

# **Profile Analysis of Multivariate Data Using the profileR Package**

**Okan Bulut**

Centre for Applied Research in Measurement and Evaluation

University of Alberta

[bulut@ualberta.ca](mailto:bulut@ualberta.ca)

---

# Outline

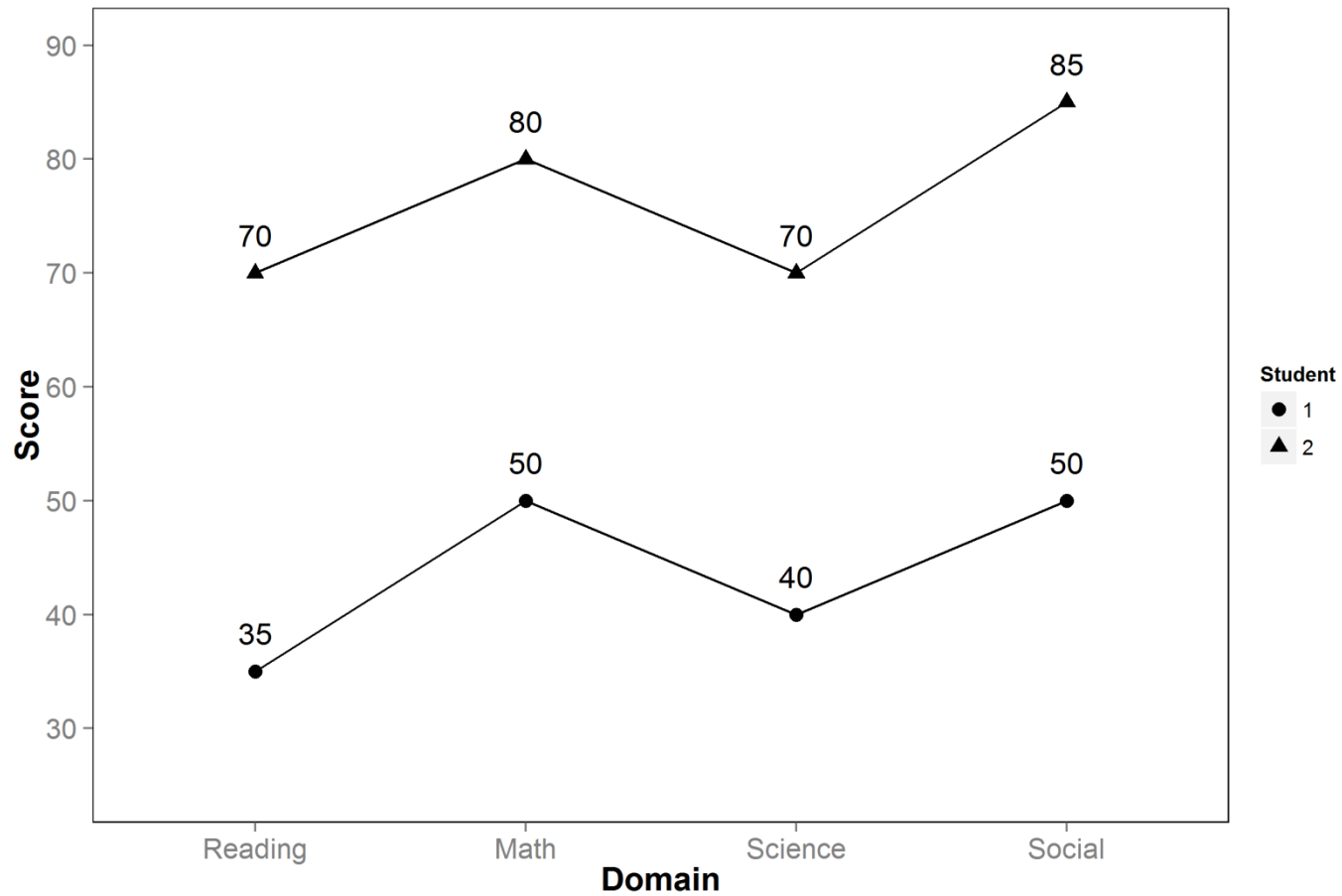


- ❑ Profile analysis
  - ❑ Profiles and profile plots
  - ❑ Profile analysis for one-sample or by group
  - ❑ Pattern and level in profiles
  
- ❑ The *profileR* package in R

# Profile Analysis

- ❑ Profile analysis is a multivariate data analysis technique that is the equivalent of a repeated measures extension of MANOVA.
  
- ❑ A profile consists of
  - ❑ a set of dependent variables on the same scale or
  - ❑ the same dependent variable measured at different time points

# Profile Plots



# Profile Analysis

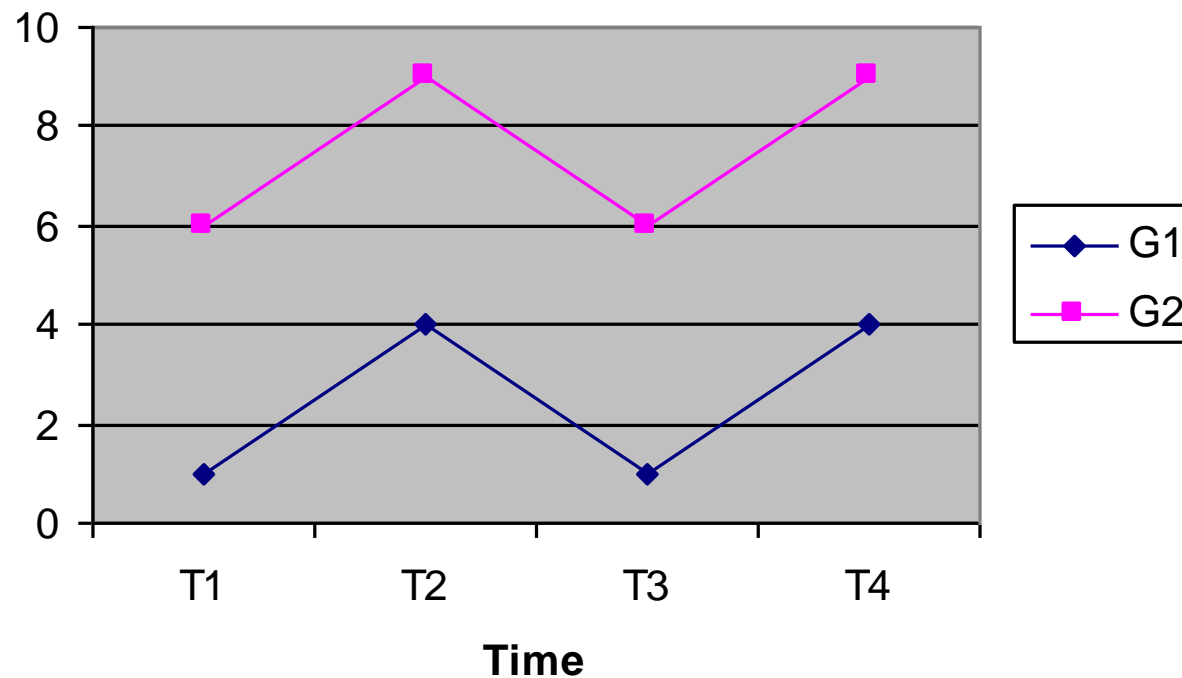
	$T_1$	$T_2$	$T_3$	$T_4$
Group <sub>1</sub>	<i>DV</i>	<i>DV</i>	<i>DV</i>	<i>DV</i>
Group <sub>2</sub>	<i>DV</i>	<i>DV</i>	<i>DV</i>	<i>DV</i>

- Profile analysis tells us about whether the groups are similar in
  - elevation,
  - dispersion,
  - shape,
  - level and pattern

# Profile Analysis Hypothesis



**Are the profiles for the two groups the same?**  
**Are the profiles “parallel”?**

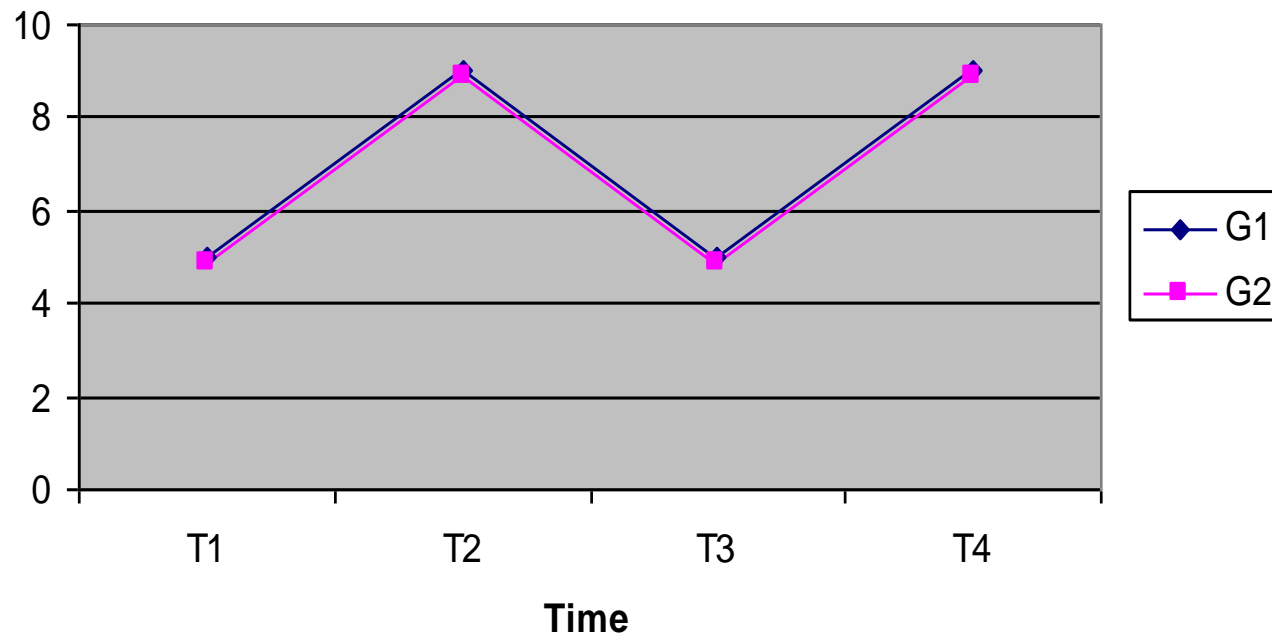


(Equivalent of the interaction effect in repeated measures ANOVA)

# Profile Analysis Hypothesis



**On average does one group score higher than the other?  
Do the groups have “equal levels”?**

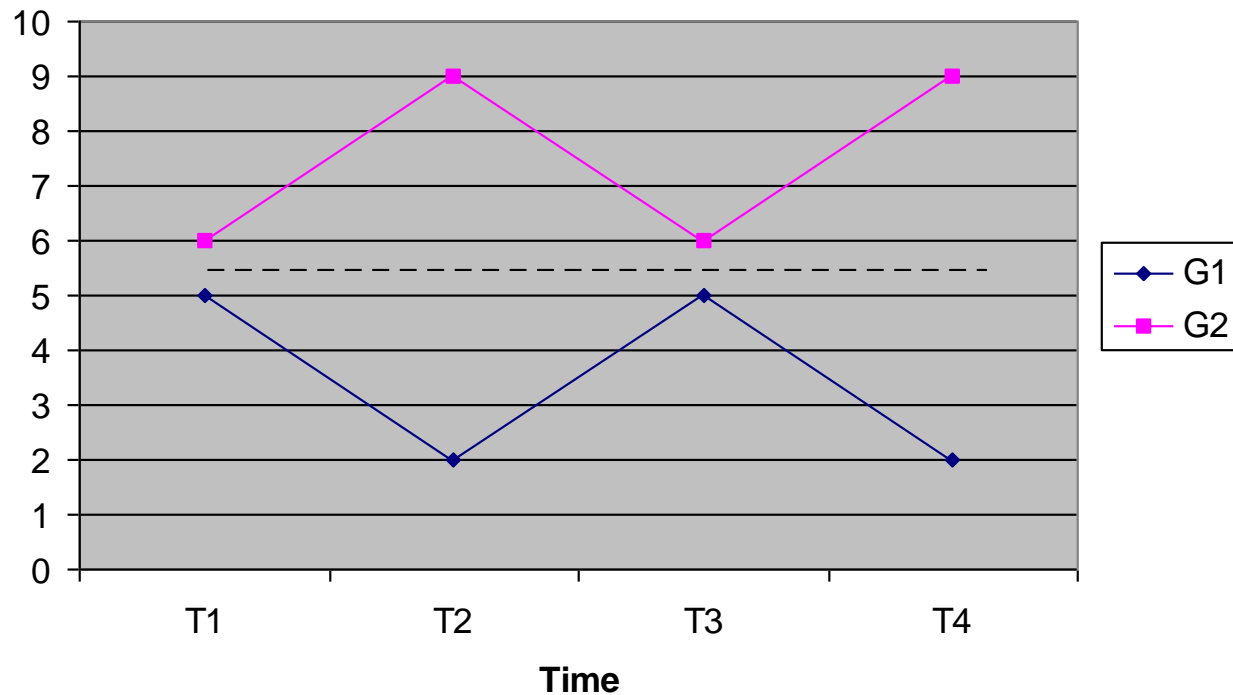


(Equivalent of between-subjects main effect in mixed-effect ANOVA)

# Profile Analysis Hypothesis



**Do all DVs elicit the same average response?**  
**Are the profiles “flat”?**



(Equivalent of within-subjects main effect in repeated measures ANOVA)



# Assumptions



1. Sample size needs to be large enough; more subjects in the smallest cell than the number of DVs
2. Multivariate normality
3. Homogeneity of variance-covariance matrices
4. Linearity

# Limitations



1. All variables should be on the same scale.
2. Manipulated group variable is the only source of causality.
3. Generalizability is limited to the sample/population being used.

# Pattern and Level in a Profile

$$\text{Pattern} = X_v - \text{Level}$$

Person	$X_1$	$X_2$	$X_3$	$X_v$
1				
2				
3				
$P$				

$$\text{Level} = \left( \sum_{v=1}^V X_v \right) / V$$

# How to use level and pattern?

1. Criterion-related profile analysis (Davison & Davenport, 2002):

$$Y'_p = \sum_v b_v X_{pv} + a$$

- $X_{pv}$  is the observed score for person p on variable v
- $Y_p$  is the criterion variable
- $b_v$  is the regression weight
- $a$  is the intercept

# How to use level and pattern?

According to Davison and Davenport (2002):

$$Y'_p = (V / k)Cov_p + VbX_{p.} + a$$

- $X_{p.}$  = Average person value (i.e., level)
- $Cov_p = (1 / V) \sum_v (X_{pv} - X_{p.})(X_{cv} - X_{c.})$  (i.e., pattern)

$$Y'_p = b_1X_{p1} + b_2X_{p2} + b_3X_{p3} + b_4X_{p4} + a.$$



$$Y'_p = \sum_v(b_v - b_{.}) (X_{pv} - X_{p.}) + \sum_v b_{.}X_{p.} + a$$



$$Y'_p = (V/k)Cov_{pc} + Vb_{.}X_{p.} + a.$$

# Example



*Raw Score and Deviation Score Profiles of 6 Hypothetical Clients Diagnosed as Either Neurotic or Psychotic*

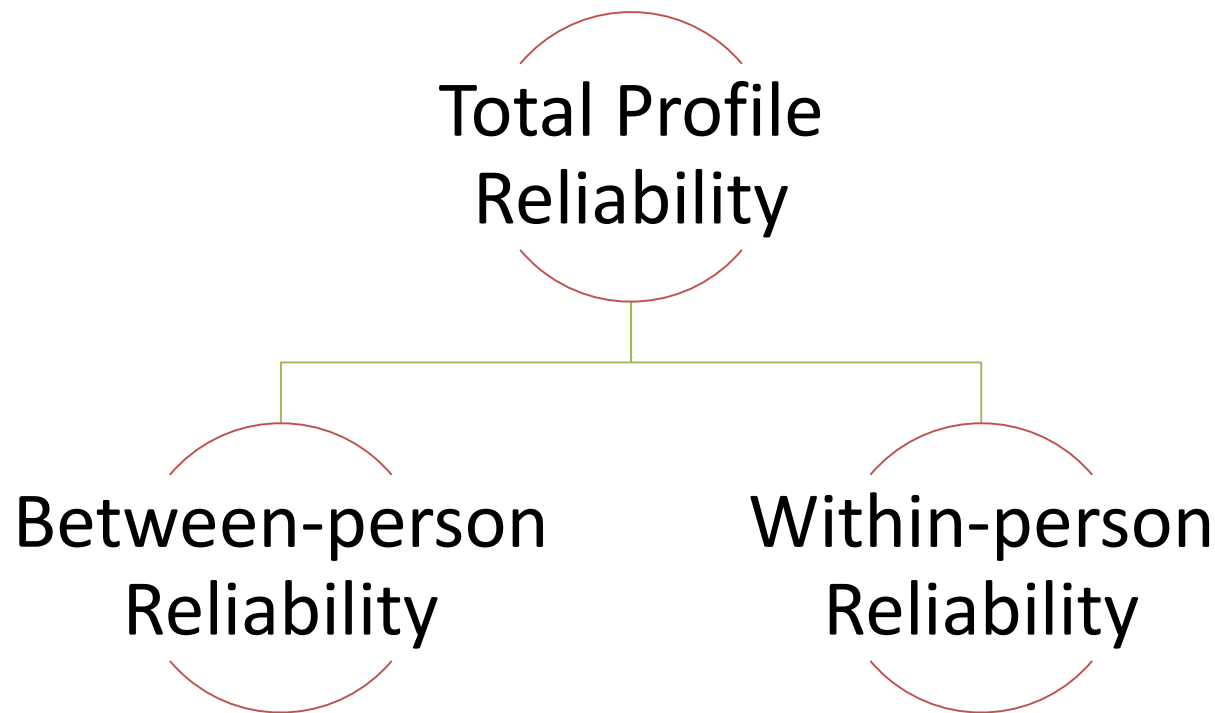
Client	IPMM scale and variable						$Cov_{pc}$
	A	H	S	B	$N-P$	$X_p$	
	Raw score						
1	90	60	90	80	1	80.00	-115.4
2	75	75	85	85	1	80.00	-80.8
3	75	60	95	75	1	76.25	-132.7
4	65	50	115	90	0	80.00	-346.2
5	60	55	100	105	0	80.00	-380.8
6	70	45	100	90	0	76.25	-328.9

*Note.* IPMM = Inventory of Personality and Mood Manifestations; A = Anxiety; H = Hypochondriasis; S = Schizophrenia; B = Bipolar Disorder;  $N-P$  = neurotic versus psychotic criterion variable (1 = neurotic, 0 = psychotic);  $X_p$  = the profile level of person  $p$ ;  $Cov_{pc}$  = the covariance between the score profile of person  $p$  and the criterion-pattern vector.

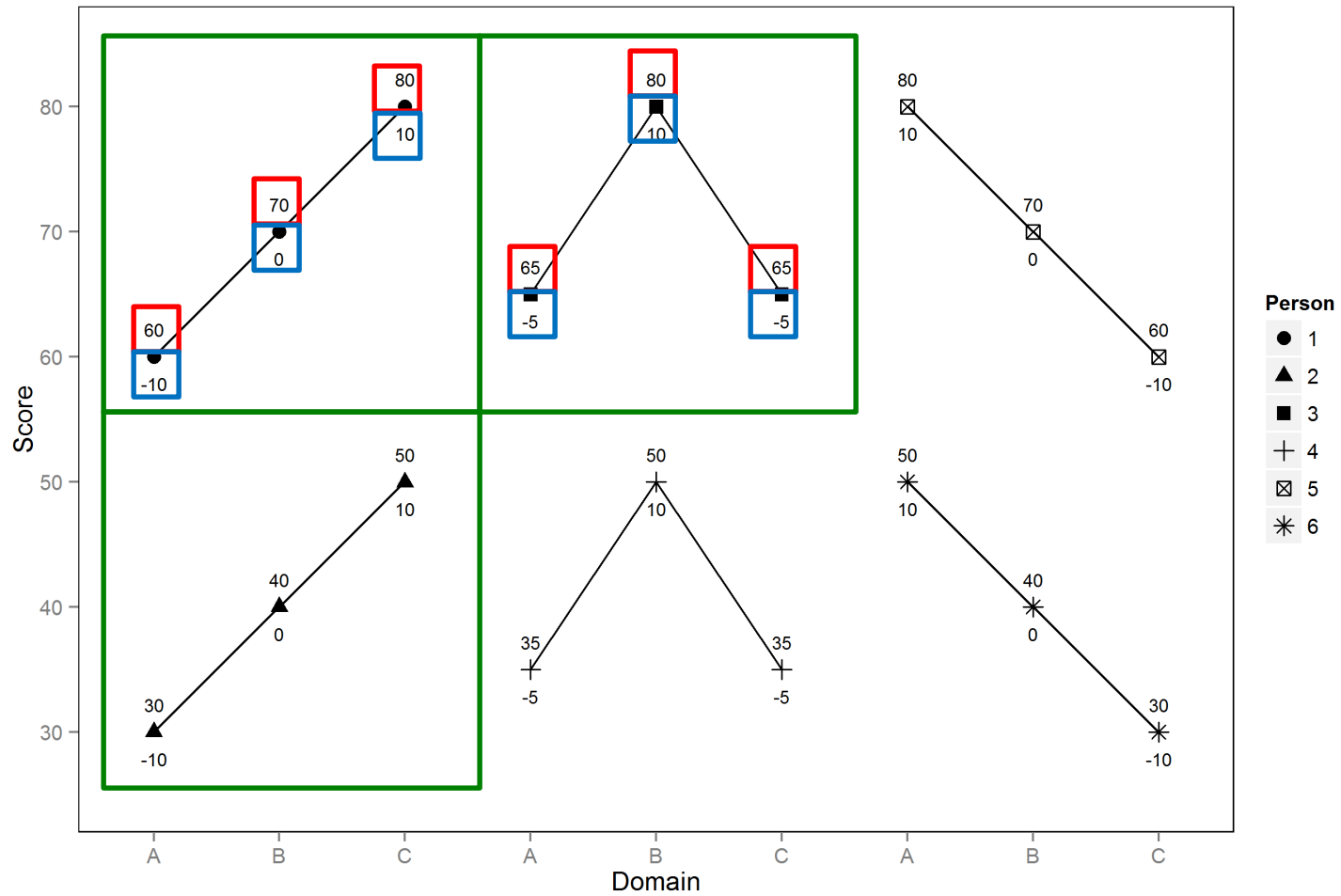
# How to Use Level and Pattern?

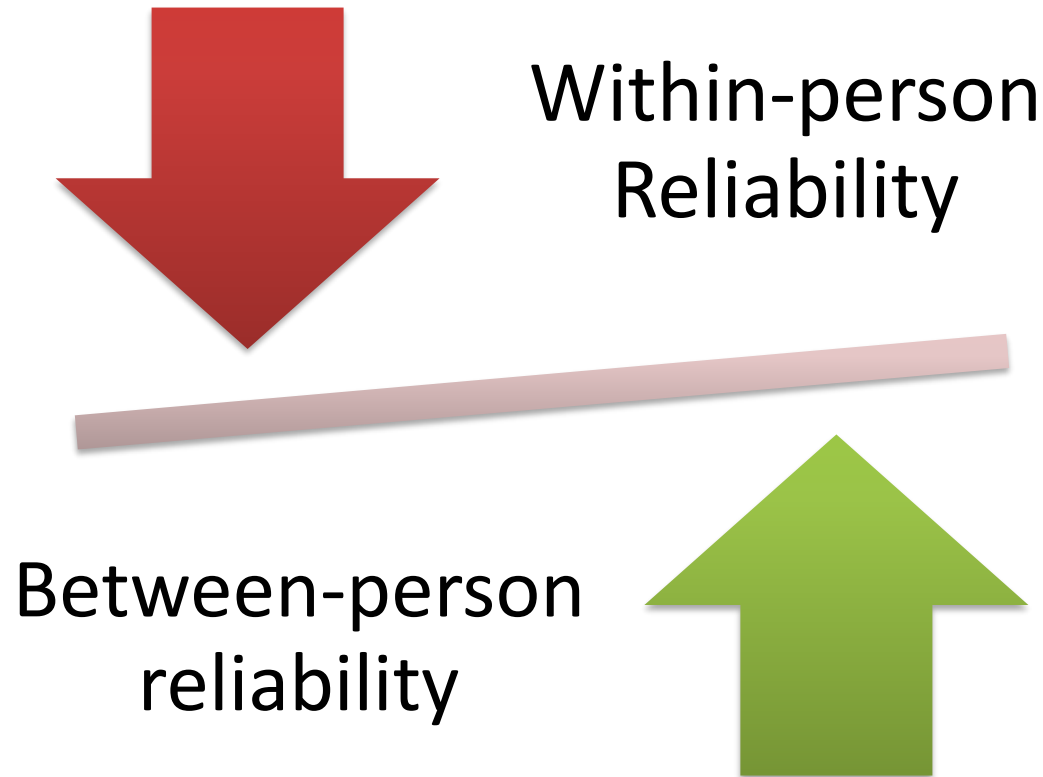


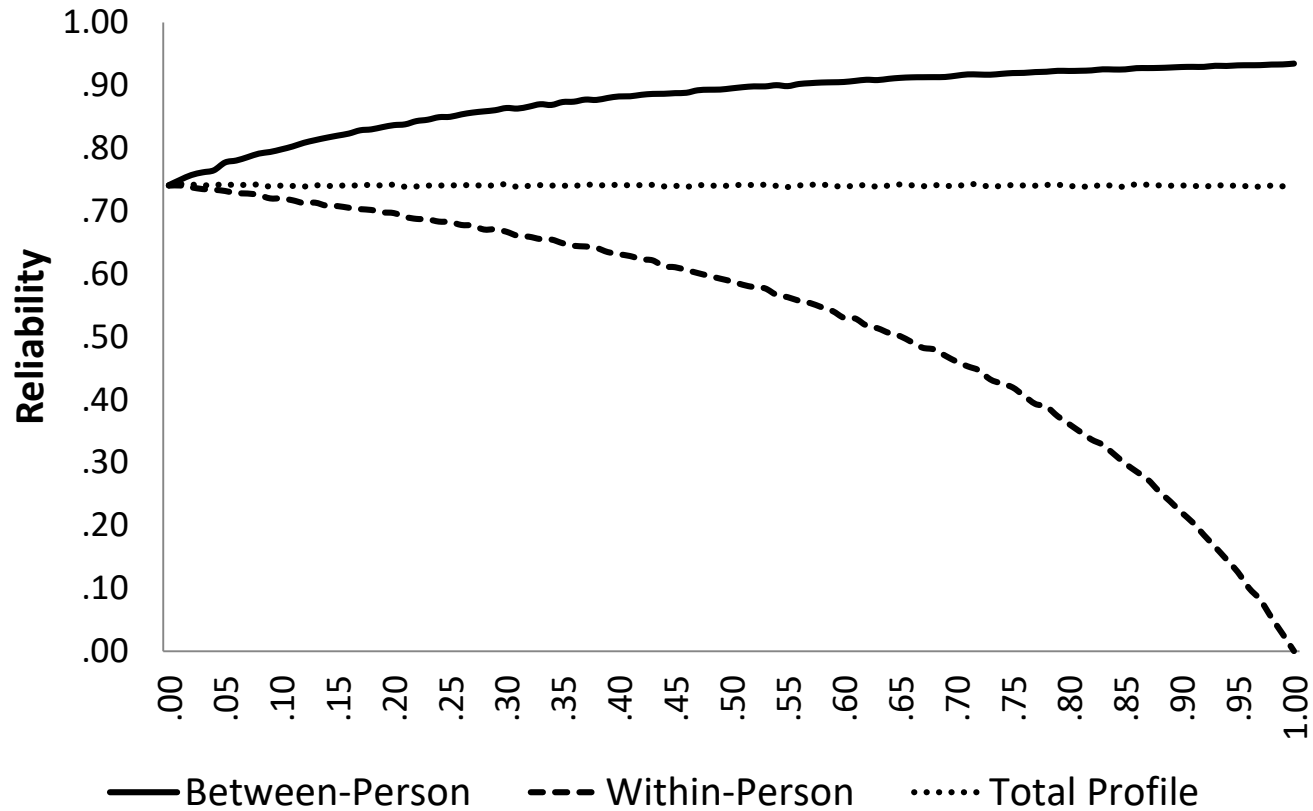
## 2. Profile reliability (Bulut, 2013):











# What else?



- ❑ Profile analysis via multidimensional scaling (Davison, 1994)
- ❑ Added-value of subscores (Davison et al., 2015)
- ❑ Within-person factor model to derive latent profiles (Davison et al., 2009)

# The *profileR* package in



## **profileR: Profile Analysis of Multivariate Data in R**

Fit profile analytic models. This includes the multivariate methods and data visualization tools to implement profile analysis

Version: 0.3-1  
Depends: [ggplot2](#), [MASS](#), [RColorBrewer](#), [reshape](#), methods, [lavaan](#), R ( $\geq 3.0.0$ )  
Published: 2015-10-19  
Author: Okan Bulut, Christopher David Desjardins  
Maintainer: Christopher David Desjardins <cddesjardins at gmail.com>  
License: [GPL-2](#) | [GPL-3](#) [expanded from: GPL ( $\geq 2$ )]  
NeedsCompilation: no  
Materials: [README](#) [NEWS](#)  
In views: [Psychometrics](#)  
CRAN checks: [profileR results](#)

### Downloads:

Reference manual: [profileR.pdf](#)  
Vignettes: [User manual](#)  
Package source: [profileR\\_0.3-1.tar.gz](#)  
Windows binaries: r-devel: [profileR\\_0.3-1.zip](#), r-release: [profileR\\_0.3-1.zip](#), r-oldrel: [profileR\\_0.3-1.zip](#)  
OS X Snow Leopard binaries: r-release: [profileR\\_0.3.tgz](#), r-oldrel: [profileR\\_0.2-1.tgz](#)  
OS X Mavericks binaries: r-release: [profileR\\_0.3-1.tgz](#)  
Old sources: [profileR archive](#)

<http://CRAN.R-project.org/package=profileR>

Centre for Research in Applied Measurement and Evaluation

# The *profileR* package in



- ❑ `install.packages("profileR")`
- ❑ `library("profileR")`

Also check out:

- ❑ <https://github.com/cddesja/profileR>

`devtools::install_github(repo = "cddesja/profileR", build_vignettes = TRUE)`



# Functions in the *profileR* package

- ❑ Criterion-related profile analysis
- ❑ Profile analysis with Hotelling's  $T^2$
- ❑ Profile analysis by groups
- ❑ Profile analysis via multidimensional scaling
- ❑ Profile reliability
- ❑ Within-person random intercept factor model
- ❑ Profile plots
- ❑ Moderated profile analysis (experimental!)



# *profileR* – Example 1

The spouse data come from a study of love and marriage. A sample of 30 husbands and their wives were asked to respond to the following questions:

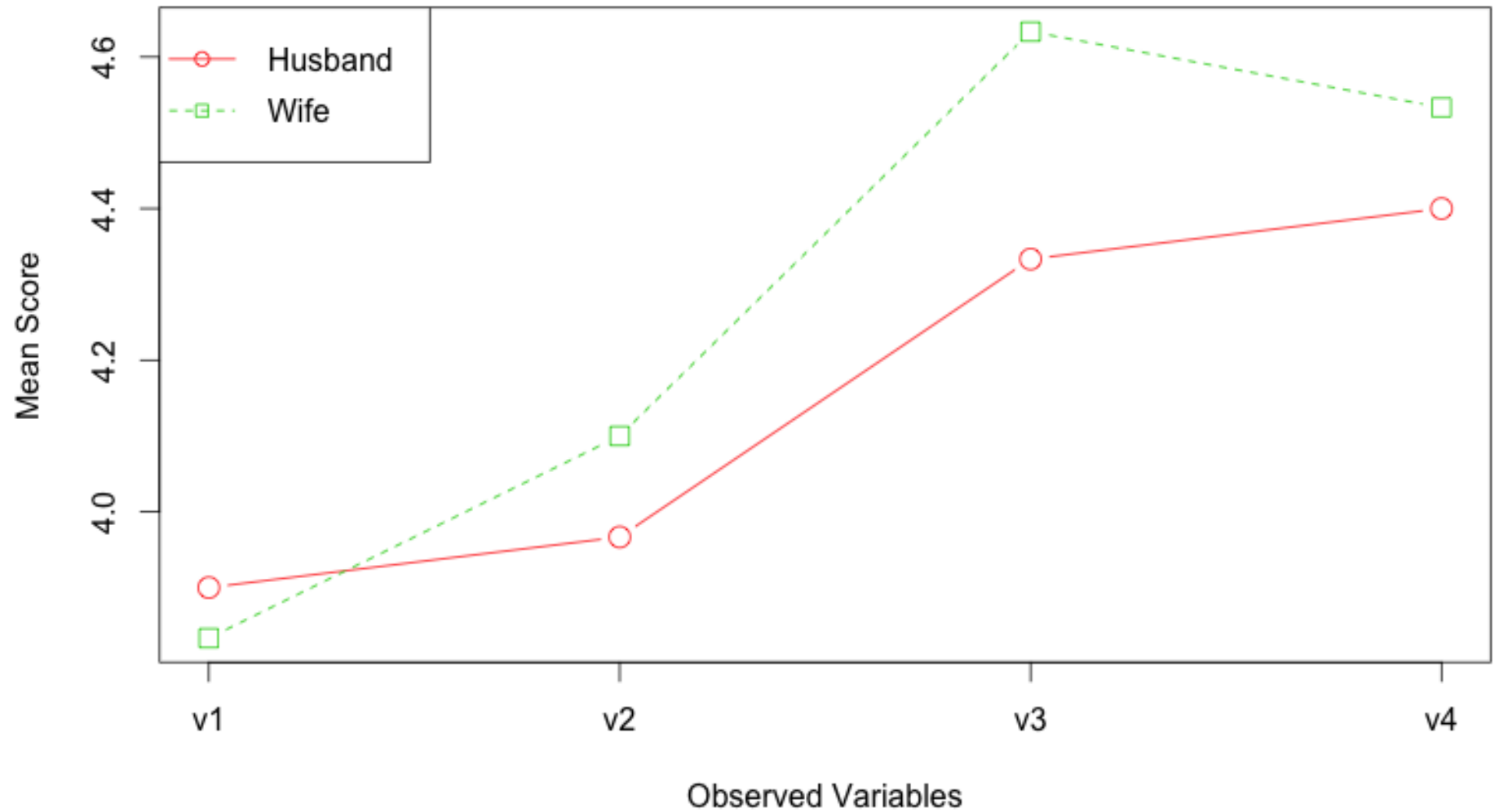
1. What is the level of passionate love you feel for your partner?
2. What is the level of passionate love that your partner feels for you?
3. What is the level of companionate love that you feel for your partner?
4. What is the level of companionate love that your partner feels for you?

The responses to all four questions are on a five-point Likert scale where 1 indicates "none at all" and 5 indicates "tremendous amount".





ID	item1	item2	item3	item4	spouse
1	2	3	5	5	Husband
2	5	5	4	4	Husband
3	4	5	5	5	Husband
4	4	3	4	4	Husband
5	3	3	5	5	Husband
6	3	3	4	5	Husband



```
> data(spouse)
> mod <- pbgi(spouse[,1:4], spouse[,5])
> summary(mod)
```

Call:

```
pbgi(x = spouse[, 1:4], y = spouse[, 5])
```

Hypothesis Tests:

	F	df1	df2	p-value
Ho: Profiles are parallel	8.016171	3	56	0.06255
Ho: Profiles are coincidental	1.532770	1	58	0.22068
Ho: Profiles are level	24.820709	3	57	0.00015

# *profileR* – Example 2



## □ Inventory of Personality and Mood Manifestation

```
> data(IPMMc)
```

```
> IPMMc
```

	A	H	S	B	R
1	75	60	50	50	1
2	60	75	45	55	1
3	60	60	55	45	1
4	50	50	75	60	0
5	45	55	60	75	0
6	55	45	60	60	0



```
> mod <- cpa(R ~ A + H + S + B, data = IPMMc)
> print(mod)
```

Call:

```
cpa(formula = R ~ A + H + S + B, data = IPMMc)
```

Coefficients

Call: glm(formula = formula, family = family, data = data,  
na.action = na.action)

Coefficients:

(Intercept)	A	H	S	B
0.500000	0.009231	0.023077	-0.009231	-0.023077

```
> anova(mod)
```

```
Call:
```

```
cpa(formula = R ~ A + H + S + B, data = IPMMc)
```

Analysis of Variance Table

	df1	df2	F value	Pr(>F)
R2.full = 0	4	1	7.87500e+00	0.2604188
R2.pat = 0	3	1	1.05000e+01	0.2221903
R2.lv1 = 0	1	1	0.00000e+00	1.0000000
R2.full = R2.lv1	3	1	1.05000e+01	0.2221903
R2.full = R2.pat	1	1	-7.21645e-15	1.0000000

# *profileR* – Example 3



## □ Entrance Examination for Graduate Studies

```
> data(EEGS)
```

```
> EEGS
```

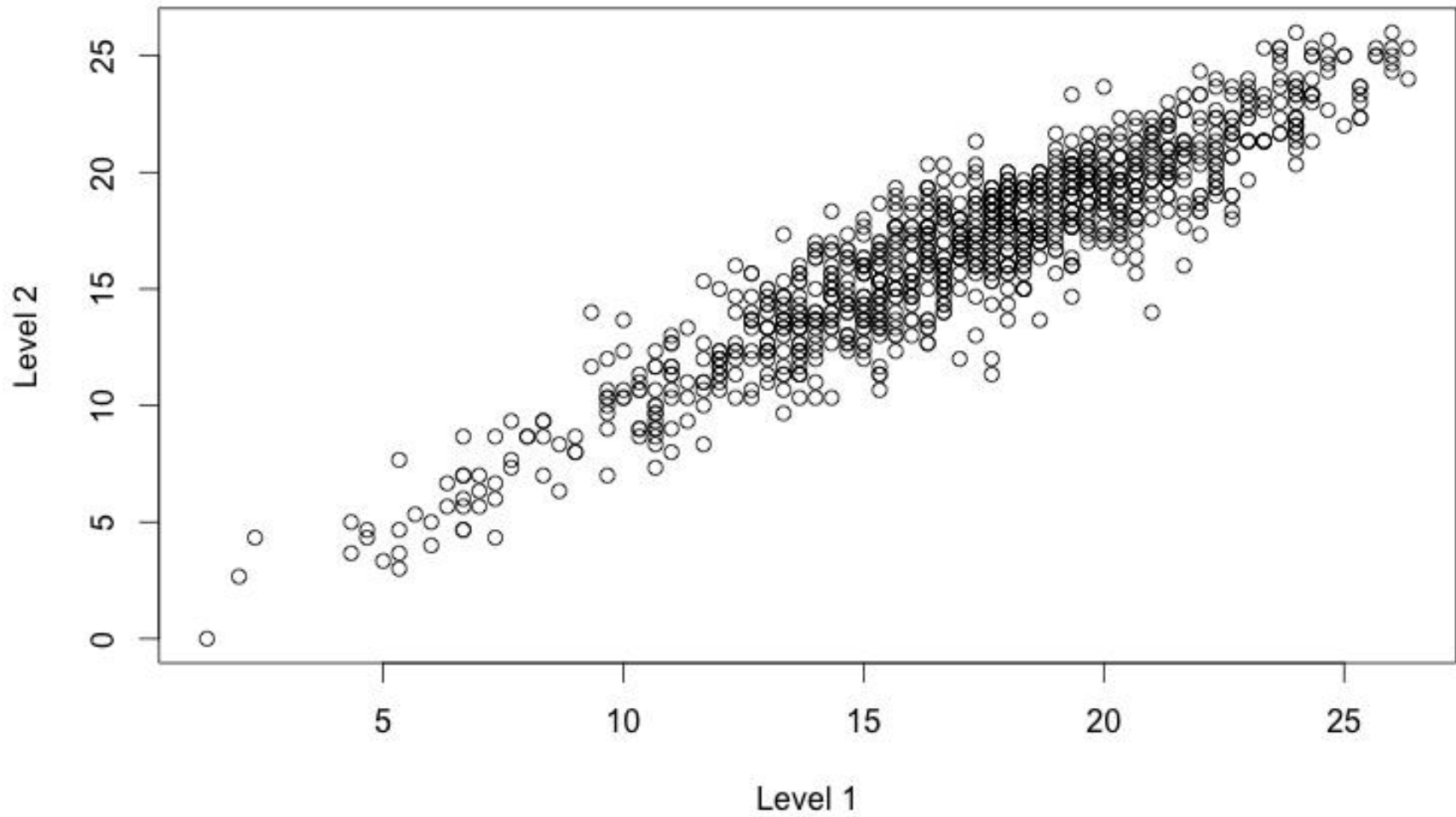
	Form1_Q1	Form2_Q1	Form1_Q2	Form2_Q2	Form1_V	Form2_V
[1,]	2	0	2	0	0	0
[2,]	4	9	0	0	3	4
[3,]	4	3	8	6	27	27
[4,]	2	6	0	0	26	29
[5,]	7	4	3	2	8	6
[6,]	18	16	1	3	14	15

```
> result <- pr(EEGS[,c(1,3,5)],EEGS[,c(2,4,6)])  
> print(result)
```

Subscore Reliability Estimates:

	Estimate	
Level	0.9245548	→ Between-person reliability
Pattern	0.9338338	→ Within-person reliability
Overall	0.9308374	→ Total profile reliability







UNIVERSITY OF  
ALBERTA

Thank you!

For further information please contact:

[bulut@ualberta.ca](mailto:bulut@ualberta.ca)