

Martin Rasmussen

**06:2006** ARBEJDSPAPIR

## Sammenligning af kvaliteten af kommunernes sagsbehandling

Forslag til metode

FORSKNINGSAFDELINGEN FOR SOCIALPOLITIK OG VELFÆRDSYDELSER

***Sammenligning af kvaliteten  
af kommunernes  
sagsbehandling  
Forslag til metode***

***Martin Rasmussen***

***Socialpolitik og velfærdsydelser  
Working Paper 06:2006***

The Working Paper Series of The Danish National Institute of Social Research contain interim results of research and preparatory studies. The Working Paper Series provide a basis for professional discussion as part of the research process. Readers should note that results and interpretations in the final report or article may differ from the present Working Paper. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including ©-notice, is given to the source.

## Om undersøgelsen

Ankestyrelsen og de regionale sociale nævn gennemfører regelmæssigt praksisundersøgelser, som har til formål at vurdere korrektheden af kommunernes afgørelser på det sociale og beskæftigelsesmæssige område. Det sker med henblik på at skabe ensartethed i kommunernes sagsbehandling og ligebehandling af borgerne. Ankestyrelsen har ønsket i højere grad at kvantificere resultaterne af disse praksisundersøgelser, blandt andet for at fremme den effekt som praksisundersøgelserne gerne skulle have på kommunernes sagsbehandlingskvalitet. Et led i denne kvantificering er sammenligning (benchmarking) af kvaliteten af kommunernes sagsbehandling. Ankestyrelsen har bedt Socialforskningsinstituttet bidrage med forslag til, hvordan kommunesammenligningen kan tilrettelægges.

I dette papir skitseres derfor metoder til kommunesammenligning. En empirisk sammenligning kan ikke foretages på nuværende tidspunkt, fordi de hidtidige praksisundersøgelser ikke indeholder nok observationer fra hver kommune. Også ved de fremtidige praksisundersøgelser bliver antallet af observationer mindre, end man kunne ønske sig til sammenligningen. Vi diskuterer, hvordan man kan øge den statistiske sikkerhed ved samtidig at analysere eksperternes vurdering af flere forskellige spørgsmål fra praksisundersøgelser i samme sammenligning. Vi foreslår også, at metoder til såkaldte 'multiple sammenligninger' bruges til dele af undersøgelserne. Det er en ældre og ikke særlig kendt metode, men den er særdeles relevant både i den konkrete anvendelse og i andre sammenhænge.

Arbejdet bag papiret er brugt som et af flere input til Ankestyrelsens mere omfattende rapport (Ankestyrelsen (2006)). Ankestyrelsens medarbejdere og specielt fuldmægtig Maria Feldfoss og fuldmægtig Britta Maar takkes for godt samarbejde.

## 1. Indledning

Ankestyrelsen og de regionale sociale nævn har gennem en årrække gennemført praksisundersøgelser af kvaliteten af kommunernes sagsbehandling. De består i, at eksperter fra fx Ankestyrelsen ser på kommunale sager, der er behandlet, og vurderer om afgørelserne og sagsbehandlingen er i overensstemmelse med loven. Hver praksisundersøgelse omfatter en konkret lovgivning, fx tildeling af førtidspension, og består af vurderinger af en række spørgsmål eller led i sagsbehandlingen. Formålet med praksisundersøgelser er (naturligvis) at øge kvaliteten i den kommunale sagsbehandling. Hidtil har det vigtigste virkemiddel været, at Ankestyrelsen på forskellig vis har informeret generelt om, hvor der opstår mange fejl.

For at øge effekten af praksisundersøgelserne vil Ankestyrelsen fremover lave sammenligning (benchmarking) af kommunerne. Meningen er – som med så megen anden benchmarking – at kommuner, der er relativt dårlige faktisk erkender det og derfor overvejer, hvad de kan gøre for at blive bedre. Det kunne muligvis ske ved at lære, hvordan de bedste kommuner gjorde.

Ankestyrelsens praksisundersøgelser er 'juridiske' i sin karakter og ikke direkte en undersøgelse af klienttilfredshed. Forskelle søges heller ikke forklaret ved baggrundsfaktorer som fx kommunens størrelse. I kommunesammenligningerne i dette papir bevares dette træk. For eksempel forklares kommuneforskelle ikke. Ideen er, at loven skal holdes, uanset at en kommune fx er lille. Socialforskningsinstituttet udfører imidlertid i samarbejde med Ankestyrelsen et andet projekt – en effektundersøgelse af praksisundersøgelserne – hvor kommuneforskelle søges forklaret. Projektet bliver baseret på samme type data.

I afsnit 2 er der mere grundigt forklaret, hvad praksisundersøgelser består af, og der gives en diskussion af, hvad der kan og ikke kan komme ud af kommunesammenligningen.

I de hidtidige praksisundersøgelser har Ankestyrelsen undersøgt godt 100 sager fra kommunerne. Typisk har der været hentet 2-3 sager ind fra hver kommune, der er udvalgt i den konkrete undersøgelse. Statistisk set er det alt for lidt til at sammenligne kommunernes tendens til at lave fejl. I de fremtidige praksisundersøgelser bliver der hentet måske 10 sager fra hver kommune. Det kan vise sig at være for lidt til at sammenligne kommuner. Hver praksisundersøgelse består dog af vurdering af flere led/spørgsmål i sagsbehandlingen ('er betingelse A opfyldt?', 'er betingelse B opfyldt?', ...), og det kan tænkes, at der er nogle spørgsmål, hvor nogle kommuner laver fejl i mange sager, fordi de ganske enkelt ikke kender en regel, mens andre kommuner næsten ingen fejl laver. Generelt må vi dog prøve at 'få flere observationer' fra hver kommune ved at se på flere spørgsmål i samme kommunesammenligning. Det kræver til gengæld lidt mere af de statistiske metoder. Der skal tages hensyn til, at nogle spørgsmål kan være sværere end andre, og at der måske er afhængighed mellem svarene i forskellige spørgsmål.<sup>1</sup>

I afsnit 3 beskrives problemet med relativt få observationer og de mulige måder at løse problemet. Indledningsvis opstilles en grundlæggende model for kommunesammenligning, og de grundlæggende statistiske forudsætninger diskuteres.

I afsnit 4 diskuteres, om hver kommune skal sammenlignes med gennemsnittet over kommuner, eller med den bedste kommune.

I afsnit 5 diskuteres 'multiple sammenligninger': antag at man undersøger om hver af 100 kommuner afviger fra (fx) gennemsnittet med et almindeligt 5 procents signifikansniveau. Det betyder, at selv hvis alle kommuner er lige gode til at sagsbehandle, så vil forventeligt 5 kommuner i den statistiske analyse blive udpeget, som om de havde for høj tendens til at lave fejl. Metoderne til multiple sammenligninger korrigerer herfor. De er et supplement til enkeltvise test for hver kommune. I nogle tilfælde skal den ene type analyse bruges, i andre tilfælde den anden.

---

<sup>1</sup> I øvrigt har Socialforskningsinstituttet tidligere foretaget benchmarkinganalyser for Ankestyrelsen, se Lindermann og Gregersen (2001) og Gregersen (2000).

Projektet laves som nævnt for Ankestyrelsen, og det skal munde ud i konkrete modeller, som Ankestyrelsen skal kunne bruge forholdsvis rutinemæssigt. I afsnit 6 konkretiseres modellerne derfor. Vi ser fx på modeller, hvor Ankestyrelsens eksperter vurderer sagsbehandlingens kvalitet hhv. kategoriseret og kontinuert, og på sammenligninger baseret på hhv. et eller flere spørgsmål. De forskellige metoder afprøves på kunstige datasæt.

I afsnit 7 forklares kort, hvordan de fremtidige praksisundersøgelser tænkes tilrettelagt.

Arbejdet bag dette arbejdsrapport har været brugt som input til en rapport, som Ankestyrelsen har lavet om de fremtidige kommunesammenligninger, se Ankestyrelsen (2006).

## **2. Uddybende om praksisundersøgelser og meningen med kommunesammenligninger**

Praksisundersøgelser foretages af Ankestyrelsen og regionale sociale nævn. Ankestyrelsen har ansvaret for praksiskoordineringen på landsplan, og de regionale nævn har ansvaret på det regionale plan. De regionale nævn er ankeinstans for borgere, der ønsker at klage over kommunernes afgørelser om retten til sociale ydelser. Ankestyrelsen træffer afgørelser i klagesager på næsten alle lovområder på det sociale og beskæftigelsesmæssige område og er øverste administrative ankeinstans. Ankestyrelsens afgørelser skaber desuden generel praksis på områder, hvor der hidtil har været tvivl.

En praksisundersøgelse omfatter et konkret område, fx tildeling af førtidspension (se på Ankestyrelsens hjemmeside [www.ast.dk](http://www.ast.dk), hvilke områder der har været taget op). Hver undersøgelse består af en vurdering af, hvor godt en række led i sagsbehandlingen er foregået, og der er en samlet vurdering af hver sag.

Ankestyrelsen og de regionale sociale nævn udfører forskellige informationsaktiviteter på baggrund af praksisundersøgelserne for at forbedre den kommunale sagsbehandling. Det drejer sig bl.a. om en rapport for hver praksisundersøgelse, tilbagemelding til deltagende kommuner, artikler i 'Nyt fra Ankestyrelsen' og undervisning.

I hver praksisundersøgelse vurderes hver sag som nævnt på en række forskellige spørgsmål. Det kan være en overordnet vurdering af, om sagen er afgjort korrekt, om borgeren er omfattet af den paragraf, som praksisundersøgelsen vedrører, om et støttebeløb er beregnet korrekt, om borgeren er blevet informeret skriftligt, eller om der er tilstrækkelig information lagt til grund for sagen. I mange praksisundersøgelser går samme type spørgsmål igen, for eksempel spørgsmålet om tilstrækkeligt dokumentationsgrundlag<sup>2</sup>. I den juridiske jargon deles disse spørgsmål op i materiel og formel korrekthed. Den materielle vurdering vedrører, om loven er opfyldt, mens den formelle korrekthed vedrører om forvaltnings- og retssikkerhedsmæssige regler er overholdt (dvs. mere generelle regler end den konkrete lov, der praksisundersøges), fx borgerinddragelse.

Når der i dette arbejdspapir bruges ordet 'fejl', skal man være opmærksom på, at det dækker over, at de juridiske eksperter vurderer, at afgørelsen eller dele heraf ikke er i overensstemmelse med lovgivningen eller gældende praksis. Vurderingen foretages på baggrund af det skriftlige materiale, som kommunerne har indsendt vedrørende den allerede afgjorte sag. I princippet kan Ankestyrelsens eksperter naturligvis tage fejl eller være påvirket af, at sagen allerede er afgjort. Såfremt der mangler lovpligtig dokumentation vurderes det som en fejl. Det kan i nogle tilfælde skyldes, at kommunen har glemt at indsende alle akter. Fejl dækker ikke blot over afvigelser fra loven, men også fra den praksis, der er fastlagt af Ankestyrelsen. Endelig kan 'fejl' også gradbøjes i eksperternes vurdering, idet en uoverensstemmelse kan udtrykkes fx 'tvivlende' eller delvist.

Som nævnt er formålet med praksisundersøgelserne at foretage en juridisk vurdering af den kommunale sagsbehandling. Det kan ved læsning af rapporter om nogle praksisun-

---

<sup>2</sup> I øvrigt peger en gennemgang af de seneste praksisundersøgelser på, at manglende dokumentation er en af de absolut hyppigste fejl.

dersøgelser virke, som om at nogle led i sagsbehandlingen bliver registreret som fejlagtige, selv om hverken borgeren, skatteyderne eller andre kan være væsentligt påvirket af 'fejlen'.<sup>3</sup> Omvendt er det i sagens natur relevant at sikre en juridisk korrekt sagsbehandling, og endelig er det vel rimeligt at antage, at borgere og skatteydere bliver *generelt* bedre stillet, desto bedre sagsbehandlingen er ud fra et juridisk synspunkt.<sup>4</sup>

De sager, der undersøges i en praksisundersøgelse kan opdeles i bevillingssager og afslagssager. Da fejlmulighederne til dels er forskellige i de to typer sager, og da der er langt flest registrerede bevillingssager, vil afslagssager fremover blive samlet i særskilte praksisundersøgelser. Den type fejl, der kan opstå i bevillingssager, har formentlig i særlig grad karakter af, at en borger bliver tildelt for generøs en ydelse fra det offentlige, mens der i afslagssager formentlig er en tendens til, at borgeren får for lidt i sager med fejl. Desuden er kravene til begrundelse over for borgeren skrapere i afslagssager.

De kommunale sager, der indgår i praksisundersøgelser, udvælges ved, at Ankestyrelsen beder om de seneste *x* afgjorte sager i de udvalgte kommuner. Det er et godt princip, men der kan for enkelte kommuner måske være problemer med repræsentativiteten, hvis det er uklart registreret, hvornår sagerne er afgjort. Mere overordnet set kan man naturligvis sagtens forestille sig, at der også opstår fejl, hvis en kommune slet ikke åbner en sag.

### **Meningen med kommunesammenligninger**

Meget abstrakt er sammenligningerne af kommunernes sagsbehandlingskvalitet en del af den generelle tendens i samfundet til i højere grad at bruge kvantitative metoder til at effektivisere den offentlige sektor. Kommunesammenligningerne kan derfor være genstand for den debat, der vedrører hele denne tendens. Så vidt jeg kan se (men jeg er ikke ekspert på området), er det centrale element, at det ofte er kamp om ressourcer til at opnå forskellige mål. Hvis man laver kvantificering på et område, kan der måske være

---

<sup>3</sup> Der er tale om små beløb i udregnet støtte i praksisundersøgelsen om § 84 (s. 21). I praksisundersøgelsen om § 77 (s. 9-10) ser det en del steder ud, som om at borgeren får det rigtige, men at udgiften "blot" er klassificeret under den forkerte paragraf.

<sup>4</sup> I øvrigt er det Ankestyrelsen og de sociale nævns lovgivningsmæssige pligt at vurdere sagsbehandlingen fra et rent juridisk synspunkt.



en tendens til, at det i sig selv trækker flere ressourcer til området på bekostning af andre områder. Rent abstrakt kunne man fx forestille sig, at de kommunale ledere ansatte urimeligt mange særligt dygtige og højtlønnede sagsbehandlere for at undgå at komme uheldigt ud i en kommunesammenligning, eller at sagsbehandlerne brugte al tiden på at sikre, at hver detalje fremgik skriftligt, frem for måske at snakke mere med borgeren om hans situation, og om hvad der er den bedste løsning. Så vidt jeg kan se, er der generelt ikke andet at gøre end at erkende dette mulige dilemma og lade være med at lade de målte størrelser få overdreven tildeling af ressourcer. Det modsatte synspunkt – nemlig at lade være med at måle, hvad der kan måles – forekommer som destruktion af viden.

I den konkrete sammenhæng virker det umiddelbart svært at se, at der i virkelighedens verden skulle være de store dilemmaer – de to ovenfornævnte eksempler er jo regulært søgte. De tiltag, man i praksis vil gøre for at forbedre sagsbehandlingen som følge af praksisundersøgelser, vil formentlig bestå i mere målrettet og effektiv undervisning, og det kan næppe siges at være noget, der trækker ressourcer fra andre områder i nævneværdigt omfang.

Et andet element i den generelle debat om kvantificering er, at selve kvantificeringen er ressourcekrævende. De ekstra ressourcer, der kræves for at kunne lave kommunesammenligninger, er relativt små i forhold til omkostningerne ved de hidtidige praksisundersøgelser og pålægges stort set kun Ankestyrelsen. Set i forhold til det enorme samlede antal sager, der behandles hvert år i landets kommuner, er det et minimum af ressourcerne, der bruges til den kvalitetssikring, der ligger i praksisundersøgelser.

Endelig er kommunesammenligninger jo ikke noget, der i sig selv forbedrer kvaliteten af sagsbehandlingen. Det samme kan siges om de eksisterende praksisundersøgelser. Meningen er, at identificere områder eller kommuner, hvor der er størst chance for at forbedre sagsbehandlingen. Kommunesammenligninger har derudover den fordel, at de bedste kommuner også udpeges, og man derigennem har mulighed for at finde ud af, hvad de bedste kommuner rent faktisk gør. Det er så muligt, at det ikke er kontrollerbare forhold, der gør nogle kommuner særligt gode. Man kan fx forestille sig, at store

kommuner med mange sager er relativt gode, fordi de har stor erfaring. Små kommuner kan jo ikke kopiere dette træk.<sup>5</sup> For nogle sagsområder kan man forestille sig, at enkelte sagsbehandlere varetager alle sager. Hvis kommunesammenligningerne kun inddrager dette enkelte område, er det ikke muligt at adskille person- og kommuneeffekter. En god kommune kan være god, fordi den mere eller mindre tilfældigt er kommet til at ansætte et naturligt sagsbehandlertal, hvilket de øvrige kommuner ikke kan kopiere. Dette fremhæver imidlertid et anonymitetsproblem ved kommunesammenligningerne, da det måske bliver enkelte sagsbehandlernes evner, der gemmer sig bag kommunesammenligningerne. Man kan naturligvis i princippet vælge mindre end fuld offentlighed omkring publiceringen af kommunesammenligningernes resultater. Omvendt er der jo meget lignende af mindst lige så personlig karakter, der er offentligt<sup>6</sup>, så forholdet er på ingen måde usædvanligt.

### **3. Få observationer – men flere spørgsmål i samme kommunesammenligning**

Lad os forestille os, at to kommuner med forskellig latent tendens til at lave fejl i et *konkret spørgsmål* i en *konkret praksisundersøgelse* sammenlignes i en statistisk test, hvor nulhypotesen er, at kommunerne er ens. Hvis forskellen i den latente fejltendens skal kunne vises statistisk, skal fejltendensen enten være stor, eller der skal være 'mange' observationer/sager fra hver kommune. Med 'mange' menes i denne sammenhæng, at de planlagte 10 sager pr. kommune ikke vil kunne vise små og almindeligt forekommende forskelle i fejltendenser. Et grundlæggende problem i kommunesammenligningerne er dette umiddelbart relativt lave antal observationer.

Der er to måder, hvorpå man kan løse problemet med de få observationer. Den ene er at lave kommunesammenligninger på de enkelte, detaljerede spørgsmål i hver praksisundersøgelse, og 'lede' blandt spørgsmålene efter kommuneforskelle. Den anden er at øge antallet af spørgsmål, der indgår i samme kommunesammenligning.

---

<sup>5</sup> Og dog: man kan forestille sig mere udbredt kommunesamarbejde, så mulighederne for specialisering øges. Eller endnu en runde kommunesammenlægninger.

<sup>6</sup> Eksempler: 1) I almindelighed er det let at finde ud af, hvem der er avanceret til chef i en virksomhed, og hvem der ikke er. 2) På mange uddannelsesinstitutioner kan karakterer ses af medstuderende. 3) For ansatte i forskningsinstitutioner er det offentligt, hvem der får offentliggjort sit arbejde i fine tidsskrifter.

Ved at se på enkelte, detaljerede spørgsmål vil man selv med få observationer kunne identificere latente fejlforskelle, hvis de er store. Sådanne store kommuneforskelle kan tænkes, hvis en kommune ganske enkelt ikke har forstået et vist element i reglerne. Eksemplerne i praksisundersøgelsen om § 84, hvor en indviklet regel betød fejl i støttebeløb giver grund til at tro, at sådanne fejlforskelle findes. Ved at se på enkelte, detaljerede spørgsmål bliver kommunesammenligningerne også anvendelige, idet man let kan se, hvilke kommuner der laver fejl i præcis hvilke spørgsmål.

Statistisk set skelner vi mellem den latente fejltendens og den faktiske fejltendens – den sidste er lig den første plus et restled, der indeholder de variationer i den faktiske fejltendens, der ikke ellers er med i modellen.

Den latente fejltendens i kommune  $k$  kan benævnes  $\alpha_k$  og den faktiske fejltendens benævnes  $I_k^*$ . Observationerne benævnes  $j_k$ , hvor  $j_k$  løber fra 1 til  $J_k$ , hvor  $J_k$  er antal observationer fra kommune  $k$ , altså fx 10 (men modellen tillader forskelligt antal observationer fra de forskellige kommuner). Modellen kan skrives som

$$(1) \quad I_{kj_k}^* = \alpha_k + e_{kj_k}$$

hvor  $e_{kj_k}$  er restleddet, og  $\alpha_k$  skal estimeres. Hvordan vi konkret kan estimere  $\alpha_k$ , giver vi eksempler på i afsnit 6. De centrale antagelser bag modellen er, at  $E(e_{kj_k} | k) = 0$ , dvs. at ingen kommune får særligt svære sager. Vi erkender imidlertid som nævnt, at  $\alpha_k$  ikke nødvendigvis kun dækker over, hvor dygtig kommunen som organisation er, men også over effekter, som kommunerne ikke har indflydelse på, fx kommunestørrelse. Endelig antages  $\text{cov}(e_{kj_{k_1}}, e_{kj_{k_2}}) = 0$ , dvs. at sagerne er ukorrelerede.

Inddrages vurderingen af flere led/spørgsmål fra praksisundersøgelserne, øges antallet af observationer. Ideen er, at kommune A måske laver flere fejl end B for hvert enkelt spørgsmål, vi ser på (dvs.  $\alpha_A > \alpha_B$  for hvert spørgsmål), men – for hvert enkelt spørgs-

mål – er forskellen ikke statistisk signifikant. Ved at se på flere spørgsmål samtidig, kan forskellen måske fremtræde signifikant. Det gælder naturligvis kun, hvis det vitterligt er sådan, at kommune A har en generel tendens til at lave flere fejl end B. Det kan lige så vel tænkes, at A måske er ringe på et område, men god på et andet område, og i så fald vil kommuneforskelle sløres (yderligere) ved at se på flere spørgsmål samtidigt.

Lad os forestille os, at vi måske ser på 4 forskellige spørgsmål samtidigt. Det kunne være spørgsmål inden for samme praksisundersøgelse, fx 'er betingelse A opfyldt?', ..., 'er betingelse D opfyldt?'. Spørgsmålene kunne også stamme fra forskellige praksisundersøgelser.

Modellen kan skrives

$$(2) \quad \begin{aligned} I_{ksj_{ks}}^* &= \alpha_{ks} + e_{ksj_{ks}} \\ &= \alpha_k + \alpha_s + e_{ksj_{ks}} \end{aligned}$$

Konstantleddet  $\alpha_{ks}$  betegner kommune  $k$ 's tendens til at lave fejl i spørgsmål  $s$ . Det er nødvendigt at antage en vis form for regelmæssighed i denne størrelse for at kunne udnytte, at vi ser på flere spørgsmål på en gang, så i andet lighedstegn antages, at  $\alpha_{ks}$  kan opsplittes lineært i et led, der betegner indflydelsen på fejltendensen, der skyldes kommunen ( $\alpha_k$ ), og et led, der skyldes, at spørgsmål  $s$  er særlig svært eller let ( $\alpha_s$ ). Som nævnt er det ofte i 'informationsspørgsmålet', at der laves fejl, så man må forvente, at  $\alpha_{\text{Information}}$  er stor. Indekset for observationer er  $j_{ks}$ , hvor  $j_{ks} = 1, \dots, J_{ks}$ , og hvor  $J_{ks}$  er antal observationer fra kommune  $k$  vedrørende spørgsmål  $s$ .

Hvis de forskellige spørgsmål i (2) tages fra samme praksisundersøgelse, kan det imidlertid til dels være en slags øjenbedrag, at man opnår flere observationer ved at inddrage flere spørgsmål. Det kan ske, hvis nogle sager ganske enkelt er mere bøvlede end andre. I så fald vil der være en stor chance for, at hvis der er fejl i fx informationsspørgsmålet, så er der også fejl i spørgsmålet om, hvorvidt borgeren har fået den rette ydelse. Hvis fejl i informationsspørgsmålet fx altid efterfølges af fejl i ydelsesspørgs-

målet, så er der ikke opnået mere information ved at tage ydelsesspørgsmålet med i kommunesammenligningen. Statistisk set er problemet, at restleddene  $e_{ksj_{ks}}$  ikke er uafhængige, dvs.  $\text{cov}(e_{ksj_{ks}}, e_{ks'j_{ks}}) \neq 0$ . Det betyder – hvis der ikke korrigeres for det – at estimerne for  $\alpha_k$  fremtræder mere sikre, end de reelt er.

Kovariansen for  $e_{ksj_{ks}}$  er illustreret nedenfor

**Tabel 1. Kovarians for  $e_{ksj_{ks}}$  i model (2) med forskellige spørgsmål fra samme praksisundersøgelse**

Kommune	Sag	Spørgsmål	Kommune Sag Spørgsmål							
			1	1	1	1	2	2	2	2
			I	Y	I	Y	I	Y	I	Y
1	1	I	$\sigma_I^2$	$\sigma_{IY}$	0	0	0	0	0	0
1	1	Y	$\sigma_{IY}$	$\sigma_Y^2$	0	0	0	0	0	0
1	2	I	0	0	$\sigma_I^2$	$\sigma_{IY}$	0	0	0	0
1	2	Y	0	0	$\sigma_{IY}$	$\sigma_Y^2$	0	0	0	0
2	1	I	0	0	0	0	$\sigma_I^2$	$\sigma_{IY}$	0	0
2	1	Y	0	0	0	0	$\sigma_{IY}$	$\sigma_Y^2$	0	0
2	2	I	0	0	0	0	0	0	$\sigma_I^2$	$\sigma_{IY}$
2	2	Y	0	0	0	0	0	0	$\sigma_{IY}$	$\sigma_Y^2$

Hver sag er indikeret i tabellen, skønt sagsnummeret ikke optræder i model (2). Derved vises, at forskellige observationer har fællestræk, der ikke er korrigeret for via et konstantled i (2), dvs. der indgår ikke et led  $\alpha_{\text{sag}}$  svarende til  $\alpha_k$  og  $\alpha_s$ . Det er heller ikke muligt at indføre  $\alpha_{\text{sag}}$  i (2). Det skyldes, at hver sag jo er knyttet til en bestemt kommune, så  $\alpha_k$  og  $\alpha_{\text{sag}}$  kan ikke adskilles i en model som

$$(2') \quad I_{ksj_{ks}}^* = \alpha_{\text{sag}} + \alpha_k + \alpha_s + e_{ksj_{ks}} \quad (\text{uidentificeret})$$

Hvis de forskellige spørgsmål derimod kommer fra hver sin praksisundersøgelse, er kovariansmatricen som følger:

**Tabel 2. Kovarians for  $e_{ksj_{ks}}$  i model (2) med forskellige spørgsmål fra forskellige praksisundersøgelser**

Kommune	Spørgsmål	Kommune Spørgsmål			
		1 PU1	1 PU2	2 PU1	2 PU2
	1 PU1	$\sigma_{P1}^2$	0	0	0
	1 PU2	0	$\sigma_{PY}^2$	0	0
	2 PU1	0	0	$\sigma_I^2$	$\sigma_{IY}$
	2 PU2	0	0	$\sigma_{IY}$	$\sigma_Y^2$

Der er her ikke noget, der skaber ensartethed mellem PU1 og PU2 fra den samme kommune (i modsætning til ovenfor, hvor sagsnummeret var det samme for forskellige spørgsmål).

Man skal være opmærksom på, at spørgsmålene i praksisundersøgelserne har forskelligt niveau. En gruppe af spørgsmål vedrører udvalgte dele af den proces, en sag skal igennem – det kan være 'er betingelse A opfyldt?', 'er betingelse B opfyldt?' osv. En anden gruppe vedrører sagen som helhed. De sidste er derfor en 'sum' af de detaljerede spørgsmål, og derved er der en automatisk afhængighed mellem detaljerede og overordnede spørgsmål. Derfor skal der i samme kommunesammenligning ikke indgå både detaljerede og overordnede spørgsmål.

**Metoder til at løse afhængighedsproblemet i (2), når spørgsmål kommer fra samme sag**

1. Test  $\sigma_{IY} = 0$ . Hvis hypotesen accepteres, så benyttes model (2), som om der er uafhængighed. Testet kan foretages med et Cochran-Mantel-Haenszel test. (Standard i SAS.) Se Agresti (2002), i Rasmussen (2005) er testet anvendt.

2. Modellér kovariansen som et led i modellen. Det kan gøres som

$$(3) \quad \begin{aligned} I_{ksj_{ks}}^* &= \alpha_k + \alpha_s + e_{ksj_{ks}} \\ e_{ksj_{ks}} &= u_{ksj_{ks}} + u_{kj_{ks}} \end{aligned}$$

Der er altså indført et led for hver sag,  $u_{kj_{ks}}$ , men det er stokastisk i modsætning til deterministisk i (2') og kan derfor estimeres. Modellen kan estimeres ved 'mixture' metoder. (Standard i SAS for lineære modeller.)

Kovariansmatricerne for  $u_{ksj_{ks}}$  og  $u_{kj_{ks}}$  kan illustreres som

**Tabel 3a. Kovarians for  $u_{ksj_{ks}}$  i model (3) (forskellige spørgsmål fra samme praksisundersøgelse)**

Kommune	Sag	Spørgsmål	Kommune								
			1		1		2		2		
			I	Y	I	Y	I	Y	I	Y	
1	I	I	$\gamma_I^2$	0	0	0	0	0	0	0	0
1	I	Y	0	$\gamma_Y^2$	0	0	0	0	0	0	0
1	2	I	0	0	$\gamma_I^2$	0	0	0	0	0	0
1	2	Y	0	0	0	$\gamma_Y^2$	0	0	0	0	0
2	I	I	0	0	0	0	$\gamma_I^2$	0	0	0	0
2	I	Y	0	0	0	0	0	$\gamma_Y^2$	0	0	0
2	2	I	0	0	0	0	0	0	$\gamma_I^2$	0	0
2	2	Y	0	0	0	0	0	0	0	$\gamma_Y^2$	0

**Tabel 3b. Kovarians for  $u_{kjs}$  i model (3) (forskellige spørgsmål fra samme praksisundersøgelse)**

Kommune	Sag	Spørgsmål	Kommune							
			Sag							
			Spørgsmål							
			1	1	1	1	2	2	2	2
			I	Y	I	Y	I	Y	I	Y
1	I	I	$\rho^2$	$\rho^2$	0	0	0	0	0	0
1	I	Y	$\rho^2$	$\rho^2$	0	0	0	0	0	0
1	2	I	0	0	$\rho^2$	$\rho^2$	0	0	0	0
1	2	Y	0	0	$\rho^2$	$\rho^2$	0	0	0	0
2	I	I	0	0	0	0	$\rho^2$	$\rho^2$	0	0
2	I	Y	0	0	0	0	$\rho^2$	$\rho^2$	0	0
2	2	I	0	0	0	0	0	0	$\rho^2$	$\rho^2$
2	2	Y	0	0	0	0	0	0	$\rho^2$	$\rho^2$

De to kovariansmatricer kan summeres og give kovariansmatricen for  $e_{kjs}$ . Det vil sige, at  $\sigma_Y^2 = \gamma_Y^2 + \rho^2$ ,  $\sigma_I^2 = \gamma_I^2 + \rho^2$  og  $\sigma_{YI} = \rho^2$ .

3. Omformuler modellen således at kommuneparametren  $\alpha_k$  umiddelbart udgår og erstattes af en parameter for hver enkelt sag. Hvis  $c$  indekserer sager, er denne betegnet  $\alpha_c$ , og modellen er

$$(4) \quad I_{cs}^* = \alpha_c + \alpha_s + e_{cs}$$

Hver enkelt observation er identificeret ved parret  $(c, s)$  dvs. sag nummer  $c$ , spørgsmål  $s$ . Det kan betale sig at lade de forskellige sager ( $c$ 'er) betegnes med et par,  $(k, j_k)$ , hvor  $k$  er kommunen og  $j_k$  er indeks for sagen i kommunen. Indekset  $j_k$  løber som før fra 1 til  $J_k$ , dvs. antal sager i kommune  $k$  (fx 10). Man kan nu finde kommuneeffekten ved at se på gennemsnittet af de 'sagsparametre', der hører til kommunen, dvs.

$$(5) \quad \alpha_k = \frac{1}{J_k} \sum_{j_k=1}^{J_k} \alpha_{(k, j_k)}$$



Restleddene  $e_{cs}$  er uafhængige, dvs.  $\text{cov}(e_{cs}, e_{cs'}) = 0$ . Det er en rimelig antagelse, for der er jo netop korrigeret for sagens sværhedsgrad via parametren  $\alpha_c$ .

#### 4. Sammenligning med gennemsnit eller bedste kommune

Givet der er estimeret en effekt fra hver kommune,  $\alpha_k$ , på den latente fejltendens, er spørgsmålet, hvordan de skal sammenlignes. I dette afsnit diskuteres kort princippet i forskellige metoder. De konkrete statistiske modeller forklares senere.

Man kan sammenligne hver kommune med landsgennemsnittet,  $\alpha_{\square}$ , eller med den bedste kommune, hvis parameter kan kaldes  $\underline{\alpha}$ . Desuden kan  $\alpha$ -parametrene rangordnes,  $\alpha_1 < \alpha_2 < \dots < \alpha_K$ , og kommunerne kan grupperes efter, hvilke andre kommuner de ikke adskiller sig fra. Tabellen illustrerer disse parvise sammenligninger.

**Tabel 4. Skitse af parvise sammenligninger**

	Gruppe 1	Gruppe 2	Gruppe 3	. . . . .
A ( $\alpha_A = \alpha_1$ )	X			
B ( $\alpha_B = \alpha_2$ )	X	X		
C ( $\alpha_C = \alpha_3$ )	X	X	X	
D ( $\alpha_D = \alpha_4$ )		X	X	
E ( $\alpha_E = \alpha_5$ )			X	
.				
.				
.				

Kommune A kan ikke afvises at have samme fejltendens som kommune B og C, men kan afvises at være bedre end D og E. Kommune B kan kun afvises at lave fejl som kommune E, og E kan afvises at være lige så god som A og B.

Sammenligning med bedste kommune viser forbedringspotentialet for alle kommuner. Imidlertid er metoden lidt vanskeligere at håndtere i nogle konkrete statistiske modeller

(se afsnit 6), og det gælder naturligvis i særlig grad, når det erindres, at kommunesammenligningerne skal kunne laves rutinemæssigt i Ankestyrelsen.<sup>7</sup>

De parvise sammenligninger giver et meget kønt billede, men det er svært at se, hvilken ekstra brugbar information man opnår, ud over hvad man opnår med sammenligninger med bedste kommune.

I afsnit 6 vises modeller, der både bruger sammenligninger med gennemsnit og med den bedste kommune.

## 5. Multiple sammenligninger

I afsnittet beskrives metoder, der ikke nødvendigvis vil blive brugt i Ankestyrelsens fremtidige rutinemæssige kommunesammenligninger. Multiple sammenligninger bygger nemlig især (i hvert fald som standardmetoder i SAS) på normalfordelte stokastiske variabler, snarere end kategoriserede variabler. Metoderne er altså som standard mere velegnede til lineære end kategoriserede modeller. Metoderne beskrives imidlertid, fordi de ikke på skrivende tidspunkt kan udelukkes, og fordi de er relevante.

Den mest anvendte metode til at teste, om kommuner adskiller sig signifikant fra hinanden, er at teste, for hver *enkelt kommune*, om kommunen adskiller sig fra den bedste kommune (eller gennemsnittet afhængigt af valget i foregående afsnit). Nulhypotesen er, at kommunerne er ens. Hvis kommunerne rent faktisk er ens, og hvis signifikansniveauet vælges til 5 procent, må man forvente, at 5 ud af 100 kommuner statistisk fremtræder, som om de er ringere end den bedste kommune. Ud fra et bestemt synspunkt er det et problem. Nemlig hvis alle de enkeltvise tests er lavet for de 100 kommuner, og man *efter* resultatet foreligger *leder* efter de dårligste kommuner, og dernæst konkluderer, at de fundne kommuner er dårlige. Omvendt er det netop en opgave for en overord-

---

<sup>7</sup> Bemærk endelig, at det i princippet kan tænkes, at den bedste enhed er baseret på meget få observationer, mens den næstbedste måske er baseret på mange observationer. Derfor kan det godt tænkes, at man statistisk kan påvise forskelle mellem en dårlig enhed og den næstbedste, men ikke forskelle mellem den dårlige enhed og den bedste enhed. I kommunesammenligningerne bliver der dog nok omtrent lige mange observationer fra hver kommune.

net kontrolinstans at se, om der er nogle kommuner, der er særligt ringe. Med metoderne til multiple sammenligninger laves tests til sådanne formål. Både enkeltvise sammenligninger og multiple sammenligninger er nyttige, men skal bruges med hver sit udgangspunkt.

*Enkeltvise sammenligninger* skal bruges, hvis man på forhånd er interesseret i en bestemt kommune og ønsker at vide, om den er dårligere end den bedste kommune. Dette synspunkt er naturligt for den pågældende kommune.

*Multiple sammenligninger* skal bruges af en overordnet myndighed, der er interesseret i, at de ringeste kommuner – uagtet hvem de er – bliver bedre.

Generelt skal der mindre til, for at en kommune afviger fra den bedste kommune i enkeltvise test end i multiple sammenligninger. Hvis en kommune bliver udpeget som dårlig i den enkeltvise test, men ikke i den multiple sammenligning, bør kommunaldirektøren altså ikke undskylde sig med, at hans kommune blot er den uheldige med mange fejl, og omvendt bør en overordnet kontrolinstans ikke sige, at den er interesseret i netop den pågældende kommune.<sup>8</sup>

For at forklare de to typer tests lidt mere formelt kan man forestille sig, at man tester, om hver af 100 kommuner har samme fejltendens som den bedste, dvs. om  $\alpha_k = \underline{\alpha}$  for hver  $k$ , og der vælges et signifikansniveau på 0,05 i de enkeltvise test. Hvis nulhypotesen er sand  $\alpha_k = \underline{\alpha}$ , er der

$$1-(1-0,05)=0,05$$

sandsynlighed for at afvise en sand nulhypotese i hvert enkeltvist test. Hvis 100 kommuner sammenlignes mod den bedste kommune, foretages i alt 100 tests af ovenstående type, der involverer størrelserne  $\alpha_k - \underline{\alpha}$ . Hvis disse størrelser var uafhængige (hvad de ikke er, se nedenfor), ville der være

---

<sup>8</sup> Og – et lidt andet eksempel – den landsholdsspiller, der brænder et straffe i en straffesparkskonkurrence bør overveje, om han ikke kan gøre det bedre, men træneren bør ikke nødvendigvis sætte ham af holdet.

$$1-(1-0,05)^{100}=0,994$$

sandsynlighed for, at den sande hypotese om ens fejltendenser blev afvist for en eller anden kommune. Hvis man imidlertid valgte 0,000513 i formlen i stedet for 0,05, ville der være

$$1-(1-0,000513)^{100}=0,05$$

sandsynlighed for, at en eller anden kommune blev afvist at være lige så god som den bedste. Det gælder altså i multiple sammenligninger om at vælge et lavt, 'enkeltvist signifikansniveau' (0,000513), så det 'multiple signifikansniveau' bliver 0,05. Hvis vi i stedet ser på konfidensintervaller for  $\alpha_k - \underline{\alpha}$  og for forklaringens skyld lader, som om  $\alpha_k - \underline{\alpha}$  er standardiseret normalfordelt, så skal konfidensgrænserne vælges som hhv. 1,96 for enkelvise sammenligninger (fordi  $\Pr(\alpha_k - \underline{\alpha} > 1,96) = 0,05$  givet nulhypotesen  $\underline{\alpha} = \alpha_k$ ) og som 3,283 (fordi  $\Pr(\alpha_k - \underline{\alpha} > 3,283 \text{ netop for } k) \approx 0,000513$  givet nulhypotesen  $\underline{\alpha} = \alpha_k$ ) for multiple sammenligninger, fordi

$$\Pr(\alpha_k - \underline{\alpha} > 3,283 \text{ for nogen } k) = \Pr(\bar{\alpha} - \underline{\alpha} > 3,283) = 0,05,$$

$$\text{hvor } \bar{\alpha} = \max\{\alpha_k \mid k = 1, 2, \dots, 100\}.$$

Nu er  $\alpha_k - \underline{\alpha}$  ikke normalfordelt med kendt varians og slet ikke uafhængige ( $\underline{\alpha}$  indgår jo for alle  $k$ ) og har endnu mindre middelværdien 0. Men ideen i multiple sammenligninger er alligevel illustreret: Det gælder om at udvide det 'enkeltvise konfidensinterval', således at observationer, der ligger ud over det udvidede 'multiple konfidensinterval', ikke blot kan betragtes, som om at den pågældende kommune er den uheldige med særligt mange fejl. I eksemplet ovenfor er konfidensintervallet udvidet fra  $[0, 1,96]$  til  $[0, 3,286]$ .

De to test kan forklares yderligere ved to nulhypoteser

$H_k^0 \quad \alpha_k = \underline{\alpha}$  for netop  $k$  (enkeltvise tests)

$H_m^0 \quad \bar{\alpha} = \underline{\alpha}$  (multipel sammenligning)

Det leder til til konfidensintervaller

$[0, \bar{\alpha}_k]$  (enkeltvise tests)

$[0, \bar{\alpha}]$  (multiple sammenligninger)

På baggrund af estimater for  $\alpha_k$ , benævnt  $\hat{\alpha}_k$ , kan man drage konklusionerne:

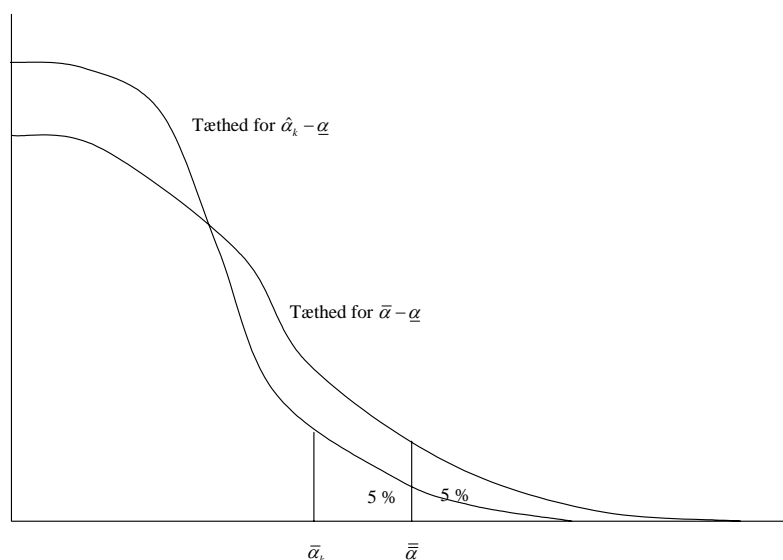
Hvis  $\hat{\alpha}_k \in [0, \bar{\alpha}_k]$  accepteres  $H_k^0$  for den konkrete kommune. Ellers afvises  $H_k^0$  for den konkrete kommune.

Hvis  $\hat{\alpha}_k > \bar{\alpha}$  for en bestemt kommune,  $k$ , så er  $k$ 's afvigelse alene nok til, at  $H_m^0$  afvises.

Bemærk, at multiple sammenligninger ikke leder til, at vi kan acceptere  $H_m^0$ . Hypotesen er fx ikke nødvendigvis accepteret, hvis fx  $\hat{\alpha}_k \in [0, \bar{\alpha}]$  alle  $k$ . Hvis man direkte ville teste  $H_m^0$ , kunne det gøres ved et traditionelt F-test.

Figuren illustrerer de to tests og markerer zoner, der leder til forskellige konklusioner.

**Figur 1. Skitse af enkeltvise og multiple sammenligninger**



En ældre standardmetode inden for multiple sammenligninger er Tukeys metode. Den bygger på en fordeling af rangen for en samling stokastiske størrelser, hvor rangen altså er  $w = \bar{\alpha} - \underline{\alpha}$ . Hvis  $\alpha$ 'erne er normalfordelt med estimat for standardafvigelsen på  $s$ , hvor  $s$  er estimeret på baggrund af  $\nu$  frihedsgrader, og hvis der er  $r$  enheder (kommuner), der sammenlignes, så er

$$(6) \quad q_{r,\nu} = \frac{w}{s}$$

fordelt efter den såkaldte 'studentized range distribution', se Dunnett (1980). Denne er tabelleret i en del bøger, se Toothaker (1993). De kritiske værdier vokser naturligvis i  $r$ , dvs. desto flere enheder vi sammenligner, og falder i  $\nu$ . Der findes adskillige andre varianter af metoder til 'multiple sammenligninger'.

Tukeys metode (Dunnett (1955, 1980)) bruges til at sammenligne alle par af kommuner. Metoden kan fx bruges til at lave sammenligninger som skitseret side 15. Metoden bygger som nævnt på den statistiske fordeling af rangen (den maksimale minus den minimale observerede værdi). Dunnett (1980) studerer tilfældet, hvor alle kommuner sammenlignes med en på forhånd given kommune. Når vi ønsker at sammenligne med den bedste kommune, skal Tukeys metode bruges, fordi det ikke på forhånd er givet, hvilken kommune der er bedst. Sammenligning med gennemsnittet er mere i stil med Dunnetts (1980) metode.

## **6. Svarkategorier i praksis og statistiske modeller**

I afsnittet præsenteres en række forskellige konkrete modeller, der kan estimere kommunernes effekt på fejltendensen,  $\alpha_k$ . Afsnittet har karakter af et 'menukort', fordi vi som nævnt endnu ikke har haft lejlighed til at lave empiriske modeller og derved udvælge en foretrukken estimation.

I de kommende praksisundersøgelser kan Ankestyrelsens eksperters vurderinger udtrykkes på forskellig måde. Der kan være tale om dikotomiske vurderinger af typen 'rigtigt/forkert' eller '0/1', kategoriserede vurderinger af type 'Helt rigtigt/stort set rigtigt/temmelig forkert/helt forkert', hvor der er en ordning mellem kategorierne, eller kontinuerte vurderinger i tal, der løber mellem 0 og 1, eller fire talværdier mellem 0 og 1.

Den statistiske behandling skal derudover tage hensyn til, om sammenligningen sker på baggrund af et eller flere spørgsmål, om der sammenlignes med den bedste kommune eller gennemsnittet af kommuner.

Endelig er der et afsnit om konsekvenserne af, at Ankestyrelsens eksperter kan tage fejl.

## 6.1 Logistiske modeller

De logistiske modeller bruges, når eksperternes vurdering kan falde enten i to kategorier eller i en lille håndfuld kategorier.

Udgangspunktet kan være tendensen til at lave fejl, her gentaget fra (2)

$$(7) \quad I_{ksjks}^* = \alpha_k + \alpha_s + e_{ksjks}$$

Denne tendens kan ikke observeres. Det observerbare er eksperternes vurdering af det pågældende spørgsmål i den pågældende sag. Tendensen til at lave fejl kan skrives som en sandsynlighed. Fra  $I_{ksjks}^*$  trækkes restleddet, så  $I_{ksjks} = I_{ksjks}^* - e_{ksjks} = \alpha_k + \alpha_s$ . (Generelt vil  $I$ 'er uden top-\* betegne dette.) Sandsynligheden for fejl er

$$(8) \quad \Pr(\text{'fejl'}) = \frac{e^{I_{ksjks}}}{1 + e^{I_{ksjks}}}$$

i den logistiske model. Det er antaget, at restleddet er fordelt efter den såkaldte 'type I extreme value'-fordeling.

Sandsynligheden for 'rigtig' er

$$(9) \quad \Pr(\text{'rigtig'}) = \frac{1}{1 + e^{I_{ksjks}}}$$

Statistisk er problemet at finde estimater for  $\alpha_k$  og  $\alpha_s$ , benævnt  $\hat{\alpha}_k$  og  $\hat{\alpha}_s$ , så der er størst mulig overensstemmelse mellem vurderingerne og de beregnede sandsynligheder.

### 6.1.1 Kommunesammenligninger baseret på et spørgsmål

Når der i kommunesammenligningen kun ses på et enkelt spørgsmål, skal spørgsmålsparameteren  $\alpha_s$  udgå af (7). Modellen til estimation af  $\alpha_k$ 'er er derfor som i (7)-(9).



### 6.1.1.1. Test mod gennemsnit eller bedste kommune

Estimationen giver som resultat estimater for kommuneeffekter  $\hat{\alpha}_k, k = 1, \dots, K$ , hvor  $K$  er antal kommuner, og tilhørende standardafvigelser  $\sigma_k$ . Den gennemsnitlige kommuneeffekt er  $\hat{\alpha}_\bullet = \frac{1}{K} \sum_k \hat{\alpha}_k$ . Den bedste kommune er  $k_0$ , hvor  $\hat{\alpha}_{k_0} < \hat{\alpha}_k$  for alle  $k$ .

Det kan testes, om kommune  $k$  er forskellig fra bedste kommune eller gennemsnittet ved hjælp af approksimative t-teststørrelser

$$(10) \quad \begin{aligned} T &= \frac{\hat{\alpha}_k - \hat{\alpha}_{k_0}}{\sqrt{\sigma_k^2 + \sigma_m^2}} && \text{i forhold til bedste kommune} \\ T &= \frac{\hat{\alpha}_k - \hat{\alpha}_\bullet}{\sqrt{\sigma_k^2 + \sigma_\bullet^2}} \approx \frac{\hat{\alpha}_k - \hat{\alpha}_\bullet}{\sigma_k} && \text{i forhold til gennemsnit} \end{aligned}$$

Det omtrentlige lighedstegn kan begrundes med, at variansen på gennemsnittet er meget lille i forhold til variansen på hver enkelt kommunes gennemsnit.

Det ellers tiltalende test i forhold til bedste kommune lider naturligvis af, at usikkerheden på  $\hat{\alpha}_{k_0}$  kan være så stor, at få kommuner kan afvises at være forskellige fra den bedste kommune. I praksis kan man derfor vælge en gruppe af de bedste kommuner og teste de øvrige mod dem.

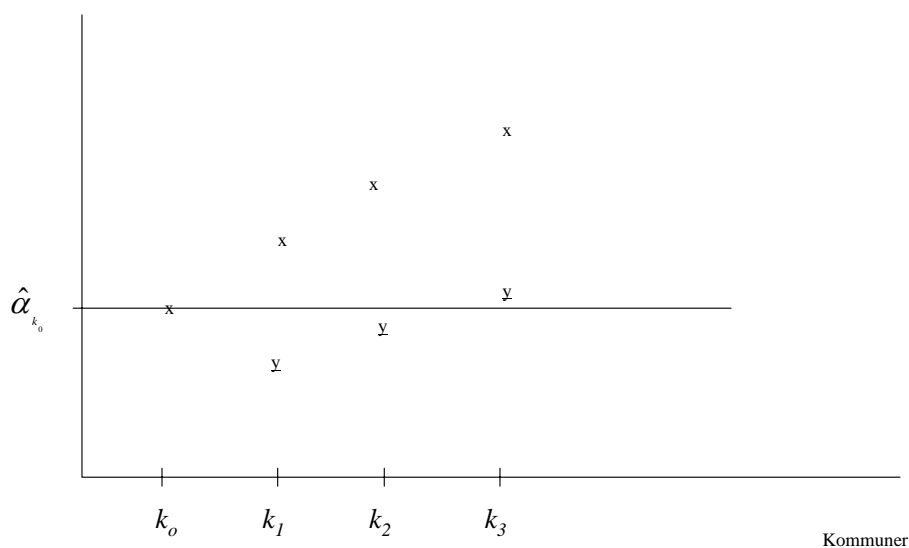
### 6.1.1.2. Konfidensintervaller og præsentation

Der kan beregnes øvre og nedre konfidensgrænser for  $\hat{\alpha}_k - \hat{\alpha}_{k_0}$  eller  $\hat{\alpha}_k - \hat{\alpha}_\bullet$ . Hvis konfidensintervallerne betegnes  $\underline{d}$  og  $\bar{d}$ , kan man til præsentationsformål beregne

$$(11) \quad \begin{aligned} \bar{\hat{\alpha}}_k &= \bar{d} + \hat{\alpha}_{k_0} \quad (\text{evt.} = \bar{d} + \hat{\alpha}_\bullet) \\ \underline{\hat{\alpha}}_k &= \underline{d} + \hat{\alpha}_{k_0} \quad (\text{evt.} = \underline{d} + \hat{\alpha}_\bullet) \end{aligned}$$

Resultaterne kan i Ankestyrelsens kommunesammenligninger præsenteres som i figur 2.

**Figur 2. Præsentation af estimerede parametre og indikation om signifikante forskelle**



I figuren aflæses de estimerede  $\alpha$ -parametre i rangordning (x-erne i figuren). Man skal huske på, at selv om masser af kommuner ikke kan afvises at have samme fejltendens som den bedste kommune, så er estimeret alligevel bedste bud på kommunens fejltendens. Dernæst kan man af figurens y-er aflæse  $\hat{\alpha}_k$ 'erne. En kommune afvises at have samme fejltendens som den bedste kommune, hvis  $\hat{\alpha}_k > \hat{\alpha}_{k_0}$ . Bemærk at  $\hat{\alpha}_k$  ikke er nedre konfidensinterval for  $\hat{\alpha}_k$ , fordi  $\hat{\alpha}_{k_0}$  ikke er en konstant, jf. (11).

Hvis man tester i forhold til gennemsnittet, skal  $\hat{\alpha}_k$  naturligvis indgå i figuren i stedet for  $\hat{\alpha}_{k_0}$ .

I øvrigt vil det måske være mere sigende at præsentere en version af figuren, hvor sandsynligheder baseret på (8) er beregnet. Man kunne beregne

$$(12) \quad \begin{aligned} \Pr(\text{'fejl'}) &= \frac{e^{\hat{\alpha}_k}}{1 + e^{\hat{\alpha}_k}} \quad \text{for middelskøn} \\ \Pr(\text{'fejl'}) &= \frac{e^{\hat{\alpha}_k}}{1 + e^{\hat{\alpha}_k}} \quad \text{for nedre skøn} \end{aligned}$$

Med kun et spørgsmål i kommunesammenligningen er der ret beset mere oplagte modeller at anvende end den logistiske model i (7)-(9), nemlig test af forskellene i sandsynligheder for binomialfordelte stokastiske variabler. Derved får man direkte middelskøn og konfidensintervaller for sandsynligheder i stedet for først at skulle beregne  $\alpha_k$ -parametre. Imidlertid er den logistiske model god, fordi den let kan udvides til flere spørgsmål, flere ordnede kategorier o.l.

### 6.1.2 Flere spørgsmål, indbyrdes uafhængige

Modellen i (7)-(9) kan bruges, og der kan beregnes  $\hat{\alpha}_k$ 'er, testes og præsenteres som i foregående afsnit.

Den kritiske uafhængighedsantagelse er

$$(13) \quad \text{cov}(e_{ksj_{ks}}, e_{ks'j_{ks}}) = 0 \text{ for } s \neq s'$$

Hvis der sammenlignes på baggrund af vurderingen af flere spørgsmål, skal vurderingen kunne behandles af samme type statistiske model – dvs. alle spørgsmål skal være fx vurderet som rigtigt/forkert eller vurderet lineært. Det er fordi, kommuneparameteren  $\alpha_k$  skal kunne tolkes på samme måde i hvert spørgsmål.

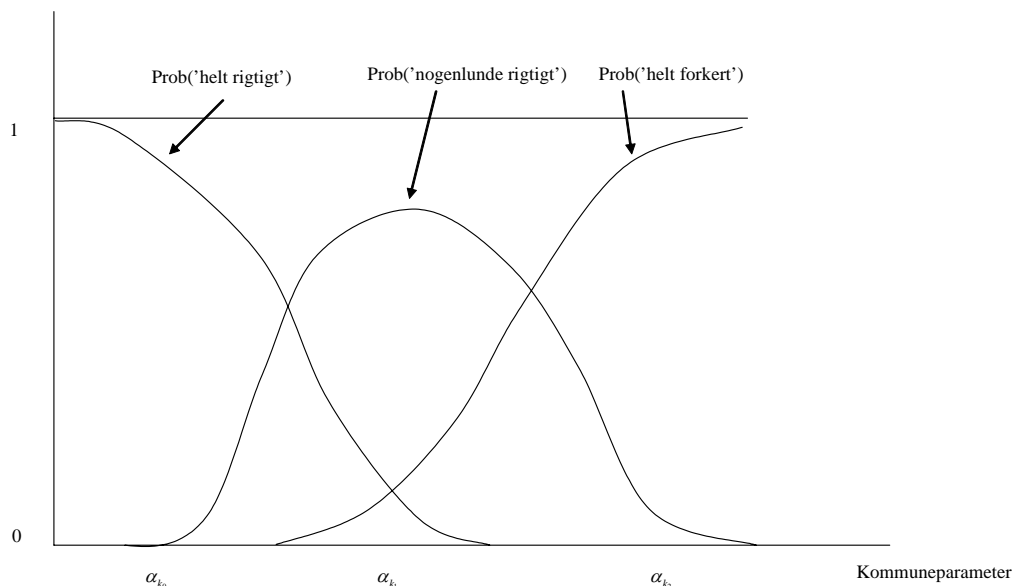
I modellen i ligning (7) ( $I_{ksj_{ks}}^* = \alpha_k + \alpha_s + e_{ksj_{ks}}$ ) er det nødvendigt, at de samme spørgsmål indgår for flere kommuner – ellers bliver det ikke muligt at adskille (identificere) kommuneeffekter fra de effekter, der skyldes, at nogle spørgsmål er sværere end andre. Se nærmere om hvordan der kan opstå identifikationsproblemer i afsnit 7.

### 6.1.3 Ordinalt ordnede, kategoriserede vurderinger

De fleste spørgsmål i praksisundersøgelserne har fire snarere end to kategorier til at vurdere kvaliteten af et bestemt led i sagsbehandlingen. De fire kategorier er ordnede, dvs. de kan rangordnes fx ved 'helt rigtigt', 'nogenlunde rigtigt', 'noget mangelfuldt' til 'helt forkert'. En standard statistisk model at behandle denne type data på er den ordinale logistiske model. Det 'ordinale' betyder, at de fire kategorier er rangordnede med 'ordenstal', men der ikke er taget stilling til, om det er en større gevinst at springe fra 'nogenlunde rigtigt' til 'helt rigtigt' end fra 'noget mangelfuldt' til 'nogenlunde rigtigt', eller om 'helt rigtigt' er dobbelt så godt eller 50 procent bedre end 'nogenlunde rigtigt'.

I nedenstående figur er modellen forsøgt forklaret. For overskuelighedens skyld er der kun skitseret tre vurderingskategorier. De tre kommuner er rangordnet, således at  $k_0$  er den bedste.

**Figur 3. Skitse af den ordinale logistiske model**



Desto ringere kommunen er (desto højere fodtegn), desto større er sandsynligheden for, at dens sagsbehandling vurderes 'helt forkert', og desto lavere er sandsynligheden for, at sagsbehandlingen vurderes 'helt rigtigt'. Sandsynligheden for mellemvurderingerne kan dog både stige og falde i den latente fejltendens. Kommune 1 har således en større sandsynlighed end kommune 0 for vurderingen 'nogenlunde rigtigt', mens kommune 2 derimod har lavere sandsynlighed for denne vurdering end for kommune 1. Man kan sige, at 'nogenlunde rigtigt' er et godt resultat set fra den ringe kommune 2's side.

Formelt er modellen meget lig (7)-(9). Der indføres blot et konstantled  $c_h$  for hver vurderingskategori,  $h$ .

$$(14) \quad I_{h,ksj_{ks}}^* = c_h + \alpha_k + \alpha_s + e_{ksj_{ks}}$$

Indekset  $h$  løber fra 1 til 4, hvis der er 4 vurderingskategorier, og konstantleddet vokser i  $h$ ,  $c_h < c_{h+1}$  ( $h = 4$  er den ringeste kategori). Rent statistisk estimeres fx

$$(15) \quad \Pr(\text{'vurderet som 'nogenlunde rigtig' eller bedre'}) = \frac{e^{I_{h,ksj_{ks}}}}{1 + e^{I_{h,ksj_{ks}}}}$$

Sandsynligheden for at blive vurderet i netop en kategori (dvs. den slags sandsynligheder, der er afbilledet i figuren), fx 'nogenlunde rigtig', er derfor

$$(16) \quad \begin{aligned} & \Pr(\text{'vurderet som 'nogenlunde rigtig' }) \\ &= \Pr(\text{'vurderet som 'nogenlunde rigtig' eller bedre'}) - \\ & \Pr(\text{'vurderet som 'helt rigtig' eller bedre'}) \end{aligned}$$

## 6.2. Lineære modeller, kardinale vurderinger

Man kan forestille sig, at man i stedet for at vurdere sagsbehandlingen kategorisk som i 6.1 vurderer sagsbehandlingen kontinuert, dvs. graden af 'forkerthed' kunne blive indplaceret på en skala fra fx 0 til 1. Hvis bedømmelsen kaldes  $V_{ksj_{ks}}$ , er modellen

$$(17) \quad V_{ksj_{ks}} = \alpha_k + \alpha_s + e_{ksj_{ks}}$$

En forskel på den lineære og kategoriserede model er, at  $V_{ksj_{ks}}$  kan observeres, hvilket  $I^*$  ikke kan. Den lineære model kan estimeres med mindste kvadraters metode. Som for modellen for kategoriserede vurderinger gælder, at hvis der kun ses på et spørgsmål ( $\alpha_s$  udgår), så findes mere direkte modeller til at sammenligne kommuner, nemlig t-test for forskelle mellem kommunegennemsnit. Ved kommunesammenligninger baseret på flere spørgsmål er de nødvendige antagelser om uafhængighed mellem restleddene som i den kategoriserede model.

Hvis sagsbehandlingen bedømmes med et tal mellem 0 og 1, og hvis man bruger modellen (17), kan man risikere, at den forudsagte værdi for nogle kommuner ligger over 1 eller under 0. Det kan løses ved alternativt at anvende modellen

$$V_{ksj_{ks}} = \frac{e^{\alpha_k + \alpha_s + e_{ksj_{ks}}}}{1 + e^{\alpha_k + \alpha_s + e_{ksj_{ks}}}}$$

(18) så

$$v_{ksj_{ks}} = \log\left(\frac{V_{ksj_{ks}}}{1 - V_{ksj_{ks}}}\right) = \alpha_k + \alpha_s + e_{ksj_{ks}}$$

Her er  $V_{ksj_{ks}}$  mellem 0 og 1, men transformationen betyder, at der kan estimeres lineært.

### 6.2.1. Vægtet firepunkts-model

I nogle tilfælde vil Ankestyrelsens eksperter ønske at sætte værdier (fx 1, 0,67, 0,33 og 0) på vurderingerne 'helt forkert', 'noget mangelfuldt', 'nogenlunde rigtigt' og 'helt rigtigt'. Hvis det fx ønskes, at kommuner, der svarer 'helt rigtigt' i stedet for 'nogenlunde rigtigt', skal præmieres yderligere ved en relativt høj  $\alpha_k$ -værdi, så kan værdierne ændres, så 'nogenlunde rigtigt' sættes til 0,5. Der kan udregnes et gennemsnit,  $M$ , for hver kommune eller hver kombination af kommune og spørgsmål (det sidste skal vælges, hvis kommunesammenligningen baseres på flere spørgsmål).

$$(19) \quad M = 0 \cdot \text{Pr}(\text{'helt rigtigt'}) + 0,33 \cdot \text{Pr}(\text{'nogenlunde rigtigt'}) \\ + 0,67 \cdot \text{Pr}(\text{'noget mangelfuldt'}) + 1 \cdot \text{Pr}(\text{'helt forkert'})$$

Dette  $M$  kan nu observeres og estimeres som funktion af  $\alpha_k$  og  $\alpha_s$ -parametre. Modellen er meget lig almindelig mindste kvadraters metode.

### 6.3 Eksperters fejl og idiosynkrasi

Som nævnt bedømmes kommunens sagsbehandling af Ankestyrelsens eksperter, der jo kan tage fejl. Desuden kan nogle eksperter kan være generelt mere strenge i deres bedømmelse end andre. Det rejser to problemer. Det ene er, om praksisundersøgelser overhovedet siger noget om, hvorvidt loven har været fulgt. Det andet er, at de estimerede kommuneforskelle i virkeligheden blot kan tænkes at dække over forskelle i eksperternes særtræk.

I afsnit 6.3.1 viser vi, at praksisundersøgelser siger noget om lovens overholdelse, hvis eksperterne blot har et minimum af uddannelse. Og da eksperterne i virkelighedens verden i sagens natur er særdeles kompetente, har vi dermed argumenteret for praksisundersøgelses værdi til trods for, at eksperterne kan tage fejl. I afsnit 6.3.2 forklares, at man kan korrigere for ekspertens identitet for at sikre sig, at kommuneforskelle ikke dækker over eksperters særtræk. Følgelig er det ikke en afgørende indvending mod praksisundersøgelser eller kommunesammenligninger, at Ankestyrelsens eksperter kan tage fejl.

### **6.3.1. Praksisundersøgelser har værdi selv om eksperter kan tage fejl**

Afsnittet falder lidt uden for arbejds papiret i øvrigt, og gælder praksisundersøgelser generelt. Vi forestiller os, at en sag eller en del af en sag kan tænkes af have en række forskellige afgørelser, og at en del af disse falder inden for loven. Vi forestiller os altså, at det i princippet objektivt kan afgøres, om en afgørelse falder inden for loven.

Lad hændelsen  $r$  betegne, at en konkret afgørelse er inden for loven. Lad  $a$  betegne den hændelse, at en ekspert A bedømmer, om afgørelsen er inden for loven, og lad tilsvarende  $b$  gælde for ekspert B.

Ekspert A kan være Ankestyrelsens ekspert, og B kan være kommunens sagsbehandler. Udgangspunktet er, at B har behandlet en sag, og hændelsen  $b$  er derfor indtruffet – B laver jo ikke med vilje fejl i behandlingen.

Vi skal vise, at

$$(20) \quad \Pr(r | ab) > \Pr(r | b)$$

dvs., der er større chance for, at behandlingen objektivt er rigtig, når både A og B mener det, end hvis kun B mente det.

Vi antager desuden, at



$$(21) \quad \Pr(a | r) > \Pr(a)$$

dvs. at der er større chance for, at A bedømmer en behandling rigtig, hvis den rent faktisk er det, end at A bedømmer en hvilken som helst sag rigtig (herunder også de objektivt forkerte). Betingelsen er meget svag og siger blot, at A ikke skal vurdere sagsbehandlinger helt tilfældigt. Betingelsen kan formentlig opfyldes blot med et minimum af uddannelse og kræver fx ikke, at Ankestyrelsens eksperter skal være dygtigere end kommunens sagsbehandlere.

Vi antager også, at

$$(22) \quad \begin{aligned} \Pr(ab | r) &= \Pr(a | r) \cdot \Pr(b | r) \\ \Pr(ab) &= \Pr(a) \cdot \Pr(b) \end{aligned}$$

Disse er uafhængighedsantagelser, der siger, at A skal vurdere sagsbehandlingen uden at lade sig påvirke af den afgørelse, der rent faktisk er truffet. Eksperten A skal ikke tænke, *'Det er nok rigtigt, når det nu er afgjort sådan i kommunen'* eller omvendt *'I dag vil jeg knageme vise mig effektiv og finde en masse fejl hos sagsbehandleren'*.

Med disse antagelser får man det ønskede (ligning (20))

$$(23) \quad \begin{aligned} \Pr(r | ab) &= \frac{\Pr(rab)}{\Pr(ab)} = \frac{\Pr(ab | r) \Pr(r)}{\Pr(a) \Pr(b)} = \frac{\Pr(a | r) \Pr(b | r) \Pr(r)}{\Pr(a) \Pr(b)} \\ &= \frac{\Pr(br)}{\Pr(r)} \frac{1}{\Pr(b)} \frac{\Pr(a | r)}{\Pr(a)} \Pr(r) = \Pr(r | b) \frac{\Pr(a | r)}{\Pr(a)} \\ &> \Pr(r | b) \end{aligned}$$

### 6.3.2. Korrektion for eksperters idiosynkrasi i kommunesammenligning

Hvis eksperterne indekseres med  $e$ , kan ekspert  $e$ 's strenghed i bedømmelsen af sagerne indgå i modellen med parameteren  $\alpha_e$ . Desto højere  $\alpha_e$ , desto strengere er  $e$  og desto større sandsynlighed er der for, at sagens bedømmes kritisk. Ligning (7) omformes til

$$(24) \quad I_{ksj_{ks}}^* = \alpha_k + \alpha_s + \alpha_e + e_{ksj_{ks}}$$

For at kunne skelne  $\alpha_k$  fra  $\alpha_e$  skal sagerne fra hver kommune behandles af forskellige eksperter. Det øger den statistiske usikkerhed i modellen, at eksperterne ikke vurderer ens. Strengt taget kunne man løse problemet ved, at alle sager fra en praksisundersøgelse blev løst af samme ekspert.

## 7. Udmøntning

Ankestyrelsen planlægger fremover at lave ca. 4 praksisundersøgelser om året, med ca. 150 sager i hver. Der vil fremover være 5 statsforvaltninger, som vil gennemføre 10 praksisundersøgelser årligt. I disse vil indgå mellem 50 og 120 sager afhængigt af regionernes størrelse. Se i øvrigt Ankestyrelsens rapporter, Ankestyrelsen (2006 a,b).

I hver praksisundersøgelse vil der indgå ca. 10 sager fra hver kommune.

Kommunerne deltager i forskellige praksisundersøgelser og kommunesammenligningen kan foretages over forskellige praksisundersøgelser. Som nævnt i afsnit 6.1.2 skal det imidlertid sikres, at kommuneeffekter og effekter fra de forskellige spørgsmål (fx forskellige praksisundersøgelser) koordineres, så de to effekter kan adskilles statistisk.

I de følgende tabeller er der eksempler på, hvordan kommunernes deltagelse i praksisundersøgelser kan koordineres, så man hhv. kan og ikke kan sammenligne kommuner.

I det følgende tilfælde kan kommune A og B sammenlignes, C og D kan sammenlignes, men fx A og C kan ikke sammenlignes, fordi man ikke ved, om forskelle skyldes, at de to praksisundersøgelser er lige lette at sagsbehandle.

**Tabel 5a. Adskillelse af kommune og spørgsmålseffekt**

Kommune	Praksisundersøgelse nummer	
	1	2
A	X	
B	X	
C		X
D		X

I det følgende tilfælde kan alle kommuner sammenlignes. Kommune A og D kan sammenlignes, fordi de er 'indirekte forbundne', i og med at B og C deltager i begge praksisundersøgelser. Via kommune B og C kan man nemlig finde forskellen i de to praksisundersøgelser sværhedsgrad, og dermed "rense" for denne forskel for kommune A og D.

**Tabel 5b. Adskillelse af kommune og spørgsmålseffekt**

Kommune	Praksisundersøgelse nummer	
	1	2
A	X	
B	X	X
C	X	X
D		X

I det følgende tilfælde kan A, B, C og D hhv. E, F, G og H sammenlignes indbyrdes, men to kommuner fra hver af de to grupper kan ikke sammenlignes.

**Tabel 5c. Adskillelse af kommune og spørgsmålseffekt**

Kommune	Praksisundersøgelse nummer			
	1	2	3	4
A	X			
B	X	X		
C	X	X		
D		X		
E			X	
F			X	X
G			X	X
H				X

I praksis bør der være en del kommuner, der deltager i mere end en praksisundersøgelse, for ellers bliver estimaterne over sværhedsgraderne i praksisundersøgelserne for statistisk usikre.

## 8. Konklusion

I arbejdsrapporten foreslås først og fremmest simple logistiske modeller til sammenligning af den juridiske korrekthed af kommunernes sagsbehandling. Af statistiske grunde kan blive nyttigt at sammenligne kommunerne ved at se på flere led af sagsbehandlingen samtidig. Der foreslås modeller til at tage højde for de problemer, der evt. opstår derved (afsnit 3).

Det foreslås også at eksperimentere med metoder til såkaldte multiple sammenligninger (afsnit 5). Metoderne er lidt mere forsigtige med at udpege kommuner med mange fejl i sagsbehandlingen. Metoderne kan bruges, hvis man ikke har speciel interesse i en konkret kommune, men i om der er nogen kommuner overhovedet, der adskiller sig fra andre.

En egentlig empirisk sammenligning af kommunerne kan ikke foretages, før nye data fra Ankestyrelsens praksisundersøgelser foreligger.

## Litteratur

- Agresti, Alan (2002):** "Categorical Data Analysis", Wiley-Interscience.
- Ankestyrelsen (2006):** "Benchmarking af kommunernes sagsbehandling", København, [www.ast.dk](http://www.ast.dk).
- Dunnet, Charles W. (1955):** "A multiple Comparison Procedure for Comparing Several Treatments with a Control", Journal of the American Statistical Association, Vol. 50, p. 1096-1121.
- Dunnet, Charles W. (1980):** "Pairwise Multiple Comparison in the Homogeneous Variance, Unequal Sample Size Case", Journal of the American Statistical Association, Vol. 75, p. 789-795.
- Gregersen, Ole (2000):** "Benchmarking af kommunernes førtidspensionspraksis", Arbejdsnotat, Socialforskningsinstituttet.
- Lindermann, Gitte og Ole Gregersen (2001):** "Benchmarking på førtidspensionsområdet - virker Den Sociale Ankestyrelses praksiskoordinering", rapport 01:2, Socialforskningsinstituttet, København, [www.sfi.dk](http://www.sfi.dk).
- Rasmussen, Martin (2005):** "Economic Incentives and Disability Benefit in Denmark - Simple Analysis of a Homogenous sample", working paper 8:2004, Socialforskningsinstituttet, København, [www.sfi.dk](http://www.sfi.dk).
- Toothaker, Larry E. (1993):** "Multiple Comparison Procedures", Sage Publications.

I øvrigt: Se [www.ast.dk](http://www.ast.dk) om praksisundersøgelser. Dokumentationen fra SAS (SAS OnlineDoc, version 8.) giver supplerende forklaring på de beskrevne tests.