Beskrivende statistisk metode – til undersøgelse af kommunesammenlægningen

Lars Ravn-Jonsen

Juni, 8:2005

akf www.akf.dk

Beskrivende statistisk metode
til undersøgelse af kommunesammenlægningen

af Lars J. Ravn-Jonsen

Juni 2005

Abstract

The object of the paper is to explain some descriptive statistical methods usable to analyse the distributional consequences of the coming merger among Danish municipalities. The methods used are box plots and calculation of the expected mean error, average and gini-coefficients as expressed in some test numbers.

The paper explains how the concept of homogenizing can be interpreted statistically: as a change from a cluster distribution towards a random distribution and further towards a systematic distribution. A cluster distribution can also be described by a positive correlation between the individuals, a random distribution by nil correlation between the individuals and a systematic distribution by a negative correlation between the individuals; consequently homogenizing can be interpreted as a decrease in correlation. A homogenizing will be expressed as a decrease in the mean error of the average, too. By the merger of smaller municipalities to fewer, but bigger municipalities, the municipal distribution of the properties will change. How this change in the distribution can be observed in box plots and test numbers is described and related to the notion of random, correlated and thereby to cluster and systematic.

In Christoffersen and Ravn-Jonsen (2005) the described methods are applied for analysing of different aspects related to the in 2007 planned fusion of the Danish municipalities. The report illustrates the potential that the graphical methods of this paper have in explaining and analysing otherwise hard accessible material by the mean of illustrations.

Forord

Dette papir redegør for en enkelt statistisk og grafisk beskrivende undersøgelsesmetode til belysning af effekten af den kommende kommunesammenlægning i 2007. Metoden skal bruges til at få indsigt i, hvordan sammenlægningerne vil påvirke fordelingen af forskellige egenskaber, fx om hvorvidt sammenlægningen medfører en homogenisering af kommunernes indbyggertal. Dette papir er oprindeligt skrevet i samarbejde med Henrik Christoffersen som metodeafsnit til en rapport om disse forhold; men da afsnittet er en let forståelig indføring til nogle statistiske metoder og derfor er alment anvendeligt ved andre analyser af samme karakter, er afsnittet udgivet som selvstændigt papir.

Lars J. Ravn-Jonsen
Juni 2005

Indholdsfortegnelse

1. Undersøgelsesmetode

I foråret 2005 kendes resultaterne af kommunernes forhandlinger om sammenlægninger i forbindelse med den forestående strukturreform. Ved sammenlægningerne af mindre kommuner til færre, men større kommuner, vil den kommunale fordeling af forskellige egenskaber ændres. Hvordan disse ændringer i fordelingen kommer til udtryk såvel grafisk som i sammenligningstørrelser er emnet for dette papir. Det er således hensigten med dette papir at beskrive redskaber til belysning af, hvordan kommunesammenlægningen vil påvirke fordelingen af en given egenskab, fx af hvorvidt kommunesammenlægningen vil medføre en homogenisering af kommunernes indbyggertal.

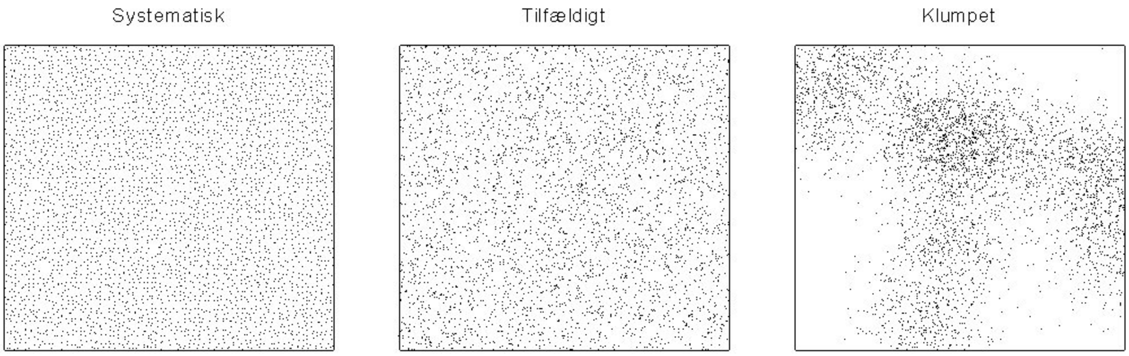
Homogenisering er en proces, der er kendt fra mejeriprodukter. Her betyder homogenisering at behandle mælken, hvor fedtet findes fordelt i klumper, således at fedtet findes jævnt i mælken. Hvis der ved homogenisering alment forstås en proces, der mere eller mindre har de egenskaber som i mejeriet, er det en proces, der kan beskrives statistisk, idet situationen før homogeniseringen kan beskrives som »klumpet« (på engelsk bruges »cluster«), den jævne fordeling i mælken vil statistisk være »tilfældig« og en endnu mere jævn fordeling, som af molekyler i et krystal vil være »systematisk«.

Inden uddybning af begreberne klumpet, tilfældig og systematisk er en forklaring af nogle statistiske størrelser nødvendig. Gennemsnit og middelværdi formodes bekendte og betegner næsten det samme, nemlig midtpunktet udregnet ved en momentberegning (første moment). Forskellen mellem de to begreber vil ikke have betydning i dette papir. Et andet statistisk begreb er spredningen. Spredningen er et mål for, hvor forskelligt et antal stikprøver vil være og angiver, hvor meget en stikprøve »typisk« vil afvige fra middelværdien. Korrelation er et udtryk for, hvor meget man ligner hinanden, fx ved bosætning: Hvis man bosætter sig tæt på andre, vil der være positiv korrelation, hvis man helst vil have stor afstand (eller en ønsker en minimumsafstand), vil der være negativ korrelation. Hvis man bosætter sig uafhængigt af andre, vil korrelationen være nul. En klumpet fordeling kan derfor også beskrives ved en positiv korrelation, tilfældig ved nul korrelation og systematisk ved negativ korrelation mellem individerne; homogenisering kan derfor også beskrives som en mindskelse af korrelationen.

En proces, der tilfældigt fordeler punkter i planet (rummet, tiden), kaldes en poisson proces (Skovgaard et al. 1999). Hvis der udlægges tilfældig prøveflade af samme størrelse i et plan, hvor punkterne er fordelt tilfældigt, vil gennemsnittet af punkter pr. prøveflade og spredningen imellem prøvefladerne være af samme størrelse. Hvis fordelingen derimod er klumpet, vil spredningen af antallet i en prøveflade være større end gennemsnittet. En fordeling kan godt have mindre spredning end gennemsnittet. Den kaldes da systematisk. I figur 1.1 er vist eksempler på de tre typer fordeling. En fordeling kan godt være systematisk på en skala og klumpet på en anden, fx vil bygninger i et villakvarter udvise en systematisk fordeling, når der ses på et kvarter, men hvis skalaen ændres, så det omslutter flere kvarterer og åbent land, vil fordelingen være klumpet.

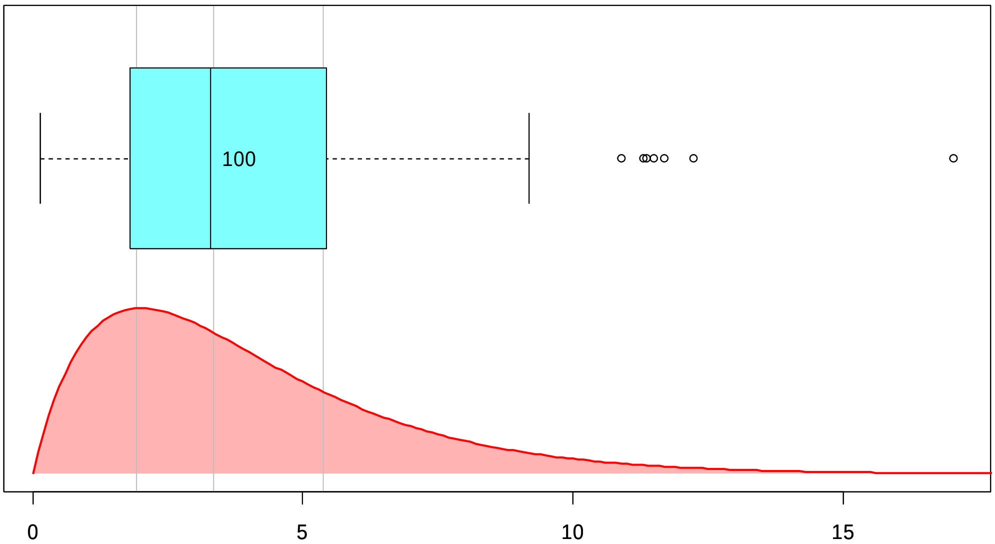
Det er ikke kun, når der er tale om punkter i et plan, der kan tales om klumpet, tilfældig eller systematisk fordeling. De samme begreber kan bruges på egenskaber, fx fordelingen af blå øjne i en befolkning. Det er umiddelbart oplagt, at mange egenskaber ved befolkningen er klumpet fordelt. Fx er bosætning klumpet, hvor klyngerne kaldes byer, og da landmænds erhverv er knyttet til det åbne land og ikke til byer, er egenskaben »landmand« klumpet fordelt og dermed også »andelen af beskæftigede i primære erhverv«. Det er da heller ikke formålet umiddelbart at diskutere, hvorvidt forskellige karakteristika ved befolkningen er tilfældigt fordelt eller ej, men at analysere, hvorvidt kommunesammenlægningen medfører, at den klumpede fordeling bliver forstærket eller mindsket; altså om der vil foregå en homogenisering eller ej. Nu vil sammenlægningen af kommuner ikke medføre, at folk er anderledes bosat, dvs. den grundlægende fordeling er den samme, det er fordelingen af egenskaber mellem kommunerne, der ændres.

Figur 1.1
Eksempel på en systematisk, en tilfældig og en klumpet fordeling af omkring 4.800 punkter på en plan



Til undersøgelse af fordelingen af egenskaber i kommunerne kan boksdiagrammer benyttes (Tukey 1977, Chambers et al. 1983). Boksdiagrammet bygger på inddeling af data i kvartiler. Dvs. datasættet sorteres efter værdi og deles i 4 lige store dele. Den værdi, der adskiller den laveste del fra resten er 1. kvartil, den værdi, der adskiller den øverste halvdel fra resten, er 3. kvartil. Den værdi, der adskiller øverste fra nederste halvdel, er 2. kvartil og kaldes som regel medianen. Medianen er således midten af data fundet efter antal. I figur 1.2 er vist et eksempel på et boksdiagram. Dette består af en central boks, der er tegnet fra 1. til 3. kvartil. Bredden af boksen har ingen betydning. Inde i boksen er medianen tegnet med en streg. På hver side af boksen er tegnet en bakkenbart. Længden af bakkenbarterne er til henholdsvis mindste og største værdi, dog tegnes bakkenbarterne aldrig længere end halvanden gange længden af boksen. Hvis det er værdier, der ligger uden for bakkenbarterne, dvs. at de er større end 3. kvartil plus halvanden gang differencen mellem 1. og 3. kvartil eller mindre end 1. kvartil minus halvanden gang differencen mellem 1. og 3. kvartil, betegnes de som udliggere og angives som enkelte punkter. Boksdiagrammet er en illustration af data, hvor det er muligt at vurdere skævhed i fordelingen og tilstedeværelse af ekstreme værdier. Hvis egenskaberne er fordelt tilfældigt på befolkningen i landet, vil egenskaberne på kommuneniveau være omtrent normalfordelte[1]. En normalfordeling er symmetrisk og det forventede antal udliggere i en normalfordeling er 0,95 og 0,34 i hver ende med en population på henholdsvis 275 og 102, altså typisk mellem 0 og 1 udligger. I figur 1.2 er vist en fordeling, der ikke er normalfordelt, og hvordan dette kan læses af boksdiagrammet. Når boksdiagrammerne for mange af de undersøgte parametre vedrørende kommunerne udviser skævhed eller har mange udliggere, er det fordi folk med disse egenskaber er bosat klumpet i kommunerne.

Figur 1.2
Illustration af boksplot



Bem.: I boksplottet er 100 tilfældige tal fra en χ^2-fordeling med 4 frihedsgrader illustreret. Der er tre ting, der viser, at fordelingen er skæv: (i) medianen ligger tættere på 1. kvartil end 3. kvartil. (ii) bakkenbarterne er af forskellig længde, og (iii) der er mange udliggere, nogle med store værdier.

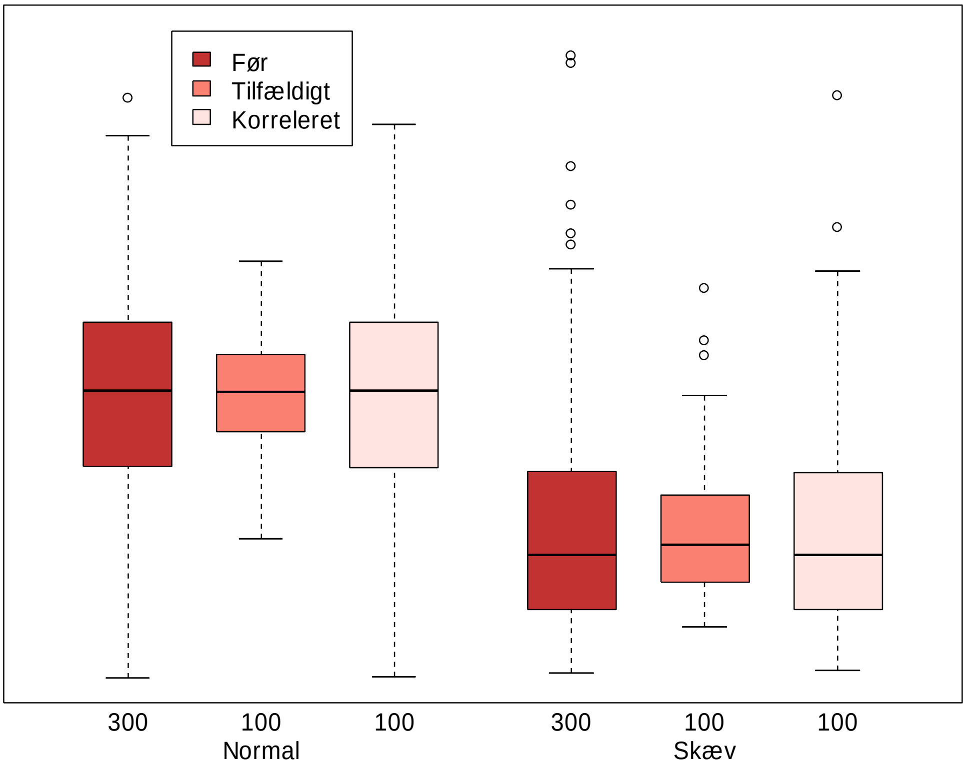
Den nederste kurve viser tætheden af en χ^2-fordeling med 4 frihedsgrader og de lodrette grå linjer denne fordelings kvartiler. Kurven illustrerer den fordeling, de 100 tilfældige tal er trukket fra, bemærk hvor tæt boksplottet rammer de sande kvartiler.

Når kommunerne sammenlæges sker der det, at landets befolkning, fra at være inddelt i 275 grupper, bliver inddelt i 102 grupper, og da befolkningen er den samme, betyder det, at grupperne bliver større. Alene det, at de grupper, befolkningen er inddelt i, bliver større, vil, hvis den foregår tilfældigt, mindske spredningen, mindske skævhed, og de fleste udliggere vil blive lagt sammen med en mere almindelig kommune og derved ikke optræde som udligger mere. Men det er ikke muligt at sammenlægge kommuner tilfældigt: dels kan kommuner kun lægge sig sammen med andre i fysisk nærhed, dels er der et ønske om en minimumsstørrelse for indbyggertal, dvs. de mindre kommuner har mere tilskyndelse til at lægge sig sammen end større. Hvis kommuner lægger sig sammen med andre, der ligner dem mest muligt, en korreleret sammenlægning, vil spredningen, skævhed og mønstret af udliggere bevares, mens antallet af udliggere vil

1 Under forudsætning af, at egenskaben bliver målt på den rigtige måde, fx som en frekvens: »andelen med lang uddannelse« eller »andelen af boliger, der er almennyttige«.

mindskes. Den tilfældige og korrelerede sammenlægning kan således observeres i boksdiagrammet. I figur 1.3 er vist eksempler på en tilfældig og korreleret sammenlægning af 300 tilfældige tal fra henholdsvis en normalfordeling og en skævfordeling (en χ^2 med 4 frihedsgrader). Sammenlægningerne kan også foregå, så de systematisk mindsker forskelle, eller så de systematisk øger forskellene. Det sidste kan fx ske ved, at udliggere ikke indgår i sammenlægninger, hvorved deres andel af kommunerne øges.

Figur 1.3
Sammenlægning af 300 tilfældige tal



Bem.: Til venstre er 300 tilfældige normalfordelte tal udtrukket og lagt sammen til 100 ved henholdsvis en tilfældig og en korreleret sammenlægning. Til højre er det samme vist for en skæv fordeling. Den tilfældige sammenlægning er foretaget ved at gruppere de 300 tilfældige tal i grupper af tre og tage gennemsnittet. Ved den korrelerede er de 300 tilfældige tal først sorteret, således at nummer 1, 2 og 3 efter størrelse er i gruppe, nr. 4, 5 og 6 efter størrelse er i gruppe osv.

Der er yderligere mulighed for at udregne nogle sammenligningsstørrelser til at bedømme, hvorvidt sammenlægningen er tilfældig, korreleret eller systematisk den ene eller anden vej. Det er her nødvendigt at skelne mellem absolutte tal, fx indbyggertallet, og relative tal som »andelen med lang uddannelse«. For absolutte tal er de beregnede sammenligningsstørrelser relativt gennemsnit og relativ spredning. For relative tal er den beregnede størrelse relativ varians (variansen er kvadratet af spredningen). For alle sammenligningsstørrelser skal relativ forstås som »efter sammenlægningen delt med før sammenlægningen«. I tabel 1.1 er forventningerne til disse sammenligningsstørrelser angivet.

Tabel 1.1
Forventede opførsel for sammenligningsstørrelser for absolutte og relative tal ved henholdsvis en tilfældig og en korreleret sammenlægning

	Tilfældig	Korreleret
Absolutte tal	Relativ spredning er omtrent samme som kvadratroden af relativt gennemsnit	Relativ spredning er omtrent samme som relativt gennemsnit
Relative tal	Relativ varians er omtrent 0,61	Relativ varians er omtrent 1

Tabel 1.2
Sammenhæng mellem sammenlægningsmønsteret og det relative gennemsnit af relative parametre

Relativt gennemsnit større end 1	Relativt gennemsnit mindre end 1
Sammenlægningen foregår overvejende mellem kommuner uden denne egenskab	Sammenlægningen foregår overvejende mellem kommuner med denne egenskab

Yderligere benyttes for relative tal det relative gennemsnit og den relative ginikoefficient som sammenligningstørrelser. Det relative gennemsnit fortæller om sammenlægningsmønstre, se tabel 1.2, og den relative ginikoefficient fortæller, om sammenlægningen foregår tilfældigt eller bevarer en klumpet fordeling. Den relative ginikoefficient er størst, hvis alle efter sammenlægningen har samme kommunegennemsnit for denne egenskab som før og vil da være 1. Hvis sammenlægningen foregår tilfældigt, og der sker den dermed forventede udjævning af kommunegennemsnit, vil ginikoefficienten være omtrent 0,5.

For uddybning af begrebet ginikoefficient og baggrunden for forventningerne som beskrevet i de foregående afsnit, henvises til boks 1.1, 1.2 og 1.3. Indholdet i boksene er af teoretisk karakter og kan overspringes.

Boks 1.1
Forventninger til absolutte tals middelværdi og spredning ved tilfældige og korrelerede sammenlægninger. For regneregler vedrørende stokastiske variabler henvises til Skovgaard et al. (1999)

Som eksempel på absolutte tal bruges indbyggertal. En kommunes størrelse før sammenlægningen er givet af en sandsynlighedsfunktion $f_1(x)$ (sandsynligheden for, at kommunen har størrelsen x). Denne funktion har middelværdien μ_1 og en varians på σ_1^2. En sammenlægning af to tilfældige kommuner vil da medføre en forventning om, at den nye sandsynlighedsfunktion for kommunestørrelsen $f_2(x)$ har middelværdien $\mu_2 = 2\mu_1$ og varians $\sigma_2^2 = 2\sigma_1^2$, dvs. man kan forvente, at $\frac{\mu_2}{\mu_1} = \frac{\sigma_2^2}{\sigma_1^2}$. Men det er ikke tilfældigt, når kommuner lægger sig sammen: dels kan man kun lægge sig sammen med andre i fysisk nærhed, dels er der et ønske om en minimumsstørrelse, dvs. de mindre har mere tilskyndelse til at lægge sig sammen end større. Hvis der tages hensyn til, at sammenlægningen ikke er tilfældig, og at to kommuner tæt på hinanden har sandsynlighed for at ligne hinanden, vil den nye varians blive $\sigma_2^2 = 2\sigma_1^2 + 2\,\mathrm{cov}_1$, hvor cov er covariansen mellem de to kommuner.

Covariansen ligger mellem $\pm$ varians, dvs. $-\sigma^2 < \text{cov} < \sigma^2$ og er lig variansen, når de to kommuner altid er 100% ens. I det ekstreme tilfælde, hvor en kommune altid lægger sig sammen med en, der er magen til sig selv, kan det forventes, at $\sigma_2^2 = 4\sigma_1^2$, altså en forventning om, at $\frac{\mu_2}{\mu_1} = \frac{\sigma_2}{\sigma_1}$. Dette svarer til en opskalering – alle størrelser og uligheder skaleres op. Der er altså to interessante størrelser til forståelse af sammenlægninger: det relative gennemsnit $m = \frac{\mu_2}{\mu_1}$ og den relative spredning $s = \frac{\sigma_2}{\sigma_1}$. s er altid mindre end m, medmindre der bevidst vælges kommunesammenlægninger for at øge forskellen. Hvis s er tæt på m, betyder det, at kommuner lægger sig sammen med kommuner, der ligner dem selv. Hvis s ligger tæt på $\sqrt{m}$, er der tale om en tilfældig sammenlægning, og hvis s er mindre end $\sqrt{m}$, er der tale om bevidst homogenisering. Dette gælder for alle absolutte tal.

Boks 1.2
Forventninger til relative tals middelværdi og varians ved tilfældige og korrelerede sammenlægninger. For regneregler vedrørende stokastiske variabler henvises til Skovgaard et al. (1999)

For alle størrelser, der er angivet med et relativt tal ϕ, kan vi antage, at dette ϕ stammer fra en stokastisk proces, der tager kommunen som input og tildeler et ϕ_x til kommunen x; lad denne proces have tæthedsfunktionen $f(x)$, der har en middelværdi μ og en varians på σ^2. Hvis to kommuner med ens indbyggertal lægger sig sammen, vil deres nye frekvens være

$$\phi_{1+2} = \tfrac{1}{2}(\phi_1 + \phi_2)$$

og deres varians

$$\begin{aligned}\sigma_{1+2}^2 &= \tfrac{1}{4}\sigma_1^2 + \tfrac{1}{4}\sigma_1^2 + 2\,\text{cov}(\tfrac{1}{2}\phi_1, \tfrac{1}{2}\phi_2)\\ &= \tfrac{1}{2}\sigma_1^2 + \tfrac{1}{2}\text{cov}(\phi_1, \phi_2)\end{aligned}$$

og altså en forventning om, at variansen er konstant, hvis de kommuner, der lægges sammen, er valgt, så de er ens. Variansen halveres, hvis kommunerne sammenlægges tilfældig, og hvis der er bevidst udjævning, skal variansen mere end halveres.

Nu er kommunerne, der lægges sammen, ikke med samme indbyggertal. Hvis de har indbyggertal på henholdsvis aI og bI, hvor $a + b = 1$, og den nye kommune derfor har I indbyggere, vil

$$\phi_{1+2} = a\phi_1 + b\phi_2$$

den nye varians

$$\begin{aligned}\sigma_{1+2}^2 &= a^2\sigma_1^2 + b^2\sigma_1^2 + 2\,\text{cov}(a\phi_1, b\phi_2)\\ &= \left(a^2 + b^2\right)\sigma_1^2 + 2ab\,\text{cov}(\phi_1, \phi_2)\end{aligned}$$

Det ses, at hvis covariansen er lig variansen, vil den nye varians være uforandret, hvorimod hvis der vælges tilfældige kommuner, vil variansen blive mindst, hvis kommunerne er ens i indbyggertal. Eksemplet kan let udvides til flere kommuner i sammenlægningen, da bliver variansen

$$\sigma^2_{samlet} = \left(a^2 + b^2 + c^2 + \cdots\right)\sigma_1^2 +$$
$$2ab\,\mathrm{cov}(\phi_1, \phi_2) + 2ac\,\mathrm{cov}(\phi_1, \phi_2) + 2bc\,\mathrm{cov}(\phi_1, \phi_2) + \cdots$$

Ved kommunesammenlægningen vil der, hvis sammenlægningen er tilfældig, derfor gælde:

$$v = \frac{\sigma^2_{efter}}{\sigma^2_{før}} \approx a^2 + b^2 + c^2 + \cdots = k$$

Den relative varians v kan altså sammenlignes med k, summen af kvadratet på de gamle kommuners indbyggertal, relativt i forhold til sammenlægningskommunens. En estimation af k som middelværdien ved den gennemførte kommunesammenlægning giver $k \sim 0,61$. Hvis sammenlægningen er sket, så ens kommuner har lagt sig sammen, dvs. $\mathrm{cov} \approx \sigma^2$, vil v være tæt på 1. v kan godt blive over 1, hvis der bevidst udvælges for at øge forskelle.

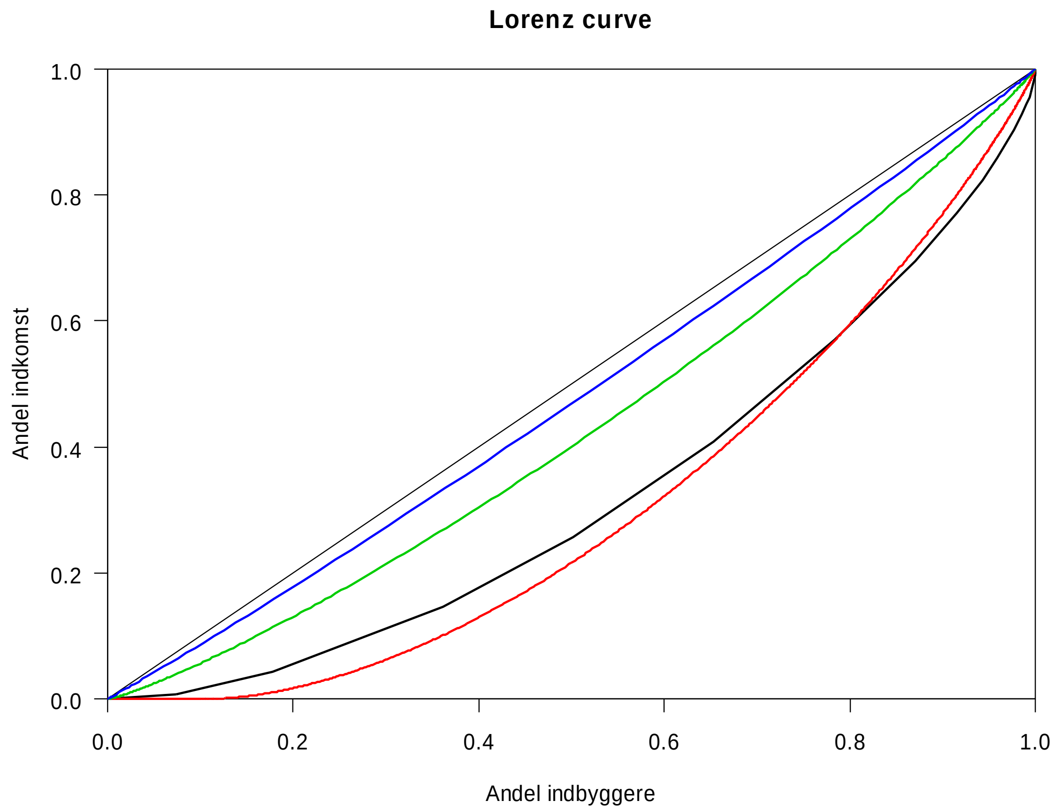
En anden teststørrelse af interesse er relativt gennemsnit:

$$f = \frac{\overline{\phi_{efter}}}{\overline{\phi_{før}}}$$

Hvis f er over 1, er det fordi overvejende kommuner med lav frekvens har lagt sig sammen, og hvis den er under 1, er det fordi overvejende kommuner med høj frekvens har lagt sig sammen.

Boks 1.3
Forklaring af ginikoefficient og forventninger til denne ved sammenlægninger

Figur 1.4
Lorenzkurve over indkomstfordelingen i Danmark med sort. Med rød er vist en normalfordeling med samme middelværdi og spredning som data for den sorte kurve, med grønt er vist en tilfældig gruppering af 10 borgere, og med blåt er vist en tilfældig gruppering på 100



Lorenzkurven for en given variabel, fx indkomst, er en kurve, der efter sortering efter indkomst, plotter andelen af samlet indkomst mod andelen af befolkningen, summeret fra den fattige side. I figur 1.4 er data for antal personer klassificeret efter personindkomst for 2002 (fra Danmarks Statistik) brugt til at konstruere den sorte Lorenzkurve. Data indeholder 4,392 millioner personer og er således ikke blot erhvervsaktive, men alle der har en indtægt. Arealet mellem diagonalen og Lorenzkurven er et udtryk for, hvor uensartet indkomsten er fordelt. Dette areal bruges til at udregne ginikoefficienten. Ginikoefficienten er defineret som forholdet mellem arealet mellem Lorenzkurven og diagonalen og arealet under diagonalen (Cowell 1995). Ginikoefficienten ligger mellem 0 og 1, hvor yderpunktet 0 opnås, når alle har samme indkomst, og 1 er, hvis én tager alt.

Når individerne samles, og der udregnes et gennemsnit for den enkelte gruppe, vil Lorenzkurven forventes at nærme sig diagonalen og ginikoefficenten falde. Hvis grupperingen sker tilfældigt, vil fordelingen af gennemsnittene for grupperne nærme sig en normalfordeling, uanset hvilken fordeling der i er udgangspunktet (den centrale grænseværdisætning). Fordelingen vil forventes at være tilnærmelsesvis normalfordelt

$N(\mu, \frac{1}{n}\sigma^2)$, hvor μ er middelværdien i den oprindelige fordeling, σ er spredningen i den oprindelige fordeling, og n er antallet af individer, der er i grupperne.

Indkomstfordelingen i Danmark, der er plottet i den sorte Lorenzkurve, har gennemsnit $\hat{\mu} = 223.101 \text{ kr.}$ og stikprøvesprednig på $s = 175.412$. En normalfordeling med samme middelværdi og spredning er tegnet som Lorenzkurve med rødt. Når den sorte og den røde ikke falder sammen, er det fordi den faktiske indkomstsfordeling i forhold til normalfordelingen, har en meget lang hale i højindkomstenden. Med grønt er med en Lorenzkurve vist, hvad der sker, når folk grupperes tilfældig med 10 personer i hver gruppe. Hvis grupperingen sker tilfældigt, vil det være svært allerede med så lille en gruppering at skelne gruppegennemsnitsfordelingen fra normalfordelingen. Med blåt er vist en tilfældig gruppering med 100 i hver gruppe.

I tabel 1.5 er ginikoefficienter for kurverne i figuren angivet for yderligere gruppering. 10.260 og 41.140 svarer til medianen i kommunernes indbyggertal henholdsvis før og efter sammenlægning. Det ses, at ved tilfældige grupperinger falder ginikoefficienten hurtigt. Hvilken reduktion i gignikoefficenten, der så kan forventes ved en kommunesammenlægning, hvis denne foregår tilfældigt, afhænger både af middelværdi og spredning; men er typisk sådan, at forholdet mellem ginikoefficienterne efter/før er omkring 0,5.

Tabel 1.5
Gininkoefoefficienter for indkomstfordelingen i Danmark, dels efter faktiske data, dels for normalfordelinger svarende til at udregne gennmsnit over en grupperet befolkning. Farver henviser til figur 1.4

Ginikoefficient	Antal	Fordeling	Farve
0,35		data	sort
0,39	1	normal	rød
0,14	10	normal	grøn
0,044	100	normal	
0,014	1000	normal	
0,0044	10260	normal	
0,0022	41140	normal	

For binomiale størrelser, som fx lang uddannelse, er forholdet nemmere: her kendes hyppigheden fra landsgennemsnittet, og spredningen er givet ved binomialfordelingen. Her kan den forventede reduktion angives som en funktion af hyppigheden. Den forventede reduktion er for alle hyppigheder omkring 0,49 (med hyppigheder mellem 0,01 og 0,99).

2. Metodens anvendelse

De beskrevne metoder er anvendt i Christoffersen og Ravn-Jonsen (2005) til analyse af forskellige aspekter i forbindelse med den kommende kommunesammenlægning i 2007. Analysen er lavet på baggrund af kommunernes tilbagemeldinger til Indenrigsministeriet ultimo 2004 – offentliggjort af Indenrigs- og Sundhedsministeriet (2005) – og data fra akf's lokaløkonomiske databank. I rapporten analyseres indbyggertal, areal, befolkningstæthed, pendlingsforhold, indkomst, skatter, og forskellige socioøkonomiske faktorer, der påvirker indtægter og udgifter i kommunen og kommunens potentiale for udvikling. Rapporten illustrer den i dette papir beskrevne grafiske metodes muligheder for at forklare og analysere ellers svært tilgængeligt materiale ved hjælp af illustrationer. I tabel 2.1 er gengivet resultaterne i hovedtræk, for flere detaljer henvises til rapporten.

Tabel 2.1
Virkninger af kommunesammenlægningerne med udgangspunkt i kommunernes indberetning til Indenrigsministeriet ultimo 2004

Parameter	Virkningen af de aftalte sammenlægninger
Indbyggertal	Kraftig homogenisering ved, at fortrinsvis små kommuner lægger sig sammen
Areal	Kraftig øgning af forskelle, meget store kommuner dannet i alle regioner, bortset fra Region Hovedstaden
Pendlingsforhold	I Jylland og på Fyn er der tendens til, at store kommuner efter sammenlægningen udgør én pendlingsregion. På Sjælland er pendlingsforholdet afhængig af afstanden til København. Kun store kommuner langt fra København udgør en pendlingsregion
Primær indkomst, pr. indbygger Indkomstskat, pr. indbygger Disponibel indkomst, pr. indbygger	De tre parametre er stærkt korrelerede, indkomstskatten er analyseret i detaljer. Sammenlægningen øger uhomogeniteten. Øgningen foregår, fordi mange kommuner med høje indtægter ikke indgår i sammenlægninger; og når de gør, da lægger de sig sammen med andre, der også har høje indtægter. Kommuner med høje indkomster findes kun i hovedstadsområdet, og der er således store regionale forskelle, men også specialet store forskelle inden for Region Hovedstaden
Andele af beskæftigede i primære erhverv	Homogenisering ud over, hvad der kan forventes ved en tilfældig sammenlægning. Homogeniseringen er foregået ved, at sammenlægningerne foregår i kommuner med høje værdier af disse parametre
Andele af beskæftigede i offentlige serviceerhverv	Homogenisering, der svarer til en tilfældig sammenlægning. Udliggere, der er i den øvre ende af fordelingen før sammenlægningen, er væk efter sammenlægningen
Andele af beskæftigede i private serviceerhverv	Sammenlægningen bevarer spredningen imellem kommunerne. Kommuner med høje værdier af denne parameter er ikke indgået i sammenlægninger, derved er den øvre hale bevaret

Parameter	Virkningen af de aftalte sammenlægninger
Andelen af almennyttige boliger, Sociale udgifter, pr. indbygger Indvandrere, pr. indbygger	Sociale udgifter bevarer spredning, og andelen af almennyttige boliger og indvandrere pr. 1000 indbyggere øger unhomogeniteten imellem kommunerne. Alle tre bevarer en øvre hale, dvs. der er få kommuner med store værdier af disse tre parametre, der indgår i sammenlægninger
Andel af arbejdsstyrken uden uddannelse Faglærtes andel af arbejdsstyrken Andel af arbejdsstyrken med kort uddannelse Andel af arbejdsstyrken med mellemlang uddannelse	Arbejdsstyrkens fordeling på uddannelsesniveau er fordelt meget forskelligt kommunerne imellem. Det ændrer sammenlægningen ikke ved: spredningen bibeholdes
Andel af arbejdsstyrken med lang uddannelse	Uhomogeniteten øges. Højtuddannede har kraftig tendens til at bosætte sig klumpet, og fordelingen har lang øvre hale både før og efter, og udliggerne indgår ikke i sammenlægninger. Fordelingen er også skæv regionalt med Region Hovedstaden med langt større andel af personer med lang videregående uddannelse, men også med store forskelle regionens kommuner imellem

3. Litteratur

Chambers, John; William Cleveland, Beat Kleiner og Paul Tukey (1983): *Graphical Methods for Data Analysis*, Wadsworth.

Christoffersen, Henrik og Lars Ravn-Jonsen (2005): *Det nye Danmarkskort – en analyse af uhomogeniteter i kommunestrukturen før og efter kommunesammenlægningen*. akf forlaget. København.

Cowell, F.A. (1995): *Measuring Inequality*, Prentice Hall/Harvester Wheatshef.

Indenrigs- og Sundhedsministeriet (2005): *Tilbagemelding fra kommuner den 18. januar 2005 med ønsker til fremtidigt tilhørsforhold i forbindelse med kommunalreformen. Indenrigs- og Sundhedsministeriet*. Tilgængeligt på internettet: http://www.im.dk//imagesupload/dokument/Kommunernes%20tilbagemelding_180105.pdf

Skovgaard, Ib; Henrik Stryhn og Marts Rudemo (1999): *Basal Biostatistik, del 1*. DSR Boghandel.

Tukey, John (1977): *Exploratory Data Analysis*, Addison-Wesley.