Om en ny type statistiske modeller og deres anvendelsesmuligheder

Af ERLING B. ANDERSEN*)

1. Om statistiske modeller

Ifølge en udbredt opfattelse er en statistiker en person, til hvem man sender sine data for at få vurderet holdbarheden af visse hypoteser, man har opstillet. I mange tilfælde bekymrer man sig kun lidt om, hvilke metoder statistikeren finder det rimeligt at benytte. Ofte føler man måske, at man ikke rigtig har mulighed for at kontrollere, hvad det egentlig er, han gør ved tallene. For en statistiker er denne opfattelse ikke særlig tilfredsstillende. Og man kan vist roligt tilføje, at klienten måske også ville være nervøs ved fremgangsmåden, såfremt han vidste, hvor hjælpeløs statistikeren ofte er konfronteret med et givet datasæt. De seneste års udvikling har da også vist, at statistikerne beskæftiger sig mere og mere med selve grundlaget for dataanalysen: modelformuleringen. Her hverken kan eller ønsker statistikeren at være enerådende, idet to hensyn af lige stor vigtighed skal tilgodeses. På den ene side skal modellen i specifik matematisk form indeholde de faktorer, som klienten ønsker data skal belyse. På den anden side må modellen udformes, så det rent statistisk metodisk kan lade sig gøre at efterprøve klientens hypoteser på det aktuelle datasæt. Når en model er opstået som følge af et sådant teamwork er der to muligheder: Modellen kan være kendt og undersøgt i den statistiske litteratur, og man behøver da blot at anvende de i litteraturen udviklede analysemetoder, når data skal analyseres.

*) Cand. stat., amanuensis ved Handelshøjskolen i København.

Om en ny type statistiske modeller og deres anvendelsesmuligheder

Af Erling B. Andersen*)

1. Om statistiske modeller

Ifølge en udbredt opfattelse er en statistiker en person, til hvem man sender sine data for at få vurderet holdbarheden af visse hypoteser, man har opstillet. I mange tilfælde bekymrer man sig kun lidt om, hvilke metoder statistikeren finder det rimeligt at benytte. Ofte føler man måske, at man ikke rigtig har mulighed for at kontrollere, hvad det egentlig er, han gør ved tallene. For en statistiker er denne opfattelse ikke særlig tilfredsstillende. Og man kan vist roligt tilføje, at klienten måske også ville være nervøs ved fremgangsmåden, såfremt han vidste, hvor hjælpeløs statistikeren ofte er konfronteret med et givet datasæt. De seneste års udvikling har da også vist, at statistikerne beskæftiger sig mere og mere med selve grundlaget for dataanalysen: modelformuleringen. Her hverken kan eller ønsker statistikeren at være enerådende, idet to hensyn af lige stor vigtighed skal tilgodeses. På den ene side skal modellen i specifik matematisk form indeholde de faktorer, som klienten ønsker data skal belyse. På den anden side må modellen udformes, så det rent statistisk metodisk kan lade sig gøre at efterprøve klientens hypoteser på det aktuelle datasæt. Når en model er opstået som følge af et sådant teamwork er der to muligheder: Modellen kan være kendt og undersøgt i den statistiske litteratur, og man behøver da blot at anvende de i litteraturen udviklede analysemetoder, når data skal analyseres.

*) Cand. stat., amanuensis ved Handelshøjskolen i København.

Lige så ofte opstår der imidlertid en model, som er ny eller lidet undersøgt. Det kan ske, at man ved visse tilnærmelser kan bringe modellen over i kendte modeller. Men i reglen må man selv i gang med at udvikle nye analysemetoder. Dette kan give anledning til en helt ny forskning på den teoretisk statistiske front. Her er det ikke mindst vigtigt, at modellerne generaliseres i deres rent statistiske formulering, idet der herigennem åbnes mulighed for anvendelse over for data, der stammer fra helt andre fagområder.

En vigtig fase af dataanalysen er modelkontrollen, hvorved forstås en konfrontation af model og data. Såfremt data og model ikke er konsistente, vil man ofte opleve, at data så at sige anviser, hvorledes modellen kan reformuleres for at være konsistent med de aktuelle data. Efter reformuleringen må hele processen selvfølgelig gennemføres igen.

Vi kan altså opstille følgende skematiske oversigt over en statistisk modelopstilling:

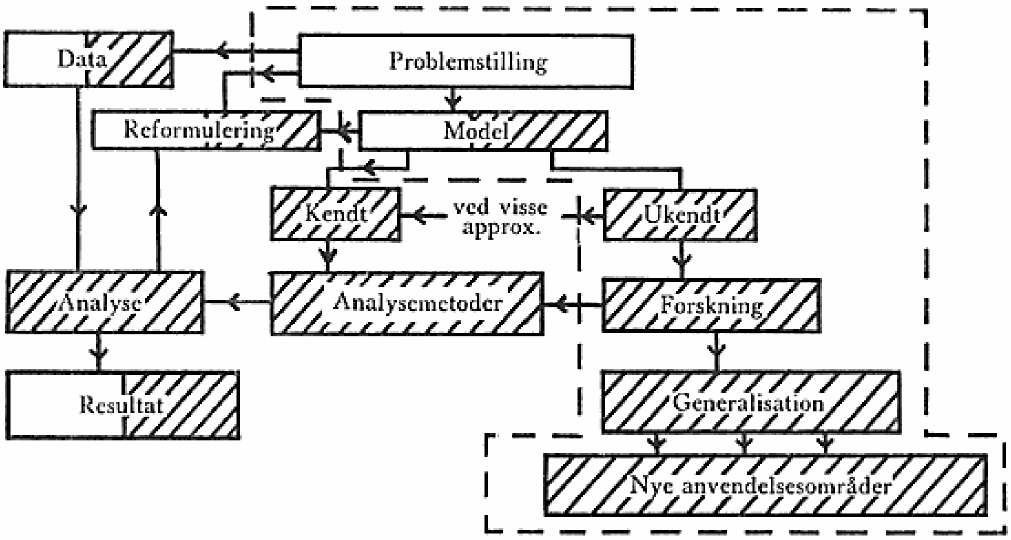

Fig. 1. Den statistiske modelformuleringsproces.

Fig. 1.

Fig. 1 kan selvfølgelig kun i hovedtræk beskrive den indviklede proces, som fører til formuleringen af en statistisk model, men kan måske bidrage til at fastholde hovedsynspunktet. Vi skal i denne artikel beskæftige os med de indrammede procesdele i forbindelse med omtalen af en type modeller, der er opstået i forbindelse med den statistiske analyse af en vis type data, der hyppigt forekommer inden for psykologiske og sociologiske undersøgelser.

Ved formuleringen og diskussionen af modellen vil vi stedse referere til den psykologiske problemstilling. Det sker dog kun af pædagogiske grunde. Alle betragtninger er generelle, og man vil let kunne oversætte sprogbrugen til andre anvendelsesområder.

2. *Den tosidige observationssituation*

Vi skal beskæftige os med observationer, der er resultatet af et samspil mellem to faktorer. Som typiske sådanne tosidige situationer kan nævnes:

Eks. 1. Udbyttet ved dyrkning af en kornart afhænger på den ene side af den dyrkede marks kvalitet eller bonitet, på den anden side af mængden eller formen, hvori kunstgødning tilsættes, f. eks. i hvilken kemisk forbindelse eller kemisk koncentration et vigtigt grundstof tilsættes.

Eks. 2. Antal trafikulykker for buschauffører i London afhænger på den ene side af den enkelte chaufførs kørefærdighed og på den anden side af trafikforholdene på chaufførens rute.

Listen kan fortsættes ad libitum. Som eksempel skal vi i det følgende benytte en situation over for hvilken den model vi skal betragte først blev anvendt af Rasch, 1960.

Eks. 3. En række personer udsættes for opgaver af varierende sværhedsgrad i forbindelse med en intelligenstest. Resultatet, målt i antal korrekte besvarelser, afhænger på den ene side af personens intelligensniveau eller dygtighed og på den anden side af opgavernes sværhedsgrad.

3. *Nogle hovedtyper af modeller.*

Vi skal først omtale nogle hovedtyper af modeller.

Ved en parametrisk model forstås en model, hvor de størrelser eller relationer, som skal undersøges, kan udtrykkes med matematiske konstanter. Hvis man skal studere, hvorledes et vist sagsforhold i gennemsnit tager sig ud, er det naturligt at vælge en model, hvor der optræder en konstant μ, der angiver middelværdien. Hvis man skal undersøge, om to faktorer varierer uafhængigt, er man i salveten, hvis uafhængighed inden for modellen er ensbetydende med, at en vis konstant ϱ har værdien 0. Sådanne matematiske konstanter kaldes parametre og er definitionsmæssigt ukendte konstanter, der bestemmer modellens udseende. I en parametrisk model for den tosidige observationssituation har vi to parametre: Θ, som er knyttet til den ene faktor, og ε, som er knyttet til den anden faktor. Det statistiske problem består i at sige noget om Θ og ε ud fra observationerne. I det følgende betragter vi kun parametriske modeller.

Det er naturligt at skelne mellem deterministiske og stokastiske modeller: Antag at vi kender værdien af parameteren for hver af de to faktorer, som påvirker observationen. Det vil f. eks. sige, at vi kender en persons intelligensniveau og en opgaves sværhedsgrad. Hvis vi har en deterministisk model, vil disse to parameterværdier bestemme observationen, det vil sige, at vi helt præcist kan forudsige resultatet. En sådan model vil være ensbetydende med en meget stærk konsekvens i personers løsning af intelligensprøver. Et enkelt blik på besvarelser af intelligensprøver modsiger en sådan opfattelse. Når vi arbejder med en stokastisk model, bestemmer værdien af de to parametre kun sandsynlighedsfordelingen for observationen, i eksempel 3 altså sandsynlighederne for korrekt og forkert løsning af opgaven. Observationer, som er underkastet en sandsynlighedslov, vil fremtræde med et vist tilfældigt præg. Det vil dog stadig være sådan, at afhængigheden af de to faktorers parametre inducerer en vis struktur i observationerne, som kan udnyttes til at bestemme parameterværdierne.

For at udnytte det tosidige i observationssituatioen vil det ofte være nødvendigt at skaffe sig et tosidigt observationsskema på den måde, at hver faktors parameterværdi fastholdes, mens den anden faktors parameterværdi varieres. I nogle situationer er dette helt naturligt tilfældet: I eks. 3 besvarer hver person alle opgaver, og hver opgave besvares af alle personer. I andre situationer må man tilrettelægge et eksperiment, som frembringer et tosidigt skema. I eksempel 1 må man benytte samme gødningssammensætning på jorder af forskellig bonitet, og for hver bonitet må man anvende alle gødningssammensætninger. Dette er netop en dyrkningssituation, som tilstræbes ved mange dyrkningsforsøg. I eksempel 2 må man (i hvert fald i en forsøgsperiode) lade alle chauffører, som indgår i eksperimentet, gennemkøre forskellige strækninger med varierende trafikforhold og lade hver af strækningerne gennemkøre af alle chaufførere.

Vi kan således i mange situationer fastholde parameterværdien for den ene faktor, mens vi varierer den anden og vice versa.

Hvis vi har et tosidigt observationsskema, indgår der n værdier $\Theta_1, \ldots, \Theta_n$ af parameteren Θ svarende til første faktor og k værdier $\varepsilon_1, \ldots, \varepsilon_k$ af parameteren ε svarende til den anden faktor i modellen. For hver kombination af et Θ_i og et ε_j har vi en observation, som vi betegner x_{ij}. (Der er selvfølgelig intet i vejen for, at vi kan have flere observationer $x_{ij1}, \ldots, x_{ijN}$ for hver kombination (Θ_i,ε_j)).

Den stokastiske model specificeres nu fuldstændigt ved at angive sandsynlighedsfordelingen for x_{ij} som funktion af Θ_i og ε_j.

Fig. 2 viser et tosidigt observationsskema sammen med de indgående parameterværdier.

Faktor I	Faktor II					
	ε_1	. . .	ε_j	. . .	ε_k	Sum
Θ_1	x_{11}	. . .	x_{1j}	. . .	x_{1k}	$x_{.1}$
.	.		.		.	.
.	.		.		.	.
.	.		.		.	.
Θ_i	x_{i1}	. . .	x_{ij}	. . .	x_{ik}	$x_{i.}$
.	.		.		.	.
.	.		.		.	.
.	.		.		.	.
Θ_n	x_{n1}	. . .	x_{nj}	. . .	x_{nk}	$x_{n.}$
Sum	x_1	. . .	$x_{.j}$	. . .	$x_{.k}$	$x_{..}$

Fig. 2. Et tosidigt ovservationsskema.

Ved formulering af modellen er det afgørende, i hvilken form observationerne foreligger. De tre eksempler viser tre meget forskellige situationer. Høstudbyttet kan f. eks. måles i vægten (ell. rummålet) af det høstede korn pr. dyrket arealenhed. Variationsområdet for observationerne vil her udgøre et kontinum på den reelle talakse, hvorfor vi kalder observationerne kontinuerte. Variationsområdet for antal trafikuheld er alle positive hele tal. Vi kalder sådanne observationer numerable, idet de positive *hele* tal kan nummereres, hvad alle positive *reelle* tal ikke kan. Besvarelsen af opgaverne ved en intelligensprøve kan kun antage to former: »rigtig« eller »forkert«. Observationer hvis variationsområde kun omfatter et endeligt antal værdier vil vi betegne diskrete. Som det vil erindres fra elementære fremstillinger af sandsynlighedsregning må kontinuerte, numerable og diskrete observationer behandles forskelligt.

I den tosidige situation er teorien for kontinuerte observationer bedst kendt. Det er her den yndede variansanalysis, der bygger på normal fordelte observationer, finder anvendelse. Knap så kendt er teorien for numerable observationer. Den bedst kendte fordeling er poissonfordelingen, som i den tosidige situation fører til en type modeller, der ikke er stort vanskeligere at arbejde med end variansanalysemodellerne. Disse modeller er behandlet af Rasch, 1960 i forbindelse med analysen af nogle læseprøver.

Det kan vises, at de grundlæggende principper for variansanalysen og den tosidige poissonanalyse er de samme, som gælder for den diskrete modeltype, vi nu skal behandle indgående.

4. *Diskrete modeller.*

Vi betragter først tilfælde, hvor hvert af x_{ij}'erne kun kan antage to værdier, som vi for nemheds skyld repræsenterer ved tallene 0 og 1. Det kan let indses, at det er uden betydning for modellen, hvorledes de to tal vælges. For at opstille en statistisk model til beskrivelse af observationer må vi specificere, hvorledes $p_{ij} = Pr\{x_{ij} = 1\}$ afhænger af Θ_i og ε_j. ($Pr\{x_{ij} = 0\}$ fås som $1-p_{ij}$!). Vi tænker os nu, at Θ_i og ε_j er reelle positive tal og stiller det rent statistiske krav, at modellen skal have en form, der tillader, at de to sæt parametre $\Theta_1, \ldots, \Theta_n$ og $\varepsilon_1, \ldots, \varepsilon_k$ kan estimeres på enkel måde med udgangspunkt i observationerne. Det vil føre for vidt her at omtale, hvorledes dette krav kan formuleres helt præcist i statistiske termer og at bevise at modellen deduceres fra dette krav. I stedet skal vi angive den meget simple model, der som den eneste tilfredsstiller vort krav, og vi skal kort omtale, hvorledes estimationen kan udføres på grundlag af den udledte tale, hvorledes estimationen kan kan udføres på grundlag af den udledte model.

Modellen har udseendet

$$(1) \qquad x_{ij} = \begin{cases} 1 & \text{med sandsynlighed } \Theta_i\varepsilon_j/(1+\Theta_i\varepsilon_j). \\ 0 & \text{–} \qquad \text{–} \qquad 1/(1+\Theta_i\varepsilon_j). \end{cases}$$

Modellen har nogle iøjnefaldende attraktive egenskaber. Hvis ε_j fastholdes, vokser sandsynligheden fra 0 til 1, når Θ_i vokser fra 0 til

$+\infty$, og tilsvarende for Θ_i. Fortolker vi i intelligensprøve-modellen Θ_i som en persons intelligensniveau, betyder det, at sandsynligheden for at løse en bestemt opgave korrekt vokser med voksende intelligensniveau. Modellen har også en invariansegenskab, som forekommer naturlig. Hvis vi multiplicerer Θ_i med en konstant og dividerer ε_j med samme konstant, ændres modellen ikke. Dette modsvarer, at en meget intelligent person kan have samme besvær med en svær opgave, som en ringe begavet person kan have med en let opgave.

Det kan vises, at rækkesummerne $x_{i.} = \sum_{j=1}^{k} x_{ij}$ og søjlesummerne $x_{.j} = \sum_{i=1}^{n} x_{ij}$ i observationsskemaet er tilstrækkelig (i statistisk terminologi: sufficiente) i statistisk forstand, når vi skal estimere Θ'ernes og ε'ernes værdi. Herved forstås, at af al den information observationerne indeholder om Θ_i'ernes og ε_j'ernes værdi er alene værdien af $x_{i.}$'erne og $x_{.j}$'erne relevant. Man kan imidlertid ikke (som det er tilfældet for variansanalysen) alene benytte $x_{.i}$'erne til estimation af Θ_i'erne og alene $x_{.j}$'erne til estimation af $\varepsilon_{.j}$'erne, hvorved de to estimationsopgaver kan løses uafhængigt af hinanden. Det primære ved estimationen af Θ_i'erne er stadig $x_{i.}$'erne, men vi er nødt til at inddrage $x_{.j}$'erne i estimationen som en slags hjælpevariable. Man kan anskueliggøre denne estimationsmetodik på følgende måde:

Antag, at en person har parameter Θ_i. Som udgangspunkt for en estimation af Θ_i har vi det totale antal rigtigt løste opgaver $x_{i.}$. Nu er det selvfølgelig af en vis betydning, hvad det er for nogle opgaver han har opnået så mange rigtige svar på. Det ville altså være en fordel, om man kendte de pågældende opgavers sværhedsgrad $\varepsilon_1, \ldots, \varepsilon_k$. Det gør vi imidlertid ikke. Det viser sig imidlertid, at man når til en statistisk tilfredsstillende fremgangsmåde, hvis man i stedet inddrager de primære størrelser $x_{.1}, \ldots, x_{.k}$ for estimationen af $\varepsilon_1, \ldots, \varepsilon_k$. Ved estimationen af ε'erne er tilsvarende $x_{.j}$'erne de primære størrelser, mens $x_{1.} \ldots, x_{n.}$ inddrages som hjælpekonstanter. Estimationsprocedurer, der bygger på den her skitserede metode, er fuldt udarbejdede, og der er skrevet programmer med henblik på elektronisk behandling af vilkårlige datasæt.

Vi skal nu generalisere modellen (1) til tilfælde, hvor observationerne kan antage flere end to mulige værdier. Typiske eksempler er spørgeskemaer, hvor en række personer skal vurdere forskellige forhold, og som svar på spørgsmålene kan vælge blandt et fast antal muligheder f. eks.:

»udmærket«, »godt«, mindre godt«, »slet«. (Eks.: »Hvad synes De om chefen?«). For at specificere en statistisk model for observationerne må vi angive 3 sandsynligheder, nemlig

(2) $\quad p_{ij}^{(1)} = Pr\{x_{ij} = 1\},\ p_{ij}^{(2)} = Pr\{x_{ij} = 2\}$
og $p_{ij}^{(3)} = Pr\{x_{ij} = 3\}$

(Den fjerde sandsynlighed $Pr\{x_{ij} = 4\}$ fås som 1 minus summen af de tre andre sandsynligheder).

Antag igen, at Θ_i og ε_j er reelle tal. Vores opgave er da at angive de 3 sandsynligheder (2) som funktion af Θ_i og ε_j. For at løse denne opgave må vi inddrage endnu et forhold i modellen, den såkaldte scoring af svarmulighederne.

Ved en scoring af et antal svarmuligheder forstås valg af en måleskala, hvorpå svarmulighederne kan måles i forhold til hinanden. Hvis vi vælger den klassiske måleskala: den reelle tallinie, er scoringen ensbetydende med at angive den relative afstand mellem svarmulighederne. Lige stor relativ afstand betegner vi æqvidistant scoring, og vi kan f. eks. vælge tallene -1.5, -0.5, $+0.5$ og $+1.5$. Mere generelt kan scoringen med den reelle tallinie som måleskala udtrykkes gennem fire scoringsværdier φ_1, φ_2, φ_3 og φ_4. Den naturlige generalisation af (1) er da

$$(3) \qquad x_{ij} = r \text{ med sandsynlighed } (\Theta_i\varepsilon_j)^{\varphi_r} \Big/ \sum_{v=1}^{4} (\Theta_i\varepsilon_j)^{\varphi_v}$$

for $r = 1, 2, 3,$ og 4. Hvis man er i den heldige situation at kende de fire φ'er, kan denne model behandles helt parallelt til (1). En mere almindelig situation er dog, at vi ikke kender φ'erne. Man kan da gå flere veje. Man kan gætte (mere eller mindre systematisk) på et sæt φ'er, f. eks. de nævnte æqvidistante og undersøge, om observationerne er konsistente med de valgte φ'er. Eller man kan opfatte φ'erne som parametre i fordelingen og estimere dem på linie med ε'erne og Θ'erne. En nødvendig forudsætning for disse fremgangsmåder er imidlertid, at måleskalaen for scoringen er kendt.

Modellen åbner mulighed for at foretage scoringen i relation til mere komplicerede måleskalaer. Det kan f. eks. tænkes, at φ'erne må opfattes som vektorer $\varphi_1 = (\varphi_{11}, \ldots, \varphi_{1m}) \ldots, \varphi_4 = (\varphi_{41}, \ldots, \varphi_{4m})$, hvorved måle-»skalaen« bliver et m-dimensional koordinatsystem. Modellen (3) vil få samme udseende, men også Θ_i'erne og ε_j'erne bliver nu m-dimensionale vektorer, og leddet $(\Theta_i\varepsilon_j)^{\varphi_v}$ må opfattes som

$$(\Theta_{i1}\varepsilon_{j1})^{\varphi_{v1}}\ (\Theta_{i2}\varepsilon_{j2})^{\varphi_{v2}} \ldots (\Theta_{im}\varepsilon_{jm})^{\varphi_{vm}}$$

Forskellen på én-dimensional og fler-dimensional scoring kan anskueliggøres ved en forsøgsvis fortolkning af, hvad der kan ligge bag de forskellige scoringer. En én-dimensional scoring kan betyde, at personerne opfatter svarmulighederne som 4 niveauer på en skala, der spænder fra en klar positiv til en klar negativ holdning til det stillede spørgsmål. Hvis vi er nødt til at anvende en 2-dimensional scoring, kan det betyde, at personerne opfatter de 4 svarmuligheder på den måde, at der først skal vælges mellem en positiv eller en negativ tilkendegivelse, og uafhængigt heraf vælger man, om man vil benytte den mere ekstreme eller den mindre ekstreme mulighed.

I mange tilfælde må man lade spørgsmålet om scoringens dimension stå åbent i starten af analysen. Modellen er velegnet til en sådan fremgangsmåde, fordi modellen, når der ses bort fra den måde leddet $(\theta_i \varepsilon_j)^{\varphi v}$ opfattes på, er den samme i det én-dimensionale og det fler-dimensionale tilfælde.

Denne fremgangsmåde er blevet anvendt ved analysen af nogle psykologiske spørgeskemaer, hvor dimensionsproblemet netop er meget centralt (jvf. Erling B. Andersen, 1964 og 1966).

5. *Erhvervsøkonomiske anvendelser.*

Som nævnt i indledningen er stokastiske modeller i deres rent statistiske formulering ikke knyttet til anvendelser inden for et bestemt fagområde. Det er således ikke svært at finde andre fagområder end psykologien og sociologien, hvor de modeller, som er omtalt i denne artikel, kan finde anvendelse. Vi skal specielt betragte mulige anvendelser inden for det erhvervsøkonomiske fagområde. Læseren vil måske finde, at de to eksempler, som er angivet nedenfor, ikke er særlig velvalgte, måske ikke engang særlig realistiske. Det er imidlertid forfatterens håb, at de to eksempler kan bidrage til at anskueliggøre de observationssituationer, over hvilke modellerne kan anvendes, og herigennem inspirere læseren til at foreslå andre anvendelser.

Eks. 4: De skal sælge en mærkevare (f. eks. et vaskemiddel) og ønsker at få noget at vide om på den ene side, hvor attraktive forskellige af mærkevarens egenskaber er, og på den anden side vil De godt vide noget om hvilke kunder, der synes om varen, således at De kan placere de interesserede kunder f. eks. geografisk eller socialt. De sender derfor en udvalgt kreds af kunder varen, og efter en vis prøvetid foretager De et rundspørge. Hver kunde konfronteres med forskellige egenskaber ved varen (vaskeevnen, skåner den husmoderens hænder, tøjets hvidhed, emballagen, den indlagte konkurrence etc.). For hver egenskab kan kunden afgive en af fire mulige vurderinger, »meget tilfreds«, »nogenlunde tilfreds«, »ikke helt tilfreds«, »utilfreds«. Deres arbejdshypotese er, at hvis kunden er interesseret, vil vedkommende være tilbøjelig til at reagere tilfreds på alle egenskaberne og vice versa. Hvis denne hypotese er korrekt, skulle modellen være velegnet til at analysere Deres data, og De vil kunne estimere på den ene side den enkelte kundes tilfredshedsgrad med varen som helhed; og samtidig hvor begejstrede kunderne som helhed er for de forskellige egenskaber, der er spurgt om.

Eks. 5: De står foran at skulle introducere Deres virksomheds produkter på nye markeder, f. eks. i visse u-lande. De er nødt til at vælge, hvor De vil sætte ind med salgskampagner, hvorfor De ønsker at vide noget om mulighederne på de forskellige markeder. De har i Deres virksomhed en række konsulenter, som De lader foretage en analyse af de mulige markeder. Hver konsulent skal for hvert marked afgive en vurdering, f. eks. i form af en udtalelse, der kan tage formerne: »gode muligheder«, »ret gode muligheder«, »kun nogenlunde muligheder«, »ringe muligheder«. De ved, at Deres konsulenter vurderer meget forskelligt, nogle er meget dristige og optimistiske af natur, mens andre er mere forsigtige, og De ønsker at tage højde for dette forhold. Vi har igen en typisk tosidig situation, hvor hver observation afhænger af på den ene side konsulentens vurderingsniveau og på den anden side markedets reelle muligheder. Da observationerne endvidere er diskrete, vil den opstillede model være velegnet til at foretage en vurdering af de forskellige markeder, der er renset for konsulenternes forskellige vurderingsniveau.

Referencer:

1. G. Rasch: Probalistic models for some intelligence and attainment tests. Danm. Pæd. Inst. 1960.

2. Erling B. Andersen: Vurdering af et psykologisk spørgeskema på grundlag af en sandsynlighedsteoretisk målingsmodel. Militær Psykologisk Tjeneste 1964.

3. Erling B. Andersen: Den diskrete målingsmodel af endelig orden. Københavns Universitet 1966.

ved varen (vaskeevnen, skåner den husmoderens hænder, tøjets hvidhed, emballagen, den indlagte konkurrence etc.). For hver egenskab kan kunden afgive en af fire mulige vurderinger, »meget tilfreds«, »nogenlunde tilfreds«, »ikke helt tilfreds«, »utilfreds«. Deres arbejdshypotese er, at hvis kunden er interesseret, vil vedkommende være tilbøjelig til at reagere tilfreds på alle egenskaberne og vice versa. Hvis denne hypotese er korrekt, skulle modellen være velegnet til at analysere Deres data, og De vil kunne estimere på den ene side den enkelte kundes tilfredshedsgrad med varen som helhed; og samtidig hvor begejstrede kunderne som helhed er for de forskellige egenskaber, der er spurgt om.

Eks. 5: De står foran at skulle introducere Deres virksomheds produkter på nye markeder, f. eks. i visse u-lande. De er nødt til at vælge, hvor De vil sætte ind med salgskampagner, hvorfor De ønsker at vide noget om mulighederne på de forskellige markeder. De har i Deres virksomhed en række konsulenter, som De lader foretage en analyse af de mulige markeder. Hver konsulent skal for hvert marked afgive en vurdering, f. eks. i form af en udtalelse, der kan tage formerne: »gode muligheder«, »ret gode muligheder«, »kun nogenlunde muligheder«, »ringe muligheder«. De ved, at Deres konsulenter vurderer meget forskelligt, nogle er meget dristige og optimistiske af natur, mens andre er mere forsigtige, og De ønsker at tage højde for dette forhold. Vi har igen en typisk tosidig situation, hvor hver observation afhænger af på den ene side konsulentens vurderingsniveau og på den anden side markedets reelle muligheder. Da observationerne endvidere er diskrete, vil den opstillede model være velegnet til at foretage en vurdering af de forskellige markeder, der er renset for konsulenternes forskellige vurderingsniveau.

Referencer:

1. G. Rasch: Probalistic models for some intelligence and attainment tests. Danm. Pæd. Inst. 1960.

2. Erling B. Andersen: Vurdering af et psykologisk spørgeskema på grundlag af en sandsynlighedsteoretisk målingsmodel. Militær Psykologisk Tjeneste 1964.

3. Erling B. Andersen: Den diskrete målingsmodel af endelig orden. Københavns Universitet 1966.