

MATERIALESAMLING

MATEMATIK I HUMANITÆRE INDSATSER - 1. OG 2. G

INDHOLD

Afsnit 1	2
Afsnit 2	6
Afsnit 3	7
Afsnit 4	14
Afsnit 5	24
Afsnit 6	27
Bilag	31
Referencer	32

Materialet er udviklet i 2020 af:

Tina Petersen og Maria Simonsen, matematikstuderende, Syddansk Universitet

Claus Michelsen, professor, Laboratorium for Sammenhængende Uddannelse og Læring, SDU
og DA Åben Virksomhed



Afsnit 1

Introduktion - Naturkatastrofer

Dette forløb handler om nødhjælpsindsatser og naturkatastrofer i Danmark og ude i verden.

Flere af nødhjælpsorganisationerne bruger denne definition af, hvad en katastrofe er (1):

- En katastrofe er en pludseligt opstået katastrofal begivenhed, der alvorligt forstyrrer driften af et samfund og forårsager menneskelige, materielle og økonomiske eller miljømæssige tab, der overskrider samfundets evne til at klare det selv med egne midler.

I medierne hører man ofte om voldsomme skovbrande, oversvømmelser, jordskælv m.v. og følgerne heraf. Generelt betegnes naturkatastrofer som katastrofer, der er forårsaget af noget naturligt (2). Den menneskeskabte påvirkning af klimaet kan dog godt føre til en revurdering af denne definition (3).

Der er forskellige typer af naturkatastrofer:

- Klimatiske naturkatastrofer, hvor klimaet forårsager katastrofer med vind og vejr, f.eks. haglstorme, hedeølger, brande, tornadoer og orkaner.
- Geologiske naturkatastrofer, der startes ved eller under jordoverfladen, f.eks. vulkanudbrud, jordskælv og laviner.
- Hydrologiske naturkatastrofer er katastrofer, der omfatter vand, f.eks. oversvømmelser og flodbølger (opstår i reglen som følge af enten en klimatisk eller geologisk hændelse).

Nogle steder på kloden er mere udsatte for naturkatastrofer end andre. Eksempelvis er der højere risiko for jordskælv og vulkanudbrud i områder beliggende nær en pladegrænse mellem jordens tektoniske plader (4). Nogle af de lande, som er hårdest ramt af naturkatastrofer, er ofte de fattige lande (f.eks. Mozambique, Armenien, Sudan, Etiopien, osv.). Det er også disse lande, som har sværest ved at gardere sig imod naturkatastrofer.

Over de seneste år har medierne rapporteret om oversvømmelser og flodbølger i Danmark og udlandet. Disse hændelser går ind under kategorien hydrologiske naturkatastrofer (hvis de forårsager en katastrofe) og forårsages ofte af enten geologiske eller klimatiske naturkatastrofer som f.eks. jordskælv, vulkansk aktivitet eller tropiske cykloner/orkaner/tyfoner (begrebet afhænger af, hvor i verden disse finder sted: (cykloner er over Det Indiske Ocean, tyfoner er ved Stillehavet, og orkaner er ved Atlanterhavet) (5).

I dette undervisningsforløb vil vi fokusere på oversvømmelser. Oversvømmelser kan f.eks. ske som følge af en tsunami eller store mængder nedbør. Tsunamier kan opstå som følge af undersøiske jordskælv, jordskred og vulkansk aktivitet (6). Ved et jordskælv forskydes to tektoniske plader pludseligt i forhold til hinanden. Denne forskydelse kan være meget voldsom, og hvis det sker under havoverfladen, kan det forårsage bølger, som ude på det dybe ocean nemt kan bevæge sig med en hastighed på 500-1.000 km/t. Sker det i nærheden af en kyst, når hastigheden op omkring 10 km/t. Højden af bølgerne afhænger af vanddybden. Ude på det dybe ocean vil højden på bølgerne være omkring 1 m, men er de nær kysterne, vil højden blive påvirket af den lave dybde og nedsatte hastighed, og derved bliver bølgerne højere. Nogle bølger kan sagtens komme op omkring 10-30 m højde.

Danmark betragtes som et land uden for farezonen for naturkatastrofer. Søger man efter de værste naturkatastrofer i Danmark, kommer der intet frem. Mange husker orkanen i 1999, som en af de voldsomste orkaner i Danmark (7). Andre vil nævne de oversvømmelser, man er begyndt at opleve i Danmark. Det skyldes både øgede nedbørsmængder og kraftig blæst, som skaber forhøjet vandstand, og det kan især skabe problemer i de mere indelukkede havområder. Eksempelvis bliver områderne omkring Limfjorden, Roskilde fjord og Vadehavet oftere udsat for oversvømmelser, hvilket har givet beboerne i områderne store problemer.

Matematik spiller en afgørende rolle, når humanitære organisationer og virksomheder prøver at løse og beskrive de problematikker, de møder i deres kontakt med virkeligheden. Det vil sige, at de oversætter problematikkerne til matematisk sprog (eksempelvis funktioner, statistik, grafteori m.m.) for derefter at løse disse matematiske problemstillinger. Til sidst tolker de på de opnåede data, og på den måde finder de hjælp til at løse problematikkerne (8).

Vi vil i de kommende afsnit fokusere på oversvømmelser ved Roskilde Fjord (9). Data, der fremgår i afsnittene, er konstrueret, så de egner sig til undervisningsbrug, men er realistiske.

Efter I har besøgt en organisation eller virksomhed, der arbejder med humanitære indsatser, fokuseres der på oversvømmelser ude i verden, og her anvendes data indsamlet af en humanitær organisation.

Opgaver

Opgave 1.1*

I forbindelse med oversvømmelser ved Roskilde Fjord gør kommunen, beboerne og frivillige et stort stykke arbejde for at forberede sig på de store mængder vand. Til dette skal der f.eks. bruges sandsække, pumper og presenning. Med så mange huse og veje, der skal sikres, skal der dog transporteres en hel masse ud. Det er frivillige, som sørger for, at materialerne bliver kørt ud i Roskilde kommune (10).

Materialerne, der skal transporteres til byerne, er:

- Fyldte sandsække (18 cm høje, 26 cm brede, 45 cm i lange og vejer 13 kg)
- Pumper til private hjem (pakket i papkasser, så de er 37 cm høje, 50 cm brede, 71 cm lange og vejer 65 kg)
- Presenning (foldet sammen, så de er 15 cm høje, 43 cm brede, 56 cm lange og vejer 9 kg).

Det transporteres ud i trailere, der er 125 cm høje, 115 cm brede og 205 cm lange, hvor højden ikke kan overskrides på grund af presenningen. Der kan lastes i alt 635 kg. I opgaven repræsenteres materialerne af centicubes:

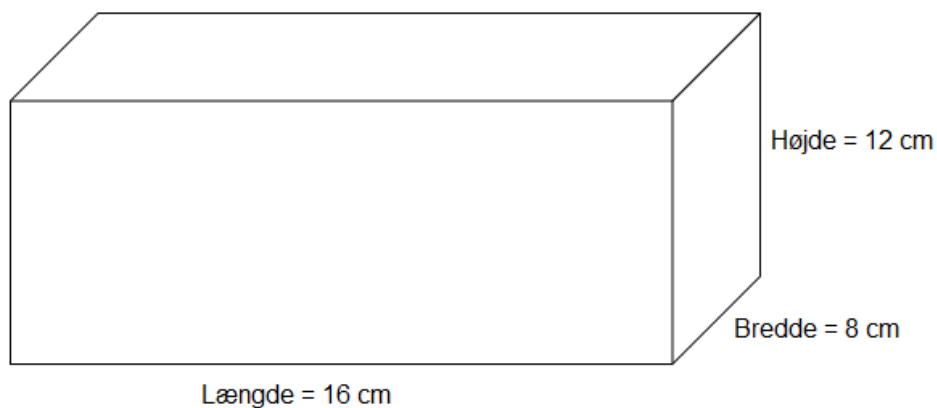
- Fyldte sandsække: 2 centicubes i højden, 2 centicubes i bredden og 4 centicubes i længden.
- Pumpe: 4 centicubes i højden, 4 centicubes i bredden og 8 centicubes i længden.
- Presenning: 1 centicube i højden, 4 centicubes i bredden og 5 centicubes i længden.

Alle tre typer materialer skal være på trailerne, da trailerne bliver kørt ud til 23 forskellige byer.

Alle de nedenstående materialer skal pakkes i trailere og transporteres til Roskilde Kommune*:

- 2.000 fyldte sandsække.
- 1.000 pumper.
- 2.700 presenninger.

Grundtegning af traileren:



Figur 1

1.1.a) Hvor mange stykker af hver type materiale kan der være på traileren, når max. lastvægten skal overholdes? Brug centicubes og grundtegningerne i bilaget. I kan udfylde skemaet nedenfor som hjælp. Der kan være flere rigtige svar.

Materialer:	Fyldte sandsække	Pumper	Presenning
Længde			
Bredde			
Højde			

1.1.b) Hvis der skal være dobbelt så mange sandsække som pumper, og 4 gange så mange presenninger som pumper, hvor mange stykker af hver type materiale kan der så være, uden at overskride vægtbegrænsningen?

1.1.c) Hvor mange af hver type materiale kan der være, hvis der skal være lige mange af hver?

1.1.d) Hvor mange trailere skal der bruges i alt for at transportere alt materiale ud?

Afsnit 2

Ligninger

Med afsæt jeres arbejde i forrige lektion arbejder vi i denne lektion videre med samme problemstilling om transport af materialer. I nedenstående opgave skal I arbejde med generalisering af ligninger, og hvordan det kan være med til at forsimple kommende transportopgaver af materialerne.

Opgaver

Opgave 2.1

Nu har I fundet ud af, hvor meget af hvert materiale der kan være i en trailer, men som nævnt i introen, er det smarte ved matematik, at man kan lave nogle generelle ligninger og derved få ny viden.

Derfor skal I nu lave generelle ligninger ud fra jeres resultater fra opgave 1. Hvor mange forskellige ligninger kan I lave? Og hvad fortæller ligningerne?

Afsnit 3

Statistik

Nødhjælpsorganisationerne står ofte i situationer, hvor de skal handle hurtigt i en situation uden fuldstændigt overblik og de første beslutninger tages ofte langt væk fra, hvor katastrofen er indtruffet. Samtidigt skal de med de økonomiske og menneskelige ressourcer, der er til rådighed, udvælge, hvor de skal sætte ind for at hjælpe bedst muligt.

Dataindsamling og statistik spiller en afgørende rolle i den forbindelse. Ofte vil man ikke have data på hver enkelt i en befolkning, men alene på en mindre gruppe. Hvis den stikprøve er udtaget rigtigt, så den afspejler hele befolkningen eller en bestemt gruppe af en befolkning, kan man bruge statistiske deskriptorer til at danne et første overblik.

Organisationerne træffer beslutninger på basis af en række forskellige input, f.eks.:

- tilgængelig statistik om situationen, typisk en stikprøve
- statistik om normale tilstande, f.eks. WHO-standarder om levevilkår og ernæring
- viden om, hvilke tiltag der hjælper bedst, typisk beskrevet ved statistik
- økonomi

Det handler dette og det næstkommende afsnit om.

Grupperede og ugrupperede datasæt

Indsamlede data kaldes for observationer. Man kunne eksempelvis indsamle observationer om alderen på de personer, som er berørt af en naturhændelse. Dette kunne resultere i følgende observationer:

36 år, 42 år, 44 år, 44 år, 50 år, 44 år, 33 år, 45 år, 45 år, 52 år, 44 år, 45 år, 50 år og 42 år.

Data kan inddeles i enten **grupperede eller ugrupperede datasæt**. Antallet af observationer, som man gør sig, betegnes med bogstavet n (totale antal observationer i undersøgelsen).

Om data er grupperede eller ikke grupperede, vil altid afhænge af indsamlingsmetoden. Det siger sig selv, at man opnår den største nøjagtighed, hvis man beder om meget præcise opgørelser, f.eks. hvis folk skal opgøre deres alder præcist (år, måneder, dage, timer etc.). Denne forskel kan eksempelvis være nødvendig i en situation, hvor man skal hjælpe nyfødte, men dette besværliggør i rigtig mange andre sammenhænge indsamlingen af data, da de færreste ved, præcist hvor gamle de

er på stående fod. Man risikerer derved at få et mindre datagrundlag og dermed større usikkerhed i det, som man ønsker at vide noget om.

Hvis det samtidig ikke har afgørende betydning for, hvad man skal sige noget om, kan man indsamle grupperede data, eksempelvis at folk alene opgør deres alder i år eller deres fødselsår. Dette vil i mange sammenhænge være nok til at sige noget om alderssammensætningen, og det vil formentlig løfte antallet af observationer.

Man kan altid gruppere ugrupperede data, men man kan ikke "u-gruppere" et grupperet datasæt.

Opgørelsen af data er altid et valg og en afvejning mellem på den ene side nøjagtighed af resultater og adgang til data og på den anden side mere unøjagtige resultater, men formentlig en lettere dataindsamling.

Desuden kan grupperede data sikre en overskuelighed, når resultater skal kommunikeres til andre. Eksempelvis vil man typisk arbejde med ikke bare år, men aldersgrupper, hvis man f.eks. skal kommunikere, hvilken del af befolkningen der er i risiko for at blive ramt af sygdom, sultkatastrofer eller lignende. Her vil det være uoverskueligt at opgøre alle aldre, og man vil ofte vælge alderstrin. Herved hjælper man læseren.

Endelig kan det være en grund til at arbejde med grupperede data, at man i nogle tilfælde har så få observationer inden for nogle kategorier, at der bliver for stor usikkerhed i ens resultater. Resultaterne vil i de allerfleste tilfælde jo være udtryk for en stikprøve – man måler ikke hele befolkningens vægt for at kunne konstatere en sultkatastrofe, men et repræsentativt udsnit. Tager vi igen eksemplet med alder, så kan det ske, at der ingen eller meget få data er for et bestemt alderstrin, f.eks. 64-årige. Derved kan vi ikke sige, hvordan den gruppe er påvirket, f.eks. i forhold til risiko for sygdom eller sult. Men hvis lægerne ved, at 64-årige generelt har samme levestandard som 63-årige og 65-årige, så kan det give god mening at gruppere data.

Tabellen til højre er et eksempel på, hvordan Folkekirkens Nødhjælp bruger data fra WHO og UNHCR. Her arbejdes der med aldersgrupper (det vil sige grupperede data), idet der regnes på, hvor stort et dagligt energiforbrug, grupperingerne har brug for. Der regnes med et gennemsnitligt energiforbrug (det vil sige en beregning af gennemsnit for grupperede data). Her

Age group	Males (kcal/day)	Females (kcal/day)	Average (kcal/day)	% of pop.
0-4 yr	1320	1250	1290	12.37%
5-9 yrs.	1980	1730	1860	11.69%
10-14 yrs.	2370	2040	2210	10.53%
15-19 yrs.	2700	2120	2420	9.54%
20-59 yrs.	2460	1990	2230	48.63
60+ yrs.	2010	1780	1890	7.24
Pregnant	285(extra)			2.4%
Lactating	500(extra)			2.6%
AVERAGE DAILY REQUIREMENT: 2100 Kcal/person/day				

Based on WHO technical Report No. 724 and UN Pop. Data, mid 1995
4/30/2020 InterWorks/UNHCR

2

har man altså valgt overblik, kommunikerbarhed frem for fuldstændig nøjagtighed, idet vi må tro på at de læger, der har opgjort data, ved, at forskelle inden for de opgjorte aldersgrupper er små, og grupperingen derfor har mindre betydning.

Vi vil selv regne på samme måde nedenfor. I eksempel 4.2 skal vi arbejde med gruppering af data. Datasættet er ret lille, hvorfor man i praksis ikke ville gruppere det, da det ikke ville skabe mere overskuelighed. Eksemplet skal derfor alene være med til at give forståelse for fremgangsmetoden i arbejdet med grupperede data, så vi kan anvende den i andre situationer.

Eksempel 3.1: Ugrupperede data

Vi genbruger datasættet fra før, der viser alderen af de berørte personer:

36 år, 42 år, 44 år, 44 år, 50 år, 44 år, 33 år, 45 år, 45 år, 52 år, 44 år, 45 år, 50 år og 42 år.

I dette tilfælde er antallet af observationer lig 14, hvilket kan noteres som $n = 14$.

Eksempel 3.2: Grupperede data

Vi kunne have målt nedbørsmængden for alle dage med nedbør i februar 2020, som blev den vådeste i Danmark nogensinde. Det kunne have ført til observationerne:

1mm; 5,5mm; 10mm; 6,6mm; 7,1mm;
13,5mm; 9,3mm; 12,2mm; 4,5mm

Her kan data f.eks. inddeles i følgende intervaller: $]0;5]$, $]5;10]$ og $]10;15]$. Ud fra observationsintervallerne og observationerne kan så vi lave følgende tabel:

Observationsinterval	Antal observationer i intervallet
]0;5]	2
]5;10]	4
]10;15]	3

I dette tilfælde er det samlede antal observationer lig 9, hvilket kan noteres som $n = 9$.

Middelværdi

Vi er interesserede i at undersøge, hvad de indsamlede data fortæller os, og hvad vi kan sige på baggrund af data. Ovenfor ser vi eksempelvis på de berørte personers alder og på nedbørsmængder i februar. Nødhjælpsorganisationerne vil altid udvælge data ud fra viden og hypoteser, f.eks. om hvor meget nedbør der typisk fører til oversvømmelser i forhold til, hvad jorden kan optage. Man ser derfor på et gennemsnit af, hvor meget regnvand der normalt falder i en måned og sammenligner observationer hermed.

Når vi finder gennemsnittet af data, finder vi **middelværdien**, som kendes ved symbolet $\bar{x}$.

Lad os beregne middelværdien for henholdsvis et grupperet og et ugrupperet datasæt. Vi begynder med et eksempel på et ugrupperet datasæt:

Eksempel 3.3: Middelværdi ugrupperet datasæt

Med afsæt i eksempel 3.1 har vi følgende observationer:

36 år, 42 år, 44 år, 44 år, 50 år, 44 år, 33 år, 45 år, 45 år, 52 år, 44 år, 45 år, 50 år, 42 år

Her finder man middelværdien ved at gøre følgende:

$$\bar{x} = \frac{36 + (42 \cdot 2) + (44 \cdot 4) + (50 \cdot 2) + 33 + (45 \cdot 3) + 52}{14} = 44 \text{ år}$$

Generalisering

Fremgangsmåden for at finde middelværdien for et ugrupperet datasæt kan generaliseres til følgende:

Middelværdi for ugrupperet datasæt:
--

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

I formelen er n antallet af observationer og $x_1, x_2, \dots$ til x_i er tallene fra vores datasæt

Lad os nu beregne middelværdien på et grupperet datasæt:

Eksempel 3.4: Middelværdi grupperede datasæt

Med afsæt i eksempel 3.2 har vi følgende tabel:

Observationsinterval	Antal observationer i intervallet
]0;5]	2
]5;10]	4
]10;15]	3

For at finde middelværdien for grupperede datasæt, skal man først finde intervalmidtpunktet, som findes ved at bruge formelen:

$$m = \frac{a + b}{2}$$

Hvor a og b er intervalendepunkterne i observationsintervallerne.

For det første observationsinterval er $a = 0$ og $b = 5$, hvorfor intervalmidtpunktet bliver $m = 2,5$. Beregnes intervalmidtpunktet for de følgende 2 observationsintervaller, så kan tabellen udvides til følgende:

Observationsinterval	Intervalmidtpunkt	Antal observationer i intervallet
]0;5]	2,5	2
]5;10]	7,5	4
]10;15]	12,5	3

Herfra er det muligt at beregne middelværdien ved brug af samme formel som for de ugrupperede observationer. I stedet for at indsætte observationerne, indsættes de beregnede intervalmidtpunkter. Dermed finder vi middelværdien på følgende måde:

$$\bar{x} = \frac{(2,5 \cdot 2) + (7,5 \cdot 4) + (12,5 \cdot 3)}{9} = 8,1 \text{ mm}$$

Hvis vi i stedet beregnede middelværdien af samme datasæt med ugrupperede data, får vi middelværdien 7,74. Ovenstående unøjagtighed vil jo følge med i vores beregninger i de kommende opgaver, når man bruger middeltallet som udgangspunkt.

Overvej i forlængelse heraf, selv om og i givet fald hvornår det, på trods af den unøjagtighed som fremkommer, alligevel kan være rimeligt at gruppere data.

Generalisering

Fremgangsmåden for at finde middelværdien for et grupperet datasæt kan generaliseres til følgende:

Middelværdi for grupperede datasæt:

Først findes intervalmidtpunktet for hvert enkelt observationsinterval med formlen:

$$m = \frac{a + b}{2}$$

Hvor m er intervalmidtpunktet, og a og b er observationsintervallets endepunkter.

Dernæst summeres der over antallet af observationer i hvert interval for at finde middelværdien ved brug af formelen:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n m_i = \frac{m_1 + m_2 + \dots + m_i}{n}$$

Hvor $m_1, m_2, \dots$ til m_i svarer til de fundne intervalmidtpunkter fra observationsintervallerne.

Opgaver

Opgave 3.1

Find middelværdien for følgende datasæt (*) over adspurgte personers alder, som blev berørt af en oversvømmelse ved Jyllinge Fjord:

9 år, 11 år, 18 år, 20 år, 23 år, 36 år, 39 år, 40 år, 41 år, 25 år, 25 år, 8 år, 41 år (dette datasæt kalder vi "Alder")

Opgave 3.2

Efter oversvømmelsen ved Isefjord og Roskilde Fjord har man målt vandstanden i de omkringliggende byer og fået nedenstående datasæt, hvor målingerne angivet i cm (11):

10,87 cm; 62 cm; 94,2 cm; 41,78 cm; 25,22 cm; 6,34 cm; 82,76 cm; 59,16 cm; 74,23 cm; 39,51 cm; 88 cm; 18,92 cm; 53 cm; 99,99 cm (dette datasæt kalder vi "Nedbør")

- Gruppér datasættet og find middelværdien for datasættet.
- Beregn herefter middelværdien af de samme ugrupperede data. Hvad betyder forskellen for nøjagtigheden af resultatet?
 - Diskutér hvorfor man i denne situation skal bruge enten grupperede eller ugrupperede data.

Opgave 3.3

Prøv selv at finde eller lave et andet datasæt og beregn middelværdien. Du må gerne vælge både positive og negative værdier eller værdier, som ikke er hele (f.eks. 0,78).

Opgave 3.4

Diskutér hvad middelværdien kan bruges til. Nævn 2 ting, som middelværdien kan fortælle om et datasæt, og i hvilke situationer i forbindelse med katastrofehjælp ved oversvømmelse ved Roskilde Fjord, hvor det er relevant at kende middelværdien af et datasæt.

Afsnit 4

Statistik

Vi fortsætter med beregning af statistiske deskriptorer i dette afsnit og bliver klogere på, hvordan statistik kan medvirke til at give nødhjælpsorganisationerne et overblik at tage beslutninger ud fra.

I et af eksemplerne oven for omtales underernæring. 0-5-årige børns vækst monitoreres i mange lande. Til dette formål anvendes WHO's vækstkurver som vækststandard for børn i hele verden. Kurverne er baseret på målinger af 888 børn i seks lande, der er målt 21 gange gennem de første leveår. For at måle, om et barn vokser og er i trivsel, holdes målinger af det enkelte barn op imod standarden.

Enhver måling er nødvendigvis et punktnedslag, men børn vokser hele tiden. Desuden er alle mennesker ikke ens – og skal ikke være det – og man kan ikke konstatere bekymring for et barns vækst eller evt. underernæring, fordi et barn ikke rammer gennemsnittet. Det normale har vide rammer, som man siger. For at tage højde for dette introducerer vi nu endnu en række statistiske deskriptorer. Vi vil nu se på, hvordan den procentvise fordeling afspejler sig. Til dette kan vi anvende en **sumkurve for grupperede observationer** og **trappediagrammer for ugrupperede observationer**. Ud fra disse diagrammer er det muligt at indtegne **kvartilsæt**, som kan bruges til at beskrive procentvise forhold i de indsamlede data.

Frekvens og kumuleret frekvens

For at kunne tegne en sumkurve eller et trappediagram skal vi først beregne **frekvensen**.

Frekvensen er et udtryk for, hvor hyppigt en observation forekommer ud af alle observationer og betegnes med bogstavet f . Hyppigheden er det antal gange, en observation forekommer, og betegnes med bogstavet h .

For at gøre det hele mere overskueligt, så lad os se på nogle eksempler for, hvordan man kan beregne frekvensen for både et grupperet og et ugrupperet datasæt.

Eksempel 4.1: Frekvens for ugrupperede data

Vi tager udgangspunkt i det ugrupperede datasæt fra eksempel 3.1. For at beregne frekvensen skal man kende hyppigheden (hvor ofte en observation forekommer) af de enkelte alderstrin ud af det totale antal observationer. Frekvensen for alderen 33 år kan beregnes ved at gøre følgende:

$$frekvens = \frac{1}{14} = 0,071 \cdot 100 = 7 \text{ pct.}$$

Gøres dette for alle alderstrin, kan de beregnede frekvenser samles i en tabel, som det ses nedenfor:

Observation	Hyppighed	Frekvens
33 år	1	7 pct.
36 år	1	7 pct.
42 år	2	14 pct.
44 år	4	29 pct.
45 år	3	22 pct.
50 år	2	14 pct.
52 år	1	7 pct.

Eksempel 4.2: Frekvens for grupperede data

Vi tager udgangspunkt i det grupperede datasæt fra eksempel 3.2. Her kan frekvensen for observationsintervallet]0;5] beregnes ved at gøre følgende:

$$frekvens = \frac{2}{9} = 0,22 \cdot 100 = 22 \text{ pct.}$$

Gøres dette for alle observationsintervaller, kan de beregnede frekvenser samles i en tabel, som det ses nedenfor:

Observationsinterval	Hyppighed (antal observationer i intervallet)	Frekvens
]0;5]	2	22 pct.
]5;10]	4	45 pct.
]10;15]	3	33 pct.

Generalisering

Fremgangsmåden for at finde frekvensen for henholdsvis et grupperet og ugrupperet datasæt kan generaliseres til følgende:

Frekvens for ugrupperede og grupperede datasæt:

$$frekvens = \frac{hyppighed}{\text{totalt antal observationer}}$$

Samme formel kan også skrives som:

$$f_i = \frac{h_i}{n}$$

Frekvensen f_i er svarende til den i 'te observation i datasættet, altså frekvensen for én af observationerne i datasættet og den tilhørende hyppighed for observationen h_i .

Er der tale om et grupperet datasæt, svarer hyppigheden til antal observationer i de enkelte observationsintervaller.

Opgaver

Opgave 4.1

Ud fra datasættet "Alder" (opg. 3.1) skal I opstille en tabel, hvor I også beregner frekvensen. Se eventuelt eksempel 3.1.

Opgave 4.2

Ud fra datasættet "Nedbør" (opg. 3.2) skal I opstille en tabel, hvor I også beregner frekvensen. Se eventuelt eksempel 3.2.

Kumuleret frekvens

Efter at have beregnet frekvensen, skal vi nu beregne den **kumulerede frekvens**, som skal bruges i diagrammerne senere. Den kumulerede frekvens får man ved at lægge frekvenserne for de enkelte observationer sammen. At kumulere betyder at "opsamle", hvorfor den kumulerede frekvens betyder, at de beregnede frekvenser samles op en efter en.

For at beregne den kumulerede frekvens, lægges frekvenserne sammen, så at de til slut giver 100 pct. Afrundinger af frekvenserne kan dog nogle gange medføre, at den totale kumulerede frekvens bliver lidt højere eller lavere end 100 pct.

Den kumulerede frekvens beregnes på samme måde for grupperede og ugrupperede data, da det er de tidligere beregnede frekvenser, som lægges sammen. Lad os igen se på nogle eksempler.

Eksempel 4.3: Kumuleret frekvens for ugrupperede data

Lad os tage udgangspunkt i tabellen fra eksempel 4.1, hvor vi beregnede frekvensen for de enkelte

alderstrin. Nu lægges frekvenserne sammen, hvorved vi får den kumulerede frekvens. F.eks. beregnes den kumulerede frekvens for 42 år på følgende måde:

$$\text{Kumuleret frekvens} = 7 \text{ pct.} + 7 \text{ pct.} + 14 \text{ pct.} = 28 \text{ pct.}$$

Gøres dette for alle frekvenser, får vi følgende tabel:

Observation	Hyppighed	Frekvens	Kumuleret frekvens
33 år	3	7 pct.	7 pct.
36 år	1	7 pct.	14 pct.
42 år	1	14 pct.	28 pct.
44 år	1	29 pct.	57 pct.
45 år	1	22 pct.	79 pct.
50 år	2	14 pct.	93 pct.
52 år	2	7 pct.	100 pct.

Eksempel 4.4: Kumuleret frekvens for grupperede data

Den kumulerede frekvens for grupperede data findes på samme måde som for ugrupperede data. Lad os tage udgangspunkt i tabellen fra eksempel 4.2, hvor vi fandt frekvenserne for de grupperede data. Nu lægges frekvenserne sammen, hvorved vi får den kumulerede frekvens i følgende tabel:

Observationsinterval	Antal observationer i	Frekvens	Kumuleret frekvens
]0;5]	2	22 pct.	22 pct.
]5;10]	4	45 pct.	67 pct.
]10;15]	3	33 pct.	100 pct.

Generalisering

Fremgangsmåden for at finde den kumulerede frekvens for henholdsvis et grupperet og ugrupperet datasæt kan generaliseres til følgende:

Kumuleret frekvens for ugrupperede og grupperede datasæt:

$$\text{kumuleret frekvens} = f_1 + f_2 + \dots + f_{i-1} + f_i = \sum_{k=1}^i f_k$$

Hvor f_k er frekvenserne for de enkelte observationer, som lægges sammen for at få den kumulerede frekvens.

Opgaver

Opgave 4.3

Med afsæt i tabellen fra datasættet ”Alder” i opgave 4.1; udvid tabellen og beregn den kumulerede frekvens.

Opgave 4.4

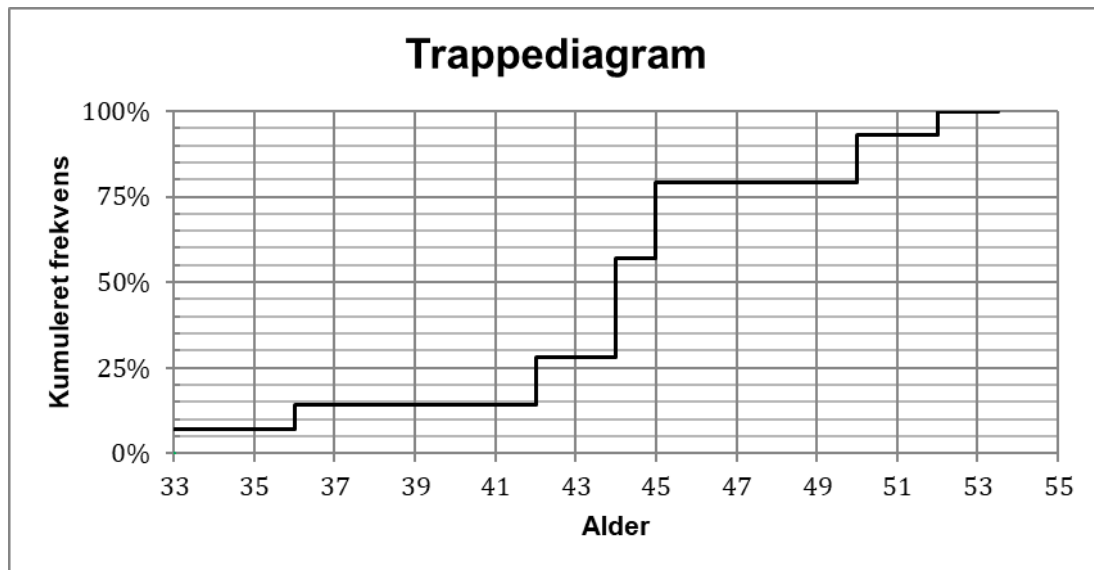
Med afsæt i tabellen fra datasættet ”Nedbør” i opgave 4.2; udvid tabellen og beregn den kumulerede frekvens.

Trappediagram og sumkurve

Ud fra vores observationer og beregnede kumulerede frekvenser kan vi lave enten en **sumkurve** eller **et trappediagram** alt afhængig af, om der er tale om et grupperet eller ugrupperet datasæt. For de ugrupperede datasæt laves der trappediagrammer, mens der for de grupperede datasæt laves sumkurver. Lad os lave et diagram over datasættene fra eksempel 1 og 2 ved at bruge de kumulerede frekvenser fra eksemplerne 4.3 og 4.4.

Eksempel 4.5: Trappediagram – ugrupperede data

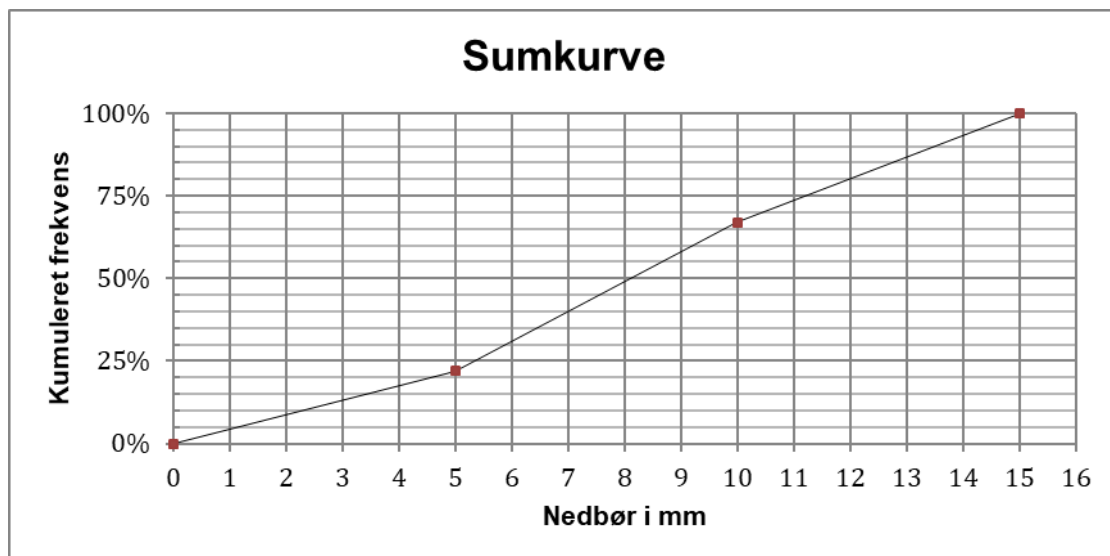
På baggrund af det ugrupperede datasæt fra eksempel 4.1 og de kumulerede frekvenser fra eksempel 4.3 er det muligt at lave følgende trappediagram:



Figur 2

Eksempel 4.6: Sumkurve – grupperede data

På baggrund af det grupperede datasæt fra eksempel 3.2 og de kumulerede frekvenser fra eksempel 4.4, er det muligt at lave følgende sumkurve:



Figur 3

Opgaver

Opgave 4.5

Ud fra observationer og den fundne kumulerede frekvens af datasættet ”Alder” (opg. 4.3) skal I fremstille et passende diagram (vælg enten trappekurve eller sumkurve).

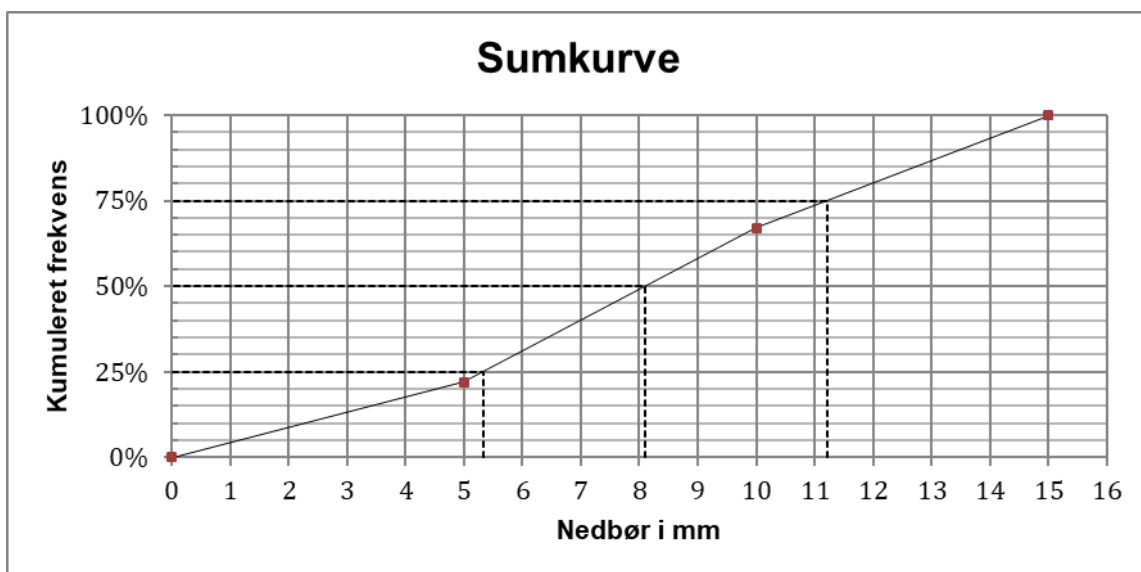
Opgave 4.6

Ud fra observationsintervallerne og den fundne kumulerede frekvens af datasættet "Nedbør" (opg. 4.4) skal I fremstille et passende diagram (vælg enten trappekurve eller sumkurve).

Kvartilsæt

Ud fra diagrammerne er det muligt at se, hvordan den procentvise fordeling afspejler sig i datasættet ved brug af kvartilsæt. Kvartilsæt inddeler datasættet i fire lige store dele og består af 3 tal, som alle kan aflæses: **Øvre kvartil**, **median** og **nedre kvartil**.

Lad os først indtegne kvartilsættene på sumkurven ovenfor og dernæst se på, hvad de betyder:



Figur 13

De stiplede linjer angiver kvartilsættene for datasættet. For at aflæse kvartilsættene starter man ved henholdsvis 25 pct., 50 pct. og 75 pct. på y-aksen. Herfra går man vandret til højre, indtil man støder på sumkurven. Når man rammer kurven, så går man lodret ned og aflæser, hvor man rammer x-aksen. Dette tal er den pågældende kvartil.

For at finde nedre kvartil går man ud fra 25 pct. på y-aksen, for at finde medianen går man ud fra 50 pct. og for at finde øvre kvartil, går man ud fra 75 pct.

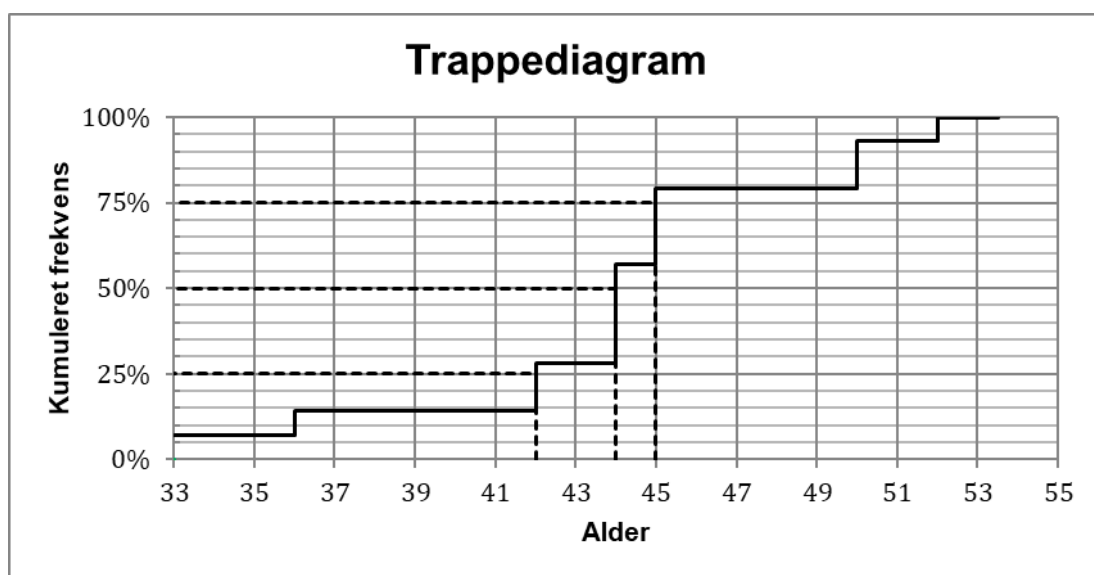
Lad os se på, hvordan man kan aflæse kvartilsættet på sumkurven og hvad det fortæller om fordelingen af data.

Medianen er det midterste tal af alle observationerne, når observationerne ordnes efter størrelse. Det vil altså sige, at medianen er det tal, som 50 pct. af observationerne er mindre end, og de resterende 50 pct. observationer er større end. I figur 13 vil det sige, at 50 pct. af nedbørsmængden, der er faldet, er 8,1 mm eller mindre.

Nedre kvartil er det tal, som 25 pct. af observationerne er mindre end, og de resterende 75 pct. observationer er større end. I den ovenstående sumkurve vil det sige, at 25 pct. af nedbørsmængden, som er faldet, er 5,3 mm eller mindre.

Øvre kvartil er det tal, som 75 pct. af observationerne er mindre end, og de resterende 25 pct. observationer er større end. I den ovenstående sumkurve vil det sige, at 75 pct. af nedbørsmængden, som er faldet, svarer til 11,2 mm eller mindre.

Trappediagrammer ser anderledes ud end sumkurverne, men principperne med kvartilsæt er de samme. Vi tager igen udgangspunkt i det trappediagram, som vi tegnede tidligere:



Figur 4

På samme måde som i sumkurven afspejler de stiplede linjer kvartilsættene. Kvartilsættene aflæses ved at starte på y-aksen, hvorfra man går vandret til højre, indtil man støder på grafen. Når man rammer grafen, går man lodret ned og det tal, som man rammer på x-aksen, svarer til den pågældende kvartil. Kvartilsættene er defineret på samme måde som for sumkurverne.

Generalisering

Betydningen af kvartilsættene for trappediagrammet og sumkurven kan beskrives på følgende måde:

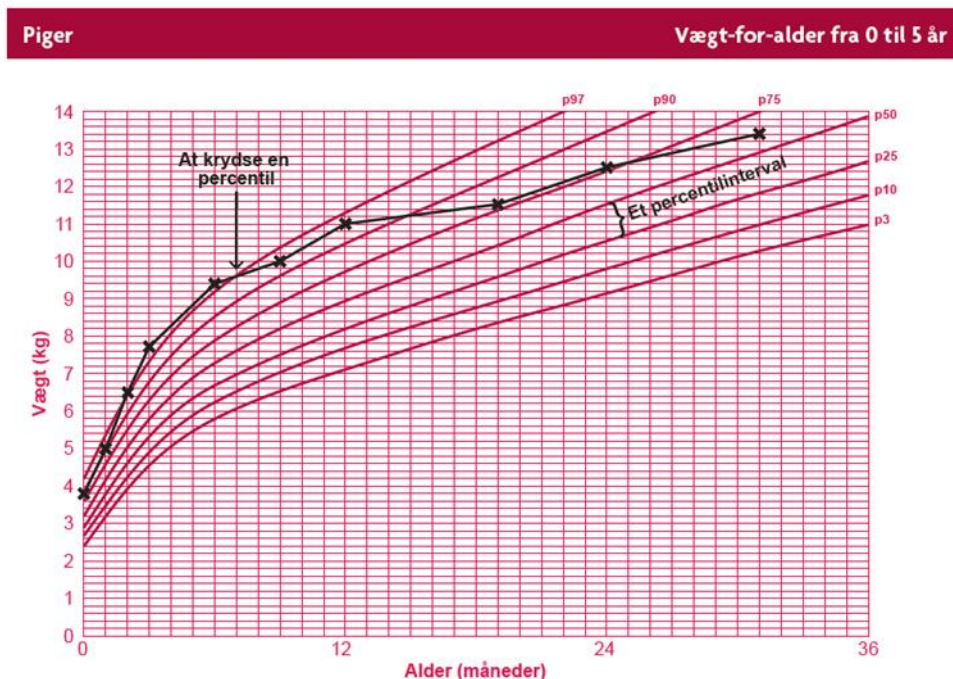
Kvartilsæt (for både sumkurve og trappediagram)

Medianen er det midterste tal af alle observationerne. Det vil altså sige, at medianen er det tal som 50 pct. af observationerne er mindre end, og de resterende 50 pct. observationer er større end.

Nedre kvartil er det tal, som 25 pct. af observationerne er mindre end, og de resterende 75 pct. observationer er større end.

Øvre kvartil er det tal, som 75 pct. af observationerne er mindre end, og de resterende 25 pct. observationer er større end.

Jf. indledningen så anvendes sumkurver og percentiler eksempelvis til at følge børns vækst. Nedenfor fremgår et eksempel på WHO's sumkurve med indtegnede percentiler og et konkret vækstmønster for et enkelt barn (12).



Opgaver

Opgave 4.7

I det diagram I lavede i opgave 4.5, skal I finde nedre kvartil, median og øvre kvartil. Hvad fortæller de om diagrammet?

Opgave 4.8

I det diagram I lavede i opgave 4.6, skal I finde nedre kvartil, median og øvre kvartil. Hvad fortæller de om diagrammet?

Opgave 4.9

Diskutér, hvad sumkurve, trappediagram og kvartilsættene kan fortælle os om et datasæt.

Afsnit 5

Opgaver

Store vandstandsstigninger i Danmark volder problemer landet over. Et fremtidsscenarie i år 2030 kunne være, at Isefjord og Roskilde Fjord går over deres bredder og oversvømmer hele 23 byer i området. Mange familier får oversvømmet deres huse og har brug for hjælp.



Figur 5: Hentet fra Google Maps

I nedenstående tabel ses antallet af indbyggere i de påvirkede byer samt det rapporterede antal af påvirkede familier i de enkelte byer (*):

Byer	Indbyggere	Påvirkede familier
Jægerspris	8.018	1.704
Tangbjerg	4.289	1.006

Skuldelev	5.326	1.166
Sønderby	1.482	202
Sæby	9.355	1.222
Veddelev	5.782	1.406
Lyndby	4.643	1.077
Herslev	2.618	411
Ølsted	6.841	1.401
Rørvig	4.021	997
Nykøbing Sjælland	7.786	1.874
Hundested	8.597	2.101
Liseleje	7.521	1.591
Ejby	7.034	1.337
Kyndby Huse	961	87
Over Dråby	3.382	623
Vellerup	3.751	812
Orø	6.482	1.399
Frederiksværk	9.936	1.582
Frederikssund	1.964	2.529
Jyllinge	1.,731	2.389
Ågerup	6.183	955
Himmelev	11.592	1.223
Totalt	150.295	29.094

Opgave 5.1

Overvej, om data (*) i ovenstående tabel bedst repræsenteres ugrupperet eller grupperet. Hvorfor?

Opgave 5.2

Beregn varians og spredning for datasættet af påvirkede familier og beskriv, hvad det forklarer.

Opgave 5.3

Beregn frekvensen af påvirkede familier i de indsamlede data, alt efter om der er valgt grupperet eller ugrupperet inddeling.

Opgave 5.4

Konstruer sumkurve eller trappediagram over datasættet og indtegn kvartilsæt. Hvor giver det bedst mening at sætte hjælp ind?

Opgave 5.5

I skal give informationer videre til kommandocentralen, som skal tage beslutninger om, hvilke byer der har brug for hjælp. Nævn de tre vigtigste informationer fra datasættet, som kan hjælpe kommandocentralen med at sende hjælpen ud.

Afsnit 6

Afsluttende opgave

Efter jeres besøg i en humanitær organisation skal I arbejde med en opgave, hvor I skal hjælpe med at finde ud af, hvor hjælpen skal sendes hen:

- 1) I skal analysere det givne datasæt, hvor I bl.a. inkluderer middelværdi, diagrammer og kvartilsæt. Her skal I bruge alt det, I har lært om i undervisningsforløbet.
- 2) I skal formidle jeres analyse i en kort artikel. Bagefter sender læreren nogle af artiklerne til den organisation, som I har besøgt. Nedenunder er det beskrevet, hvad artiklen skal indeholde.

Artikel

I skal fremstille en kort artikel ud fra jeres datasæt, som skal indeholde følgende:

- 1) En introduktion til det land datasættet drejer sig om.
- 2) Præsentation af jeres analyse:
 - Fremstil relevante tabeller og diagrammer, hvor I beskriver, hvad det fortæller. (brug det I har lært i afsnit 3 og 4 om statistik til at vælge, hvilken type diagrammer det giver mest mening at bruge til at analysere jeres datasæt).
 - Kom med en forklaring om, hvor I vil hjælpe i landet.
- 3) Jeres artikel skal være velformuleret og lavet i et flot design, så det kan være med til at sætte fokus på nødhjælp.

Til slutopgaven skal I benytte datasættet fra Malawi på de følgende sider, med mindre den humanitære organisation, I har besøgt, har givet jer et alternativt datasæt.

Introduktion til opgaven

I begyndelsen af marts 2019 ramte et hårdt vejrsystem Malawi med kraftig regn ledsaget af en kraftig vind. Vejrsystemet blev intensiveret i cyklonen Idai og ramte Malawi for anden gang. Den



Figur 7: Hentet fra Google Maps

kraftige og vedvarende regn førte til alvorlige oversvømmelser på tværs af nogle distrikter i det sydlige Malawi. Mere end 868.900 mennesker blev ramt, herunder blev mere end 86.980 fordrevet,



Figur 8: Hentet fra Google Maps

og det resulterede ifølge regeringen i 60 dødsfald og 672 sårede. I alt blev 15 distrikter og 2 byer påvirket (13).

Datasæt

Nedenstående datasæt er hentet fra www.reliefweb.int. Datasættet giver et overblik over de ramte distrikter i Malawi og viser, hvor hårdt populationen er påvirket.

Distrikt	Population i 2018	Påvirkede	Påvirkede
Balaka	438.379	964	4.338
Blantyre	451.220	2.142	9.639
Blantyre City	800.264	8.249	37.121
Chikwawa	564.684	13.602	61.209
Dedza	830.512	8.637	38.867
Machinga	735.438	29.301	131.855
Mangochi	1.148.611	14.927	67.172
Mulanje	684.107	1.909	8.591
Mwanza	130.949	2.912	13.104
Neno	138.291	1.993	8.969
Nsanje	299.168	18.000	81.000
Ntcheu	659.608	5.674	25.533
Phalombe	429.450	22.848	102.816
Thyolo	721.456	10.312	46.404
Zomba City	105.013	789	3.551
Zomba	746.724	29.892	134.514
Chiradzulu	356.875	20.937	94.215
Totalt	9.240.749	193.088	868.895

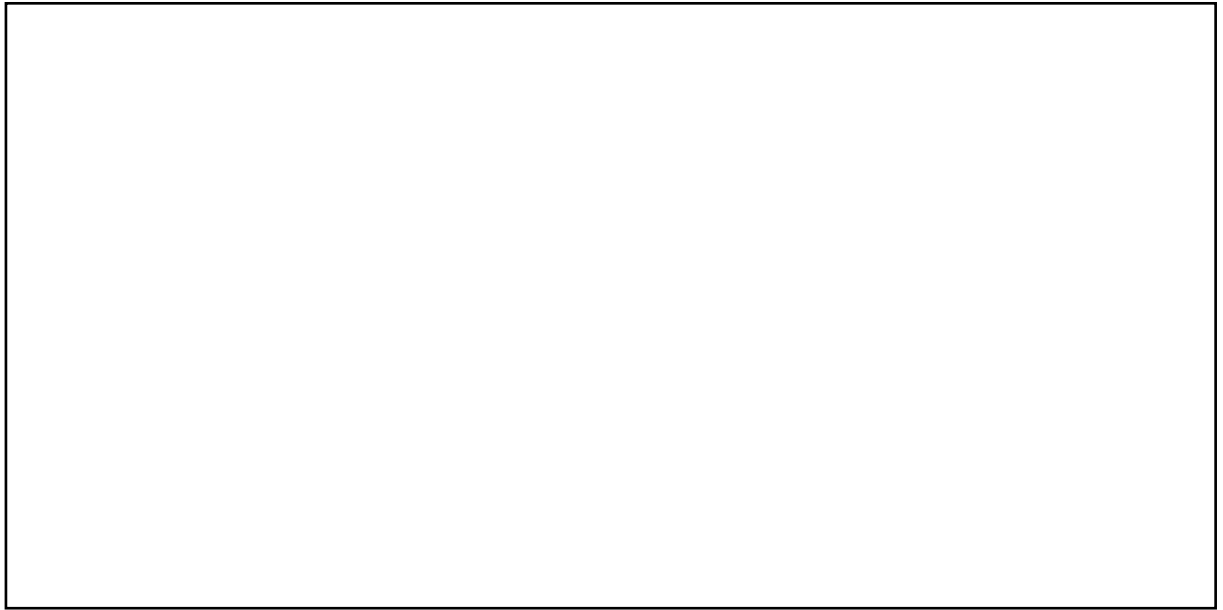
- Er der stor forskel på, hvor mange der er ramt i byerne?
- Konstruer et passende diagram til datasættet.
- Overvej, hvad I skal bruge for at kunne gøre det.
- Hvad fortæller kvartilsættet om jeres datasæt?
- Ved hvilket kvartilsæt vil I anbefale organisationen at indsætte hjælp?

Prøv så vidt muligt at tænke kørsel, økonomi og ressourcer ind i jeres overvejelser, selv om I ikke har specifikke data til det.

Bilag

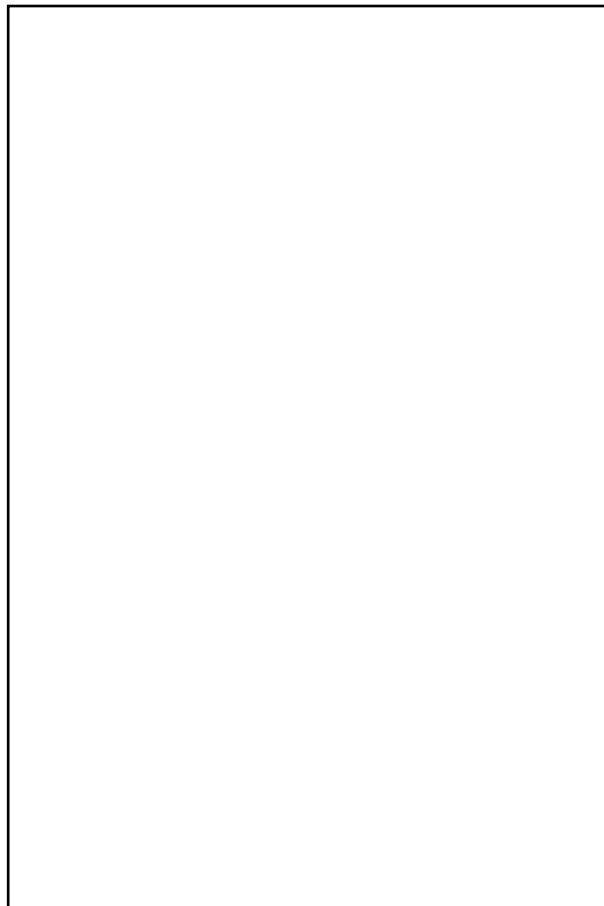
Længde: 16 cm

Bredde: 8 cm



Bredde: 8 cm

Højde: 12 cm



Referencer

- 1) Bl.a. International Røde Kors: <https://www.ifrc.org/en/what-we-do/disaster-management/about-disasters/what-is-a-disaster/>
 - 2) <https://da.wikipedia.org/wiki/Naturkatastrofe>, <https://faktalink.dk/titelliste/kata>
 - 3) <https://videnskab.dk/naturvidenskab/mennesket-udleder-op-til-100-gange-mere-co2-end-vulkaner>
 - 4) <https://videnskab.dk/miljo-naturvidenskab/jordskaelv-og-vulkaner-i-danmark>
 - 5) <https://videnskab.dk/sporg-videnskaben/tyfoner-cykloner-orkaner-hvad-er-forskellen>
 - 6) <https://e-hvordan.dk/hvordan-opstaar-en-tsunami/>
 - 7) https://da.wikipedia.org/wiki/Decemberorkanen_1999
 - 8) Hoff Kjeldsen, Tinne (2011): Hvad er matematik? Akademisk forlag, side 146
 - 9) <https://www.tv2lorry.dk/frederikssund/ny-prognose-stormflod-truer-ved-roskilde-fjord>
 - 10) <https://www.berlingske.dk/samfund/urds-vandmasser-har-toppet-i-roskilde-fjord>
 - 11) <https://sn.dk/Frederikssund/Et-vaagent-oeje-med-vandstand-i-fjordene/artikel/912967?fbclid=IwAR2vJjhd0fD2EbmY7ExCCTGwUSnWnMsmbYdpsUi46Li2lmnRQrV03zRTFlw>
 - 12) https://www.sst.dk/da/udgivelser/2015/~/_media/A72D478EDC6F4298ACEE1E9AE545BF83.ashx
 - 13) <https://reliefweb.int/sites/reliefweb.int/files/resources/FINAL%20Malawi%20Flood%20Response%20Plan%20March%20-%20May%202019%20-28March2019.pdf>
- *) Disse data tager udgangspunkt i et virkeligt scenarie og er realistiske, men konstruerede til undervisningsbrug.