

Deskriptiv statistik: Ugrupperede observationer

Forestil dig, at du skal lave en undersøgelse af folks rejsevaner. Du udsender et spørgeskema til en af de nye 1.g klasser med 25 elever. Et af spørgsmålene lyder *Hvor mange lande har du besøgt?* og du får svarene:

8, 4, 16, 8, 9, 6, 16, 19, 7, 6, 4, 8, 11, 8, 9, 6, 9, 10, 11, 8, 14, 4, 6, 7, 10

Hvert tal kaldes en *observation*, og alle observationerne udgør tilsammen et *observationssæt*.

Når resultaterne af undersøgelsen skal præsenteres, er det vigtigt at gøre dette på en overskuelig og hensigtsmæssig måde. Dette vil vi beskæftige os med i den deskriptive statistik.

Tabel

En simpel måde at præsentere observationssættet er i en tabel. De væsentligste størrelser at præsentere er:

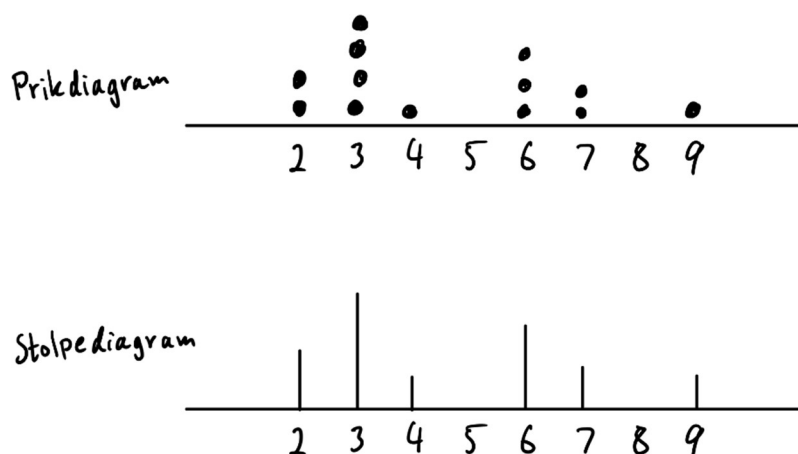
- *Størrelse*, N : Det samlede antal observationer
- *Hyppighed*, h : Antal gange den enkelte observation optræder
- *Frekvens*, f : Hyppigheden omsat til decimaltal eller procent, beregnes som $\frac{h}{N}$
- *Kumuleret hyppighed / frekvens*: Summen af alle hyppigheder / frekvenser op til og med den givne observation

Prikdiagram og stolpediagram

Et observationssæt præsenteres ofte grafisk med et *prikdiagram* eller et *stolpediagram* (søjlediagram). I prikdiagrammet tegnes en stabel af prikker svarende til hyppigheden for hver observation, og i stolpediagrammet tegnes stolper svarende til enten hyppigheden eller frekvensen.

Eksempel 1

Obs	Hyp	Frek
2	2	15 %
3	4	31 %
4	1	8 %
5	0	0 %
6	3	23 %
7	2	15 %
8	0	0 %
9	1	8 %
Total	13	100 %



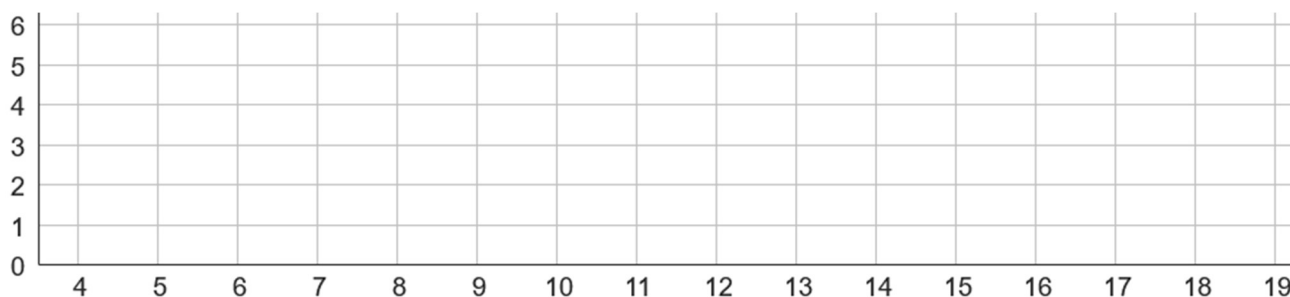
Opgave 1 Udfyld tabellen på baggrund af observationssættet på side 1. Skriv observationerne i rækkefølge med den mindste først.

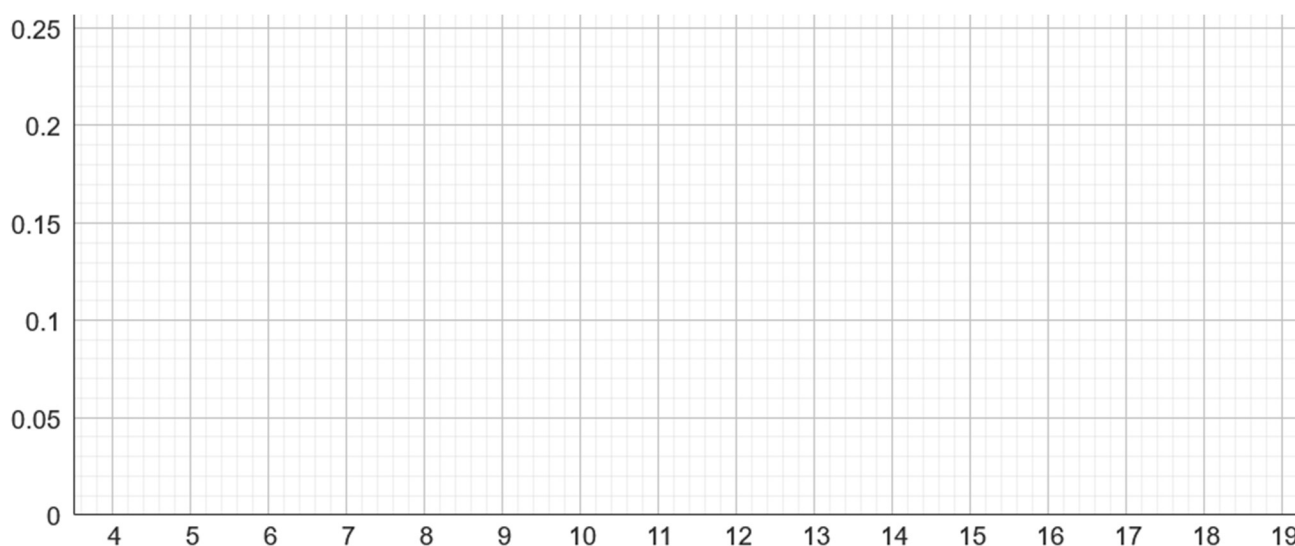
Observation	Hyppighed	Kum. hyppighed	Frekvens	Kum. frekvens
Total	25		100%	

Opgave 2 Hvor stor en del af eleverne har besøgt mindst 7 og højst 11 lande?

Hint: Overvej hvor mange der har besøgt hhv. mindre end 7 eller højst 11.

Opgave 3 Udfyld prikdiagrammet og stolpediagrammet på baggrund af observationssættet.





Opgave 4 (i dit hæfte)

Til et fodboldhold skal der indkøbes fodboldstøvler til 20 spillere. Træneren spørger derfor hver enkelt spiller, hvilken skostørrelse han har. Her er resultatet:

44, 42, 47, 44, 43, 43, 45, 42, 44, 42, 43, 44, 50, 45, 43, 41, 47, 42, 46, 44

- Lav en tabel med observationer, hyppigheder, kumulerede hyppigheder, frekvenser og kumulerede frekvenser.
- Tegn et prikdiagram over observationssættet.
- Hvor mange procent af spillerne har en skostørrelse på højst 42?
- Hvor mange procent af spillerne har en skostørrelse på mindst 45?
- Hvor mange procent af spillerne har en skostørrelse på mindst 43 og højst 46?

Deskriptorer

For at beskrive observationssættet anvendes en række *deskriptorer*, som er en række tal der hver især udtrykker en egenskab ved observationssættet. Her kan bl.a. nævnes:

- *Størrelse*, N : Det samlede antal observationer
- *Minimum*, min : Den mindste observation
- *Maksimum*, max : Den største observation
- *Variationsbredde*: Forskellen mellem den største og mindste observation, $max - min$
- *Typetal*: Den hyppigste observation (observationen med størst hyppighed)

Opgave 5 Angiv de relevante deskriptorer for observationssættet fra side 1.

Størrelse	Minimum	Maksimum	Variationsbredde	Typetal

Kvartiler

Ud over de ovennævnte deskriptorer benyttes også ofte observationssættets *kvartilsæt*. Dette bestemmes ved først at stille alle observationerne op i rækkefølge fra mindst til størst.

Opgave 6 Skriv observationssættet fra side 1 i rækkefølge fra mindst til størst (brug evt. din tabel fra opgave 1 for at gøre det lettere for dig selv)

Observationssættets *median*, m , også kaldet anden kvartil, er den midterste observation i rækken.

Opgave 7 Hvad er medianen for observationssættet i opgave 6?

Observationssættets *nedre kvartil*, Q_1 , og *øvre kvartil*, Q_3 , bestemmes som den midterste observation i hhv. den nederste halvdel og den øverste halvdel af observationssættet efter medianen er bestemt. Pas på: Hvis der er et lige antal observationer, er kvartilen i stedet *gennemsnittet af de to midterste observationer*.

Opgave 8 Hvad er de to kvartiler for observationssættet i opgave 6?

Q_1 :	Q_3 :
---------	---------

Kvartilbredden bestemmes som forskellen mellem øvre og nedre kvartil, $Q_3 - Q_1$.

Opgave 9 Hvad er kvartilbredden af observationssættet i opgave 6?

En *outlier* i observationssættet er en observation, der ligger mere end 1,5 gange kvartilbredden *under* nedre kvartil eller *over* øvre kvartil.

Opgave 10

a) Beregn 1,5 gange kvartilbredden fra opgave 9.

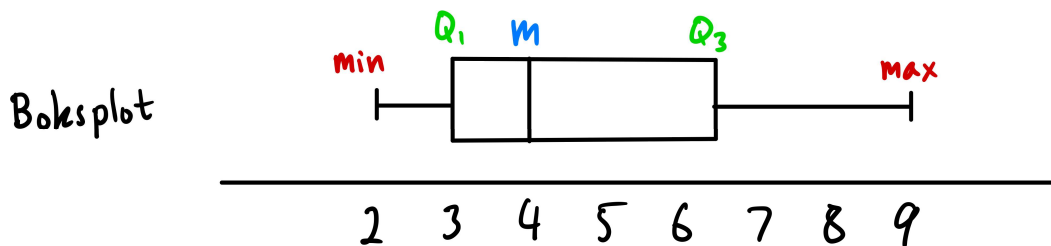
b) Beregn grænserne for outliers ved at trække tallet fra spørgsmål a fra Q_1 og ved at lægge samme tal til Q_3 .

$Q_1 - (\text{svar fra a}) :$	$Q_3 + (\text{svar fra a}) :$
-------------------------------	-------------------------------

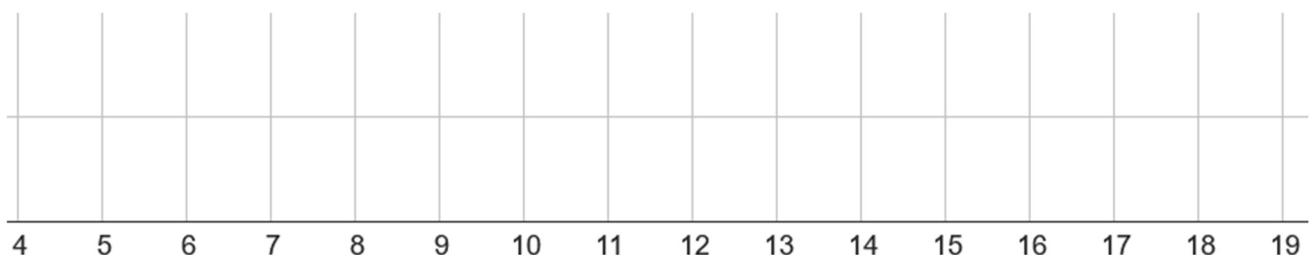
c) Er der outliers i observationssættet, og i så fald hvilke observationer er outliers? (Husk at en outlier er en observation der ligger *uden for* området bestemt i spørgsmål b)

Boksplot

Et observationssæts minimum, maksimum og kvartilsæt vises ofte grafisk i et *boksplot*, illustreret her med tallene fra eksempel 1.



Opgave 11 Tegn et boksplot for observationssættet i opgave 6.



Middelværdi, varians og spredning

En anden hyppigt anvendt deskriptor er observationssættets *middelværdi*, \bar{x} . Dette svarer til gennemsnittet af samtlige observationer og kan enten regnes ved at lægge samtlige observationer sammen og dividere med observationssættets størrelse eller ved at gange de enkelte observationer med frekvenserne og lægge sammen, vist herunder med tallene fra eksempel 1:

$$\bar{x} = \frac{2+2+3+3+3+3+4+6+6+6+7+7+9}{13} \blacktriangleright 4.69231$$

$$\bar{x} = 2 \cdot 0.15 + 3 \cdot 0.31 + 4 \cdot 0.08 + 6 \cdot 0.23 + 7 \cdot 0.15 + 9 \cdot 0.08 \blacktriangleright 4.7$$

Afvigelsen på decimalerne skyldes afrunding af frekvenserne.

Opgave 12 Beregn middelværdien af observationssættet i opgave 1.

Ud over middelværdien omtales også ofte observationssættets *variens*, *Var*, og *spredning*, σ . De siger begge noget om, hvordan observationerne er spredt i forhold til middelværdien. Variansen og spredningen beregnes som:

$$Var = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2$$

$$\sigma = \sqrt{Var}$$

Her er beregningerne vist med tallene fra eksempel 1:

$$Var = (2-4.69231)^2 + (3-4.69231)^2 + (4-4.69231)^2 + (6-4.69231)^2 + (7-4.69231)^2 + (9-4.69231)^2$$

$\blacktriangleright 36.1834$

$$\sigma = \sqrt{36.1834} \blacktriangleright 6.01526$$

Nu har du alle værktøjerne til at besvare opgaverne på næste side.

På sidste side vises en oversigt over de omtalte deskriptorer.

Opgaver fra plus B stx



Øvelse 8.1.2 - Ugrupperede observationer

Betragt følgende observationssæt:

- 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
- 1, 3, 5, 5, 8, 8, 9, 9, 10
- 1, 2, 2, 2, 3, 3, 6, 6, 8, 10

1. Bestem for hvert af observationssættene: Størrelse, min, max, variationsbredde, kvartilsæt, kvartilbredde, middelværdi og spredning.
2. Tegn boksplottene for de tre observationssæt.



Øvelse 8.1.3 - Ugrupperede observationer

På en arbejdsplads med 15 ansatte har lønforhandlinger betydet, at de ansattes månedsløn er blevet ændret med følgende størrelser (tallene er i tusinde kroner):

-2, -2, -1, -1, -1, -1, -1, -1, 0, 0, 0, 3, 6, 6, 10

Et negativt tal angiver en lønnedgang og 0 betyder, at lønnen er uændret.

1. Bestem observationsættets middelværdi og median.
2. Hvilket af de to tal, synes du, beskriver observationsættet bedst? Begrund dit svar.
3. Bestem kvartilbredden og find eventuelle outliers.



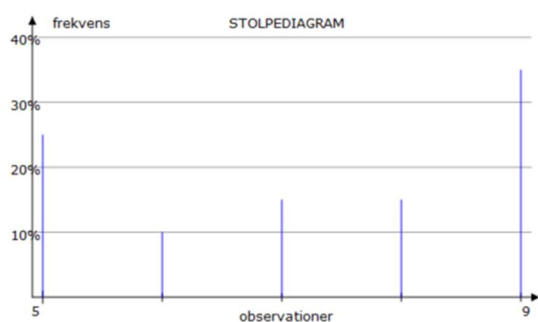
Øvelse 8.1.4 - Ugrupperede observationer

Bestem middelværdi og spredning for hvert af følgende observationssæt:

1.

Observation	4	5	8	10	11	16
Frekvens	9 %	25 %	17 %	34 %	10 %	5 %

2.



Øvelse 8.1.5 - Ugrupperede observationer

I øvelse 8.1.1 så vi på skostørrelserne for 20 fodboldspillere. Observationssættet er her opstillet i numerisk rækkefølge:

41, 42, 42, 42, 42, 43, 43, 43, 43, 44, 44, 44, 44, 44, 45, 45, 46, 47, 47, 50

1. Bestem størrelse, *min*, *max*, variationsbredde, kvartilsæt, kvartilbredde, middelværdi og spredning for observationssættet.
2. Er der nogen outliers?
3. Tegn et boksplot for observationssættet.

Oversigt over deskriptorer

Størrelse : Antallet af observationer

min : Den mindste observation

max : Den største observation

Variationsbredde : $max - min$

Median, m : Den "midterste observation"

Nedre kvartil, Q_1 : Medianen af observationssættets nederste halvdel

Øvre kvartil, Q_3 : Medianen af observationssættets øverste halvdel

Kvartilbredde : $Q_3 - Q_1$

Kvartilsæt : (Q_1, m, Q_3)

Udvidet kvartilsæt : (min, Q_1, m, Q_3, max)

Outlier : En observation under $Q_1 - 1,5 \cdot (Q_3 - Q_1)$ eller over $Q_3 + 1,5 \cdot (Q_3 - Q_1)$

Middelværdi : $\mu = \frac{\text{sum af observationer}}{\text{antallet af observationer}}$

Varians : Middelværdien af kvadraterne på forskellen mellem observationerne og deres middelværdi

Spredning : Kvadratroden af variansen, $\sigma = \sqrt{Var}$

Eksempel på kvartilsæt

ulige antal observationer: $\underbrace{2, 2, 3}_{n.\text{halvdel}}, \underbrace{5, 6, 6, 7}_{\emptyset.\text{halvdel}}$ kvartilsæt = (2; 5; 6)

lige antal observationer: $\underbrace{2, 3, 5}_{n.\text{halvdel}}, \underbrace{6, 6, 7}_{\emptyset.\text{halvdel}}$ kvartilsæt = (3; 5,5; 6)

Et boksplot er en grafisk fremstilling af min , max og kvartiler:

