

Statistik

Formålet	1
Mindsteværdi	1
Størsteværdi	1
Ikke grupperede observationer	2
Median og kvartiler defineres ved ikke grupperede observationer således:	2
Middeltal defineres ved ikke grupperede observationer således:	2
Boksplot (eller kassedigram):	2
Grupperede observationer	3
hyppighed (antal)	3
Sumkurve	3
Frekvensen	3
Kumuleret hyppighed	3
Kumuleret frekvens	3
Median og kvartiler defineres ved grupperede observationer således:	4
Middeltal ved grupperede observationer	5
Illustration af forskel på middeltal og median	5
Histogram	5
Boksplot bruges også ved grupperede observationer	5

Formålet

med statistik er at få overblik over et stort talmateriale.

Vi anvender nogle såkaldte **deskriptorer**, der beskriver talmaterialet.

De enkelte tal i talmaterialet kaldes **observationer**.

Hele talmaterialet kaldes **observationssættet**

Vi vil her lære betydningen af følgende såkaldte deskriptorer:

Mindsteværdi (den mindste observation)

Størsteværdi (Den største observation)

Middeltal (gennemsnit),

Median eller andet kvartil (midten af observationerne, altså grænsen efter første halvdel)

Første kvartil eller **nedre kvartil** (Grænsen efter første fjerdedel af observationerne)

Tredje kvartil eller **øvre kvartil** (Grænsen efter tredje fjerdedel af observationerne)

Kvartilsæt (de 3 tal 1. kvartil, median og 3. kvartil)

Første kvartil og nedre kvartil er det samme. Ligeledes er tredje kvartil lig øvre kvartil.

Når vi arbejder med statistik kan vi gøre det på to måder. Nogen gange grupperer vi observationerne i intervaller. Så kalder vi det grupperede observationer.

Kvartiler og middeltal defineres forskelligt ved grupperede observationer og ved ikke grupperede observationer.

Ikke grupperede observationer

Median og kvartiler defineres ved ikke grupperede observationer således:

Observationerne sorteres i stigende orden

Ved et **ulige** antal observationer defineres **medianen** som den midterste observation.

Ved et **lige** antal observationer defineres **medianen** som midtpunktet af de 2 midterste observationer. Fx Medianen for 4 personer i alderen 4, 6, 8 og 10 år er 7 år.

Medianen kaldes også **anden kvartil** eller 2. kvartil.

Første kvartil eller 1. kvartil eller **nedre kvartil** defineres som medianen for de observationer, der er før hele observationssættets median.

Tredje kvartil eller 3. kvartil eller **øvre kvartil** defineres som medianen for de observationer, der er efter hele observationssættets median.

Kvartilsættet udgøres af 1. kvartil, 2.kvartil og 3. kvartil

Middeltal defineres ved ikke grupperede observationer således:

Middeltallet ved ikke grupperede observationer er summen af alle observationer divideret med antal observationer.

Eksempel:

Vi betragter 7 personer med følgende aldre: **3, 5, 9, 10, 12, 14, 20**

Median: 10

1. kvartil: 5

3. kvartil: 14

Kvartilsæt: 5, 10, 14

Mindsteværdi: 3

Størsteværdi: 20

Middeltal: $(3+5+9+10+12+14+20) / 7 = 10,4$ afrundet.

Boksplot (eller kassediagram):

En boksplot er en grafisk fremstilling af nogle observationer.

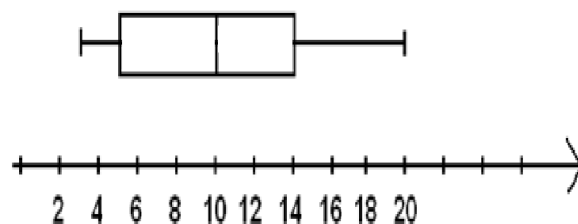
Boksplotten viser mindsteværdien, de 3 kvartiler og størsteværdien.

I en boksplot indgår et rektangel, hvor de to tider angiver 1. og 3. kvartil.

Parellelt med disse sider er inde i rektanglet et linjestykke, som viser 2. kvartil (medianen)

Vinkelret herpå, tværs gennem rektanglet, er et linjestykke, hvis endepunkter viser mindste og størsteværdi.

Boksplot



Grupperede observationer

Når der er mange forskellige observationer vælger man ofte at gruppere observationerne, fx:

Alder:	0 → 29	30 → 59	60 → 89	90 → 120
Alder:	[0;30[[30;60[[60;90[[90;120[
hyppighed (antal)	3500	3000	2000	1500

Sumkurve

Ved grupperede observationer tegner man ofte en såkaldt sumkurve. Inden vi ser hvordan må vi lige høre om yderligere et par deskriptorer:

Frekvensen angiver hvor stor en brøkdelt, der er i et interval i forhold til alle observationer.

$$\text{Frekvens} = \text{hyppighed} / \text{antal observationer i alt}$$

Frekvens kan angives i % eller som decimalbrøk, fx 35% eller 0,35

Kumuleret hyppighed, der beregnes ved at lægge hyppigheder sammen.

Kumuleret frekvens, der beregnes ved at lægge frekvenser sammen.

Hvis frekvenserne er afrundede tal, kan det være mere nøjagtigt i stedet at dividere kumuleret hyppighed med antal observationer i alt.

Her ses et skema med beregninger af de forskellige deskriptorer:

Alder:		0 → 29	30 → 59	60 → 89	90 → 120
Alder:		[0;30[[30;60[[60;90[[90;120[
Antal (hyppighed)	h	3500	3000	2000	1500
Kumuleret hyppighed	H	3500	3500+3000 = 6500	6500+2000 = 8500	8500+1500 = 10000
Frekvens (Benyttes ved tegning af et såkaldt histogram, som vi senere skal se)	f	$\frac{3500}{10000}$ = 0,35 = 35%	$\frac{3000}{10000}$ = 0,30 = 30%	20%	15%
Kumuleret frekvens (Benyttes ved tegning af en såkaldt sumkurve, som vi senere skal se)	F	35%	$\frac{6500}{10000}$ = 0,65 = 65%	$\frac{8500}{10000}$ = 0,85 = 85%	100%
Interval-midtpunkt Midtpunkt af aldersintervallet (Benyttes ved beregning af middeltal)		15	$\frac{(30+60)}{2}$ = 45	$\frac{(60+90)}{2}$ = 75	105

h, f, H og F er hyppigt anvendte forkortelser for hyppighed, frekvens, kumuleret hyppighed og kumuleret frekvens.

Den kumulerede frekvens **F** er en funktion, der fortæller hvor mange %, der er under en bestemt alder. Fx ses at 65% er under 60 år.

Grafen for denne funktion kaldes en **sumkurve** og sumkurver tegnes altid med rette linjestykker mellem støttepunkterne.

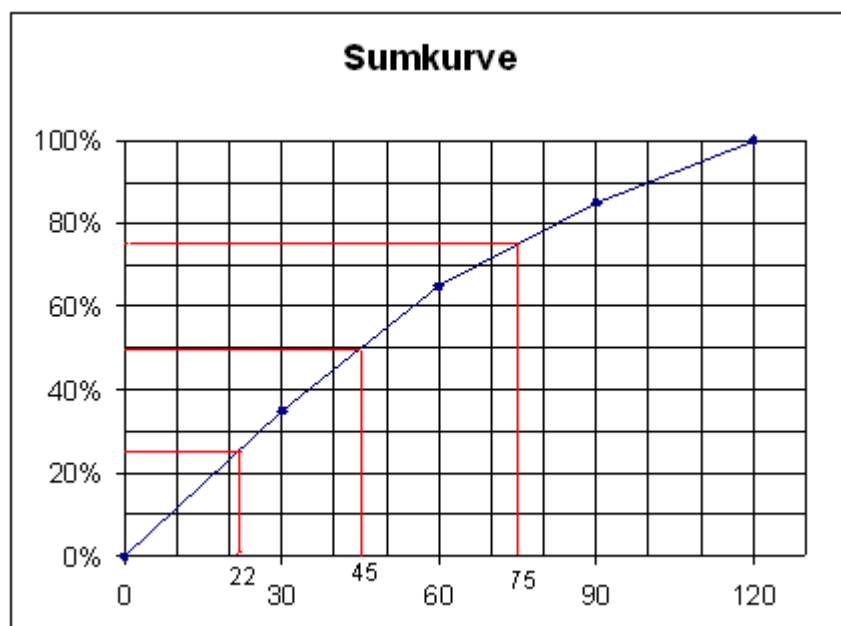
Skemaet fortæller ikke, hvordan de grupperede observationer fordeler sig i hvert interval, men man har vedtaget at betragte det som om, de fordeler sig jævnt i hvert interval.

Det er derfor, der tegnes rette linjestykker mellem støttepunkterne.

De rette linjestykker er udtryk for, at man betragter observationerne, som om de fordeler sig jævnt i hvert interval.

Støttepunkter:

Alder	0	30	60	90	120
F	0%	35%	65%	85%	100%



Median og kvartiler defineres ved grupperede observationer således:

Ved grupperede observationer aflæser vi kvartilsættet ved hjælp af sumkurven:

1. kvartil: Gå vandret fra 25% på 2.aksen til sumkurven og så lodret til 1.aksen.

Medianen: Gå vandret fra 50% på 2.aksen til sumkurven og så lodret til 1.aksen.

3. kvartil: Gå vandret fra 75% på 2.aksen til sumkurven og så lodret til 1.aksen.

Ved ovenstående sumkurve bliver kvartilsættet: 22 år, 45 år, 75 år

Middeltal ved grupperede observationer

Ved grupperede observationer kan man som tidligere nævnt ikke vide, hvordan observationerne fordeler sig i hvert interval, og man har vedtaget at betragte det som om, de fordeler sig jævnt.

Ved beregning af middeltallet kommer det ud på det samme, som at betragte det som om, alle observationer i hvert interval ligger midt i intervallet.

Derfor beregnes middeltallet ved for hvert interval at gange intervalmidtpunktet med intervallets frekvens og derefter lægge alle produkterne sammen

I ovenstående eksempel er intervalmidtpunkterne: **15, 45, 75 og 105**.

$$\begin{aligned}\text{Middeltallet} &= 15 \cdot 35\% + 45 \cdot 30\% + 75 \cdot 20\% + 105 \cdot 15\% \\ &= 15 \cdot 0,35 + 45 \cdot 0,30 + 75 \cdot 0,20 + 105 \cdot 0,15 = \mathbf{49,5}\end{aligned}$$

Illustration af forskel på middeltal og median

Man kan forestille ovennævnte aldersfordeling i en lille kommune.

Lad os antage at nogen af de 90 → 120 årige fraflytter kommunen.

I stedet flytter nogen endnu ældre på mellem 120 og 130 til kommunen i samme antal.

Derfor ændrer vi det sidste interval til 90 → 130.

Vil denne ændring ændre på medianen og/eller Middeltallet?

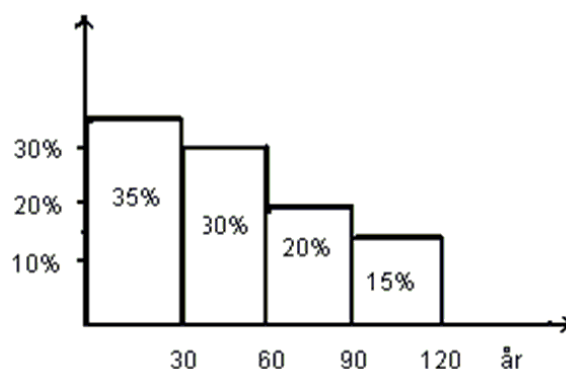
Medianen forbliver den samme, kun sidste del af sumkurven ændres. Det stykke, hvor medianen aflæses ændres ikke.

Derimod ændres middeltallet, fordi det sidste intervalmidtpunkt bliver ændret.

De meget gamle trækker i middeltallet, men ikke i medianen.

Histogram

Frekvensen (eller hyppigheden) kan fremstilles grafisk i et såkaldt histogram:



Boksplot bruges også ved grupperede observationer

og defineres på samme måde som ved ikke-grupperede observationer.