Phil Ender

UCLA Statistical Consulting Group

July 31, 2014
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

Ex 1: Parallelism

Variables

grp = 1

grp = 2

Phil Ender

Profile Analysis
Example of Nonparallel Profiles

Ex 2: Nonparallel

Variables

\[\text{grp } = 1\]
\[\text{grp } = 2\]
2. Test of Levels (Separation)

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

Ex 3: Separation

Variables

 grp = 1
 grp = 2
Example of No Separation

Ex 4: No Separation

mean

y1 y2 y3 y4

Variables

grp = 1

grp = 2
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

Ex 5: Flatness

Variables

grp = 1
grp = 2

y1 y2 y3 y4

mean
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 Analysis Overview
Profile Plots

Example 1
Example 2

Profile Plot

Variables
setosa	versicolor
virginica

Phil Ender
<table>
<thead>
<tr>
<th>type</th>
<th>sl</th>
<th>sw</th>
<th>pl</th>
<th>pw</th>
</tr>
</thead>
<tbody>
<tr>
<td>setosa</td>
<td>5.006</td>
<td>3.428</td>
<td>1.462</td>
<td>.246</td>
</tr>
<tr>
<td>versicolor</td>
<td>5.936</td>
<td>2.77</td>
<td>4.26</td>
<td>1.326</td>
</tr>
<tr>
<td>virginica</td>
<td>6.588</td>
<td>2.974</td>
<td>5.552</td>
<td>2.026</td>
</tr>
</tbody>
</table>
. 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

<table>
<thead>
<tr>
<th>Source</th>
<th>Statistic</th>
<th>df</th>
<th>F(df1, df2) = F</th>
<th>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>type</td>
<td>W</td>
<td>2</td>
<td>8.0 288.0 199.15</td>
<td>0.0000 e</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>8.0 290.0 53.47</td>
<td>0.0000 a</td>
<td></td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>8.0 286.0 580.53</td>
<td>0.0000 a</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>4.0 145.0 1166.96</td>
<td>0.0000 u</td>
<td></td>
</tr>
</tbody>
</table>

Residual | 147

Total    | 149
Test of Parallelism - Part 1

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

<table>
<thead>
<tr>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>r2</td>
<td>0</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>r3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
. 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

<table>
<thead>
<tr>
<th>Source</th>
<th>Statistic</th>
<th>df</th>
<th>F(df1, df2) = F</th>
<th>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>type</td>
<td>W 0.0412</td>
<td>2</td>
<td>6.0 290.0 189.92</td>
<td>0.0000 e</td>
</tr>
<tr>
<td></td>
<td>P 0.9691</td>
<td>6</td>
<td>292.0 45.75 0.0000 a</td>
<td></td>
</tr>
<tr>
<td></td>
<td>L 23.0505</td>
<td>6</td>
<td>288.0 553.21 0.0000 a</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R 23.0397</td>
<td>3</td>
<td>146.0 1121.27 0.0000 u</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>147</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>group</th>
<th>1 month</th>
<th>2 months</th>
<th>3 months</th>
<th>4 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>estrogen</td>
<td>12.9825</td>
<td>11.17286</td>
<td>8.924643</td>
<td>8.827857</td>
</tr>
</tbody>
</table>
**Profile Analysis Overview**

**Profile Plots**

**Example 1**

**Example 2**

---

### Preliminary Manova

```
.manova dep1 dep2 dep3 dep4 = group
```

**Number of obs = 45**

<table>
<thead>
<tr>
<th>Source</th>
<th>Statistic</th>
<th>df</th>
<th>F(df1, df2)</th>
<th>F</th>
<th>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>W 0.7623</td>
<td>1</td>
<td>4.0</td>
<td>40.0</td>
<td>3.12</td>
</tr>
<tr>
<td></td>
<td>P 0.2377</td>
<td>4.0</td>
<td>40.0</td>
<td>3.12</td>
<td>0.0253</td>
</tr>
<tr>
<td></td>
<td>L 0.3117</td>
<td>4.0</td>
<td>40.0</td>
<td>3.12</td>
<td>0.0253</td>
</tr>
<tr>
<td></td>
<td>R 0.3117</td>
<td>4.0</td>
<td>40.0</td>
<td>3.12</td>
<td>0.0253</td>
</tr>
</tbody>
</table>

| Residual | 43 |

| Total    | 44 |
Test of Parallelism - Part 1

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

<table>
<thead>
<tr>
<th></th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>r2</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>r3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
</tr>
</tbody>
</table>
Test of Parallelism - Part 2

.manovatest group, ytrans(c1)

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

<table>
<thead>
<tr>
<th>Source</th>
<th>Statistic</th>
<th>df</th>
<th>F(df1, df2) = F</th>
<th>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>W 0.9095</td>
<td>1</td>
<td>3.0 41.0 1.36</td>
<td>0.2684 e</td>
</tr>
<tr>
<td></td>
<td>P 0.0905</td>
<td>3</td>
<td>41.0 1.36 0.2684 e</td>
<td></td>
</tr>
<tr>
<td></td>
<td>L 0.0995</td>
<td>3</td>
<td>41.0 1.36 0.2684 e</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R 0.0995</td>
<td>3</td>
<td>41.0 1.36 0.2684 e</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results not significant, therefore profiles are parallel
Test of Levels - Part 1

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

<table>
<thead>
<tr>
<th></th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Transformation of the dependent variables
(1) dep1 + dep2 + dep3 + dep4

<table>
<thead>
<tr>
<th>Source</th>
<th>Statistic</th>
<th>df</th>
<th>F(df1, df2) = F</th>
<th>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>W 0.8448</td>
<td>1</td>
<td>1.0 43.0</td>
<td>7.90</td>
</tr>
<tr>
<td></td>
<td>P 0.1552</td>
<td>1</td>
<td>1.0 43.0</td>
<td>7.90</td>
</tr>
<tr>
<td></td>
<td>L 0.1837</td>
<td>1</td>
<td>1.0 43.0</td>
<td>7.90</td>
</tr>
<tr>
<td></td>
<td>R 0.1837</td>
<td>1</td>
<td>1.0 43.0</td>
<td>7.90</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results significant, therefore group levels are different
We demonstrate the test of flatness even though the profiles show significant separation.
. 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
. manovatest, test(xm) ytrans(c1)

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

Test constraint: (1) _cons = 0

<table>
<thead>
<tr>
<th>Source</th>
<th>Statistic</th>
<th>df</th>
<th>F(df1, df2) = F</th>
<th>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>manovatest</td>
<td>W 0.8012</td>
<td>1</td>
<td>3.0 41.0 3.39</td>
<td>0.0268</td>
</tr>
<tr>
<td></td>
<td>P 0.1988</td>
<td>3</td>
<td>3.0 41.0 3.39</td>
<td>0.0268</td>
</tr>
<tr>
<td></td>
<td>L 0.2481</td>
<td>3</td>
<td>3.0 41.0 3.39</td>
<td>0.0268</td>
</tr>
<tr>
<td></td>
<td>R 0.2481</td>
<td>3</td>
<td>3.0 41.0 3.39</td>
<td>0.0268</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results are significant, therefore profiles are not flat
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    group      |  -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  
```

Phil Ender
Profile Analysis
Group profiles are parallel
Group levels differ significantly
Lines are not flat
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.
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: