

## **Review:**

**Brigitte Le Roux & Henry Rouanet, *Multiple Correspondence Analysis*, London, Sage, 2010, 115 s.**

Jan Fredrik Hovden

*“Modellen må følge data, og ikke omvendt. Dette er ... et trist trekk trekk ved tilpassingen av matematikken til samfunnsvitenskapene; en overflod av a priori spesifiserte modeller, som deretter blir konfronterte med data gjennom såkalte tester. Snart tjener testen til å rettferdiggjøre en modell på et felt der det finnes flere parameter å tilpasse enn de en selv har fastsatt utifra data. Testen tjener til å strengt forkaste de klokeste innvendinger en forsker kan ha. Men det vi har bruk for er en rigorøs metode som er i stand til å trekke ut strukturer med utgangspunkt i data.”*

(Benzécri 1973:II:6).

Som sitatet fra Jean-Paul Benzécri, korrespondanseanalysens franske far<sup>1</sup> antyder, har den franske *Analyse des données*-tradisjonen en nokså annen tilnærming til statistisk analyse enn den rådende angloamerikanske tradisjonen. “Modellen må følge data, ikke omvendt” er ett av Benzécri’s kjente sitater, og kan illustrere den induktive filosofien som fundamental i MCA. I *Geometric Data Analysis* (Le Roux and Rouanet 2004:5), boken som *Multiple correspondence analysis* (2010) på mange måte er kortversjonen av, gjør Brigitte Le Roux og Henry Rouanet (sistnevnte døde dessverre i 2008) et nært beslektet skille mellom “multivariat analyse” som *kvantitativ, matrise- og utvalgsorientert*, og “geometrisk analyse” (GDA) som *geometrisk, formell og deskriptivt orientert*. Noen viktige differensierende prinsipp for GDA (der enkel og multipel korrespondanseanalyse - heretter benevnt som CA og MCA - og prinsippal-komponentanalyse - PCA - er de tre sentrale metodene innen tilnærmingen) er slik for det første en *geometrisk tilnærming* der data blir representert som punkt i et geometrisk rom (i motsetning til vanlig multivariat analyse, der tall er basis og slutt-

---

<sup>1</sup> Denne æren må selvsagt dels deles med andre, som f.eks Ronald Fisher hvis arbeider om optimal skalering fra 1940-tallet inneholder de grunnleggende formlene for enkel korrespondanseanalyse. For mer om metodens statistiske forhistorie, se Le Roux & Rouanet (2004:11).

produkt av prosedyrene), metodene er basert på *lineær algebra* (fremfor matrisealgebra)<sup>2</sup> og der den angloamerikanske tradisjonen har en "utvalgs-orientert" tilnærming til data og et medfølgende fokus på signifikanstesting, legger GDA vekt på at den *deskriptive analysen* kommer først), eller igjen med Benzécri sine ord: "Statistikk er ikke sannsynlighetslære!". Dette betyr *ikke* at MCA ikke har prosedyrer knytt til statistisk inferens eller ikke kan brukes som en konfirmerende (versus eksplorerende) teknikk, en vanlig misforståelse som vi skal komme tilbake til. Men kombinasjonen av den ubrutte linjen tilbake til Benzécri og Bourdieu-tradisjonen (førstnevnte var veileder på Le Roux sin doktoravhandling, og begge var nære medarbeidere av Bourdieu i hans senere år) og den konsekvente tydeliggjøringen av MCA som en analysemetode med tydelig andre underliggende prinsipper enn den vanlige angloamerikanske tradisjonen de fleste av oss er kjent med bidrar til å skille - og løfte opp – Le Roux og Roaunet sine arbeid om korrespondanseanalyse fra andre engelskspråklige innføringsbøker (se f.eks Greenacre 1984; Greenacre 2007; Clausen 2009).

Først litt historikk og en kort tilstandsrapport: Enkel korrespondanseanalyse (CA) - for analyse av en krystabell - ble etablert av Benzécri i 1963 og etterhvert kombinert med andre metoder og generalisert til en multivariat analysemetode (multipel korrespondanseanalyse – MCA) på tidlig syttitallet. Utover syttitallet levde korrespondanseanalysen i "storslått isolasjon" (s3) som en utbredt metode i fransk samfunnsforskning, men nesten ukjent i andre land. Fra midten på åttitallet blir imidlertid metoden mer kjent for et engelskspråklig publikum. Pierre Bourdieu (sammen med Monique de Saint Martin) publiserte sitt første arbeid der korrespondanseanalyse ble brukt i 1976 med "Anatomie du Goût", og med den engelskspråklige publikasjonen av *Distinksjonen* (1984) blir metoden for alvor kjent for et internasjonalt publikum samtidig som de første engelskspråklige lærebøkene dukker opp (Greenacre 1984; Lebart, Morineau et al. 1984). Rommet av sosiale posisjoner og livsstiler i *Distinksjonen* er fremdeles det mest kjente eksempelet på metoden (ironisk nok basert på en nokså "ustandard" fremgangsmåte for MCA<sup>3</sup>), og betydningen av Bourdieus arbeid for å utvikle metoden som sosiologisk redskap og å gjøre den kjent for et ikke-fransk publikum kan vanskelig overvurderes, noe som Le Roux og Roaunet også understreker. Her bør også nevnes Bourdieus understreking av den nære sammenhengen mellom korrespondanseanalysens logikk og den sosiale verdens logikk (spesielt står Bourdieus bruk av feltbegrepet fra syttitallet og fremover i et symbiotisk

<sup>2</sup> Betegnelsen "matrisealgebra" blir i matematisk teori vanligvis brukt om de mest grunnleggende matriseoperasjoner, som gange, deling, pluss/minus etc., mens "lineær algebra" vanligvis refererer til mere avanserte operasjoner knytt til vektorer, egenverdier, egenvektorer, dekomponering etc.

<sup>3</sup> Bourdieu sier selv i klartekst på s126 i den engelske utgaven av *Distinksjonen*: "The figures presented here are *not* plane diagrams of correspondence analyses, although various such analyses were drawn on in order to construct them..." (min uthevelse). Som det går fram av bokens appendiks 1 og 2, er kartet ikke et direkte resultat av en MCA på et enkelt datasett, men derimot resultatet av kombinerende av MCA-analyser av flere ulike datasett (inkludert to ulike surveyer gjort av Bourdieu selv i 1963 og 1967-68, en survey gjort av INSEE i 1967 og en gjort av det franske kulturministeriet i 1973 - begge to sistnevnte om kulturelle praksiser), og manuell innplassering av kategorier basert på ei nærlesing av annen statistikk. Kartet har slik også element av det Bourdieu med sin analyse av det politiske rommet i samme bok kaller et "provisorisk diagram" og et "teoretisk skjema" (Bourdieu 1984:452-3).

forhold til MCA).<sup>4</sup> Interessant nok er den franske geometriske tilnærmingen til statistisk analyse etter Le Roux og Rouanets vurdering fremdeles “nokså isolert” innen feltet av multivariat statistikk, og for MCA sin del er situasjonen også “nokså dårlig”, da metoden “er fremdeles sjeldent diskutert og derfor underbrukt” (s3). *Multiple Correspondence Analysis* med sin tilgjengelige men samtidige grundige form burde slik ha potensiale til å gi metoden større utbredelse og samtidig oppklare en del misforståelser om metoden.

Før gjennomgangen av bokens innhold er det på plass med en advarsel: Undertegnede er sosiolog og ikke statistiker. Selv om jeg gjennom bruk av MCA gjennom 15 år har opparbeidet en elementær forståelse av den statistiske basis for analysemetoden, kan jeg ikke vurdere det statistiske håndverket i *Multiple Correspondence Analysis*. Gitt forfatterene og det faktum at boken er utgitt på SAGEs kvalitetsbevisste serie “Quantitative Applications in the Social Sciences” er det imidlertid ingen grunn til å tro at dette noe annet enn uklanderlig. Derimot har jeg en del erfaring med MCA som et praktisk, sosiologisk verktøy brukt på “real-life data”, noe som passer bra da Le Roux og Rouanet presenterer sin lille bok som en *ikke-teknisk* og forskningsrettet introduksjon til multipel korrespondanseanalyse. Om det store omfanget av formler og matematiske utlegninger i boken av den jevne samfunnsviter vil bli oppfattet som ikke-teknisk er nok tvilsomt, men det er klart at Le Roux og Rouanet lykkes på forbildelig vis å i et kort format (115s) formidler både det matematiske grunnlaget og den praktiske logikken i metoden på en måte som burde være tilgjengelig for de fleste med et minimum av statistisk grunnlag og interesse.

Boken er inndelt i seks kapitler. *Introduksjonskapittelet* drøfter kort MCAs hovedprinsipp og historikk som dels skissert innledningsvis, i tillegg til å gi en smakebit på fremgangsmåten og hovedbegrepene i analysemetoden gjennom et praktisk eksempel om kulturpreferanser i Storbritannia (Le Roux, Rouanet et al. 2008). Det grunnleggende resultat fra MCA er to geometriske rom (eller punktskyer, fr. *nuage*), rommet av individer og rommet av kategorier (eller “rommet av egenskaper”, som Bourdieu ofte kalte det). *Kapittel 2* introduserer individskyens geometri, og med dette også det matematiske grunnlaget for utregninger av skyens *inertia* (eller varians), avstander mellom punkt, punkts relative og absolutte bidrag og prinsipalaksene, mens *kapittel 3* viderefører dette i en introduksjon til de sentrale prinsipper, begrep og statistikker i MCA. Den påfølgende eksemplifiseringen av metoden gjennom en detaljert analyse og tolking av tidligere nevnte smaksdata er ett glimrende *exemplum* som mange nok vil finne svært oppklarende. Det er også bra at boken både dekker MCA av indiktor-matriser og Burt-tabeller, da dette ofte fremstår som forvirrende alternativ for de som skal gjøre slike analyser for første gang. Kapittelet drøfter også kort spesifikk MCA (SpeMCA) og klassespesifikk MCA (CSA). Førstnevnte er rett og slett “real-life” MCA

---

<sup>4</sup> “I use Correspondence Analysis very much, because I think that it is essentially a relational procedure whose philosophy fully expresses what in my view constitutes social reality. It is a procedure that ‘thinks’ in relations, as I try to do it with the concept of field.” (Bourdieu and Kraus 1991). For mer om Bourdieus forhold til korrespondanseanalysen se (Le Roux, Rouanet et al. 2000).

på data der uønskede kategorier - som kategorier med relativt svært få individ (<5%), "vet ikke"-kategorier og manglende svar - blir tatt ut av analysen som aktive punkt (som påvirker rommets konstruksjon) og blir gjeninnført som passive punkt - også ofte kalt *supplementærpunkt* (som ikke gjør det). Klassespesifikk MCA er derimot enkelt sagt en MCA analyse av et subsett av individene i det originale datasettet, en fremgangsmåte som Bourdieu viser i sine kart over "smaksvariasjoner" hos henholdsvis de dominerende fraksjoner og småborgerskapet i *Distinksjonen* (Bourdieu 1984:262,340).

*Kapittel 4* omhandler *strukturert dataanalyse*. Utgangspunktet er at vi vanligvis har mer informasjon om individene enn det er fornuftig å bruke som aktive punkt i MCA, da analysemetoden forutsetter en viss homogenitet i kategoriene (s 11) som blir brukt til å konstruere rommet ("All in all, doing a Correspondence Analysis is nothing more than diagonalizing a matrix. The point is just to pick up the right matrix to diagonalize" – Benzécri). Vanligvis ønsker en derfor vanligvis å bruke noe av denne ekstra informasjonen (som Le Roux og Rouanet logisk nok, men litt forvirrende kaller "strukturende faktorer") – f.eks kjønn – til å supplere tolkingen av kartet, vanligvis i form av supplementærpunkt<sup>5</sup>. Slike punkt har imidlertid en viktig begrensing: de viser oss bare den *gjennomsnittlige* posisjonen til en kategori, ikke *spredningen* (subskyen) til individene som har dette kjennemerket. Dette er en begrensing Bourdieu selv pekte på i de ovennevnte analyser av smakens variasjoner i de dominerte klasser og småborgerskapet, og et problem han forsøkte løse ved å tegne bokser i kartet som illustrerte spredningen av individene til ulike yrkeskategorier i rommet (Le Roux, Rouanet et al. 2000:3). Av denne årsak er bruk av supplementærvariabler i MCA bare det "første steg" i en strukturert dataanalyse (s68), og for en mer nyansert analyse har Le Roux og Rouanet senere utviklet et utmerket - om fremdeles lite utbredt<sup>6</sup> - verktøy for dette, *konsentrasjonsellipser*, som antyder subskyens geometri i kartet. Hoveddelen i kapitlet er derfor viet disse ellipsene.<sup>7</sup>

*Kapittel 5* tar oppgjør med en nevnt vanlig misforståelse om korrespondanseanalyse, nemlig at dette primært er en induktiv/beskrivende teknikk og ikke egner seg til *statistisk inferens*. Selv om Benzécri som sitatet innledningsvis illustrerer var svært skeptisk til *test*-mentaliteten som underliggjer mye av angloamerikansk-statistikk, er metoder og statistikker for slik inferens etterhvert blitt en naturlig del av MCA sin verktøykasse, f.eks tester om en kategori (Lebart, Morineau et al. 1995:181-4) har en plassering signifikant ulik nullpunktet utgjort av en prinsipalakse. I GDA er

<sup>5</sup> I studier av livsstiler bør en for eksempel ikke blande indikatorer på sosiale posisjoner og indikatorer for livsstiler som aktive punkt i samme analyse, men enten lage et rom av sosiale posisjoner (utifra kapitaltilganger) og så projisere inn livsstilene som supplementærvariabler – eller omvendt. Det beste er selvsagt å gjøre begge deler - det som Lebart og Morineau (1984:111) kaller for en "resiprok" analyse.

<sup>6</sup> Dette skyldes ikke bare uvitenhet fra de praktiserende, men også at slike ellipser kun er tilgjengelig i fullform i programpakken SPAD ([www.cisia.fr](http://www.cisia.fr)), og delvis i R ([www.r-project.org](http://www.r-project.org)), og ikke i mer utbredte program som SPSS eller STATA.

<sup>7</sup> Siste delen av kapitlet som omhandler "additive" subskyer (der subskyer definert av to eller flere egenskaper) og utregningen av "struktureffekter" overgikk imidlertid over min forstand, og kunne trolig tjent på en klarere motivering og substansiell tolking av eksempelet.

imidlertid slike tester alltid sekundære den statistiske beskrivelsen (“Beskrivelse først, inferens senere!” – Benzécri) og innlemmet i et *kombinatorisk rammeverk* som er mer i harmoni med metodens filosofi og geometriske basis. Kapittelet presenterer både typikalitet- og homogenitet-tester, der førstnevnte omhandler i hvilken grad en subskye er atypisk for populasjonen den er trukket fra, mens sistnevnte type test er rettet om spørsmålet om to subskyer/kategorier har en signifikant ulik plassering en kryssende prinsipalakse eller en annen subskye/kategori i planet, jfr. eksemplet med Lebart et. al ovenfor. Le Roux og Rouanet demonstrerer også hvordan konfidensellipsene kan brukes til slike tester, i form av å tegne konfidenssoner rundt rommets geometriske gjennomsnitt (s 89). *Siste kapittel* gir to opplysende eksempler på mer avansert bruk av MCA ved en gjennomgang av Bourdieus siste publiserte arbeid med bruk av metoden (“The Field of Publishers in France” (Bourdieu 1999) og en analyse av det norske maktfeltet (Hjellbrekke, Korsnes et al. 2007). Appendikset har også instruktive råd om egnet programvare til MCA.

Vi omtalte innledningsvis den induktive filosofien som fundamental i MCA, med vekt på beskrivelse av strukturer fremfor modellering, og knytt til dette finnes et annet særtrekk med metoden sammenlignet med mange andre multivariate metoder: dens relative modellfrihet. I motsetning til f.eks enkel faktoranalyse forutsetter MCA f.eks ikke at dataene er normalfordelte eller lineært relaterte. Selv om Benzécri utsagn om at korrespondanse-analysen gjør det mulig å skue direkte inn i Guds skaperverk<sup>8</sup> er i overkant sterk, er MCA et verktøy som i praksis er svært sensitivt for de underliggende strukturene i data, og dette skaper både muligheter (originale, nyanserte og innsiktsfulle analyser) og utfordringer (*really screwing things up*) for forskeren.<sup>9</sup> Heldigvis gir Le Roux og Rouanet i boken som nevnt ikke bare utmerkede eksempler på korrespondanseanalyse i praksis ved steg-for-steg analyser av flere ulike datasett, men også noen praktiske tommelfingerregler for å hindre de vanligste feilene (som behovet for å utelate kategorier under en viss relativ størrelse og å sikre at variablene i analysen er balanserte ved å ikke ha alt for ulikt antall kategorier).

Fra egen erfaring med slike analyser skulle jeg imidlertid gjerne sett at boken var utvidet med enda flere slike håndverkspraktiske råd. Først og fremst tror jeg mer om hvordan man skal tolke karakteristiske mønstre i kategori- og individskyen – hva de sier oss om våre data – ville være til uvurderlig hjelp for de med lite erfaring med metoden, da disse strukturene kan informere oss om alvorlige problemer med de valgte variablene (som hestesko/Guttman-effekten som oppstår når det eksisterer et kvasi-funksjonelt forhold mellom disse) og behov for å gjøre separate analyser av underutvalg av individ (se f.eks Volle 1993:155-161). Men det er selvsagt en grense for hvor mye en kan ta med på 115 sider. For den jevne samfunnsforsker gir *Multiple*

---

<sup>8</sup> “The essence of things created by God” (van Meter, Schiltz et al. 1994:129), også referert av Michael Greenacre i kurs i korrespondanseanalyse March 2002 på Zentralarchiv für Empirische Sozialforschung, Universität zu Köln.

<sup>9</sup> Sistnevnte skal også være noe av grunnen til at MCA møtte en del motstand fra statistikere i Frankrike den første tiden (van Meter, Schiltz et al. 1994:133).

*Correspondence Analysis* en svært vellykket og etterlengtet kortintroduksjon til analysemetoden og filosofien bak.

For de som ønsker dykke tyngre ned i analysemetoden tilbyr Le Roux to andre gode ressurser: For det første boken *Geometric Data Analysis* (Le Roux and Rouanet 2004), som med sine 475 sider er det moderne standardverket innen MCA/GDA og har supplerende/utvidet informasjon om det aller meste som dekkes i *Multiple Correspondence Analysis*. For det andre har Le Roux en glimrende nettside på både fransk og engelsk (<http://www.math-info.univ-paris5.fr/~lerb/>), som i tillegg til supplerende materiale til boken (i form av fargeversjoner av kartene og nedlastbare datafiler fra noen av eksemplene) også har en FAQ for vanlige spørsmål knytt til bruk av korrespondanseanalyse.<sup>10</sup>

Jan Fredrik Hovden, Førsteamanuensis, Universitetet i Bergen,  
Forsker II, Høgskulen i Volda

## Referanser

- Benzécri, J. P. 1973. *L'analyse des données*. Paris, Dunod.
- Bourdieu, P. 1984. *Distinction : a social critique of the judgement of taste*. London: Routledge & Kegan Paul.
- Bourdieu, P. 1999. Une révolution conservatrice dans l'édition. *Actes de la recherche en sciences sociales* 126-127: 3-28.
- Bourdieu, P. & M. de Saint Martin 1976. L'Anatomie du goût. *Actes de la recherche en sciences sociales* 5: 2-212.
- Bourdieu, P. & B. Krais 1991. 'Meanwhile, I have come to know all the diseases of sociological understanding'. In *The Craft of Sociology*. P. Bourdieu, J.-C. Chamboredon & J.-C. Passeron, 247-259. Berlin: de Gruyter.
- Clausen, S.-E. 2009. *Multivariate analysemetoder for samfunnsvitere*. Oslo: Universitetsforlaget.
- Greenacre, M. 1984. *Theory and applications of correspondence analysis*. London: Academic Press.
- Greenacre, M. J. 2007. *Correspondence analysis in practice*. London: Chapman & Hall.
- Hjellbrekke, J., O. Korsnes, et al. 2007. The Norwegian Field of Power anno 2000. *European Societies* 9 (2): 245-273.
- Le Roux, B. & H. Rouanet 2004. *Geometric Data Analysis*. Dordrecht: Kluwer Academic Publishers.
- Le Roux, B. & H. Rouanet 2010. *Multiple Correspondence Analysis*. London: Sage.
- Le Roux, B., H. Rouanet, et al. 2000. The Geometric Analysis of Questionnaires: The Lesson of Bourdieu's La Distinction. *Bulletin de Méthodologie Sociologique* 65: 5-15.
- Le Roux, B., H. Rouanet, et al. 2008. Class and cultural division in the UK. *Sociology* 42: 1049-1071.
- Lebart, L., A. Morineau, et al. 1995. *Statistique exploratoire multidimensionnelle*. Paris: Dunod.
- Lebart, L., A. Morineau, et al. 1984. *Multivariate descriptive statistical analysis correspondence analysis and related techniques for large matrices*. New York: Wiley.
- van Meter, K. M., M.-A. Schiltz, et al. 1994. Correspondence Analysis: A History and French Sociological Perspective. In *Correspondence analysis in the social sciences*. M. Greenacre & J. Blasius. London: Academic Press.
- Volle, M. 1993. *Analyse des données*. Paris: Economica.

<sup>10</sup> Når dette ble skrevet inneholdt denne FAQ bare et spørsmål og svar, men Le Roux har lovet å oppdatere denne etterhvert.