

Errata:

Chapter 3.2:

The symbols and numbers used for table 3-1 are:

		case g*	
		presence	absence
case g	presence (1) or (+)	d = conjoint presence	b = mismatch
	absence (0) or (-)	c = mismatch	a = conjoint absence

Errata:

Chapter 4.2:

The correct syntax for QUICK CLUSTER after eliminating the missing cases is:

```
get file="c:\koeln\spss\jkult.sav".
```

```
compute valid=sum.8(v39.01 to v39.08).
```

```
fre var=valid.
```

```
select if (valid >= 0).
```

```
fre var=valid.
```

```
FLIP
```

```
VARIABLES=v39.01 v39.02 v39.03 v39.04 v39.05 v39.06 v39.07 v39.08
```

```
.
```

```
CLUSTER var001 to var543
```

```
<---(and not var
```

```
541!!!)
```

```
  /METHOD Baverage single complete
```

```
  /MEASURE= correlation
```

```
  /PRINT SCHEDULE cluster(3)
```

```
  /PRINT DISTANCE
```

```
  /ID case_lbl
```

```
  /save cluster(2,7)
```

```
  /PLOT DENDROGRAM.
```

Chapter 4.4:

The results for complete linkage, single linkage and baverage linkage are wrong. The correct results for the 3 cluster solutions are:

	Cluster 1	Cluster 2	Cluster 3
single	{v39.03, v39.05, v39.02, v39.07}	{v39.01, v39.08}	{v39.04, v39.06}
complete	{v39.03, v39.05}	{v39.01, v39.08, V39.02, V39.07}	{v39.04, v39.06}
baverage	{v39.03, v39.05}	{v39.01, v39.08}	{ V39.02, V39.07. V39.04, v39.06}

Chapter 4.7:

The formulas for the Mojena's rule are wrong.

The correct formula for **Mojena's rule I** is:

$t_{1k} = (v_{k+1} - \bar{v})/s$ with

$$\bar{v} = \sum_{i=1}^{n-1} v_i / (n-1)$$

$$s = \sqrt{\frac{1}{n-2} \sum_i (v_i - \bar{v})^2}$$

The t_1 s are simple the standardized coefficient. They can be computed with a simple SPSS-Syntax. The correct test statistics for the examples are:

Mojena's test statistic t_1 for the example of chapter 4-1.

CLUSTERS	ZCOEFF = T1
5,00	-,85553
4,00	-,47529
3,00	-,09506
2,00	1,42588

The results for the example of chapter 4-2 are:

NC	ZCOEFF
7,00	2,10532
6,00	,17292
5,00	-,07677
4,00	-,09305
3,00	-,55987
2,00	-,70100
1,00	-,84755

The correct formula for Mojena's rule II is as described in the skript. However note: Only p (in CLUSTAN: $0.5*n$) values are used in the computation of the regression slope and the regression constant. ALMO uses all previous cases.

Properties: Rule I does not depend on the number of cases, rule II only in so far as the number of cases that are used in estimation is concerned.

Sorry for these errors!

Errata:

Chapter 5.6:

The formula of Catell's coefficient of profile similarity r_p (Catell 1949) is:

$$r_p(i \in C_I, k \in C_K) = \frac{\chi_{0,5}^2 - \chi^2}{\chi_{0,5}^2 + \chi^2} \quad df = p$$

with

$$\chi^2 = \frac{n_i \cdot n_k}{n_i + n_k} \sum_j \frac{(\bar{x}_{ij} - \bar{x}_{kj})^2}{s_{j/ik}^2}.$$