ISBN 9788770668279

Projekt 2.9 Sumkurver som funktionsudtryk – anvendt til Lorenzkurver og Ginikoefficienter (især for B- og A-niveau)

En sumkurve fremkommer ifølge definitionen, ved at vi forbinder en række *punkter* afsat i et koordinatsystem med *rette linjestykker*. Punkternes 1. koordinater er intervalendepunkterne, fra det første venstre intervalendepunkt til det sidste højre intervalendepunkt. Da ingen af de rette linjestykker derfor kan være lodrette, vil sumkurven være graf for en funktion: Til hvert tal *x* i definitionsmængden svarer præcis én y-værdi.
Lad os for et givet datamateriale kalde denne funktion for $s(x)$. I det gennemgående eksempel i grundbogens afsnit 4 er definitionsmængden for $s(x)$ lig med intervallet $[0;90]$. Af tabellen over de kumulerede procenter kan vi se funktionsværdierne hørende til intervalendepunkterne, fx:

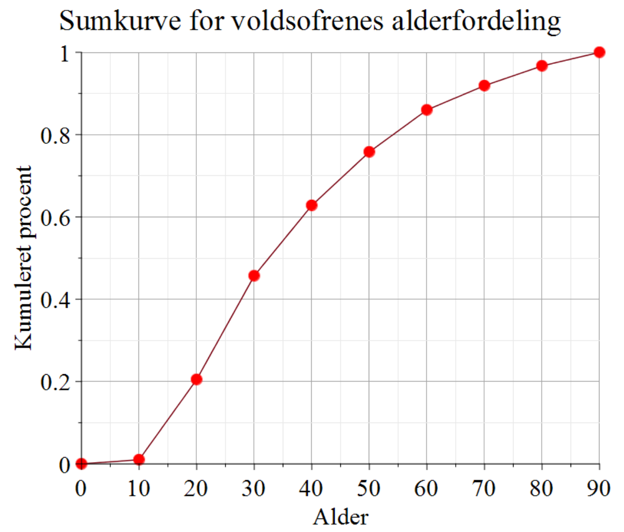
$s(0)=0$, $s(30)=45,7$ og $s(90)=100$

Vi kan ved grafisk aflæsning bestemme andre funktionsværdier, fx $s(35)$ eller $s(62)$.

Vi kan tilsvarende løse en ligning som
$s(x)=0.6$
ved grafisk afløsning

Øvelse 1
Anvend din egen sumkurve, eller grafen der er gengivet her, til at bestemme de to værdier og løse ligningen

Det er klart, at din aflæsning ikke alene er en tilnærmet værdi – det er stort set altid tilfældet – men at den er forbundet med en vis usikkerhed.



1 Stykkevis lineære funktioner

Ikke alle funktioner har en regneforskrift. I kapitel 1 har vi set eksempler på grafer, fx over ledighedens udvikling, hvor det er helt urealistisk at lede efter et funktionsudtryk. Men sumkurven består af linjestykker og vi har i kapitel 1 lært at bestemme forskriften for lineære funktioner bestemt ved to punkter. Så vi må kunne bestemme forskrifterne for hvert linjestykke, og derefter stykke disse sammen.
Funktionen $s(x)$ kaldes for en *stykkevis lineær funktion* (på engelsk: *piecewise linear function*). De fleste værktøjsprogrammer kan give os forskriften umiddelbart, men lad os i hånden udregne de første to stykker.
Første linjestykke er bestemt af punkterne $(0,0)$ og $(10,0.01)$.
Dette giver os, at konstantleddet (*b*-tallet) er 0.

Hældningskoefficienten for linjen udregnes som $a=\dfrac{\Delta y}{\Delta x}=\dfrac{0.01}{10}=0.001$.

Konklusion: Forskriften for første linjestykke på grafen er: $s_1(x)=0.001\cdot x$

$s_1(x)$ er imidlertid kun defineret i første interval. Dette angiver vi således:

$$s_1(x)=0.001\cdot x,\quad 0\leq x<10$$

Øvelse 2
a) Hvilke to punkter bestemmer det næste linjestykke?
b) Bestem en forskrift for den lineære funktion $s_2(x)$, hvis graf går gennem de to punkter
c) Opskriv forskriften for $s_2(x)$ på samme form som vi gjorde ovenfor med $s_1(x)$

Du kan i appendiks finde en vejledning i at få værktøjsprogrammerne til at bestemme den samlede forskrift for den stykkevis lineære funktion. Forskriften skrives således:

$$s(x)=\begin{cases} 0.001\cdot x, & 0\leq x<10 \\ 0.0195\cdot x-0.185, & 10\leq x<20 \\ 0.0252\cdot x-0.299, & 20\leq x<30 \\ 0.0171\cdot x-0.056, & 30\leq x<40 \\ 0.0130\cdot x+0.108, & 40\leq x<50 \\ 0.0102\cdot x+0.248, & 50\leq x<60 \\ 0.00590\cdot x+0.506, & 60\leq x<70 \\ 0.00480\cdot x+0.583, & 70\leq x<80 \\ 0.00330\cdot x+0.703, & 80\leq x<90 \end{cases}$$

Læg mærke til, at de to første lineære udtryk præcis er de, som vi beregnede ovenfor. Betingelserne som $10\leq x<20$ osv. skrives med lidt forskellig notation i de forskellige værktøjsprogrammer, men de betyder grundlæggende det samme: Alle tal i intervallet fra og med 10 til 20 (som ikke er med). Her har vi valgt at opskrive forskriften med skarpt ulighedstegn til højre, pga den oprindelige tabelopstilling. Men det kan variere.
Når værktøjsprogrammet har beregnet forskriften lader vi den normalt være skjult i de beregninger vi foretager og de ligninger vi løser. Bliver vi fx spurgt om, hvor gamle den ældste femtedel af de voldsramte er, så afkoder vi først, at en femtedel svarer til 20%, og da det drejer sig om de ældste 20%, er spørgsmålet:

Ved hvilket årstal når den kumulerede sumfunktion op på 80%?

Når vi herefter skal svare på dette spørgsmål, så kunne vi hente vi det aktuelle lineære udtryk ud, og løse den tilsvarende ligning – af grafen ser vi, at det er det sjette udtryk, så ligningen er:

$0.0102\cdot x+0.248=0.8$, med løsningen: $x=54{,}12$

Men i stedet anvender vi den stykkevis lineære funktion, $s(x)$, og opstiller og løser blot følgende ligning:

$s(x)=0.8$, der naturligvis har samme løsning: $x=54.12$

Svaret er således: Den ældste femtedel er 54,1 år eller ældre

Øvelse 3

a) Hvor gamle er den yngste tiendedel af de voldsramte?
b) Hvor stor en procentdel af de voldsramte ligger i aldersgruppen fra 15 til 25 år?

2 Anvendelse af sumkurver: Lorenz-kurver og Gini-koefficienter

I kapitel 2, afsnit 2 omtalte vi begreberne *absolut* og *relativ fattigdom*. Det er begreber, der kan karakterisere det enkelte individs situation i et givet samfund. Man er imidlertid ofte interesseret i at finde begreber, der kan udtrykke graden af lighed og ulighed i hele samfundet. Det skal være begreber, der kan anvendes til at sammenligne, både horisontalt (dvs med andre lande), og vertikalt (dvs med situationen tidligere år).

Den mest udbredte metode er anvendelsen af de såkaldte *Lorenz-kurver*, og det mest udbredte begreb er den såkaldte *Gini-koefficient*, der beregnes ud fra disse kurver. Dette er begreber og metoder, der trækker på en viden om sumkurver, og de er således oplagte eksempler på anvendelsen af denne del af den beskrivende statistik. Vi vil gennemgå metoden og begreberne ud fra et taleksempel – man vil let kunne generalisere til andre taleksempler.

Den grundlæggende ide er et meget enkelt tankeeksperiment: Alle landets indbyggere stilles i rækkefølge efter deres disponible indkomst. Rækken starter med landets fattigste og slutter med landets rigeste. Man grupperer dernæst rækken i 10 grupper, første gruppe er de fattigste 10%, næste gruppe er de hvis indkomst hører til de 10-20% fattigste osv op til den sidste gruppe, der er de 10% rigeste. En opdeling i tiendedele kaldes en opdeling i *deciler*. Det er den mest almindelige, men man kan også møde opdeling i femtedele (20% i hver gruppe), der kaldes opdeling i *kvintiler*, og i hundrededele (1% i hver gruppe), der kaldes *percentiler*. Andelen af befolkningen afsættes på 1. aksen, så fx 40 svarer til 40% af befolkningen.

ISBN 9788770668279

L&R Uddannelse
EGMONT

For hver af de ti grupper optæller man dernæst, hvor stor en andel de har af den samlede indkomst i landet. Disse andele kumuleres og afsættes som punkterne i en sumkurve, som vi har set i afsnit 3
Vi illustrerer nu med følgende oversigt over indkomstfordelingen i Danmark

Ækvivaleret Indkomst fordelt på deciler. 2014

Gruppe	Gruppe beskrivelse	Andel af indkomsten	Gennemsnit i gruppen	Indkomstinterval i gruppen
		pct.	1.000 kr.	
1. decilgruppe	De 10 pct. med lavest indkomst	2,80	70,4	< 125
2. decilgruppe	De 10-20 pct. med lavest indkomst	5,64	142,1	125 - 156
3. decilgruppe	De 20-30 pct. med lavest indkomst	6,62	166,8	156 - 178
4. decilgruppe	De 30-40 pct. med lavest indkomst	7,54	190,0	178 - 202
5. decilgruppe	De 40-50 pct. med lavest indkomst	8,51	214,4	202 - 227
6. decilgruppe	De 40-50 pct. med højest indkomst	9,51	239,5	227 - 253
7. decilgruppe	De 30-40 pct. med højest indkomst	10,61	267,4	253 - 283
8. decilgruppe	De 20-30 pct. med højest indkomst	11,98	301,8	283 - 323
9. decilgruppe	De 10-20 pct. med højest indkomst	14,00	352,6	323 - 391
10. decilgruppe	De 10 pct. med højest indkomst	22,79	574,0	391+
1. percentilgrp.	Den ene procent med lavest indkomst ...	-0,52	-129,8	< 19
100. percentilgrp.	Den ene procent med højest indkomst ...	5,55	1 397,5	756+

Kilde: Flere tal og oplysninger om indkomstfordeling på www.statistikbanken.dk/10518

Oversigten er hentet fra Danmarks Statistiks årlige publikation Indkomster, Indkomster 2014, s 76. *Du kan tilgå denne side, hvor fx alle tidligere års opgørelser også ligger, via dette link:* http://www.dst.dk/da/Statistik/Publikationer/VisPub?cid=015219

Tabellen angiver den faktiske indkomstfordeling af de disponible indkomster i Danmark – dvs. indkomsten efter skat og overførsler – i 2014. Som det fremgår af tabellen er der ulighed – den fattigste tiendedel har kun 2,8 % af indkomsterne, mens den rigeste tiendele har 22,8 % af indkomsterne.

Øvelse 4

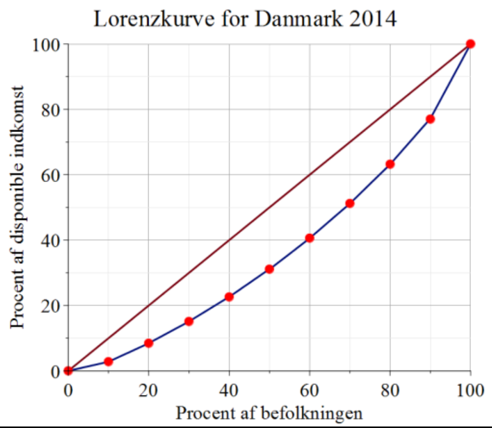
a) Opstil nu selv en tabel som i afsnit 3.3 over:
– procentandel af befolkningen
– højre intervalendepunkt
– procentandel af samlede indkomst
– kumuleret procent
b) Tegn en sumkurve over indkomstfordelingen
c) Hvordan ville en fuldstændig lige fordeling, hvor alle har samme indkomst, se ud? Opstil tabellen og tegn sumkurven for den lige fordeling i samme koordinatsystem som sumkurven over den faktiske indkomstfordeling.

ISBN 9788770668279

Resultatet af øvelse 5 skal give dig et grafisk billede som dette, hvor vi også har indtegnet punkterne.

I sådanne grafiske fremstillinger anvender vi ikke de indbyggede kommandoer i værktøjsprogrammerne, der umiddelbart kan give os en sumkurve, dels fordi vi ønsker at tegne grafen for ligefordelingen i samme billede til sammenligning, og dels fordi vi ønsker at arbejde videre med dette billede, der illustrerer graden af ulighed.

Diagonalen repræsenterer den maksimale lighed.

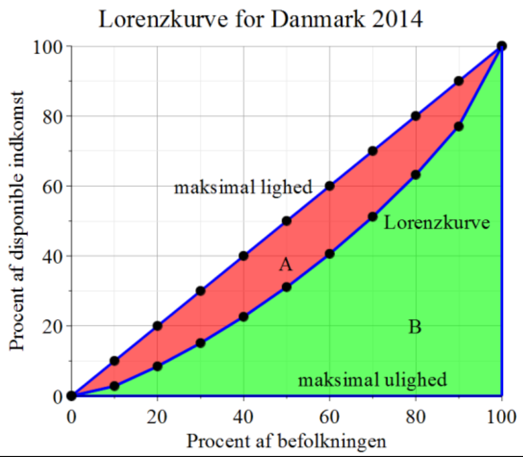


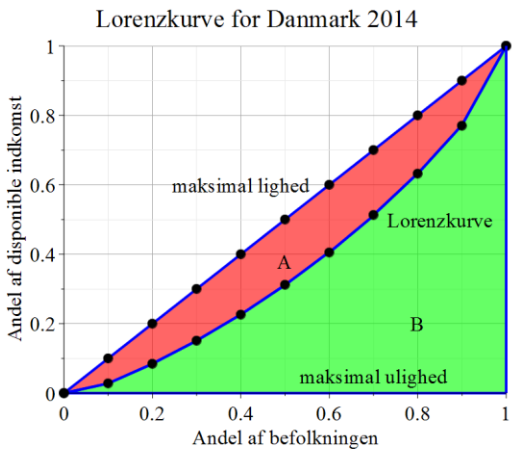
Øvelse 5
a) Beskriv hvilken indkomstfordeling, der ville svare til maksimal ulighed
b) Hvad er den grafiske repræsentation af den maksimale ulighed?

Lorenz-kurverne danner udgangspunkt for det mest anvendte matematiske begreb til beskrivelse af uligheden i et samfund, nemlig *Gini-koefficienten*. Du kan læse mere om Lorenz-kurver og Gini-koefficienten i B-bogens kapitel 14, fagligt samarbejde matematik og samfundsfag, side 50ff. Du kan få adgang til via bogens **website.**

I det grafiske billede af Lorenz-kurven har vi samtidig billedet af den maksimale lighed – nemlig diagonalen – og billedet af den maksimale ulighed – nemlig en kurve der følger 1. aksen. Størrelsen af området mellem Lorenz-kurven og diagonalen må derfor give os et mål for uligheden. De to illustrationer er identiske, blot har vi i den højre omregnet procenttal til decimaltal. På figurerne har vi markeret kurverne, der repræsenterer maksimal lighed og ulighed.

Vi har endvidere markeret de to områder A og B. A og B står for arealerne af pågældende områder.





Definition: Gini-koefficient
En Lorenz-kurve afgrænser sammen med kurverne for maksimal lighed og maksimal ulighed to områder i første kvadrant, med arealerne henholdsvis A og B, som angivet på figuren.

Gini-koefficienten er lig med tallet: $G = \dfrac{A}{A+B}$

ISBN 9788770668279

Øvelse 6: G er uafhængig af skalering

Vis, at vi får samme værdi af G uanset hvilken af de to figurer vi anvender:

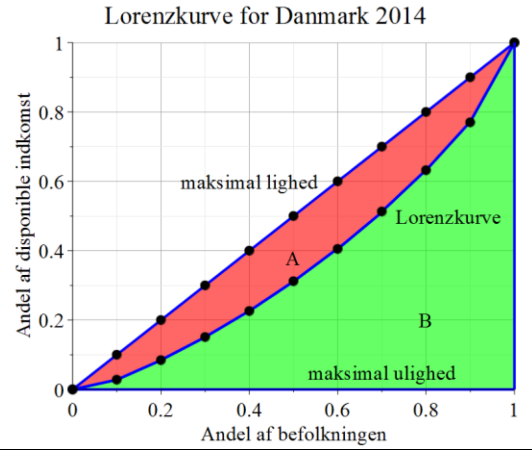
Øvelse 7: G er et mål for ulighed

Anvend figuren til højre.

a) Argumenter for, at tallet G altid ligger i intervallet $[0;1]$

b) Argumenter for, at $A+B=\frac{1}{2}$, og vis, at $G=2\cdot A$

c) Argumenter ud fra denne formel for, at en lille værdi af G svarer til, at Lorenzkurven ligger tæt på kurven for maksimal lighed, og en stor værdi af G svarer til, at Lorenz-kurven ligger langt væk fra kurven for maksimal lighed.



Øvelse 8: Beregning af Gini-koefficienten

a) Vis at $G=1-2B$ (*hint: udnyt, at* $2A+2B=1$)

b) Lorenz-kurven består af rette linjestykker. Argumenter for at området B kan opdeles i en række trapezer.

c) Find formlen for arealet af et trapez, og argumenter for denne formel

d) Anvend trapezformlen til at bestemme arealet af området B. (Du må gerne tage dit værktøjsprogram i brug)

e) Anvend endelig formlen i a) til at bestemme G. Du skal få et tal i nærheden af 0,283 .

f) Undersøg på nettet hvad Gini-koefficienten er for et antal udvalgte lande.

g) Undersøg via adressen på Danmarks Statistik, hvordan Ginikoefficienten har udviklet sig i Danmark.