Profile Analysis

2014 Boston Stata Users Group

Phil Ender

UCLA Statistical Consulting Group

July 31, 2014

What is Profile Analysis?

Profile analysis is a multivariate technique for analyzing the shape (profile) of variables across groups.

Profile analysis is a "true" multivariate approach which uses separate correlated response variables. The data are arranged in wide form. The response variable scales should be commensurate.

We will use **manova** and **manovatest** to perform profile analysis.

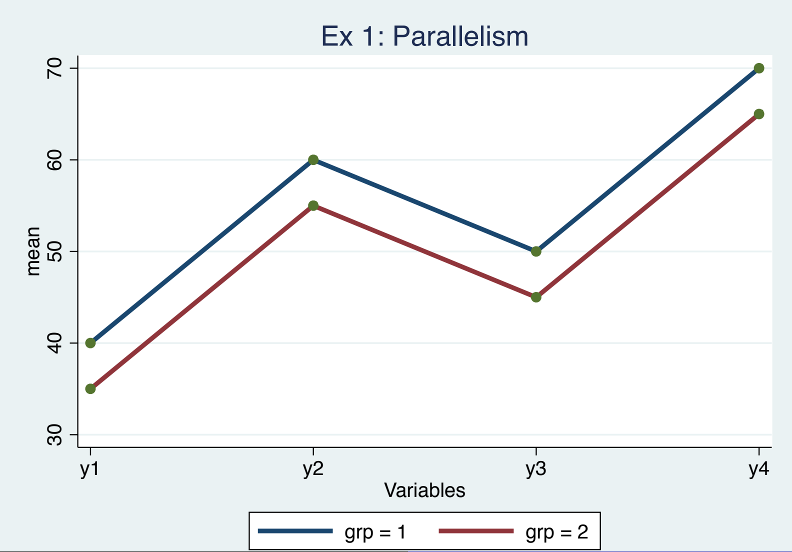
The Three Parts to Profile Analysis

- Test of Parallelism
- Test of Levels (Separation)
- Test of Flatness

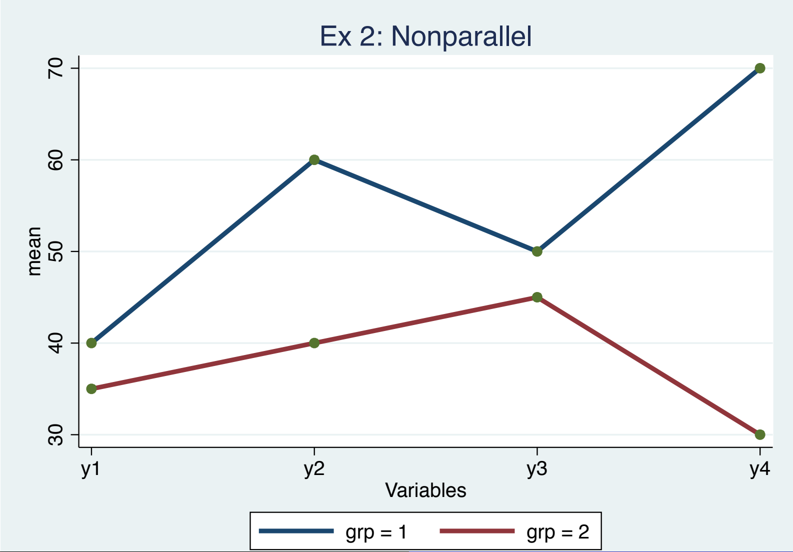
1. Test of Parallelism

Tests that each of the segments of the profiles are pairwise parallel.

Example of Parallel Profiles



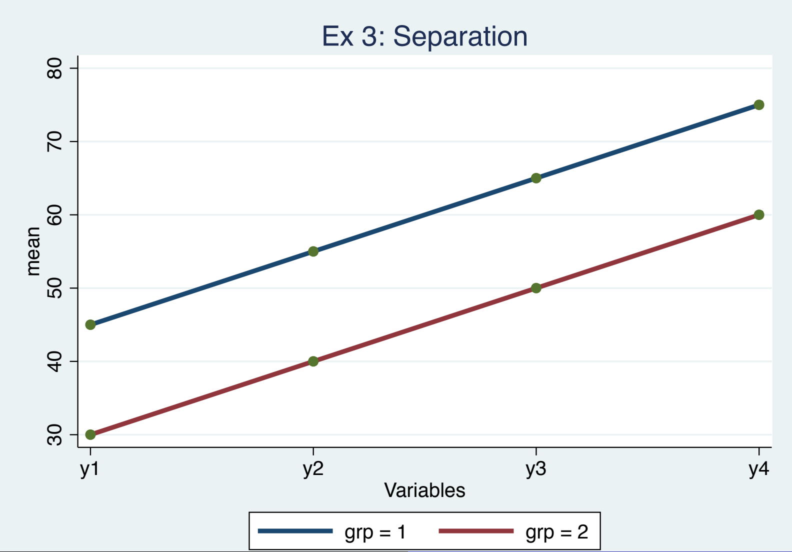
Example of Nonparallel Profiles



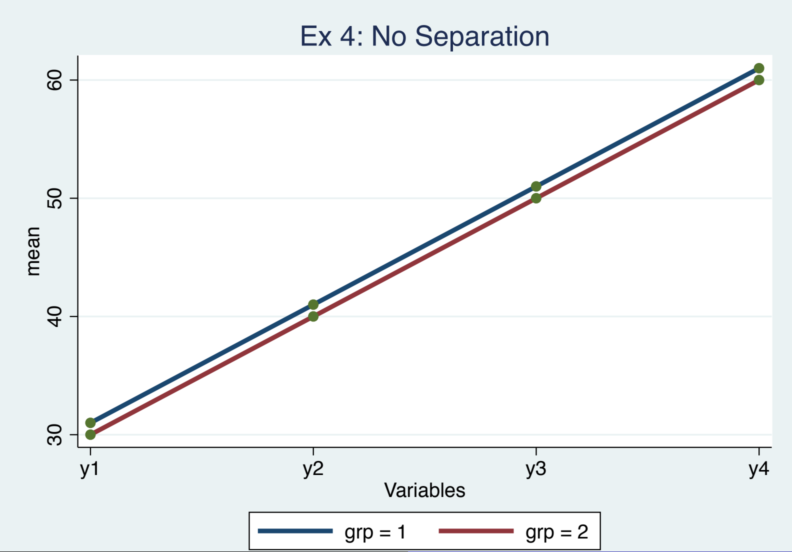
2. Test of Levels (Separation)

If profiles are pairwise parallel then test whether the profiles of the groups are separated.

Example of Separation



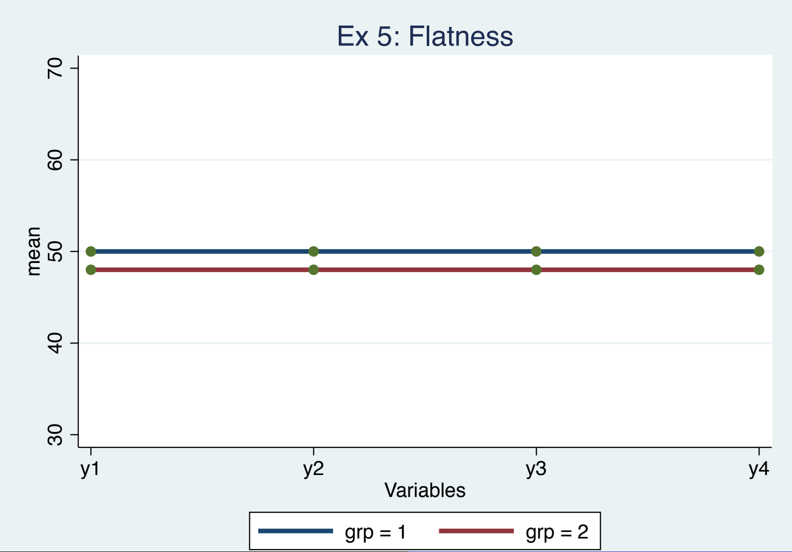
Example of No Separation



3. Test of Flatness

If profiles are parallel and not separated then test whether the profiles are flat, that is, the levels are the same across variables.

Example of Flatness



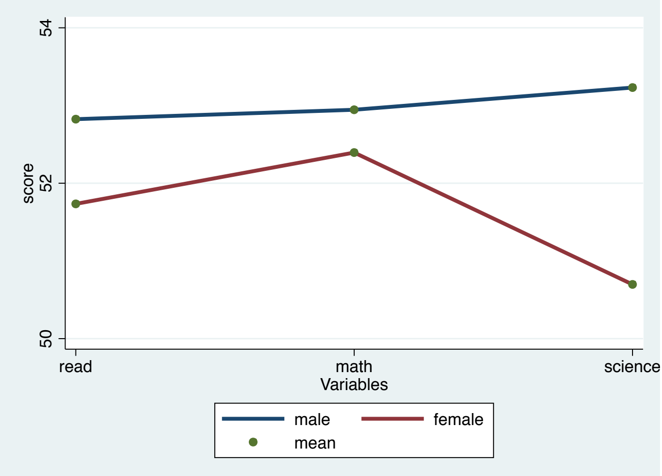
-profileplot- command

The command profileplot.ado is a user written convenience command that plots profiles for multiple groups.

```
profileplot varlist [if] [in] , by(varname) ///
   [median xlabel(x-axis_labels) xtitle(title_string) ///
   msymbol(marker_symbol) * ]
```

Using -profileplot-

```
profileplot read math science, by(female) ytitle(score)
```

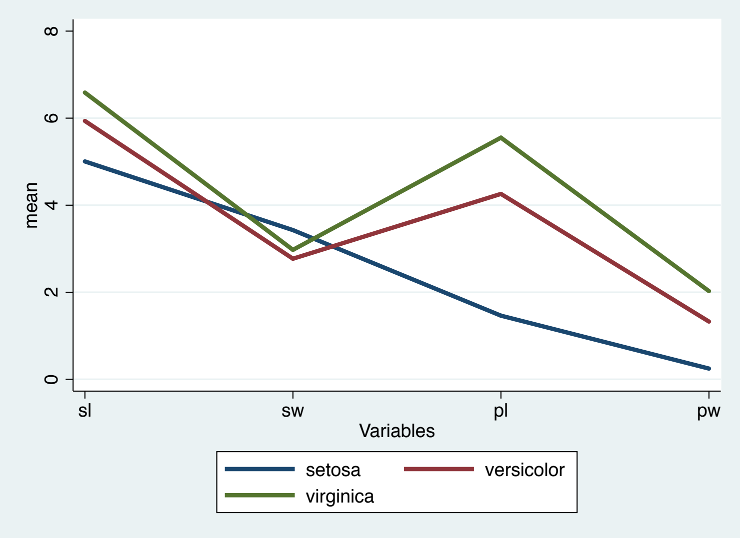


Example 1 - Fisher's Iris Data

Three varieties of Iris
Setosa (n=50)
Versacolor (n=50)
Virginica (n=50)

Four response variables:
Sepal length
Sepal width
Petal length
Petal width

Profile Plot



Group Means

type	sl	sw	pl	pw
setosa	5.006	3.428	1.462	.246
versicolor	5.936	2.77	4.26	1.326
virginica	6.588	2.974	5.552	2.026

Preliminary Manova

```
. manova sl sw pl pw = type

               Number of obs =     150

               W = Wilks' lambda      L = Lawley-Hotelling trace
               P = Pillai's trace     R = Roy's largest root

      Source |  Statistic     df   F(df1,    df2) =   F   Prob>F
    -----------+--------------------------------------------------
        type | W    0.0234     2      8.0   288.0   199.15 0.0000 e
             | P    1.1919            8.0   290.0    53.47 0.0000 a
             | L   32.4773            8.0   286.0   580.53 0.0000 a
             | R   32.1919            4.0   145.0  1166.96 0.0000 u
             |--------------------------------------------------
    Residual |               147
    -----------+--------------------------------------------------
       Total |               149
```

Test of Parallelism - Part 1

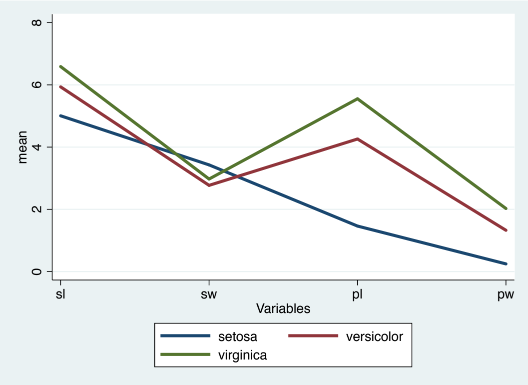
```
. matrix c1 = (1,-1,0,0\0,1,-1,0\0,0,1,-1)

     c1   c2   c3   c4
r1    1   -1    0    0
r2    0    1   -1    0
r3    0    0    1   -1
```

Test of Parallelism - Part 2

```
. manovatest type, ytrans(c1)
 Transformations of the dependent variables
 (1)     sl - sw
 (2)     sw - pl
 (3)     pl - pw
                 W = Wilks' lambda      L = Lawley-Hotelling trace
                 P = Pillai's trace     R = Roy's largest root
     Source |  Statistic     df   F(df1,    df2) =   F   Prob>F
  -----------+--------------------------------------------------
       type | W   0.0412      2      6.0    290.0   189.92 0.0000 e
            | P   0.9691             6.0    292.0    45.75 0.0000 a
            | L  23.0505             6.0    288.0   553.21 0.0000 a
            | R  23.0397             3.0    146.0  1121.27 0.0000 u
            |--------------------------------------------------
   Residual |              147
             e = exact, a = approximate, u = upper bound on F
```

Results are statistically significant, therefore profiles are not parallel

Conclusions for Example 1



Iris varieties profiles are not parallel

Example 2

These data are adapted from a 1996 study (Gregoire, Kumar, Everitt, Henderson and Studd) on the efficacy of estrogen patches in treating postnatal depression.

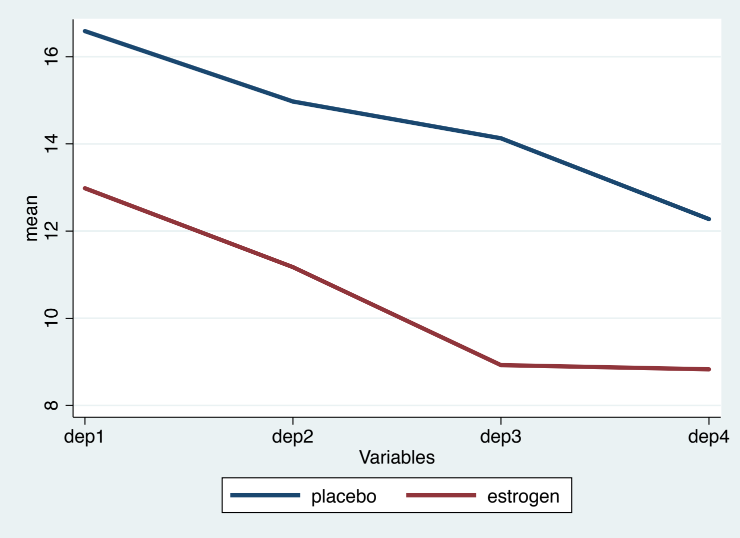
Women were randomly assigned to either a placebo control group (n=17) or estrogen patch group (n=24). The Edinburgh Postnatal Depression Scale (EPDS) data were collected monthly for six months once the treatment began. Higher scores on the EDPS are indicative of higher levels of depression. Only patients with complete data for four months were used in this example.

Group Means

The response variables are the four monthly scores on the depression scale.

group	1 month	2 months	3 months	4 months
placebo	16.58824	14.97294	14.12882	12.27471
estrogen	12.9825	11.17286	8.924643	8.827857

Profile Plot



Preliminary Manova

```
. manova dep1 dep2 dep3 dep4 = group

             Number of obs =       45

      Source |  Statistic     df   F(df1,    df2) =   F   Prob>F
  -----------+----------------------------------------------
       group | W   0.7623      1     4.0    40.0      3.12 0.0253 e
             | P   0.2377            4.0    40.0      3.12 0.0253 e
             | L   0.3117            4.0    40.0      3.12 0.0253 e
             | R   0.3117            4.0    40.0      3.12 0.0253 e
             |----------------------------------------------
    Residual |               43
  -----------+----------------------------------------------
       Total |               44
```

Test of Parallelism - Part 1

```
. matrix c1 = (1,-1,0,0\0,1,-1,0\0,0,1,-1)

     c1   c2   c3   c4
r1    1   -1    0    0
r2    0    1   -1    0
r3    0    0    1   -1
```

Test of Parallelism - Part 2

```
. manovatest group, ytrans(c1)

 Transformations of the dependent variables
 (1)   dep1 - dep2
 (2)   dep2 - dep3
 (3)   dep3 - dep4

     Source |  Statistic     df   F(df1,    df2) =   F   Prob>F
 -----------+----------------------------------------------------
      group | W    0.9095      1      3.0    41.0      1.36 0.2684 e
            | P    0.0905             3.0    41.0      1.36 0.2684 e
            | L    0.0995             3.0    41.0      1.36 0.2684 e
            | R    0.0995             3.0    41.0      1.36 0.2684 e
            |----------------------------------------------------
   Residual |                43
```

Results not significant, therefore profiles are parallel

Test of Levels - Part 1

```
. matrix c2 = (1,1,1,1)

     c1   c2   c3   c4
r1    1    1    1    1
```

Test of Levels - Part 2

```
. manovatest group, ytrans(c2)

 Transformation of the dependent variables
 (1)   dep1 + dep2 + dep3 + dep4

      Source |  Statistic     df   F(df1,     df2) =   F   Prob>F
  -----------+----------------------------------------------
       group | W   0.8448      1     1.0     43.0      7.90 0.0074 e
             | P   0.1552            1.0     43.0      7.90 0.0074 e
             | L   0.1837            1.0     43.0      7.90 0.0074 e
             | R   0.1837            1.0     43.0      7.90 0.0074 e
             |----------------------------------------------
    Residual |                  43
```

Results significant, therefore group levels are different

Test of Flatness - Part 1

We demonstrate the test of flatness even though the the profiles show significant separation.

Test of Flatness - Part 2

```
. manovatest, showorder

 Order of columns in the design matrix
      1: (group==0)
      2: (group==1)
      3: _cons
```

We use the xm matrix to select constant

```
. matrix xm = (0,0,1)

    b0  b1  b2
    c1  c2  c3
r1   0   0   1
```

Test of Flatness - Part 3

```
. manovatest, test(xm) ytrans(c1)

 Transformations of the dependent variables
 (1) dep1 - dep2 (2) dep2 - dep3 (3) dep3 - dep4

 Test constraint:  (1)  _cons = 0

      Source |  Statistic     df   F(df1,    df2) =   F   Prob>F
  -----------+----------------------------------------------------
  manovatest | W   0.8012      1      3.0    41.0     3.39 0.0268 e
             | P   0.1988             3.0    41.0     3.39 0.0268 e
             | L   0.2481             3.0    41.0     3.39 0.0268 e
             | R   0.2481             3.0    41.0     3.39 0.0268 e
             |----------------------------------------------------
    Residual |                43
```

Results are significant, therefore profiles are not flat

Test of Flatness - Part 4

The test of flatness tests whether the constants (intercepts) for each of the dependent variables are equal.

Neither **manova** nor **manovatest** will display the separate intercepts. You can view these using the **mvreg** command.

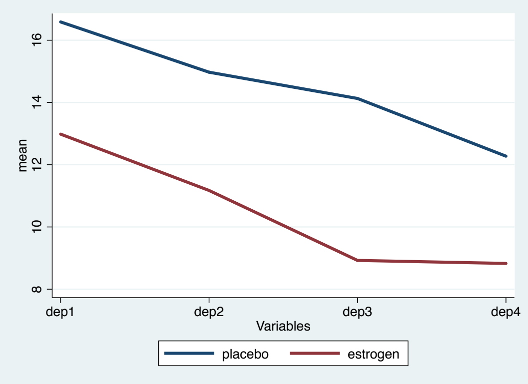
The relationship between **mvreg** and **manova** is analogous to the relationship between **regress** and **anova**, which leads to what looks like an item from the Miller Analogies Test.

mvreg : manova :: regress : anova

Use mvreg to display constants

```
. mvreg dep1 dep2 dep3 dep4 = group

             |      Coef.   Std. Err.      t    P>|t|
-------------+--------------------------------------------------
dep1 g  roup |  -3.605735   1.721864    -2.09   0.042
       _cons |   16.58824   1.358225    12.21   0.000
-------------+--------------------------------------------------
dep2   group |  -3.800084   1.918374    -1.98   0.054
       _cons |   14.97294   1.513234     9.89   0.000
-------------+--------------------------------------------------
dep3   group |  -5.204181   1.624143    -3.20   0.003
       _cons |   14.12882   1.281141    11.03   0.000
-------------+--------------------------------------------------
dep4   group |  -3.446849     1.5799    -2.18   0.035
       _cons |   12.27471   1.246242     9.85   0.000
```

Conclusions for Example 2



Group profiles are parallel
Group levels differ significantly
Lines are not flat

Alternative Approach

There are alternatives to performing profile analysis with wide data in a multivariate framework. One alternative would be to stack the response variables. Once the data are in long form, use a linear mixed model to test the various profile analysis hypotheses.

Summary

Profile analysis is an interesting multivariate method that appears to have fallen out of favor. It is rarely seen in the current research literature but it can still be useful in certain situations.

Reference:

Morrison, D.F. (1976) *Multivariate statistical methods* (2nd ed.). New York, NY: McGraw-Hill.