

# Principal component analyse (PCA) til typebestemmelse

Thomas Teasdale, som er lektor i psykologi, dr. med. ved Københavns Universitet, har i bogen om "Undersøgelsesmetoder i klinisk psykologi" (Munksgaard 1992) forsøgt at forklare matematikken ved faktoranalyse. Han har til dette formål vist en fiktiv opstilling af den korrelationsmatrix (Tabel 1), som er opstået, når man vil måle intelligens med seks forskellige test eller item (A, B, C, D, E, F). Man kan i Tabel 1 se, at de seks item korrelerer positivt med hinanden i større eller mindre grad.

Fra denne korrelationsmatrix i Tabel 1 har Teasdale foretaget den matrix algebra som findes i PCA (principal component analysis), nemlig den matematiske metode som Hotelling beskrev i 1933, hvorved man går fra korrelationskoefficient til egenvektoren (eigenvalue) som udtrykker den varians som de enkelte item indeholder.

Figur 1 viser de egenvektorer eller eigenvalues som Teasdale har udregnet i sin fiktive opstilling.

Summen af disse eigenvalues er 6 (=antallet af komponenter)

Således har 1. komponent en eigenvalue på 3.1, 2. komponent på 1.3, 3. komponent på 0.43, 4. komponent på 0.41, 5. komponent på 0.39 og 6. komponent på 0.36. Disse værdier er anført i Figur 1, men i Figur 1 er også angivet den procentvise varians som hver af disse seks komponenter er ansvarlig for.

I Figur 1 er "explained variance" i procent angivet op ad ordinataksen. Ud ad abscisseaksen er anvist de seks komponenter. Således forklarer 1. komponent 51.7% af variansen, 2. komponent forklarer 21.7% af variansen, hvilket vil sige, at de to første principale komponenter tilsammen forklarer 73.4% af variansen, hvorfor de resterende komponenter er ret betydningsløse.

På selve abscissen i Figur 1 er anført "ramified hierarchy of typological components" for at kunne henvise til Russels typebeskrivelse.

Den første principale komponent, som forklarer lidt over 50% af variansen, bliver kaldt den generelle intelligensfaktor, hvor alle seks test (A, B, C, D, E og F) er positivt korrelerede, hvad vi allerede vidste fra resultatet i Tabel 1. Men som det fremgår af Tabel 2, angives dette nu mere præcist ved hjælp af faktor loadings som blot angiver korrelationen mellem de enkelte test og selve komponenten.

Den næste principale komponent er bi-direktorial, som det fremgår af Tabel 2, idet item A, B og C har positive loadings, medens item D, E og F har negative loadings.

Loadings er således beslægtet med korrelationskoefficienter og varierer mellem -1.0 og 1.0.

Det Teasdale herefter viser er, at såfremt man foretager en egentlig eksplorativ faktoranalyse med rotation, så opnår man blot det resultat som ses i Tabel 3. Herved består den roterede

faktor 1 af A, B og C med høje (signifikante) loadings. Den næste roterede faktor 2 består af D, E og F med høje signifikante loadings, dvs. loadings højere end 0.30. Den eksplorative faktoranalyse er statistisk med "signifikante" loadings, medens PCA blot på god matematisk vis direkte angiver fortegnet på loadings. Denne faktoranalytiske rotation har alene sikret, at alle loadings bliver positive!

Mange opfatter resultatet af denne PCA analyse således, at den første principale komponent "måler" det generelle begavelsesniveau, fordi alle seks item eller test har positive loadings. Russels typebeskrivelse er en god måde til at illustrere, at PCA ikke er en metode til at belyse det rent måletekniske.

Russel benyttede i sit eksempel dette, at være en typisk englænder. Hvis vi her antager, at den typiske englænder især har en sproglig begavelse, medens den typiske europæer fra kontinentet især har en ikke-sproglig begavelse, så nytter det, ifølge Russel, ikke at tage alle 6 test eller item (A, B, C, D, E, F) i betragtning, idet man da ofte vil se, at den typiske englænder har højt score på A og B, men ikke C, og lave scores på D, E og F og så bliver atypisk, hvis man bruger alle seks kriterier på at være en typisk englænder. Ifølge Russel må man et trin væk fra den 1. komponent og se på de verbale test eller item blandt de item i den næste komponent med positive loadings (A, B og C). Dette eksempel viser også, at summen af alle seks item eller test (A+B+C+D+E+F) ikke er et sufficient mål for intelligens.

Til evaluering af om total score af en samling test eller item er et sufficient mål for intelligens eller depression, må man foretage en item-response teori (IRT) analyse (se næste regneeksempel).

PCA er således en metode til at undersøge om nogle item i en skala både korrelerer med mange af de andre item i skalaen, men især om der foreligger en dual komponent med hvilken man er i stand til at klassificere eller typebeskrive snarere end at være i stand til at foretage en egentlig måling.

Inden for depression er typebestemmelse af item vigtig, når man skal klassificere de antidepressive præparater i de sederende versus non-sederende, og det måletekniske har betydning ved at evaluere den egentlige antidepressive virkning.

### Referencer:

Teasdale TW. Psykometriske aspekter af kvantitativ testning. I: Østergaard L (ed) Undersøgelsesmetoder i klinisk psykologi. København, Munksgaard 1992 pp 112-135

Russell B. My philosophical development. London, Routledge, 1956

Child D. The essentials of factor analysis. London, Continuum, 3rd Edition 2006.



**Tabel 1**  
**Korreltions-**  
**matrix**

Inter-korrelationskoefficienter af de 6 items A, B, C, D, E, F

	A	B	C	D	E	F
A	-					
B	0.62	-				
C	0.58	0.60	-			
D	0.31	0.29	0.28	-		
E	0.32	0.33	0.29	0.60	-	
F	0.30	0.31	0.29	0.63	0.59	-

**Tabel 2**

Faktorloadings for de to første principale komponenter

	A	B	C	D	E	F
Komponent 1	0.72	0.73	0.70	0.72	0.73	0.72
Komponent 2	0.45	0.47	0.48	-0.49	-0.43	-0.47

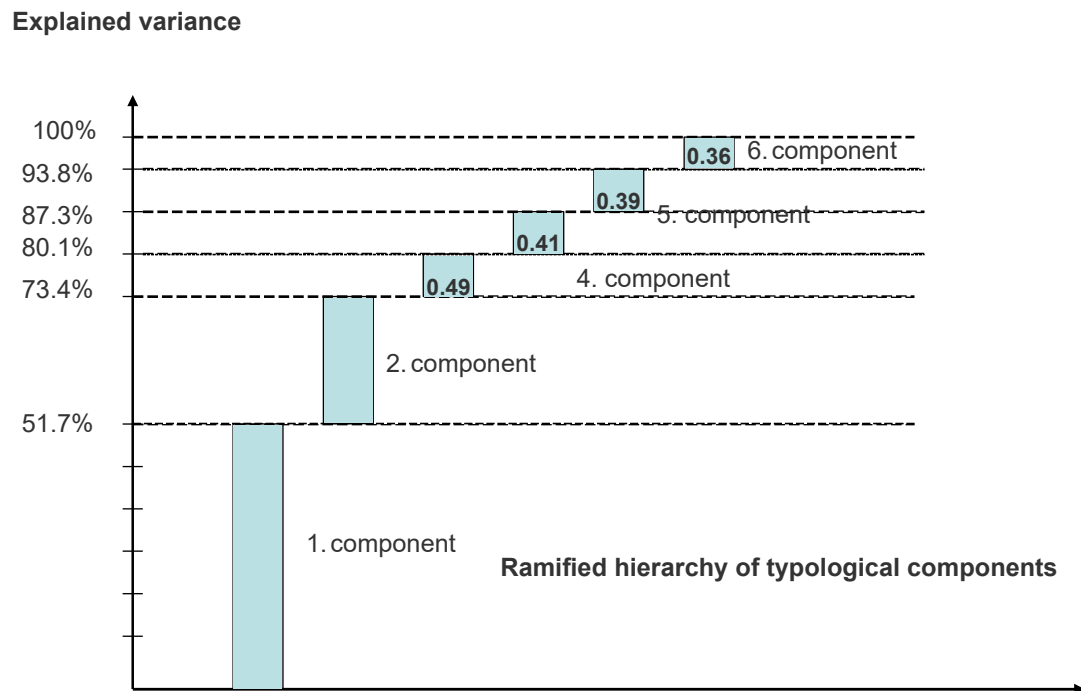
**Tabel 3**

Explorativ faktor-rotation

	Faktor loadings					
Roterede faktorer	A	B	C	D	E	F
Komponent 1	0.83	0.84	0.83	0.16	0.20	0.18
Komponent 2	0.19	0.19	0.16	0.85	0.82	0.84

Figur 1

De beregnede eigenvalues, f.eks. 3.1 for 1.komponent og de korresponderende procentværdier (explained variance)



For yderligere information:

**Bech, P. Klinisk psykometri,  
Munksgaard Danmark, 2011**