.656784    -1.97   0.052    -6.569622    .0258718
_Isubject_26 |  -5.896875   1.656784    -3.56   0.001    -9.194622   -2.599128
_Isubject_27 |   1.978125   1.656784     1.19   0.236    -1.319622    5.275872
         age |    .784375   .0775011    10.12   0.000     .6301129    .9386371
          as |  -.3048295   .1214209    -2.51   0.014    -.5465118   -.0631473
       _cons |   19.12187   1.098768    17.40   0.000     16.93483    21.30892
------------------------------------------------------------------------------
```

...and so we get the same p-value as the analysis of variance. We get the same estimated difference between the 2 slopes as we got in the response feature analysis last class only now we get, arguably, the 'correct' standard error for this estimate.

Notice here that the sex comparison is not listed as it is a part of the between subject comparisons. In this study, we have no interest in this comparison since we clearly detected an interaction. So the absence of the sex comparison in the regression analysis is of no consequence.

The assessment that includes possible nonlinearity could be done with the actual polynomials or with orthogonal polynomials.

```
regr dist i.subject age a2 a3 c.age#sex c.a2#sex c.a3#sex

      Source |       SS       df       MS              Number of obs =     108
-------------+------------------------------           F( 32,    75) =   12.18
       Model |  769.564289    32   24.048884           Prob > F      =  0.0000
    Residual |  148.127841    75  1.97503788           R-squared     =  0.8386
-------------+------------------------------           Adj R-squared =  0.7697
       Total |   917.69213   107  8.57656196           Root MSE      =  1.4054

------------------------------------------------------------------------------
        dist |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     subject |
           2 |     -4.375   .9937399    -4.40   0.000    -6.354631   -2.395369
             [ 3 to 26 deleted]
          27 |   -41.5483   64.15452    -0.65   0.519    -169.3506    86.25406
             |
```

```
         age |  -8.648437   11.65447    -0.74   0.460    -31.86534    14.56846
          a2 |   .8242187    1.08112     0.76   0.448    -1.329483    2.977921
          a3 |  -.0234375   .0327342    -0.72   0.476    -.0886473    .0417723
             |
   sex#c.age |
      female |   11.54238   18.25905     0.63   0.529    -24.83151    47.91627
             |
    sex#c.a2 |
      female |   -1.04581    1.69379    -0.62   0.539    -4.420013    2.328393
             |
    sex#c.a3 |
      female |   .0300663   .0512846     0.59   0.559    -.0720979    .1322305
             |
       _cons |   54.09375   40.94997     1.32   0.191    -27.48277    135.6703
------------------------------------------------------------------------------
```

Then you would remove c.a3#sex and then c.a2#sex.

```
regr dist i.subject alin aquad acub c.alin#sex c.aquad#sex c.acub#sex
```

```
      Source |       SS       df       MS              Number of obs =     108
-------------+------------------------------           F( 32,    75) =   12.18
       Model |  769.564289    32   24.048884           Prob > F      =  0.0000
    Residual |  148.127841    75  1.97503788           R-squared     =  0.8386
-------------+------------------------------           Adj R-squared =  0.7697
       Total |   917.69213   107  8.57656196           Root MSE      =  1.4054

------------------------------------------------------------------------------
        dist |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     subject |
           2 |     -4.375   .9937399    -4.40   0.000    -6.354631   -2.395369
         [3 to 26 deleted]
          27 |     -1.375   .9937399    -1.38   0.171    -3.354631    .6046313
             |
        alin |    .784375    .078562     9.98   0.000     .6278714    .9408786
       aquad |    .203125   .1756701     1.16   0.251    -.1468277    .5530777
        acub |    -.05625    .078562    -0.72   0.476    -.2127536    .1002536
             |
  sex#c.alin |
      female |  -.3048295   .1230831    -2.48   0.016    -.5500236   -.0596355
             |
 sex#c.aquad |
      female |  -.2144886   .2752221    -0.78   0.438    -.7627591    .3337819
             |
  sex#c.acub |
      female |   .0721591   .1230831     0.59   0.559    -.1730349    .3173531
             |
       _cons |      27.75   .7026802    39.49   0.000     26.35019    29.14981
------------------------------------------------------------------------------
```

With orthogonal polynomials, you can test the cubic and quadratic directly without the need to refit.

Now let us move to a look at another type of repeated measures study; the cross over study. We will see that these types of designs have 2 error terms like split unit studies. This time, we will see that for cross over studies, the factors - treatment and order are within subject comparisons while the order*treatment interaction is a between subject comparison.

Lets take a 'simple' example. A group of 12 children with asthma were randomized to one of 2 sequence groups: A: Formoterol (F) first; Salbutamol (S) second or B: Salbutamol first; Formoterol second. The outcome is peak expiratory flow (PEF). The data is in forsal.dta
[pef0 =pef for F, pef1 =pef for S; grp=0 for A grp=1 for B) Notice that, since the same number of subjects received the A order as the B order, the treatment comparison does not reflect any order differences. Our error term for testing treatment differences should not reflect either subject differences or order differences. If the data is shaped as ...

```
list pef0 pef1 grp
```

```
     +-------------------+
     | pef0   pef1   grp |
     |-------------------|
  1. |  310    270     1 |
  2. |  385    370     0 |
  3. |  400    310     0 |
  4. |  310    260     1 |
  5. |  410    380     0 |
     |-------------------|
  6. |  370    300     1 |
  7. |  410    390     1 |
  8. |  380    350     1 |
  9. |  320    290     0 |
 10. |  250    210     1 |
     |-------------------|
 11. |  330    365     0 |
 12. |  340    260     0 |
     +-------------------+
```

… one might be tempted to perform a paired t test. This method would provide a standard error that does not reflect subject differences but would not remove order differences from the standard error.

```
ttest pef1=pef0
```

```
Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    pef1 |      12    312.9167    16.63396    57.62174    276.3056    349.5278
    pef0 |      12      351.25    14.25053    49.36529    319.8848    382.6152
---------+--------------------------------------------------------------------
    diff |      12   -38.33333    9.541245    33.05184   -59.33347   -17.33319
------------------------------------------------------------------------------
     mean(diff) = mean(pef1 - pef0)                               t =  -4.0176
 Ho: mean(diff) = 0                         degrees of freedom =       11

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.0010         Pr(|T| > |t|) = 0.0020          Pr(T > t) = 0.9990
```

If we adopt a notation of y_{ijk} : the PEF for treatment i [0=F, 1=S], order j [0=first, 1=second] and group k [0=A, 1=B], we have:

A		B	
y_{000}	y_{110}	y_{011}	y_{101}

we can see that $d_{jk} = y_{1jk} - y_{0jk}$ is the treatment comparison that does not reflect subject differences. Now d_{j1} is the second period PEF – the first period PEF and
d_{j0} is the first period PEF – the second period PEF

We then get a t test for order:

```
gen diff=(pef1-pef0)/2
ttest diff,by(grp)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |       6       -17.5    9.309493    22.80351   -41.43081    6.430815
       1 |       6   -20.83333    3.515837    8.612007   -29.87108   -11.79559
---------+--------------------------------------------------------------------
combined |      12   -19.16667    4.770622    16.52592   -29.66674   -8.666597
---------+--------------------------------------------------------------------
    diff |            3.333333     9.95127               -18.83948    25.50615
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   0.3350
Ho: diff = 0                                     degrees of freedom =       10

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.6277         Pr(|T| > |t|) = 0.7446          Pr(T > t) = 0.3723
```

And, with this little trick:

```
replace diff=-diff if grp==1
(6 real changes made)
```

… we get a t test for treatments

```
ttest diff,by(grp)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |       6       -17.5    9.309493    22.80351   -41.43081    6.430815
       1 |       6    20.83333    3.515837    8.612007    11.79559    29.87108
---------+--------------------------------------------------------------------
combined |      12    1.666667    7.476816    25.90045   -14.78969    18.12303
---------+--------------------------------------------------------------------
    diff |           -38.33333     9.95127               -60.50615   -16.16052
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -3.8521
Ho: diff = 0                                     degrees of freedom =       10

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0016         Pr(|T| > |t|) = 0.0032          Pr(T > t) = 0.9984
```

Here is a test to compare the 2 groups. We will see that this the test of whether the treatment comparison depends on the order given [below]

```
gen sp=(pef0+pef1)/2

ttest sp,by(grp)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |       6    346.6667    15.56795    38.13354     306.648    386.6854
       1 |       6       317.5    25.05827    61.37996    253.0857    381.9143
---------+--------------------------------------------------------------------
combined |      12    332.0833    14.73514    51.04402    299.6515    364.5152
---------+--------------------------------------------------------------------
    diff |            29.16667    29.50047               -36.56448    94.89781
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   0.9887
Ho: diff = 0                                     degrees of freedom =       10

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.8269         Pr(|T| > |t|) = 0.3461          Pr(T > t) = 0.1731
```

The paired t-test is not a correct analysis here.

The next t test assesses whether there is an order difference. In other words,whether mean PEF for first in order is different from mean PEF for second in order (ignoring the treatment).

Then the next t test assesses whether there is a treatment difference. In other words, whether mean PEF for those on active is different from mean PEF for those on placebo (ignoring order)

Lastly, we have a t test comparing the two groups. This is sometimes called the validity test. [explanation below]
Now reshape the dataset as long and add a variable for the order [ord]

```
gen id = _n
reshape long pef, i(id) j(tr)
gen ord = tr
replace ord = 1-ord if grp==1

list id grp pef tr ord
     +---------------------------+
     | id   grp   pef   tr   ord |
     |---------------------------|
  1. |  1     1   270    1     1 |
  2. |  1     1   310    0     0 |
  3. |  2     0   370    1     0 |
  4. |  2     0   385    0     1 |
  5. |  3     0   400    0     1 |
     |---------------------------|
  6. |  3     0   310    1     0 |
  7. |  4     1   310    0     0 |
  8. |  4     1   260    1     1 |
  9. |  5     0   410    0     1 |
 10. |  5     0   380    1     0 |
     |---------------------------|
 11. |  6     1   370    0     0 |
 12. |  6     1   300    1     1 |
 13. |  7     1   390    1     1 |
 14. |  7     1   410    0     0 |
 15. |  8     1   380    0     0 |
     |---------------------------|
 16. |  8     1   350    1     1 |
 17. |  9     0   290    1     0 |
 18. |  9     0   320    0     1 |
 19. | 10     1   250    0     0 |
 20. | 10     1   210    1     1 |
     |---------------------------|
 21. | 11     0   365    1     0 |
 22. | 11     0   330    0     1 |
 23. | 12     0   340    0     1 |
 24. | 12     0   260    1     0 |
     +---------------------------+
```

The little table below may help to see the issues now.

	Group A	Group B

comparison	***F first***	***S second***	***F second***	***S first***
treatment	-1	1	-1	1
order	-1	1	1	-1
treat X order	1	1	-1	-1

```
anova pef grp /id|grp tr ord
```

```
                           Number of obs =      24     R-squared     =  0.9176
                           Root MSE      = 24.3755     Adj R-squared =  0.8106

                  Source |  Partial SS    df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |  66204.1667    13  5092.62821       8.57     0.0009
                         |
                     grp |  5104.16667     1  5104.16667       0.98     0.3461
                  id|grp |  52216.6667    10  5221.66667
              -----------+----------------------------------------------------
                      tr |  8816.66667     1  8816.66667      14.84     0.0032
                     ord |  66.6666667     1  66.6666667       0.11     0.7446
                         |
                Residual |  5941.66667    10  594.166667
              -----------+----------------------------------------------------
                   Total |  72145.8333    23  3136.77536
```

Notice that the F test here is the same as the t test shown above. This analysis of variance looks much like the one we saw with the split unit studies only now both the treatment comparison and the order comparison are within subject comparisons. We can see that the order*treatment interaction is in fact identical to the the comparison between Group A and Group B and is indeed a between subject comparison.

At first blush, this can be quite concerning. If the comparison between the two treatments in the second phase is different from the comparison between the two treatments in the first phase, then the study is in trouble. But this comparison is precisely the order*treatment interaction and this comparison is a part of the between subject comparisons and is not estimated as precisely as either the treatment comparison or the order comparison.

In designing such studies, it is crucial then to configure the study in such ways as to avoid this trouble. A lengthy wash out period may be required between the first and second phases [this may be difficult to reconcile if the asthma patients cannot be 'on placebo' for long!] There is considerable debate in the biostatistics/epidemiology literature as to whether the investigator should test for order*treatment interaction. Perhaps the absence of such an interaction should be a part of the assumptions for the use of such a design. Certainly a non-significant test for this interaction is problematic given that such a test may well have very low power and the study may not have been designed in such a way that detection such an interaction was a priority, in any case.

A regression analysis gives [almost] the same information as the analysis of variance:

```
regr pef i.id tr ord
```

```
      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F( 13,    10) =    8.57
       Model |  66204.1667    13  5092.62821           Prob > F      =  0.0009
    Residual |  5941.66667    10  594.166667           R-squared     =  0.9176
-------------+------------------------------           Adj R-squared =  0.8106
       Total |  72145.8333    23  3136.77536           Root MSE      =  24.376

------------------------------------------------------------------------------
         pef |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          id |
           2 |       87.5   24.37553     3.59   0.005     33.18793    141.8121
      [3 to 11 deleted]
          12 |         10   24.37553     0.41   0.690    -44.31207    64.31207
             |
          tr |  -38.33333    9.95127    -3.85   0.003    -60.50615   -16.16052
         ord |   3.333333    9.95127     0.33   0.745    -18.83948    25.50615
       _cons |      307.5   18.61712    16.52   0.000     266.0185    348.9815
------------------------------------------------------------------------------
```

```
disp (38.3333/9.95127)^2
14.838684
```

Notice that the square of the t statistic is the same as the F from the analysis of variance. So both methods yield the same main result. This regression approach does not provide for the order X treatment interaction as this comparison is buried within the comparisons between the subjects.

Notice, also, that the residual sum of squares here is the same as the within subjects sum of squares from the analysis of variance. The regression analysis gives us identical results to the two t tests done earlier.

Now lets consider a cross over study with an ordinal outcome. This example is taken from Rosner (2011 7th Ed p686). The data is in pain.dta. Participants with tennis elbow were randomized to either Group A: active then placebo or Group B: placebo then active. The outcome is from a pain scale: 1-worse 2-same 3-slight improvement 4-moderate improvement 5- mostly improved 6- completely improved.
Rosner gives an analysis approach like the asthma study treating the ordinal outcomes as though they are interval outcomes:

```
reshape wide ove imp ord, i(id) j(tr)
gen diff=(ove1-ove0)/2
ttest diff,by(grp)
replace diff=-diff if grp==2
ttest diff,by(grp)
gen sump=(ove0+ove1)/2
ttest sump,by(grp)
```

The data from each group looks like:

```
        Group A: active first/placebo second
           |              active
  placebo | 1    2     3     4     5     6
----------+---------------------------------
        6 |      1                    4     1
        5 |      3     2      4      3
        4 |            3      1      1
        3 |      1            2      3
        2 |      2     3      2      4
        1 |      1            1
-------------------------------------------

        Group B: placebo first/active second
           |              active
   placebo| 1    2     3     4     5     6
----------+---------------------------------
        6 |
        5 |            1            3     3
        4 |                         2     1
        3 |      1     1      3     5
        2 |      6     4      3     5     3
        1 |                   1
-------------------------------------------
```

You might want to consider a series of proportional odds models:

```
ologit ove ord grp tr
ologit ove ord tr
ologit ove tr
```

...except these are not correct either.

Lets take a look at these 2 tables above. With group A, we can see that 18 participants did 'better' with placebo while 17 participants did better with active (7 participants said change from baseline was the 'same'). With group B, we can see that 2 participants did better with placebo while 30 participants did better with active (10 participants said change from baseline was the same). There was apparently a 2 week 'wash-out' between the 2 periods.

For each group, a simple sign test [or maybe a signed rank test] delivers the obvious. [try them] For those assigned to Group A, there is no difference in the scale between active and placebo while for Group B, there is a difference [active appears to be superior to placebo]
Alternately, suppose we consider the outcome [imp] given by score > 3 (i.e. moderately, mostly or completely = 1; worse, no change or slight =0). To use mcc, we interpret 'case' as active, 'control' as 'placebo'; 'exposed' as improve [imp=1] 'not exposed' as not improve [imp=0]. The groups are coded A: grp=1 ; B:grp=2

```
mcc imp1 imp0 if grp==1

                | Controls               |
Cases           |   Exposed   Unexposed  |      Total
----------------+------------------------+-----------
        Exposed |        14          12  |         26
      Unexposed |         9           7  |         16
----------------+------------------------+-----------
          Total |        23          19  |         42

McNemar's chi2(1) =      0.43    Prob > chi2 = 0.5127
Exact McNemar significance probability       = 0.6636

Proportion with factor
        Cases       .6190476
        Controls     .547619     [95% Conf. Interval]
                    ---------    --------------------
        difference  .0714286     -.1651367   .3079938
        ratio       1.130435      .7829503   1.632138
        rel. diff.  .1578947      -.275903   .5916925

        odds ratio  1.333333      .5156253   3.583017   (exact)

mcc imp1 imp0 if grp==2

                | Controls               |
Cases           |   Exposed   Unexposed  |      Total
----------------+------------------------+-----------
        Exposed |         9          20  |         29
      Unexposed |         1          12  |         13
----------------+------------------------+-----------
          Total |        10          32  |         42

McNemar's chi2(1) =     17.19    Prob > chi2 = 0.0000
Exact McNemar significance probability       = 0.0000

Proportion with factor
        Cases       .6904762
        Controls    .2380952     [95% Conf. Interval]
                    ---------    --------------------
        difference   .452381      .2642126   .6405493
        ratio            2.9       1.71136   4.914221
        rel. diff.    .59375      .4148524   .7726476

        odds ratio        20      3.198859   828.9558   (exact)
```

Again, we get the same message. Here the odds ratio is the odds of improvement for those receiving active divided by the odds of improvement for those receiving placebo. For group A, the estimated odds ratio is 1.33 [0.51 3.59] while for Group B, the estimated odds ratio is 20 [3.19, 828.96]
We have an indication that validity is questionable. The 2 odds ratio estimates are very different. WE do not have a test of significance [yet].

In situations like this, it is often argued that only the first period data can be used in which case a 6 by 2 table can be determined based on placebo first versus active first.

```
tab tr ove if ord==0,exact

           |                                ove
        tr |         1          2          3          4          5          6 |     Total
-----------+---------------------------------------------------------------------+----------
         0 |         1         21         10          3          7          0 |        42
         1 |         0          8          8         10         15          1 |        42
-----------+---------------------------------------------------------------------+----------
     Total |         1         29         18         13         22          1 |        84

           Fisher's exact =                 0.005

ologit ove tr if ord==0
```

```
tab tr imp if ord==0,exact

           |          imp
        tr |         0          1 |     Total
-----------+----------------------+----------
         0 |        32         10 |        42
         1 |        16         26 |        42
-----------+----------------------+----------
     Total |        48         36 |        84

           Fisher's exact =                 0.001
   1-sided Fisher's exact =                 0.000
```

The first table using the actual ordinal outcome while the second table uses imp [whether or not score >3]. So we have a salvage job in this instance.

If we consider conditional logistic regression, we then might consider tr (0=placebo; 1=active) ord (0=first 1; 1=second) and grp (1=A; 2=B)

```
clogit imp tr ord grp,group(id) or
note: multiple positive outcomes within groups encountered.
note: 42 groups (84 obs) dropped due to all positive or
      all negative outcomes.
note: grp omitted due to no within-group variance.

Conditional (fixed-effects) logistic regression   Number of obs   =         84
                                                  LR chi2(2)      =      21.50
                                                  Prob > chi2     =     0.0000
Log likelihood = -18.361396                       Pseudo R2       =     0.3693

------------------------------------------------------------------------------
         imp | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tr |   5.163978   2.880329     2.94   0.003      1.73064    15.40855
         ord |   3.872983   2.160247     2.43   0.015      1.29798    11.55642
------------------------------------------------------------------------------
```

Oh, Oh... grp has been omitted and grp is the order*treatment interaction. But then we must remember that grp is a between subject comparison. Like linear regression, clogit is unable to assess such a comparison [in this instance because the conditioning process is analogous to the removal of between subject comparisons in an analysis of variance]

Accordingly, this model and the corresponding fit is suspect [indeed, it is surely discredited]. This is a decent example of a situation is which you cannot ignore a note: message. We must rethink our process.
Alas, all we can do is reproduce our 'classical' analysis.

```
clogit imp tr if grp==1,group(id) or
note: multiple positive outcomes within groups encountered.
note: 21 groups (42 obs) dropped due to all positive or
      all negative outcomes.

Conditional (fixed-effects) logistic regression   Number of obs   =         42
                                                  LR chi2(1)      =       0.43
                                                  Prob > chi2     =     0.5120
Log likelihood =  -14.34107                       Pseudo R2       =     0.0148

------------------------------------------------------------------------------
         imp | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tr |   1.333333   .5879447     0.65   0.514      .561816    3.164341
------------------------------------------------------------------------------
```

```
clogit imp tr if grp==2,group(id) or
note: multiple positive outcomes within groups encountered.
note: 21 groups (42 obs) dropped due to all positive or
      all negative outcomes.

Conditional (fixed-effects) logistic regression   Number of obs   =         42
                                                  LR chi2(1)      =      21.07
                                                  Prob > chi2     =     0.0000
Log likelihood = -4.0203257                       Pseudo R2       =     0.7238

------------------------------------------------------------------------------
         imp | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tr |         20    20.4939     2.92   0.003     2.684157    149.0226
------------------------------------------------------------------------------



cci 26 10 16 32,exact
                                                        Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+------------------------
           Cases |        26          10  |         36       0.7222
        Controls |        16          32  |         48       0.3333
-----------------+------------------------+------------------------
           Total |        42          42  |         84       0.5000
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |               5.2       |    1.844848    15.02845 (exact)
 Attr. frac. ex. |          .8076923       |      .45795    .9334595 (exact)
 Attr. frac. pop |          .5833333       |
                 +-------------------------------------------------
                                 1-sided Fisher's exact P = 0.0004
                                 2-sided Fisher's exact P = 0.0008

logit imp tr if ord==0,or

Logistic regression                               Number of obs   =         84
                                                  LR chi2(1)      =      12.80
                                                  Prob > chi2     =     0.0003
Log likelihood = -50.962919                       Pseudo R2       =     0.1116

------------------------------------------------------------------------------
         imp | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tr |        5.2   2.505793     3.42   0.001     2.022198    13.37159
------------------------------------------------------------------------------
```