matsamf

thomas schausen & morten damsgaard-madsen

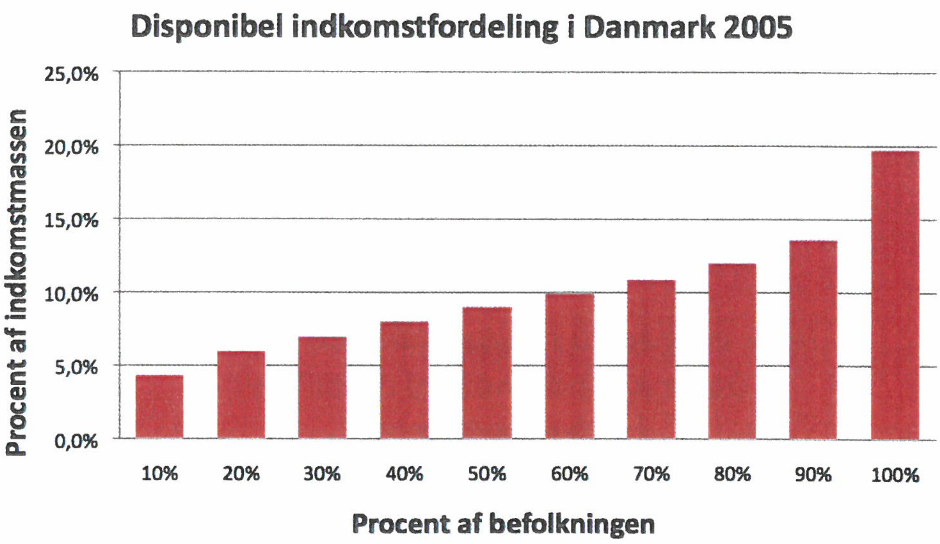
SYSTIME

Ginikoefficienten

Vi vil nu indføre et nyt håndgribeligt kvantitativt mål for ulighed, nemlig Ginikoefficienten, der er udviklet af den italienske samfundsforsker Corrado Gini (1884 – 1965).

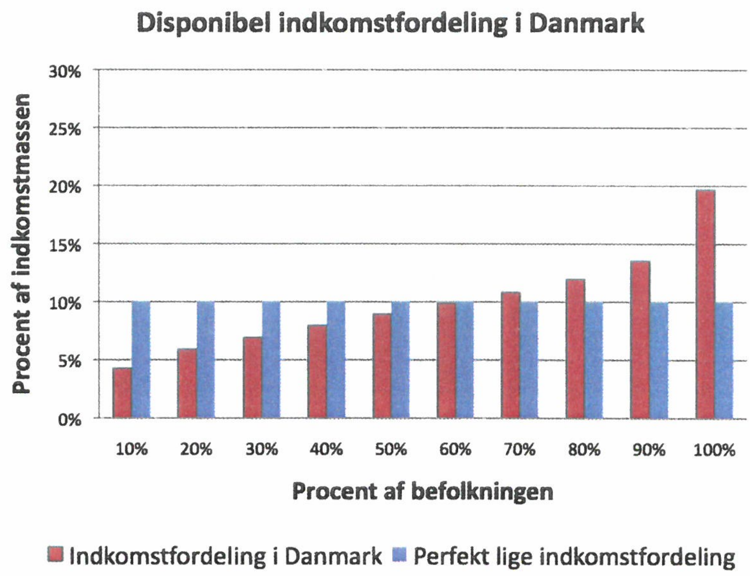
Deciler

For at redegøre for Ginikoefficienten er det nødvendigt at indføre nogle begreber, hvor det første er *deciler*. Deciler er beslægtet med kvartiler, hvor en population – eksempelvis en befolkning – inddeles i 4 lige store dele, mens en median inddeler en befolkning i to lige store dele. Tilsvarende inddeler deciler en population i ti lige store dele. Vi er interesserede i at inddele befolkningen i ti lige store dele, *ordnet* således at den første tiendedel er de fattigste, den næste tiendedel er de næstmest fattige og så videre indtil den sidste tiendedel, som er de rigeste. Til at illustrere dette vises et eksempel.



Figur 2.5. *Histogram, hvor den danske befolkning er inddelt i ti lige store grupper ordnet i forhold til indkomstfordelingen på førsteaksen. Søjlens højde viser hver gruppes samlede indkomst. Hvis befolkningens indkomst var ligeligt fordelt, ville søjlerne være lige høje. Kilde OECD.*

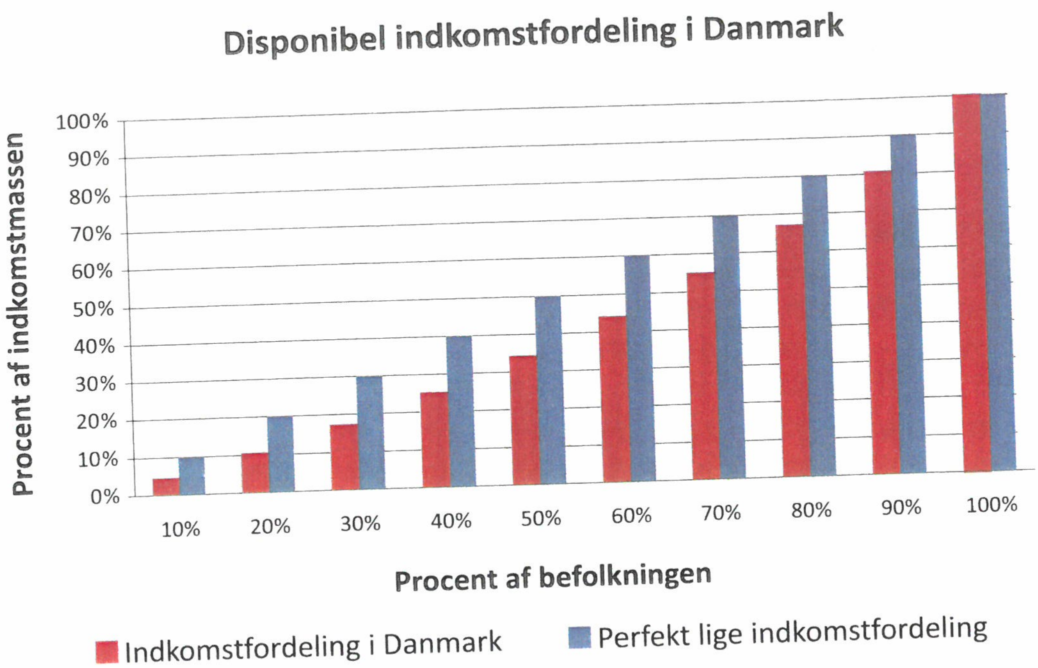
Figuren viser tydeligt, at den rigeste tiendedel har en langt større disponibel indkomst end den fattigste tiendedel, men for at vurdere, hvordan det ville have set ud, hvis de ti deciler havde den samme indkomst, vises et nyt histogram.



Figur 2.6. *De røde søjler er de samme som i figur 2.5, mens de blå repræsenterer en indkomstfordeling, hvor alle deciler har den samme indkomst, hvilket svarer til at alle personer har den samme indkomst. Bemærk, at der med "Perfekt lige indkomstfordeling" menes, at indkomsten er fuldstændig jævnt fordelt og udtrykker således ikke en holdning til indkomstfordelingen.*

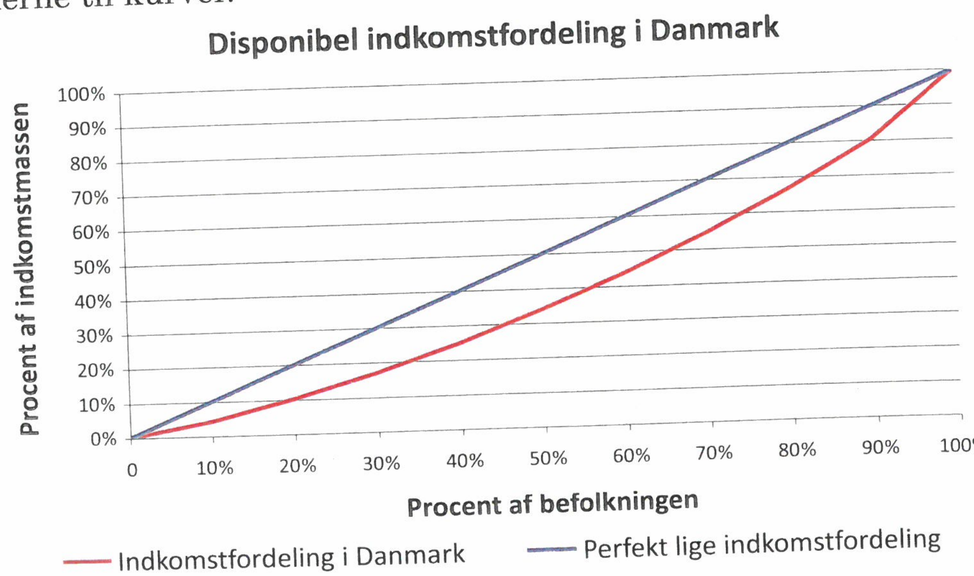
Lorenzdiagram

Det næste begreb, der er nødvendigt for at forklare Ginikoefficienten, er et *Lorenzdiagram*. Dette er en videreudvikling af decilerne præsenteret i histogrammerne ovenfor. Første skridt er at "stable" de røde søjler. Den første røde søjle i figur 2.6 går igen i figur 2.7, mens den anden søjle ved 20 % af befolkningen i figur 2.7 er summen af de to første søjler i figur 2.6. Med andre ord "stables" de to første røde søjler ovenpå hinanden, hvorved søjle to fås i figur 2.7.



Figur 2.7. *Samme data som i figur 2.6. Her er søjlerne blot "stablet", således at den fattigste tiendedels samlede indkomst er den første røde søjle, mens den næste røde søjle er summen af den fattigste og næstfattigste tiendedel og så videre. De blå søjler viser igen den perfekte lige indkomstfordeling. Forskellen på figur 2.6 og figur 2.7 svarer til forskellen mellem frekvens og kumuleret frekvens i deskriptiv statistik.*

Sidste skridt i vejen mod at lave et Lorenzdiagram er at ændre søjlerne til kurver.



Figur 2.8. *Et Lorenzdiagram med de samme data fra Danmark som i figur 2.7. Her er søjlerne erstattet af linjer.*

I et samfund, hvor indkomsten er perfekt lige, vil indkomstfordelingen være som den blå kurve i figur 2.8, mens den røde kurve viser den reelle indkomstfordeling i Danmark 2005. Hvis vi vurderer arealerne under de to kurver, har vi et meget konkret grundlag til at vurdere uligheden i et samfund.

Ginikoefficienten er defineret som arealet imellem de to kurver som andel af arealet under den blå kurve. Disse arealer vil vi nu beregne og som kronen på værket beregne Ginikoefficienten. Dette vil vi gøre på to måder.

Beregning af Ginikoefficient

Vi skal nu beregne arealet mellem de to kurver i figur 2.8. Dette gøres ved at beregne arealerne under den blå kurve og den røde kurve hver for sig.

Den blå kurve er en ret linje med en hældning på 1 samt skæring med y-aksen i (0, 0). Derfor har den regneforskriften:

$$P(x) = 1 \cdot x + 0 = x$$

P repræsenterer den procentvise fordeling af indkomstmassen, såfremt denne var perfekt ligeligt fordelt. x repræsenterer den procentvise del af befolkningen, hvor x løber fra 0 til 1, svarende til fra ingen del af befolkningen (0) til hele befolkningen (1).

Vi har da, at arealet A_p under P i [0, 1] er

$$A_p = \int_0^1 x dx = \left[\frac{1}{2}x^2\right]_0^1 = \frac{1}{2}1^2 - \frac{1}{2}0^2 = \frac{1}{2}$$

Det beregnede areal er af en retvinklet trekant, der har kateterne 1. Derfor kan arealet også beregnes meget let med formlen

$$T = \frac{1}{2} \cdot h \cdot g,$$

hvor både højden h og grundlinjen g er 1. Vi har da, at

$$A_p = \frac{1}{2} \cdot 1 \cdot 1 = \frac{1}{2},$$

hvilket jo stemmer overens med ovenstående.

Ginikoefficienten ved hjælp af stykvis lineære funktioner

Ved at iagttage den røde kurve i figur 2.8 ses det, at den "knækker" ved 10 %, 20 % og så videre, idet hældningen varieres. Derfor består den røde kurve af 10 stykvis lineære funktioner. Metoden er at bestemme regneforskrifterne i hvert af de ti intervaller for den røde kurve, beregne de ti arealer og endelig lægge dem sammen.

Datamaterialet fra OECD

Gennemsnitlig indkomst for decilgrupper i 2005	
1. decil	10.408
2. decil	14.398
3. decil	16.832
4. decil	19.380
5. decil	21.787
6. decil	24.046
7. decil	26.345
8. decil	29.090
9. decil	32.965
10. decil	47.864

Figur 2.9. *Data fra OECD, der ligger til grund for de foregående figurer. Data er disponible indkomster i US $ opgjort efter købekraftparitetsprincippet (PPP).*

Tallene tastes ind i et regneark.

	A	B	C	D	E	F	G	H	I	J	K	L
1		Gennemsnitlig indkomst for decilgrupper i 1000 kr.	Frekvens	Kumuleret frekvens = søjlens højde i figur 2.4	Interval nummer	y-værdi ved interval-start	y-værdi ved interval-stutning	x-værdi ved interval-start	x-værdi ved interval-stutning	$a=\frac{y_2-y_1}{x_2-x_1}$	b	Areal i interval under den røde kurve
2				0		y_1	y_2	x_1	x_2			
3	1. decil	10.408	4,3%	4,3%	1.	0,0%	4,3%	0%	10%	0,428	-	0,002
4	2. decil	14.398	5,9%	10,2%	2.	4,3%	10,2%	10%	20%	0,592	-0,016	0,007
5	3. decil	16.832	6,9%	17,1%	3.	10,2%	17,1%	20%	30%	0,692	-0,036	0,014
6	4. decil	19.380	8,0%	25,1%	4.	17,1%	25,1%	30%	40%	0,797	-0,068	0,021
7	5. decil	21.787	9,0%	34,1%	5.	25,1%	34,1%	40%	50%	0,896	-0,107	0,030
8	6. decil	24.046	9,9%	44,0%	6.	34,1%	44,0%	50%	60%	0,989	-0,154	0,039
9	7. decil	26.345	10,8%	54,8%	7.	44,0%	54,8%	60%	70%	1,084	-0,211	0,049
10	8. decil	29.090	12,0%	66,8%	8.	54,8%	66,8%	70%	80%	1,197	-0,290	0,061
11	9. decil	32.965	13,6%	80,3%	9.	66,8%	80,3%	80%	90%	1,356	-0,417	0,074
12	10. decil	47.864	19,7%	100,0%	10.	80,3%	100,0%	90%	100%	1,969	-0,969	0,090
13	**Total**	**243.114,5**										
14										**Samlet areal**		**0,387**

Figur 2.10. *Udregning af regneforskrifter for de stykvise lineære funktioner i de 10 intervaller samt beregning af arealer under kurven i hvert af intervallerne. Endelig summeres arealerne til 0,387. Farvelægningen henviser til udregningerne af a og b nedenfor.*

Indkomstfordeling i kolonne B summeres i celle B13. Herefter omregnes indkomstfordelingen til frekvens efter samme metode som, når hyppighed omregnes til frekvens i deskriptiv statistik, nemlig ved at dividere indkomsten for decil med den samlede indkomst i celle B13. Frekvensværdierne i kolonne C lægges sammen til kumuleret frekvens i kolonne D. Herved har vi begyndelses- og slutpunkt for hvert interval.

For hvert interval har vi begyndelsespunktets x-værdi i kolonne H og y-værdien i kolonne F. Endvidere har vi intervallets slutpunkt givet ved, at x-værdien er i kolonne I og y-værdien i kolonne G.

Når disse punkter kendes, kan hældningskoefficienten a for regneforskriften i hvert interval beregnes med sætningen angivet i celle J1. Eksempelvis beregnes hældningen a i første interval:

$$a=\frac{y_2-y_1}{x_2-x_1} \Rightarrow a=\frac{0{,}043-0}{0{,}1-0}=0{,}428$$

hvilket netop er værdien i celle J3.

Konstanten b beregnes ved at isolere b i ligningen $y_1 = a \cdot x_1 + b$ $\Leftrightarrow b = y_1 - a \cdot x_1$. Vi ser på udregningen af b i det andet interval – celle K4

$$b = y_1 - a \cdot x_1 \Rightarrow b = 0,043 - 0,592 \cdot 0,1 = -0,016$$

Efter at have udregnet a og b har vi regneforskrifterne. Næste opgave er at udregne arealerne i hvert interval ved hjælp af integralregning. Til dette indfører vi symboler for arealerne.

A_1: Arealet i det første interval, hvor x løber fra 0% til 10% under "Den røde kurve" i figur 2.8. Tilsvarende hedder de følgende arealer $A_2, A_3, ..., A_{10}$

Stamfunktionerne til $a \cdot x + b$ bestemmes

$$\int (a \cdot x + b)dx = \frac{1}{2} \cdot a \cdot x^2 + b \cdot x + k,$$

når vi skal beregne arealet er det bestemte integral derfor:

$$A_1 = \int_0^{0,1} (a \cdot x + b)dx = \left[\frac{1}{2} \cdot a \cdot x^2 + b \cdot x\right]_0^{0,1} = \frac{1}{2} \cdot a \cdot 0,1^2 + b \cdot 0,1 - \left(\frac{1}{2} \cdot a \cdot 0^2 + b \cdot 0\right)$$

I kolonne L er de ti arealer udregnet og lagt sammen til 0,387 i celle L14.

Nu kan vi beregne Ginikoefficienten ved hjælp af definitionen:

$$\text{Ginikoefficient} = \frac{\text{Perfekt lige fordeling} - \text{Aktuel fordeling}}{\text{Perfekt lige fordeling}}, \tag{1}$$

hvor begge fordelinger henviser til de beregnede arealer, hvor

$$\text{Perfekt lige fordeling} = 0,5 \text{ og } \text{Aktuel fordeling} = 0,387$$

Ved at bruge (1) fås $\text{Ginikoefficient} = \frac{0,5 - 0,387}{0,5} = 0,227$

Herved har vi beregnet Ginikoefficienten til 0,227.

OPGAVE 2.1

Beregn Ginikoefficienten for par i Danmark i 2008 ud fra figur 2.11.

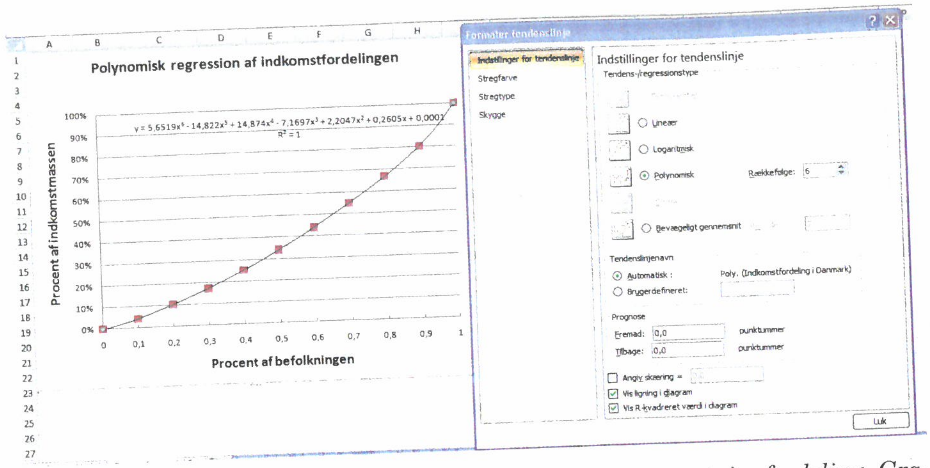
Procentvis fordeling af indkomstmassen. 2008

	Personer i alt	Familier		
		Enlige	Par	Alle
Indkomst i alt	pct.			
1. decil	0,5	1,8	2,8	1,6
2. decil	3,8	5,1	4,8	3,7
3. decil	5,3	6,2	6,2	4,5
4. decil	6,5	6,9	7,6	5,7
5. decil	8,1	7,8	8,7	7,0
6. decil	9,6	9,2	9,7	8,6
7. decil	11,1	10,8	10,7	10,9
8. decil	12,7	12,6	12,0	13,6
9. decil	15,0	14,9	13,9	16,6
10 decil	27,5	24,7	23,7	27,9

Figur 2.11. *Fordeling af indkomstmassen før fradrag af skat for personer fordelt på enlige og par. Kilde: Danmarks Statistik. Bemærk, at de angivne værdier svarer til "frekvens" i figur 2.10.*

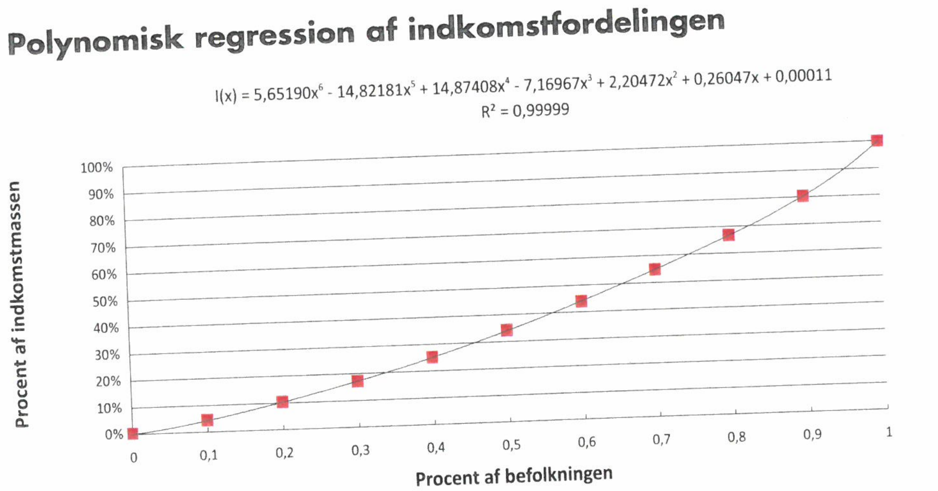
Ginikoefficienten ved hjælp af regression

Vi har nu set, hvordan vi kan udregne Ginikoefficienten eksakt, men det krævede en del arbejde. Det er mindre arbejdskrævende at udregne Ginikoefficienten ved hjælp af regression i et regneark. Her vil vi bruge en polynomisk regression.



Figur 2.12. *Polynomisk regression af indkomstmassens procentvise fordeling. Graden af polynomiet bestemmes i rullemenuen "Rækkefølge" til 6, der er den maksimale grad, Excel kan regne med.*

Ud af regression i figur 2.12 fås en regneforskrift for den aktuelle fordeling af indkomstmassen:



Figur 2.13. *Polynomisk regression af indkomstfordeling med et 6'te gradspolynomium. Læg mærke til, at det er en meget pæn regression med en R^2 meget tæt på 1.*

Vi har nu regneforskriften $I(x)$, hvor I repræsenterer den procentvise fordeling af indkomstmassen , som den var ifølge OECD i 2005, og x stadig repræsenterer procent af befolkningen, der også her løber fra 0 til 1. For at få et korrekt resultat er det vigtigt at få tilpas mange decimaler med fra regnearket. Ved at gå ind i "formater tendenslinje etiketten" kan dette antal tilpasses som eksempelvis ovenfor, hvor det er sat til 5 decimaler. Nu kan vi ved hjælp af integralregning beregne arealet under den blå kurve.

Vi indfører symbolerne:
A_I : Arealet under kurven repræsenterende indkomstfordelingen i Danmark.

A_P : Arealet under kurven repræsenterende den perfekt lige indkomstfordeling.

$$A_I = \int_0^1 (5{,}65190x^6 - 14{,}82181x^5 + 14{,}87408x^4 - 7{,}16967x^3 + 2{,}20472x^2 + 0{,}26047x + 0{,}00011)dx$$
$$= 0{,}385$$

Nu beregnes Ginikoefficienten

$$\text{Gini} = \frac{A_P - A_I}{A_P} \Rightarrow \text{Gini} = \frac{0{,}5000 - 0.385}{0{,}5000} = 0{,}230$$

Spørgsmålet er, hvor meget vores forsimplede metode med regression tager fejl i forhold til den eksakte metode. Dette spørgsmål kan besvares ved at udregne den relative difference.

$$\frac{\text{Gini}_{\text{Eksakt}} - \text{Gini}_{\text{Regression}}}{\text{Gini}_{\text{Eksakt}}} \cdot 100\% = \frac{0{,}230 - 0{,}227}{0{,}227} \cdot 100\% = 1{,}6\%$$

Differencen er rimelig lille, hvilket kan retfærdiggøre brug af regression. I øvrigt vil polynomiet altid ligge under den stykvise lineære funktion, hvorved Ginikoefficienten bliver en lille smule større, når den beregnes ved regression end ved den præcise måde med stykvis lineære funktioner.

OPGAVE 2.2

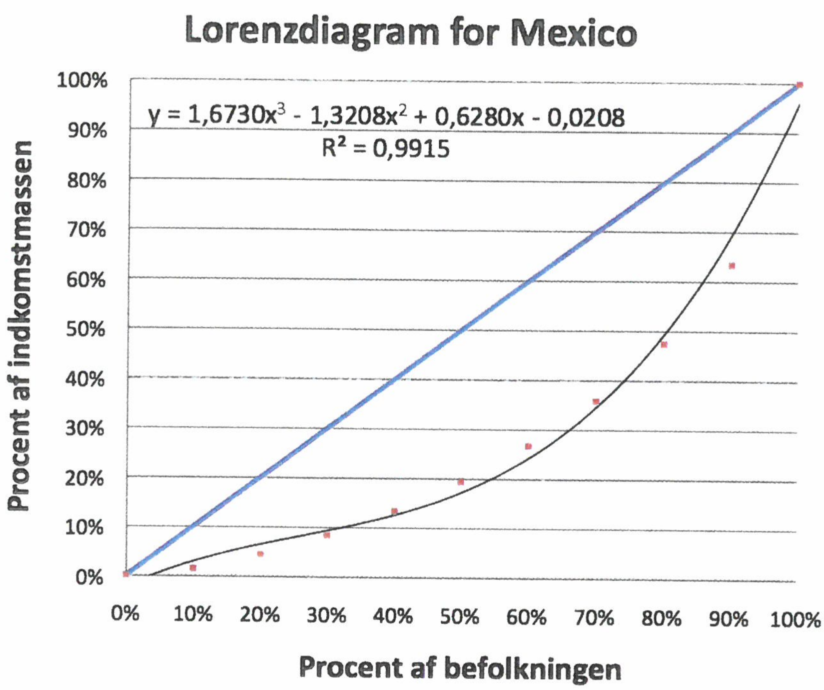
Beregn Ginikoefficienten for par ved hjælp af regression. Brug data fra figur 2.11.

Ginikoefficienten som mål for ulighed

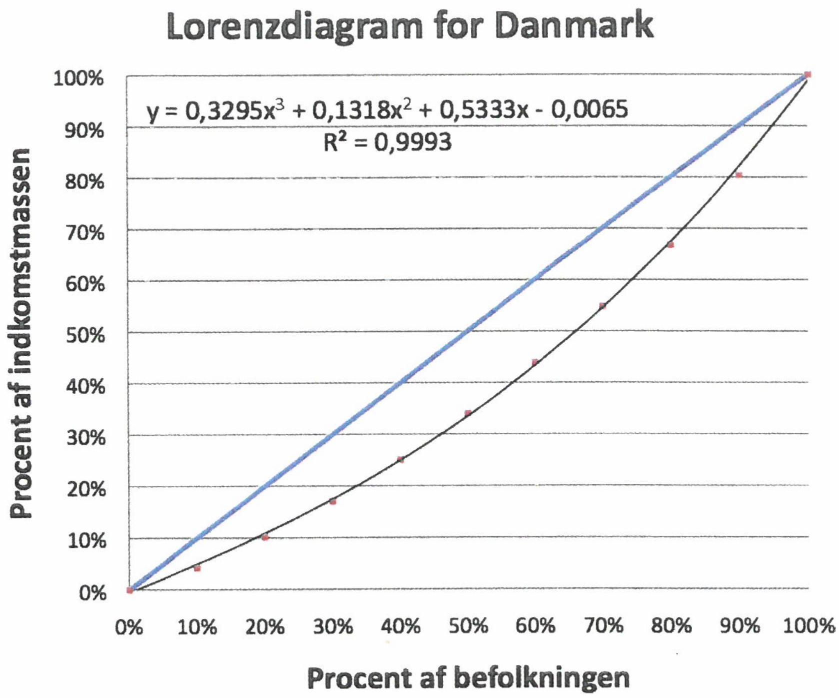
Efter at have set på, hvordan Ginikoefficienten udregnes, vil vi nu se nærmere på, hvordan den kan anvendes til at beskrive og vurdere ulighed.

Som beskrevet ovenfor er Ginikoefficienten defineret som *arealet imellem de to kurver som andel af arealet under den blå kurve* (se figur 2.8 ovenfor). Det vil altså sige, at Ginikoefficienten er større, jo større afstand der er mellem den vandrette linje og den kurve, der viser den virkelige indkomstfordeling.

For at vise forskellen kan vi se på to lande, nemlig Mexico og Danmark.



Figur 2.14. *Lorenzdiagram og Ginikoefficient for Mexico.*



Figur 2.15. *Lorenzdiagram og Ginikoefficient for Danmark.*

Som man kan se på figurerne, er arealet mellem den blå og den røde kurve meget større for Mexico end for Danmark. Vi har også, at Ginikoefficienten for de to lande er hhv. 0,47 og 0,23. Vi kan altså konstatere, at uligheden i Mexico er meget større end den er i Danmark.

For at forstå årsagerne til den store forskel i Ginikoefficienten mellem Danmark og Mexico er det nødvendigt at studere de to lande nærmere, specielt når det gælder opbygningen af skattesystemet og organiseringen af velfærdsydelserne. I denne bog vil vi ikke gå nærmere ind i denne analyse, men se på nogle generelle problematikker omkring anvendelsen af Ginikoefficienten som mål for ulighed.

Fordelen ved at bruge Ginikoefficienten er, at den angiver ulighed som en andel af landets indkomst. Den er derfor uafhængig af, hvor stort et land der beskrives, og hvor stor landets samlede indkomst er. På denne måde er det muligt at sammenligne ulighed i mange forskellige lande. Det betyder dog, at man skal supplere Ginikoefficienten med andre mål for at få et mere præcist billede af fattigdom og ulighed. Ofte bruger man BNP og BNP pr. indbygger som supplerende mål, da de angiver landets rigdom og den gennemsnitlige rigdom i landet. Men ingen af disse mål siger noget om fordelingen. Selv om et land har et meget stort BNP pr. indbygger kan uligheden godt være stor, da der er tale om et simpelt gennemsnit og ikke en opdeling i deciler.

Hvis vi ser på vores diskussion i starten af kapitlet, skal vi også være opmærksomme på, at Ginikoefficienten normalt måler indkomstfordelingen og ikke formuefordelingen (man kan dog også udregne en Ginikoefficient for denne). Heller ikke den uformelle økonomi er indregnet.

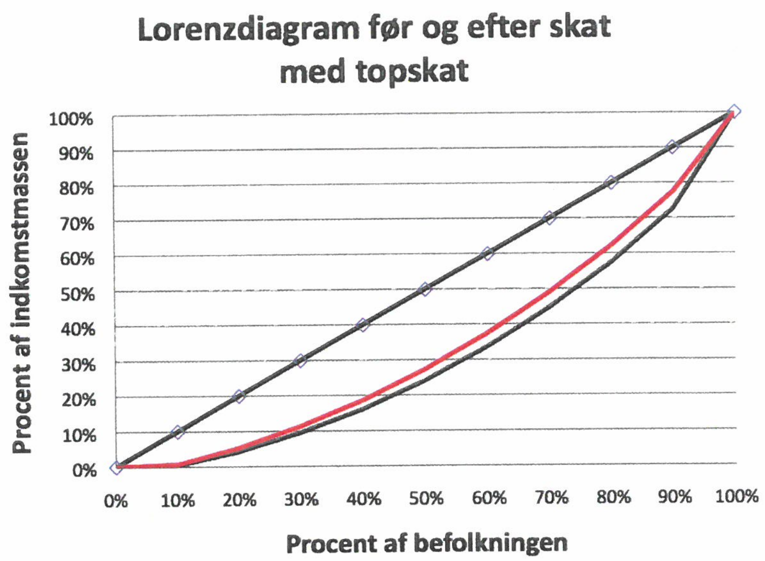
En sidste problemstilling, som vi ikke har berørt, er, at man måler ulighed på et givet tidspunkt og ikke ulighed i livsindkomst. I alle samfund vil det være normalt, at indkomsten varierer med alderen, og der vil derfor altid være en forskel alene af denne grund.

OPGAVE 2.3

Beregn Ginikoefficienten i Mexico ved hjælp af den polynomiske regression, der er angivet i figur 2.14.

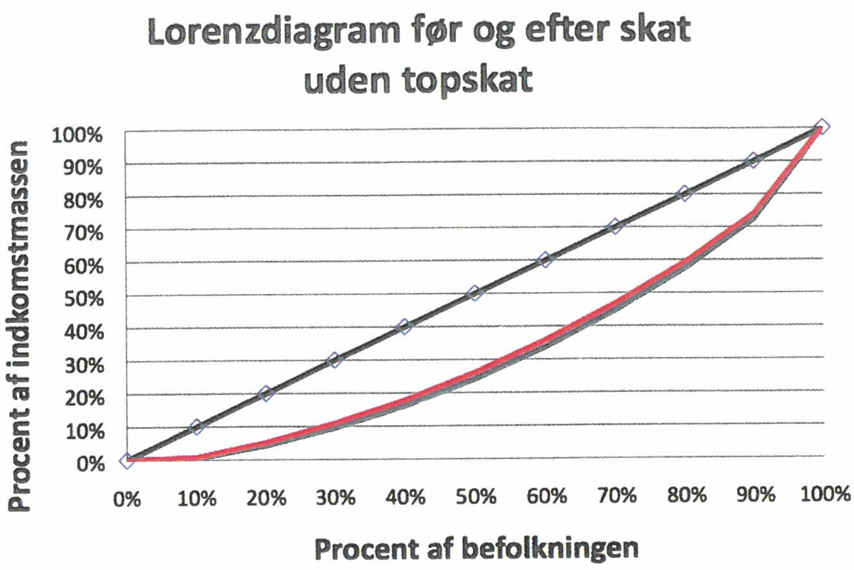
Skat og ulighed

I starten af kapitlet havde vi også en diskussion af skattens betydning. Denne diskussion vil vi nu se nærmere på ved hjælp af et par eksempler. I den nedenstående figur ses virkningen af skatten i et Lorenzdiagram:



Figur 2.16. *Den blå kurve er fordelingen af indkomstmassen før skat, mens den røde kurve er fordeling af indkomstmassen efter skat. Indkomstmassen gælder for* ***Personer*** *i 2008 (se figur 2.11). Kilde: Danmarks Statistik. Den beregnede skat følger omtrent det danske system. Ginikoefficienten er 0,374 før skat og 0,327 efter skat.*

I Lorenzdiagrammet kan vi direkte se skattens *omfordelende effekt*, dvs. at skatten gør uligheden mindre i samfundet. Det ses ved, at den røde kurve ligger tættere på den sorte kurve, hvorved arealet imellem de to kurver bliver mindre. Det betyder også, at ændringer i skattesystemet har betydning for uligheden, som vi kan se i den følgende figur:



Figur 2.17. *Samme grundlag som figur 2.16 bortset fra, at topskatten er afskaffet. Det ses, at den røde kurve nærmer sig den blå kurve en anelse i forhold til den foregående figur. Herved bliver arealet mellem den røde og sorte kurve større, og Ginikoefficienten stiger fra 0,327 til 0,347.*

Ulighed i international sammenhæng

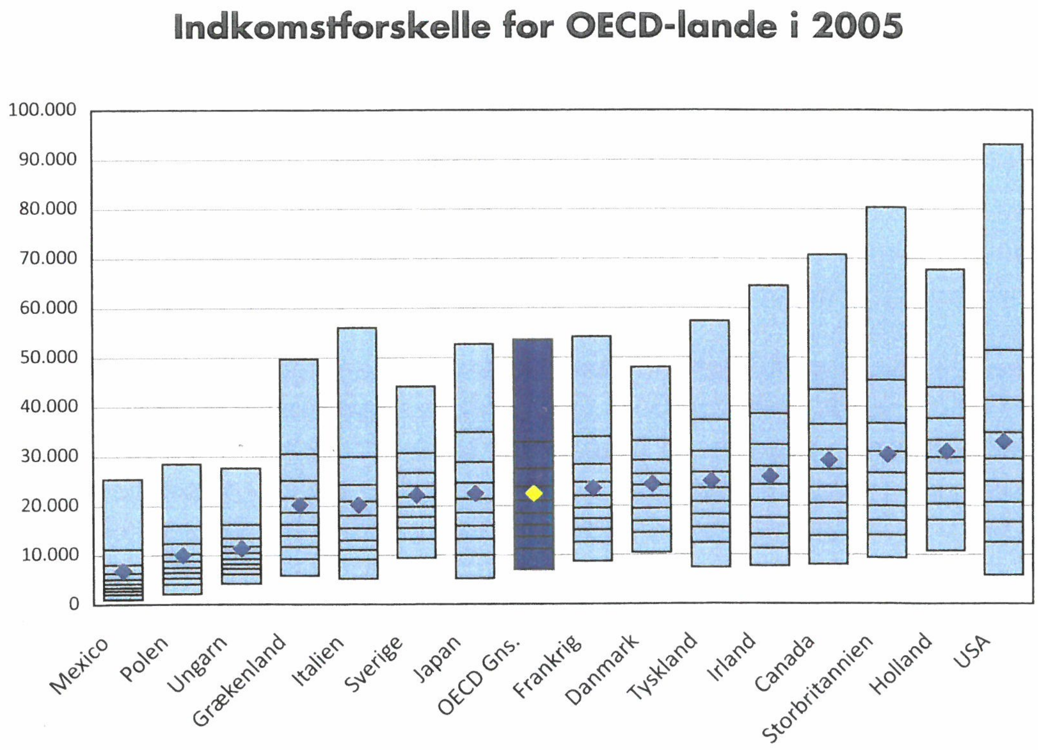
Lad os som afslutning se lidt på, hvordan uligheden ser ud, når man sammenligner en række forskellige lande. Vi så ovenfor, at der er stor forskel på uligheden i hhv. Mexico og Danmark, men hvordan er det samlede billede?

Når man opgør indkomstforskelle i forskellige lande er der to grundlæggende problemer.

For det første er der forskel i valutakurserne. Det betyder, at der kan være meget store forskelle på, hvor meget indkomsten i et land reelt er værd i forhold til andre lande. Som eksempel kan vi se på den danske krone. Danmark er et lille land, som importerer en stor del af produktionen (læs mere om dette i kapitel 4 om *importkvoten*). Hvis kronekursen er lav, betyder det, at de importvarer som forbrugerne køber, vil være dyrere end de er ved en høj kronekurs. Indkomsten for en dansker vil derfor være mindre værd i denne situation.

For det andet er der forskel i inflation og priser på varerne. Danmark er et af de lande i verden, der har de højeste priser på forbrugsgoder – f.eks. var prisen på Apples iPad 2 ved introduktionen i marts 2011 den højeste i de 26 lande, hvor produktet blev lanceret.

For at tage højde for disse forskelle anvender man i internationale sammenligninger ofte begrebet købekraftsparitet. Det hedder på engelsk *Purchasing Power Parity*, og man ser derfor ofte betegnelsen *PPP* anvendt. Kort fortalt går det ud på, at man sammenligner prisen på en udvalgt række af forbrugsgoder og justerer for inflation og valutakurs. Denne måde at beregne på giver derfor et bedre billede af indkomstforskelle og ulighed mellem lande. I den nedenstående figur kan man se data fra 2005 fra en række lande.

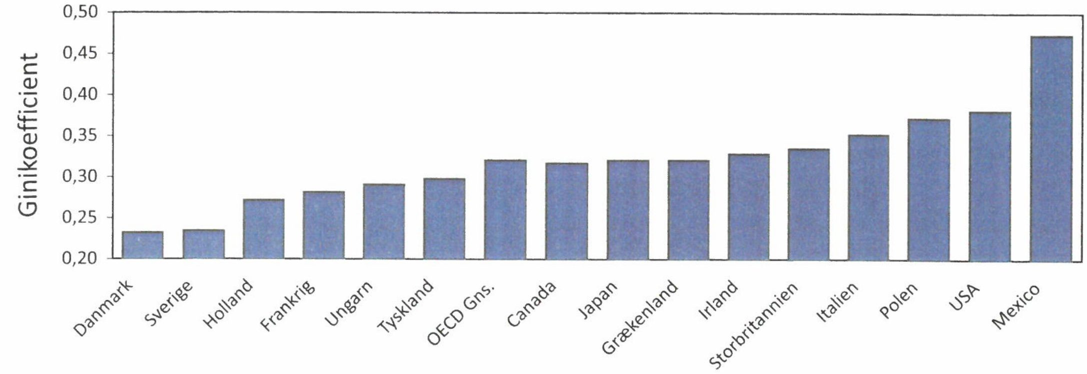


Figur 2.18 *PPP-justerede indkomstforskelle i OECD-lande 2005 i $. Indkomsten er angivet i deciler, og diamanten angiver gennemsnitsindkomsten. Kilde: OECD 2008.*

Denne figur giver os et billede af indkomstforskellene på et givet tidspunkt, men vi er også interesserede i at vide noget om, hvordan indkomstforskellene har udviklet sig over tid. Eller sagt med andre ord: Er uligheden i verden blevet større eller mindre?

Her er Ginikoefficienten en stor hjælp. På trods af de metodiske indvendinger, man kan have, er den meget velegnet til at se på store datamængder, da den udtrykker uligheden ved hjælp af et enkelt tal. Lad os først se på indkomstforskellene i midten af 2005 udtrykt ved Ginikoefficienten:

Forskelle i Ginikoefficienter for OECD-lande i 2005

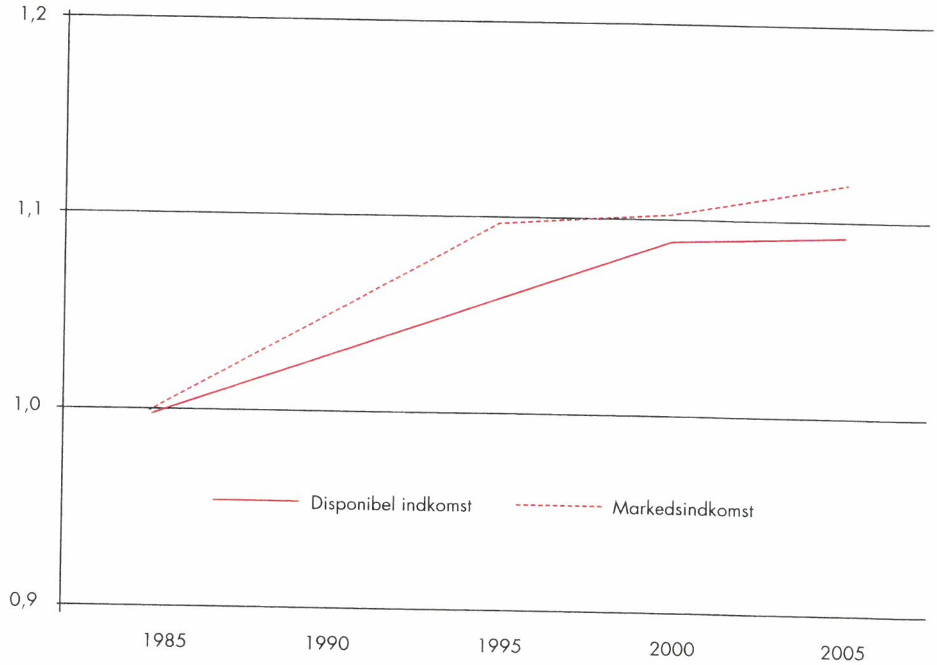


Figur 2.19. *Ginikoefficienter for OECD-lande i midten af 2005. Kilde: OECD 2008.*

Som det ses, ligger Danmark nederst i denne opgørelse med en Ginikoefficient på omkring 0,23. Vi er altså sammen med Sverige det mest lige land i verden. Vi bemærker også, at 0,23 er det tal, som vi udregnede tidligere i kapitlet på baggrund af OECD's opgørelser.

Den næste figur viser udviklingen i ulighed. Som det ses, er uligheden stigende blandt landene i denne opgørelse, både målt som indkomst før og efter skat.

Gennemsnitlig udvikling i Ginikoefficient for 15 OECD-lande



Figur 2.20. *Udviklingen af ulighed siden 1980. Indeks: Midten af 1980'erne=1,0. Baseret på data fra de 15 OECD-lande, som man har data for siden midten af 1980'erne (Canada, Danmark, Finland, Frankrig, Tyskland, Grækenland, Italien, Japan, Luxembourg, Holland, New Zealand, Norge, Sverige, England og USA). Kilde: OECD 2008.*

Opsummering

Til sidst kan vi kort opsummere, hvad vi har gjort.

- Vi har set på, hvordan man forstår begrebet ulighed ud fra forskellige ideologiske opfattelser.
- Vi har diskuteret, hvordan ulighed kan måles.
- Vi har introduceret Lorenzdiagrammet.
- Med udgangspunkt i Lorenzdiagrammet har vi introduceret Ginikoefficienten som et konkret mål for ulighed.
- Vi har vist to forskellige måder at udregne Ginikoefficienten på og diskuteret fordele og ulemper ved disse tilgange.
- Vi har diskuteret den praktiske anvendelse af Ginikoefficienten.