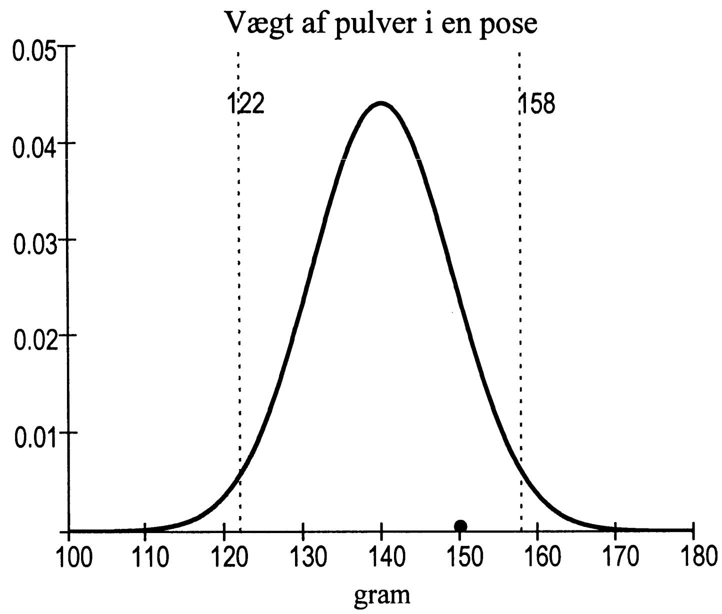
Statistisk beskrivelse og test



2005 Karsten Juul

Kapitel 1. Intervalhyppigheder

Afsnit 1.1: Histogram

En elevgruppe på et gymnasium har spurgt 100 tilfældigt valgte elever på gymnasiet om hvor lang tid det tager dem at komme fra deres hjem til skolen. Hver af de 100 elever fik en seddel hvor de skulle skrive hvor mange minutter det tager dem at komme i skole. Her er de 100 tal eleverne skrev:

10	37	37	8	12	6	6	39	28	13
8	3	10	8	6	11	12	10	10	12
8	5	7	6	11	19	15	13	6	9
6	26	22	5	14	18	5	17	34	16
12	10	6	13	7	13	26	7	29	2
19	11	30	13	23	16	10	34	7	6
23	10	9	47	3	36	7	13	6	6
18	25	7	15	7	2	22	8	5	40
5	17	14	9	6	29	14	5	17	25
24	4	18	14	14	10	22	3	12	11

Hvordan kan man få overblik over disse tal?

For at få overblik over disse tal talte elevgruppen op hvor mange af de spurgte elever der der havde en transporttid i intervallet 0-5, hvor mange der havde en transporttid i intervallet 5-10, osv.

Skal 5 tælles med i intervallet 0-5 eller i intervallet 5-10?

Eleverne går ud fra at halvdelen af de elever der har skrevet 5, har en transporttid mellem $4\frac{1}{2}$ og 5, og at den anden halvdel har en transporttid mellem 5 og $5\frac{1}{2}$. Derfor talte de 5 med som $\frac{1}{2}$ i hvert af intervallerne 0-5 og 5-10.

Resultatet af optællingen blev følgende:

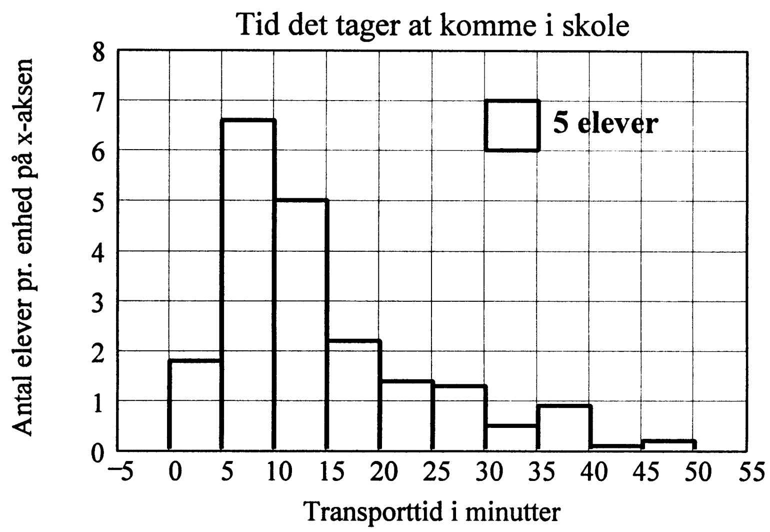
"Tid i min"	Antal elever"
"0 - 5"	9.0
"5 - 10"	33.0
"10 - 15"	25.0
"15 - 20"	11.0
"20 - 25"	7.0
"25 - 30"	6.5
"30 - 35"	2.5
"35 - 40"	4.5
"40 - 45"	0.5
"45 - 50"	1.0

Tabellen er taget direkte fra det matematikprogram der har beregnet den. Normalt skriver man ikke anførselstegn om intervaller og overskrifter.

Hvordan kan tabellens tal anskueliggøres på en figur?

For at anskueliggøre tallene der står i tabellen, tegnede eleverne en figur på følgende måde:

> Over hvert interval tegnede de et rektangel hvis **areal** var lig antallet af elever hvis transporttid lå i intervallet.

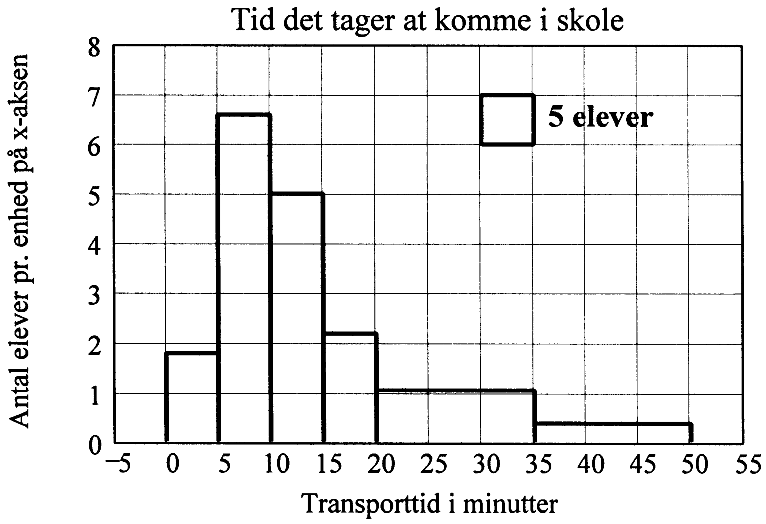


Beregning af rektanglernes højder

Arealet af rektanglet over intervallet 10-15 skal altså have arealet 25, og rektanglets bredde er $15-10=5$. Derfor gælder om højden h at $5 \cdot h = 25$, så $h = \frac{25}{5} = 5$. Højderne af de andre rektangler er udregnet på tilsvarende måde.

Forenklet figur

I intervallerne til højre for 20 er der ikke ret mange elever, så det er nok ret tilfældigt om det er det ene eller det andet rektangel der er størst. Derfor tegnede eleverne en ny figur hvor de seks intervaller blev erstattet af intervallerne 20-35 og 35-50:



1.1.1 Nogle fagudtryk

Ved hjælp af fagudtryk kan undersøgelsen på de første sider beskrives sådan:

- De minuttal eleverne har givet som svar, udgør observationssættet.
- Observationssættets størrelse er 100.
- Tabellen angiver hyppighedsfordelingen for det grupperede observationssæt.
- Observationsintervallerne er 0-5, 5-10, osv.
- De tilsvarende intervalhyppigheder er 9, 33 osv.
- På figuren er hyppighedsfordelingen anskueliggjort ved hjælp af et histogram.

Opgave 1. Observationssæt

Angiv om det er a), b), c) eller d) der er det korrekte svar på spørgsmålet "Hvad er observationssættet i eksemplet med transporttider?" .

a) Alle skolens elever.

b) De tilfældigt valgte elever som blev spurgt om deres transporttid.

c) 100.

d) De transporttider eleverne oplyste.

Opgave 2. Histogram

Nogle elever slår intervallerne 0-5 og 5-10 sammen til intervallet 0-10 og vil tegne et histogram svarende til denne nye gruppering. Hvor højt et rektangel skal de tegne over intervallet 0-10?" .

Opgave 3. Histogram

På en anden skole spurgte man også 100 elever om deres transporttid i minutter. Man fik observationssættet nedenfor.

a) Gruppér observationerne og tegn et histogram.

b) Hvordan kan gymnasiet og boligerne tænkes at være placeret når histogrammet ser sådan ud?

22	16	22	18	33	18	33	15	20	21
34	22	31	19	22	21	24	31	21	11
26	23	19	11	18	33	32	31	22	18
32	14	30	18	34	20	34	20	20	19
19	33	33	18	19	32	19	11	18	20
15	19	18	33	19	30	33	33	18	23
34	24	19	20	32	17	25	19	30	26
21	21	25	34	21	19	23	28	16	19
33	32	21	33	21	20	32	32	19	31
21	18	19	31	31	18	14	20	18	17

Opgave 4. Histogram

Et gymnasiums elever bor jævnt fordelt i et stort byområde, og gymnasiet ligger midt i området.

Tegn et eksempel på hvordan transporttids-histogrammet kan tænkes at se ud, og skriv en begrundelse for udseendet.

Opgave 5. Histogram

Et gymnasium ligger ved en landevej, og gymnasiets elever bor jævnt fordelt langs landevejen.

Tegn et eksempel på hvordan transporttids-histogrammet kan tænkes at se ud, og skriv en begrundelse for udseendet.

1.1.2 Intervalhyppigheder og histogram

En række målte tal

$$x_1,\ x_2,\ x_3,\ \dots,\ x_n$$

fx længder, vægte eller tidsrum, får man ofte bedre overblik over ved at skrive nogle intervaller

$$a_0 - a_1,\quad a_1 - a_2,\quad a_2 - a_3,\quad \dots\quad a_{m-1} - a_m$$

og for hvert interval skrive hvor mange af de målte tal der ligger i intervallet. Dette antal, altså antallet af målte tal i intervallet, kaldes intervalhyppigheden.

Et målt tal der er lig et intervalendepunkt, tæller med som en halv i hvert af intervallerne. Begrundelsen er at når et måleresultat fx er 8.3, vil den målte størrelse i halvdelen af tilfældene ligge i intervallet 8.25-8.30, og i halvdelen af tilfældene i intervallet 8.30-8.35.

Intervalhyppighederne kan anskueliggøres ved at tegne et histogram. Dette tegnes sådan: På x-aksen afsættes intervalendepunkterne, og over hvert interval tegnes et rektangel hvis areal er lig antallet af målte tal i intervallet.

Da rektanglets areal er lig antallet af målte tal i intervallet, kan rektanglets højde beregnes ved at dividere antallet af målte tal i intervallet med intervallets længde. (Længden af et interval er højre endepunkt minus venstre endepunkt).

Afsnit 1.2: Middelværdi

Der findes mange slags gennemsnit. Den slags gennemsnit der bruges når en elevs karaktergennemsnit udregnes, kaldes middelværdi.

Middelværdien af nogle tal fås ved at lægge tallene sammen og dividere resultatet med antallet af tal.

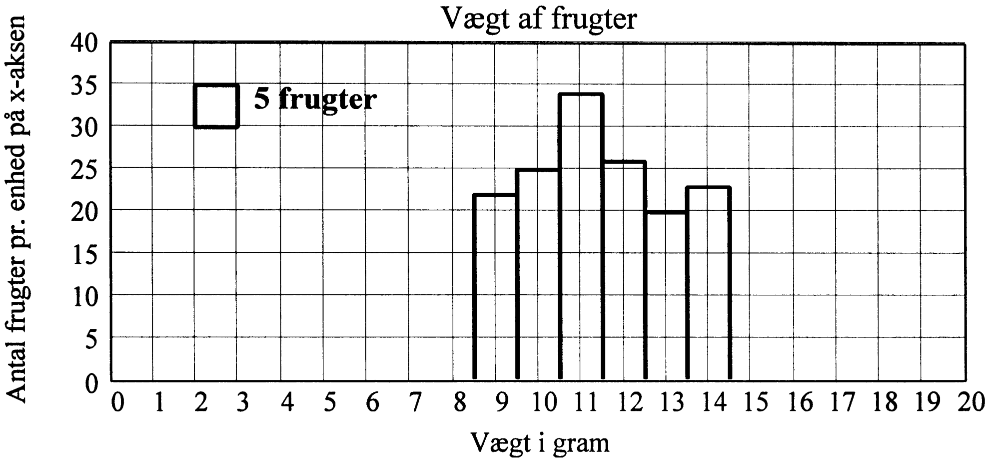
Opgave 6. Udregning af middelværdi af nogle tal

"Hvor mange skal der være i en gruppe når I har gruppearbejde?" De 27 elever i en klasse har besvaret dette spørgsmål. Ni foretrak en gruppestørrelse på to, tolv foretrak tre og seks foretrak fire.

Udregn middeltallet af de 27 svar.

Opgave 7. Oplæg til udregning af middelværdi ud fra intervalhyppigheder

Hver af frugterne i en kasse er blevet vejet. Resultatet fremgår af histogrammet.



Når vi skal udregne middelværdien af frugterne ud fra histogrammet, går vi ud fra observationerne mellem 9.5 og 10.5 er jævnt fordelt, så vi kan lade som om de alle har vægten 10.0: Hvis vi lægger tallene 9.7, 10.0 og 10.3 sammen, så får vi det samme som hvis vi lægger tallene 10.0, 10.0 og 10.0 sammen. For de andre intervaller gælder det tilsvarende.

Bestem middelværdien af frugternes vægte.

Hvis vægtene afrundes til et helt antal gram, gælder: Mange frugter vejer 9 gram, men ingen vejer 8 gram. Hvad kan tænkes at være grunden til dette?

1.2.1 Sådan udregnes middelværdien ud fra intervalhyppighederne

Når man udregner middelværdien for et grupperet observationssæt, så regner man som om det for hvert observationsinterval gjaldt at alle observationerne i intervallet lå midt i intervallet.

For det grupperede observationssæt der er givet ved histogrammet på side 3 (se også tabellen side 2), udregner man altså middelværdien sådan:

$$\frac{2.5\cdot 9+7.5\cdot 33+12.5\cdot 25+17.5\cdot 11+27.5\cdot 16+42.5\cdot 6}{100}=14.7$$

Middelværdien er $\underline{\underline{14{,}7\,\text{gram}}}$.

Opgave 8. Udregning af middelværdi ud fra intervalhyppigheder

"Hvor lang tid bruger du på et sæt hjemmeopgaver i matematik?" Dette spørgsmål blev besvaret af 98 elever fra 2.g . Svarene fremgår af tabellen.

Timer pr. sæt	0.5-1	1-2	2-3	3-5
Antal elever	4	18	35	41

Udregn middelværdien for dette observationssæt.

Afsnit 1.3: Kumuleret hyppighed

På side 2 er hyppighedsfordelingen for elevernes transporttider vist både ved hjælp af en tabel og ved hjælp af et histogram.

Det ses at 42 elever har en transporttid på 10.0 minutter eller derunder. Man siger at den kumulerede hyppighed af 10.0 er 42.

1.3.1 Kumuleret hyppighed

Den kumulerede hyppighed af et tal er antallet af observationer der er mindre end eller lig tallet.

Opgave 9. Uddybning vedr. kumuleret hyppighed

Følgende spørgsmål drejer sig om observationssættet der er beskrevet på side 2.

Hvor mange elever har en transporttid på 25.0 minutter eller derunder, og hvor mange har en transporttid på 60.0 minutter eller derunder?

Hvad er den kumulerede hyppighed af 25.0, og hvad er den kumulerede hyppighed af 60.0?

Bestem den kumulerede hyppighed af hvert af tallene 45 og -10.

Opgave 10. Oplæg til kumuleret hyppighed

Følgende spørgsmål drejer sig om observationssættet der er beskrevet i opgave 7.

Bestem den kumulerede hyppighed af 10.5 .

Bestem den kumulerede hyppighed af 11.0 .

1.3.2 Kumuleret hyppighed af et tal der ikke er et intervalendepunkt

Når man beregner kumulerede hyppigheder for grupperede observationssæt, så går man ud fra at observationerne ligger helt jævnt fordelt i ethvert af observationsintervallerne.

Hvis et observationsinterval går fra 30 til 40, så regner man altså som om en hundrededel af observationerne ligger i intervallet fra 33.2 til 33.3 da dette intervals længde er en hundrededel af hele intervallets længde. Ligger der 17 observationer mellem 30 og 40, så regner man altså som om der ligger 17 hundrededele af en observation (altså 0.17 observation) mellem 33.2 og 33.3 .

Opgave 11. Kumuleret hyppighed af et tal der ikke er et intervalendepunkt

For et observationssæt gælder

36 af observationerne har størrelse 8 eller derunder, og

12 af observationerne ligger i intervallet 8-10.

Hvad er den kumulerede hyppighed af 8,5?

Hvad er den kumulerede hyppighed af 8,2?

1.3.3 Kumuleret hyppighedsfordeling

Lad H betegne funktionen bestemt ved at

$$H(x) = \text{kumuleret hyppighed af } x \ , \qquad x \in \mathbb{R} .$$

Denne funktion kaldes den kumulerede hyppighedsfordeling.

Opgave 12. Tegne graf for kumuleret hyppighedsfordeling

Et grupperet observationssæt er givet ved at observationsintervallerne er 0-6, 6-10 og 10-16, og at disses intervalhyppigheder er hhv. 300, 400 og 300. Lad H betegne den kumulerede hyppighedsfordeling

$H(x) =$ kumuleret hyppighed af x , $x \in R$.

Udfyld tabellen og tegn grafen for H.

x	-4	-2	0	2	4	6	8	10	12	14	16	18	20
H(x)													

Opgave 13. Bruge graf for kumuleret hyppighedsfordeling

Vægtfordelingen for en sending æbler fremgår af figuren nedenfor.

a) Hvor mange æbler vejer under 160 gram?

b) Hvor mange æbler vejer under 140 gram?

c) Hvor mange æbler vejer mellem 140 og 160 gram?

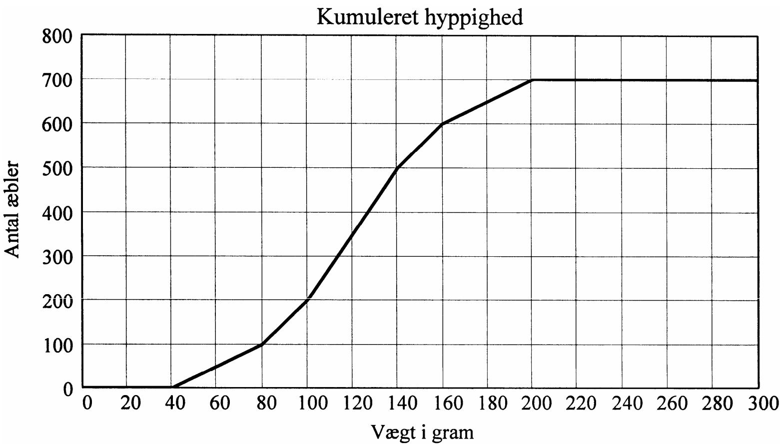
d) Hvor mange æbler er der?

e) Hvor mange æbler vejer over 160 gram?

f) Lav en tabel der viser intervalhyppighederne for intervallerne

40-80, 80-100, 100-120, 120-140, 140-160, 160-200.

g) Bestem middelværdien af æblernes vægte.

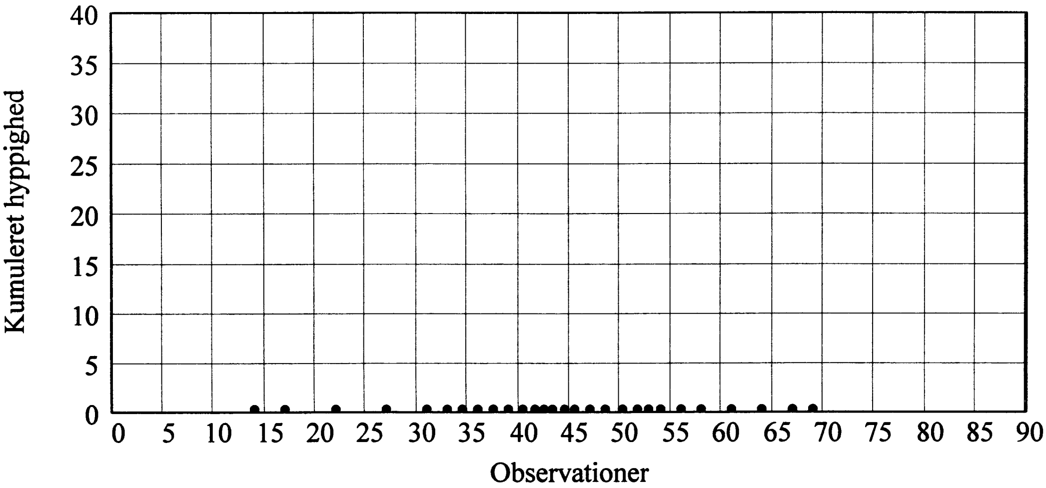


Opgave 14. Indsigt vedr. graf for kumuleret hyppighedsfordeling

På figuren er et observationssæt angivet på førsteaksen. På figuren på næste side skal du tegne grafen for den kumulerede hyppighed svarende til en gruppering i intervallerne

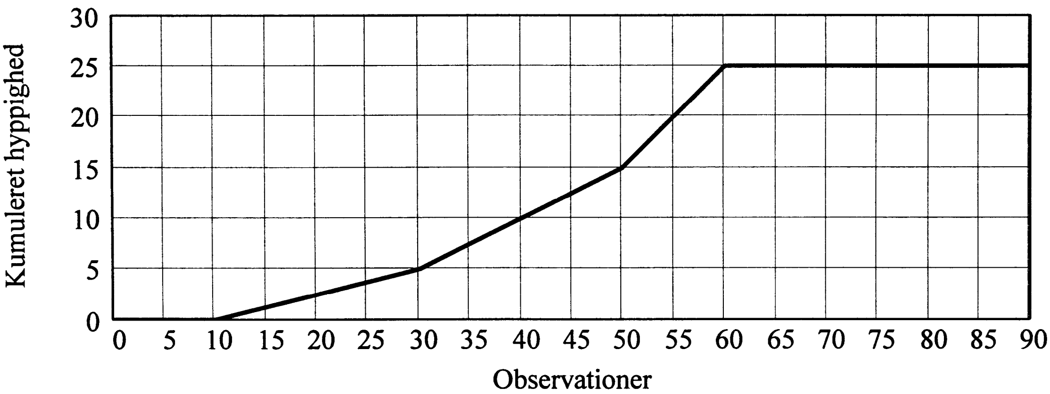
10-30, 30-40, 40-45, 45-55, 55-70.

Hvordan ser man på grafen hvor observationerne ligger tættest?



Opgave 15. Indsigt vedr. graf for kumuleret hyppighedsfordeling

I foregående opgave var tallene i et observationssæt angivet med prikker på førsteaksen, og du skulle tegne grafen for den kumulerede hyppighedsfordeling. I denne opgave er det omvendt: Her er grafen for den kumulerede hyppighedsfordeling givet, og du skal på førsteaksen afsætte prikker som angiver tallene i et observationssæt som har den viste hyppighedsfordeling. Der er mange observationssæt som har den viste hyppighedsfordeling, men du skal kun angive et af dem.



1.3.4 Et sprogproblem

Hvis nogle frugter måles og resultatet angives som et helt antal cm, kan sætningen

26 af frugterne er under 8 cm

både betyde

(1) 26 af frugterne er under 8.00··· cm (fx 7.98··· cm)

og

(2) 26 af frugterne er målt til under 8 cm (dvs. er under 7.5 cm).

I disse noter er betydningen (1).

1.3.5 Er 8 med blandt de der er "under 8"?

Når betydningen (1) i 1.3.4 bruges, har det ingen betydning for vores udregninger om 8 regnes med eller ej. Vi regner jo som om observationerne var helt jævnt fordelt i et observationsinterval, og der er uendelig mange tal i intervallet, så i vores udregninger er 0% af frugterne 8.00··· cm.

Opgave 16. Fordelingsfunktion for grupperede observationer

Betragt det grupperede observationssæt som er givet ved grafen i opgave 15.

Hvor mange observationer x opfylder at $x \leqq 50$?

Hvor mange observationer x opfylder at $x < 50$?

Kapitel 2. Intervalfrekvenser

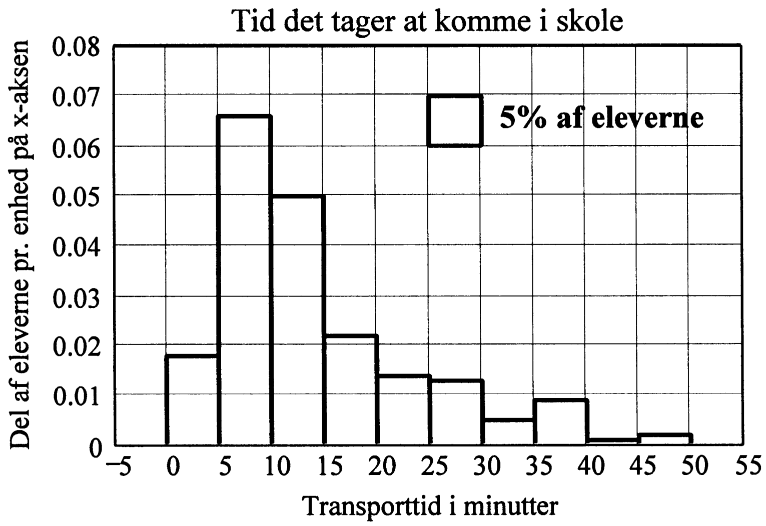
Afsnit 2.1. Histogram der viser intervalfrekvenser

Af tabellen på side 2 ses at 33 af de spurgte elever har en transporttid der ligger i intervallet 5-10 minutter. Da man har spurgt 100 elever, er det altså 33% af de spurgte elever der har en transporttid i intervallet 5-10 minutter. Dette udtrykkes ved at sige at intervallet 5-10 har intervalfrekvensen 33% (eller 0.33).

Tabellen nedenfor angiver frekvensfordelingen for det grupperede observationssæt.

"Tid i min"	Del af eleverne"
"0 - 5"	0.090
"5 - 10"	0.330
"10 - 15"	0.250
"15 - 20"	0.110
"20 - 25"	0.070
"25 - 30"	0.065
"30 - 35"	0.025
"35 - 40"	0.045
"40 - 45"	0.005
"45 - 50"	0.010

Histogrammet nedenfor anskueliggør frekvensfordelingen.



Opgave 17 Intervalfrekvenser

Antag at man havde spurgt 352 elever, og at tabellen og histogrammet ovenfor viser intervalfrekvenserne for disse elevers transporttider. Hvor stort et antal elever havde så en transporttid mellem 5.0 og 15.0 minutter?

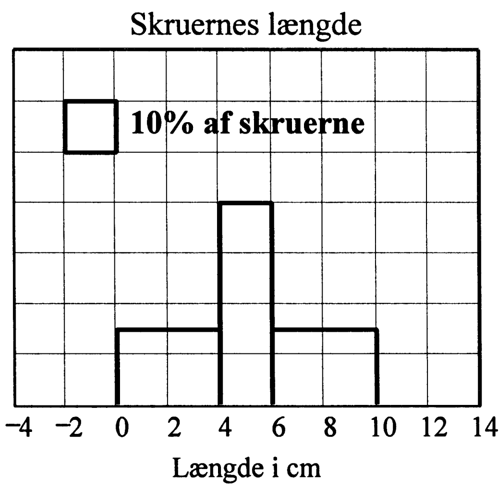
2.1.1 Intervalfrekvens

Intervalfrekvensen for et interval fås ved at dividere antallet af observationer der ligger i intervallet, med det samlede antal observationer. Intervalfrekvensen kan skrives som en brøk eller som et decimaltal eller som en procent.

Afsnit 2.2. Udregning af middelværdi ud fra intervalfrekvenser

Opgave 18 Oplæg til udregning af middelværdi ud fra intervalfrekvenser

Uden på en tom pose som har indeholdt forskellige skruer, er der et histogram som det der er tegnet nedenfor.



a) Hvis observationssættets størrelse er 50, hvor mange skruer er der så i hvert af intervallerne 0-4, 4-6 og 6-10. Og hvad er middelværdien af skruerne?

b) Hvis observationssættets størrelse er 100, hvor mange skruer er der så i hvert af intervallerne 0-4, 4-6 og 6-10. Og hvad er middelværdien af skruerne?

c) Uden at foretage udregninger skal du begrunde at følgende to lighedstegn gælder:

$$\frac{2\cdot 15+5\cdot 20+8\cdot 15}{50}=\frac{2\cdot 15}{50}+\frac{5\cdot 20}{50}+\frac{8\cdot 15}{50}=2\cdot\frac{15}{50}+5\cdot\frac{20}{50}+8\cdot\frac{15}{50}$$

d) Tallene 15, 20 og 15 kaldes intervalhyppigheder. Hvad kaldes tallene $\frac{15}{50}$, $\frac{20}{50}$ og $\frac{15}{50}$?

e) Hvis de tre intervallers intervalfrekvenser i stedet var hhv. 0.2, 0.5 og 0.3, hvad var så middelværdien af skruernes længde?

2.2.1 Beregning af middelværdi ud fra frekvensfordeling

Middelværdien af et grupperet observationssæt kan beregnes ud fra frekvensfordelingen ved hjælp af følgende metode:

1. For hvert interval ganges tallet i intervallets midte med intervallets frekvens.
2. Disse resultater lægges sammen.

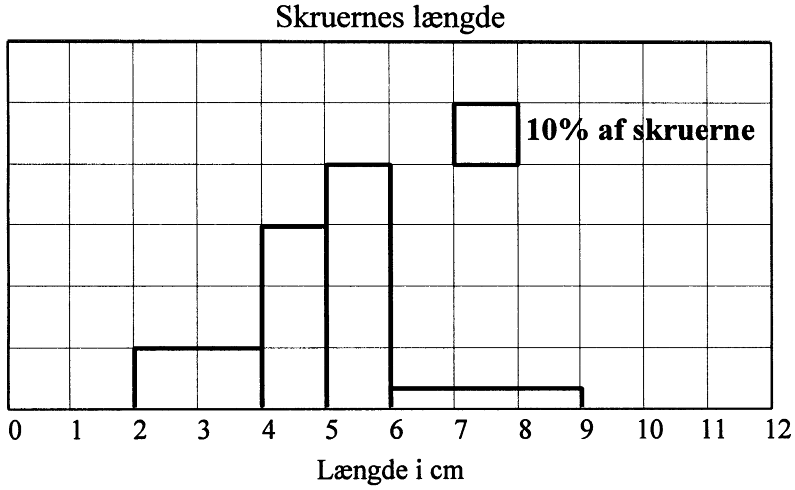
Det tal man får, er middelværdien.

2.2.2 Bemærkning om beregning af middelværdi ud fra intervalfrekvenser

Når man udregner middelværdien ud fra intervalfrekvenserne, så skal man altså ikke dividere summen med noget. Fra opgave 18 ved vi at dette skyldes at de intervalfrekvenser man ganger med, er fremkommet ved at dividere intervalhyppighederne med observationssættets størrelse. Der er altså blevet divideret før der blev lagt sammen.

Opgave 19 Beregning af middelværdi ud fra histogram

Uden på en tom pose som har indeholdt forskellige skruer, er der et histogram som det der er tegnet nedenfor. Udregn middelværdien af skruernes længder.



Afsnit 2.3 . Kumuleret frekvens

På side 11 er frekvensfordelingen for elevernes transporttider vist både ved hjælp af en tabel og ved hjælp af et histogram.

Det ses at 67% af eleverne har en transporttid under 15.0 minutter. Man siger at den kumulerede frekvens af 15.0 er 67%.

2.3.1 Kumuleret frekvens

Ved den kumulerede frekvens af et tal forstås den procentdel (eller brøkdel) af observationerne der er mindre end eller lig tallet.

Den kumulerede frekvens kan skrives som en brøk eller som et decimaltal eller som en procent.

Opgave 20 Uddybning vedr. kumuleret frekvens

Følgende spørgsmål drejer sig om observationssættet der er beskrevet på side 11.

Hvor mange procent af eleverne har en transporttid under 30.0 minutter, og hvor mange har en transporttid under 60.0 minutter?

Hvad er den kumulerede frekvens af 30.0?, og hvad er den kumulerede frekvens af 60.0?

Bestem den kumulerede frekvens af hvert af tallene 40 og -5.

2.3.2 Kumuleret frekvensfordeling

Lad F betegne funktionen bestemt ved at

$$F(x) = \text{kumuleret frekvens af } x \;, \qquad x \in \mathbb{R} \,.$$

Denne funktion kaldes den kumulerede frekvensfordeling.

Opgave 21 Tegne graf for kumuleret frekvensfordeling

Betragt det grupperede observationssæt som er givet ved histogrammet i opgave 19. Lad F være den kumulerede frekvensfordeling

$$F(x) = \text{kumuleret frekvens af } x \;, \qquad x \in \mathbb{R} \,.$$

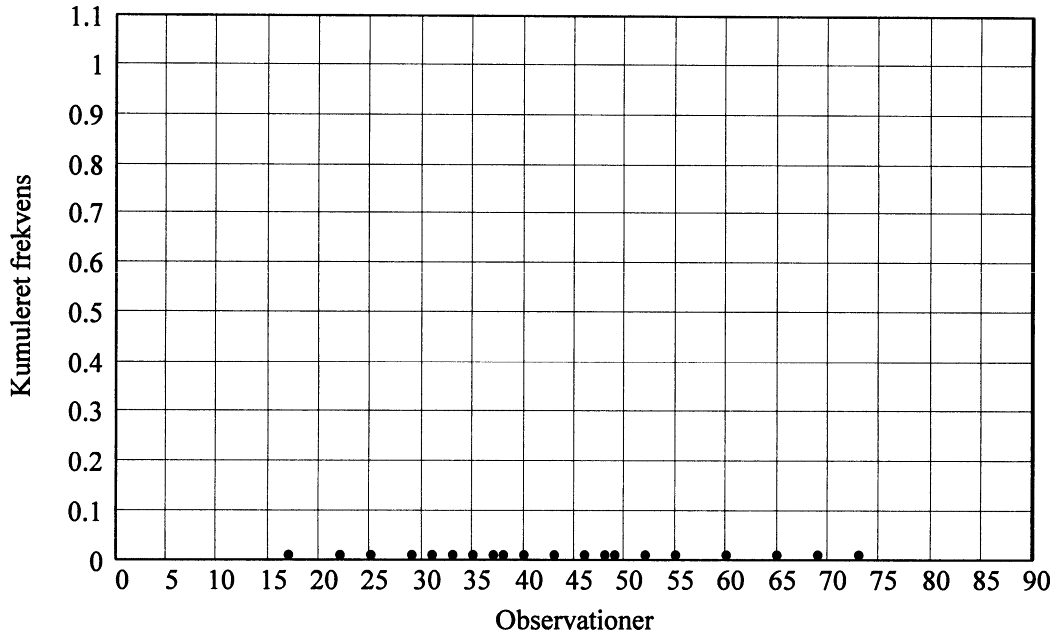
Udfyld tabellen og tegn grafen for F.

x	1	2	3	4	4.5	5	5.5	6	7	8	9	10
F(x)												

Opgave 22 Indsigt vedr. graf for kumuleret frekvensfordeling

På figuren på næste side er et observationssæt angivet ved at der på førsteaksen er sat en prik ved hvert tal som er en observation. På figuren skal du tegne grafen for den kumulerede frekvensfordeling svarende til en gruppering i intervallerne

15-30, 30-50, 50-75.



Afsnit 2.4. Formler vedr. middelværdi

Opgave 23 Træning i forståelse af symboludtryk

Tabellen viser frekvensfordelingen for et grupperet observationssæt, hvis **middelværdi er 5.2** .

Interval	2-4	4-5	5-7	7-10
Intervalfrekvens	f_1	f_2	f_3	f_4

a) Bestem tallet $f_1 + f_2 + f_3 + f_4$.

b) Bestem tallet $3 \cdot f_1 + 4.5 \cdot f_2 + 6 \cdot f_3 + 8.5 \cdot f_4$.

c) Lad F betegne den kumulerede frekvens. Afgør hvilken af følgende ligninger der er korrekt :

$$F(4) = f_1 + f_2 \;,\quad F(4.5) = f_1 + f_2 \;,\quad F(5) = f_1 + f_2 \;.$$

Opgave 24 Opstille formler vedr. middelværdi

Tabellen viser frekvensfordelingen for et grupperet observationssæt.

Interval	$]a_0;a_1]$	$]a_1;a_2]$	$]a_2;a_3]$
Intervalfrekvens	f_1	f_2	f_3

Lad x_1, x_2 og x_3 betegne midtpunkterne af de tre intervaller.

a) Skriv en ligning $x_1 = \cdots$ som viser hvordan x_1 kan beregnes ud fra tallene i tabellen. På ligningens højreside må ud over regnetegn og cifre kun stå nogle af talbetegnelserne a_0, a_1, a_2, a_3, f_1, f_2 og f_3.

b) Lad $\overline{x}$ betegne middelværdien af observationssættet. Skriv en ligning $\overline{x} = \cdots$, der viser hvordan middelværdien kan beregnes. På ligningens højreside må ud over regnetegn kun stå nogle af talbetegnelserne a_0, a_1, a_2, a_3, f_1, f_2, f_3, x_1, x_2 og x_3.

Opgave 25 Opstille formler vedr. middelværdi

Tabellen nedenfor viser hyppighedsfordelingen for et grupperet observationssæt hvis størrelse er n.

Interval	$]a_0;a_1]$	$]a_1;a_2]$	$]a_2;a_3]$
Intervalhyppighed	h_1	h_2	h_3

Lad x_1, x_2 og x_3 betegne midtpunkterne af de tre intervaller.

a) Skriv en ligning $\overline{x} = \cdots$ der viser hvordan middelværdien $\overline{x}$ kan beregnes. På ligningens højreside må ud over regnetegn kun stå nogle af talbetegnelserne n, a_0, a_1, a_2, a_3, h_1, h_2, h_3, x_1, x_2 og x_3.

b) Skriv en ligning $f_2 = \cdots$ der viser hvordan intervalfrekvensen f_2 for det andet interval kan beregnes. På ligningens højreside må ud over regnetegn kun stå nogle af talbetegnelserne n, a_0, a_1, a_2, a_3, h_1, h_2, h_3, x_1, x_2 og x_3.

Opgave 26 Bevis vedr. middelværdi

Vi forestiller os nu at de to foregående opgaver drejer sig om de samme observationer. Lad dig inspirere af din besvarelse af spørgsmål c) i opgave 18 og bevis at det udtryk du i opgave 25 fandt for middelværdien $\overline{x}$, er lig det udtryk du i opgave 24 fandt for middelværdien.

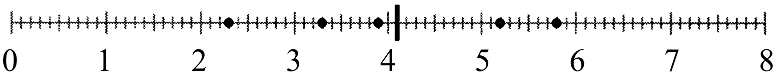
Skriv beviset her:

Kapitel 3. Varians og spredning

Opgave 27 Indfører begreberne varians og spredning

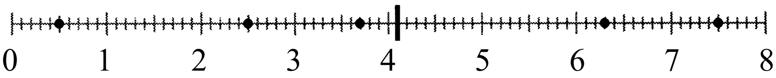
På den øverste tallinje er vist et observationssæt bestående at 5 tal.

a) Bestem middelværdien af disse tal.



På tallinje nr. 2 er vist et andet observationssæt, som også består af 5 tal.

b) Bestem middelværdien af disse tal.



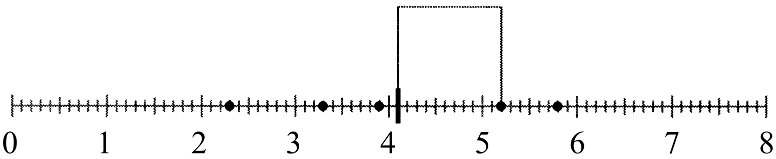
Tallene på tallinje nr. 2 ligger mere spredt end tallene på den øverste tallinje. Dette kan også udtrykkes ved at sige at de gennemsnitligt ligger længere væk fra middelværdien.

På figuren nedenfor er anskueliggjort

kvadratet på afstanden fra middelværdien til observationen 5.2 .

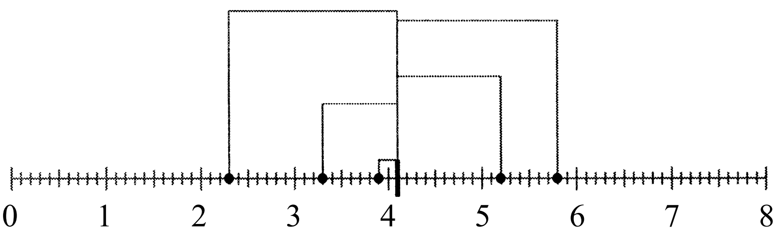
Arealet af det viste kvadrat bruges som et mål for hvor langt 5.2 ligger fra middelværdien.

c) Bestem kvadratet på afstanden fra middelværdien til observationen 5.2 .



På figuren nedenfor er for hvert af de 5 tal tegnet det tilhørende kvadrat.

d) Bestem middelværdien af kvadraterne på observationernes afstande fra middelværdien.



Middelværdien af kvadraterne på observationernes afstande til middelværdien kaldes variansen. Variansen er altså et mål for hvor spredt observationerne ligger.

e) Bestem variansen af de 5 tal på tallinje nr. 2.

Kvadratroden af variansen kaldes spredningen.

f) Bestem spredningen for observationssættet på den øverste tallinje, og for observationssættet på tallinje nr. 2.

Opgave 28 Varians og spredning for et grupperet observationssæt

Når man udregner varians og spredning for et grupperet observationssæt, så regner man som om alle observationer i et observationsinterval lå i intervallets midtpunkt.

Tabellen nedenfor viser hyppighedsfordelingen for et grupperet observationssæt. Det ses at der er 9 observationer. For at udregne variansen skal man altså udregne middelværdien af 9 kvadrater.

Bestem varians og spredning for det grupperede observationssæt.

Interval	3-5	5-6	6-8
Intervalhyppighed	2	4	3

Opgave 29 Spredning

Tegn to histogrammer der viser to frekvensfordelinger A og B, sådan at B har større spredning end A.

Opgave 30 Formel til beregning af varians

I denne opgave betegner V variansen for observationssættet i opgave 25.

Skriv en ligning der viser hvordan variansen kan beregnes.

$$V = \cdots$$

Opgave 31 Udlede regel for varians

Et observationssæt består af de n tal $x_1, x_2, ..., x_n$.

Følgende ligning viser hvordan middelværdien $\overline{x}$ kan beregnes:

$$\overline{x} = \frac{x_1 + x_2 + ... + x_n}{n} \ .$$

Skriv en ligning der viser hvordan variansen kan beregnes:

$$V_x =$$

Opskriv udtrykket for middelværdien af tallene $k \cdot x_1, \ k \cdot x_2, \ ..., \ k \cdot x_n$, hvor k er en konstant. Omskriv udtrykket til et udtryk på formen

En konstant gange udtrykket for middelværdien af tallene $x_1, x_2, ..., x_n$.

Skriv udregningerne her:

Opskriv udtrykket for variansen af tallene $k \cdot x_1$, $k \cdot x_2$, ..., $k \cdot x_n$, hvor k er en konstant. Omskriv udtrykket til et udtryk på formen

En konstant gange udtrykket for variansen af tallene $x_1, x_2, ..., x_n$.

Skriv udregningerne her:

Hermed er vist følgende regel om varians:

Sætning 1 Varians af tallene $k \cdot x_1$, $k \cdot x_2$, ..., $k \cdot x_n$

Når k er en konstant, gælder:

Variansen af tallene $k \cdot x_1$, $k \cdot x_2$, ..., $k \cdot x_n$ er ______ gange variansen af tallene $x_1, x_2, ..., x_n$.

Opgave 32 Oplæg vedr. spredning

a) To terninger A og B blev kastet. Forskellen på øjetallene blev 5.

Hvilke muligheder er der for hvad A kan have vist?

Hvilke muligheder er der for hvad A og B kan have vist? Skriv mulighederne som talpar (,) hvor A's øjetal står først.

b) Hvis forskellen på øjetallene havde været 4, hvilke muligheder er der så for hvad A og B kunne have vist?

c) Terningerne kastes ekstremt mange gange. Så kan man gå ud fra at alle muligheder (,) forekommer ca. lige tit.

Hvis det en million gange forekommer at forskellen på øjetallene er 5, hvor mange gange vil forskellen så være 4?

d) Hvilke tal kan forekomme som forskel på øjetallene?

e) Bestem frekvensen for hver af disse.

f) Bestem middelværdi og spredning.

g) Udfør nogle gange eksperimentet der består i at kaste to terninger og notere forskellen på øjetallene.

h) Bestem middelværdien for din stikprøve.

i) Bestem ud fra din stikprøve spredningen på to måder:
   - Ved at bruge den rigtige middelværdi fra spørgsmål f).
   - Ved at bruge middelværdien fra spørgsmål h).

3.1 Beregning af middelværdi og spredning ud fra stikprøve

Tallene der skal undersøges
Der er fremstillet et stort antal dimser. De har ikke præcis samme længde, men det vides at middelværdien af deres længder er 14.0 cm.

Varians bestemt ud fra alle tallene
For at bestemme variansen af længderne skal man for hver dims måle længden x og udregne kvadratet $(x-14.0)^2$. Variansen er middelværdien af alle disse kvadrater.

Varians bestemt ud fra stikprøve når middelværdi er kendt
I stedet for at måle længden af alle dimser nøjes man normalt med tilfældigt at udtage en del af dimserne og udregne middelværdien af de tilsvarende kvadrater $(x-14.0)^2$. Dette er et rimeligt gæt på variansen af alle dimsernes længder.

Varians bestemt ud fra stikprøve når middelværdi ikke er kendt
Hvis man ikke ved at middelværdien er 14.0, så må man i $(x-14.0)^2$ erstatte 14.0 med et gæt på middelværdien. Middelværdien $\overline{x}$ af de udtagne dimsers længder er et rimeligt gæt på middelværdien af alle dimsernes længder.

For en stikprøve er middelværdien af kvadraterne $(x-\overline{x})^2$ dog altid mindre end middelværien af kvadraterne $(x-14.0)^2$. Dette skyldes at tallene i stikprøven gennemsnitligt ligger tættere på deres egen middelværdi end på middelværdien af alle. (Se opgave 32 i).

Hvis man bruger middelværdien af stikprøvens kvadrater $(x-\overline{x})^2$ som gæt på variansen af alle tallene, så er ens gæt altså gennemsnitligt lidt for lille. Det er muligt at udregne et bedre gæt på følgende måde:

Man dividerer ikke summen af kvadraterne $(x-\overline{x})^2$ med deres antal n, men i stedet med $n-1$. Så får man et resultat der er lidt større, og dette resultat er det bedste gæt på variansen af alle tallene.

Sætning 2 Beregning af middelværdi og spredning ud fra stikprøve

En stikprøve bestående af n tal $x_1, x_2, ..., x_n$ er udtaget tilfældigt af en stor samling tal.

Ud fra tallene i stikprøven kan udregnes et gæt på middelværdien af tallene i den store samling. Det bedste gæt er det tal $\overline{x}$ som udregnes sådan:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

Og det bedste gæt på spredningen af tallene i den store samling er det tal s som udregnes sådan:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1}}.$$

Opgave 33 Beregning af middelværdi og spredning ud fra stikprøve

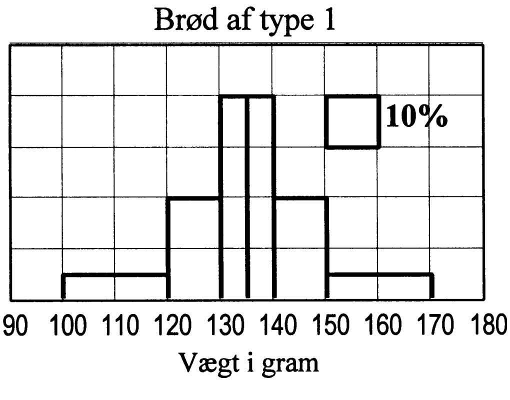
Ved en fest med 428 deltagere vil man bestemme middelværdi og spredning for deltagernes alkoholpromille. For 10 tilfældigt udtagne deltagere måles promillen. Resultaterne blev følgende:

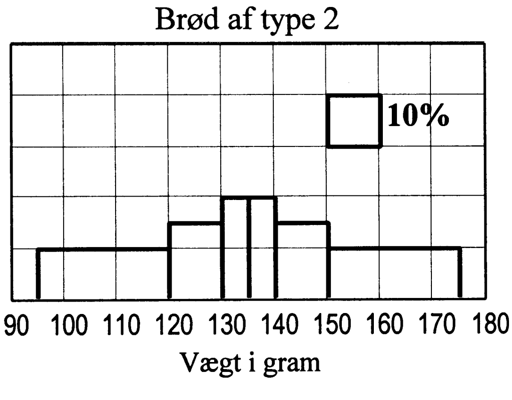
$$(2.9 \;\; 1.6 \;\; 1.3 \;\; 2.5 \;\; 3.1 \;\; 2.9 \;\; 1.7 \;\; 2.5 \;\; 2.5 \;\; 1.7)$$

Brug formlerne fra sætning 2 til at beregne det bedst mulige gæt på middelværdien af de 428 deltageres promiller og det bedst mulige gæt på spredningen af de 428 deltageres promiller.

Opgave 34 Øvelse som er af betydning for det følgende

En bager sælger to typer brød. Figurerne nedenfor viser frekvensfordelingerne for de to typer brøds vægt.





a) Hvorfor kan man her bestemme middelværdierne uden at regne?

b) Afgør uden at regne for hvilken af de to typer brød spredningen er størst?

c) Hvor mange procent af brødene af type 1 afviger mindre end 10 gram fra middelværdien?

En kunde der vil vide hvor store de to typer brød er, køber og vejer et af hver type.

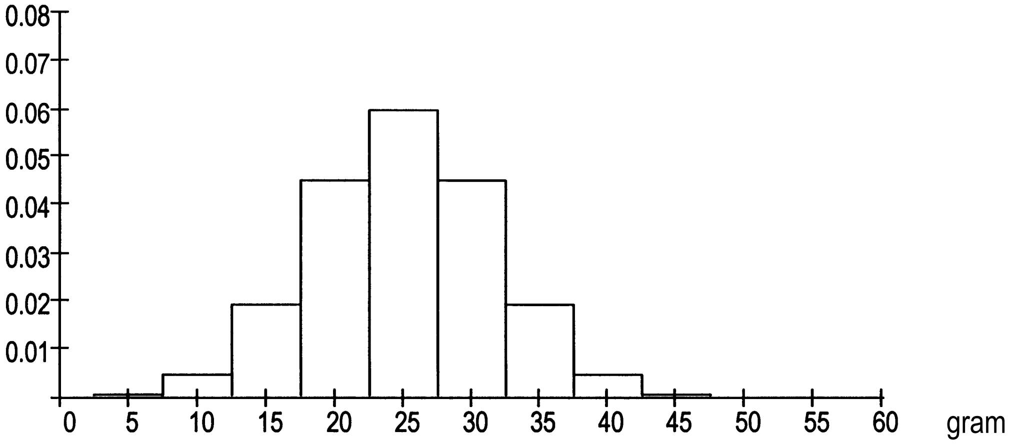
d) Er der over 50% chance for at det købte brød af type 1 har en vægt der afviger mindre end 10 gram fra middelværdien?

e) Er der over 50% chance for at det købte brød af type 2 har en vægt der afviger mindre end 10 gram fra middelværdien?

Kapitel 4. Tæthedsfunktion og fordelingsfunktion

Opgave 35 Forbereder indførelse af tæthedsfunktion

Figuren nedenfor viser frekvensfordelingen for vægten (i gram) af nogle frugter.



a) Hvad er arealet af rektanglet svarende til intervallet $]12.5;15.0]$?

b) Hvor mange procent af frugterne vejer ifølge figuren mellem 12.5 og 15.0 gram?

c) Hvor mange procent af frugterne vejer ifølge figuren mellem 15.0 og 17.5 gram?

d) Hvor mange procent af frugterne vejer ifølge figuren mellem 17.5 og 20.0 gram?

e) Hvor mange procent af frugterne vejer ifølge figuren mellem 20.0 og 22.5 gram?

f) Betragt de fire intervaller:

$(*)$ $\quad]12.5;15.0]$, $]15.0;17.5]$, $]17.5;20.0]]$, $]20.0;22.5]$.

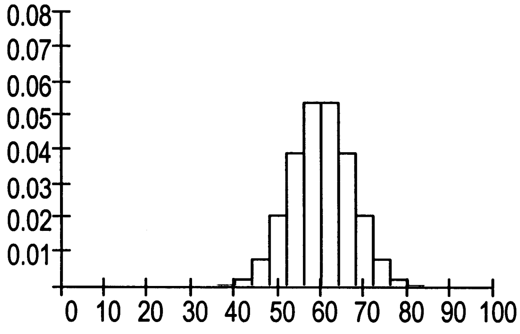
Ifølge figuren gælder: Når vi bevæger os fra venstre mod højre i denne række af intervaller, så er frekvensen konstant før og efter 17.5, mens der sker en kraftig stigning ved 17.5.

Dette skyldes den forenkling der er foretaget ved gruppering i de store intervaller $]12.5;17.5]$ og $]17.5,22.5]$. I virkeligheden stiger frekvensen nok også fra det første til det andet, og fra det tredje til det fjerde af intervallerne $(*)$.

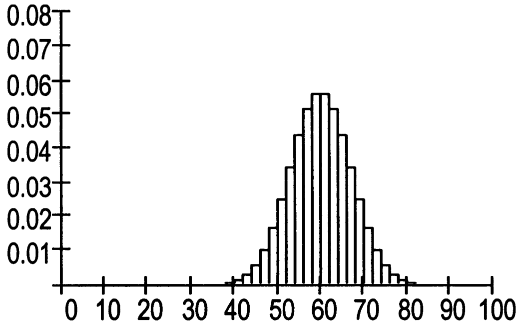
Over hvert af intervallerne $(*)$ skal du på figuren tegne et rektangel sådan at arealerne af disse rektangler kunne være mere i overensstemmelse med virkeligheden end svarene på spørgsmålene b) - e). Husk at de frekvenser for intervallerne $]12.5;17.5]$ og $]17.5,22.5]$ som histogrammet angiver, er i overensstemmelse med virkeligheden.

4.1 Hvad er en tæthedsfunktion

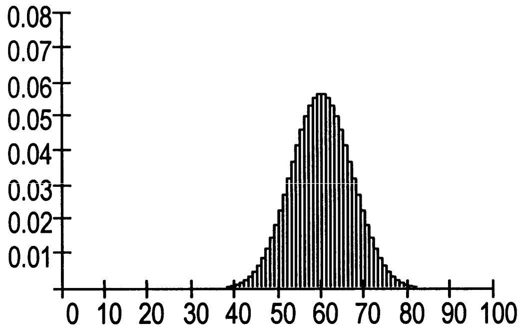
Der foreligger et meget stort antal observationer. Det første histogram nedenfor viser den frekvensfordeling der fås når man grupperer observationerne i intervaller af længde 4. De næste to histogrammer er fremkommet ved gruppering i intervaller af længde 2 og 1. Når intervallængden er tæt på 0, så er histogrammet tæt på den afrundede figur i det fjerde koordinatsystem. Kurven i dette koordinatsystem er grafen for tæthedsfunktionen.



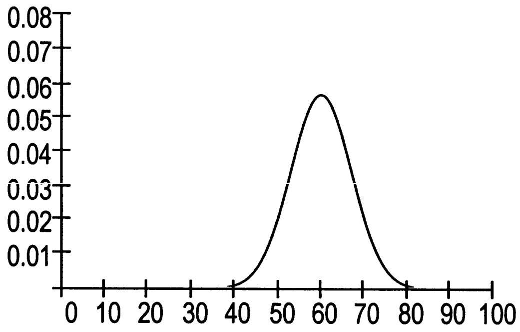
Intervallernes bredde er 4



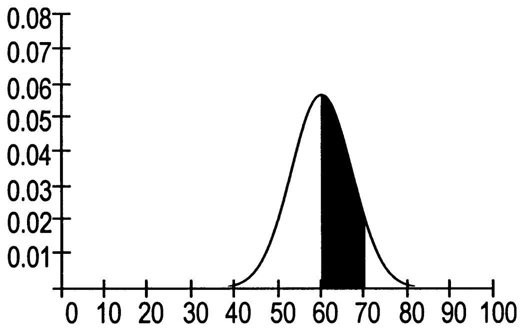
Intervallernes bredde er 2



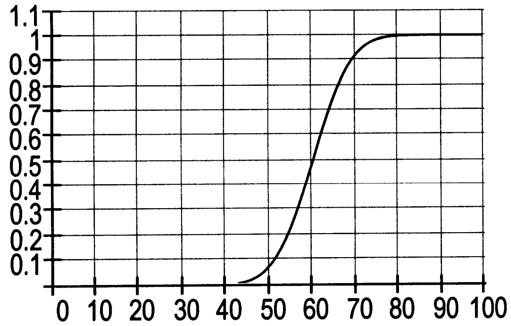
Intervallernes bredde er 1



Graf for tæthedsfunktionen



Graf for tæthedsfunktionen



Graf for fordelingsfunktionen

4.2 Brug af tæthedsfunktion

Tæthedsfunktionen angiver frekvensen af et interval på samme måde som et histogram gør, dvs. ved et areal. Frekvensen af intervallet 60-70 er altså arealet af det sorte område i det femte koordinatsystem ovenfor. Hvis arealet er 0.42, så indeholder intervallet 60-70 altså 42% af observationerne.

4.3 Hvad er en fordelingsfunktion

Hvis man tegnede graferne for de kumulerede frekvensfordelinger svarende til de tre histogrammer på foregående side, ville man se at de nærmede sig til kurven i det sidste koordinatsytem. Denne kurve er grafen for fordelingsfunktionen.

4.4 Brug af fordelingsfunktion

Fordelingsfunktionen F angiver frekvenser på samme måde som den kumulerede frekvensfordeling, dvs. fx er F(60) den brøkdel af observationerne som ligger til venstre for 60.0 på x-aksen. På grafen aflæses at $F(60) = 0.50$ og $F(70) = 0.92$. Da $0.92 - 0.50 = 0.42$, ligger 42% af observationerne i intervallet 50-60, og arealet af det sorte område i det femte koordinatsystem på foregående side er 0.42.

Opgave 36 Fordelingsfunktion

Fordelingsfunktionen F svarende til nogle observationer har forskriften

$$F(x) = \frac{1}{1 + 63 \cdot 1.2^{-x}}.$$

På figuren er vist grafen for den tilhørende tæthedsfunktion f.

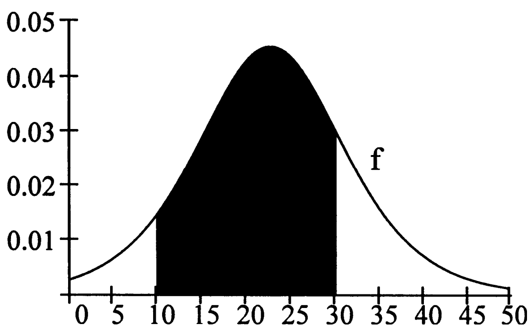
a) Man kan udregne $F(15)$ ved at indsætte 15 i forskriften ovenfor. Får man herved udregnet den procentdel af observationerne der er lig 15, den procentdel af observationerne der er mindre end 15, eller den procentdel af observationerne der er større end 15?

Brug fordelingsfunktionen F til at finde svarene på følgende tre spørgsmål:

b) Hvor mange procent af observationerne ligger mellem 10 og 30 på x-aksen?

c) Hvad er arealet af det sorte område.

d) Hvor mange procent af observationerne ligger til højre for 40 på x-aksen (ikke kun den viste del af x-aksen)?



Kapitel 5. Normalfordeling

5.1 Hvad er en normalfordeling?

På side 23 er i to af koordinatsystemerne vist en klokkeformet graf for en tæthedsfunktion. Grafer af nøjagtig denne form (bortset fra bredde, højde og placering) optræder i mange sammenhænge. Når tæthedsfunktionens graf har nøjagtig denne form, siges observationerne at være <u>normalfordelt</u>.

5.2 Hvor bruges normalfordelinger?

– Når samme størrelse måles flere gange, så fås en række tal der ikke er helt ens. Disse tal er typisk normalfordelt.

– Hvis man for nogle dyr eller planter af samme slags måler vægt, længde, omkreds, temperatur eller lignende, så fås en række tal der typisk er normalfordelt.

– Når en vare der skal sælges, hældes i poser eller flasker, vil der ikke komme præcis lige meget i hver. Måles indholdet, fås en række tal der typisk er normalfordelt.

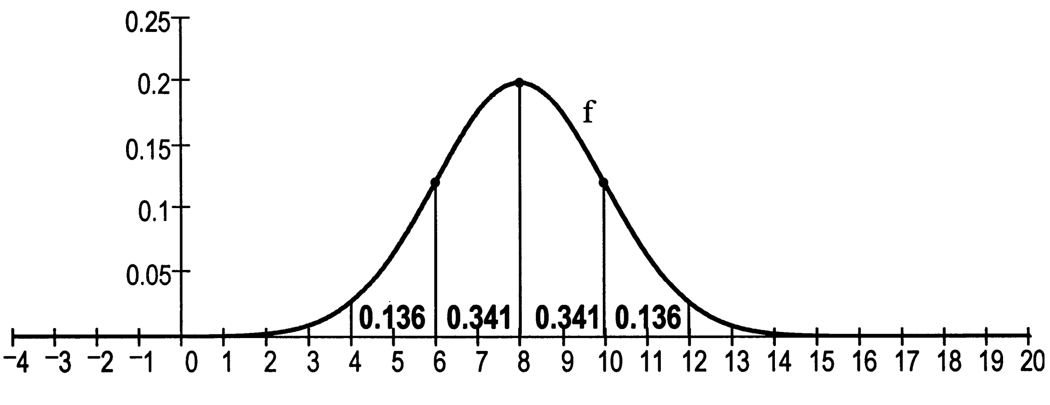
– I et utal af andre situationer der minder om en af disse tre, er tallene også normalfordelt.

5.3 Hvorfor kan normalfordelinger bruges i alle disse sammenhænge?

At normalfordelingen optræder i alle disse sammenhænge, skyldes at en sum af mange uafhængige, tilfældige størrelser altid vil være normalfordelt.

5.4 Middelværdi og spredning for normalfordeling

Som nævnt i det foregående er der mange forskellige tæthedsfunktioner for normalfordelinger. Man angiver hvilken af dem man taler om, ved at angive middelværdien μ og spredningen σ. På figuren nedenfor er vist tæthedsfunktionen f for en normalfordeling med $\mu = 8$ og $\sigma = 2$.



På grafen er tegnet tre prikker der markerer toppunktet og de to punkter hvor grafen skifter mellem at krumme opad og nedad. For alle normalfordelinger er førstekoordinaterne til disse punkter $\mu-\sigma$, μ og $\mu+\sigma$.

For alle normalfordelinger gælder desuden:

– Grafen er symmetrisk om den lodrette linje gennem toppunktet.
– Intervallerne $[\mu-\sigma;\mu]$ og $[\mu;\mu+\sigma]$ indeholder (angivet med én decimal) hver 34.1% af observationerne.
– Intervallerne $[\mu-2\sigma;\mu-\sigma]$ og $[\mu+\sigma;\mu+2\sigma]$ indeholder (angivet med én decimal) hver 13.6% af observationerne.

Opgave 37 Middelværdi og spredning for normalfordeling

Spørgsmålene i denne opgave vedrører den normalfordeling som er angivet på figuren ovenfor, og de skal besvares ved hjælp af de oplysninger om normalfordeling der står øverst på denne side.

a) Hvor mange procent af observationerne ligger til venstre for $\mu-2\sigma$ på x-aksen?

b) Hvor mange procent af observationerne ligger mellem 6 og 10 på x-aksen?

c) Hvor mange procent af observationerne ligger til højre for 8 på x-aksen?

d) Den del af området mellem grafen og x-aksen som ligger til højre for linjen med ligningen $x=12$, er uendelig langt. Hvad er arealet af dette område?

Opgave 38 Middelværdi og spredning for normalfordeling

Denne opgave skal besvares ved hjælp af de oplysninger om normalfordeling der står øverst på denne side.

En maskine fylder rødvin på literkartoner. Erfaringsmæssigt ved man at hvis man udtager mange kartoner og for hver af dem noterer hvor mange liter den indeholder, så vil de noterede tal være normalfordelt med middelværdien 0.995 liter og spredningen 0.005 liter.

Hvor mange procent (helt tal) af kartonerne indeholder under 1.000 liter rødvin?

Opgave 39 Repetition af beregning af varians ud fra stikprøve

Brug sætning 2 side 20-21 til at besvare følgende spørgsmål:

En stikprøve $x_1, x_2, ..., x_8$ udtages af en stor samling tal. Summen af kvadraterne $(x_1-\overline{x})^2, (x_2-\overline{x})^2, ..., (x_8-\overline{x})^2$ er 19.3 .

Hvilket tal er det bedst mulige gæt på variansen af tallene i den store samling?

Opgave 40 Repetition af tæthedsfunktion og fordelingsfunktion

Brug oplysningerne side 23-24 til at besvare følgende spørgsmål:

a) En graf viser tæthedsfunktionen for vægten af en bestemt type fisk.

Hvilken forbindelse er der mellem denne graf og histogrammer der viser frekvens-fordelinger for vægten af denne type fisk?

b) Når man har grafen for tæthedsfunktionen, hvordan kan man så se hvor stor en brøkdel af fiskene der vejer mellem 400 gram og 500 gram? (I svarets formulering kan indgå en skitse).

c) Lad G være fordelingsfunktionen for vægten af den pågældende type fisk. Det er oplyst at $G(500) = 47\%$.

Hvad siger dette om fiskenes vægt?

d) Hvordan kan man ved hjælp af fordelingsfunktionen G beregne hvilken brøkdel af fiskene der vejer mellem 400 gram og 500 gram?

Opgave 41 Situationer hvor normalfordeling kan bruges

I 5.2 på side 25 er omtalt tre typer situationer hvor man typisk får tal der er normalfordelt.

Giv for hver af disse typer et konkret eksempel på hvad du kunne måle for at få nogle tal der sandsynligvis er normalfordelt.

Computer

Man kan få computeren til at foretage beregninger og tegne grafer vedrørende normalfordelinger. I det følgende forudsættes at du har fået at vide hvordan dette gøres.

Opgave 42 Normalfordeling

Ejeren af en fadølsrobot kan indstille hvor meget robotten skal hælde i hvert glas. Robotten er ikke præcis, så der kommer ikke samme mængde i alle glas. Erfaringsmæssigt ved man at hvis man noterer mængderne i mange af disse glas, så vil de noterede tal være normalfordelt med spredningen 0.06 liter. Middelværdien afhænger af hvad robotten er indstillet til at hælde i hvert glas.

Ejeren påstår at robotten er indstillet sådan at middelværdien er 0.50 liter, men en kunde får kun 0.40 liter i sit glas og påstår at robotten ikke er indstillet så middelværdien er 0.50 liter.

I hvor stor en procentdel af glassene vil robotten hælde 0.40 liter eller mindre hvis det er rigtigt at middelværdien er 0.50?

Antag at ejeren for at snyde har indstillet robotten så middelværdien er 0.45 liter. I hvor stor en procentdel af glassene vil der så blive hældt 0.40 liter eller mindre?

Tegn i samme koordinatsystem graferne for tæthedsfunktionerne svarende til middelværdierne 0.50 og 0.45.

Markér (fx ved en prik) 0.40 på x-aksen.

Sørg for at figuren er omhyggeligt udført (passende udsnit, passende tal på akserne, overskrift, tekst ved x-aksen osv.)

Tilføj tekst hvor der står hvad det er der regnes ud.

Udskriv besvarelsen.

Opgave 43 Normalfordeling

På samme måde som i opgave 42 skal du lave en omhyggelig besvarelse af spørgsmålene i denne opgave. Husk figur og tekst.

Et stort antal elever regner samme prøve. I stedet for en sædvanlig karakter får eleverne et pointtal. Teoretisk set skulle pointtallene være normalfordelt med middelværdi 50.0 og spredning 10.0 .

Bestågrænsen er ca. 44.0 point.

Hvor mange procent af eleverne består?

Der uddeles et diplom til de bedste elever. Ca. 10% får diplom.

Hvilket pointtal er diplomgrænse?

Et år begås en fejl så middelværdien bliver 60.0 . Hvor mange procent får dette år et pointtal over bestågrænsen 44.0 point?

Diplomgrænsen var offentliggjort før fejlen blev opdaget. Hvor mange procent fik et pointtal der lå over den diplomgrænse du bestemte ovenfor?

Opgave 44 Normalfordeling

Man har givet 1254 børn en intelligenstest. Deres intelligenskvotienter var normalfordelt med middelværdi 100 og spredning 13.

Børn med intelligenskvotient i intervallet 90-110 kaldes almindeligt begavede.

Hvor mange procent af børnene var almindeligt begavede?

Hvor mange procent af børnene havde en intelligenskvotient over 120?

Lav en illustration af samme type som i de to foregående opgaver.

Opgave 45 Normalfordeling

En maskine fremstiller metalkugler. Deres vægte (målt i gram) er normalfordelt med middelværdi $\mu = 28.60$ og spredning $\sigma = 0.91$.

Hvor mange procent af kuglerne vejer under 26.00 gram?

Kugler hvis vægt afviger mere end 3% fra 28.60 gram, bliver kasseret.

Hvor mange procent af kuglerne bliver kasseret?

Lav en illustration med tæthedsfunktion og markering af de grænser en kugles vægt skal ligge imellem for at kuglen ikke bliver kasseret.

5.5 Sådan kan opgaver løses ved at bruge fordelingsfunktionen $F(x)$

Nogle målte tal er normalfordelt med middelværdi 6.5 og spredning 1.2 .

På computeren starter vi med at taste at F er fordelingsfunktionen for denne fordeling. Herefter kan vi bruge F til udregninger.

Hvilken procentdel af de målte tal er mindre end 5?

Svaret er

$$F(5.0) = 0.10565 \ , \ \text{altså } 11\%$$

da $F(x)$ er den procentdel af de målte tal der er mindre end x.

Hvilken procentdel af de målte tal er større end 5?

Svaret er

$$100\% - 11\% = 89\%$$

da 11% af de målte tal er mindre end 5 ifølge den første udregning.

Hvilken procentdel af de målte tal ligger mellem 4 og 5?

Svaret er

$$F(5.0) - F(4.0) = 0.08704 \ , \ \text{altså } 9\%$$

da $F(5.0)$ er den procentdel af de målte tal der er mindre end 5.0, og $F(4.0)$ er den procentdel af de målte tal der er mindre end 4.0 .

Om hvilket tal a gælder at 5% af de målte tal er mindre end a ?

Svaret er det tal a hvorom der gælder at

$$F(a) = 0.05$$

da $F(a)$ er den procentdel af de målte tal som er mindre end a .

Vi får computeren til at løse denne ligning. Svaret er

$$a = 4.5262 \ , \ \text{altså } 4.5 .$$

Om hvilket tal a gælder at 80% af de målte tal er større end a ?

Svaret er det tal a hvorom der gælder at

$$F(a) = 0.20$$

da $100\% - 80\% = 20\%$ af de målte tal er er mindre end a .

Vi får computeren til at løse denne ligning. Svaret er

$$a = 5.4901 \ , \ \text{altså } 5.5 .$$

Opgave 46 Brug af fordelingsfunktion

Længderne af nogle muslinger er normalfordelt med middelværdien 4.7 cm og spredningen 0.8 cm.

Hvor mange procent af muslingerne har en længde over 3.7 cm?

Hvor mange procent af muslingerne har en længde der afviger mindre end 0.5 cm fra middelværdien?

En dag får du ordre til at frasortere de 5% af muslingerne der er længst. Hvor lange er de muslinger du skal frasortere?

En anden dag får du ordre til at frasortere de 5% af muslingerne hvis længder afviger mest fra middelværdien. Hvilke længder har de muslinger du skal frasortere?

Få tegnet grafen for fordelingens tæthedsfunktion, og markér på figuren de to længder a og b hvorom der gælder at a er mindre end b, at middelværdien ligger midt mellem a og b, og at 95% af længderne ligger mellem a og b.

Hvor stort er arealet af det område under grafen der svarer til længder mellem a og b?

Hvor stort er arealet af det område under grafen der svarer til længder under a?

Opgave 47 Tæthedsfunktion for middelværdier

Dette er en fortsættelse af opgave 46.

Vi udtager tilfældigt 4 muslinger, lægger dem i en æske, og skriver middelværdien af deres længder på æsken. Dette gør vi mange gange. De tal vi har skrevet på æskerne, er normalfordelt med middelværdi 4.7 cm og spredning 0.4 cm.

Få tegnet grafen for disse tals tæthedsfunktion i det koordinatsystem hvor du fik tegnet grafen for tæthedsfunktionen fra opgave 46.

Hvor stort er arealet af det område under den nye graf der svarer til længder mellem de tal a og b du fandt i opgave 46?

Opgave 48 Normalfordelte tilvækster

Nogle dyr vejes den 1. maj, og de samme dyr vejes igen den 1. juni. For hvert dyr udregnes slutvægt minus startvægt, altså tilvækst i vægt. Disse tilvækster er normalfordelt med middelværdi -2.1 gram og spredning 1.4 gram .

Når man for et af dyrene har udregnet tilvæksten i vægt (altså slutvægt minus startvægt), hvordan kan man så se på resultatet om dyret er gået ned i vægt? Hvis dyret er gået 2 gram ned i vægt, hvad har dyrets tilvækst i vægt så været?

Få tegnet grafen for tilvæksternes tæthedsfunktion.

Hvor mange procent af dyrene har øget deres vægt i løbet af maj måned?

Hvad er arealet af den del af området under grafen som ligger til højre for y-aksen?

Hvor mange procent af dyrene er gået mere end 3.0 gram ned i vægt?

Udregn det tal k hvorom der gælder at 5% af dyrene har øget deres vægt mere end k gram.

Kapitel 6. Test

Afsnit 6.1: Signifikansniveau og forkastelsesgrænser

6.1.1 Sådan udføres en test

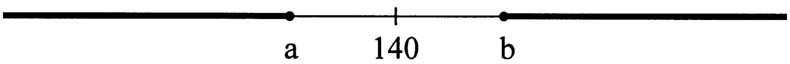
Hypotesen vi vil teste

En maskine pakker pulver i poser. Posernes vægte, målt i gram, er normalfordelt med spredning $\sigma = 9$. Middelværdien μ afhænger af hvordan maskinen er indstillet.

Det påstås at middelværdien er 140 gram. Vi vil undersøge denne påstand, dvs. vi vil *teste hypotesen* $\mu = 140$.

Sådan kan vi udregne forkastelsesgrænserne a og b

På figuren er vist to tal a og b som ligger lige langt fra middelværdien 140.



Vi vil bestemme a og b så der gælder:

> Hvis hypotesen $\mu = 140$ er korrekt, så ligger 5% af posernes vægte uden for grænserne a og b.

Dette kan gøres sådan:

> De to tal a og b er bestemt ved at 2.5% af vægtene er $\leq a$ og 2.5% af vægtene er $\geq b$. På computeren taster vi at F skal være <u>fordelingsfunktionen</u> for normalfordelingen med $\mu = 140$ og $\sigma = 9$. Ved at lade computeren løse ligningerne
>
> $$F(a) = 2.5\% \quad \text{og} \quad 100\% - F(b) = 2.5\%$$
>
> fås $a = 122$ og $b = 158$.

Figuren på næste side viser grafen for <u>tæthedsfunktionen</u> f for normalfordelingen med $\mu = 140$ og $\sigma = 9$. Grænserne a og b er markeret på figuren.

Test af hypotesen

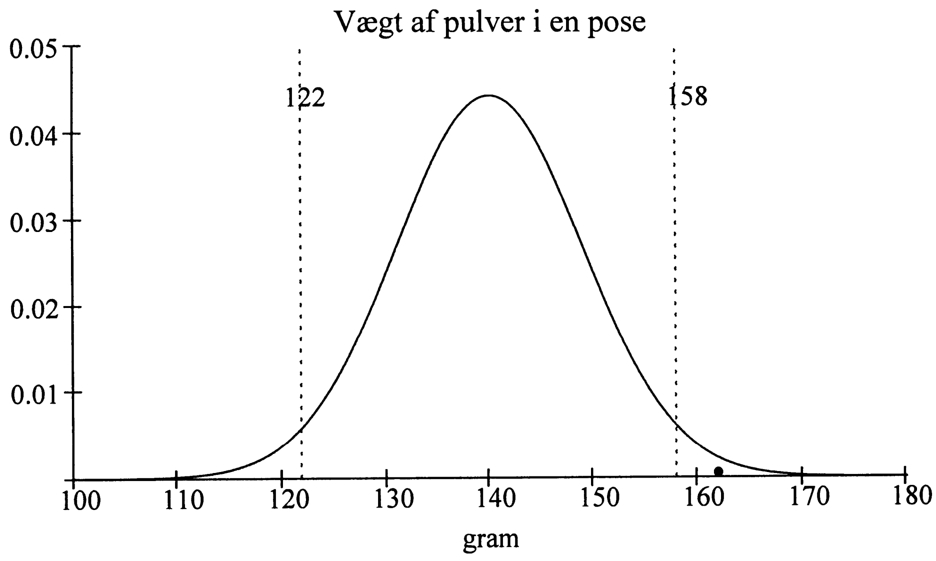
De to små områder under grafen uden for grænserne 122 og 158 har begge arealet 0.025, dvs. 5% af posernes vægte ligger på x-aksen uden for disse grænser.

Vi køber en pose og vejer den. Dens vægt er 162 gram.

Da vægten ligger uden for grænserne 122 og 158, *forkaster vi hypotesen.*

Begrundelsen for dette er følgende:

> Hvis figuren var rigtig (dvs. hvis $\mu = 140$), så var det usandsynligt at vi ville få en pose hvis vægt lå uden for grænserne.



Signifikansniveau

Ovenfor valgte vi forkastelsesgrænserne a og b sådan at der gjaldt:

Hvis hypotesen er rigtig, så ligger **5%** af alle vægtene uden for disse grænser.

Undertiden vælger man en anden procent end 5%, fx **1%**. Den procent man vælger, kaldes *signifikansniveauet*. For hypotesen $\mu = 140$ om poserne gælder altså:

På signifikansniveau 5% forkastes hypotesen.

Opgave 49 Udregne forkastelsesgrænserne a og b

I 6.1.1 er omtalt nogle poser hvis vægte, målt i gram, er normalfordelt med spredning $\sigma = 9$. Vi testede hypotesen $\mu = 140$, dvs. hypotesen at middelværdien er 140 gram. Testen blev udført på signifikansniveau 5%, dvs. forkastelsesgrænserne a og b blev valgt sådan at hvis hypotesen $\mu = 140$ er korrekt, så ligger 5% af posernes vægte uden for grænserne a og b.

Bestem de forkastelsesgrænser a og b som vi skulle bruge hvis vi ville teste hypotesen på signifikansniveu 1%.

I 6.1.1 står at den købte pose vejede 162 gram. Forkastes hypotesen hvis vi vælger et signifikansniveau på 1% ?

Hvis hypotesen er rigtig, hvilken procentdel af alle posernes vægte ligger så uden for de grænser du bestemte i første spørgsmål i denne opgave?

Forestil dig at hypotesen er korrekt, og at mange kunder tester hypotesen ved at købe en pose og veje den. Hvilken procentdel af disse kunder får en vægt der ligger uden for forkastelsesgrænserne?

Opgave 50 Er det sandsynligt at forkert hypotese forkastes?

I 6.1.1 er omtalt nogle poser hvis vægte, målt i gram, er normalfordelt med spredning $\sigma = 9$. Vi fandt frem til at hypotesen $\mu = 140$ skulle forkastes hvis den pose vi købte, lå uden for grænserne $a = 122$ og $b = 158$.

Antag at posernes middelværdi er 150 gram. Hvor mange procent af poserne har så en vægt mellem 122 gram og 158 gram?

Hvis vi køber en pose og vejer den, er det så mest sandsynligt at vi vil forkaste den forkerte hypotese $\mu = 140$?

6.1.2 Det er en dårlig test

På figuren øverst på side 32 ses at hvis den korrekte værdi af μ er fx 155, så er det sandsynligt at den udtagne poses vægt ligger mellem grænserne 122 og 158. Selv om den rigtige middelværdiværdi 155 ligger langt fra 140, er det altså sandsynligt at hypotesen $\mu = 140$ ikke forkastes. Derfor er det en dårlig test.

I afsnit 6.3 står hvordan man kan lave en bedre test. Afsnit 6.2 går ud på at udlede en regel som skal bruges i afsnit 6.3 .

Opgave 51 Test

Prisen for en vare er er ikke den samme i alle forretninger. Prisen er normalfordelt med spredning 800 kr.

I en tilfældig forretning ses at prisen er 4995 kr. Kan man ud fra denne oplysning forkaste hypotesen $\mu = 5800$ på signifikansniveau 5%?

Lav en illustration af samme type som figuren på foregående side.

Afsnit 6.2: Spredning af middelværdier

6.2.1 "En række tilfældige tal"

Når vi i disse noter taler om en række tilfældige tal t_1, t_2, t_3 - - - , så tænker vi på et uoverskueligt stort antal tal som vi kun kan undersøge ved stikprøver, fx vægtene af æblerne i en plantage.

Da rækkefølgen af tallene t_1, t_2, t_3 - - - er tilfældig, kan de første tal i rækken bruges som en tilfældig stikprøve.

Sætning 3 Spredning af tallene $k \cdot x_1$, $k \cdot x_2$, $k \cdot x_3$, $- - -$

Hvis en række tilfældige tal

$$x_1,\ x_2,\ x_3,\ ---$$

har spredningen σ, så vil rækken af tallene

$$k \cdot x_1\ ,\ k \cdot x_2\ , k \cdot x_3\ ,\ ---$$

hvor k er positiv, have spredningen $k \cdot \sigma$.

Bevis for sætning 3

Hvis spredningen af

$x_1, x_2, x_3, ---$

er σ, så er variansen $V = \sigma^2$. Af sætning 1 side 19 følger at variansen af

$k \cdot x_1 , \; k \cdot x_2 , k \cdot x_3 , ---$

er $k^2 \cdot V = k^2 \cdot \sigma^2$. Så er spredningen lig kvadratroden af dette, dvs. $k \cdot \sigma$.

Sætning 4 Varians af summer

Hvis en række tilfældige tal

$x_1, x_2, x_3, ---$

har varians V_x, og en række tilfældige tal

$y_1, y_2, y_3, ---$

der er uafhængige af x'erne, har varians V_y, så vil summerne

$s_1 = x_1+y_1 , \quad s_2 = x_2+y_2 , \quad s_3 = x_3+y_3 , \quad ---$

have varians $V_x + V_y$.

Bevis for sætning 4 springes over.

Opgave 52 Varians af sum

Nogle frugter lægges i hver sin æske. Frugternes vægte (målt i gram) har variansen 16, og de tomme æskers vægte (målt i gram) har variansen 9.

Hvad er variansen af de fyldte æskers vægte?

Hvad er spredningen af vægtene af hhv, frugterne, æskerne og de fyldte æsker?

Opgave 53 Varians

Udgiften ved anskaffelse af en bestemt type vare består af pris og forsendelse:

udgift = pris + forsendelse .

Prisen afhænger af varens vægt, og betalingen for forsendelse fås ved at gange prisen med 0.5 :

forsendelse = 0.5 · pris .

Tallene

$x_1, x_2, x_3, ---$

er priser på en række tilfældigt udtagne varer, og tallene

$y_1, y_2, y_3, ---$

er forsendelse for disse varer.

Summerne

$s_1 = x_1+y_1 , \quad s_2 = x_2+y_2 , \quad s_3 = x_3+y_3 , \quad ---$

er udgifterne ved anskaffelse af de pågældende varer.

Det er oplyst at x'erne har variansen 4, og at y'erne har variansen 1.

Hvorfor kan sætning 4 ikke bruges til at udregne variansen af s'erne?

Brug en anden regel til at udregne variansen af s'erne.

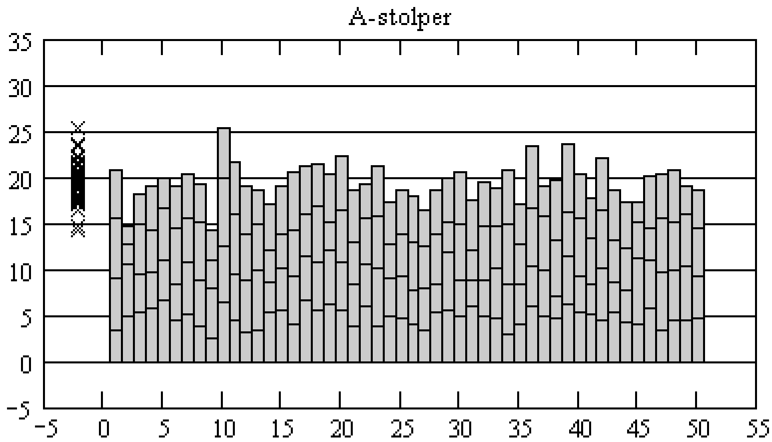
6.2.2: Spredning af summer

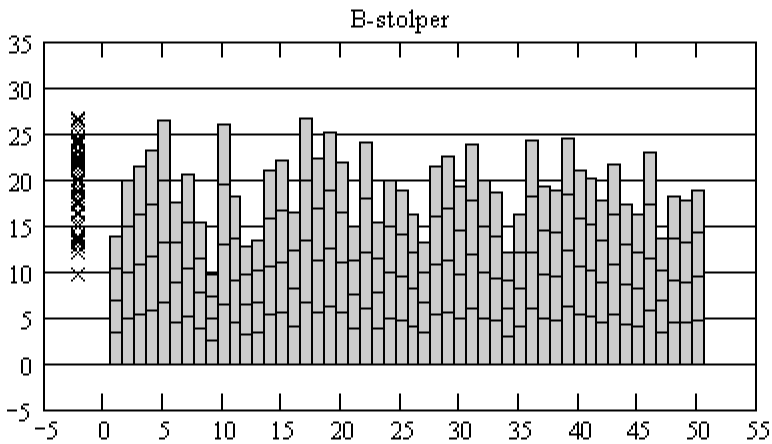
Til et spil fremstilles klodser hvis længder er normalfordelt med middelværdi 5 og spredning 1.

Ved hjælp af klodserne laves to typer stolper som vi kalder A-stolper og B-stolper:

– En **A-stolpe** laves ved at anbringe 4 tilfældigt valgte klodser i forlængelse af hinanden. I det øverste af koordinatsystemerne nedenfor er vist nogle af A-stolperne.

– En **B-stolpe** laves ved at vælge en tilfældig klods og anbringe 3 andre klodser af samme længde i forlængelse af den. I det nederste af koordinatsystemerne er vist nogle af B-stolperne.





For begge typer stolper gælder at middelværdien af deres længder er 20, for de består jo af 4 klodser, og klodsernes gennemsnitlige længde er 5.

For B-stolperne er spredningen af højderne 4. Af sætning 3 følger nemlig at spredningen er 4 gange spredningen af klodslængderne, og denne er 1.

For A-stolperne er spredningen mindre end for B-stolperne. Dette ses tydeligt ved at sammenligne de to figurer. Grunden er følgende:
– Blot længden af den første klods i en B-stolpe er langt fra middelværdien, så får man en stolpehøjde langt fra middelværdien.
– Selv om længden af den første klods i en A-stolpe er langt fra middelværdien, behøver stolpehøjden ikke være langt fra middelværdien da der ofte er nogle af de andre klodser i stolpen der ødelægger det.

Det viser sig at A-stolpehøjdernes spredning kun er 2.

Hvis hver stolpe bestod af 9 klodser, så ville spredningen være 9 for B-stolperne og 3 for A-stolperne.
Hvis hver A-stolpe bestod af n klodser, ville A-stolpehøjderne have spredning $\sqrt{n}$.

Sætning 5 Spredning af summer

Hvis en række tilfældige tal

$$x_1,\ x_2,\ x_3,\ ---$$

har spredning σ, så vil summer

$$s_1 = x_1+x_2+...+x_n\ ,\quad s_2 = x_{n+1}+x_{n+2}+...+x_{2n}\ ,\quad ---$$

af n af disse have spredningen $\sigma\cdot\sqrt{n}$.

Bevis for sætning 5

Tallene

$$x_1,\ x_2,\ x_3,\ ---$$

har varians σ^2, så tallene

$$s_1 = x_1+x_2+...+x_n\ ,\quad s_2 = x_{n+1}+x_{n+2}+...+x_{2n}\ ,\quad ---$$

har ifølge sætning 4 variansen

$$\underbrace{\sigma^2+\sigma^2+...+\sigma^2}_{n} = n\cdot\sigma^2$$

og dermed spredningen

$$\sqrt{n\cdot\sigma^2} = \sqrt{n}\cdot\sqrt{\sigma^2} = \sqrt{n}\cdot\sigma\ .$$

Opgave 54 Spredning af summer

Nogle appelsiners vægte har en spredning på 45 gram. Appelsinerne sælges i net med 15 i hver.

Hvad er spredningen af nettenes vægte?

Sætning 6 Spredning af middelværdier

Hvis en række tilfældige tal

$$x_1,\ x_2,\ x_3,\ ---$$

har spredning σ, så vil middelværdierne

$$\overline{x}_1 = \frac{x_1 + x_2 + \ldots + x_n}{n}\ ,\quad \overline{x}_2 = \frac{x_{n+1} + x_{n+2} + \ldots + x_{2n}}{n}\ ,\quad ---$$

af n af disse have spredningen $\dfrac{\sigma}{\sqrt{n}}$.

Bevis for sætning 6

Af sætning 3 følger at når nogle tal ganges med et positivt tal, så bliver deres spredning ganget med dette tal, så da spredningen af tallene

$$s_1 = x_1+x_2+\ldots+x_n\ ,\quad s_2 = x_{n+1}+x_{n+2}+\ldots+x_{2n}\ ,\quad ---$$

er $\sigma\cdot\sqrt{n}$, vil spredningen af tallene

$$\overline{x}_1 = \frac{1}{n}\cdot\left(x_1 + x_2 + \ldots + x_n\right)\ ,\quad \overline{x}_2 = \frac{1}{n}\cdot\left(x_{n+1} + x_{n+2} + \ldots + x_{2n}\right)\ ,\quad ---$$

være

$$\frac{1}{n}\cdot\left(\sigma\cdot\sqrt{n}\right) = \sigma\cdot\frac{\sqrt{n}}{n} = \sigma\cdot\frac{1}{\sqrt{n}}\ .$$

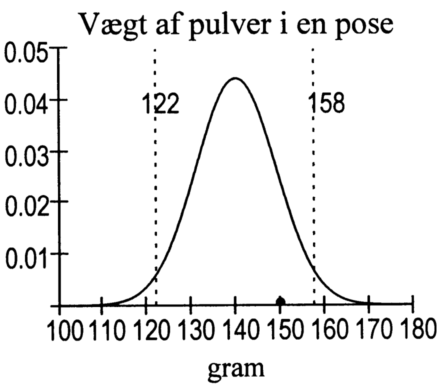
Afsnit 6.3: Forbedret test

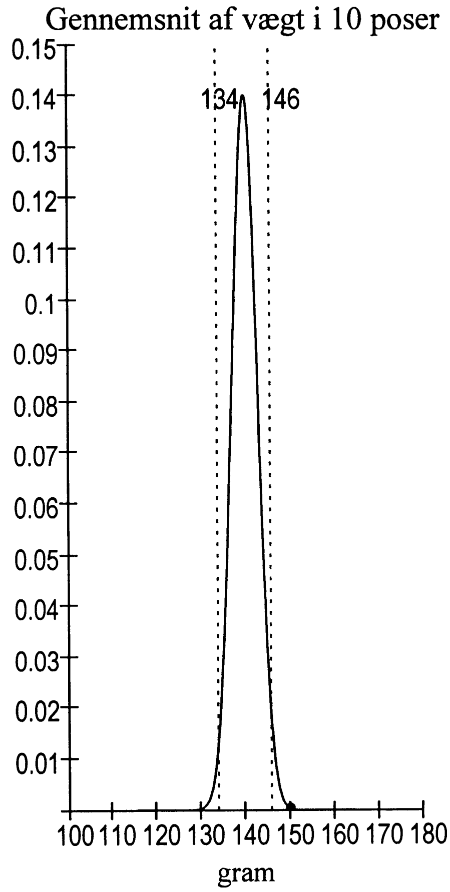
6.3.1: Forbedret test

I 6.1.1 blev omtalt en test af hypotesen $\mu = 140$ vedrørende vægten af nogle poser med pulver. Antag at den rigtige middelværdi er 150.

Den venstre figur på næste side viser grafen for tæthedsfunktionen for tal der er normalfordelt med middelværdi 140 og spredning 9. Af denne figur ses at det ikke er sandsynligt at vi får forkastet den forkerte hypotese. Dette skyldes at spredningen $\sigma = 9$ er for stor.

Testen kan forbedres ved at bruge sætning 6: Hvis vi i stedet for vægten af én poses indhold bruger middelværdien af vægtene af 10 posers indhold, så bliver spredningen $\sigma = \frac{9}{\sqrt{10}} \approx 2.85$. Grafen for tæthedsfunktionen svarende hertil er vist på den højre figur nedenfor. Af denne figur ses at nu er det sandsynligt at få forkastet den forkerte hypotese.





Opgave 55 Tæthedsfunktion

Hvorfor må grafen for tæthedsfunktionen nødvendigvis blive højere samtidig med at den bliver smallere?

Opgave 56 Forkastelsesgrænser m.m.

Antag at vi ved undersøgelsen af poserne fra 6.3.1 benyttede middelværdien af vægtene af 100 poser. Hvad ville så være grænserne for at forkaste på signifikansniveau 5%?

Forestil dig at hypotesen $\mu = 140$ er korrekt, og at mange personer tester den ved at bestemme 100 posers vægte og se om disses middelværdi ligger uden for de grænser du bestemte i første spørgsmål. Hvor mange procent af disse personer forkaster hypotesen?

Computer

I det følgende forudsættes at du har fået at vide hvordan du kan gøre følgende på computeren:

- Taste en stikprøve (dvs. en række tal) sådan at du kan få udført flere udregninger med tallene uden at taste dem igen.
- Nemt udregne middelværdi og spredning ud fra stikprøve.

Opgave 57 Test

Længderne målt i cm af nogle fisk er normalfordelt med spredning 2 cm. Den nøjagtige middelværdi afhænger af hvor gunstige vækstbetingelserne har været. Man vil teste hypotesen $\mu = 16$.

Hvis nogen mange gange udtager en stikprøve på 30 fisk og nedskriver middelværdien af deres længder, hvad vil spredningen af de nedskrevne tal så være?

Én gang udtager vi en stikprøve på 30 fisk og udregner middelværdien $\overline{x}$ af deres længder. Bestem de forkastelsesgrænser a og b som tallet $\overline{x}$ skal ligge udenfor for at hypotesen forkastes på signifikansniveau 5%.

Kan følgende stikprøve forkaste hypotesen på signifikansniveau 5%:

$$\begin{pmatrix} 16.3 & 17.3 & 16.8 & 17.5 & 16.7 & 17.4 & 17.3 & 17.4 & 16.9 & 17.4 \\ 17.0 & 17.5 & 16.8 & 16.8 & 16.6 & 16.5 & 17.3 & 17.1 & 17.3 & 17.1 \\ 17.0 & 15.9 & 16.4 & 16.9 & 16.4 & 16.8 & 16.9 & 16.7 & 16.4 & 17.2 \end{pmatrix}$$

Opgave 58 Test

Nogle varers vægte er normalfordelt med spredning 62 gram. Vægtenes middelværdi kontrolleres dagligt ved at udtage en stikprøve på 20 styk.

Hvad er spredningen σ for middelværdier af stikprøver på 20 styk?

En dag har dagens stikprøve middelværdien $\overline{x} = 807$.

Afgør om hypotesen $\mu = 830$ bliver forkastet på signifikansniveau 5%.

Opgave 59 Test

Nogle perlers vægte, målt i gram, er normalfordelt med spredning 1.3. For at teste hypotesen $\mu = 7$ udtages følgende stikprøve:

$$\begin{pmatrix} 8 & 6 & 6 & 6 & 7 & 5 & 8 & 6 & 5 & 4 \end{pmatrix}$$

Forkastes hypotesen?

Afsnit 6.4: Spredning af differenser

Når to firmaer A og B fremstiller samme type vare, kan man være interesseret i at undersøge om varerne fra A og B gennemsnitligt har samme holdbarhed. I denne og mange andre situationer skal man teste om der er forskel. Til disse test skal bruges nogle formler for spredning af differenser. Disse formler udledes i dette afsnit.

Sætning 7 Varians af differenser

Hvis en række tilfældige tal

$$x_1,\ x_2,\ x_3,\ ---$$

har varians V_x, og en række tilfældige tal

$$y_1,\ y_2,\ y_3,\ ---$$

der er uafhængige af x'erne, har varians V_y, så vil differenserne

$$d_1 = x_1 - y_1\ ,\quad d_2 = x_2 - y_2\ ,\quad d_3 = x_3 - y_3\ ,\quad ---$$

have varians $V_x + V_y$.

(Bemærk at der er *plus* mellem V_x og V_y selv om der er *minus* mellem x_i og y_i).

Bevis for sætning 7 springes over.

Sætning 8 Spredning af differenser

Hvis en række tilfældige tal

$$x_1,\ x_2,\ x_3,\ ---$$

har spredningen σ_x, og en række tilfældige tal

$$y_1,\ y_2,\ y_3,\ ---$$

der er uafhængige af x'erne, har spredning σ_y, så vil differenserne

$$d_1 = x_1 - y_1\ ,\quad d_2 = x_2 - y_2\ ,\quad d_3 = x_3 - y_3\ ,\quad ---$$

have spredning $\sqrt{\sigma_x{}^2 + \sigma_y{}^2}$.

Bevis for sætning 8

Varianserne af x'erne og y'erne kalder vi V_x og V_y.

Hvis spredningen af x'erne og y'erne er hhv. σ_x og σ_y, så er

$$V_x = \sigma_x{}^2 \quad \text{og} \quad V_y = \sigma_y{}^2 .$$

Af sætning 7 følger at variansen V_d af differenserne

$$d_1 = x_1 - y_1 \;, \quad d_2 = x_2 - y_2 \;, \quad d_3 = x_3 - y_3 \;, \quad ---$$

er

$$V_d = V_x + V_y = \sigma_x{}^2 + \sigma_y{}^2 \,.$$

Differensernes spredningen σ_d er derfor

$$\sigma_d = \sqrt{V_d} = \sqrt{\sigma_x{}^2 + \sigma_y{}^2} \,.$$

Sætning 9 Spredning af differens af middelværdier

Hvis vi har to uafhængige rækker tilfældige tal

$$x_1,\ x_2,\ x_3,\ --- \qquad \text{og} \qquad y_1,\ y_2,\ y_3,\ ---$$

hvis spredninger er hhv. σ_x og σ_y, så vil middelværdier

$$\overline{x}_1 = \frac{x_1 + x_2 + \ldots + x_m}{m} \;, \quad \overline{x}_2 = \frac{x_{m+1} + x_{m+2} + \ldots + x_{2m}}{m} \;, \quad ---$$

og

$$\overline{y}_1 = \frac{y_1 + y_2 + \ldots + y_n}{n} \;, \quad \overline{y}_2 = \frac{y_{n+1} + y_{n+2} + \ldots + y_{2n}}{n} \;, \quad ---$$

af x'er og y'er have differenser

$$d_1 = \overline{x}_1 - \overline{y}_1 \;, \quad d_2 = \overline{x}_2 - \overline{y}_2 \;, \quad d_3 = \overline{x}_3 - \overline{y}_3 \;, \quad ---$$

hvis spredning er

$$\sigma_d = \sqrt{\frac{\sigma_x{}^2}{m} + \frac{\sigma_y{}^2}{n}} \,.$$

Bevis for sætning 9

De to rækker af middelværdier

$$\overline{x}_1,\ \overline{x}_2,\ \overline{x}_3,\ ---$$

og

$$\overline{y}_1,\ \overline{y}_2,\ \overline{y}_3,\ ---$$

som er nævnt i sætningen, har ifølge sætning 6 spredningerne

$$\sigma_{\overline{x}} = \frac{\sigma_x}{\sqrt{m}} \qquad \text{og} \qquad \sigma_{\overline{y}} = \frac{\sigma_y}{\sqrt{n}} \,.$$

Af sætning 8 fås derfor at differenserne

$$d_1 = \overline{x}_1 - \overline{y}_1 \;, \quad d_2 = \overline{x}_2 - \overline{y}_2 \;, \quad d_3 = \overline{x}_3 - \overline{y}_3 \;, \quad ---$$

har spredningen

$$\sigma_d = \sqrt{\sigma_{\bar{x}}^2 + \sigma_{\bar{y}}^2} = \sqrt{\left(\frac{\sigma_x}{\sqrt{m}}\right)^2 + \left(\frac{\sigma_y}{\sqrt{n}}\right)^2} = \sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}} .$$

Afsnit 6.5: Test af om der er forskel

6.5.1 Hvordan kan vi teste om der er forskel?

For en bestemt type batterier regnes med at deres holdbarheder (målt i timer) er normalfordelt med spredning 1.6 . Vi vil undersøge om to typer af disse batterier har forskellig holdbarhed.

Først måler vi holdbarhederne $x_1, x_2, ..., x_{10}$ af 10 batterier af den ene type og holdbarhederne $y_1, y_2, ..., y_{10}$ af 10 batterier af den anden type. Så beregner vi differensen $z = \bar{x} - \bar{y}$. Vi udregner kun én differens af denne type, men vi ved at hvis vi udregnede en masse differenser af denne type, så ville disse være normalfordelt med spredningen

$$\sigma_d = \sqrt{\frac{1.6^2}{10} + \frac{1.6^2}{10}} = 0.72 .$$

Hvis der ikke er forskel på holdbarhederne vil differensernes middelværdi μ_d være lig 0, så vi vil teste hypotesen

$$\mu_d = 0 ,$$

og lad os vælge et signifikansniveauet på 5%. Vi bestemmer derfor forkastelsesgrænserne a og b sådan at 5% af differenserne ligger uden for disse grænser hvis hypotesen er rigtig. Hvis det beregnede tal z ligger uden for disse grænser, så forkastes hypotesen om at der ikke er forskel på de to typers holdbarhed.

Opgave 60 Udregning af a og b fra 6.5.1

Hvis hypotesen er rigtig, er differenserne normalfordelt med middelværdi 0, og spredningen er i eksemplet udregnet til 0.72 .

Brug fordelingsfunktionen F for denne normalfordeling til at finde forkastelsesgrænserne a og b.

Opgave 61 Udregning af z fra eksempel 6.5.1

De 10 batterier af den ene type havde holdbarhederne

5.5 8.1 8.7 10.6 9.0 5.3 5.8 11.8 9.2 8.0

og de 10 batterier af den anden type havde holdbarhederne

5.6 5.4 4.4 5.3 8.8 8.5 8.2 5.3 6.3 9.6

Udregn differensen $z = \overline{x} - \overline{y}$.

I opgave 60 blev forkastelsesgrænserne udregnet. Forkastes hypotesen om at der ikke er forskel på holdbarhederne?

Opgave 62 Test af om der er forskel

Når man måler mængden (i mg/kg) af et bestemt giftstof i fisk af en bestemt type, plejer måleresultaterne at være normalfordelt med spredning 0.04 .

I et vandløb har man to år i træk målt mængden af gift i en række fisk af den nævnte type:

1. år: 0.81 0.83 0.71 0.82 0.72 0.71 0.75 0.76 0.80 0.76 0.71 0.77

2. år: 0.76 0.78 0.81 0.73 0.76 0.70 0.65 0.81 0.73 0.68

Kan vi på signifikansniveau 5% forkaste hypotesen: Mængden af giftstof er ikke ændret?

Afsnit 6.6: Testning når spredningen er ukendt

6.6.1 Hvordan man kan teste uden at kende spredningen på forhånd

Vi er nødt til at kende spredningen hvis vi vil teste på de måder vi har brugt i det foregående, men ofte kender man ikke spredningen på forhånd.

En af måderne at løse dette problem på er at foretage så mange målinger at man ud fra disse kan udregne et tilstrækkeligt godt gæt på spredningen.

Der skal helst være 30 eller flere målinger.

Spredningen udregnes ved hjælp af sætning 2 side 20-21.

Opgave 63 Test når spredningen ikke kendes på forhånd

For at undersøge om nogle dyr gennemsnitligt ændrer deres vægt i en bestemt periode, udvælges tilfældigt 100 dyr. For hver af disse bestemmes

vægttilvækst = slutvægt - startvægt

Enheden er kg.

Vi går ud fra at vægttilvæksterne er normalfordelt.

Ved hjælp af de to formler i sætning 2 side 20-21 beregnes:

Middelværdi af de 100 tal: 0.877

Gæt på spredningen: 3.28

Bemærk at det ikke er spredningen af middelværdierne der er angivet.

Kan man på signifikansniveau 5% forkaste hypotesen: Gennemsnitligt har dyrene ikke ændret vægt?

Opgave 64 Test når spredningen ikke kendes på forhånd

I hvert af to lande vælges tilfældigt 50 fjortenårige. De får hver en matematikprøve som bedømmes med et pointtal fra 0 til 100. Pointtallene står i det elektroniske dokument MatPrøve.

Afgør om man på signifikansniveau 5% kan forkaste hypotesen: Der er ikke forskel på de to lande mht. hvor gode de fjortenårige er til at løse opgaver af den stillede type.

(Du skal ikke taste lange udregninger. Det er meningen at computeren skal lave arbejdet).