Kogebog til "Beskrivende Statistik"

Når man skal lave en statistisk opgave fra bunden, skal man først gøre sig klart om man skal bruge "Diskrete variable" eller "Grupperede variable".

Forskellen på de "Diskrete variable" og de "Grupperede variable" er følgende:

Diskrete variable bruger man, når man gentagne gange har registreret bestemte størrelser. Tag for eksempel følgende eksempel, hvor man har målt hvor mange gange man har registreret familier med netop 1 barn, 2 børn osv. Bemærk, at en familie ikke her kan have 1,8 børn. Alle registreringer er på NETOP 1, 2, 3 ... osv. børn.

Grupperede variable bruger man, hvis de registreringer man har ikke er "præcise". Eller med andre ord – hvis de målinger man foretager, ikke er ens, men derimod godt kan indpasses i et interval. Tag for eksempel følgende eksempel, hvor man godt kan læse lektier i 2,8 timer – også selvom der ikke er nogen præcis værdi på 2,8 timer. Men her kan man indpasse registreringen i intervallet]0;3] timer. (Læses: "Fra 0 til 3 timer".)

I dette notat vil definitioner være skrevet med blå skrift, og eksempler vil være skrevet med rødt.

Først en (lang) række definitioner og eksempler derpå:

For bedre at kunne forstå definitionerne, betragtes følgende eksempel, som er taget fra "HTX Mat B2", af Martinus et.al.

I en lille studieretning er der 9 (ni) elever. De har fået karaktererne: 7, 4, 10, 7, 7, 12, 4, 7 og 02 i matematik.

For bedre at kunne overskue datasættet (og desuden til brug for senere udregninger) ordnes datasættet i en mængde, så karaktererne (elementerne i datasættet) stiger numerisk fra venstre mod højre.

Datasættet kaldes for X, og da der er 9 elementer, siger man, at populationen, n, er lig med 9.

 $X = \{2, 4, 4, 7, 7, 7, 7, 10, 12\}, n = 9.$

Definition: Elementer:

Elementer dækker over hvert eneste tal i et datasæt. Altså de værdier, som undersøgelsen kan antage. Bemærk, at observationerne ikke kan "falde udenfor". I eksemplet er det f.eks. ikke muligt at give karakteren 6,5. Eftersom der arbejdes med et diskret datasæt, bør elementerne sorteres efter stigende talværdi, skrevet i en liste fra venstre mod højre.

 $X = \{x_1, x_2, \dots, x_n\}$ $X = \{2, 4, 4, 7, 7, 7, 7, 10, 12\}, n = 9$

Det kan være praktisk – særligt hvis der er mange observationer – at registrere antallet af forekomster af hver enkelt slags observation. Antallet af forskellige slags observationer skrives som k.

I dette eksempel vises antal observationer af de forskellige karakterer, som er givet på holdet. Der er ikke givet "-3" og "00", men resten af karakterskalaen er benyttet. Det betyder, at der er registreret 5 forskellige karakterer. Således er k = 5 i dette eksempel.

Det kan være en rigtig god idé at indsætte værdier i en tabel (regneark) for overskuelighedens skyld.

Karakter	Karakter i Excel-formler
<i>x</i> _{<i>i</i>}	<i>x</i> _{<i>i</i>}
2	2
4	4
7	7
10	10
12	12
k=5	=SAMMENKÆDE("k=";TÆL(A3:A7))

Definition: Hyppighed:

Hyppighed (h_i) defineres som antallet af observationer, som "rammer" en bestemt talværdi. Det kan også sige at være det antal gange en bestemt talværdi forekommer blandt elementerne i et datasæt. Eftersom hyppigheden skal beskrives for hvert enkelt element, skrives det: h_i , hvor det lille *i* repræsenterer hver enkelt talværdi, som de forskellige elementer kan antage.

Hvor mange gange optræder en værdi i observationssættet? En simpel optælling. Bemærk, at observationssættet nogle gange kan stå for sig selv, som f.eks. 7, 7, 10, 7, 7, 12, 4, 7, 02, osv. I dette tilfælde må man selv tælle hvor mange gange f.eks. en bestemt karakter er givet. Resultatet af denne optælling, indsættes i skemaet. Andre gange får man bare at vide at karakteren 7 er givet 4 gange, og i dette tilfælde indsætter man ganske enkelt 4 i skemaet udfor karakteren 7.

Og så er det altid en god idé at summere (lægge sammen) det totale antal observationer som en kontrol for at der er regnet rigtigt og at observationerne er optalt korrekt. I dette eksempel er der 9 observationer i alt.

Bruger man Excel, kan man bruge SUM-funktionen. Stå i den celle, hvor resultatet skal stå. Tryk på "Sum"-knappen. (Den med $\sum på.$) Nu skulle Excel gerne selv vælge cellerne med hyppighederne. Hvis Excel ikke gætter rigtigt, så skal man selv vælge (markere) alle cellerne med hyppighederne ... Afslut med "Enter"-tasten.

Karakter	Hyppighed	Hyppighed i Excel-formler
<i>x</i> _{<i>i</i>}	h_i	h _i
2	1	1
4	2	2
7	4	4
10	1	1
12	1	1
k=5	n=9	=sammenkæde("n=";sum(B3:B7))

Definition: Frekvens (Den relative hyppighed):

Frekvens (f_i) defineres som hyppigheden i procent af det samlede antal elementer i et datasæt.

 $f_i = \frac{\text{Hyppighed af en bestemt talværdi}}{\text{Population}} = \frac{h_i}{n}$

Frekvensen eller den relative hyppighed. Leverer data til pindediagrammet. Udregnes som en bestemt hyppighed divideret med det totale antal observationer. F.eks. for "karakteren 7" gælder: Der er observeret 4 forekomster! 4 divideret med det totale antal observationer (9) = 0,44.... Sådan udregnes frekvensen for alle hyppigheder, hver for sig.

Bruger man Excel, bør man oprette en formel! Det gøres smartest i den øverste celle – altså den øverste af de celler, hvor man skal have udregninger. Her skriver man: "=B2/B7". MEN det vil ikke gå godt, fordi når man kopierer denne formel ned, vil alle referencer flytte med. Men da man ALTID (i de 5 udregninger) skal dividere med det totale antal (9) skal referencen til B7 låses. Dette gøres ved at, når man lige har skrevet "B7" i formlen, trykker på "F4"-tasten. Derved kommer der dollartegn på referencen: "=B2/\$8\$7". Denne formel er korrekt, for når man trækker den ned i fyldhåndtaget, får man det rigtige resultat i alle cellerne.

Karakter	Hyppighed	Frekvens	Frekvens i Excel-formler
<i>x</i> _{<i>i</i>}	h_i	f_i	f_i
2	1	0,11	=B3/sum(\$B\$3:\$\$B7)
4	2	0,22	=B4/sum(\$B\$3:\$\$B7)
7	4	0,44	=B5/sum(\$B\$3:\$\$B7)
10	1	0,11	=B6/sum(\$B\$3:\$\$B7)
12	1	0,11	=B7/sum(\$B\$3:\$\$B7)
k=5	n=9	1,00	=sum(C3:C7)

Definition: Summeret frekvens:

Her tager man frekvenserne og lægger sammen. Data i denne kolonne kan bruges til at finde ud af, hvor stor en andel er mindre end eller lig med en bestemt værdi? (Det er her, fraktilerne kommer ind i billedet.) Denne kolonne forsyner data til et trappediagram, fraktiler og kvartilsættet.

Det at "lægge frekvenserne sammen" vil sige, at man hele tiden summerer for hver enkelt observation. Begynd øverst og husk at et trappediagram i denne forbindelse ALTID går opad og går mellem 0 og 1. Således vil man i eksemplet se, at for den øverste række giver resultatet: 0+0,11=0,11.

Række nr. 2 giver resultatet: (sidste resultat)+(frekvens for næste obs.) = 0,11+0,22=0,33.

Række nr. 3 giver resultatet: (sidste resultat)+(frekvens for næste obs.) = 0,33+0,44=0,77. Osv...

Bemærk, at resultatet i 3. række (karakteren 7) fortæller os, at i 77 % af de givne karakterer blev der højst givet 7, mens der (aflæses i 2. række) at 33 % af karaktererne var højst 4 osv. Det er derfor, at det er i det summerede trappediagram, at man aflæser fraktiler og kvartiler.

Bruger man Excel, kan man lave en formel. I eksemplet i den øverste række indtastes: "=C3". Det er ikke nødvendigt med mere, da der jo ikke er nogle rækker oven over, som kan bidrage til resultatet.

I række 2 er det lidt mere besværligt. Tast "=D3+C4". Denne formel vil tage den næste frekvens og lægge til det forrige resultat. Denne formel kan trækkes ned i fyldhåndtaget, og vil give det rigtige resultat for hver enkelt række.

Karakter	Hyppighed	Frekvens	Summeret frekvens	Summeret frekvens i Excel-formler
<i>x</i> _{<i>i</i>}	h_i	f_i	F_{i}	F_i
2	1	0,11	0,11	=C3
4	2	0,22	0,33	=D3+C4
7	4	0,44	0,78	=D4+C5
10	1	0,11	0,89	=D5+C6
12	1	0,11	1,00	=D6+C7
k=5	<i>n</i> =9	1,00		

Alternativt (hvilket er marginalt mere besværligt), kan man udregne den akkumulerede hyppighed og benytte den til at udregne den summerede frekvens.

Ordet "Akkumuleret" er præcis det samme som "summeret".

Karakter	Hyppighed	Summeret hyppighed	Summeret hyppighed i Excel-formler	Frekvens	Summeret frekvens	Summeret frekvens i Excel-formler
<i>x i</i>	h_i	H_i	H_i	f_i	F_{i}	F_i
2	1	1	=B3	0,11	0,11	=C3/SUM(\$B\$3:\$B\$7)
4	2	3	=C3+B4	0,22	0,33	=C4/SUM(\$B\$3:\$B\$7)
7	4	7	=C4+B5	0,44	0,78	=C5SUM(\$B\$3:\$B\$7)
10	1	8	=C5+B6	0,11	0,89	=C6/SUM(\$B\$3:\$B\$7)
12	1	9	=C6+B7	0,11	1,00	=C7/SUM(\$B\$3:\$B\$7)
k=5	<i>n</i> =9			1,00		

Definition: Produkt:

Produkterne bruges til at finde middelværdien (μ) – også kaldet middeltallet eller gennemsnittet. Produktet bruges mest til at finde middelværdi (=gennemsnit). Hver række bidrager til produktet med værdien (i dette tilfælde karakteren) ganget med hyppigheden. Dette skyldes, at når f.eks. karakteren 7 er givet 4 gange, så er den samlede "vægt i regnskabet" at der i alt er givet 28 karakterpoints ved hjælp af karakteren 7.

Produktet $x_i \cdot h_i$ giver den enkelte talværdis vægt i kraft af antallet af gange talværdien er blevet observeret.

Bruger man Excel, kan man lave en formel. I eksemplet i den øverste række indtastes: "=A3*B3". Afslut med "Enter"-tasten. Denne formel kan man trække direkte ned i fyldhåndtaget for at lave resten af rækkerne.

I eksemplet er karakteren (0)2 er givet én gang. I det samlede regnskab giver dette <u>2</u> karakterpoints. 2 gange er der givet karakteren 4. Dette bidrager til det samlede regnskab med 2 gange $4 = \underline{8}$. etc. Når man så har fundet produktet for hver værdi, lægges disse sammen. Det vil altså sige, at man har fundet ud af, at der i hele undersøgelsen er givet 60 karakterpoints.

Bruger man Excel, kan man igen bruge SUM-funktionen. Stå i den celle, hvor resultatet skal stå. Tryk på "Sum"-knappen. (Den med $\sum på$.) Nu skulle Excel gerne selv vælge cellerne med hyppighederne. Hvis Excel ikke gætter rigtigt, så skal man selv vælge (markere) alle cellerne med hyppighederne.. Afslut med "Enter"-tasten.

Da det vides, at der er 9 elever divideres 60 med 9 for at finde middeltallet $\mu = 6,67$.

Skal man beregne middelværdien, lægger man alle produkterne sammen. Dette tal vil i dette eksempel være den totale sum af karakterpoints, som er givet.

Dette skrives matematisk:

$$x_1 \cdot h_1 + x_2 \cdot h_2 + \ldots + x_n \cdot h_n = \sum_{i=1}^n x_i \cdot h_i$$
 eller i dette eksempel: $\sum_{i=1}^5 x_i \cdot h_i$, da der er 5 forskellige (anvendte) karakterer.

Middelværdien skrives da som:

$$\mu = \frac{\sum_{i=1}^{n} x_i \cdot h_i}{n} \text{ eller i dette eksempel: } \mu = \frac{\sum_{i=1}^{5} x_i \cdot h_i}{9} = \frac{2 \cdot 1 + 4 \cdot 2 + 7 \cdot 4 + 10 \cdot 1 + 12 \cdot 1}{9} = \frac{2 + 8 + 28 + 10 + 12}{9} = \frac{60}{9} \approx 6,67.$$

Karakter	Hyppighed	Summeret hyppighed	Frekvens	Summeret frekvens	Produkt	Produkt i Excel-formler	
<i>x</i> _{<i>i</i>}	h_i	H_i	f_i	F_{i}	$x_i \cdot h_i$	$x_i \cdot h_i$	
2	1	1	0,11	0,11	2	=A3*B3	
4	2	3	0,22	0,33	8	=A4*B4	
7	4	7	0,44	0,78	28	=A5*B5	
10	1	8	0,11	0,89	10	=A6*B6	60
12	1	9	0,11	1,00	12	=A7*B7	$\mu = \frac{60}{2} \approx 6.0$
k=5	n=9		1,00		60	=SUM(G3:G7)	9

Alternativt kan man beregne middelværdien vha. et andet produkt

Man kan også finde frem til middeltallet på en anden måde. I stedet for at gange værdien med hyppigheden, lægge alle resultaterne sammen og til sidst dividere resultatet med antallet af observationer (som blev gjort i 5. kolonne), kan man lige så godt tage værdierne og gange med frekvensen. Dette svarer jo til, at man dividerer med antallet af observationer for hver enkelt række med det samme. Dette er den mest benyttede måde – i hvert fald i Excel, fordi man sparer en udregning i den anden ende. Lægger man alle produkterne sammen i 6. kolonne, får man nemlig direkte udregnet middeltallet.

Dette skrives matematisk:

 $\mu = x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n = \sum_{i=1}^n x_i \cdot f_i \text{ eller i dette eksempel:}$ $\mu = \sum_{i=1}^5 x_i \cdot f_i = 2 \cdot 0,11 + 4 \cdot 0,22 + 7 \cdot 0,44 + 10 \cdot 0,11 + 12 \cdot 0,11 = 0,22 + 0,88 + 3,08 + 1,1 + 1,32 = 6,6$

Karakter	Hyppighed	Summeret hyppighed	Frekvens	Summeret frekvens	Produkt	Produkt i Excel-formler
<i>x</i> _{<i>i</i>}	h_i	\overline{H}_{i}	f_i	F_i	$x_i \cdot f_i$	$x_i \cdot f_i$
2	1	1	0,11	0,11	0,2222	=A3*D3
4	2	3	0,22	0,33	0,8889	=A4*D4
7	4	7	0,44	0,78	3,1111	=A5*D5
10	1	8	0,11	0,89	1,1111	=A6*D6
12	1	9	0,11	1,00	1,3333	=A7*D7
k=5	n=9		1,00		6,6667	=SUM(F3:F7)

Således er der lavet en tabel, som giver det fulde overblik over de ting, som skal udregnes:



Definition: Maksimum:

Maksimum (Maks) defineres som den største talværdi, som optræder blandt elementerne i et datasæt.

$X = \{2, 4, 4, 7, 7, 7, 7, 10, 12\}, n = 9$

Maks = 12

Idet elementet (12) er den største talværdi blandt de ni elementer, så er Maks = 12. Hvis der havde været givet f.eks. 4 12-taller, så ville *Maks* stadig være lig med 12. *Maks* er en absolut værdi.

For fremtidige eksempler, vil de ni celler som indeholder datasættet i Excel

være navngivet til "*Datasæt*". Ellers kan man skrive de ni celler som f.eks.: A1:A9. Excel-eksempler er primært konstrueret til at være generiske. Dvs., at i dette (meget) simple eksempel, så vil det virke meget besværligt at sætte så meget i gang for så lidt. Man skal dog tænke på, at datasæt ofte vil være af en ganske anden – og noget større – størrelse.

I Excel:

=MAKS(*Datasæt*) (På forfatterens computer, virker det IKKE, hvis man bruger MAX). Denne funktion hedder "MAX ..." på engelsk.

Definition: Minimum:

Minimum (Min) defineres som den mindste talværdi, som optræder blandt elementerne i et datasæt.

 $X = \{2, 4, 4, 7, 7, 7, 7, 10, 12\}, n = 9$

Min = 2

Idet elementet (2) er den mindste talværdi blandt de ni elementer, så er Min = 2. Hvis der havde været givet f.eks. 5 02-taller (Bemærk, at der kun regnes med 2. Et foranstillet 0 er ikke væsentligt i matematikken.), så ville *Min* stadig være lig med 2. *Min* er en absolut værdi.

I Excel:

=MIN(*Datasæt*) Denne funktion hedder "MIN ..." på engelsk.

Definition: Middelværdi (μ) :

Middelværdien, der skrives som symbolet μ , defineres som gennemsnittet af alle populationens talværdier, som optræder i et datasæt. Er der tale om en stikprøve i stedet for en population, benyttes symbolet \overline{x} .

$$X = \{x_1, x_2, \ldots, x_n\}$$

$$\mu = \frac{\sum_{i=1}^{n} x_{i}}{n} = \frac{x_{1} + x_{2} + \ldots + x_{n}}{n}$$

 $X = \{2, 4, 4, 7, 7, 7, 7, 10, 12\}, n = 9$

$$\mu = \frac{2 + 4 + 4 + 7 + 7 + 7 + 7 + 10 + 12}{9} = \frac{60}{9} \approx 6,7$$

Har man allerede lavet skemaet/regnearket, som indeholder observationerne, hyppighederne og frekvenserne, er det allerede beskrevet, hvordan man kan udregne middelværdien.

Middeltallet er dog letpåvirkeligt over for meget små eller meget store værdier. Eksempel kan vise, at hvis der i undersøgelsen pludselig indgår en karakter på -3, så vil middeltallet falde kraftigt, og man kan her vurdere om det er rimeligt, da karakteren -3 ikke er (voldsomt) repræsentativ i datasættet. (Den falder lidt uden for det normale.)

I Excel:

=MIDDEL(*Datasæt*) Denne funktion hedder "AVERAGE ..." på engelsk.

Definition: Variationsbredde:

Variationsbredden defineres som afstanden mellem minimum og maksimum.

Variationsbredden = Maks – Min

Variations bredden = 12 - 2 = 10

I Excel:

=MAKS(*Datasæt*) –MIN(*Datasæt*)

Definition: Median:

Medianen er den værdi, som deler et ordnet observationssæt i to lige store dele. Dvs. hvis alle elementer i et datasæt ordnes efter stigende talværdier, så er medianen det midterste elements talværdi. Er der et lige antal observationer, så er medianen lig med gennemsnittet af de to midterste elementer.

$X = \{2, 4, 4, 7, 7, 7, 7, 10, 12\}, n = 9$

Da der er et ulige antal observationer (9), må det femte element fra venstre (eller fra højre) være det midterste element. Værdien af det femte element er 7. Derfor er medianen = 7.

I Excel:

=MEDIAN(*Datasæt*) Denne funktion hedder "MEDIAN ..." på engelsk.

Definition: Skævhed:

Hvis middelværdien (μ) er større end medianen, siger man at datasættet er højreskævt. Hvis middelværdien (μ) er mindre end medianen, siger man at datasættet er venstreskævt.

Middelværdien (μ) er udregnet til at være 6,7. Medianen er bestemt til at være 7.

Da middelværdien er mindre end medianen, er datasættet venstreskævt. Det betyder samtidig at størstedelen af datasættet ligger til højre for medianen.

I Excel:

Der eksisterer faktisk en funktion i Excel, som hedder "SKÆVHED". Ser man i hjælpefilen for Excel, så er funktionen forklaret som: "Returnerer skævheden for en stokastisk variabel. Skævhed er den grad af asymmetri, der er i en fordeling omkring dens middelværdi. Positiv skævhed indikerer en fordeling med en asymmetrisk hale, der hælder mod mere positive værdier. Negativ skævhed indikerer en fordeling med en asymmetrisk hale, der hælder mod mere negative værdier."

Så det er altså ikke DÉN, man skal bruge.

=SKÆVHED(Datasæt)

Til gengæld findes der også funktionen: "SKÆVHED.P", som hjælpefilen beskriver som: "Returnerer skævheden af en distribution baseret på en population: en karakteristik af graden af asymmetri for en distribution omkring dens middelværdi".

=SKÆVHED.P(Datasæt)

Her er skævheden beregnet som: $v = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i - \overline{x}^3}{\sigma}$. Denne formel vil naturligvis returnere et tal.

Grundlaget for denne formel er ikke kendt, så konklusionen er, at et tilstrækkeligt resultat vil være at vurdere, om datasættet er højre- eller venstreskævt.

Definition: Typetal:

Typetallet er den eller de observationer, som forekommer oftest (flest gange). Det vil automatisk være det eller de tal, som har den største hyppighed.

 $X = \{2, 4, 4, 7, 7, 7, 7, 10, 12\}, n = 9$

Da karakteren 7 har hyppigheden 4, og der ikke er andre tal med den (eller større) hyppighed,

så har karakteren 7 den største hyppighed alene.

Derfor er typetallet = 7.

I Excel:

=HYPPIGST(*Datasæt*) Denne funktion hedder "MODE ..." på engelsk.

Definition: Kvartiler:

Kvartiler er nogle ganske bestemte fraktiler (se senere). Kvartiler opdeler datasættet i 4 (fire) lige store dele.

Der er således 5 kvartiler, som benævnes Q_0 , Q_1 , Q_2 , Q_3 og Q_4 .

 Q_0 er det samme som datasættets minimum. Q_2 er det samme som datasættets median. Q_4 er det samme som datasættets maksimum.

Da maksimum og minimum normalt er underforstået, så er det normalt kun Q_1 , Q_2 og Q_3 der skal findes, når det bliver spurgt til kvartilsættet. Medtager man Q_0 og Q_4 , kaldes det for "Det udvidede kvartilsæt".

Hvis der er et LIGE antal elementer i datasættet, så beregnes Q_1 som medianen af den første halvdel af datasættet, og Q_3 beregnes som medianen af den sidste halvdel af datasættet.

Hvis der er et ULIGE antal elementer i datasættet, medtages medianen ikke i hverken den første eller sidste halvdel af datasættet, og kvartilerne beregnes som tidligere beskrevet.

(Det skal nævnes, at der desværre ikke er enighed i verden om, hvordan man beregner kvartilerne, så der kan være andre acceptable metoder, som vil give et anderledes resultat.)

I eksemplet er der 9 (ni) elementer. Så derfor er der et ulige antal elementer i datasættet.

Første halvdel af datasættet (eksklusiv den midterste værdi, da der er et ulige antal elementer) består af $\{2, 4, 4, 7\}$, og 1. kvartil (Q_1) beregnes derfor som:

$$Q_1 = \frac{4+4}{2} = 4$$

Sidste halvdel af datasættet (eksklusiv den midterste værdi, da der er et ulige antal elementer) består af $\{7, 7, 10, 12\}$, og 3. kvartil (Q_3) beregnes derfor som:

$$Q_3 = \frac{7+10}{2} = 8,5$$

I Excel:									
	=KVAR	ΓIL.I	MEDTA	G(Matrix;kvartilnumme	r (0-4))				
	F.eks. =KVARTIL.MEDTAG(Datasæt;3), hvilket vil returnere tallet 7. I forhold til tidligere udregning er dette ikke korrekt!								
	=KVARTIL (Matrix;kvartilnummer (0-4)) Denne funktion hedder "QUARTILE" på engelsk.								
	F.eks. =KVARTIL (Datasæt;3), hvilket vil returnere tallet 8,5. Dette resultat stemmer i forhold til tidligere udregning, så vil man udregne resultaterne i Excel, anbefales det at benytte =KVARTIL (Matrix;kvartilnummer (0-4)).								
	Det kan b skal inklu	oetal idere	e sig at l es må væ	ave en lille tabel i Excel ere op til den enkelte opg	, hvor k gave.	vartilerne beregnes. Hvilke	af dem, som		
			Е	F	G	Formel i kolonne 'G'			
			:		:				
	25		0	Q ₀ - 0. Kvartil: (Min)	2	=KVARTIL(Datasæt;E25)			
	26		1	Q ₁ - 1. Kvartil:	4	=KVARTIL(Datasæt;E26)			
	27		2	Q ₂ - 2. Kvartil: (Median)	7	=KVARTIL(Datasæt;E27)			
	28		3	Q ₃ - 3. Kvartil:	7	=KVARTIL(Datasæt;E28)			
	29		4	Q ₄ - 4. Kvartil: (Maks)	12	=KVARTIL(Datasæt;E29)			
	Her er id længst ud	een 1 le til	ned talle højre. D	ene i kolonne 'E', at ang Da en navngiven referenc	ive det a e ALTI	andet argument til formlen, D er absolut, kan man nøjes	som er vist s med at		
	skrive for	rmle	n i celle	G2 og derefter trække d	en ned.				

Definition: Kvartilbredde:

Kvartilbredden defineres som afstanden mellem1. kvartil (Q_1) og 3. kvartil (Q_3).

Kvartilbredden = Q3 - Q1 = 8, 5 - 4 = 4, 5

I Excel:

=KVARTIL(*Datasæt*;E28)- KVARTIL(*Datasæt*;E26)

Definition: Boksplot:

Et boksplot er et plot, som viser de fem kvartiler, hvor kvartilbredden er fremhævet.

Da man ofte er interesseret i at få at vide, hvor den midterste halvdel af et datasæt ligger, så er et boksplot en god måde at vise det på. Særligt, hvis man skal sammenligne to datasæt er boksplot en rigtig godt værktøj.

Laver man et boksplot i Excel, indtaster man elementerne i en kolonne, markerer elementerne og indsætter en graf. Denne graf skal være af typen "Kasse med hale, som findes under menuen: "Alle diagrammer". I modsætning til de fleste andre programmer, vil boksplottet vises lodret i Excel. Det er fint! Det betyder ingenting, hvilken vej det vender.



Definition: Fraktiler:

En fraktil er en procentgrænse, hvor andelen af elementer, som er mindre end eller lig med fraktilen, er mindst lige så stor som fraktilens procentværdi.

Fraktilen beregnes som værdien for x_i , givet: $i \ge \frac{P}{100} \cdot n$, hvor *i* er det mindste tal, som opfylder uligheden.

P er fraktilens procent, og n er - som sædvanlig - populationen, altså antallet af elementer i datasættet.

(Dette er én blandt flere accepterede måder at beregne fraktiler på. De forskellige måder giver forskellige resultater, men har alle det tilfælles, at de bliver mere ens jo flere elementer, der er i et datasæt.

0,1-fraktilen er altså den værdi, hvor det først sker, at *i* er større end eller lig med 0,1.

0,8-fraktilen er den værdi hvor *i* rammer (eller overskrider) 0,8.

Dette kan aflæses i den summerede frekvens i tabellen eller direkte i trappediagrammet (beskrives senere). I trappediagrammet kan en fraktil findes ved at afmærke fraktilen op ad frekvens-aksen og derfra tegne en vandret streg ud. Det "trappetrin", som den vandrette streg støder på, svarer til den fraktilens værdi.

Der er nogle fraktiler, som er lidt specielle. Det er dem, man benytter oftest, og de har endda specielle navne:

0,0-fraktil	0. kvartil	Minimum
0,25-fraktil	1. kvartil	Nedre kvartil
0,50-fraktil	2. kvartil	Median
0,75-fraktil	3. kvartil	Øvre kvartil
1,0-fraktil	4. kvartil	Maksimum

De tre (midterste) kvartiler kaldes tilsammen for kvartilsættet eller (for alle fem) det udvidede kvartilsæt.

 $X = \{2, 4, 4, 7, 7, 7, 7, 10, 12\}, n = 9$

5 % fraktilen bestemmes:

$$i \ge \frac{5\%}{100\%} \cdot 9 = 0,45 \Longrightarrow \underline{i=1}$$

Da der skal rundes op til nærmeste hele element, skal dette skal forstås som det i = 1. element - altså x_1 .

5 % fraktilen er altså det samme som værdien af $x_1 = 2$.

10 % fraktilen bestemmes:

$$i \ge \frac{10\%}{100\%} \cdot 9 = 0,90 \Longrightarrow \underline{i=1}$$

Da der skal rundes op til nærmeste hele element, skal dette skal forstås som det i = 1. element - altså x_1 .

10 % fraktilen er altså det samme som værdien af $x_1 = 2$.

Dette svarer til, at 10 % af karaktererne, som er givet er 02 eller derunder.

80 % fraktilen bestemmes:

 $i \ge \frac{80\%}{100\%} \cdot 9 = 7,20 \Longrightarrow \underline{i=8}$

(Da der skal rundes op til nærmeste hele element,

skal dette skal forstås som det i = 8. element - altså x_{g} .

80 % fraktilen er altså det samme som værdien af $x_8 = 10$.

Dette svarer til, at 80 % af karaktererne, som er givet er 10 eller derunder.

I Excel:											
	=FRAK	TIL.	MEDTAG(Matrix;K	(0-1))							
	F.eks. =FRAKTIL.MEDTAG(<i>Datasæt</i> ;0,1), hvilket vil returnere tallet 3,6.1 forhold til										
	tidligere	udre	egning er dette ikke k	orrekt!							
	=FRAKTIL.UDELAD(Matrix;kvartilnummer (0-1))										
	F.eks. =	FRA	KTIL.UDELAD(Dat	<i>asæt</i> :0.1). hv	vilket vil r	eturnere tallet 2. De	ette resultat				
	stemmer	i fo	rhold til tidligere udre	egning, så vil	man udro	egne resultaterne i H	Excel, anbefales				
	det at be	nytte	e =FRAKTIL.UDELA	AD(Matrix;k	vartilnum	mer (0-1)).					
	Det kan	beta	le sig at lave en lille t	abel i Excel,	hvor frak	tilerne beregnes. H	vilke af dem, som				
	skal bere	egne	s må være op til den e	enkelte opgav	/e.						
		Б	E	G							
		÷		:	FRAKTIL	FRAKTIL.MEDTAG	FRAKTIL.UDELAD				
	25 ····	0,05	5 % Fraktil	#NUM!	2,8	2,8	#NUM!				
	26 ····	0,1	10 % Fraktil	2	3,6	3,6	2				
	27	0,8	80 % Fraktil	10	8,2	3,6	10				
	D		11 1 0 1			1, 0, 1, 11, 1	0 1.4				
	Det ses,	at E	xcel kan beregne frak	tiler med ikk	e mindre	end tre forskellige i	funktioner.				
	FKAKI Tilaunal	IL, F	KAKTIL.MEDIAG	Og FKAKII	L.UDELA	AD. ARTH MEDTAC	ag da giyar bagga				
	forkorto	roou	de er der ingen forski Itatar i forhold til da i	el pa rkak i	IL Og FK	AKTIL.WEDTAG	, og de giver begge				
	Sormon	dori	mod n ⁸ ED A KTIL II	DELAD of	sonn er 10 viver den	de rigtige regultator	to ud of tro				
	sel man	Lda	n forsta udragning fo	DELAD, sa g	giver della	ae figuge fesultater	fail anatår fordi				
	Erroal ile		n lørste udregning, 18	to that from do in		se: #INUIVI. Denne I	leji opstar tordi				
	EXCELIK forbinda	ке ка 100)	an interpoiere resulta	dornunktor	alteå nor	nar. Typisk vir deni					
	Domas	1se)	opsia i de eksireme y	$\frac{1}{2}$	ansa nær	veu 0 % og ved 10	υ %0.				
	Denne I	unkti	ion neader PERCEN	eritte j ba	i engelsk.						

Definition: Varians:

For en **population**, *n*, beregnes variansen som den gennemsnitlige kvadratafvigelse.

$$VAR_{p} = \frac{1}{n} \cdot \sum_{i=1}^{n} h_{i} \cdot \left(x_{i} - \mu\right)^{2} = \frac{h_{1} \cdot \left(x_{1} - \mu\right)^{2} + h_{2} \cdot \left(x_{2} - \mu\right)^{2} + \dots + h_{n} \cdot \left(x_{n} - \mu\right)^{2}}{n}, \text{ hvor } \mu \text{ er populationens middelværdi.}$$

For en **Stikprøve**, beregnes variansen som:

$$VAR_{s} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} h_{i} \cdot \left(x_{i} - \overline{x}\right)^{2} = \frac{h_{1} \cdot \left(x_{1} - \overline{x}\right)^{2} + h_{2} \cdot \left(x_{2} - \overline{x}\right)^{2} + \dots + h_{n} \cdot \left(x_{n} - \overline{x}\right)^{2}}{n-1}, \text{ hvor } \overline{x} \text{ er stikprøvens middelværdi.}$$

Da de enkelte observationer kan være både større end middelværdien og mindre end middelværdien, ville man ikke kunne bruge deres sum som et spredningsmål. Derfor har man valgt at kvadrere (sætte dem i anden potens) afvigelserne og lægge dem sammen.

Dette refereres ofte til som SAK, som står for Summen af Afvigelsernes Kvadrater.

Ved at dividere med antallet af elementer i datasættet får man den gennemsnitlige kvadratafvigelse, som kaldes for variansen.

Da man ved en stikprøve ikke kender den sande middelværdi, dividerer man i stedet med et mindre tal, så variansen bliver større.

Ved meget store datasæt på f.eks. 10000 elementer, har det ikke den store indflydelse på variansen om man dividerer med 10000 eller 9999. Det viser sig også at være ok, da stikprøvens middelværdi må formodes at være meget tæt på den sande middelværdi ved så stor en stikprøve. Heraf følger det, at ved en lille stikprøve, så har det imidlertid en meget stor betydning.

 $X = \{2, 4, 4, 7, 7, 7, 7, 10, 12\}, n = 9$

Ser man på datasættet som kun beskrivende for den lille studieretnings karakterer i matematik, så beregnes variansen som en population:

$$VAR_{p} = \frac{h_{1} \cdot (x_{1} - \mu)^{2} + h_{1} \cdot (x_{2} - \mu)^{2} + h_{1} \cdot (x_{3} - \mu)^{2} + h_{1} \cdot (x_{4} - \mu)^{2} + h_{1} \cdot (x_{5} - \mu)^{2}}{n}$$

$$VAR_{p} = \frac{1 \cdot (2 - 6, 67)^{2} + 2 \cdot (4 - 6, 67)^{2} + 4 \cdot (7 - 6, 67)^{2} + 1 \cdot (10 - 6, 67)^{2} + 1 \cdot (12 - 6, 67)^{2}}{9}$$

$$VAR_{p} = \frac{(-4, 67)^{2} + 2 \cdot (-2, 67)^{2} + 4 \cdot (0, 33)^{2} + (5, 33)^{2}}{9}$$

$$VAR_{p} = \frac{21,8089 + 2 \cdot 7,1289 + 4 \cdot 0,1089 + 11,0889 + 28,4089}{9} = \frac{21,8089 + 14,2578 + 0,4356 + 11,0889 + 28,4089}{9}$$

$$VAR_{p} = \frac{76,00}{9}$$

$$VAR_{p} = \frac{76,00}{9}$$

$$VAR_{p} = 8,4$$
I Excel:

=VARIANS.P(Matrix) Denne funktion hedder "VAR.P ..." på engelsk. =VARIANS.P(Datasæt)

Bemærk, at de kvadrerede afvigelser skal multipliceres med hyppigheden af de enkelte elementer.

Betragter man derimod datasættet som en stikprøve, som skal beskrive alle eleverne i klassen og ikke kun den lille studieretning, så kender man ikke den sande middelværdi, og man skal derfor dividere med 9-1=8 i stedet for 9.

$VAR_{ps} = \frac{h_1 \cdot (x_1 - \mu)^2 + h_1 \cdot (x_2 - \mu)^2 + h_1 \cdot (x_3 - \mu)^2 + h_1 \cdot (x_4 - \mu)^2 + h_1 \cdot (x_5 - \mu)^2}{n - 1}$	
n-1	
$VAR_{s} = \frac{1 \cdot (2 - 6, 67)^{2} + 2 \cdot (4 - 6, 67)^{2} + 4 \cdot (7 - 6, 67)^{2} + 1 \cdot (10 - 6, 67)^{2} + 1 \cdot (12 - 6, 67)^{2}}{9 - 1}$	
$VAR_{s} = \frac{(-4,67)^{2} + 2 \cdot (-2,67)^{2} + 4 \cdot (0,33)^{2} + (3,33)^{2} + (5,33)^{2}}{8}$	
1	
$VAR_{s} = \frac{21,8089 + 2 \cdot 7,1289 + 4 \cdot 0,1089 + 11,0889 + 28,4089}{8} = 21,8089 + 14,2578 + 0,4989 + 0,4989 $	$\frac{356+11,0889+28,4089}{8}$
\$	0
$VAR_s = \frac{76,00}{8}$	
\Diamond	
$\underline{VAR_s = 9,5}$	
I Excel:	
=VARIANS.S(Matrix) ' Denne funktion hedder "VAR.S" på engelsk.	
=VARIANS.S(<i>Datasæt</i>)	

Variansen er kvadratet på spredningen. Så hvis man vælger at tage kvadratroden af variansen, er det naturligvis fordi man gerne vil have et spredningsmål med samme enhed som de elementer, som er i datasættet.

Spredningen kaldes ofte for den gennemsnitlige afvigelse.

I statistik benytter man ofte græske bogstaver for de sande værdier. Hvis der beregnes på en hel population, vil de beregnede værdier være sande, og derfor bruges det græske bogstav μ (my), som er det tilsvarende bogstav for det latinske 'm', som igen er valgt for middelværdi. På samme måde er det græske bogstav σ (sigma), som er det tilsvarende bogstav for det latinske 's', som igen er valgt for spredning.

Hvis datasættet er normalfordelt, så vil:

95 % af datasættet vil ligge i intervallet : $\mu \pm 1,96 \cdot \sigma$ 95,4 % af datasættet vil ligge i intervallet : $\mu \pm 2 \cdot \sigma$ 90 % af datasættet vil ligge i intervallet : $\mu \pm 1,645 \cdot \sigma$

Hvis datasættet IKKE er normalfordelt, så vil (ifølge Chebyshevs teorem):

Mindst 75 % af datasættet ligge i intervallet : $\mu \pm 2 \cdot \sigma$

Mindst 89 % af datasættet ligge i intervallet : $\mu \pm 3 \cdot \sigma$

Mindst 94 % af datasættet ligge i intervallet : $\mu \pm 4 \cdot \sigma$

⁽Normalfordelingen er en af de vigtigste sandsynlighedsfordelinger og benævnes også Gaussfordelingen. Den er kontinuert og kan principielt omfatte alle reelle tal. Den er symmetrisk og kan entydigt bestemmes ved observationssættets middelværdi og varians.

Normalfordelingen bruges som en "model" af hvordan et stort antal statistiske elementer fordeler sig omkring deres gennemsnit. Hvis man for eksempel måler højden eller vægten af hver enkelt person i en stor, ensartet gruppe af mennesker, vil de fleste ligge omkring et vist gennemsnit, mens meget store eller små personer er mere sjældne. Tegner man resultaterne ind i en graf, med højde eller vægt hen ad den vandrette abscisse-akse og f.eks. procent op ad den lodrette ordinat-akse, får grafen den karakteristiske klokkeformede facon, der kan være mere eller mindre fladtrykt eller sammenpresset. Toppunktet ligger ved hele gruppens gennemsnit, markeret på den vandrette akse, og "bulen" omkring toppunktet svarer til det flertal af de målte personer der ligger tæt på gennemsnitet. På begge sider af dette store midterfelt af "gennemsnitsfolk" falder kurven, som tegn på at jo længere væk man kommer fra gennemsnittet, jo sjældnere støder man på folk med sådan en højde eller vægt. Disse dele af kurven omtaler matematikere ofte som haler (ental: en hale).)

Definition: Spredning (Standardafvigelsen):

Spredningen er kvadratroden af variansen.

For en **population**, beregnes spredningen som :

$$\sigma = \sqrt{VAR_p}$$
 eller $\sigma^2 = VAR_p$.

For en **Stikprøve**, beregnes variansen som:

 $s = \sqrt{VAR_s}$ eller $s^2 = VAR_s$.

 $X = \{2, 4, 4, 7, 7, 7, 7, 10, 12\}, n = 9$

Igen er der forskel. Betragtes datasættet som en population, fås følgende spredning:

$$\sigma = \sqrt{VAR_p} = \sqrt{8,4} \approx 2,9$$

I Excel:

=STDAFV.P(Matrix) Denne funktion hedder "STDEV.P ..." på engelsk.

= STDAFV.P(*Datasæt*)

Ser man derimod datasættet som en stikprøve, skrives spredningen således:

$$s = \sqrt{VAR_s} = \sqrt{9,5} \approx 3,1$$

I Excel:

=STDAFV.S(Matrix) Denne funktion hedder "STDEV.S ..." på engelsk.

= STDAFV.S(*Datasæt*)

Det vil være en rimelig antagelse at karaktererne er normalfordelte. Idet karaktererne er normalfordelte, vil man kunne forvente at 95,4 % af elementerne ligger i intervallet:

 $\mu \pm 2 \cdot \sigma$

I dette eksempel bliver det til:

 $6,67 \pm 2 \cdot 2,9$, hvilket svarer til intervallet: [0,8;12,5].

Da både minimum og maksimum er inkluderede i dette interval må hele datasættet ligge i intervallet.

Grunden til at det lige netop er forventeligt at 95,4 % af elementerne ligger i intervallet er, at X er højst 2 standardafvigelser fra μ . Denne forklaring – inklusive bevis – er ikke yderligere beskrevet i dette notat.

Diagrammer for diskrete variable

Typisk vælger man at afbilde data for diskrete variable i to diagrammer. Et **pindediagram** og et **trappediagram**.

I pindediagrammet afsætter man værdierne ud af abscisseaksen og frekvensen op ad ordinataksen. Bemærk at der gælder for BÅDE pindediagrammet og trappediagrammet at ordinataksen ALTID går mellem 0 og 1. (Trappediagrammet går netop fra 0 til 1, medens pindediagrammet – som i nedenstående eksempel – kan være endnu mindre, afhængigt af frekvensernes størrelse.)



(Her er et diagram, som er lavet i Excel. Bemærk, at i egenskaber for dataserie er mellemrummet mellem søjlerne sat til størst muligt (mellemrumsbredde = 500), for at gøre søjlerne så smalle som muligt (pinde). Det er således ikke muligt (I Excel) at lave et helt korrekt pindediagram.

I trappediagrammet nedenunder vises den summerede frekvens. Her afsætter man værdierne ud af abscisseaksen. Sørg for at have den mindste værdi ved ordinataksen. Værdien for den summerede frekvens aftegnes som en vandret streg fra værdiens begyndelse til næste værdi. Her gentages processen for den næste summerede frekvens. Til sidst forbindes de vandrette streger med lodrette streger, så det kommer til at ligne en trappe. (Heraf navnet... [©]) Bemærk desuden, at trappediagrammet altid vil gå opad, da man jo altid lægger sammen. Trappetrinene vil ALTID blive højere og højere, jo længere man kommer hen mod højre. De kan ALDRIG blive lavere!



Her er et trappediagram, som er lavet i Excel. Bemærk, at modsat pindediagrammet, så er mellemrummet mellem søjlerne sat til 0, for at få søjlerne til at støde op mod hinanden.

Det er lidt en tilsnigelse at kalde det et trappediagram, da alle de lodrette streger går hele vejen ned mellem søjlerne. I Excel kan man sagtens fjerne stregerne, hvis man gerne vil.

Følgende er en oversigt over de deskriptorer, som beskriver datasættet $X = \{2, 4, 4, 7, 7, 7, 7, 10, 12\}, n = 9$

Karakter	Hyppighed	Frekvens	Summeret frekvens	Produkt	Produkt
<i>x</i> _{<i>i</i>}	h_i	f_i	F_i	$x_i \cdot h_i$	$x_i \cdot f_i$
2	1	0,11	0,11	2	0,22
4	2	0,22	0,33	8	0,89
7	4	0,44	0,78	28	3,11
10	1	0,11	0,89	10	1,11
12	1	0,11	1,00	12	1,33
k=5	n=9	1,00		60	6,67

Populationen:	<i>n</i> =9	Variationsbredde:	10	Kvartilbredde:	4,5	Typetal:	7
Maksimum:	Maks = 12	Median:	7	5 % fraktil:	2	Varians:	$VAR_{\rm p} = 8,4$
Minimum:	Min = 2	1. kvartil:	$Q_1 = 4$	10 % fraktil:	2	Spredning:	$\sigma = 2,9$
Middelværdi:	$\mu = 6,67$	3. kvartil:	$Q_3 = 8,5$	80 % fraktil:	10	Datasættet er	r venstreskævt

Desuden vises de tre grafiske fremstillinger:





Grupperede variable

Som nævnt i indledningen, er det ofte, at de indsamlede data er så forskellige, at man inddeler dem i grupper eller intervaller. I modsætning til de diskrete variable, kan man her tilgodese alle værdier og ikke kun enkelte værdier.

De grupperede variable vil blive forklaret vha. et eksempel. Igen er eksemplet taget fra "HTX Mat B2", af Martinus et.al.

I en klasse med 27 elever er elevernes højder målt (i *cm*). Resultatet af denne måling er som følger – ordnet efter højde.

158, 160, 165, 165, 166, 167, 167, 168, 169, 169, 169, 170, 171, 173, 175, 175, 175, 176, 176, 177, 179, 180, 181, 181, 190, 191 & 195.

Alle de statistiske deskriptorer beregnes på nøjagtig samme måde, som det blev gjort i kapitlet om Diskrete Variable. Der er dog lige et par ting, som skal overvejes nøje, inden man begynder.

Fra og med 158 (*cm*) til og med 195 (*cm*) er der 38 mulige heltalsværdier, så det kan hurtigt blive til en meget stor tabel. Måske ikke noget udregningsproblem med f.eks. Excel, men det er også svært at overskue en så stor tabel. Desuden er der kun 27 elever, så – især hvis man tænker på, at flere elever kan have samme højde – må der være en del højder, som ikke bliver "brugt". Dette siges også at være nul-forekomster i datasættet.



Så det er tydeligt, at dette ikke er nogen god fremstilingsmåde til den slags data.

Man kan i stedet vælge at gruppere datasættet i et antal intervaller. Der er ikke nogen facitliste til, hvor mange intervaller der bør være. Det kan være grupperinger à 5 eller à 10, hvis det passer til data. Nogle synes godt om at lave \sqrt{n} intervaller fordelt over variationsbredden, hvor *n* er antallet af observationer. Denne sidstnævnte metode kan godt blive lidt "kunstig", da man normalt ville forvente grupperinger à 5 eller à 10.

Konklusionen må være, at det er op til det enkelte tilfælde, hvordan det er smartest at inddele variationsbredden, så det giver mest mening.

I dette eksempel grupperes højderne først i intervaller à 10 cm. Sidenhen i intervaller à 5 cm.

Sættes intervalbredden til 10, bliver variationsbredden delt op i følgende 5 intervaller: [150;160], [160;170], [170;180], [180;190] og [190;200].

Bemærk intervalgrænserne! F.eks. 160 er kun med i det nederste interval. Det er sådan set ligegyldigt om en værdi er inkluderet i det nedre eller det øvre interval, så længe værdien kun er med én gang. Ellers ville alle med en højde på 160 *cm* jo blive registreret to gange.

Mange større datasæt er kun tilgængelige som grupperede datasæt, og her vil man ofte komme frem til et andet resultat for deskriptorerne, end hvis man havde adgang til kildedata. Her ved man jo ikke, hvordan data er fordelt i de enkelte intervaller, så her vil man principielt altid lave en fejl. Man antager normalt altid at data er fordelt jævnt i et interval. Ser man f.eks. på intervallet 160 til 170, hvor hyppigheden er 10, så vil man antage at hyppigheden er 5 i både intervallet 160 til 165 og fra 165 til 170, men det fremgår jo tydeligt af ovenstående pindediagram, at det bestemt ikke er tilfældet.

Sætning:

Når kildedata ikke er kendt, antages data at være jævnt fordelt i et interval.

Skemaet, som blev brugt i eksemplet med diskrete variable laves igen.

Lidt om, hvordan skemaet kan laves i Excel, når der er tale om grupperede variable:

I det følgende gives en komplet fremgangsmåde til, hvordan man kan skabe skemaet med grupperede variable. Først og fremmest kan det være en god idé, at der er nem adgang til kildedata. F.eks. er det smart, hvis man udnævner et regneark (et af fanebladene nederst i projektmappen) til at indeholde data. Dette faneblad bør navngives, så det fremgår, hvilke data der er tilgængelige her. Der kan jo være flere datasæt.

• Dobbeltklik på navnet "Ark *x*" for at få mulighed for at skrive et nyt arknavn.



Dernæst er det en god idé (næsten et krav), at man navngiver det område, som data er placeret i.

- Markér det relevante område.
- Klik oppe i navneboksen.
- Skriv navnet på området og afslut med <ENTER>.

		• •												ndik (MIN					
	Hjerr	Induer																	
Set ind •	ж ⊪∋- ∢	Calibri F K U	* 11 * 🖽 * Skriftype	- A	· . =	= = = Just	#> - #⊒ #⊒ sing	8 9 - 9 - 9	tandard 9 - 96 o Tai	- - 158 43	Beting	et Form ng * som ta Typog	atér Cellet bel - rafier	ypografi ,	Slet * Celler	Σ - • -	AT Satisfies	Q ng og ekg "	
abe				fr 1	58														
	158 160 165		¢	0	E		F	G	н	1	J	K	L	м	N	0	P	Q	
	166 167 167 168						Na	vng	giv	cell	eon	nråd	e						
	169 169 169																		
	171 173 175																		
	175 175 176																		
	176 177 179																		
	180 181 181																		
	191 195																		
		E	0																

Denne navngivning betyder, at det fremover ikke er nødvendigt at finde og markere (eller huske og skrive) de celler, som data står i. Nu kan de (i dette eksempel) bare refereres til som: "Tabel".

Et lille tip: Man kan "fremtidssikre" sin projektmappe ved at markere og navngive EN HEL KOLONNE. Derved kan man i fremtidige opgaver bare indsætte sine data i denne kolonne. Så behøver man ikke at finde, indsætte og navngive tingene igen og igen ...

Tilføj nu et nyt regneark. Det er her, hvor selve skemaet kommer til at stå.

- Klik på det lille "+" (plus) nede ved siden af de(t) eksisterende regneark.
- Dobbeltklik på det nye fremkomne faneblad.
- Skriv navnet på arket og afslut med <ENTER>.



Alt er nu klart til at begynde at oprette selve arket. Overskrifter indtastes og der gøres klar til at indsætte data. Bemærk, at kolonnen: "Højde i cm" faktisk består af fem kolonner. Bare rolig! Det kommer ikke til at kunne ses – bortset fra at det kommer til at se virkelig godt ud. Det er kun fordi der er tale om grupperede variable. De fem kolonner skal bruges til at vise (og bruge) intervalgrænserne på en snedig måde. Overskrifterne som hører til den første (de første fem) kolonner formateres med "Flet og centrer"-funktionen.



Bemærk, at intervalgrænserne forberedes med kantede parenteser og semikolonner. Det er jo tanken, at denne projektmappe senere skal kunne anvendes som en skabelon, så jo mere der er lavet på forhånd, desto bedre.

Vil man ændre skrifttyper (f.eks. kursiv skrift eller hævet/sænket skrift) i dele af en tekst, kan det lade sig gøre ved at åbne dialogboksen "Formatér celler" og vælge fanebladet "Skrifttype".

Markér de karakterer, som skal ændres. Åbn dialogboksen "Formatér celler" og vælge fanebladet "Skrifttype". Luk dialogboksen.

Kolonnebredder ændres ved at trække i mellemrummet mellem kolonnebogstaverne.

🖬 😙 · 근 · · Mappe1 (version	2).xlsb[Gendannet auto	ma Michel Mand	ix (MIM - Underviser - U	nord) 🖭	- 🗆 ×		
Filer Hjem Indsæt Sidelayout	Formler Data (Gennemse Vis	Hjælp Acrobat Ç	Fortæl mig det	P₄ Del		
$\begin{array}{c c} & & \\ & &$		Standard • 5	Betinget formatering + Formatér som tabel + Celletypografi +	Formatér *	$\sum \cdot \hat{x} \cdot \cdot$ $\Psi \cdot \cdot$		
oukipsnoider is skriittype	is Justering i	a Iai iai	Typograner	Celler	Realgening		
A1 \rightarrow $X \swarrow f_X$	Højde i cm						
ABCDEF	G	н	I.	J	K L		
1 Højde i cm Intervalmidtpunkt	Interval hyppighed	Intervalfrekvens	Summeret frekvens	Produkt			
$\begin{array}{c} 2 [x_{i-1}; x_i] \qquad m_i \end{array}$	h _i	Ji	r _i	$m_i f_i$			
3 4] :]							
5] ;]							
6];]							
7];]							
8 ;]							
10							
11		,					
12							
🔹 Data Skema 🕂	A Data Skema ÷ i i i i i i i i i i i i i i i i i i						
Markér destinationsområde, og tryk på Enter		ndstillinger f	or visning 🔠 🗉	巴	+ 100 %		

Umiddelbart er den eneste ændring – i forhold til skemaet med de diskrete variable, at der er tilføjet en ekstra kolonne: "Interval midtpunkt: m_i ". Naturligvis udover ændringen i første kolonne, der bruges til at beskrive de forskellige intervaller.

Nu udfyldes skemaet med oplysningerne om intervallerne.

⊟ూం	 Mappe1 (version 	2).xlsb[Gendannet autor	ma Michel Mand	ix (MIM - Underviser - U	nord) 🖪	
Filer Hjem I	ndsæt Sidelayout	Formler Data (Gennemse Vis	Hjælp Acrobat Ç		P₄ Del
Sæt ind - V Udklipsholder	New Ro • 11 • A	A Justering	Standard • 5	Betinget formatering + Formatér som tabel + Celletypografi + Typografier	Formatér * Celler	∑ · Å▼ · ↓ · , ○ · ≪ · Redigering ^
К1 - Е	$\times \checkmark f_x$					~
A B C D E 1 Hojde i cm 2]x ₁ ; x _i] 3 3	F Intervalmidtpunkt <i>m</i> i	G Interval hyppighed h _i	H Intervalfrekvens <i>f</i> i	l Summeret frekvens <i>F</i> _i	J Produkt m _i f _i	K L
7 100 190 8 190 200 9 10 10 11 12	Skema (+)		ndstillinger 1	i (+ 100 %

 $1. \ kolonne \ (de \ fem \ første \ kolonner \ i \ regnearket):]x_{i-1};x_i] \ "Elevernes \ højde"$

De værdier, som undersøgelsen kan antage. I dette tilfælde er det de målte højder af de 27 elever. Bemærk igen, at observationerne nu godt kan "falde udenfor ganske bestemte værdier" – f.eks. kan man sagtens være 182 *cm* høj, selvom 182 *cm* ikke er en eksakt observation. Så bliver man bare indskrevet i intervallet]180;190].

2. kolonne : *m_i* "Intervalmidtpunkt"

Her beregnes midtpunkterne af værdi-intervallerne. Dette er nødvendigt for at have en fast værdi at gå ud fra. Det medfører naturligvis også en vis usikkerhed i resultaterne, at man beregner værdierne ud fra intervalmidtpunktet, for man kan jo principielt ikke vide, om alle observationer i et bestemt interval alle ligger over eller under intervalmidtpunktet. I det virkelige liv må man gå ud fra at observationerne er ligeligt fordelt – som allerede nævnt, og derfor kan der ses bort fra denne usikkerhed. Har man dog mistanke om, at datamaterialet er for usikkert, kan man jo indføre mindre intervaller. Dette vil medføre en større præcision i resultaterne.

Intervalmidtpunkterne beregnes ganske enkelt som:

 $m_i = Intervalstartværdi + \frac{Intervalslutværdi - Intervalstartværdi}{2}$

F.eks. er intervalmidtpunktet for intervallet $]160;170] = 160 + \frac{170 - 160}{2} = 160 + \frac{10}{2} = 160 + 5 = 165$

Dette noteres i Excel som:



Pas på!!! Intervallerne behøver ikke at være lige store. Hverken i opgaverne eller i det virkelige liv. Så lad være med at stole blindt på, at man bare kan tage det samme interval og lægge til det forrige, men i stedet udregne hvert interval med den ovenstående formel.

Her er det vigtigt at vide, at i det tilfælde, hvor intervallerne er uens, så regner man snarere med arealet af "frekvensen", for så kan man operere med forskellige intervaller.

∎ ఈెం	 Mappe1 (version 	2).xlsb[Gendannet auto	ma Michel Mand	ix (MIM - Underviser - U	nord) 🖭	
Filer Hjem I	Indsæt Sidelayout	Formler Data	Gennemse Vis	Hjælp Acrobat Ç	Fortæl mig det	A₁ Del
Sæt ind - V Udklipsholder 5	New Ro + 11 → A (<u>U</u> +) ⊞ + Skrifttype	$\mathbf{A}^{*} = \mathbf{E}^{b} = \mathbf{E}^{b}$ $\mathbf{E}^{*} = \mathbf{E}^{b} = \mathbf{E}^{b}$ $\mathbf{A}^{*} = \mathbf{E}^{b} = \mathbf{E}^{b}$ $\mathbf{A}^{*} = \mathbf{E}^{b} = \mathbf{E}^{b}$ Justering	Standard • 5	Betinget formatering + Formatér som tabel + Celletypografi + Typografier	Formatér • Celler	$\sum_{i=1}^{n} \frac{1}{i} \sum_{j=1}^{n} \frac{1}{i} \sum_{j$
G6 - :	$\times \checkmark f_x$					*
A B C D E	F Intervalmidtpunkt	G Interval hyppighed	H Intervalfrekvens	l Summeret frekvens	J Produkt	K L 🔺
2 $[x_{i-1}; x_i]$	m _i	h _i	f_i	F_i	$m_i f_i$	
3	155					
5] 160 ; 170]	165					
6] 170 ; 180]	175					
7] 180 ; 190]	185					
8] 190 ; 200]	195					
9						
11						
12						
A Data	Skema 🕀			1 4		
Klar			ndstillinger f	or visning	巴	+ 100 %

3. kolonne : *h_i* "Intervalhyppighed"

Hvor mange gange optræder en værdi i observationssættet? Forklaringen er nøjagtig den samme, som for de diskrete variable, dog med den forskel, at hvis der er 20 forekomster af højden 179 *cm* og ikke andre observationer i intervallet]170;180], så vil de 20 forekomster i teorien være jævnt fordelt i hele intervallet.

Nu er det tydeligt, hvorfor første kolonne inddeles i fem mindre kolonner. Selve det "synlige" af kolonnen udgøres af rammen. Netop fordi der er plads til at indtaste intervalgrænserne i individuelle celler, kan de nu benyttes i udregninger af f.eks. intervalmidtpunkterne.

Når den første linje (række 4) er indtastet kan den trækkes ned i fyldhåndtaget i stedet for at skrive det hele fire gange mere.

Nu skal data indsættes fra tabellen, som i begyndelsen blev lagt ind i det andet regneark. Gudskelov blev tabellen også navngivet, så det er ikke nødvendigt at tænke over, hvor den ligger henne.

Problemet med grupperede observationer er, at de skal sorteres ind i de forskellige intervaller. Heldigvis er intervalgrænserne indtastede som værdier, så de kan bruges i denne sortering.

Der findes en indbygget funktion i Excel, som kan udføre netop denne operation.

Den hedder: TÆL.HVISER(Kriterieområde1;Kriterie1[; ... Kriterieområde*n*;Kriterie*n*]) Denne funktion kan søge efter (og tælle) værdier i forskellige celleområder.

I dette tilfælde er det relativt nemt! Kriterieområdet er altid det samme – nemlig "Tabel", hvilket området med data blev navngivet til i begyndelsen.

Det er værre med kriterierne! Der må være to kriterier for hver linje, idet værdien skal være større end mindsteværdien og mindre ens størsteværdien for det pågældende interval.

Kriterier indtastes i gåseøjne. Skal en værdi f.eks. være større end 10, indtastes ">10", som kriterium.

Dette virker noget besværligt, for så skal man jo manuelt ind og ændre hver eneste linje til at søge efter de forskellige intervaller. Nu er det godt, at der er gjort et godt forarbejde, for man kan godt lave et kriterium, som forholder sig til indholdet i en celle.

Samme eksempel som før: Der skal sorteres efter værdier, som er større end 10.

10 indtastes i celle B8. Nu kan kriteriet skrives som: ">"&B8

Tilbage til eksemplet med elevernes højder:

I celle G4 indtastes: =TÆL.HVISER(Tabel;">"&B4;Tabel;"<="&D4)

Det huskes, at "Tabel" er navnet på det område, hvor alle data står. ">"&B4 er referencen til intervallets nedre grænse, som står i celle B4. "<="&D4 er referencen til intervallets øvre grænse, som står i celle D4. Husk lighedstegnet foran formlen. Nu vil Excel søge (og tælle) højder, som er mellem 150 cm og 160 *cm*. (Den burde returnere tallet 2).

121	 Mappe1 (version 	2).xlsb[Gendannet autor	ma Michel Mand	lix (MIM - Underviser - U	nord) 🖭	- 0	/ (× /
Filer Hjem	Indsæt Sidelayout	Formler Data C	Gennemse Vis	Hjælp Acrobat Ç	Fortæl mig det		R C-1
Saet Ind - V	N P s New Ro - 11 - A K ∐ - ⊡ - ۞ - Skrifttyne		Gandard ♥ Standard ♥ ∰ + % ∞ % #	Beenget formatering ~ Formater som tabel ~ Celletypografi ~	Formatér ~	∑ ~ <u>∧</u> ⊤ ~ ⊎ ~ <u>)</u> ~ <i>∢</i> ~ Redigering	
		Just		iypograner	cener	neargening	
ER.TOM *	$X \checkmark f_x$	=TÆL.HVISER(Tabe	el;">"&B4Tabel;"<=	="&D4)			~
ABCDE	F	G	н	1	J	к	L 🔺
1 Højde i cm	Intervalmidtpunkt	Interval hyppighed	Intervalfrekvens	Summeret frekvens	Produkt		
2 $]x_{i-1}; x_i]$	mi	h _i	f_i	Fi	$m_i f_i$		
3 4] 150; 160] 5] 160; 170]	=TÆL.HVISER(T	abel;">"&B4Tabel;"< ieområde1; kriterier1; [<="&D4) (kriterieområde2; krit	terier2]; [kriterieområde3	;)		
6] 170 ; 180] 7] 100 100]	175						
7 J 180 ; 190 J 8 J 190 · 200 J	185						
9 10	175						
11							
12							-
A Data	a Skema 🕂			4			Þ
Rediger			Indstillinger f	for visning 🔠 🗉	─ -	+ +	100 %

Når formlen er indtastet korrekt, kan den trækkes ned i fyldhåndtaget.

⊟ 5°¢°	 Mappe1 (version 	2).xlsb[Gendannet auto	ma Michel Ma	ndix (MIM - Underviser - U	nord) 🖭	- (0/)	< /
Filer Hjem	Indsæt Sidelayout	Formler Data (Gennemse Vis	Hjælp Acrobat Ç	Fortæl mig det	۶4 De	
Sæt ind - *	s New Ro + 11 - A K <u>U</u> + ⊡ + △ + , Skrifttype		Standard ▼ ♀ ♀ % 000 *% ↓% Tal	Betinget formatering ~ Formatér som tabel ~ Celletypografi ~ Typografier	Formatér •	∑ ~ Å▼ ~ ↓ ~ ,○ ~ ∢ ~	~
42	: X	,				· · · · · · · · · · · · · · · · · · ·	~
A B C D E	F	G	н	1	J	κL	
1 Højde i cm	Intervalmidtpunkt	Interval hyppighed	Intervalfrekver	is Summeret frekvens	Produkt		1
2 $]x_{i4}; x_i]$	m _i	h _i	f_i	Fi	$m_i f_i$		
3 4] 150 ; 160]	155	2		_			
5] 160 ; 170]	165	10					
6 J 170 ; 180 J 7 J 180 - 100 J	175	10					
8 1 190 · 200 1	185	2					
9		-					
10							
11							
12							Ŧ
 Data 	Skema 🕂			1 4		•	
Klar			ndstillinge	er for visning 🔠 🗉	<u> </u>	+ 100 9	%

Afslut kolonnen "Interval hyppighed" med at summere antallet af de enkelte observationer. Dette gøres af flere årsager. Dels som en kontrol, hvis man nu kender det totale antal observationer, men – og det er langt mere vigtigt – fordi denne værdi skal bruges i de næste udregninger.



4. kolonne : *f_i* "Intervalfrekvens"

Frekvensen eller den relative hyppighed. Forklaringen er nøjagtig den samme, som for de diskrete variable. Det er antallet af observationer divideret med det totale antal observationer (populationen).

I Excel vil en normal cellereference være relativ. Dvs. at når man flytter eller kopierer en celle med en reference i, så vil en reference til en anden celle flytte sig tilsvarende. Vil man undgå det, kan man lave en reference "absolut" i stedet for relativ. Dette gøres ved at tilføje dollartegn (\$) foran bogstaver og tal i referencen. Det gøres nemt med genvejstasten <F4>, som cykler gennem de forskellige kombinationer af fastlåsning af en celle. I dette tilfælde skal blot bruges total fastlåsning, så når man trykker en enkelt gang på <F4>, vil en reference (man skal stå på den i formlen) ændre sig. F.eks. vil en reference til celle "K5" ændre sig til "\$K\$5" efter en enkelt tryk på <F4>. Kommer man til at trykke for mange gange på <F4>, så bare bliv ved at trykke på <F4>, indtil der står "\$K\$5".

∎ ອະ∂	- т м	appe1 (version	2).xlsb[Gendannet auto	oma Michel Mar	idix (MIM - Underviser - U	Inord) 🖭	- 🛛	/(×/
Filer Hjem	Indsæt	Sidelayout	Formler Data	Gennemse Vis	Hjælp Acrobat 🤇	Fortæl mig det		ନ୍∔ Del
Sæt ind - V Udklipsholder 5	F <u>K</u> ∐⊸ Ska	<u>• 11</u> • Â	$ \begin{array}{c} A \\ A \\ \hline \\ A \\ \hline \\ \hline \\ \hline \\ \hline \\ \hline \\ \hline \\$	 Standard ▼ Standard ▼	Betinget formatering ~ Formatér som tabel ~ Celletypografi ~ Typografier	智 Indsæt ~ 图 Slet ~ 同 Formatér ~ Celler	∑ • Å⊤ · ↓ · ,○ • ∢ • Redigering	^
G9	: x	$\checkmark f_x$	=G4/\$G\$9					~
A B C D 1 Hojde i c 2] <i>x</i> _{<i>i</i>-1} ; <i>x</i> _{<i>i</i>}	E m Interv	F almidtpunkt <i>m</i> i	G Interval hyppighed h _i	H I Intervalfrekven <i>f</i> i	s Summeret frekvens <i>F</i> _i	J Produkt m _i f _i	К	
3 4] 150 ; 16 5] 160 ; 17 6] 170 ; 18	i0] i0]	155 165 175	2 10 10	=G4/\$G\$9				
7] 180 ; 19 8] 190 ; 20 9	0]	185 195	3 2 27					
10 11 12								
 ← → □ Peger 	Data Ske	ema (+)		ndstillinge	i 4	₽ -	+ +	► 100 %

Igen trækkes formlen ned, når den er korrekt indtastet og igen summeres værdierne for kontrol. Summen skal give 1,0000.

⊡ 5 • ∂•	 Mappe1 (version 	2).xlsb[Gendannet autor	ma Michel Man	dix (MIM - Underviser - Ur	nord) 🖭	- 0	/(× /
Filer Hjem	Indsæt Sidelayout	Formler Data C	Gennemse Vis	Hjælp Acrobat Ç	Fortæl mig det		우 Del
Sæt ind - * Udklipsholder 5	es New Ro - 11 - A K <u>U</u> - ⊞ - <u>∆</u> - Skrifttype		Tal →	Betinget formatering ~ Formatér som tabel ~ Celletypografi ~ Typografier	Formatér * Celler	∑ ~ Â▼ ~ ↓ ~ ♀ ◆ ~ Redigering	^
Н9 т	: $\times \checkmark f_x$	=SUM(H4:H8)					~
A B C D 1 Hojde i cm 2]x _{i,1} ;x _i] 3 4] 150 ; 160 5] 160 ; 170 6] 170 ; 180	E F Intervalmidtpunkt m _i 155 165 175	G Interval hyppighed h _i 2 10 10	H Intervalfrekvens <i>f</i> i 0,0741 0,3704 0,3704	Summeret frekvens F _i	J Produkt m _i f _i	K L	
7] 180 ; 190 8] 190 ; 200 9 10 11 12] 185] 195	3 2 27	0,1111 0,0741 1,0000				
Klar Da	ta skema (+)		ndstillinger	for visning	<u> </u>		100 %

5. kolonne : *F_i* "Summeret Frekvens"

Her tager man frekvenserne og lægger sammen. Igen er forklaringen præcis den samme som for de diskrete variable.

I eksemplet i celle I4 indtastes: "=H4". Det er ikke nødvendigt med mere, da der jo ikke er nogle rækker oven over, som kan bidrage til resultatet.

I række 2 er det lidt mere besværligt. I celle 15 tastes "=H5+I4". Denne formel vil tage den næste frekvens og lægge til det forrige resultat. Denne formel kan trækkes ned i fyldhåndtaget, og vil give det rigtige resultat for hver enkelt række.

Igen skal sidste værdi (i dette tilfælde i række 8) være lig med 1,0000.



6. kolonne : $m_i \cdot f_i$ "Produkt"

Produktet bruges mest til at finde middelværdi (=gennemsnit). Hver række bidrager til produktet med værdien (i dette tilfælde højden) ganget med frekvensen.

Men da højderne er angivet i intervaller er det nødvendigt at bruge intervalmidtpunkterne og gå ud fra, at observationerne er rimelig ligelig fordelt. Det er dog allerede antaget, så det er ikke noget problem.

I dette tilfælde ville man sige, at der er 10 elever, som er mellem 160 og 170 *cm* høje. Det vides ikke bedre end at sige at de gennemsnitligt er 165 *cm* høje. Nogle er lavere og nogle er højere, og det går nok nogenlunde op. De 165 *cm* multipliceres med den udregnede frekvens for de observerede 10 elever. I det samlede regnskab giver dette 11,4845. Når man så har fundet produktet for hver værdi, lægges disse sammen. Det vil altså sige, at man har fundet ud af, at eleverne i hele undersøgelsen i gennemsnit er 172,4 *cm* høje.

<mark>ມ</mark> ່າ∘∂	Mappe1 (versio	n 2).xlsb[Gendannet auto	ma Michel Ma	ndix (MIM - Underviser - U	nord) 🖭	- 0	/ (× /
Filer Hjem	Indsæt Sidelayout	Formler Data	Gennemse Vis	Hjælp Acrobat Ç	Fortæl mig det		₽ ₄ Del
Sæt ind - V Udklipsholder 15	imes New Ro ▼ 11 - ▼ F K <u>U</u> ▼ H ▼ A	A A A A A A A A A A A A A A A A	Tal ▼ ▼ ♀ % 000 % % 5 Tal 5	Betinget formatering ~ Formatér som tabel ~ Celletypografi ~ Typografier	Endsæt • Slet • Formatér • Celler	∑ - Â⊤ - ↓ - Redigering	^
J9	$: \times \checkmark f_x$	=SUM(J4:J8)					~
A B C D	E F m Intervalmidtpunk	G t Interval hyppighed	H	s Summeret frekvens	J	K L	
2]x _{i-1} ;x _i] <i>m_i</i>	h _i	f_i	F _i	$m_i f_i$		
3 4] 150 ; 16 5] 160 ; 17	i0] 155	2	0,0741	0,0741	11,48148		
6] 170 ; 18	0] 175	10	0,3704	0,8148	64,81481		
7] 180 ; 19 8] 190 ; 20	0] 185 0] 195	3	0,1111 0,0741	0,9259 1,0000	20,55556 14,4444		
9		27	1,0000		172,4074		
11 12							_
	A Data Skema 🕀 : A Data Data Skema						
Klar			Indstillinge	er for visning 🔠 🗉	<u> </u>	+	100 %

Diagrammer for grupperede variable

Typisk vælger man at afbilde data for grupperede variable i to diagrammer. Et **søjlediagram** (kaldes også for et histogram) og en **frekvenskurve**.

Søjlediagrammet laves på samme måde som pindediagrammet. Den eneste forskel er udseendet på søjlerne, idet søjlerne her skal være så brede, at de støder op mod hinanden. Dette gøres for at illustrere at det gælder for HELE intervallet. I søjlediagrammet afsætter man værdierne ud af abscisse-aksen og frekvensen op ad ordinat-aksen.

Bemærk at der – ligesom for de diskrete observationer - gælder for BÅDE søjlediagrammet og frekvenskurven at ordinat-aksen ALTID går mellem 0 og 1. (Frekvenskurven går netop fra 0 til 1, medens søjlediagrammet – som i nedenstående eksempel – kan være endnu mindre, afhængigt af frekvensernes størrelse.)



(Her er et diagram, som er lavet i Excel. Bemærk, at i egenskaber for dataserie er mellemrummet mellem søjlerne sat til 0, for at gøre søjlerne så brede som muligt.

Højde i cm	Højde	Intervalmidtpunkt	Interval hyppighed	Intervalfrekvens	Summeret Frekvens	Produkt
$]x_{i-1};x_i]$		m_i	h_i	f_i	F_i	$m_i \cdot f_i$
					0,0000	
] 150 ; 160]	150	155	2	0,0741	0,0741	11,4815
] 160 ; 170]	160	165	10	0,3704	0,4444	61,1111
] 170 ; 180]	170	175	10	0,3704	0,8148	64,8148
] 180 ; 190]	180	185	3	0,1111	0,9259	20,5556
] 190 ; 200]	190	195	2	0,0741	1,0000	14,4444
	200		27			172,4

Bemærk, at der i skemaet indsættes en ny kolonne, "Højde". Den er ikke strengt nødvendig for at løse opgaven, men den kan være en stor hjælp. Specielt hvis frekvenskurven skal indtegnes i Excel, så er den faktisk uundværlig. I frekvenskurven vises den summerede frekvens. Her afsætter man værdierne ud af

abscisse-aksen. Sørg for at have den mindste værdi ved ordinat-aksen. Værdien for den summerede frekvens aftegnes som en kurve, som er stykvis lineær (en ret linie) mellem punkterne. Lad kurven begynde i (150,0). Bemærk, at det er hér hvor den ekstra kolonne bruges, således at stregen går fra (150,0) til (160;0,0741) og videre fra (160;0,0741) til (170;0,4444) osv.



Her er en frekvenskurve, som er lavet i Excel. Bemærk, at punkterne går direkte fra (150,0) og til (160;0,0741) osv. Dette er ikke en Excel standard, for der vil kurven begynde midt mellem 150 og 160. Men hvis man går ind i Excel i diagrammet og højreklikker på *x*-aksen og vælger "Formater akse", der kan man vælge at "Lodret akse krydser" "Mellem aksemærker", og så virker det!

Mht. fraktiler og dermed også kvartiler, er de lidt sværere at aflæse – både i tabellen og kurven. Har man indtegnet kurven på millimeterpapir, kan man lave en ok aflæsning.



Skal man udregne fraktilen nøjagtigt, kan man benytte følgende metode...

Denne sidste øvelse vil dog ofte være overflødig.

Dette er en af de få matematiske discipliner, hvor det kan være mere formålstjenstligt at aflæse end at beregne.

Husk, at data i forvejen er approksimerede, idet data antages at være jævnt fordelt i hvert interval. Derfor må det påpeges at den ekstra præcision der er givet ved at beregne de eksakste fraktiler ikke kan ændre på den usikkerhed, som er iboende i måden at behandle data på.

Derfor vil det være ekstremt få tilfælde, hvor denne metode vil komme i brug.

Et par eksempler med grundig Excel forklaring.

Diskrete variable.

Af en virksomheds ulykkesstatistik fremgår, at antal ulykker pr. dag varierer fra 0 til 5. I en periode på 200 dage ser ulykkesstatistikken således ud:

Antal ulykker pr. dag	Antal dage
0	97
1	63
2	31
3	8
4	0
5	1
I alt	200

Der laves et skema med tilhørende udregninger i Excel: (Celler med formler er skrevet med rødt.)

Antal ulykker pr. dag	Antal dage (Hyppighed)
Xi	h _i
0	97
1	63
2	31
3	8
4	0
5	1
Antal kategorier: k = 6	Antal observationer: Population i alt: n = 200

Teksten og tallene indtastes.

- Teksterne med blå skrift laves således:
 - Skriv teksten. Marker "i" et. 1.
 - 2. Klik på den lille pil nederst i menufeltet "Skrifttype". (Eller genvej: CTRL+SHIFT+B) for at åbne dialogboksen. (Eller genvej: CTRL+1) for at åbne dialogboksen. irmatér celle
 - 3. Vælg "Sænket skrift". (Eller genvej: ALT+N – gælder kun i dialogboksen)
 - 4. Vælg blå skriftfarve. Enten i dialogboksen eller i menuen. 5. Centrer teksten.
- Tal Jur

H 6

ind -Jdklipsholder 🕞 à

ies New Roma

Juster

Hjem

ж

6 Sæt

~*

De røde tekster indeholder både almindelig tekst og tal, som er udregnet i Excel. Først en forklaring af de funktioner, som udregner tallene: Antal kategorier er antallet af de muligheder, som man kan observere. De er alle listet i tabellen. I dette tilfælde er det ikke svært at overskue at der er seks forskellige muligheder, men man kan sagtens forestille sig et eksempel, hvor der er mange flere. Lad det være givet at cellen med den sorte tekst: "Antal ulykker pr, dag" er placeret i position A1. Det er rart med en underoverskrift, hvor betegnelse eller enhed er angivet. Det er her gjort med blå skrift.

ß

Dernæst kommer antallet af observationer af de forskellige kategorier. De står derfor lodret under hinanden i cellerne A3 til A8. I Excel skrives dette som A3:A8. Altså cellerne fra A3 til A8 (begge inklusive) skrives som A3:A8.

Skal man vide, hvor mange celler, som indeholder TAL, kan man bruge Excel-funktionen: TÆL(celler). Her er det vigtigt at vide, at der KUN tælles celler, som indeholder TAL-værdier. Så den kan altså ikke anvendes til at tælle celler med almindelig tekst.

Enhver funktion, som man indtaster i Excel, skal begynde med et lighedstegn.

T&EL(celler) bruges således: Stå i den celle, hvor resultatet af optællingen skal stå. Skriv: "=T&EL(". Herefter skal man angive de celler, som indeholder tallene. Det kan gøres på flere måder. De mest almindelige er, at man skriver det – altså som A3:A8. Man kan også markere med musen. Så trækkes der hen over cellerne. Husk at afslutte med en slut-parentes.

Det endelige resultat bør være: =TÆL(A3:A8). Denne funktion hedder "COUNT ..." på engelsk.

Som tidligere nævnt, kan man godt kombinere flere tal og tekster i en celle. Til dette bruges funktionen: SAMMENKÆDE(celle/"tekst"; celle/"tekst"; ...; celle/"tekst"). Bemærk, at ren tekst skal skrives i anførselstegn. Denne funktion hedder "CONCATENATE ..." på engelsk.

Resultatet i dette tilfælde giver: =SAMMENKÆDE("Antal kategorier: k = ";tæl(A3:A8)).

Alt dette er fint, men hvis man ønsker at regne med denne talværdi, kan det blive lidt besværligt, fordi indholdet af denne celle vil (af Excel) blive betragtet som en tekst og ikke som et tal. Det vil sige, at man i de senere udregninger, hvor dette tal skal bruges igen, skal regne helt forfra.

Man kan i stedet formatere sig ud af problemet. Selvfølgelig er der ingen indbyggede formateringer, som tilbyder ar skrive "Antal kategorier: k =" foran et tal, så den må man selv lave.

- Brugerdefineret formatering laves således:
 - 1. Stå i cellen, hvor antallet af kategorier skal stå. Lav beregningen med funktionen: =TÆL(Matrix)

Klip

Sæt nd + 💞 Formatp

Formatér celler

Calibri

Tal Justering Skrifttype Kant Fyld Beskyt

Eksempel AKKUM.HOVEDSTOL

Celler i standardformat har ikke noget bestemt talforma

• 11 • A A = = = ≫•• ₽ Ombryd tekst

F K U • ⊞ • 💩 • ▲ • ≡ ≡ ≡ 🖽 🖽 Flet og centrer • 🍄 • % 0

Standard

- 2. Klik på den lille pil nederst i menufeltet "Tal".
- 3. Vælg "Brugerdefineret i rullemenuen.
- I feltet "Type" skriv:
 "Antal kategorier: k = "0 nøjagtig som på linjen ovenover.

rmatér celler		?	×
Tal Justering	Skrifttype Kant Fyld Beskyttelse		
ategori: Standard Tal Valuta Revision Data	Eksempel AKKUM.HOVEDSTOL Type:		
Klokkeslæt Procent Brøk Vide zil belig	"Antal kategorier: k = "0 Standard 0 0,000		^
Speciel Brugerdefineret	"		
	#,##0 kr.;[Rød]-#,##0 kr.	Slet	~
kriv koden for talfor	 matet. Benyt en af de koder, som allerede findes, som udgang	spunkt.	
	ОК	Annu	iller

Denne operation foretages tilsvarende for cellen ved siden af, for at formatere cellen, hvor populationen, n, udregnes.

OK Annuller

Hvis man laver professionelle regneark, er det næsten en krav, at man navngiver sine celler! Det er der flere grunde til. For det første, så er det meget nemmere at huske at momssatsen står i en celle ved navn: "Moms", end at huske at den står i cellereferencen: "EK472". For det andet vil en reference til en celle, som er navngivet ALTID være absolut. Dette skyldes naturligvis, at der aldrig kan forekomme mere en én celle, som har det samme navn.

- For at navngive cellen med den udregnede population (celle B9), gør følgende:
 - 1. Stå i cellen, som skal navngives.
 - 2. Klik på navnefeltet.
 - 3. Skriv cellens nye entydige navn. (Her: "Population") 2
 - 4. Afslut med ENTER. (Dette er vigtigt!)
- Det kan varmt anbefales at navngive så mange celler som muligt.
- Husk at alle referencer til navngivne celler er absolutte, så det er ikke et mål i sig selv at ALLE celler skal navngives.
- Sørg for at navngive med så logiske navne som muligt.
- Husk, at alle fremtidige referencer til den navngivne celle SKAL skrives nøjagtig som den er navngivet altså med en kombination af små og store bogstaver.

Et andet godt tip er at gemme kildedata i et andet regneark i projektmappen.

I mange opgaver, er kildedata givet i hyppigheder, altså ligesom i dette eksempel, hvor det er givet, at der er 97 forekomster af 0 ulykker på en dag etc.

Det kan dog hænde – f.eks. hvis der er tale om måledata fra et fysikforsøg – at man simpelthen har givet kildedata som de enkelte observationer.

I dette eksempel er dette simuleret ved at der i et andet regneark er lavet 200 enkelte observationer, som passer med de givne hyppigheder. Det giver nemlig mulighed for at bestemme hyppighederne på en smart måde i Excel:

Begynd med at oprette et nyt regneark og navngiv de to regneark, som er indeholdt i projektmappen.

Teksten og tallene indtastes.

- Nye regneark oprettes og navngives således:
 - 1. Klik på det lille "+" nederst til venstre i vinduet.
 - 2. Der står nu "Ark*n*", hvor *n* er antallet af regneark i projektmappen. Erstat "Ark*n*" med navnet "Kildedata".
 - 3. Dobbeltklik på navnet i fanebladet "Ark1". Skriv det nye navn: "Deskriptorer".

Som allerede nævnt, så er der for eksemplets skyld, indtastet de 200 observationer i arket: "Kildedata".

8																				×
Fler	Hem in	dsæt Sidela	yout Formle	r Data Ge	nnemse Vis	Udvikler	Acrobat 🖓	Fortel mig. huge du	vilgare								Michel M	andix (MIM - Unde	rviser - Esnant) 🙎	. D-1
		" 🗇 "	C Lineal	✓ Formellirje	Q 📑	0	ೆ" 🗆 1	Dpdel 0	D'Vis side om side		187 17									
Normal	Vis Side I	Insperdefinerede	🖂 Giterinjer	2 Oveskifter	Zeorn 100%	Zoom på	Nyt Arranger	lays Sigui		un af vinduat	Skift Makes	ier								
_	Projektrappents	ninger		Vis	Zeon	narearingen is	mose are n	00**	indue		Make	ner .								~
J20		×	0																	
			6				6				K	1.1			0		0			
1	0	1			2		1	0 1			1		IVI	IN	0	r	Q.	ĸ	3	-11
2	1	2	1	2	1		0	1 0	0		0									-11
3	0	2		1			0	0 1	1		0									-11
4	1	0	0		0		1	2 0	0		1									-11
5	0	1	0	3	1		3	5 0	1		1									-11
6	2	0	0	1	0)	0	2 1	0		3									
7	1	2	0	0 0	0)	2	0 0	1		0									
8	0	0	3	0	1		2	0 2	0		1									
9	0	1	1	1	0)	0	0 0	0		1									
10	0	0	0	0 0	0)	2	0 2	1		2									
11	1	0	1	2	1		0	3 0	1		0									
12	1	2	0	1	. 0)	2	1 2	0		0									
13	0	0	1	0	0)	0	0 1	1		1									
14	3	0	2	1	0)	1	1 2	0		0									-11
15	0	0	0	0 0	1		0	2 0	1		1									-11
16	1	0	0	2	1		0	3 1	1		0									-11
17	0	1	1	0	0		2	1 2	0		2									-11
18	1	2	0	1	1		0	0 0	2		0									-11
19	1	0	3	0	2		0	0 2	1		1									-11
20	1	0	1	0	0) 	2	1 1	0		0									-11
21																				-11
22																				-11
23																				
24																				
	Desk	riptorer KG	dedata	۲													1001	80 50		
Cal.	-	_	_	_	_	_		_	_		_				_			B E *	+	120.%

F	iler Hjem Indsi	et Sidelayout	Formler	Data Gennemse	١
Sa	► Klip E Kopiér → eet d → ✓ Formatpensel	Times New Roma •	11 т А́а́		
	Udklipsholder 🕞	Skrifttype	5	Just	erin
Pc	opulation 🝷 🗄 >	<i>√ f</i> x =	SUM(B3:B8)		
	A		В		
1	Antal ulykker pr. d	ag A	ntal dage (H	yppighed)	
1 2	Antal ulykker pr. d x _i	ag A	ntal dage (H h _i	yppighed)	
1 2 3	Antal ulykker pr. d x _i 0	ag A	ntal dage (H <u>h</u> i 97	yppighed)	
1 2 3 4	Antal ulykker pr. d	ag A	ntal dage (H <u>h</u> i 97 63	yppighed)	
1 2 3 4 5	Antal ulykker pr. d	ag A	ntal dage (H h _i 97 63 31	yppighed)	
1 2 3 4 5 6	Antal ulykker pr. d	ag A	ntal dage (H h _i 97 63 31 8	yppighed)	
1 2 3 4 5 6 7	Antal ulykker pr. d x _i 0 1 2 3 4	ag A	ntal dage (H <u>h</u> i 97 63 31 8 0	yppighed)	
1 2 3 4 5 6 7 8	Antal ulykker pr. d x _i 0 1 2 3 4 5	Ag A	Initial dage (H h _i 97 63 31 8 0 1	yppighed)	



Ydermere, så er OMRÅDET som indeholder de 200 værdier også navngivet med navnet "Kildedata". Det er gjort ved at markere alle 200 celler, og så navngive, som allerede beskrevet. Dette er særlig smart, for så kan man fremover bare referere til navnet "Kildedata" i stedet for at skrive referencen eller trække med musen fra A1 til J20.

Vend tilbage til det oprindelige regneark ("Deskriptorer") ved at klikke på fanen nede ved statuslinjen.

Ved at have de 200 observationer i et navngivet område, kan Excel bestemme hyppighederne ved at indtaste følgende formel i celle B3 (den gule celle).

=TÆL.HVIS(Kildedata;A3) Denne funktion hedder "COUNTIF ..." på engelsk

Denne formel kan trækkes ned til og med celle B8. Bemærk, at referencen til området med kildedata er konstant, mens referencen til celle A3 (altså den kategori der søges) skifter til A4 i cellen neden under etc.

Denne funktion er meget anvendelig. Den leder i et celleområde ("Kildedata" efter forekomster af det, som står i celle A3 ("0")) og foretager en optælling af "hits" og returnerer dette tal. I dette eksempel er dette tal netop hyppigheden af de enkelte tal fra 0 til 5.

Samlet ser det således ud:

Antal ulykker pr. dag	Antal dage (Hyppighed)	Formler i kolonne 'B'
<i>x</i> _{<i>i</i>}	h _i	
0	97	=TÆL.HVIS(Kildedata;A3)
1	63	=TÆL.HVIS(Kildedata;A4)
2	31	=TÆL.HVIS(Kildedata;A5)
3	8	=TÆL.HVIS(Kildedata;A6)
4	0	=TÆL.HVIS(Kildedata;A7)
5	1	=TÆL.HVIS(Kildedata;A8)
Antal kategorier: k = 6	Antal observationer: Population i alt: n = 200	=SUM(B3:B8)

Nu tilføjes den næste kolonne. I den ønskes frekvenserne af de enkelte observationer at blive beregnet.

Hvis projektmappen er konstrueret korrekt, vil formlerne blive nemmere og nemmere at arbejde med.

Lav overskrifterne i kolonne C. For at formatere cellerne hurtigt, kan det anbefales at trække overskrifterne fra kolonne B. Derved bliver tekst OG formater kopieret. Rediger nu teksten, så står den tilbage med det rigtige format.

Skriv følgende formel i celle C3:

=B3/Population

Den formel vil tage den værdi som står i celle C3 og dividere med den absolutte reference til cellen, som er navngivet Population. Bemærk, at idet der tastes referencen "Population", så vil Excel reagere med at lave en hurtig rullemenu med alle de elementer i Excel, som begynder med "P", "o" etc. efterhånden som man indtaster. Før eller siden, vil elementet "Population" dukke op, fordi en celle tidligere er navngivet således.

For at spare lidt tid og kræfter, men ikke mindst for at undgå stavefejl i referencerne, kan man klikke i rullemenuen på linjen: "Population", for at indsætte denne reference. Afslut med ENTER.

dag	Antal dage (Hyppighed)	Frekvens	
	h,	Si.	
	97	=B3/Po	
	63	C POISSON.FO	RDELING
	31	Population	
	8	(6) POTENS	
	0	Va POISSON	_
	1		
k = 6	Antal observationer: Population i alt: n = 200		

Træk formlen ned til og med celle C8 vha. fyldhåndtaget. (Det er den lille sorte klat nederst i et markeret celleområde – også hvis der kun er markeret en enkelt celle). Nu er alle frekvenserne beregnet. Lav evt. en kontrolsum i celle C9 for at addere alle frekvenserne. Resultatet skulle meget gerne give 1.

Antal ulykker pr. dag	Antal dage (Hyppighed)	Frekvens	Formler i kolonne 'C'
<i>x</i> _{<i>i</i>}	h _i	f_i	
0	97	0,485	=B3/Population
1	63	0,315	=B4/Population
2	31	0,155	=B5/Population
3	8	0,040	=B6/Population
4	0	0,000	=B7/Population
5	1	0,005	=B8/Population
Antal kategorier: k = 6	Antal observationer: Population i alt: n = 200	Kontrolsum = 1,000	=SUM(C3:C8)

Bemærk, at der i kontrolsummen er tre decimaler på 1-tallet. Hvis cellen er formateret på samme måde som hhv. celle A9 og B9, så vil der ikke være tre decimaler. Det kan ordnes ved at lave følgende brugerdefinerede formatering:

Formatér celler	?	×
Tal Justering Skrifttype Kant Fyld Beskyttelse		
Standard Tal Validation Dato Klocksizet Brokent Brak Vidensizebig Tekst Buggerdefineet Exempel Kontroisum = 1,000 Dype: Kontroisum = 0,000 Forent Brak Buggerdefineet Brak Vidensizebig Brokent Buggerdefineet Total Standard S	Sjet kt.	^
ОК	Annu	ller

Hvor man tvinger tre decimaler frem ved at skrive "0,000" efter teksten. Antallet af decimaler kan altså kontrolleres på denne måde ved at skrive antallet af ønskede decimaler med antallet af 0'er.

Næste kolonne indeholder den summerede frekvens. Her lægger man den aktuelle frekvens sammen med summen af alle de tidligere frekvenser. Det er vigtigt at begynde ovenfra og arbejde sig ned. Her er det lidt vigtigt at huske, at i den første kategori, så er der jo ikke nogen tidligere frekvens at arbejde med, så det svarer til at lægge den første frekvens til 0 (nul). Dermed er det jo bare den første frekvens. Derfor er formlen i den øverste række lidt mere enkel end i de øvrige rækker.

Bemærk, at der her bruges udelukkende relative referencer (altså ingen navngivne celler og ingen dollartegn), således at referencerne "flytter med", når den næst øverste formel trækkes ned vha. fyldhåndtaget.

Antal ulykker pr. dag	Antal dage (Hyppighed)	Frekvens	Summeret frekvens	Formler i kolonne 'D'
<i>x</i> _{<i>i</i>}	h _i	f_i	F_i	
0	97	0,485	0,485	=C3
1	63	0,315	0,800	=D3+C4
2	31	0,155	0,955	=D4+C5
3	8	0,040	0,995	=D5+C6
4	0	0,000	0,995	=D6+C7
5	1	0,005	1,000	=D7+C8
Antal kategorier: k = 6	Antal observationer: Population i alt: n = 200	Kontrolsum = 1,000		

Nu mangler blot produktet. Som beskrevet tidligere, så kan produktet beregnes på to måder. Her vælges det at beregne produktet på baggrund af frekvensen, således at middelværdien kan beregnes direkte.

Antal ulykker pr. dag	Antal dage (Hyppighed)	Frekvens	Summeret frekvens	Produkt	Formler i kolonne 'E'
x i	h _i	f_i	F _i	$x_i \cdot f_i$	
0	97	0,485	0,485	0	=A3*C3
1	63	0,315	0,800	0,315	=A4*C4
2	31	0,155	0,955	0,31	=A5*C5
3	8	0,040	0,995	0,12	=A6*C6
4	0	0,000	0,995	0	=A7*C7
5	1	0,005	1,000	0,025	=A8*C8
Antal kategorier: k = 6	Antal observationer: Population i alt: n = 200	Kontrolsum = 1,000			

Nu er skemaet færdigt. Nu mangler blot beregningen af de forskellige særlige deskriptorer, som middelværdi, median etc.

De er allerede beskrevet, på nær "Skævhed", som lige skal beskrives lidt yderligere.

Skævheden afhænger af forholdet mellem medianen og middelværdien.
Det kan antage to former:
Hvis middelværdien (µ) er større end medianen, siger man at datasættet er højreskævt.
Hvis middelværdien (µ) er mindre end medianen, siger man at datasættet er venstreskævt.
Da der kun er de to muligheder, kan dette vises i Excel vha. "HVIS-funktionen".

HVIS-funktionen er defineret således:

=HVIS(Logisk_test;Værdi_hvis_sand;Værdi_hvis_falsk)

Dette skal forstås på denne måde: Den logiske test er et spørgsmål eller en påstand, som kan besvares med "Sand" eller "Falsk". Der kan f.eks. stå resultatet af en temperaturmåling i celle A1. Der ønskes at skelne mellem om temperaturen er større end eller lig med 20°C. Såfremt dette er tilfældet, altså at påstanden om at temperaturen er større end eller lig med 20°C er "Sand", defineres temperaturen som "Varm". Hvis temperaturen er under 20°C, altså at påstanden om at temperaturen er større end eller lig med 20°C.

Ovenstående lille eksempel kan beskrives med en HVIS-funktion i Excel. Den målte temperatur tænkes at stå i celle A1 og vurderingen om hvorvidt det er en varm eller kold dag, skal være i celle C1.

Marker celle C1 og indtast funktionen: =HVIS(A1>=20;"Varm dag";"Kold dag"). I dette tilfælde er det to tekster, som er resultatet af vurderingen, men det kunne lige så godt have været tal eller formler. Husk, at tekster skal skrives i anførselstegn.

Mht. spørgsmålet om, hvorvidt datasættet i eksemplet er venstre- eller højreskævt, kan det beregnes som følger:

Middelværdi: µ:	0,77	=MIDDEL(Kildedata)
Median:	1	=MEDIAN(Kildedata)
Skævhed:	Datasættet er venstreskævt	=HVIS(I3>I4;"Datasættet er højreskævt";"Datasættet er venstreskævt")

Sammen med de andre deskriptorer, som allerede er beskrevet, kan den endelige beskrivelse af datasættet se således ud:

På baggrund af følgende observationer om antallet af ulykker pr. dag:

Filer	Hjam	Indu	et Sidela	yout Formier	Data Ge	nemer Vis	Udvikter J	kcrobat 🖓 Fo	rtel mig, hved du v	il gane		Γ.
Set Ind-	¥ Kip I⊴ Kopitr → ∳ Formatpe	nsel	EK U -	- 11 - A - 12 - 1	x = = =	* P(Dmbryd tekst Net og centrer	Standard	56 23 Betings	t Formatér 1g = som tabel =	Normal Bemærk!	
Ud	ik kp sholder	- 1	9	kriftlype	6	Justering		Tal	-			
K13												
1	A		в	C	D	E	F	G	н	1.1	J	
1		0	1	0	0	2	1	0	1	0	1	4
2		1	2	1	2	1	0	1	0	0	(2
3		0	2	0	1	2	0	0	1	1	(١
4		1	0	0	0	0	1	2	0	0	1	4
5		0	1	0	3	1	3	5	0	1	1	1
6		2	0	0	1	0	0	2	1	0	1	3
7		1	2	0	0	0	2	0	0	1	(د
8		0	0	3	0	1	2	0	2	0	1	1
9		0	1	1	1	0	0	0	0	0	1	1
10		0	0	0	0	0	2	0	2	1		2
11		1	0	1	2	1	0	3	0	1	(2
12		1	2	0	1	0	2	1	2	0	(_ ۱
13		0	0	1	0	0	0	0	1	1	1	1
14		3	0	2	1	0	1	1	2	0	(۱ د
15		0	0	0	0	1	0	2	0	1	1	1
16		1	0	0	2	1	0	3	1	1	(۱ د
17		0	1	1	0	0	2	1	2	0	1	2
18		1	2	0	1	1	0	0	0	2	(2
19		1	0	3	0	2	0	0	2	1	1	4
20		1	0	1	0	0	2	1	1	0	(د ا

Bestemmes følgende deskriptorer:

					Populationen: n:	200	Middelværdi: µ:	0,77
							Skævhed:	Datasættet er venstreskævt
					0. Kvartil: Minimum:	0		
Antal ulykker pr. dag	Antal dage (Hyppighed)	Frekvens	Summeret frekvens	Produkt	1. kvartil:	0	Variations bredde:	5
x _i	h_i	f_i	F _i	$\mathbf{x}_i \cdot \mathbf{f}_i$	2. Kvartil: Median:	1	Kvartilbre dde:	1
0	97	0,485	0,485	0	3. kvartil:	1		
1	63	0,315	0,800	0,315	4 Kvartil: Maksimum:	5	Typetal:	0
2	31	0,155	0,955	0,31	4. Itvartii. Maksiinuili.	5	Typetai.	
3	8	0,040	0,995	0,12				
4	0	0,000	0,995	0	5 % Fraktil:	0	Varians: (VARp):	0,8271
5	1	0,005	1,000	0,025	10 % Fraktil:	0	Spredning: σ :	0,909450383
Antal kategorier: k = 6	Antal observationer: Population i alt: n = 200	Kontrolsum = 1,000			80 % Fraktil:	1,2		