

≠ [Ulighed]

Dennis Pipenbring

12. september 2011

Indhold

1	Forord	3
2	Indledning	4
3	Observation og hyppighed	4
4	Frekvens	4
5	Middeltal	6
6	Summerede frekvenser	8
6.1	Histogram	9
6.2	Sumkurve	10
6.3	Boksplot	12
6.4	Lorenzdiagram	13
6.5	Gini-koefficient	17
7	Diskrete observationer	19
7.1	Pindediagram	20
7.2	Trappediagram	20

1 Forord

Materialet er udarbejdet så de studerende selvstændigt kan arbejde med det. Det er således ikke meningen, at store dele af forløbet skal gå med tavleundervisning. Lærerens opgave er at definere hvilke opgaver eleverne skal besvare evt. i samspil med samfundsfag. Udvalget af opgaver skal være således at besvarelsen som helhed udgør en rapport i matematik. Hvis der samarbejdes med samfundsfag er det vigtigt at der indgår ekstra samfundsfagligt materiale, til at belyse et emne om ulighed f.eks. levevilkår.

2 Indledning

Statistik er et redskab til at beskrive et datamateriale. Ofte vil et datamateriale være for omfattende til, at det er muligt at overskue og konkludere noget om det blot ved at se på selve datamaterialet. Derfor er der udviklet forskellige metoder, modeller og tabeller, som gør det muligt at overskue store mængder af data, og på baggrund af disse modeller og tabeller er det muligt at drage konklusioner på baggrund af datamaterialet. Hvis man ikke forstår at læse tabeller og hvis man ikke ved, hvordan man skal forstå de statistiske begreber, kan man *ikke* selv drage konklusioner eller forstå andres konklusioner. Så man vil meget let blive snydt når man læser, hører eller ser andres udtalelser, når disse tager udgangspunkt i statistisk materiale.

3 Observation og hyppighed

Den mindste enhed i et datasæt er en observation. Det kan være observationer af næsten alt hvad, man kan tænke sig, f.eks. kan det være en families indkomst. I dette tilfælde vil der være tale om 2.741.498 observationer. Observationerne er i dette tilfælde **grupperede** af hensyn til overskueligheden.

Indkomst	Familier i alt
Under 200.000 kr.	698.872
200.000-299.999 kr.	516.181
300.000-399.999 kr.	378.214
400.000-499.999 kr.	248.121
500.000-599.999 kr.	237.841
600.000 kr. og derover	662.269
I alt	2.741.498

Kilde: Danmarks Statistik - statistikbanken.dk

Efter at have lavet disse observationer, ses på hyppigheden af de forskellige observationsværdier. Dette betyder, at der spørges:

"Hvor ofte (eller hvor hyppigt) er der familier der har en indkomst mellem 200.000 og 299.999 kr."

Svar: "Der er 516.181 familier der har en indkomst mellem 200.000 og 299.999 kr."

I dette tilfælde er '200.000 - 299.999 kr.' observationsværdien og hyppigheden er 516.181.

4 Frekvens

Det næste begreb er frekvens. Frekvensen er defineret som hyppigheden divideret med det totale antal observationer.

$$\text{Frekvensen} = \frac{\text{Hyppigheden}}{\text{Det totale antal observationer}} \cdot 100\%$$

I eksemplet fra før, er frekvensen af 'antal familier der har en indkomst mellem 200.000 og 299.999 kr.' følgende:

$$\text{Frekvensen af 200.000 - 299.999 er: } \frac{516.181}{2.741.498} \cdot 100\% = 18,8\%$$

På samme måde udregnes de øvrige frekvenser. I eksemplet er frekvensen for de øvrige hyppigheder.

Indkomst	Familier i alt	Frekvens
Under 200.000 kr.	698.872	25,5 %
200.000-299.999 kr.	516.181	18,8 %
300.000-399.999 kr.	378.214	13,8 %
400.000-499.999 kr.	248.121	9,1 %
500.000-599.999 kr.	237.841	8,7 %
600.000 kr. og derover	662.269	24,2 %
I alt	2.741.498	100,1 %

Kilde: Danmarks Statistik - statistikbanken.dk

Det er værd at bemærke at summen af frekvenserne altid skal være 100 %. Hvis det ikke er tilfældet, er der lavet en regnefejl. Der kan godt være tale om en fejl der skyldes afrunding, men så vil resultat ofte være $100\% \pm 0,5\%$.

Opgave 4.1 Lav en hyppighedstabel over nedenstående data. Tabellen skal indeholde hyppigheden og frekvensen for observationerne.

Indkomst	Par i alt
Under 200.000 kr.	37.264
200.000-299.999 kr.	151.005
300.000-399.999 kr.	148.708
400.000-499.999 kr.	158.789
500.000-599.999 kr.	203.937
600.000 kr. og derover	624.154
I alt	1.323.857

Kilde: Danmarks Statistik - statistikbanken.dk

Opgave 4.2 Lav en hyppighedstabel over nedenstående data. Tabellen skal indeholde hyppigheden og frekvensen for observationerne.

Indkomst	Enlige i alt
Under 200.000 kr.	661.608
200.000-299.999 kr.	365.176
300.000-399.999 kr.	229.506
400.000-499.999 kr.	89.332
500.000-599.999 kr.	33.904
600.000 kr. og derover	38.115
I alt	1.417.641

Kilde: Danmarks Statistik - statistikbanken.dk

Opgave 4.3 Kommentér forskellen mellem indkomsten for enlige og par. Hvorfor er det vigtigt at omregne til frekvens (procent)?

5 Middeltal

Det næste begreb er middeltal - også kaldet gennemsnit eller middelværdi. Middeltallet udregnes ved at summere produkterne af middeltallet for de enkelte intervaller med frekvensen af de enkelte observationsværdier.

Indkomst	Middeltal for familier i alt	Frekvens
Under 200.000 kr.	136.975 kr.	25,5 %
200.000-299.999 kr.	247.782 kr.	18,8 %
300.000-399.999 kr.	345.991 kr.	13,8 %
400.000-499.999 kr.	448.028 kr.	9,1 %
500.000-599.999 kr.	550.331 kr.	8,7 %
600.000 kr. og derover	993.927 kr.	24,2 %
Samlet middeltal	443.208 kr.	100 %

Kilde: Danmarks Statistik - statistikbanken.dk

Kendes gennemsnittet ikke for de enkelte intervaller udregnes middeltallet ved at summere produkterne af **intervalmidtpunkt** af observationsværdiintervallet og frekvensen for hver observationsværdi. Intervalmidtpunktet for det sidste interval sættes til 800.000 kr.

Eksempel 5.1 Middeltallet for indkomsten for familier i alt bliver:

$$136.975 \cdot 25,5\% + 247.782 \cdot 18,8\% + \dots + 993.927 \cdot 24,2\% = 443.208 \text{ kr}$$

Bemærk at middeltallet skal være et tal der ligger mellem det højeste og det laveste tal i datasættet, ellers er der lavet en regnefejl.

Opgave 5.2 Udregn middeltallet for par og enlige.

Indkomst	Middeltal for par
Under 200.000 kr.	125.147 kr.
200.000-299.999 kr.	248.740 kr.
300.000-399.999 kr.	349.629 kr.
400.000-499.999 kr.	451.612 kr.
500.000-599.999 kr.	551.606 kr.
600.000 kr. og derover	924.824 kr.
Samlet middeltal	646.334 kr.

Kilde: Danmarks Statistik - statistikbanken.dk

Indkomst	Middeltal for enlige
Under 200.000 kr.	137.641 kr.
200.000-299.999 kr.	247.386 kr.
300.000-399.999 kr.	343.634 kr.
400.000-499.999 kr.	441.658 kr.
500.000-599.999 kr.	542.663 kr.
600.000 kr. og derover	1.082.991 kr.
Samlet middeltal	253.520 kr.

Kilde: Danmarks Statistik - statistikbanken.dk

6 Summerede frekvenser

Den summerede frekvens er summen af alle tidligere frekvenser, dette lader sig bedst illustrere ved et eksempel.

Indkomst	Familier i alt	Frekvens	Summeret frekvens
Under 200.000 kr.	698.872	25,5 %	25,5 %
200.000-299.999 kr.	516.181	18,8 %	25,5 % + 18,8 %
300.000-399.999 kr.	378.214	13,8 %	44,3 % + 13,8 %
400.000-499.999 kr.	248.121	9,1 %	58,1 % + 9,1 %
500.000-599.999 kr.	237.841	8,7 %	67,2 % + 8,7 %
600.000 kr. og derover	662.269	24,2 %	75,9 % + 24,2 %
I alt	2.741.498	100,1 %	

Kilde: Danmarks Statistik - statistikbanken.dk

Tabellen skrives ikke som i tabellen ovenfor. De summerede frekvenser lægges sammen resultatet skrives. Så kommer tabellen til at se således ud:

Indkomst	Familier i alt	Frekvens	Summeret frekvens
Under 200.000 kr.	698.872	25,5 %	25,5 %
200.000-299.999 kr.	516.181	18,8 %	44,3 %
300.000-399.999 kr.	378.214	13,8 %	58,1 %
400.000-499.999 kr.	248.121	9,1 %	67,2 %
500.000-599.999 kr.	237.841	8,7 %	75,9 %
600.000 kr. og derover	662.269	24,2 %	100,1 %
I alt	2.741.498	100,1 %	

Kilde: Danmarks Statistik - statistikbanken.dk

Hvis den summerede frekvens ikke ender med at være $100\% \pm 0,5\%$ er der hvert fald lavet mindst en fejl, eller der er afrundet for meget.

Den summerede frekvens kan bruges til f.eks. at fastslå at 58,1% af familierne har en indkomst, der er mindre end 400.000. Eller at fastslå at $100\% - 67,2\% = 32,8\%$ af familierne har en indkomst, der er over 500.000.

Opgave 6.1 Udregn den summerede frekvens for enlige og par.

Opgave 6.2 Hvor mange procent af parene hhv. de enlige har en indkomst over 400.000 kr?

6.1 Histogram

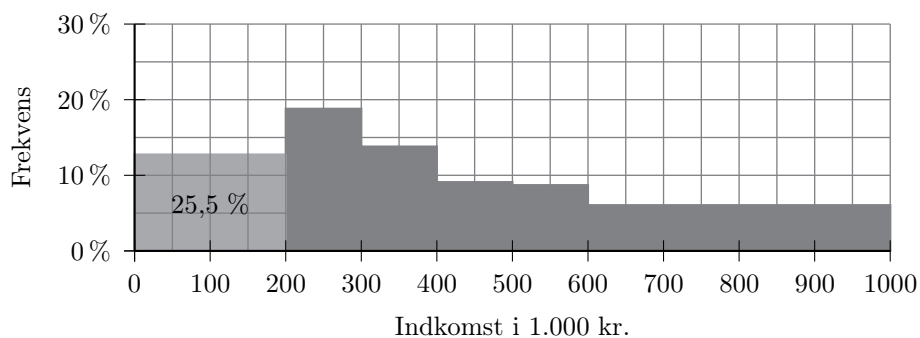
Et diagram er en grafisk måde at illustrere frekvenser/summerede frekvenser og observationsværdier. Der er to forskellige typer af diagrammer: Histogram og sumkurve. Histogram bruges til at illustrere sammenhængen mellem frekvens og observationsværdi, mens sumkurven bruges til at illustrere sammenhængen mellem summeret frekvens og observationsværdi.

Med udgangspunkt i følgende hyppighedstabel tegnes et histogram.

Indkomst	Familier i alt	Frekvens	Summeret frekvens
Under 200.000 kr.	698.872	25,5 %	25,5 %
200.000-299.999 kr.	516.181	18,8 %	44,3 %
300.000-399.999 kr.	378.214	13,8 %	58,1 %
400.000-499.999 kr.	248.121	9,1 %	67,2 %
500.000-599.999 kr.	237.841	8,7 %	75,9 %
600.000 kr. og derover	662.269	24,2 %	100,1 %
I alt	2.741.498	100,1 %	

Kilde: Danmarks Statistik - statistikbanken.dk

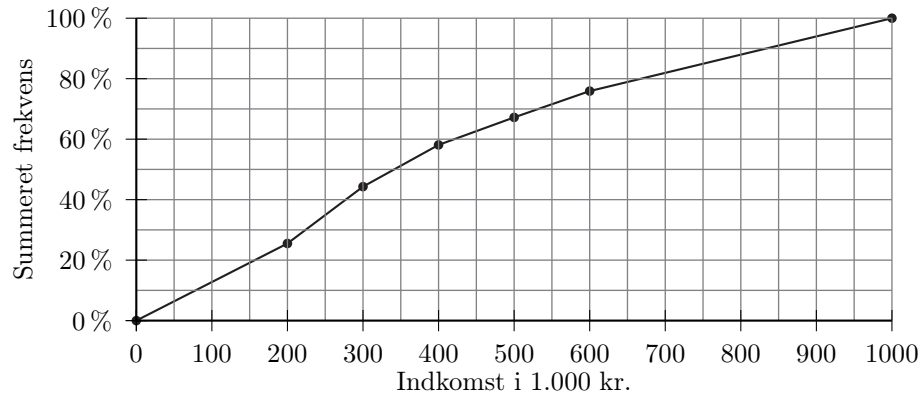
I histogrammet vælges en standard for bredden af søjlerne. I dette tilfælde er bredden af en sølje, der dækker over 100.000 kr., 1 cm. Det betyder, at det første interval, som går fra 0 til 200.000, og derfor dækker over et interval på 200.000, har en bredden på 2 cm, og procentsatsen skal derfor divideres med 2. På histogrammet er denne sølje *lys grå*. Det betyder med andre ord, at hvert tern svarer til 2,5 %. Det sidste interval som er sat fra 600.000 - 1.000.000 dækker over 4 cm og derfor skal procentsatsen divideres med 4.



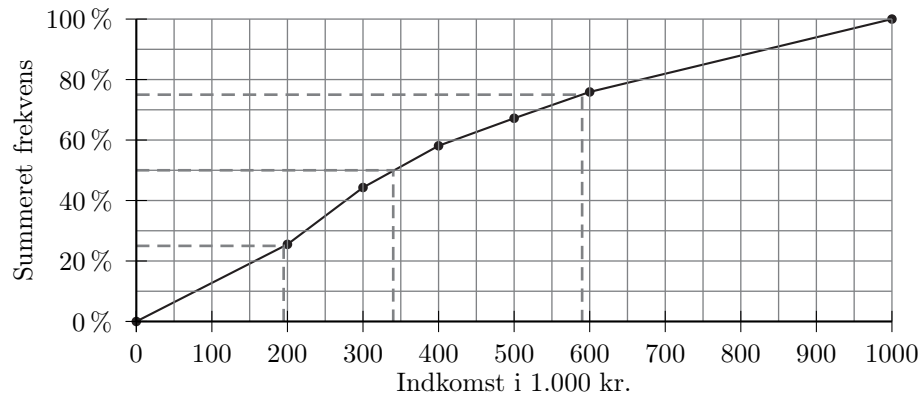
Opgave 6.3 Tegn histogrammet for de enlige og parenes indkomst.

6.2 Sumkurve

Sumkurven konstrueres ved at afmærke den summerede frekvens som y -værdien og endepunkterne af de tilhørende intervaller som x -værdien i et koordinatsystem. Der ud over afmærkes startpunktet af det første interval på x -aksen, som i dette tilfælde er 0. På grafen er disse punkter markeret med \bullet . Herefter sættes streger mellem punkterne.



Nu kan kvartilerne aflæses ved at gå ud fra hhv. 25 %, 50 % og 75 % til kurven og derefter ned på x -aksen og aflæse kvartilerne.



Kvartil	Indkomst
Nedre kvartil	195.000 kr.
Median	340.000 kr.
Øvre kvartil	590.000 kr.

En sumkurve kan også anvendes til at aflæse hvor mange procent der er indenfor et givet interval. På sumkurven ses at 35 % har en indkomst der er 250.000 kr. eller under og der er 62 % der har en indkomst der er 450.000 kr. eller under. Det betyder at der er (62 % - 35 %) 27 % der har en indkomst mellem 250.000 kr. og 450.000 kr.

Opgave 6.4 Tegn sumkurven for de enliges og parenes indkomst.

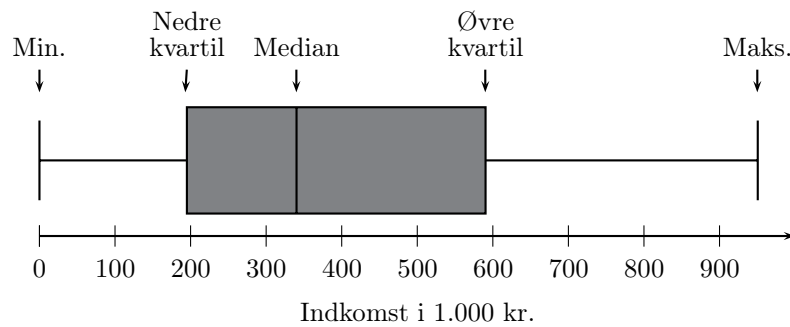
Opgave 6.5 Aflæs kvartilerne for de enliges og parenes indkomst.

Opgave 6.6 Udregn hvor mange procent af hhv. enlige og par, der har en indkomst mellem 250.000 kr. og 450.000 kr.

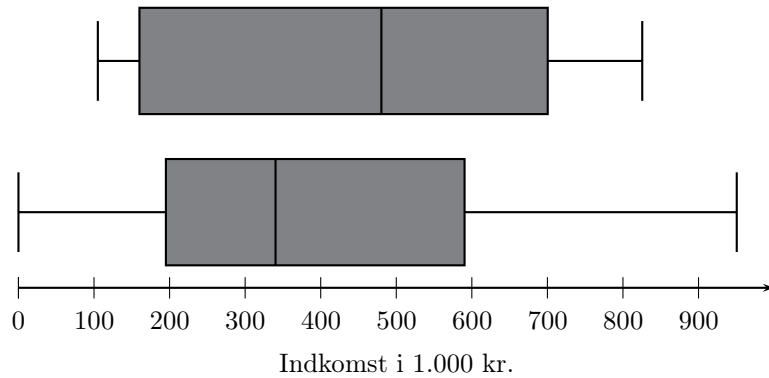
6.3 Boksplot

Boksplot kan tegnes ud fra en sumkurve, hvis den mindste og den største observationsværdi i datasættet er kendt. De resterende deskriptorer aflæses på sumkurven.

Boksplottet konstrueres ved at tegne en x -akse, der er identisk med x -aksen i sumkurven. Ca. 1 cm over x -aksen ved den mindste observation tegnes en lodret streg på ca. 1 cm. Tilsvarende for de andre deskriptorer. Derefter forbindes midtpunktet af strengen for den mindste observation med midtpunktet af strengen for nedre kvartil med en vandret streg. Derefter forbindes endepunkterne af stregerne for nedre kvartil og median og øvre kvartil med to vandrette streger. Til sidst forbindes midtpunktet af strengen for den største observation med midtpunktet af strengen for øvre kvartil med en vandret streg. Så kommer boksplottet for indkomsten for alle familier til at se således ud:



Meget ofte anvendes boksplottet til at sammenligne to eller flere datasæt. Ét boksplot konstrueres for hver af de to datasæt, og indtegnes i samme koordinatsystem.

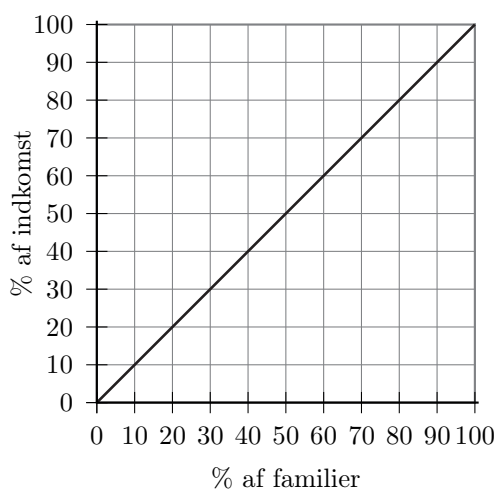


Opgave 6.7 Tegn boksplottet for indkomsten for enlige og par i samme koordinatsystem. Antag at minimum for begge grupper er 0 kr. og at maksimum er 1.000.000 for begge grupper.

Opgave 6.8 Kommentér forskellen i indkomsten for enlige og par med udgangspunkt i boksplottet.

6.4 Lorenzdiagram

Lorenzdiagrammet viser hvor stor uligheden er f.eks. i forhold til indkomsten. I nedenstående eksempel er der fuldstændig *lighed* idet de 10 % af familierne, som har den laveste indkomst, også har 10 % af indkomsten. Og 20 % af familierne, som har den laveste indkomst, også har 20 % af indkomsten osv.



For at tegne Lorenzdiagrammet skal den samlede indkomst for de enkelte grupper beregnes. For at beregne den, ganges den middeltallet indkomst i gruppen med antallet af familier i gruppen.

Indkomst	Antal familier	Middeltal for familier i alt
Under 200.000 kr.	698.872	136.975 kr.
200.000-299.999 kr.	516.181	247.782 kr.
300.000-399.999 kr.	378.214	345.991 kr.
400.000-499.999 kr.	248.121	448.028 kr.
500.000-599.999 kr.	237.841	550.331 kr.
600.000 kr. og derover	662.269	993.927 kr.
Samlet middeltal	2.741.498	443.208 kr.

Kilde: Danmarks Statistik - statistikbanken.dk

Indkomst	Samlede indkomst
Under 200.000 kr.	95.727.992.200 kr.
200.000-299.999 kr.	127.900.360.542 kr.
300.000-399.999 kr.	130.858.640.074 kr.
400.000-499.999 kr.	111.165.155.388 kr.
500.000-599.999 kr.	130.891.275.371 kr.
600.000 kr. og derover	618.510.900.363 kr.
Samlet indkomst	1.215.053.845.584 kr.

Kilde: Danmarks Statistik - statistikbanken.dk

Indkomstfrekvensen er den samlede indkomst divideret med den totale indkomst for alle familier.

Eksempel 6.9 Indkomstfrekvensen for gruppen 200.000-299.999 er

$$\frac{127.900.360.542}{1.215.053.845.584} \cdot 100\% = 10,53\%$$

Indkomst	Familier i alt	Samlede indkomst	Indkomstfrekvens
Under 200.000 kr.	698.872	95.727.992.200	7,88 %
200.000-299.999 kr.	516.181	127.900.360.542	10,53 %
300.000-399.999 kr.	378.214	130.858.640.074	10,77 %
400.000-499.999 kr.	248.121	111.165.155.388	9,15 %
500.000-599.999 kr.	237.841	130.891.275.371	10,77 %
600.000 kr. og derover	662.269	618.510.900.363	50,90 %
I alt	2.741.498	1.215.053.845.584	100 %

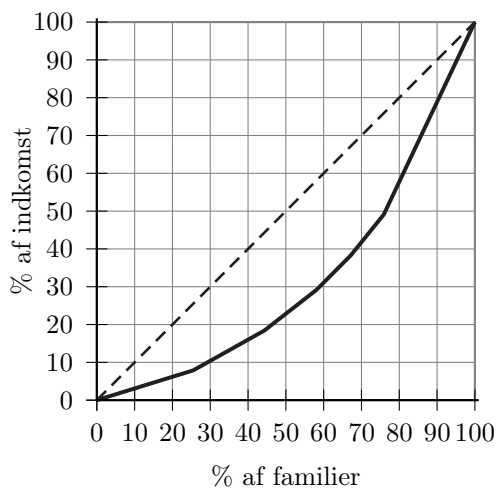
Kilde: Danmarks Statistik - statistikbanken.dk

Ud fra ovenstående tabel og tabellen for familierne kan følgende tabel konstrueres:

Indkomst	Summeret frekvens af familier	Summeret frekvens af indkomst
Under 200.000 kr.	25,5 %	7,88 %
200.000-299.999 kr.	44,3 %	18,40 %
300.000-399.999 kr.	58,1 %	29,17 %
400.000-499.999 kr.	67,2 %	38,32 %
500.000-599.999 kr.	75,9 %	49,10 %
600.000 kr. og derover	100,1 %	100,0 %

Kilde: Danmarks Statistik - statistikbanken.dk

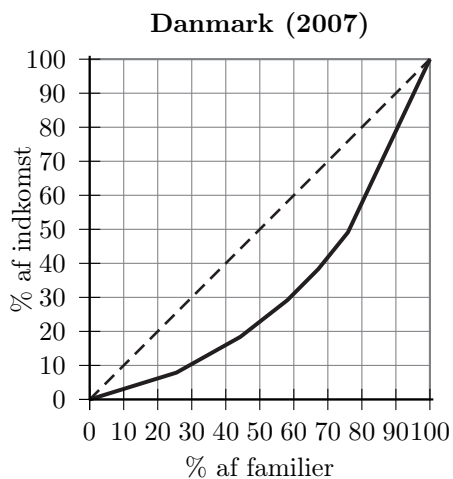
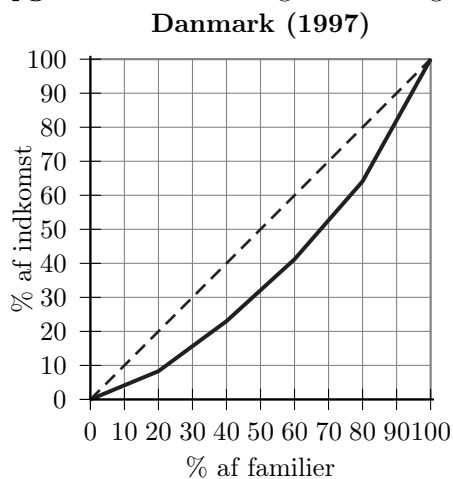
Med udgangspunkt i ovenstående tabel kan Lorenzdiagrammet for alle familier i Danmark tegnes.



Opgave 6.10 Tegn Lorenzdiagrammet for indkomsten for hhv. par og enlige.

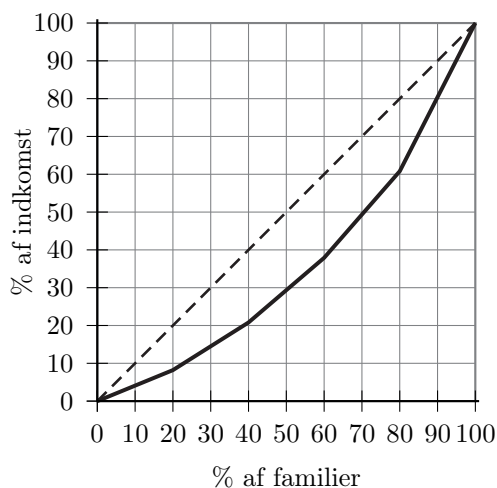
Opgave 6.11 Sammenlign Lorenzdiagrammet for hhv. alle familier, par og enlige og diskuter årsager til forskellene.

Opgave 6.12 Sammenlign Lorenzdiagrammet for 2007 med 1997.

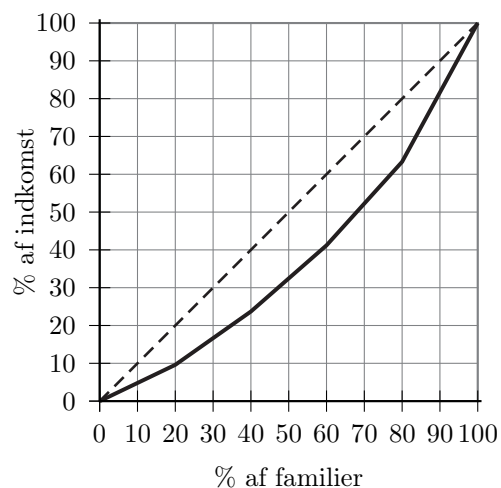


Opgave 6.13 Sammenlign Lorenzdiagrammerne fra følgende lande. *Kilde: The World Bank*

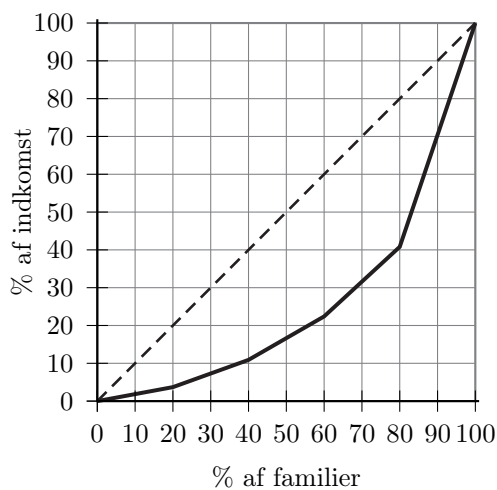
Albanien (2004)



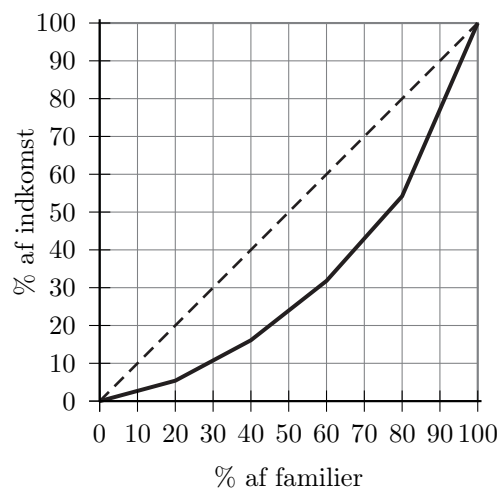
Finland (2000)



Chile (2003)

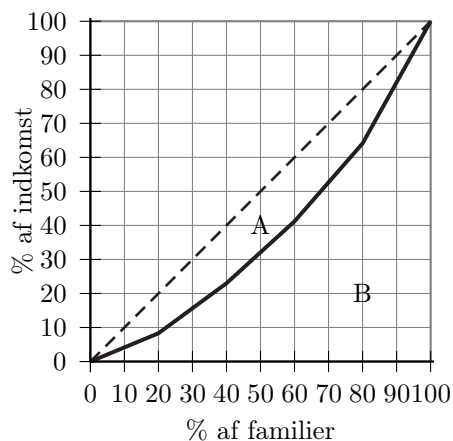


USA (2000)



6.5 Gini-koefficient

Gini-koefficienten G er defineret som forholdet mellem arealet mellem linien der beskriver lighed og Lorenzkurven (A) og arealet mellem linien der beskriver lighed og x -aksen. Arealet mellem Lorenzkurven og x -aksen kaldes B .



Formlen for Gini-koefficienten er altså

$$G = \frac{A}{A + B}$$

Da $A + B = 0.5$ er

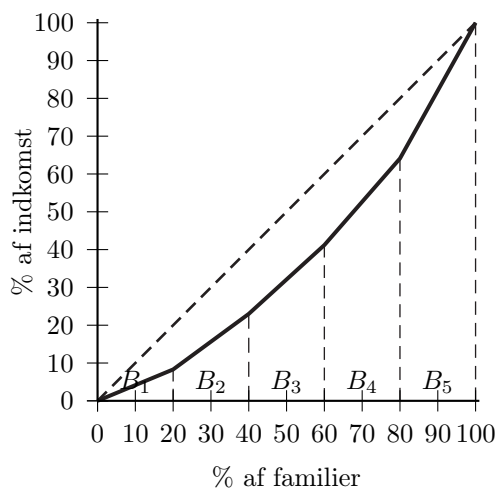
$$G = \frac{A}{0.5} = 2A$$

Da det er nemmest at udregne arealet B omskrives formelen idet $A = 0.5 - B$ så

$$G = 2(0.5 - B) = 1 - 2B$$

Eksempel 6.14 Arealet B beregnes ved at bruge tallene for % familier og % indkomst.

% familier	0	20	40	60	80	100
% indkomst	0	8,3	23	41,2	64,1	100



Arealet af B kan nu beregnes som summe af arealerne B_1 til B_5 .

$$B = 0,2 \cdot \frac{0,083 + 0}{2} + 0,2 \cdot \frac{0,23 + 0,083}{2} + 0,2 \cdot \frac{0,412 + 0,23}{2} + 0,2 \cdot \frac{0,641 + 0,412}{2} + 0,2 \cdot \frac{1 + 0,641}{2} = 0,3731$$

Nu kan Gini-koefficienten beregnes

$$G = 1 - 2 \cdot 0,3731 = 0,2538$$

Dette tal ganges ofte med 100. Gini-koefficienten for Danmark i 1997 var 25,4.

Opgave 6.15 Beregn Gini-koefficienten for familier i Danmark i 2007.

Indkomst	Summeret frekvens af familier	Summeret frekvens af indkomst
Under 200.000 kr.	25,5 %	7,88 %
200.000-299.999 kr.	44,3 %	18,40 %
300.000-399.999 kr.	58,1 %	29,17 %
400.000-499.999 kr.	67,2 %	38,32 %
500.000-599.999 kr.	75,9 %	49,10 %
600.000 kr. og derover	100,1 %	100,0 %

Kilde: Danmarks Statistik - statistikbanken.dk

7 Diskrete observationer

Nogle observationer er diskrete, det vil sige at de ikke bliver inddelt i intervaller. Et eksempel kunne være alder.

Alder	Antal
0 år	61.256
1 år	64.240
2 år	63.697
3 år	64.030
4 år	63.974
5 år	63.316
6 år	64.440
7 år	66.129
8 år	65.334
9 år	65.414
10 år	66.601
11 år	66.901
I alt	775.332

Kilde: Danmarks Statistik - statistikbanken.dk

Antallet af børn kan omregnes til procent af det samlede antal børn mellem 0 og 11 år.

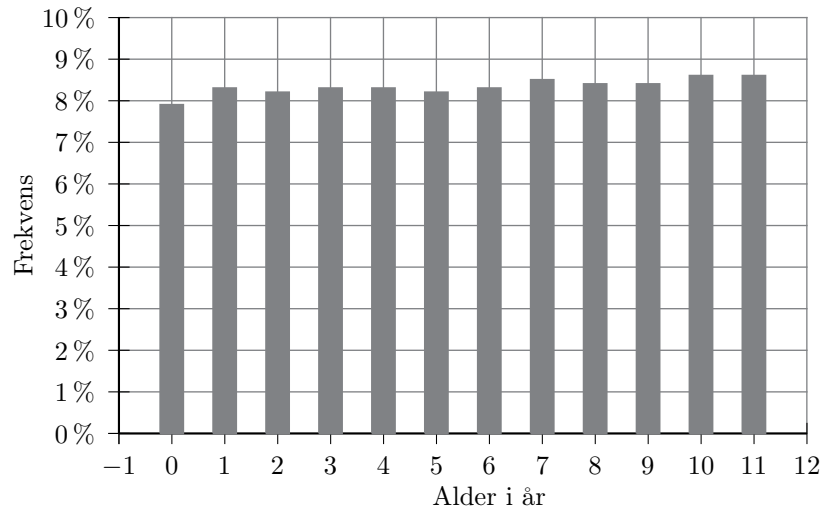
Eksempel 7.1 Ud af det samlede antal børn udgør de 5 årige $\frac{63.316}{775.332} \cdot 100\% = 8,2\%$

Alder	Antal	Frekvens
0 år	61.256	7,9 %
1 år	64.240	8,3 %
2 år	63.697	8,2 %
3 år	64.030	8,3 %
4 år	63.974	8,3 %
5 år	63.316	8,2 %
6 år	64.440	8,3 %
7 år	66.129	8,5 %
8 år	65.334	8,4 %
9 år	65.414	8,4 %
10 år	66.601	8,6 %
11 år	66.901	8,6 %
I alt	775.332	100 %

Kilde: Danmarks Statistik - statistikbanken.dk

7.1 Pindediagram

Pindediagrammet tegnes ved at angive observationsværdierne på x -aksen. På y -aksen angives frekvensen.



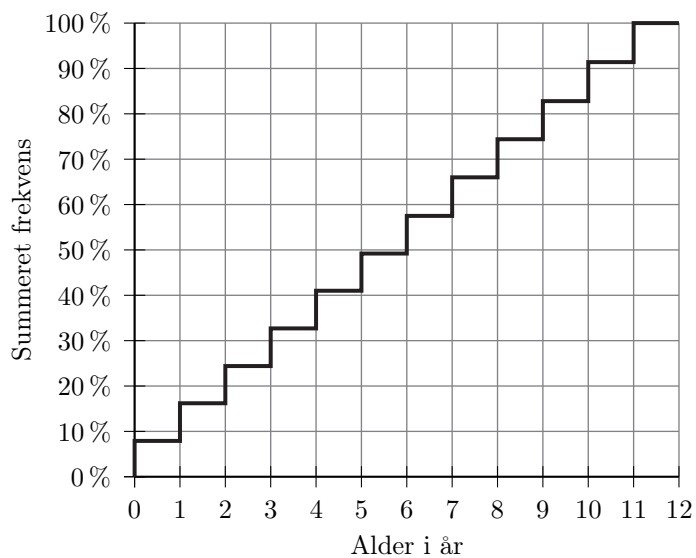
På pindediagrammet ses det, at antallet af børn er nogenlunde konstant til svagt faldende.

7.2 Trappediagram

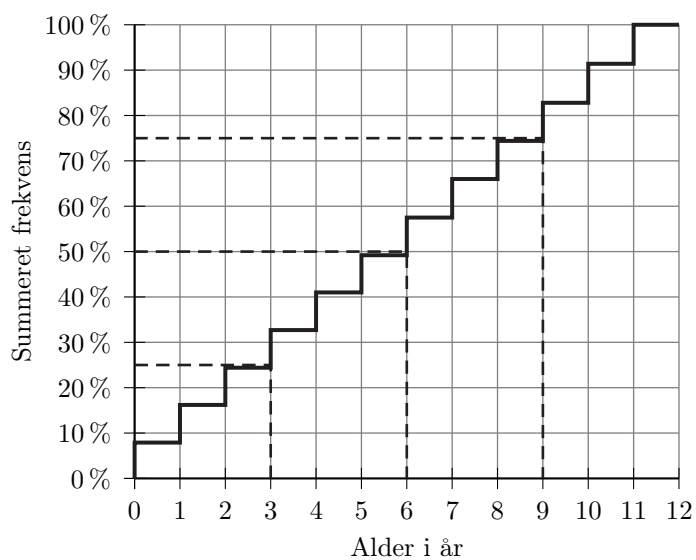
På trappediagrammet angives observationsværdierne på x -aksen og den sumerede frekvens på y -aksen.

Alder	Antal	Frekvens	Summeret frekvens
0 år	61.256	7,9 %	7,9 %
1 år	64.240	8,3 %	16,2 %
2 år	63.697	8,2 %	24,4 %
3 år	64.030	8,3 %	32,7 %
4 år	63.974	8,3 %	41,0 %
5 år	63.316	8,2 %	49,2 %
6 år	64.440	8,3 %	57,5 %
7 år	66.129	8,5 %	66,0 %
8 år	65.334	8,4 %	74,4 %
9 år	65.414	8,4 %	82,8 %
10 år	66.601	8,6 %	91,4 %
11 år	66.901	8,6 %	100,0 %
I alt	775.332	100 %	

Kilde: Danmarks Statistik - statistikbanken.dk



Kvartilerne aflæses på trappediagrammet ved hhv. 25 %, 50 % og 75 %.



Kvartil	Alder i år
Nedre kvartil	3 år
Median	6 år
Øvre kvartil	9 år

Opgave 7.2 Lav et trappediagram over antallet af børn i alderen 0 til 11 år, som bor hos deres enlig far eller mor.

Alder	Antal
0 år	3.482
1 år	5.347
2 år	6.735
3 år	7.768
4 år	8.733
5 år	9.607
6 år	10.474
7 år	11.237
8 år	11.374
9 år	11.778
10 år	12.223
11 år	12.700
I alt	111.458

Kilde: Danmarks Statistik - statistikbanken.dk

Opgave 7.3 Aflæs kvartilsættet for børn i alderen 0 til 11 år, som bor hos deres enlige far eller mor. Sammenlign dette med kvartilsættet for alle børn i alderen 0 til 11 år.