

# **Praktisk Statistik til Mikroøkonomi og Makroøkonomi**

**Til brug sammen med  
Mikroøkonomi Teori og beskrivelse  
og  
Makroøkonomi Teori og beskrivelse**

**LIMEDESIGN**

Formålet med dette appendiks er at præsentere centrale redskaber i den praktiske statistik, dvs. redskaber til at illustrere centrale egenskaber ved et datamateriale. Afnittet behandler forskellige metoder til beregning og vurdering af niveaumål og spredningsmål. Niveaumålene har til formål at præsentere gennemsnit eller "midt-værdier" i et datamateriale. Spredningsmålene søger at opstille tal for spredningen eller ensartetheden i et talmateriale. Endvidere præsenteres metoder til beskrivelse af vækst og udvikling. Særlig fokus rettes mod konstruktionen af prisindeks. Undervejs gives vejledning i tabelmæssig og grafisk præsentation af datamaterialer. I fremstillingen benyttes næsten udelukkende kvantitative datamaterialer, dvs. datamaterialer, hvor data kan angives ved tal og hvor regneoperationer meningsfuldt kan anvendes. Mange af de viste metoder lader sig i praksis nemmest gennemføre ved brug af lommeregner eller regnearksprogrammer. Det falder dog uden for denne fremstillings rammer at give egentlig instruktion i lommeregneranvendelse og anvendelse af regneark.

## Niveaumål

### Middeltal for observationssæt

Statistik beskæftiger sig med systematisk indsamling og bearbejdning af datamaterialer. De registrerede data kaldes observationer og en samling af sammenhørende observationer kaldes et observationssæt.

Et eksempel kunne være de karakterer, som 16 studerende opnåede ved en eksamen:

00 7 12 00 -3 4 7 7 02 00 10 4 7 4 4 7

Beregningen af deltagernes karaktergennemsnit sker ved at dividere summen af karaktererne med antallet af deltagere:

$$\frac{00 + 7 + 12 + 00 + (-3) + 4 + 7 + 7 + 02 + 00 + 10 + 7 + 4 + 4 + 4 + 7}{16} = \frac{72}{16} = 4,5$$

Et gennemsnit beregnet på denne måde kaldes også et aritmetisk gennemsnit eller et middeltal:

Gennemsnittet eller middeltallet ( $\bar{x}$ ) for et observationssæt beregnes ved at dividere summen af observationsværdierne ( $x_1 + x_2 + \dots + x_n = \sum_{i=1}^n x_i$ ) med antallet af observationer ( $n$ ):

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

I eksemplet ovenfor er observationsværdierne ( $x$ -værdierne) altså karakterværdierne ( $x_1 = 00$ ,  $x_2 = 7$ , ...,  $x_n = x_{16} = 7$ ), og  $n$  er antallet af studerende, dvs. 16.

*Øvelse 1: Følgende talmateriale viser det antal minutter, som de studerende på et hold havde brugt til forberedelse før en lektion i mikroøkonomi:*

10 49 0 60 13 40 45 0 70 30 55 64 45

*Beregn det gennemsnitlige antal minutter brugt på forberedelse.*

## Grupperede observationssæt. Fordelinger

Ofte vil man møde datamateriale, som er grupperet efter et eller flere kendetegn. Der er derfor behov for metoder til at beregne gennemsnit med udgangspunkt i grupperede observationssæt.

Antag fx at vi ønsker at beregne karaktergennemsnit for et meget stort antal deltagere, som er grupperet efter den opnåede karakter. De grupperede observationer fremgår af følgende tabel:

**Tabel 1 Studerende fordelt efter eksamenskarakter**

Observations- værdi nr. $i$	Karakter $x_i$	Antal i gruppen (hyppighed) $h_i$
1	-3	1
2	00	29
3	02	33
4	4	44
5	7	68
6	10	36
7	12	7

I Tabel 1 er observationssættet grupperet således, at vi direkte kan aflæse, hvor hyppigt de enkelte karakterer forekommer. Et sådant observationssæt kaldes også en fordeling<sup>1</sup>.

En fordeling består af et antal kendetegn (her forskellige karakterer) med tilhørende hyppigheder, der angiver, hvor mange gange i datamaterialet, det pågældende kendetegn forekommer.

Det samlede karaktergennemsnit beregnes stadig som den samlede karaktersum divideret med det samlede antal deltagere. I tabel 2 er vist karaktersummen i de enkel-

<sup>1</sup> Man taler her om en empirisk fordeling for at understrege, at fordelingen er baseret på indsamlede oplysninger, dvs. en empirisk undersøgelse.

te karaktergrupper. Eksempelvis har de 44 deltagere, som har opnået karakteren 4, en samlet karaktersum på  $4 \cdot 44 = 176$ .

**Tabel 2 Beregning af karaktersum**

Observations- værdi nr. j	Karakter $x_j$	Antal i gruppen (hyppighed) $h_j$	Karaktersum i gruppen $x_j \cdot h_j$
1	-3	1	-3
2	00	29	0
3	02	33	66
4	4	44	176
5	7	68	476
6	10	36	360
7	12	7	84
I alt	-	218	1159

Den samlede karaktersum får vi ved at lægge alle tallene i den sidste kolonne sammen:

$$\text{Samlet karaktersum} = \sum_{j=1}^7 x_j \cdot h_j = -3 + 0 + 66 + \dots + 360 + 14 = 1159$$

Det samlede karaktergennemsnit vil derfor være:

$$\bar{x} = \frac{\sum_{j=1}^7 x_j \cdot h_j}{n} = \frac{\sum_{j=1}^7 x_j \cdot h_j}{\sum_{j=1}^7 h_j} = \frac{1159}{218} = 5,3$$

Undertiden vil man med fordel kunne beregne middeltal eller gennemsnit ved at lave fordelingen om til en frekvensfordeling. I eksemplet ovenfor sker det ved at omregne hyppighederne til frekvenser, dvs. andele af det samlede antal deltagere, som opnåede en bestemt karakter. Da der er 44 ud af 218 deltagere, der opnåede karakteren 4, vil frekvensen af 4-taller være:

$$f_4 = \frac{44}{218} = 0,201835$$

Frekvensen af 4-taller er da 0,201835. Det betyder, at godt 20 pct. af deltagerne fik karakteren 4.

**Tabel 3 Hyppighedsfordeling og frekvensfordeling**

Observations- værdi nr. i	Karakter $x_i$	Antal i gruppen (hyppighed) $h_i$	Andel i gruppen (frekvens) $f_i$	Procentandele
1	-3	1	0,004587	0,46
2	00	29	0,133028	13,30
3	02	33	0,151376	15,14
4	4	44	0,201835	20,18
5	7	68	0,311927	31,20
6	10	36	0,165138	16,51
7	12	7	0,032110	3,21
I alt	-	218	1	100,00

Med udgangspunkt i frekvensfordelingen kan middeltallet eller gennemsnittet beregnes ved at gange de enkelte karakterværdier med deres frekvens. Således kan 4-tallernes "bidrag" til gennemsnittet beregnes som  $4 \cdot 0,201835 = 0,80734$ . Det samlede gennemsnit bliver:

$$\bar{x} = \sum_{j=1}^7 x_j \cdot f_j = (-3) \cdot 0,004587 + 0 \cdot 0,133028 + \dots + 12 \cdot 0,0032110 = 5,3$$

Gennemsnittet eller middeltallet ( $\bar{x}$ ) for en fordeling, der er opdelt i k grupper kan beregnes som:

$$\bar{x} = \frac{x_1 \cdot h_1 + x_2 \cdot h_2 + \dots + x_k \cdot h_k}{n} = \frac{\sum_{j=1}^k x_j \cdot h_j}{n} = \frac{\sum_{j=1}^k x_j \cdot h_j}{\sum_{j=1}^7 h_j}$$

hvor

$h_j$  er hyppigheden for observationsværdi  $x_j$

$k$  er antal forskellige observationsværdier (antal grupper)

$n$  er antal observationer i alt.

eller som:

$$\bar{x} = x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_k \cdot f_k = \sum_{j=1}^k x_j \cdot f_j$$

hvor  $f_j$  er frekvensen for observationsværdi  $x_j$

Øvelse 2: I en lille provinsby har en trafikomlægning givet anledning til en del trafikulykker med personskader. I 2006 blev der for hver ulykke optalt følgende antal personer med skader:

1 2 3 3 4 2 6 1 3 1 2 2 3 1 2 2 1 3 4 6 6 4 1 3 7 8 1 3 3 2  
2 2 2 1 2 2 7 6 6 5 3 6 4 3 9 1 6 3 5 7 1 2 7 3 4

- Opstil en hensigtsmæssig fordeling, der giver overblik over ulykkernes omfang.
- Udregn fordelings middeltal.

Øvelse 3: I tabellen nedenfor er vist fordelingen af husstandene i en gade efter antal børn.

**Tabel 4 50 husstande fordelt efter antal børn**

Antal børn i husstanden	Antal husstande
0	5
1	15
2	11
3	8
4	6
5	5

- Beregn frekvensfordelingen ud fra ovenstående tabel.
- Udregn det gennemsnitlige antal børn pr. husstand ud fra frekvensfordelingen.

## Typetal

Betragt følgende tabel, der viser fordelingen af sygedage på et skoleår blandt eleverne i en skoleklasse.

**Tabel 5 30 elever fordelt efter antal sygedage**

Antal sygedage i skoleåret	Antal elever
0	3
1	9
2	8
3	5
4	2
5	2
50	1

Ud fra tabellen kan vi beregne middeltallet, dvs. det gennemsnitlige antal sygedage til 3,6.

I tabellen kan vi se, at der er 5 elever, der har et fravær over gennemsnittet, mens de resterende 25 elever har et fravær under gennemsnittet. Man kan derfor kritisere gennemsnittet for ikke at udtrykke det "normale" sygefravær i klassen. Årsagen til dette er, at fordelingen er "skæv" i den forstand, at eleven med 50 sygedage har en meget stor indflydelse på middeltallet.

Vil vi have et udtryk for det "normale" sygefravær, som ikke i så høj grad er præget af enkelte ekstreme observationer, kan vi opgøre typetallet (modus).

Typetallet i en fordeling er den (de) observationsværdi(er), der har den største hyppighed eller den største frekvens.

I tabellen ovenfor er et sygefravær på 1 dag den oftest forekommende observationsværdi. Derfor er typetallet 1 sygedag.

*Øvelse 4: Efter sæsonafslutningen har en fodboldklub opgjort antallet af advarsler, som hver af deres førsteholdsspillere er blevet tildelt i sæsonens kampe. Resultatet er følgende:*

0 2 1 2 0 0 0 0 12 0 3 0 1 1 2 1 1 4 8 2 1 2 1 3 4 2 1 0 1 0

a. Opstil en hyppighedsfordeling til beskrivelse af talmaterialet.

b. Udregn fordelingens middeltal og typetal.

*Øvelse 5: Bestem typetallet for talmaterialet i Øvelse 3.*

## Median

Et andet mål for det gennemsnitlige niveau i et observationssæt er medianen. Medianen opgøres som den midterste observationsværdi, når observationerne opstilles i størrelsesorden. Det vil sige den værdi, der opdeler talmaterialet i to lige store dele henholdsvis over og under værdien.

Betragt fx følgende tal, der viser legemsvægten i kg for 13 drenge på et fodboldhold:

56 61 57 49 58 58 51 54 48 62 51 50 61

Ordnet efter vægt ser talmaterialet sådan ud:

48 49 50 51 51 54 56 57 58 58 61 61 62

Medianen vil her være 56 kg, idet der er 6 drenge, der vejer mindre end 56 kg, og 6 drenge, der vejer mere end 56 kg.

Hvis antallet af observationer i et talmateriale er et lige tal, er der ikke en enkelt observation, som kan siges at være den midterste observation. Her kan medianen findes ved at tage gennemsnittet af de to midterste observationers værdier.

Vender vi tilbage til eksemplet med sygefraværet, er bestemmelsen af medianen ikke helt så enkel. Vi kan ikke her finde et antal sygedage, der opdeler materialet på en måde, så der er lige mange elever med et lavere og med et højere sygefravær. For at finde en entydig metode til bestemmelse af medianen, ser vi på frekvensfordelingen:

**Tabel 6 30 elever fordelt efter antal sygedage**

Antal sygedage i skoleåret	Antal elever	Frekvens	Summeret frekvens	Summeret procentandel
0	3	0,100	0,100	10,0
1	9	0,300	0,400	40,0
2	8	0,267	0,667	66,7
3	5	0,167	0,833	83,3
4	2	0,067	0,900	90,0
5	2	0,067	0,967	96,7
50	1	0,033	1,000	100,0
I alt	30	1	-	-

I tabel 6 er frekvensfordelingen beregnet ved at dividere antallet af elever i hver gruppe med 30.

I tabellens sidste kolonne er vist den summerede frekvens (kumulerede frekvens). Den summerede frekvens fås ved at lægge frekvenserne sammen:

Den summerede frekvens for 1 sygedag er  $0,100 + 0,300 = 0,400$  og det betyder at 40 pct. af eleverne havde 0 eller 1 sygedag. Den summerede frekvens for 2 sygedage er  $0,100 + 0,300 + 0,267 = 0,667$ . 66,7 pct. af eleverne havde altså 2 eller færre sygedage.

Medianen er et udtryk for den midterste observation, når observationerne opstilles i størrelsesorden. Medianen i ovenstående fordeling er derfor 2 sygedage.

Vi kan aflæse medianen i kolonnen med de summerede frekvenser ved at finde den observationsværdi  $x_j$ , hvor den summerede frekvens første gang når op på eller overstiger 0,5.

Medianen bestemmes som den mindste observationsværdi, der har en summeret frekvens på mindst 0,5.

*Øvelse 6: Bestem median og typetal for talmaterialet i Øvelse 2.*

*Øvelse 7: Bestem medianen for talmaterialet i Øvelse 3.*

*Øvelse 8: Bestem medianen for talmaterialet i Øvelse 4.*



## Intervalgrupperede fordelinger

I mange talmaterialer er antallet af forskellige observationsværdier meget stort. Indsamles eksempelvis oplysninger om befolkningens skattepligtige indkomster, vil man få flere hundredetusinde forskellige observationsværdier. Skal et sådant talmateriale præsenteres overskueligt, er der behov for en gruppering af observationsværdierne i intervaller. Fastlæggelsen af, hvilke intervaller, som er hensigtsmæssige, må bero på en vurdering af det konkrete talmateriale. Generelt må intervallerne opfylde følgende krav:

- *Intervallerne er entydigt afgrænsede.* Der må ikke herske tvivl om, i hvilket interval en observationsværdi skal placeres.
- *Intervallerne dækker hele talmaterialet.* Det skal være muligt at placere en hvilken som helst tænkelig observationsværdi i et af intervallerne.
- *Et overskueligt antal intervaller.* Overskueligheden kan nemt forsvinde, hvis man vælger for mange intervaller. Omvendt må man have et tilstrækkeligt antal intervaller til at fordelingsstrukturen tydeligt fremgår. Som en praktisk tommelfingerregel kan anbefales, at man stræber mod et antal intervaller på mellem 5 og 15.

Følgende er et eksempel på en konkret intervalgrupperet fordeling.

**Tabel 7 Fordeling af medarbejdere efter ugentligt antal overtimer**

Interval nr. j	Antal ugentlige overtimer (interval) $]x_{j-1}; x_j]$	Antal personer i intervallet (hyppighed) $h_j$	Intervalmidtpunkt $m_j = \frac{x_{j-1} + x_j}{2}$	Samlet overtimetotal i intervallet $m_j \cdot h_j$
1	$[0; 2,5]$	31	1,25	38,75
2	$]2,5; 5,0]$	48	3,75	180,00
3	$]5,0; 7,5]$	26	6,25	162,50
4	$]7,5; 10,0]$	14	8,75	122,50
5	$]10,0; 12,5]$	8	11,25	90,00
6	$]12,5; 15,0]$	2	13,75	27,50
7	$]15,0; 17,5]$	1	16,25	16,25

I tabellens anden kolonne er intervallerne specificeret. I første interval placeres medarbejdere med et ugentligt overtimetal på 2,5 timer eller lavere. I andet interval placeres medarbejdere med et overtimetal, som er større end 2,5, men ikke over 5,0, osv. Bemærk at den entydige intervalinddeling er sikret ved at lade intervallerne være åbne ved den nedre grænse og lukkede<sup>2</sup> ved den øvre grænse.

<sup>2</sup> At et interval er åbent ved nedre grænse betyder, at den nedre grænse ikke er med. Det symboliseres med en intervalklamme, hvis grene vender væk fra tallet. At et interval er lukket ved øvre grænse betyder, at den øvre grænse er med. Det symboliseres med en intervalklamme, hvis grene vender hen mod tallet.

Tredje kolonne i tabel 7 viser hyppighederne i de enkelte intervaller. Fx ses, at der er 48 medarbejdere, der har et ugentligt overtimetal, som er højere end 2,5, men ikke over 5,0.

Beregningen af middeltallet – dvs. det gennemsnitlige overtimetal for virksomhedens medarbejdere – foregår efter samme metode som tidligere. Det vil sige ved at opgøre det samlede antal overtimer og dividere med antallet af medarbejdere i virksomheden.

Vi støder dog her på det problem, at vi ikke kender medarbejdernes præcise overtimetal. Vi ved, hvor mange der er placeret i de forskellige intervaller, men vi ved ikke, hvor i intervallet de er placeret.

Hvis det kan antages – og det vil ofte være tilfældet – at de enkelte medarbejders timetal er jævnt fordelt over intervallet, kan problemet klares på følgende måde. Det gennemsnitlige overtimetal inden for hvert interval vil da være intervalmidtpunktet. I beregningerne kan vi lade som om, at alle personer i de enkelte intervaller har et overtimetal svarende til intervalmidtpunktet. I tabellens sidste kolonne er vist det samlede overtimetal i hvert interval beregnet som intervalmidtpunktet ( $m_j$ ) ganget med antal personer i det pågældende interval ( $h_j$ ).

Virksomhedens samlede overtimetal får vi da ved at summere alle tallene i denne kolonne:

$$\text{Samlet overtimetal} = \sum_{j=1}^7 m_j \cdot h_j = 38,75 + 180,00 + \dots + 27,50 + 16,25 = 637,50$$

Det samlede antal medarbejdere finder vi i tabellens tredje kolonne ved at lægge antallet af personer i de enkelte intervaller sammen:

$$\text{Samlet antal medarbejdere} = n = \sum_{j=1}^7 h_j = 31 + 48 + \dots + 2 + 1 = 130$$

Det samlede overtimegennemsnit vil derfor være

$$\bar{x} = \frac{\sum_{j=1}^7 m_j \cdot h_j}{n} = \frac{\sum_{j=1}^7 m_j \cdot h_j}{\sum_{j=1}^7 h_j} = \frac{637,5}{130} = 4,9$$

Som tilfældet var med de ikke-intervalgrupperede fordelinger, kan middeltallet også beregnes ud fra frekvensfordelingen.

Frekvensen for de enkelte intervaller findes ved at dividere intervallets hyppighed (antal medarbejdere i det pågældende interval) med det samlede antal observationer (det samlede antal medarbejdere).

Frekvenser og summerede frekvenser er vist i tabel 8.

**Tabel 8** Intervalgrupperet frekvensfordeling

Antal ugentlige overtimer (interval) $]x_{j-1}; x_j]$	Antal personer i intervallet (hyppighed) $h_j$	Intervalmidtpunkt $m_j = \frac{x_{j-1} + x_j}{2}$	Interval- frekvens $f_j$	Summeret frekvens $F_j$
[0;2,5]	31	1,25	0,2385	0,2385
]2,5;5,0]	48	3,75	0,3692	0,6077
]5,0;7,5]	26	6,25	0,2000	0,8077
]7,5;10,0]	14	8,75	0,1077	0,9154
]10,0;12,5]	8	11,25	0,0615	0,9770
]12,5;15,0]	2	13,75	0,0154	0,9923
]15,0;17,5]	1	16,25	0,0077	1,0000

Herefter kan gennemsnittet beregnes ved at gange de enkelte intervalmidtpunkter med deres frekvens. Således er første intervals "bidrag" til gennemsnittet  $1,25 \cdot 0,2385 = 0,2981$ . Gennemsnittet eller middeltallet bliver:

$$\bar{x} = \sum_{j=1}^7 m_j \cdot f_j = 1,25 \cdot 0,2385 + 3,75 \cdot 0,3692 + \dots + 16,25 \cdot 0,0077 = 4,9$$

Gennemsnittet eller middeltallet ( $\bar{x}$ ) for en intervalgrupperet fordeling kan beregnes som:

$$\bar{x} = \frac{m_1 \cdot h_1 + m_2 \cdot h_2 + \dots + m_k \cdot h_k}{n} = \frac{\sum_{j=1}^k m_j \cdot h_j}{n} = \frac{\sum_{j=1}^k m_j \cdot h_j}{\sum_{j=1}^7 h_j}$$

hvor

$h_j$  er hyppigheden for interval nr. j

k er antal intervaller

n er antal observationer i alt

eller som:

$$\bar{x} = m_1 \cdot f_1 + m_2 \cdot f_2 + \dots + m_k \cdot f_k = \sum_{j=1}^k m_j \cdot f_j$$

hvor  $f_j$  er frekvensen for interval nr. j

$m_j$  er intervalmidtpunktet i interval nr. j.  $m_j = \frac{x_{j-1} + x_j}{2}$ , hvor  $x_{j-1}$  er intervallets nedre grænse og  $x_j$  intervallets øvre grænse.

Øvelse 9: En beboerforening har i forbindelse med overvejelser om igangsætning af fritidsaktiviteter for børn indsamlet oplysninger om aldersfordelingen for de 100 børn, der bor i foreningens område. Resultatet er følgende tabel:

**Tabel 9 Fordeling af børn efter alder**

Alder	Antal børn
0-3 år	15
4-7 år	25
8-11 år	35
12-15 år	25

Beregn middeltallet, dvs. den gennemsnitlige alder for børnene i området. Overvej nøje, hvordan intervalmidtpunkterne bør fastlægges.

Øvelse 10: I Statistisk Tiårsoversigt 2006 findes følgende fordeling af de personer, som i løbet af 2005 blev registreret som ledige. De ledige er fordelt efter ledighedsperiodens varighed.

**Tabel 10 Fordeling af ledige efter ledighedsperiodens varighed**

Ledighedsperiodens varighed	Antal tusinde personer
Under 74 dage	267
74 – 146 dage	118
147 – 219 dage	75
220 – 292 dage	46
293 dage og derover	38

Opstil frekvensfordelingen og beregn herudfra middeltallet. Forklar, hvad middeltallet udtrykker.

Øvelse 11: Ved en undersøgelse af størrelsesfordelingen af 48 medlemsvirksomheder i en brancheforening indsamles følgende oplysninger om antal beskæftigede i de 48 virksomheder:

12	23	44	97	4	34	56	54	34	23	45	78
88	3	12	16	82	34	50	22	5	9	24	40
60	62	34	12	19	91	9	48	34	12	16	35
24	34	56	16	7	4	5	7	21	24	28	2

- Foretag en hensigtsmæssig intervalgruppering af observationerne.
- Beregn middeltallet ud fra intervalgrupperingen.
- Beregn middeltallet direkte ud fra grundmaterialet.
- Forklar forskellen mellem resultaterne i b. og c.

I stedet for typetallet (modus) for en intervalgrupperet fordeling bestemmes typeintervallet, dvs. det interval, hvor der er flest observationer. I ovenstående fordeling af antal overtimer vil typeintervallet således være  $]2,5 ; 5,0]$ . Det betyder, at det er mest "almindeligt", at det ugentlige antal overtimer ligger mellem 2,5 og 5,0.

Typeintervallet i en intervalgrupperet fordeling med lige store intervaller er den (de) observationsinterval(ler), der har den største hyppighed eller den største frekvens.

*Øvelse 12: Bestem typeintervallerne i Øvelse 9 og Øvelse 10.*

Medianen for en intervalgrupperet fordeling er ligesom i en ugrupperet fordeling den observationsværdi, der opdeler talmaterialet i to lige store dele henholdsvis over og under værdien. Vi har tidligere brugt den summerede frekvens som et redskab til at finde medianen, nemlig som den mindste observationsværdi, der har en summeret frekvens på mindst 0,5.

Forsøger vi at bruge denne metode på fordelingen af medarbejdere efter ugentligt overtimetal i tabel 8, kan vi se, at medianen må befinde sig i intervallet fra 2,5 til 5,0 overtimer. Hvis vi vil bestemme medianen mere præcist, kan vi bruge vores forudsætning om, at observationerne er jævnt fordelt i de enkelte intervaller. I tabel 8 kan vi se, at den summerede frekvens stiger fra 0,2385 til 0,6077, når intervallet fra 2,5 til 5,0 gennemløbes. Er observationerne jævnt fordelt, vil frekvensen 0,5 opnås, når en andel på  $\frac{0,5 - 0,2385}{0,6077 - 0,2385}$  er gennemløbet. Derfor kan medianen bestemmes som:

$$2,5 + \frac{0,5 - 0,2385}{0,6077 - 0,2385} \cdot (5,0 - 2,5) = 4,27.$$

Medianen bestemmes som den mindste observationsværdi, der har en summeret frekvens på mindst 0,5. I en intervalgrupperet fordeling findes medianen ved først at identificere det interval, hvor den summerede frekvens passerer 0,5. Kaldes den summerede frekvens ved intervallets start for  $F_{\text{START}}$  og den summerede frekvens ved intervallets afslutning for  $F_{\text{SLUT}}$  kan medianen bestemmes som:

$$\text{Median} = \text{Nedre intervalgrænse} + \frac{0,5 - F_{\text{START}}}{F_{\text{SLUT}} - F_{\text{START}}} \cdot \text{Intervalbredden}$$

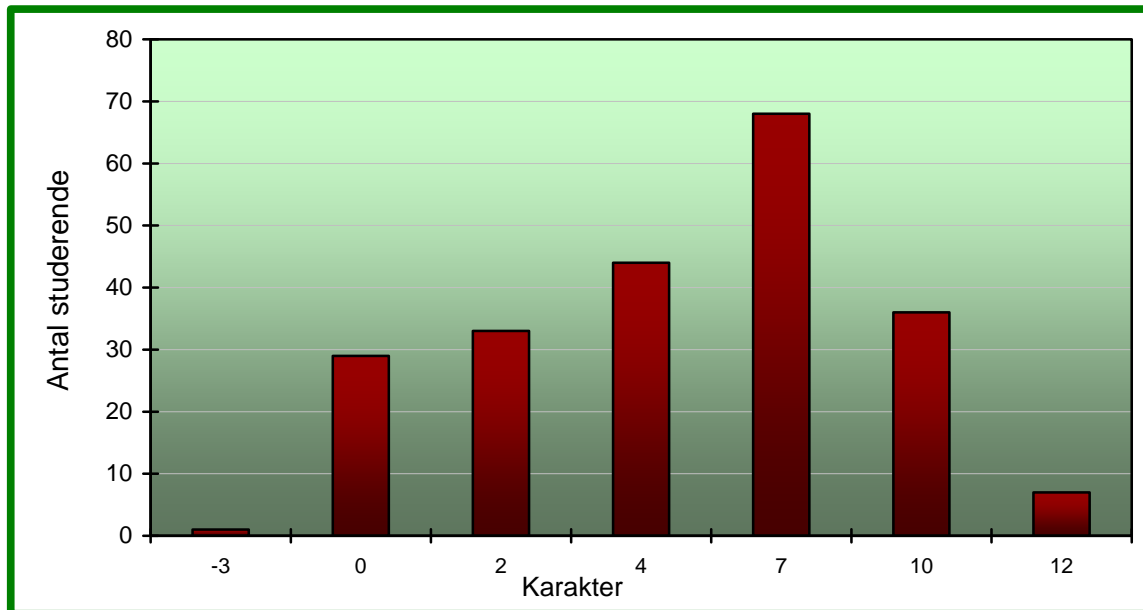
*Øvelse 13: Bestem medianen i Øvelse 9 og Øvelse 10.*

## Grafisk afbildning af fordelinger

Ofte kan det være en god ide at illustrere en fordeling i et diagram. Almindeligvis anvendes et søjlediagram, hvor søjlernes højde markerer hyppigheden (eller eventuelt frekvensen) for hver observationsværdi eller observationsinterval.

I figur 1 er fordelingen fra tabel 1 taget som eksempel.

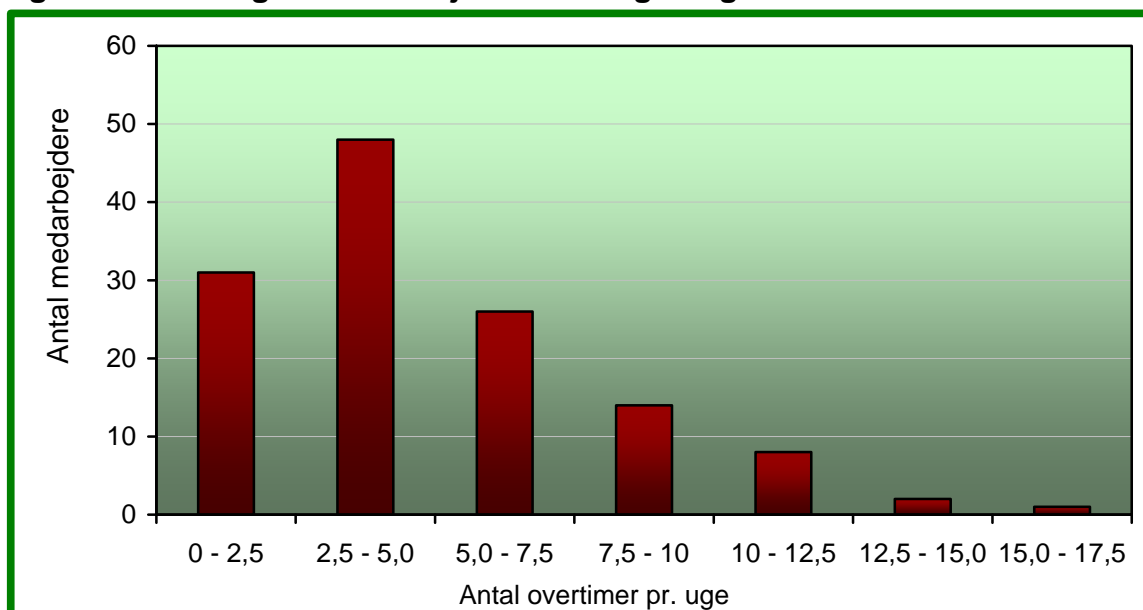
**Figur 1 Fordeling af studerende efter eksamenskarakter**



Kilde: Tabel 1

Figur 2 viser et eksempel på søjlediagram for en intervalgrupperet fordeling med udgangspunkt i fordelingen fra tabel 7.

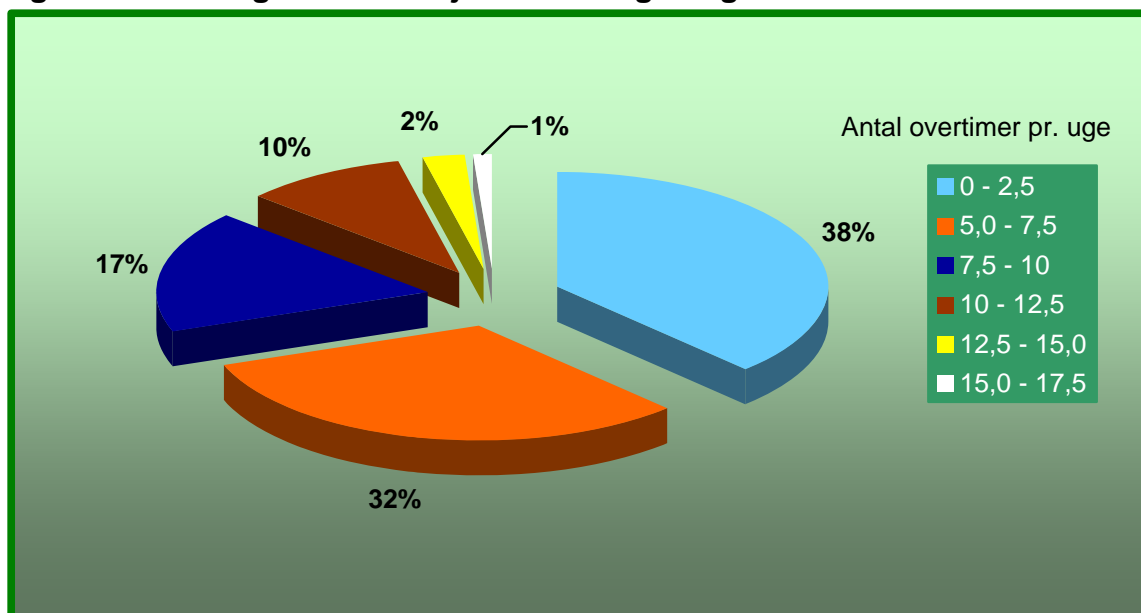
**Figur 2 Fordeling af medarbejdere efter ugentligt antal overtimer**



Kilde: Tabel 7

Hvis det specielt har interesse at præsentere den relative eller procentvise fordeling af observationsværdier, kan et cirkeldiagram som vist i figur 3 være at foretrække.

**Figur 3 Fordeling af medarbejdere efter ugentligt antal overtimer**



Kilde: Tabel 7

## Spredningsmål

Meget ofte er det relevant at beregne tal, der siger noget om spredningen eller variationen af observationsværdierne i et talmateriale. Som for niveaumålene er der flere forskellige mulige angrebsvinkler, når man skal beregne spredningstal.

Det mest simple udtryk for spredningen i et talmateriale er at finde den mindste og den største observationsværdi. I starten af dette appendiks så vi på de karakterer, som 16 studerende opnåede ved en eksamen:

00 7 12 00 -3 4 7 7 02 00 10 4 7 4 4 7

I dette talmateriale er 00 den mindste observationsværdi og 12 den højeste talværdi. Forskellen mellem største og mindste observationsværdi i et talmateriale kaldes variationsbredden. I vores talmateriale er variationsbredden  $12 - 00 = 12$ .

Variationsbredden i et talmateriale er udtryk for den maksimale spredning i et talmateriale.

Variationsbredde = Største observationsværdi – mindste observationsværdi

Problemet med variationsbredden er, at beregningen udelukkende baseres på de to mest ekstreme observationsværdier i et talmateriale. Variationsbredden siger derfor ikke nødvendigvis noget om den almindelige spredning i talmaterialet.

Et andet spredningsmål, hvor alle observationsværdier inddrages i beregningen er standardafvigelsen ( $s$ ), som er et skøn over observationsværdiernes gennemsnitlige afvigelse fra middeltallet. Beregningen sker ved hjælp af formlen

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Størrelsen under kvadratrodstegnet kaldes variansen ( $s^2$ ) og er et skøn over gennemsnittet af observationernes kvadrerede afvigelser fra middeltallet.

Lad os bruge talmaterialet med eksamenskaraktererne som eksempel på beregningen af varians og standardafvigelse. Tidligere har vi beregnet middeltallet i talmaterialet til 4,5.

Først vises beregningen af variansen:

$$s^2 = \frac{(00 - 4,5)^2 + (7 - 4,5)^2 + (12 - 4,5)^2 + \dots + (7 - 4,5)^2}{16 - 1} = \frac{242}{15} = 16,133$$

Variansen kan ikke give nogen praktisk fortolkning. Variansberegningens rolle i beregningen er at kvadrere alle afvigelser fra middeltallet, så afvigelserne får samme fortegn. Ellers ville positive og negative afvigelser fra middeltallet udligne hinanden.

Herefter beregnes standardafvigelsen ved at tage kvadratroden af variansen:

$$s = \sqrt{s^2} = \sqrt{16,133} = 4,017$$

Den gennemsnitlige afvigelse fra middelkarakteren på 4,5 er derfor ca. 4 karakterpoint.

Standardafvigelsen ( $s$ ) for et observationssæt beregnes som observationsværdiernes gennemsnitlige afvigelse fra middeltallet:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

$n$  er antal observationer

For grupperede fordelinger sker beregningen ved hjælp af de enkelte observationsgruppers eller observationsintervallers hyppigheder:



Standardafvigelsen ( $s$ ) for en grupperet fordeling beregnes som observationsværdiernes gennemsnitlige afvigelse fra middeltallet:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 \cdot h_1 + (x_2 - \bar{x})^2 \cdot h_2 + \dots + (x_k - \bar{x})^2 \cdot h_k}{n - 1}} = \sqrt{\frac{\sum_{j=1}^k (x_j - \bar{x})^2 \cdot h_j}{n - 1}}$$

$n$  er samlet antal observationer

$k$  er antal forskellige observationsværdier (antal grupper)

I en intervalgrupperet fordeling bruges intervalmidtpunkterne  $m_j$  i stedet for  $x_j$ .

Beregningen af standardafvigelse for en grupperet fordeling kan illustreres med fordelingen af overtimer i tabel 7. Middeltallet for det ugentlige antal overtimer er tidligere beregnet til 4,9.

**Tabel 11 Fordeling af medarbejdere efter ugentligt antal overtimer**

Interval nr. $j$	Interval-midtpunkt $m_j$	Antal personer $h_j$	Afgivelse fra middeltal $m_j - \bar{x}$	Kvadrat på afvigelse $(m_j - \bar{x})^2$	Produkt $(m_j - \bar{x})^2 \cdot h_j$
1	1,25	31	-3,65	13,3225	412,9975
2	3,75	48	-1,15	1,3225	63,4800
3	6,25	26	1,35	1,8225	47,3850
4	8,75	14	3,85	14,8225	207,5150
5	11,25	8	6,35	40,3225	322,5800
6	13,75	2	8,85	78,3225	156,6450
7	16,25	1	11,35	128,8225	128,8225
Sum		$n = 130$	-	-	1339,4250

Variansen er derfor  $s^2 = \frac{1339,4250}{130 - 1} = 10,383$

og standardafvigelsen er  $s = \sqrt{s^2} = \sqrt{10,383} = 3,222$

Vores beregninger af middeltal og standardafvigelse viser, at de 130 medarbejdere i gennemsnit havde 4,9 arbejdstimer pr. uge, og at medarbejderne i gennemsnit afveg 3,2 timer fra dette gennemsnit.

*Øvelse 14: Bestem standardafvigelsen i Øvelse 9.*

*Øvelse 15: Bestem standardafvigelsen i Øvelse 10.*

I nogle tilfælde vil beregningen af standardafvigelsen føre til misvisende resultater. Det kan ske, når et talmateriale indeholder en eller flere meget afvigende observationer. Hvis vi beregner standardafvigelsen på talmaterialet i tabel 5 (s. 294) får vi et resultat på 8,9, hvilket ikke giver meget mening.

I sådanne tilfælde kan vi opnå en bedre beskrivelse af fordelings spredning ved at beregne det såkaldte kvartilsæt. Kvartilsættet består af tre tal, der inddeler et sorteret talmateriale i fire lige store dele.

Kvartilsættets tre tal kaldes henholdsvis Første kvartil, Anden kvartil og Tredje kvartil. Første kvartil (nedre kvartil) er den observationsværdi, hvor en fjerdedel af talmaterialets værdier er lavere. Anden kvartil er den observationsværdi, hvor to fjerdedele eller halvdelen af talmaterialets værdier er lavere. Anden kvartil bliver hermed det samme som medianen. Endelig er tredje kvartil eller øvre kvartil den observationsværdi, hvor tre fjerdedele af materialet er lavere og en fjerdedel er højere.

Forskellen mellem øvre kvartil og nedre kvartil kaldes *kvartilafstanden* (eller interkvartilbredden). Kvartilafstanden udtrykker det område, hvor den midterste halvdel af observationerne befinder sig.

Nedre og øvre kvartil kan defineres ved hjælp af de summerede frekvenser:

Nedre kvartil (Første kvartil) bestemmes som den mindste observationsværdi, der har en summeret frekvens på mindst 0,25. I en intervalgrupperet fordeling findes medianen ved først at identificere det interval, hvor den summerede frekvens passerer 0,25. Kaldes den summerede frekvens ved intervallets start for  $F_{\text{START}}$  og den summerede frekvens ved intervallets afslutning for  $F_{\text{SLUT}}$ , kan den nedre kvartil bestemmes som:

$$\text{Nedre kvartil} = \text{Nedre intervalgrænse} + \frac{0,25 - F_{\text{START}}}{F_{\text{SLUT}} - F_{\text{START}}} \cdot \text{Intervalbredden}$$

Øvre kvartil (Tredje kvartil) bestemmes som den mindste observationsværdi, der har en summeret frekvens på mindst 0,75. I en intervalgrupperet fordeling findes medianen ved først at identificere det interval, hvor den summerede frekvens passerer 0,75. Kaldes den summerede frekvens ved intervallets start for  $F_{\text{START}}$  og den summerede frekvens ved intervallets afslutning for  $F_{\text{SLUT}}$ , kan den øvre kvartil bestemmes som:

$$\text{Øvre kvartil} = \text{Nedre intervalgrænse} + \frac{0,75 - F_{\text{START}}}{F_{\text{SLUT}} - F_{\text{START}}} \cdot \text{Intervalbredden}$$

Forsøger vi at bruge denne metode på fordelingen af medarbejdere efter ugentligt overtimetal i tabel 8, kan vi se, at nedre kvartil må befinde sig i intervallet fra 2,5 til 5,0 overtimer. Derfor kan nedre kvartil bestemmes som:

$$2,5 + \frac{0,25 - 0,2385}{0,6077 - 0,2385} \cdot (5,0 - 2,5) = 2,58.$$

Øvre kvartil befinder sig derimod i intervallet fra 5,0 til 7,5 overtimer og kan bestemmes til

$$5,0 + \frac{0,75 - 0,6077}{0,8077 - 0,6077} \cdot (7,5 - 5,0) = 6,78$$

Med baggrund i disse beregninger kan vi skønne, at de midterste 50 pct. af observationerne har en værdi mellem 2,58 og 6,78. Kvartilafstanden kan beregnes til:

$$\text{Øvre kvartil} - \text{Nedre kvartil} = 6,78 - 2,58 = 4,20.$$

$$\text{Kvartilafstand} = \text{Øvre kvartil} - \text{Nedre kvartil}$$

*Øvelse 16: Bestem kvartilafstanden i Øvelse 9.*

Kvartilerne findes ved at inddele et ordnet talmateriale i fire lige store dele. En finere inddeling kan opnås ved at inddele i fem lige store dele (kvintiler) eller i ti lige store andele (deciler). Kvintiler og deciler bruges af Danmarks Statistik blandt andet i opgørelser af indkomstfordelinger.

## Grafisk illustration af spredningen. Lorenz-kurven

I mange tilfælde kan det være en god ide at vise spredningen i et talmateriale ved en grafisk illustration.

**Tabel 12 Fordeling af familieindkomster før skat. Danmark 2004**

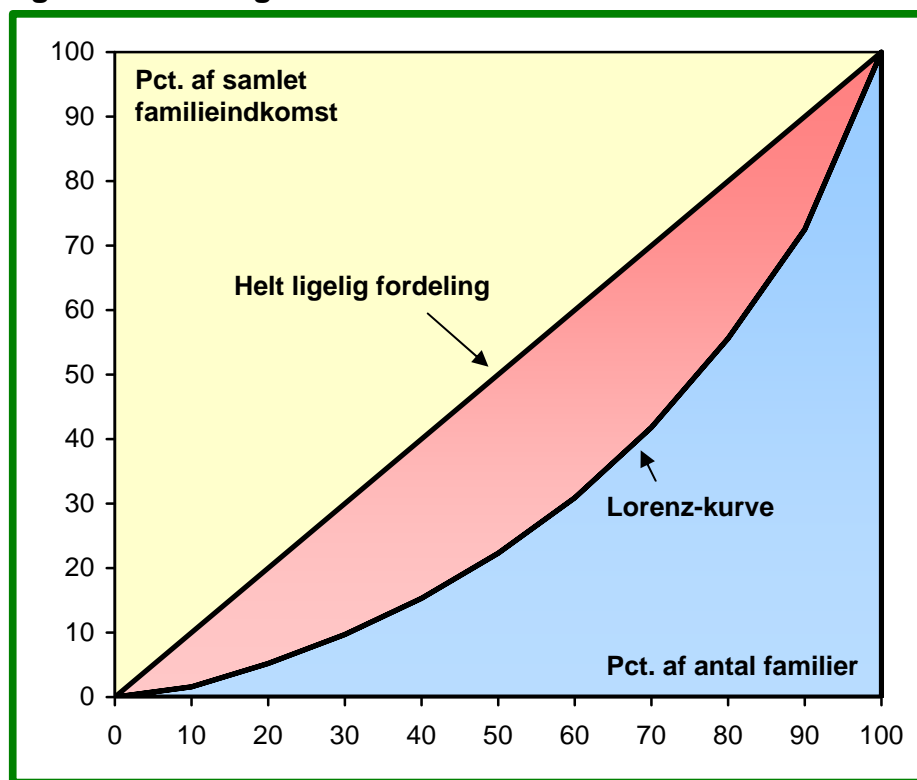
Indkomst-interval (1000 kr. pr. år)	Procentandel af familier i intervallet	Procentandel af indkomst i intervallet	Summeret procentandel af familier	Summeret procentandel af indkomst
0 – 112,4	10	1,6	10	1,6
112,4 – 148,9	10	3,6	20	5,2
148,9 – 183,7	10	4,5	30	9,7
183,7 – 229,6	10	5,6	40	15,3
229,6 – 284,0	10	7	50	22,3
284,0 – 351,3	10	8,6	60	30,9
351,3 – 453,0	10	10,9	70	41,8
453,0 – 558,1	10	13,8	80	55,6
558,1 – 696,3	10	16,9	90	72,5
Mere end 696,3	10	27,5	100	100,0
I alt	100	100,0	-	-

Kilde. Indkomster 2004. Danmarks Statistik 2006.

I tabel 12 er vist en intervalgrupperet fordeling af danske familiers indkomster i 2004. Familierne er sorteret efter indkomstens størrelse og herefter inddelt i ti lige store grupper, dvs. i deciler. Tabellen skal læses på den måde, at i 1. decil, dvs. den tiendedel af familierne med de laveste indkomster, var højeste familieindkomst på 73.000 kr. i 2004. Tilsammen havde familierne i 1. decil en andel på 1,6 pct. af de samlede indkomster. Vi kan også se af tabellens kolonner for de summerede procentandele at de første fem deciler, dvs. den halvdel af familierne med de laveste indkomster, tilsammen havde 22,3 pct. af de samlede indkomster. De sidste 77,7 pct. af den samlede indkomst tilfalder derfor den halvdel med de højeste indkomster.

Tabellen tyder derfor på en betydelig spredning i fordelingen af familieindkomster. Et grafisk billede af denne skævhed kan opnås ved at tegne tabellens to sidste kolonner som en kurve i et koordinatsystem med de summerede procentandele af familierne (tabellens lysegrønne kolonne) markeret på den vandrette akse og de summerede procentandele af de samlede indkomster (tabellens lyseblå kolonne) markeret på den lodrette akse. Herved fremkommer Lorenz-kurven som vist i figur 4.

**Figur 4 Fordeling af familieindkomster 2004. Lorenz-kurve**



Kilde: Tabel 12

I figur 4 er udover Lorenz-kurven vist diagonallinien, der viser hvordan Lorenz-kurven ville forløbe, hvis alle familier havde lige stor indkomst. I så fald ville der være præcis 10 pct. af den samlede indkomst i hver tiendedel af familierne. Den mest ulige fordeling, der kan tænkes, hvis vi ser bort fra muligheden af negativ indkomst vil være, hvis én familie havde al familieindkomsten og alle øvrige familier havde en indkomst på 0. Tegnede man en Lorenz-kurve på dette grundlag ville det med en mere detaljeret gruppeinddeling end tiendedele give en vinkelret Lorenz-

kurve, der fulgte den vandrette akse, lige indtil vi stødte på familien med indkomst. Herefter ville kurven gå næsten lodret op mod figurens øverste højre hjørne.

Man kan derfor i Lorenz-kurven få et mål for skævheden eller spredningen i fordelingen ved at undersøge, hvor stort det røde areal er i forhold til det blå areal. Jo større det røde areal er, jo skævere fordeling.

Sammenligner man Lorenz-kurver for forskellige fordelinger, vil en visuel betragtning ofte være nok til at afgøre, hvilken fordeling, der er mest lige. Der findes dog nogle nøgletal, nemlig den maksimale udjævningsprocent og Gini-koefficienten, der kan bruges til at sætte tal på skævheden i en Lorenz-kurve.

*Den maksimale udjævningsprocent* er den procentdel af de samlede indkomster, der skal flyttes fra familier med indkomst over gennemsnittet til familier med indkomster under gennemsnittet for at få en helt lige indkomstfordeling. Danmarks Statistik har beregnet den maksimale udjævningsprocent i fordelingen af familieindkomster i 2004 til 29,2 pct.

*Gini-koefficienten* er et mål for graden af ulighed. En helt lige fordeling vil have en Gini-koefficient på 0, mens koefficienten vil være 1 for den maksimalt skæve fordeling. I figuren kan Gini-koefficienten fortolkes som den andel det røde areal udgør af summen af det blå areal og det røde areal. Danmarks Statistik har beregnet Gini-koefficienten i fordelingen af familieindkomster i 2004 til 0,36.

*Øvelse 17: Bestem medianen for fordelingen af familieindkomster i tabel 12.*

*Øvelse 18: I tabel 13 vises fordelinger af familieindkomster 2004, opdelt på enlige og par. I tabellen er antallet af par og antallet af enlige fordelt på kvintiler, dvs i femtede dele med 20 pct. af familierne i hver kvintil. 1. kvintil består af de 20 pct. af familierne med lavest indkomst.*

**Tabel 13 Fordeling af familieindkomster før skat. Danmark 2004**

	Procentandel af familier i intervallet	Procentandel af indkomst i intervallet <i>Par</i>	Procentandel af indkomst i intervallet <i>Enlige</i>
1. kvintil	20	9,7	8,8
1. kvintil	20	15,2	15,0
1. kvintil	20	19,0	18,3
1. kvintil	20	22,6	23,0
1. kvintil	20	33,6	35,0
Maksimal udjævningsprocent	-	16,2	18,0
Gini-koefficient	-	0,24	0,26

Kilde. Statistisk Tiårsoversigt 2006. Danmarks Statistik 2006.

Tegn i samme diagram to Lorenz-kurve for fordelingen af indkomster hos par henholdsvis enlige. Sammenlign skævheden i de to fordelinger på baggrund af tabel og Lorenz-kurver.

## Sammenligning af niveaumål og spredningsmål

Ved vurdering af niveaumål og spredningsmål må man nøje overveje, om de er relevante i forhold til formålet med beregningen.

Middeltallet og standardafvigelsen er den mest udbredte metode til at bedømme niveau og spredning i et talmateriale. Det er allerede nævnt, hvorledes indflydelsen fra ekstreme observationer kan gøre, at middeltallet bliver et misvisende mål for det "normale" niveau i en fordeling. Ligeledes kan ekstreme observationer betyde, at standardafvigelse og variationsbredde giver et forkert billede af den typiske spredning i et talmateriale.

Den vigtigste anvendelse af middeltal og standardafvigelser er at skabe grundlag for sammenligning af flere talmaterialers niveau og spredning. Her er det væsentligt at være omhyggelig med at sørge for, at middeltal og standardafvigelse er opgjort på en måde, der gør dem anvendelige i forhold til den sammenligning, man ønsker.

I dette afsnit skal vi give et par eksempler på simple metoder til at forbedre sammenligneligheden af middeltal og standardafvigelse.

### Standardberegning

I tabel 14 vises fordelingen af medarbejdere efter uddannelse og timeløn i to virksomheder A og B:

**Tabel 14 Fordeling af medarbejdere efter uddannelse og timeløn**

Medarbejder gruppe	Virksomhed A		Virksomhed B	
	Timeløn kr.	Antal personer	Timeløn kr.	Antal personer
Ufaglærte	100	20	120	50
Faglærte	200	30	230	30
Ingeniører	300	50	340	20

Ud fra tabellens tal kan den gennemsnitlige timeløn i de to virksomheder beregnes ved at bruge formlen for middeltallet.

Gennemsnitlig løn i virksomhed A:

$$\bar{x}_A = \frac{100 \cdot 20 + 200 \cdot 30 + 300 \cdot 50}{100} = \frac{23000}{100} = 230 \text{ kr.}$$

Gennemsnitlig løn i virksomhed B:

$$\bar{x}_B = \frac{120 \cdot 50 + 230 \cdot 30 + 340 \cdot 20}{100} = \frac{19700}{100} = 197 \text{ kr.}$$

Kan vi nu konkludere, at lønniveauet er højere i virksomhed A end i virksomhed B? Svaret på dette spørgsmål kan være både ja og nej alt efter hvilket formål, der er med beregningen.

Hvis virksomhederne beregner salgspriser på baggrund af de gennemsnitlige løn-omkostninger pr. time, og formålet med beregningen er at finde et grundlag for en prisberegning, vil det være en holdbar konklusion at sige, at lønniveauet er højere i virksomhed A end i virksomhed B.

Men hvis formålet er at finde ud af hvilken af de to virksomheder, der betaler de højeste lønninger, vil det ikke være en korrekt konklusion at svare ja. Det fremgår tydeligt af tabellen, at virksomhed B betaler højere timeløn til alle tre medarbejderkategorier. Når den gennemsnitlige timeløn ender med at blive højest i virksomhed B, er forklaringen, at medarbejderstaben i de to virksomheder er sammensat meget forskelligt. Hvis vi skal lave en troværdig sammenligning, må vi have et mere ensartet grundlag at sammenligne på. Det kan vi få ved en såkaldt standardberegning, hvor vi beregner middeltallet med et ensartet "vægtgrundlag". I dette tilfælde må hyppigheden af de tre medarbejderkategorier gøres ensartet. Hvis vi lader "standard" være medarbejdersammensætningen i virksomhed A, kan vi lave en standardberegning af gennemsnitslønnen i virksomhed B:

Standardberegnet gennemsnitlig løn i virksomhed B:

$$\bar{x}_B = \frac{120 \cdot 20 + 230 \cdot 30 + 340 \cdot 50}{100} = \frac{26300}{100} = 263 \text{ kr.}$$

Hvis medarbejdersammensætningen i virksomhed B havde været den samme som i virksomhed A, ville det gennemsnitlige lønniveau have været 263 kr. pr. time, mens den gennemsnitlige løn i virksomhed A blot er 230 kr. pr. time. Så hvis vi ønsker at finde svaret på spørgsmålet: "Hvilken af de to virksomheder betaler de bedste lønninger?", vil svaret være virksomhed B.

*Øvelse 19: I tabel 15 vises fordelingen af lejelejligheder efter størrelse og husleje i to kommuner Bøgedal og Egehøj:*

**Tabel 15 Fordeling af lejelejligheder efter størrelse og husleje**

Lejligheds- type	Bøgedal kommune		Egehøj kommune	
	Gennem- snitlig husleje (kr. pr. md.)	Antal lejligheder	Gennem- snitlig husleje (kr. pr. md.)	Antal lejligheder
1-2 værelser	3000	100	3300	400
3-4 værelser	5000	300	5500	200
5 eller flere værelser	8000	500	8800	100



- a. Beregn den gennemsnitlige husleje i hver af de to kommuner.
- b. Brug fordelingen af lejligheder i Bøgedal kommune som standard og foretag en standardberegning af den gennemsnitlige husleje i hver af de to kommuner.
- c. Sammenlign huslejeniveauet i de to kommuner med udgangspunkt i svarene på a. og b.

## Variationskoefficient

Som nævnt er standardafvigelsen det mest almindelige redskab til at beskrive spredningen i et talmateriale. Det er dog ikke altid, at standardafvigelser for to talmaterialer uden videre kan sammenlignes.

Antag at vi i en måned har indsamlet daglige lukkekurser for AP Møller Mærsk B-aktier og aktier i Vestas A/S. Hvis vi ønsker at finde ud af hvilken af de to aktier, der har den mest stabile kurs, kan det være relevant at beregne standardafvigelsen. Lad os antage at standardafvigelsen for kursen på AP Møller Mærsk B-aktier er opgjort til 1080 kr., mens der i samme periode er målt en standardafvigelse for kursen på Vestas-aktier på 10,8 kr. Standardafvigelsen for kursen på AP Møller Mærsk B-aktier er derfor 100 gange større end standardafvigelsen på Vestas-aktier. Men det vil ikke være rimeligt, at konkludere at AP Møller Mærsk B-aktier er mere ustabile end Vestas-aktier. Forklaringen er, at gennemsnitskursen på AP Møller Mærsk B-aktier er meget højere end gennemsnitskursen på Vestas-aktier. I dette tilfælde har vi derfor behov for et relativt spredningstal, der sætter standardafvigelsen i forhold til gennemsnitsniveauet. En simpel måde at gøre dette på er at beregne variationskoefficienten:

$$\text{Variationskoefficient} = \frac{\text{Standardafvigelse}}{\text{Middeltal}} = \frac{s}{\bar{x}}$$

Hvis gennemsnitskurserne for AP Møller Mærsk B-aktier og Vestas-aktier i den pågældende måned er henholdsvis 54000 og 216, kan variationskoefficienterne beregnes til:

$$\text{Variationskoefficient for AP Møller Mærsk B-aktier} = \frac{1080}{54000} = 0,02$$

$$\text{Variationskoefficient for Vestas-aktier} = \frac{10,8}{216} = 0,05$$

På Vestas-aktien var der derfor en standardafvigelse på 5 pct. af middeltallet, mens der på Mærsk-aktien kun var tale om en standardafvigelse på 2 pct. af middeltallet.

Når vi tager hensyn til at kurserne på Mærsk aktier ligger væsentligt højere end kursen på Vestas-aktier, bliver konklusionen derfor, at Mærsk-aktien den pågældende måned har ligget mere stabilt i kurs end Vestas-aktien.



# Procentændring og indekstal

## Procentregning

I tabel 16 er vist tal for omsætningsudviklingen for tre virksomheder i perioden fra 2004 til 2006.

**Tabel 16 Omsætning i tre virksomheder 2004-2006**

Omsætning mio.kr.	2004	2005	2006
Virksomhed A	72,5	80,0	87,0
Virksomhed B	125,0	117,5	122,5
Virksomhed C	135,0	147,0	162,0

Vi kan af tabellen se, at omsætningen har udviklet sig forskelligt. I virksomhed A er omsætningen steget med  $87,0 - 72,5 = 14,5$  mio.kr. fra 2004 til 2006. Omsætningen i virksomhed B er derimod faldet med 2,5 mio.kr., mens den i C er steget med 27 mio.kr.

Vil man undersøge, hvor omsætningsændringen er kraftigst, er en beregning af *forskelle* i omsætning ikke tilstrækkelig. Man må tage hensyn til omsætningens niveau for at kunne vurdere, om en stigning på 14,5 mio.kr. er en stor eller lille stigning. Dette opnås ved at beregne den *relative* eller *procentvise* omsætningsændring:

$$\text{Procentvis ændring i omsætning} = \frac{\text{Omsætning efter ændring} - \text{Omsætning før ændring}}{\text{Omsætning før ændring}} \cdot 100$$

Ved beregning af procentvise ændringer sættes forskellen altså i forhold til det oprindelige niveau. Med tallene fra tabel 16 kan vi beregne følgende procentændringer fra 2004 til 2006:

$$\text{Virksomhed A: } \frac{87,0 - 72,5}{72,5} \cdot 100 = 20 \text{ pct.}$$

$$\text{Virksomhed B: } \frac{122,5 - 125,0}{125,0} \cdot 100 = -2 \text{ pct.}$$

$$\text{Virksomhed C: } \frac{162,0 - 135,0}{135,0} \cdot 100 = 20 \text{ pct.}$$

Af procentberegningen fremgår, at omsætningen er steget lige kraftigt i virksomhed A og virksomhed C. Dette kan være vanskeligt at se direkte i tabel 16.

## Indekstal

Indekstal angiver værdien af en størrelse, udtrykt som procent af værdien i en nærmere angivet basisperiode (eller basistidspunkt). En størrelse, der fra 2001 til 2005 er steget med 12 pct., kan således angives ved indeks 112, når 2001 anvendes som basisperiode. Indekstal anvendes navnlig, hvor interessen samler sig om at sammenligne relative ændringer i forskellige størrelser. Der skelnes mellem simple indeks og sammensatte indeks (vejede indeks).

## Simple indeks

En talrække omregnes til indekstal ved at dividere alle tal i rækken med værdien i *basisperioden* (eller på basistidspunktet) og gange med 100. Basisperioden (eller basistidspunktet) kan principielt vælges frit, men vil ofte være den første periode (tidspunkt) i talrækken.

$$\text{Simpelt indeks} = \frac{\text{Aktuel værdi}}{\text{Basisværdi}} \cdot 100$$

I tabellen nedenfor er vist en omregning af omsætningsudviklingen i virksomhed A, B og C til simple indeks, idet 2004 er valgt som basisperiode. Dette indebærer, at 2004-værdierne automatisk har værdien 100.

**Tabel 17 Simple indeks for omsætningen i tre virksomheder 2004-2006**

	Omsætning, mio. kr.			Indeks for omsætning (2004 = 100)		
	2004	2005	2006	2004	2005	2006
<b>A</b>	72,5	80,0	87,0	$\frac{72,5}{72,5} \cdot 100 = 100$	$\frac{80,0}{72,5} \cdot 100 = 110$	$\frac{87,0}{72,5} \cdot 100 = 120$
<b>B</b>	125,0	117,5	122,5	$\frac{125,0}{125,0} \cdot 100 = 100$	$\frac{117,5}{125,0} \cdot 100 = 94$	$\frac{122,5}{125,0} \cdot 100 = 98$
<b>C</b>	135,0	147,0	162,0	$\frac{135,0}{135,0} \cdot 100 = 100$	$\frac{147,0}{135,0} \cdot 100 = 108,9$	$\frac{162,0}{135,0} \cdot 100 = 120$

Indekstallene viser den procentvise afvigelse fra værdien i basisperioden. Af tallene i tabellen kan fx læses, at omsætningen er henholdsvis 10 pct. større, 6 pct. mindre og 8,9 pct. større i 2005 sammenlignet med 2004. Men bemærk, at den procentvise omsætningsændring fra 2005 til 2006 for virksomhed A *ikke* er 10 pct., men derimod

$$\frac{120 - 110}{110} \cdot 100 = \text{ca. } 9 \text{ pct.}$$

Indekstal som i tabel 15 kaldes *simple* eller *usammensatte* indeks.

## Sammensatte indeks. Prisindeks

I tilfælde, hvor den gennemsnitlige udvikling for flere talrækker ønskes udtrykt i ét enkelt tal, anvendes såkaldte *sammensatte* eller *vejede* indeks.

Ved beregningen af sammensatte indeks må man tage hensyn til de enkelte talrækkes betydning for helheden. Man har i denne forbindelse brug for et *vægtgrundlag*, som angiver den vægt, man bør tillægge hver enkelt talrække.

Nedenfor er som eksempel på dannelse af vejede indeks vist den metode, der anvendes i beregningen af forbrugerprisindekset. Dette indeks har til formål at angive den gennemsnitlige udvikling i priserne på de varer, som indgår i husholdningernes forbrugsudgifter. Mens der i praksis indgår et meget stort antal varepriser i beregningen, er eksemplet forenklet til kun at omfatte 4 varer. Som vægtgrundlag anvendes fordelingen af de samlede udgifter i år I. Det betyder, at vi i beregningen lader som om, at de indkøbte mængder er uændrede i perioden fra År I til År III.

**Tabel 18 Prisudviklingen for fire varer.**

	Priser			Mængder
	År I	År II	År III	År I
	- kr. pr. stk. -			(År I = 100)
Vare A	2,00	2,00	2,20	40
Vare B	10,00	12,00	14,00	16
Vare C	5,00	5,50	6,00	20
Vare D	15,00	13,50	12,00	4

De vejede indeks fremkommer ved, at vi beregner, hvordan udgifterne ville have været i hvert af årene, hvis man havde indkøbt mængderne fra år I til priserne i det aktuelle år. Disse udgifter sættes derfor i forhold til udgifterne i basisåret:

Vejet indeks år I:

$$\frac{2,00 \cdot 40 + 10,00 \cdot 16 + 5,00 \cdot 20 + 15,00 \cdot 4}{2,00 \cdot 40 + 10,00 \cdot 16 + 5,00 \cdot 20 + 15,00 \cdot 4} \cdot 100 = \frac{400}{400} \cdot 100 = 100$$

Vejet indeks år II:

$$\frac{2,00 \cdot 40 + 12,00 \cdot 16 + 5,50 \cdot 20 + 13,50 \cdot 4}{2,00 \cdot 40 + 10,00 \cdot 16 + 5,00 \cdot 20 + 15,00 \cdot 4} \cdot 100 = \frac{436}{400} \cdot 100 = 109$$

Vejet indeks år III:

$$\frac{2,20 \cdot 40 + 14,00 \cdot 16 + 6,00 \cdot 20 + 12,00 \cdot 4}{2,00 \cdot 40 + 10,00 \cdot 16 + 5,00 \cdot 20 + 15,00 \cdot 4} \cdot 100 = \frac{480}{400} \cdot 100 = 120$$

Af beregningen af de vejede indeks kan udledes, at priserne for de fire varer i gennemsnit i år II og år III er henholdsvis 9% og 20% højere end priserne i år I.

Den anvendte beregningsmetode kaldes *Laspeyres indeks*. Med symboler kan beregningerne af Laspeyresindekset med basisperiode år I udtrykkes ved:

$$P_{t,I}^{La} = \frac{\sum_{t=I}^{III} p_t \cdot m_I}{\sum_{t=I}^{III} p_I \cdot m_I} \cdot 100$$

hvor P er prisindeks, t angiver det aktuelle år, p er pris og m er mængde.

Af formlerne kan ses, at Laspeyres indekset viser den relative udgiftsvækst i forhold til basisåret, hvis man i år t havde købt de mængder, man købte i år I til de priser, der gælder i år t. Det betyder, at det vejede indeks er en standardberegning, hvor vi bruger varemængderne i basisåret som standard.

Ved anvendelsen af vejede prisindeks må man hele tiden have for øje, at der er tale om gennemsnitstal. Prisudviklingen kan for en husholdning, hvis forbrugssammensætning afviger fra vægtgrundlaget for det vejede indeks, adskille sig væsentligt herfra.

Ligeledes bør man være opmærksom på, at forbrugssammensætningen ændrer sig, blandt andet fordi priserne på de forskellige varer ikke udvikler sig ens. Folk vil ofte spare på de varer, der stiger i pris, og udvide forbruget af varer, som bliver billigere. Det betyder, at et Laspeyres prisindeks kan give for store prisstigninger, idet man ser bort fra, at folk reagerer på prisændringer. Hvis et vejet prisindeks ønskes anvendt over længere perioder, bør vægtgrundlaget med mellemrum ændres.

Hvis man ønsker at anvende nye vægte (løbende vægte), kan det såkaldte *Paasche indeks* vælges. Beregningsformlen er:

$$P_{t,I}^{Pa} = \frac{\sum_{t=I}^{III} p_t \cdot m_t}{\sum_{t=I}^{III} p_I \cdot m_t} \cdot 100$$

Lad os antage, at der er givet følgende oplysninger om priser og mængder for vare A, B, C og D i årene I, II og III:

**Tabel 19 Priser og mængder for fire varer i år I, II og III**

	Pris			Mængder		
	År I	År II	År III	År I	År II	År III
	- kr. pr. stk. -			stk		
Vare A	2,00	2,00	2,20	40	44	42
Vare B	10,00	12,00	14,00	16	15	14
Vare C	5,00	5,50	6,00	20	18	16
Vare D	15,00	13,50	12,00	4	5	7

Med udgangspunkt i tabel 19 kan Paasche prisindekset for år II beregnes ved:

$$P_{II,I}^{Pa} = \frac{2 \cdot 44 + 12 \cdot 15 + 5,5 \cdot 18 + 13,5 \cdot 5}{2 \cdot 44 + 10 \cdot 15 + 5 \cdot 18 + 15 \cdot 5} \cdot 100 = \frac{434,5}{403} \cdot 100 = 107,8$$

Af formelen og beregningen ses, at Paasche prisindekset viser den relative udgiftsvækst i forhold til basisåret, hvis man sammenligner udgiften i år t (beregningsåret) med den udgift, man i år I (basisåret) havde haft, hvis man i år I havde købt de samme varemængder, som man købte i år t.

*Øvelse 20: Beregn Paasche prisindeks for år I og III med udgangspunkt i tabel 19. Forklar hvorfor indeksene er forskellige fra de tilsvarende Laspeyres prisindeks.*

Ligesom ved Laspeyres prisindekset er der den ulempe ved Paasche prisindekset, at man ser bort fra, at folk reagerer på prisændringer. Det betyder, at Paasche indekset kan have en tendens til at undervurdere prisændringer, idet man bruger nye vægte, dvs. forudsætter, at forbrugerne har reageret på prisændringerne, inden de er forekommet.

Endvidere er Paasche prisindeks "dyrere i drift" end Laspeyres prisindeksene, idet beregningen kræver løbende information om såvel priser og mængder, mens beregningen af Laspeyres prisindeksene blot kræver mængdeoplysninger fra basisåret, og løbende prisoplysninger.

Det forhold, at Laspeyres prisindeksene ved at bruge faste vægte kan overvurdere prisændringerne, mens Paasche-prisindekset ved at bruge løbende vægte omvendt kan undervurdere prisændringerne, betyder, at et mere korrekt udtryk for prisudviklingen ofte vil ligge et sted imellem. Dette har givet anledning til ideen om, at beregne et gennemsnit af de to prisindeks, Fisher's indeks:

$$P_{t,I}^{Fi} = \sqrt{P_{t,I}^{La} \cdot P_{t,I}^{Pa}}$$

Herved har vi foretaget en beregning af *det geometriske gennemsnit* af Laspeyres og Paasche prisindekset.

*Øvelse 21: Beregn Fisher prisindeks for år I, II og III med udgangspunkt i tabel 19 og prisindeksberegningerne i tekst og øvelse 20.*

Det er ikke alene prisændringer, som kan udtrykkes ved hjælp af indekstal. Også udviklingen i mængder og omsætning kan udtrykkes på denne måde.

Laspeyres, Fishers og Paasche mængdeindeks (Q) kan beregnes ved formlerne

$$Q_{t,I}^{La} = \frac{\sum_{t=I}^{III} p_t \cdot m_t}{\sum_{t=I}^{III} p_I \cdot m_I} \cdot 100 \qquad Q_{t,I}^{Pa} = \frac{\sum_{t=I}^{III} p_t \cdot m_t}{\sum_{t=I}^{III} p_t \cdot m_I} \cdot 100$$

$$Q_{t,I}^{Fi} = \sqrt{Q_{t,I}^{La} \cdot Q_{t,I}^{Pa}}$$

Øvelse 22. Beregn Laspeyres, Paasche og Fisher mængdeindeks for år I, II og III med udgangspunkt i tabel 19.

For størrelser, der kan opfattes som et produkt af en mængde og en pris, gælder følgende sammensætning, idet fælles basisperiode for indeksene må forudsættes:

$$\text{Omsætnings- eller værdiindeks (PQ)} = \frac{\text{Prisindeks} \cdot \text{Mængdeindeks}}{100}$$

idet det dog gælder, at det ene af de indeks, man ganger sammen, skal være et Laspeyres, og det andet skal være et Paasche indeks. Værdiindekset kan således beregnes på følgende to måder:

$$PQ_{t,I} = \frac{P_{t,I}^{La} \cdot Q_{t,I}^{Pa}}{100} = \frac{P_{t,I}^{Pa} \cdot Q_{t,I}^{La}}{100}$$

Øvelse 23. Beregn Værdiindeks for år I, II og III med udgangspunkt i tabel 19 og de foretagne indeksberegninger.

## Beregning af gennemsnitlig vækstrate

Ved analyse af udviklingsforløb, der strækker sig over flere perioder, har man ofte behov for at beregne den gennemsnitlige vækst eller procentændring pr. periode. Antag fx at de samlede udgifter i en kommune over en treårig periode er steget med 15 pct. det første år, faldet 14 pct. det andet år og steget 17 pct. det tredje år. Ved beregning af den gennemsnitlige procentstigning for de tre år kunne man fristes til at bruge formelen for middeltallet (det aritmetiske gennemsnit), dvs. finde summen af de tre tal og dividere med tre (antallet af perioder). Denne metode er imidlertid ikke korrekt, idet man ikke får taget hensyn til, at de tre procentstigninger ikke har samme udgangspunkt.

Den korrekte måde at finde den gennemsnitlige vækstrate, er at omregne til vækstfaktorer (rentefaktorer). Vækstfaktorerne i de tre perioder er de faktorer man skal gange med for at komme fra foregående periodes niveau til den efterfølgende periodes niveau. I vores eksempel er vækstfaktorerne 1,15, 0,86 og 1,17. Herefter beregnes det geometriske gennemsnit af de tre vækstfaktorer ved at tage kubikroden af produktet af de tre vækstfaktorer:

$$\text{Gennemsnitlig vækstfaktor} = \sqrt[3]{1,15 \cdot 0,86 \cdot 1,17} = 1,0499 \approx 1,05$$

Den gennemsnitlige vækstfaktor i perioden er derfor 1,05, hvilket betyder at den gennemsnitlige procentændring kan beregnes ved:

$$\text{Gennemsnitlig procentstigning} = (1,05 - 1) \cdot 100 = 5 \text{ pct.}$$

Hvis procentændringerne over  $n$  perioder er  $p_1, p_2, p_3, \dots, p_n$ , kan den gennemsnitlige procentstigning pr. periode beregnes som:

$$P_G = \left( \sqrt[n]{\left(1 + \frac{p_1}{100}\right) \cdot \left(1 + \frac{p_2}{100}\right) \cdot \left(1 + \frac{p_3}{100}\right) \cdots \left(1 + \frac{p_n}{100}\right)} - 1 \right) \cdot 100$$

En tilsvarende teknik benyttes, hvis man vil beregne gennemsnitlige procentstigninger pr. periode ud fra værdierne ved en længere periodes start og slutning. Antag fx at et forbrugerprisindeks over en tiårig periode er steget fra 125 til 185. Den gennemsnitlige årlige prisstigning i procent kan beregnes som

$$\text{Gennemsnitlig årlig prisstigning} = \left( \sqrt[10]{\frac{185}{125}} - 1 \right) \cdot 100 = 4,0 \text{ pct.}$$

Hvis en størrelse over  $n$  perioder er steget fra  $X_{\text{START}}$  til  $X_{\text{SLUT}}$  kan den gennemsnitlige procentstigning pr. periode beregnes som:

$$P_G = \left( \sqrt[n]{\frac{X_{\text{SLUT}}}{X_{\text{START}}}} - 1 \right) \cdot 100$$

Øvelse 24: I Danmark er bruttonationalproduktet i faste priser i årene fra 2000 til 2004 opgjort som vist i tabel 20:

**Tabel 20 Bruttonationalproduktet (BNP) i Danmark 2000-2004**

Mia.kr. 2000 priser	2000	2001	2002	2003	2005
BNP	1294,0	1303,1	1309,2	1318,2	1342,9

Kilde: Statistisk Tiårsoversigt 2006, Danmarks Statistik.

- Beregn den procentvise årlige vækst i BNP i forhold til sidste år for hvert af årene 2001, 2002, 2003 og 2004. Svarene gives med 1 decimal.
- Opstil et simpelt indeks, der viser udviklingen i BNP i årene fra 2000 til 2004. Svarene gives med 1 decimal.
- Beregn den gennemsnitlige årlige vækst i BNP i perioden fra 2000 til 2004. Svarene gives med 1 decimal.

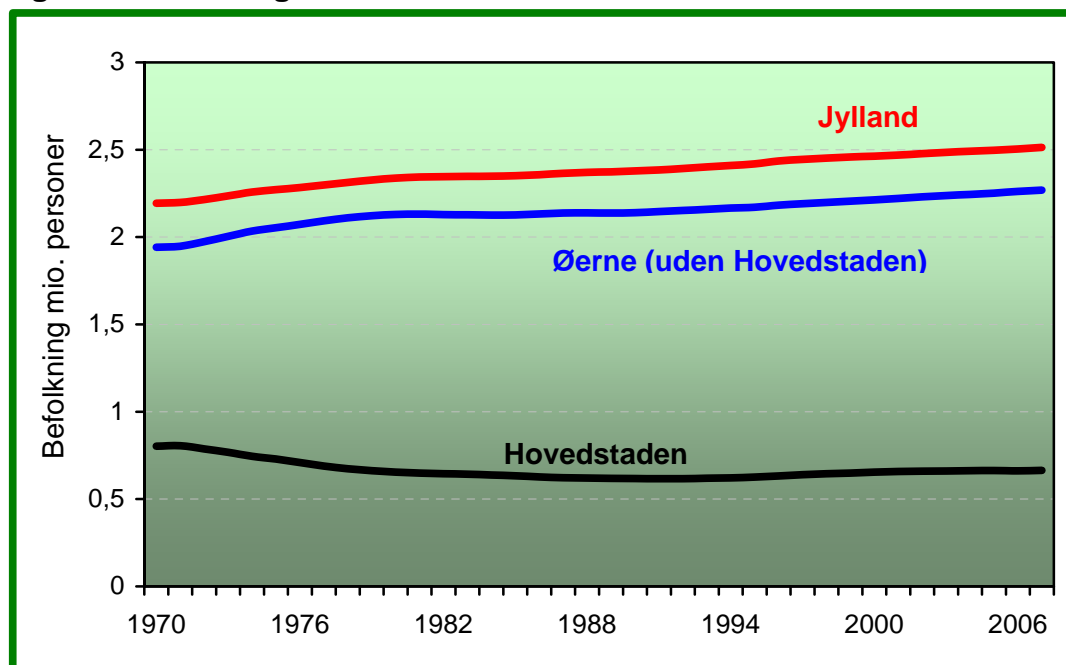
Øvelse 25: I et land er inflationsraten (vækstprocenten i det generelle prisniveau fra året før) i tre på hinanden følgende år opgjort til 12,1 pct., 5,1 pct. og 3,5 pct. Beregn (med 1 decimal) den gennemsnitlige inflationsrate i de tre år.

## Grafisk afbildning af udviklingsforløb

Til grafisk afbildning af udviklingsforløb er det almindeligt at bruge kurvediagrammer. Kurvediagrammer kan varieres på utallige måder. Nedenfor vises fire forskellige måder at illustrere, hvorledes den danske befolknings fordeling på hovedlandsdele har udviklet sig i perioden 1970-2007.

I figur 5 er vist et diagram, hvor udviklingen i befolkningstallet i hver af de tre hovedlandsdele er markeret med tre kurver. I figuren kan vi sammenligne niveauerne for befolkningerne i de tre hovedlandsdele. Vi kan se at befolkningstallet i Hovedstaden (Københavns og Frederiksberg kommune) er faldet over perioden, mens de to andre hovedlandsdele har oplevet stigende befolkningstal.

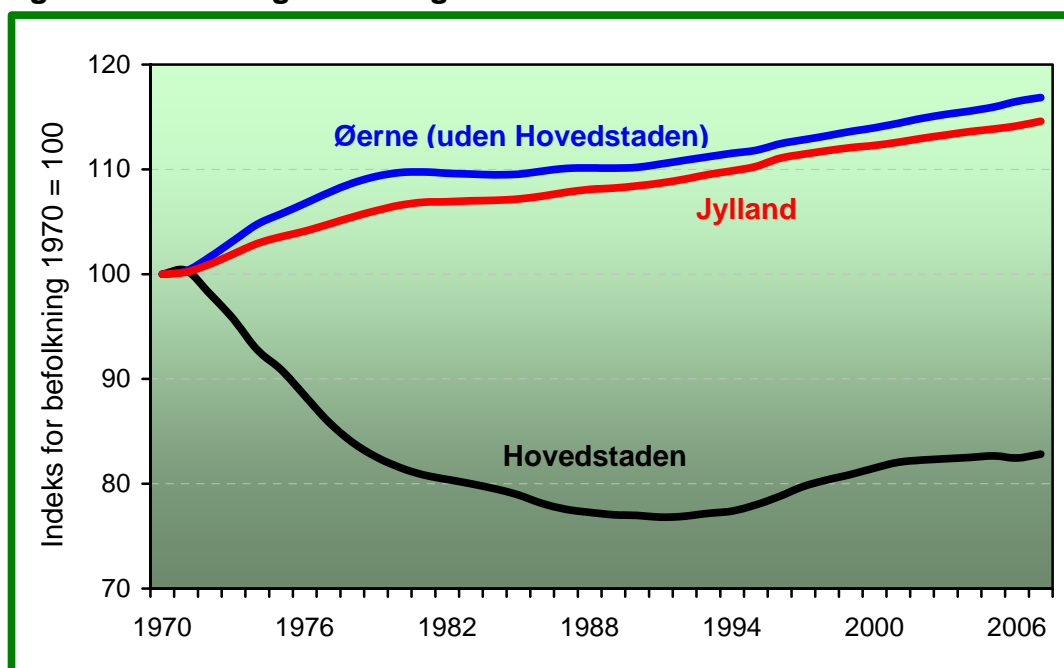
**Figur 5** Befolkningen i hovedlandsdele 1970-2007



Kilde: Danmarks Statistik

Hvis interessen retter sig mod at finde ud af, hvor kraftig befolkningsændringen har været i de tre hovedlandsdele, kan en omregning til indekstal give en mere tydelig figur. Det er vist i figur 6, der er baseret på samme datamateriale som figur 5, men befolkningstallene i de enkelte hovedlandsdele er omregnet til indekstal med basisår i 2005. I figuren kan vi se at Øerne (uden Hovedstaden) har haft en større procentvis befolkningsstigning end Jylland. Det kan også ses, at Hovedstaden i perioden har haft den procentvis største befolkningsændring blandt de tre hovedlandsdele.

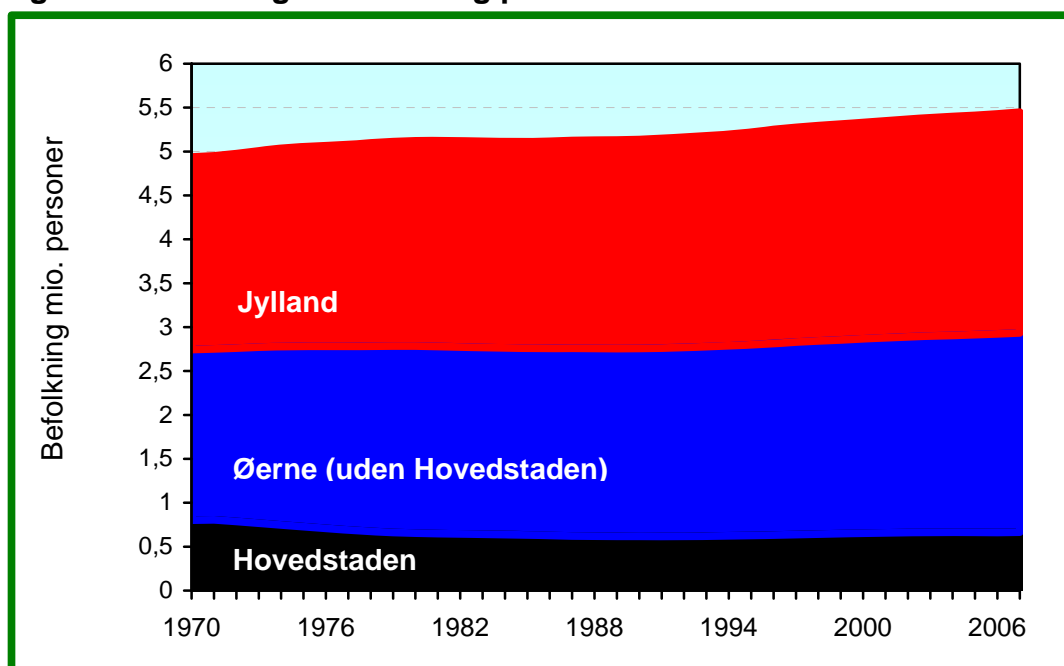


**Figur 6 Befolkningsudviklingen i hovedlandsdele 1970-2007**

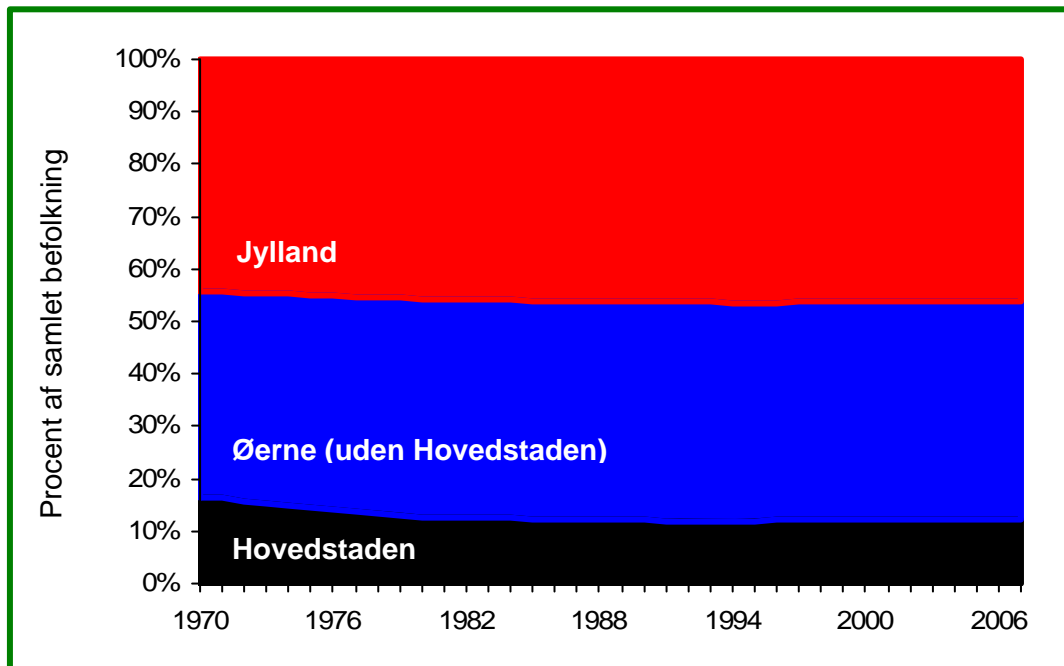
Kilde: Danmarks Statistik

I figur 7 er vist et såkaldt "stabledt kurvediagram", hvor de tre kurver er lagt "oven i hinanden". En gevinst ved dette diagram er, at man kan følge udviklingen i summen af befolkningstal i de tre hovedlandsdele, dvs. den samlede befolkning i Danmark.

Hvis hovedinteressen retter sig mod den relative fordeling af befolkningen på hovedlandsdele, kan man få et mere tydeligt billede ved at tegne et stabledt kurvediagram med de procentvise befolkningsandele i hver af de tre hovedlandsdele. Dette er vist i figur 8.

**Figur 7 Befolkningens fordeling på hovedlandsdele 1970-2007**

Kilde: Danmarks Statistik

**Figur 8 Befolkningens fordeling på hovedlandsdele 1970-2007**

Kilde: Danmarks Statistik

# Opstilling af tabeller og figurer

## Tabeller

Tabeller anvendes til at præsentere små og store mængder af information på en simpel og overskuelig måde. Tabelopstillingen kan illustreres skematisk som vist i Tabel 21.

**Tabel 21 Tabeloverskrift**

	Tabelhoved				
Tabellens forspalte			Søjle (kolonne)		
		Celle			
	Række			Række	
			Søjle (kolonne)		
Sumrække					

Anm:

Fodnoter til række, søjle eller celle

Kilde: Udgiver, titel, evt. år.

I præsentationer med flere tabeller bør hver tabel have sit eget *tabelnummer*, som anvendes, når man henviser til teksten.

*Tabeloverskriften* bør give en klar og tydelig beskrivelse af, hvad tabellen viser. Hvis tabellen dækker et begrænset geografisk område bør dette nævnes i overskriften. Hvis tabellen dækker en bestemt periode eller et bestemt tidspunkt, bør disse nævnes i overskriften.

*Tabellens forspalte* bør give kort og tydelig information om indholdet af de enkelte rækker. Der kan være tale om et talinterval, et geografisk område eller andre kategorier.

*Tabelhovedet* indeholder tilsvarende oplysninger om indholdet af de enkelte søjler. Ofte vil tabelhovedet opdeles i to rækker, hvor den øverste række giver generel information om flere eller alle søjler, mens den nederste række indeholder søjleoverskrifter.

*Sumrækker og sumsøjler.* I mange tabeller vil det være relevant at vise rækker eller søjler med sammentællinger. I tabel 19 er der blot vist en sammentællingsafdeling, nemlig sumrækken, hvor samlede værdier for de enkelte søjler kan præsenteres.

*Måleenhed.* Det er vigtigt at anvendte måleenheder (fx antal personer, km, pct. af BNP eller indekstal) tydeligt fremgår af tabellen. Er måleenheden fælles for alle tabellens tal kan den placeres i overskriften, i tabelhovedets generelle del eller i tabellens øverste venstre hjørne (det mørkeblå felt i tabel 17). Hvis måleenheden er forskellig i de enkelte rækker, søjler eller celler, må måleenheden markeres ved de enkelte rækker, søjler eller celler.

*Anmærkninger (Anm.) og fodnoter.* Hvis der er behov for yderligere information om tabellens indhold end der er plads til i tabeloverskriften, kan denne information gives i en anmærkning under tabellen. Anmærkninger vil ofte indeholde definitioner af begreber anvendt generelt i tabellen. Hvis der er behov for supplerende bemærkninger til enkelte rækker, søjler eller celler, bør disse gives i form af fodnoter under tabellen. Disse fodnoter nummereres fortløbende, således at henvisningen står i den enkelte række, søjle eller celle.

*Kildeangivelse* bør tydeligt fremgå, så det er muligt for læseren at kontrollere de anvendte oplysninger og eventuelt søge yderligere information. Er tabellen baseret på konstruerede eksempler eller egne beregninger eller undersøgelser, bør dette nævnes.

Selvom der i præsentationer altid vil blive givet en supplerende forklaring til tallene i en tabel, er det en god ide at udforme tabeller, så de kan læses uafhængigt af kommentarerne i præsentationen.

Ligeledes er det vigtigt at være omhyggelig med at gøre tabellen overskuelig. En tabel bør derfor kun indeholde tal, der er relevante i sammenhængen. Store tabeller kan ofte med fordel opdeles i flere mindre tabeller. Valget af måleenhed bør nøje overvejes. Mange cifre og decimaler giver en uoverskuelig tabel.

## **Grafiske afbildninger**

Tabeller kan ofte med fordel erstattes af grafiske fremstillinger. Figurer er lettere at forstå og kan bruges til at give et visuelt overblik over et talmateriale.

Figurer kan udformes på utallige måder. Det er ikke hensigten med dette afsnit at give en oversigt over forskellige diagramtyper, men blot at nævne nogle generelle retningslinier for præsentationen af grafiske afbildninger. Disse retningslinier er i vid udstrækning fælles med de retningslinier, der ovenfor er givet om tabelpræsentation.

En figur bør være forsynet med

*Figurnummer*

*Overskrift*

*Måleenhed(er)*

*Anmærkninger og fodnoter efter behov.*

*Kildeangivelse.*

I søjlediagrammer, kurvediagrammer og andre diagrammer, der benytter sig af et koordinatsystem, bør der være en tydelig markering på akserne af *hvad*, der måles, og af hvilken *måleenhed*, der anvendes. Hvis en akse bruges til at måle årstal eller andre tidsangivelser, kan angivelse af måleenhed dog undlades, hvis tidsangivelsen er markeret tydeligt. Der bør altid være en præcis markering af inddelingen af akserne.

Ligesom ved tabelopstillingen, er det en god ide, at udforme figurer, så de kan læses uafhængigt af kommentarerne i præsentationen. Ligeledes er det vigtigt at være omhyggelig med at gøre figuren overskuelig. For meget information eller for mange talserier kan gøre en figur uoverskuelig.

## Svar på øvelserne i kapitlet:

Øvelse 1:

$$\bar{x} = \frac{10 + 49 + 0 + 60 + 13 + 40 + 45 + 0 + 70 + 30 + 55 + 64 + 45}{13} = \frac{481}{13} = 37 \text{ minutter}$$

Øvelse 2:

Antal skadede	Antal ulykker	Antal skadede i gruppen
1	10	10
2	13	26
3	12	36
4	5	20
5	2	10
6	7	42
7	4	28
8	1	8
9	1	9
I alt	55	189

$$\text{Middeltal} = \bar{x} = \frac{189}{55} = 3,44 \text{ skadede pr. ulykke}$$

Øvelse 3:

a.

Antal børn i husstanden	Antal husstande	Frekvens
0	5	0,10
1	15	0,30
2	11	0,22
3	8	0,16
4	6	0,12
5	5	0,10
I alt	50	1,00

$$b. \bar{x} = 0 \cdot 0,10 + 1 \cdot 0,30 + 2 \cdot 0,22 + 3 \cdot 0,16 + 4 \cdot 0,12 + 5 \cdot 0,10 = 2,2 \text{ børn pr. husstand}$$

## Øvelse 4:

a.

Antal advarsler	Antal spillere (hyppighed)	Antal advarsler i gruppen
0	9	0
1	9	9
2	6	12
3	2	6
4	2	8
8	1	8
12	1	12
I alt	30	55

b.  $\bar{x} = \frac{55}{30} = 1,83$  advarsler pr. spiller. Typetal er 0 og 1 advarsel.

## Øvelse 5:

Typetallet er 1 barn pr. husstand.

## Øvelse 6:

Medianen er 3 personer med skader. Typetallet er 2 personer med skader.

## Øvelse 7:

Medianen er 2 børn pr. husstand.

## Øvelse 8:

Medianen er 1 advarsel pr. spiller.

## Øvelse 9:

Alder	Intervalmidtpunkt	Antal børn	Intervalmidtpunkt · antal børn
0-3 år	2	15	30
4-7 år	6	25	150
8-11 år	10	35	350
12-15 år	14	25	350
I alt	-	100	880

$$\bar{x} = \frac{880}{100} = 8,8 \text{ år}$$

## Øvelse 10:

Ledigheds- periodens varighed	Intervalmidt- punkt	Antal tusinde personer	Intervalmidtpunkt · antal tusinde per- soner
Under 74 dage	37,5	267	10012,5
74-146 dage	110,5	118	13039
147-219 dage	183,5	75	13762,5
220-292 dage	256,5	46	11799
Mindst 293 dage	329	38	12502
I alt	-	544	61115

$\bar{x} = \frac{61115}{544} = 112,34$  dage. Middeltallet udtrykker den gennemsnitlige varighed af ledigheden. Mindre afvigelser kan forekomme, da fastlæggelsen af intervalmidtpunkter kan diskuteres.

## Øvelse 11:

a.

Antal beskæftigede	Antal virksomheder	Antal beskæftigede	Antal virksomheder
0-9	10	50-59	10
10-19	8	60-69	8
20-29	8	70-79	8
30-39	7	80-89	7
40-49	4	90-99	4

b. 32,42 ansatte pr. virksomhed. c. 32,27 ansatte pr. virksomhed.

d. Ved intervalgrupperingen mistes noget præcis information. De 32,42 forudsætter en jævn fordeling af observationerne inden for intervallet. 32,27 er derfor det korrekte middeltal.

Øvelse 12: Typeintervallet for talmaterialet i Øvelse 9 er 8-11 år, mens typeintervallet i Øvelse 10 er under 74 dage.

Øvelse 13: Medianen for talmaterialet i Øvelse 9 er 9,1 år, mens medianen i Øvelse 10 er 77,1 dage. Mindre afvigelser kan forekomme, da fastlæggelsen af intervalmidtpunkter kan diskuteres.



Øvelse 14:

Interval- midtpunkt	Afvigelse fra mid- deltal	Antal børn	Kvadrat på afvigelse fra middeltal · antal børn
2	-6,8	15	693,6
6	-2,8	25	196,0
10	1,2	35	50,4
14	5,2	25	676,0
I alt	-	100	1616,0

$$s = \sqrt{\frac{1616}{100 - 1}} = 4,0 \text{ år}$$

Øvelse 15:

Interval- midtpunkt	Afvigelse fra mid- deltal	Antal personer	Kvadrat på afvigelse fra middeltal · antal personer
37,5	-74,84	267000	1495473835,0
110,5	-1,84	118000	399500,8
183,5	71,16	75000	379780920,0
256,5	144,16	46000	955976857,6
329	216,66	38000	1783779113,0
I alt	-	544000	4615410226,0

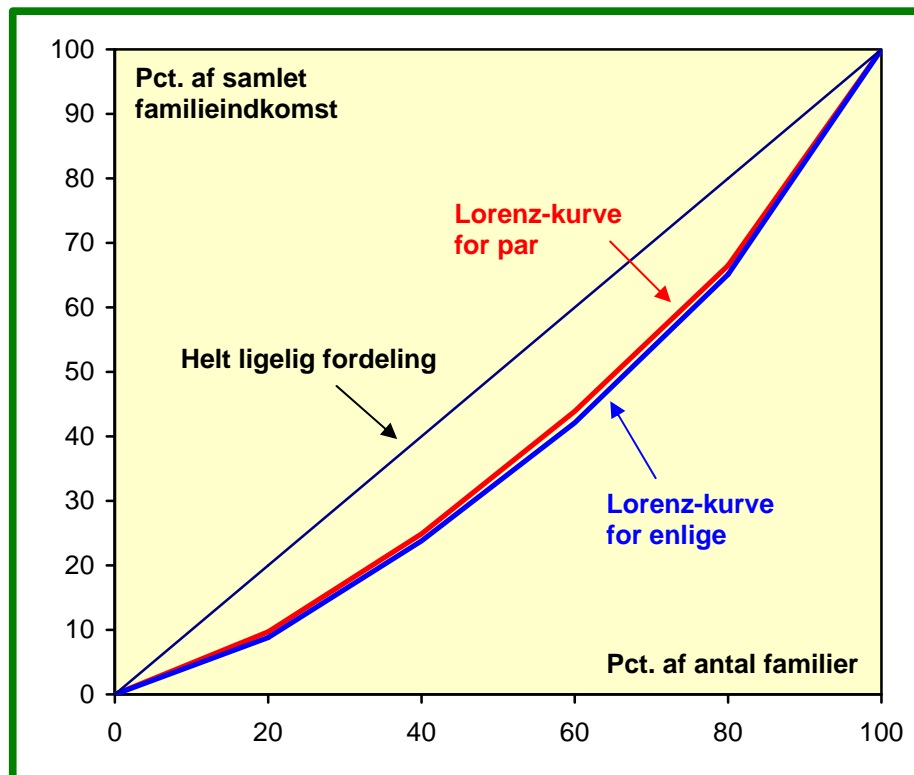
$$s = \sqrt{\frac{4615410226}{544000 - 1}} = 92,1 \text{ dage.}$$

Mindre afvigelser kan forekomme, da fastlæggelsen af intervalmidtpunkter kan diskuteres.

Øvelse 16: Øvre kvartil for talmaterialet i Øvelse 9 er 12,0 år, mens nedre kvartil er 5,6 år. Kvartilafstanden er derfor  $12,0 - 5,6 \text{ år} = 6,4 \text{ år}$ .

Øvelse 17: Medianen kan findes, som indkomsten hos den familie, der bringer den summerede frekvens op på 0,5. Det svarer i tabel 12 til højeste familieindkomst i 5. decil, hvor den summerede procentandel af familierne når 50. Medianindkomsten er her 284,0 eller 284.000 kr.

Øvelse 18: Da den maksimale udjævningsprocent og Gini-koefficienten begge er større for indkomstfordelingen blandt enlige end for indkomstfordelingen blandt par, er indkomstfordelingen blandt enlige mere ulige end den er blandt par. Dette billede understreges af at Lorenz-kurven for enlige ligger en smule længere fra diagonalen end Lorenz-kurven for par.



Øvelse 19:

a. Gennemsnitlig husleje i Bøgedal kommune:

$$\bar{x}_B = \frac{100 \cdot 3000 + 300 \cdot 5000 + 500 \cdot 8000}{900} = \frac{5800000}{900} = 6444 \text{ kr. pr. måned}$$

Gennemsnitlig husleje i Egehøj kommune:

$$\bar{x}_E = \frac{400 \cdot 3300 + 200 \cdot 5500 + 100 \cdot 8800}{700} = \frac{3300000}{700} = 4714 \text{ kr. pr. måned}$$

b. Når fordelingen af lejligheder i Bøgedal kommune bruges som standard, vil gennemsnitshuslejen i Bøgedal stadig være 6444 kr. pr. måned. Men den standardregnede husleje i Egehøj kommune vil blive:

$$\bar{x}_E = \frac{100 \cdot 3300 + 300 \cdot 5500 + 500 \cdot 8800}{900} = \frac{6380000}{900} = 7089 \text{ kr. pr. måned}$$

c. Når den faktiske gennemsnitshusleje i Egehøj kommune er lavere end i Bøgedal kommune, skyldes det at der er relativt mange små lejligheder i Egehøj. Korrigeres der for den forskellige fordeling på små og store lejligheder er huslejeniveauet klart større i Egehøj kommune sammenlignet med Bøgedal kommune.

Øvelse 20:

År I:

$$P_{I,I}^{Pa} = \frac{2 \cdot 40 + 10 \cdot 16 + 5 \cdot 20 + 15 \cdot 4}{2 \cdot 40 + 10 \cdot 16 + 5 \cdot 20 + 15 \cdot 4} \cdot 100 = 100$$

År III:

$$P_{III,I}^{Pa} = \frac{2,2 \cdot 42 + 14 \cdot 14 + 6 \cdot 16 + 12 \cdot 7}{2 \cdot 42 + 10 \cdot 14 + 5 \cdot 16 + 15 \cdot 7} \cdot 100 = \frac{468,4}{409} \cdot 100 = 114,52$$

Paasche-prisindeksene er lavere end de tilsvarende Laspeyres-indeks, da man bruger aktuelle mængder som vægtgrundlag. Herved forudsættes, at man allerede før en prisstigning på en vare, reducerer sit forbrug af varen, samt at man øger mængden af varer, der falder i pris, inden prisændringen.

I Laspeyres-prisindeksene bruges basisperiodens mængder som vægtgrundlag. Det indebærer en forudsætning om, at man ikke reagerer på prisændringer.

Øvelse 21:

$$P_{I,I}^{Fi} = \sqrt{P_{I,I}^{La} \cdot P_{I,I}^{Pa}} = \sqrt{100 \cdot 100} = 100$$

$$P_{II,I}^{Fi} = \sqrt{P_{II,I}^{La} \cdot P_{II,I}^{Pa}} = \sqrt{109 \cdot 107,8} = 108,4$$

$$P_{III,I}^{Fi} = \sqrt{P_{III,I}^{La} \cdot P_{III,I}^{Pa}} = \sqrt{120 \cdot 114,52} = 117,23$$

Øvelse 22:

År I:

$$Q_{I,I}^{La} = \frac{2 \cdot 40 + 10 \cdot 16 + 5 \cdot 20 + 15 \cdot 4}{2 \cdot 40 + 10 \cdot 16 + 5 \cdot 20 + 15 \cdot 4} \cdot 100 = 100$$

$$Q_{I,I}^{Pa} = \frac{2 \cdot 40 + 10 \cdot 16 + 5 \cdot 20 + 15 \cdot 4}{2 \cdot 40 + 10 \cdot 16 + 5 \cdot 20 + 15 \cdot 4} \cdot 100 = 100$$

$$Q_{I,I}^{Fi} = \sqrt{Q_{I,I}^{La} \cdot Q_{I,I}^{Pa}} = \sqrt{100 \cdot 100} = 100$$

År II:

$$Q_{II,I}^{La} = \frac{2 \cdot 44 + 10 \cdot 15 + 5 \cdot 18 + 15 \cdot 5}{2 \cdot 40 + 10 \cdot 16 + 5 \cdot 20 + 15 \cdot 4} \cdot 100 = \frac{403}{400} \cdot 100 = 100,75$$

$$Q_{II,I}^{Pa} = \frac{2 \cdot 44 + 12 \cdot 15 + 5,5 \cdot 18 + 13,5 \cdot 5}{2 \cdot 40 + 12 \cdot 16 + 5,5 \cdot 20 + 13,5 \cdot 4} \cdot 100 = \frac{434,5}{436} \cdot 100 = 99,66$$

$$Q_{II,I}^{Fi} = \sqrt{Q_{II,I}^{La} \cdot Q_{II,I}^{Pa}} = \sqrt{100,75 \cdot 99,66} = 100,2$$

År III:

$$Q_{III,I}^{La} = \frac{2 \cdot 42 + 10 \cdot 14 + 5 \cdot 16 + 15 \cdot 7}{2 \cdot 40 + 10 \cdot 16 + 5 \cdot 20 + 15 \cdot 4} \cdot 100 = \frac{409}{400} \cdot 100 = 102,25$$

$$Q_{III,I}^{Pa} = \frac{2,2 \cdot 42 + 14 \cdot 14 + 6 \cdot 16 + 12 \cdot 7}{2,2 \cdot 40 + 14 \cdot 16 + 6 \cdot 20 + 12 \cdot 4} \cdot 100 = \frac{468,4}{480} \cdot 100 = 97,58$$

$$Q_{III,I}^{Fi} = \sqrt{Q_{III,I}^{La} \cdot Q_{III,I}^{Pa}} = \sqrt{102,25 \cdot 97,58} = 99,89$$

Øvelse 23:

År I:

$$PQ_{I,I} = \frac{P_{I,I}^{La} \cdot Q_{I,I}^{Pa}}{100} = \frac{100 \cdot 100}{100} = 100$$

$$PQ_{I,I} = \frac{P_{I,I}^{Pa} \cdot Q_{I,I}^{La}}{100} = \frac{100 \cdot 100}{100} = 100$$

År II:

$$PQ_{II,I} = \frac{P_{II,I}^{La} \cdot Q_{II,I}^{Pa}}{100} = \frac{109 \cdot 99,66}{100} = 108,6$$

$$PQ_{II,I} = \frac{P_{II,I}^{Pa} \cdot Q_{II,I}^{La}}{100} = \frac{107,8 \cdot 100,75}{100} = 108,6$$

År III:

$$PQ_{III,I} = \frac{P_{III,I}^{La} \cdot Q_{III,I}^{Pa}}{100} = \frac{120 \cdot 97,58}{100} = 117,1$$

$$PQ_{III,I} = \frac{P_{III,I}^{Pa} \cdot Q_{III,I}^{La}}{100} = \frac{114,52 \cdot 102,25}{100} = 117,1$$

Øvelse 24:

Mia.kr. 2000 priser	2000	2001	2002	2003	2005
BNP	1294,0	1303,1	1309,2	1318,2	1342,9
a. Vækst fra foregående år		0,7	0,5	0,7	1,9
b. Simpelt indeks	100,0	100,7	101,2	101,9	103,8

c. Den gennemsnitlige procentvise årlige vækst kan beregnes på to måder:

$$\left( \sqrt[4]{\frac{1342,9}{1294}} - 1 \right) \cdot 100 = 1,0$$

eller som

$$\left( \sqrt[4]{1,007 \cdot 1,005 \cdot 1,007 \cdot 1,019} - 1 \right) \cdot 100 = 1,0$$

For at få korrekt resultat i den sidste beregning skal man have mere end 1 decimal med i procentstigningerne, der bruges til at beregne vækstfaktorerne.

Øvelse 23:

$$\text{Gennemsnitlig inflationsrate} = \left( \sqrt[3]{1,121 \cdot 1,051 \cdot 1,035} - 1 \right) \cdot 100 = 6,8 \text{ pct.}$$