

Matematik og samfundsfag – Gini-koefficienten

Den såkaldte *Gini-koefficient*, introduceret i 1912 i en artikel af den italienske statistiker, demograf og sociolog *Corrado Gini*, er et mål for indkomstuligheden i et samfund. For at beregne gini-koefficienten, skal man først have fat i den såkaldte *Lorenz-kurve*, opkaldt efter den amerikanske økonom *Max Otto Lorenz*.

Man inddeler befolkningen i ti grupper med lige mange personer i hver. Disse kaldes *decilgrupper*. Den første gruppe indeholder de 10% af befolkningen, som har den laveste indkomst. Den anden decilgruppe indeholder de 10% af befolkningen, som har de laveste indkomster bortset fra decilgruppe 1. Tredje decilgruppe indeholder de 10% af befolkningen, som har den laveste indkomst, bortset fra decilgruppe 1 og 2, etc. indtil decilgruppe 10, som indeholder de 10% af befolkningen, som har den højeste indkomst. Som population kan man for eksempel bruge hele befolkningen i et land, befolkningen i en kommune og lignende. Man kan i princippet også undersøge personerne i en organisation, med henblik på at undersøge uligheden i indkomst heri. Vi skal i det kommende gennemregne et eksempel for at forstå processen med at fremstille en Lorenz-kurve og beregne gini-koefficienten. Det skal også lige nævnes, at man kan diskutere, hvad der skal medregnes i indkomsten. Et godt bud er den såkaldte *disponible indkomst*, hvor forskellige skatter er fraregnet. Se linket [1] sidst i dokumentet for en definition heraf.

Eksempel 1

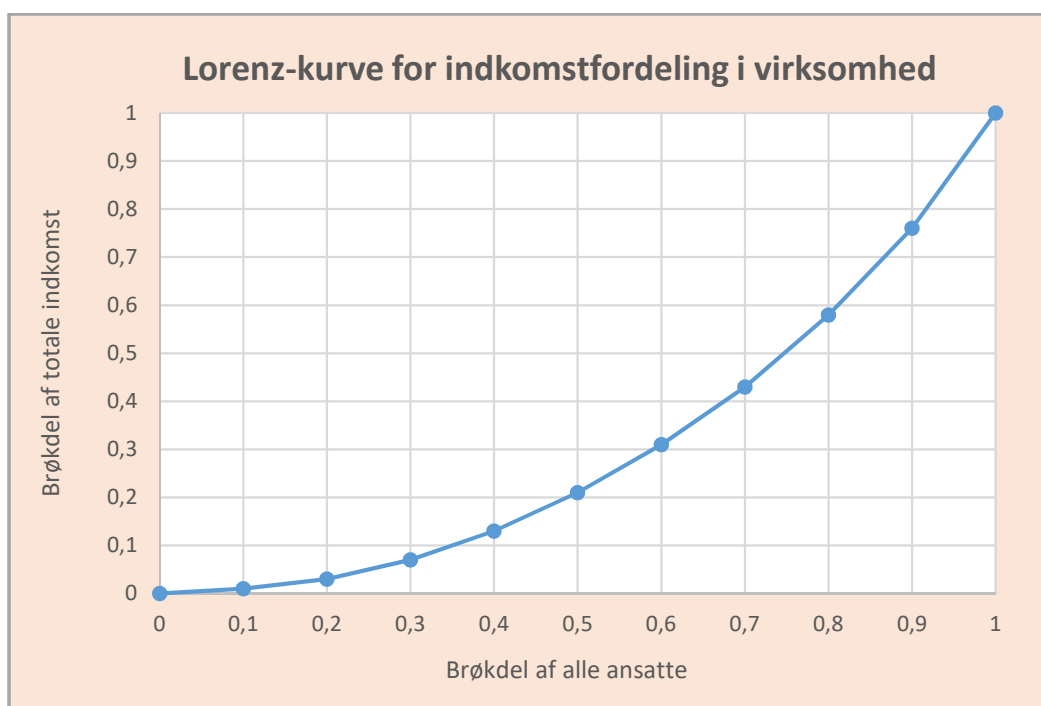
Lad os kigge på gruppen af ansatte i en stor virksomhed. De ansatte inddeles i decilgrupper efter, hvor stor deres disponible indkomst er.

Decilgruppe	1	2	3	4	5	6	7	8	9	10
Samlede indkomster i mio. kr.	5	7	10	13	15	20	25	30	35	40
Kumulerede samlede indkomster i mio. kr.	5	12	22	35	50	70	95	125	160	200
Kumulerede brøkdele af den totale indkomst	0,025	0,060	0,110	0,175	0,250	0,350	0,475	0,625	0,800	1,000

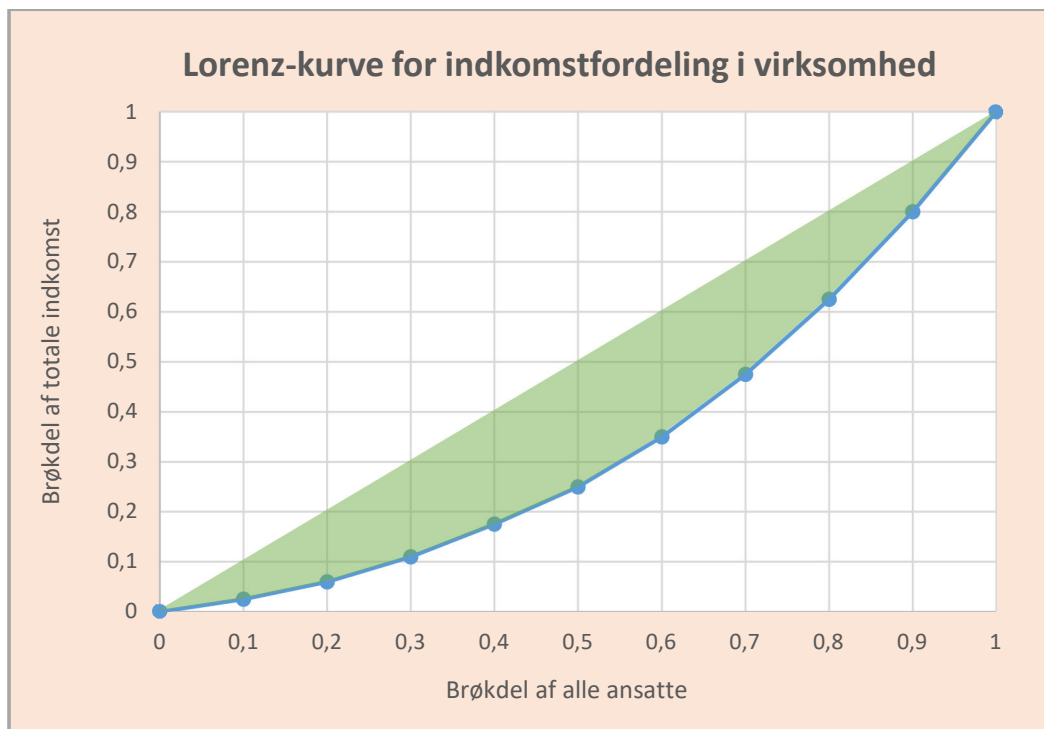
De ansatte i 1. decilgruppe har en sammenlagt disponibel indkomst på 5 mio. kr. De ansatte i 2. decilgruppe har en sammenlagt disponibel indkomst på 7 mio. kr., etc. De er nedskrevet i anden række. I 3. række er de kumulerede disponible indkomster udregnet. Således fås værdien ud for for eksempel decilgruppe 4 ved at lægge de samlede disponible indkomster for decilgrupperne 1, 2, 3 og 4 sammen: $5 + 7 + 10 + 13 = 35$, regnet i mio. kr. Denne værdi kunne naturligvis være regnet lettere ud ved at tage den forrige kumulerede værdi fra decilgruppe 3 og lægge sammen med den samlede indkomst i decilgruppe 3 alene: $22 + 13 = 35$.

I den sidste række i tabellen er udregnet de kumulerede brøkdeler af den totale indkomst. For det første er den totale indkomst her 200 mio. kr. Den kan fås ved at lægge alle tallene i række 2 sammen eller alternativt aflæse den sidste værdi i række 3 med de kumulerede indkomster. Den kumulerede brøkdel af den totale indkomst i for eksempel decilgruppe 4 fås ved at tage den kumulerede indkomst i denne decilgruppe og dividere med den totale indkomst: $35/200 = 0,175$. Tilsvarende i de øvrige decilgrupper.

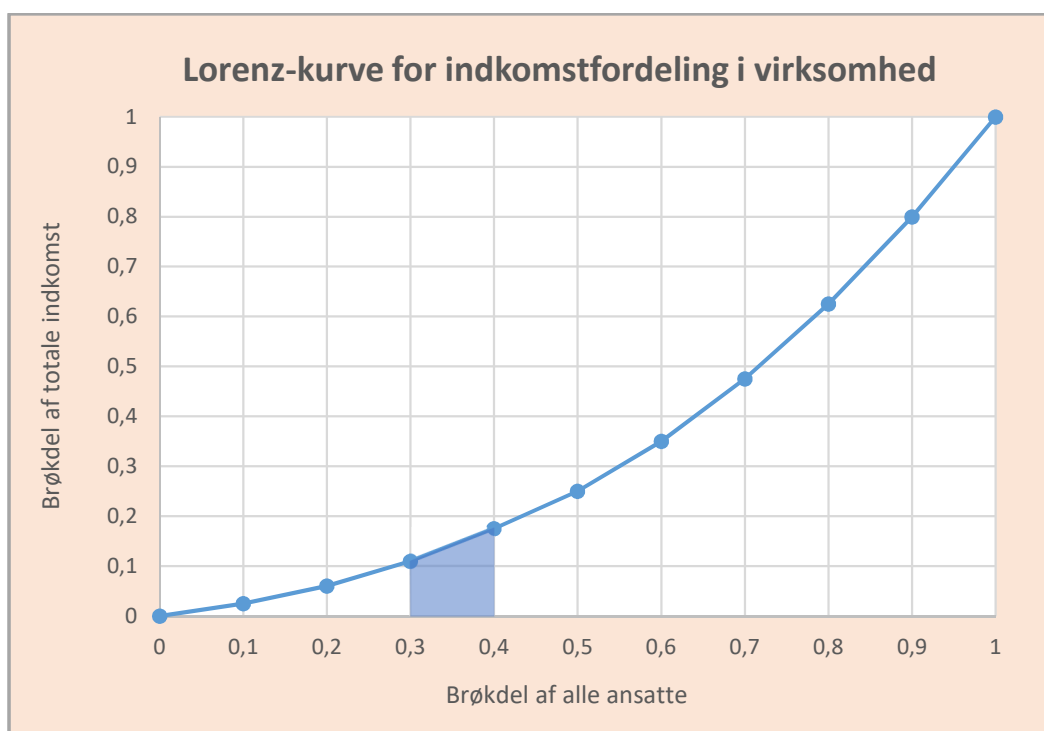
Vi er nu klar til at tegne Lorenz-kurven. Vi laver et koordinatsystem med brøkdeler af befolkningen (her ansatte) på 1. akse og brøkdelen af den samlede indkomst på 2. akse. På 1. akse er det fornuftigt at afsætte tallene $0; 0,1; 0,2; \dots; 0,9; 1,0$ fordi vi har grupperet de ansatte i grupper af 10% (decilgrupper), og 10% er jo det samme som 0,1. 2. akse går ligeledes fra 0 til 1, og her kan man også passende inddele i skridt på 0,1. For det første afsættes nu punkterne $(0,0)$ og $(1,1)$ i koordinatsystemet. Dernæst afsætter vi over 0,1 på 1. akse den kumulerede brøkdel af den totale indkomst, der står under decilgruppe 1. Det er her 0,015. Vi har altså punktet $(0,1; 0,025)$. På tilsvarende vis afsættes punkterne $(0,2; 0,060)$, $(0,3; 0,110)$, \dots , $(0,9; 0,800)$. Punktet $(0,4; 0,175)$ fortæller i øvrigt, at de 40% af de ansatte, der har den laveste indkomst, tilsammen har en indkomst, som udgør 17,5% af den totale indkomst i virksomheden. Vi forbinder nabopunkterne med rette linjestykker. Bag dette valg ligger faktisk den antagelse, at alle personer i en decilgruppe har den samme indkomst (overvej!).



Området i diagrammet begrænset af diagonalen fra $(0,0)$ til $(1,1)$ og Lorenz-kurven er markeret med grønt på figuren på næste side. Det tal, man får ved at dividere arealet af området med $1/2$ (= arealet af trekanten under diagonalen) og derefter gange resultatet med 100, kaldes *gini-koefficienten*. Vi skal snart indse, at jo større gini-koefficienten er, jo større ulighed er der i befolkningens indkomst.



En fornuftig måde at udregne arealet af det grønne område på er ved at tage arealet af hele trekanten under diagonalen (det er $1/2$) og fratække arealet af det polygonale område under Lorenz-kurven. Sidstnævnte består af 10 trapezer. Arealet af trapezet mellem 0,3 og 0,4 på 1.aksen fås således ved at tage gennemsnittet af y -værdierne i 0,3 og 0,4, dvs. gennemsnittet af 0,110 og 0,175 og gange resultatet med bredden af trapezet, som er $0,4 - 0,3 = 0,1$. Det giver arealet $(0,110 + 0,175)/2 \cdot 0,1 = 0,01425$.



Her får vi følgende udregninger:

Areal under Lorenz-kurve =

$$\begin{aligned} & \frac{0+0,025}{2} \cdot 0,10 + \frac{0,025+0,060}{2} \cdot 0,10 + \frac{0,060+0,110}{2} \cdot 0,10 + \frac{0,110+0,175}{2} \cdot 0,10 + \\ & \frac{0,175+0,250}{2} \cdot 0,10 + \frac{0,250+0,350}{2} \cdot 0,10 + \frac{0,350+0,475}{2} \cdot 0,10 + \\ & \frac{0,475+0,625}{2} \cdot 0,10 + \frac{0,625+0,800}{2} \cdot 0,10 + \frac{0,800+1}{2} \cdot 0,10 = 0,337 \end{aligned}$$

Arealet af det grønne område fås nu ved at trække tallet fra $1/2$: $\frac{1}{2} - 0,337 = 0,163$. Vi får herefter gini-koefficienten ved at dividere med arealet af trekanten under diagonalen, som er 0,5, og derefter gange med 100: $0,163/0,5 \cdot 100 = 32,6$. Dette tal er altså et mål for uligheden i virksomheden. Jo større gini-koefficienten er, jo større ulighed er der i virksomheden, hvad angår indkomsten.

$(1) \quad \text{Gini-koefficient} = \frac{\text{Areal af grønne område}}{\frac{1}{2}} \cdot 100$

I appendikset sidst i dette dokument kan du se en beskrivelse af, hvordan du kan udregne tallene i tabellen, tegne Lorenz-kurven og udregne gini-koefficienten mere automatisk.

□

Opgave 1 (Egenskaber)

For at få større indsigt i Lorenz-kurvens og gini-koefficientens egenskaber, kan du forsøge at besvare spørgsmålene nedenfor.

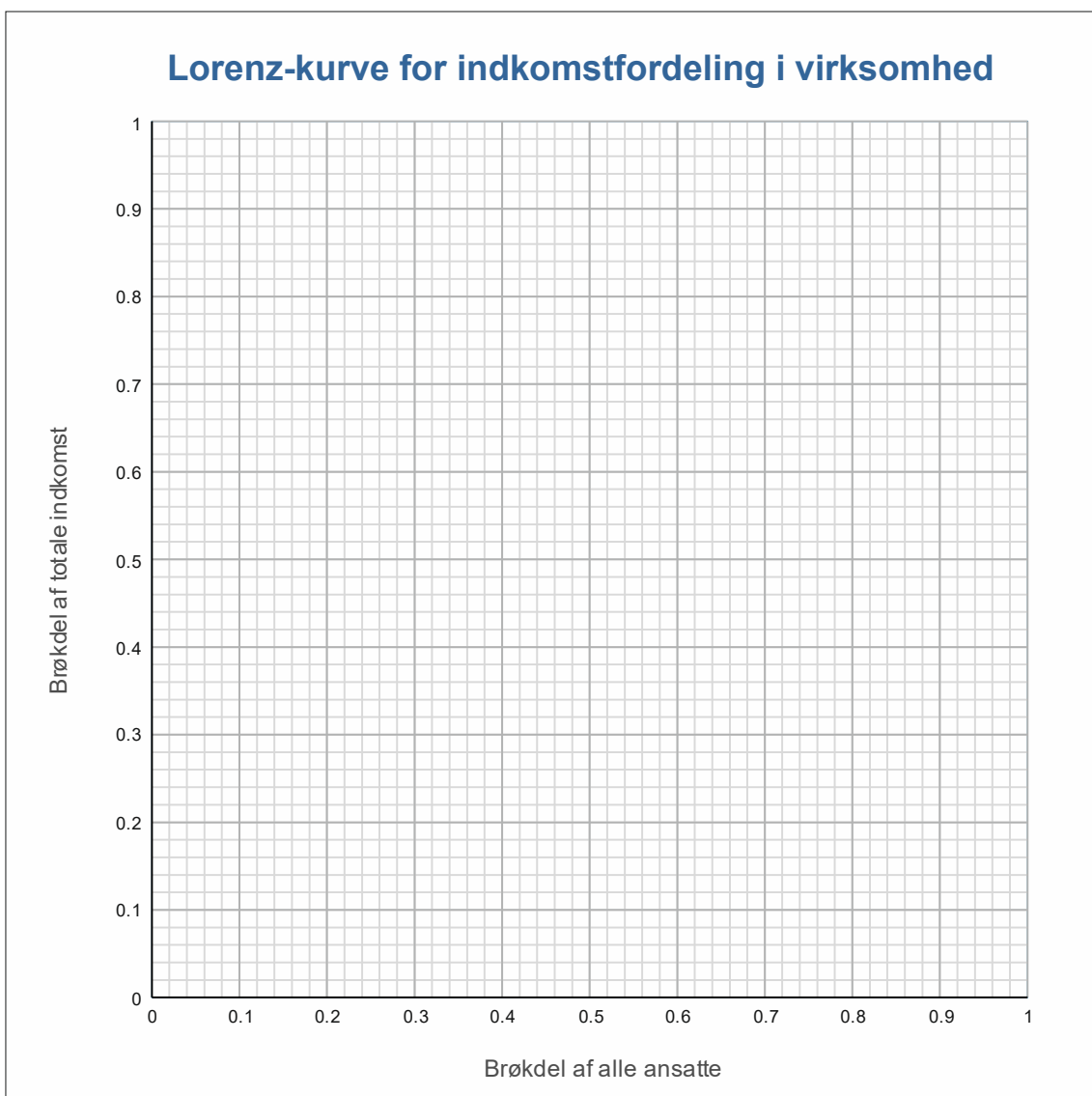
- a) Hvorfor er det klart, at Lorenz-kurven altid må være voksende?
- b) Hvorfor vil Lorenz-kurven altid ligge på eller under diagonalen?
- c) Hvorfor kan linjestykkerne på Lorenz-kurven aldrig aftage i hældning, når man bevæger sig mod højre på kurven?
- d) Hvad kan man sige om fordelingen af indkomster, hvis Lorenz-kurven er sammenfaldende med diagonalen, og hvilken værdi for gini-koefficienten vil man få i dette tilfælde?
- e) Hvordan vil Lorenz-kurven se ud for en befolkning, hvor der er maksimal ulighed, og hvor stor vil gini-koefficienten være i det tilfælde?
- f) (Svær) Som nævnt forbindes punkterne i Lorenz-kurven med rette linjestykker. Forklar hvorfor det normalt vil være mere rimeligt at lave en blød kurve igennem punkterne, hvor hældningen hele tiden vokser?
- g) (Svær) Hvis du har et CAS-værktøj til rådighed, kan du fodre værktøjet med de 11 punkter på Lorenz-kurven og lave et fit med et polynomium. Benyt det til at udregne en anden værdi for gini-koefficienten baseret på den nye Lorenz-kurve.

Opgave 2

I en virksomhed fordeler de samlede indkomster i decilgrupperne sig som det fremgår af tabellen nedenfor.

Decilgruppe	1	2	3	4	5	6	7	8	9	10
Samlede indkomster i mio. kr.	1	2	4	6	8	10	12	15	18	24
Kumulerede samlede indkomster i mio. kr.										
Kumulerede brøkdele af den totale indkomst										

- Udfyld de tomme felter i tabellen efter principperne givet i eksempel 1.
- Tegn Lorenz-kurven i koordinatsystemet nedenfor.
- Bestem gini-koefficienten for medarbejdernes indkomstfordeling i virksomheden.



Opgave 3* (Den gennemsnitlige indkomst kan også bruges ...)

I eksempel 1 var den samlede indkomst for personerne i hver decilgruppe til rådighed. Undertiden har man kun opgivet den *gennemsnitlige* indkomst for personerne i hver decilgruppe. Det er for eksempel tilfældet forskellige steder i databanken hos Danmarks Statistik. Påvis ved matematiske argumenter, at man lige så godt kan bruge den gennemsnitlige indkomst i hver decilgruppe til at beregne de kumulerede brøkdele af den totale indkomst – som skal bruges for at tegne Lorenz-kurven og beregne gini-koefficienten.

Hjælp: Betegn eventuelt de samlede indkomster for personerne i hver decilgruppe med henholdsvis a_1, a_2, \dots, a_{10} , og betegn den gennemsnitlige indkomst i hver decilgruppe med henholdsvis b_1, b_2, \dots, b_{10} . Lad desuden n betegne antallet af personer i hver decilgruppe. Udnyt at $a_k = n \cdot b_k$ for alle $k = 1, 2, \dots, 10$.

□

Man kan udregne gini-koefficienter for indkomster i såvel byer, lande, virksomheder, etc. Bare man har relevant data til rådighed. I den følgende opgave skal du se, hvordan man kan bruge Danmarks statistik til at udregne gini-koefficienten for de disponible indkomster i Danmark for et givet år.

Opgaver 4 (Ulighed i Danmark – Danmarks statistik)

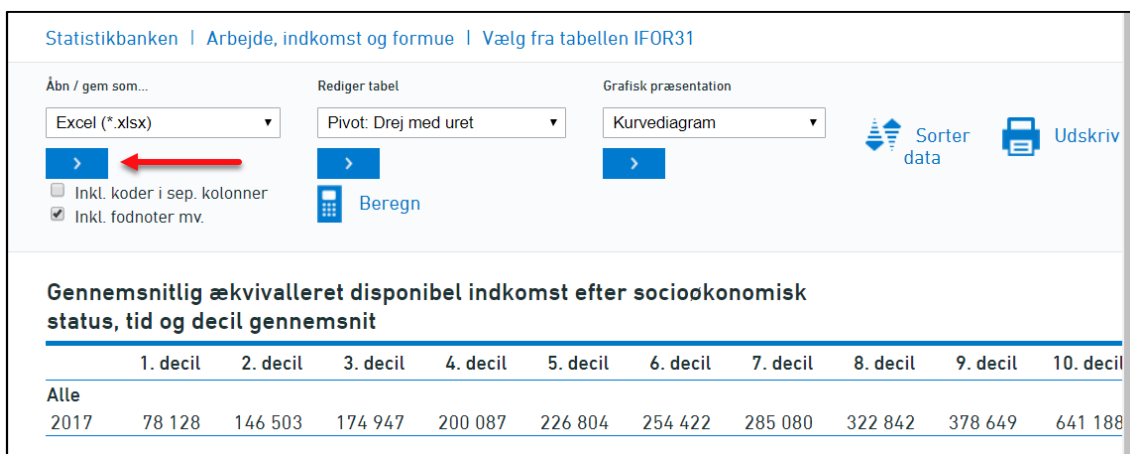
1. Benyt link [4] til at tilgå den side i *Statistikbanken* hos Danmarks Statistik (DST), som viser *Gennemsnitlig ækvivaleret disponibel indkomst efter decil og socioøkonomisk status*. Vi kan for eksempel få de ti deciler for året 2017 i Danmark via de valg som ses på figuren nedenfor.

The screenshot shows the Danish Statistical Database (Statistikbanken) interface. The page title is "IFOR31: GENNEMSNITLIG ÆKVIVALERET DISPONIBEL INDKOMST EFTER DECIL OG SOCIOØKONOMISK STATUS". The unit is "kr.". There are three dropdown menus for selection:

- DECIL GENNEMSNIT (10):** A list of deciles from 1 to 10.
- SOCIOØKONOMISK STATUS (17):** A list of socio-economic status categories including "Alle", "Selvstændige", "Lønmodtagere i alt", "Lønmodtagere med ledelsesarbejde", "Lønmodtagere på højeste niveau", "Lønmodtagere på mellem niveau", "Lønmodtagere på grund niveau", "Lønmodtagere i øvrigt", "Lønmodtagere uden nærmere angivelse", "Arbejdsløse", and "Midlertidigt ude af arbejdsstyrken".
- ÅR:** A list of years from 2007 to 2017.

At the bottom, it shows "Antal valgte tal til tabellen: 10 (Vælg max. 10000)" and two buttons: "ANNULLER" and "VIS TABEL".

2. Tryk på *Vis tabel*, hvorved følgende data viser sig:



Statistikbanken | Arbejde, indkomst og formue | Vælg fra tabellen IFOR31

Abn / gem som... Rediger tabel Grafisk præsentation

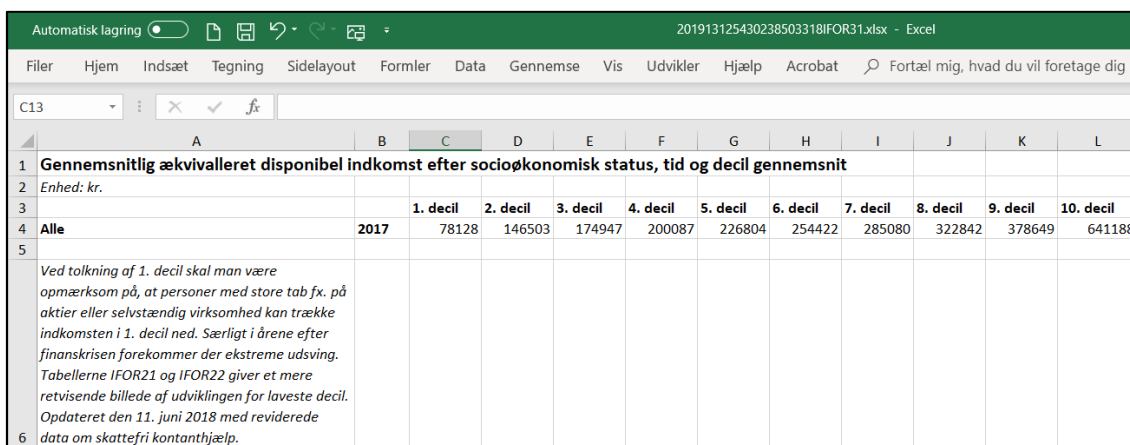
Excel (*.xlsx) Pivot: Drej med uret Kurvediagram

Inkl. koder i sep. kolonner Inkl. fodnoter mv. Beregn Sorter data Udskriv

Gennemsnitlig ækvivaleret disponibel indkomst efter socioøkonomisk status, tid og decil gennemsnit

	1. decil	2. decil	3. decil	4. decil	5. decil	6. decil	7. decil	8. decil	9. decil	10. decil
Alle										
2017	78 128	146 503	174 947	200 087	226 804	254 422	285 080	322 842	378 649	641 188

3. Tryk på knappen ud for Excel (vist med rød pil). Det giver data i et Excel regneark. Ud for hver decilgruppe står gennemsnittet af indkomsterne for indbyggerne med en indkomst i den givne decilgruppe. For eksempel ser vi, at indbyggerne med de 10% laveste disponible indkomster har en *gennemsnitlig* disponibel indkomst på 78128 kr. Ligeså for de øvrige decilgrupper.



Automatisk lagring 201913125430238503318IFOR31.xlsx - Excel

File Hjem Indsæt Tegning Sidelayout Formler Data Gennemse Vis Udvikler Hjælp Acrobat Fortæl mig, hvad du vil foretage dig

	A	B	C	D	E	F	G	H	I	J	K	L
1	Gennemsnitlig ækvivaleret disponibel indkomst efter socioøkonomisk status, tid og decil gennemsnit											
2	Enhed: kr.											
3			1. decil	2. decil	3. decil	4. decil	5. decil	6. decil	7. decil	8. decil	9. decil	10. decil
4	Alle	2017	78128	146503	174947	200087	226804	254422	285080	322842	378649	641188
5	Ved tolkning af 1. decil skal man være opmærksom på, at personer med store tab fx. på aktier eller selvstændig virksomhed kan trække indkomsten i 1. decil ned. Særligt i årene efter finanskrisen forekommer der ekstreme udsving. Tabellerne IFOR21 og IFOR22 giver et mere retvisende billede af udviklingen for laveste decil. Opdateret den 11. juni 2018 med reviderede data om skattefri kontanthjælp.											

4. Benyt nu Excel-teknikken skitseret i appendikset efter denne opgave til at udregne de kumulerede brøkdeler af den totale indkomst, tegne Lorenz-kurven og udregne Gini-koefficienten. Husk at de gennemsnitlige indkomster i hver decilgruppe er lige så gode at bruge som de samlede indkomster i hver decilgruppe, ifølge opgave 3!
5. Når du har udført punkt 4, kan du eventuelt tage data for et andet år ned fra Danmarks Statistik med henblik på at udregne gini-koefficienten og sammenligne. Er der blevet mere ulighed de senere år? NB! Hvis du er smart, kan du copy/paste rækken med data fra det nye år ind i det Excel-regneark, du fra punkt 4. Så skal du ikke til at gentage processen!
6. Man kan også få Gini-koefficienterne for et givet år direkte ved at bruge link [5] bagerst i dette tillæg. Det giver ikke helt den samme Gini-koefficient, som under punkt 4. Det skyldes, at DST har anvendt mange flere punkter til deres Lorenz-kurve, og fordi de har taget hensyn til særlige effekter i fordelings endepunkter.

Appendiks: Lorenz-kurve og gini-koefficient i Excel

Vi skal se, hvordan de udregninger, vi foretog i eksempel 1, smart kan udføres i Excel.

1. Tast data ind i et Excel-regneark, som vist herunder:

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2													
3	Decilgruppe		1	2	3	4	5	6	7	8	9	10	
4	Samlede indkomster i mio. kr.		5	7	10	13	15	20	25	30	35	40	
5	Kum. samlede indkomster i mio. kr.												
6	Kum. brøkdele af totale indkomst												

2. Skriv et 0 i celle B5. Skriv herefter formelen: $=C4+B5$ i celle C5, som vist her:

The screenshot shows the same Excel spreadsheet as above, but with the following changes:

- Cell B5 now contains the value 0.
- Cell C5 contains the formula $=C4+B5$.

3. Tryk på **Enter** for at udføre formelen. Nedkopier formelfeltet ved at trække vandret i den lille sorte firkant i nederste højre hjørne af cellen.

The screenshot shows the result of the formula in cell C5, which is 5. The drag handle (a small square) is visible in the bottom-right corner of cell C5, indicating that the formula is being copied horizontally to the right.

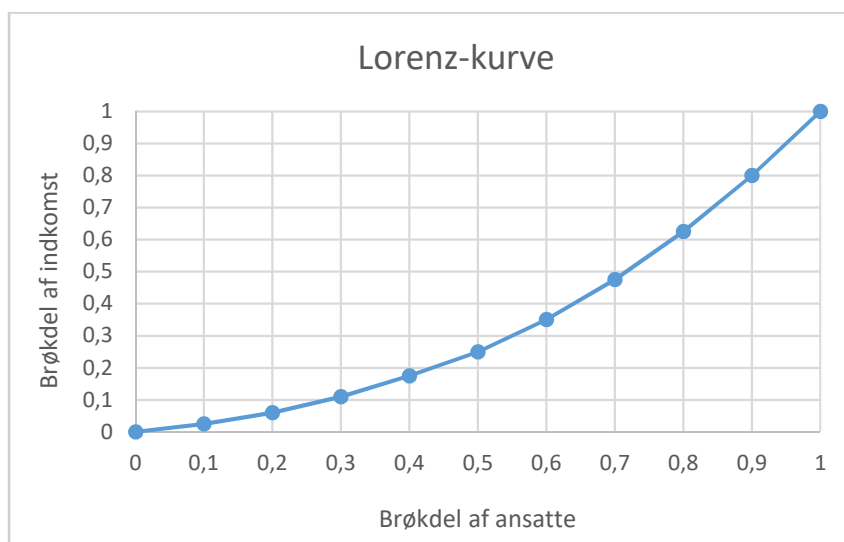
4. Skriv endnu et 0, denne gang i celle B6. Skriv herefter formelen: $=C5/SL\$5$ i celle C6, som vist herunder: \$-tegnene er vigtige for at feltet L8 ikke nedkopieres senere ...

The screenshot shows the final step of the calculation. Cell B6 now contains the value 0, and cell C6 contains the formula $=C5/SL\$5$.

13. Vi skal have lavet et diagram: Markér felterne fra B1 til L1 og hold derefter **Ctrl**-tasten nede, mens du markerer felterne fra B6 til L6. Mens de to rækker er markeret vælges menuen *Indsæt* og det diagram, som fremgår af figuren herunder:

Decilgruppe	1	2	3	4	5
Samlede indkomster i mio. kr.	5	7	10	13	15
Kum. samlede indkomster i mio. kr.	0	5	12	22	35
Kum. brøkdeler af totale indkomst	0	0,025	0,060	0,110	0,175
Trapez-arealer	0,00125	0,00425	0,00850	0,01425	0,02125
Gini-koefficient	32,6				

14. Efter at have rettet akserne til og valgt aksetitler, kan det se således ud:



□

Hvis du har lyst til at læse mere, kan du læse [3], hvor der også er omtalt et andet indeks, kaldet *den maksimale udjævningsprocent*, som på engelsk kaldes *Robin Hood index* eller *Hoover index*). Se også [2] for alternative ulighedsmål.

Links

- [1] <http://www.dst.dk/ext/1872554511/0/velfaerd/Disponibel-indkomst--pdf>
- [2] <https://da.wikipedia.org/wiki/Gini-koefficient>
- [3] https://www.dst.dk/ext/arbejde-loen-og-indkomst/Gini_2--pdf
- [4] <http://www.statistikbanken.dk/IFOR31>
- [5] <http://statistikbanken.dk/ifor41>