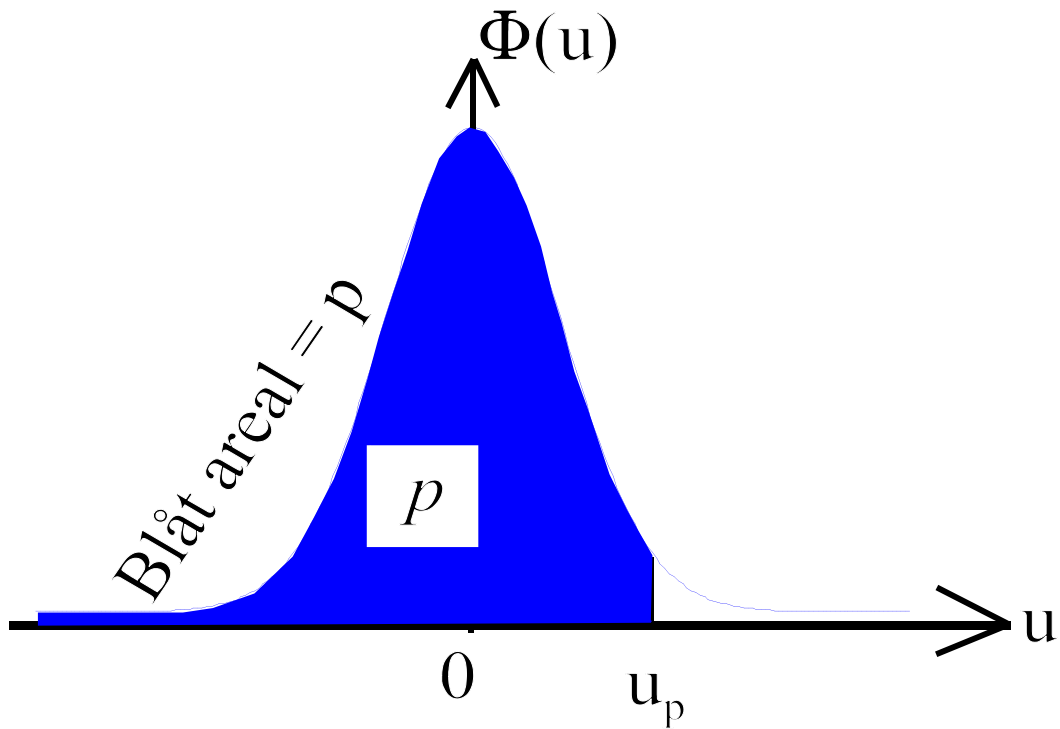


MOGENS ODDERSHEDE LARSEN

STATISTISKE GRUNDBEGREBER



8. udgave 2007

FORORD

Der er i denne bog søgt at give letlæst og anskuelig fremstilling af de statistiske grundbegreber til brug ved en indledende undervisning i statistik. De væsentligste definitioner og sætninger forklares derfor fortrinsvist ved hjælp af figurer og gennemregnede praktiske eksempler. Ønskes en mere matematisk uddybende forklaring, bevis for sætninger osv. kan dette ofte findes i et særskilt tillæg til bogen, som findes på nettet under titlen “Supplement til statistiske grundbegreber”.

Sidst i hver kapitel findes en række opgaver.

Bagerst i bogen findes en facitliste til alle opgaverne.

I appendix findes en række oversigter over de vigtigste metoder og formler.

Læsning: Bogen er bygget således op, at der hurtigt nås frem til normalfordelingen og de vigtige normalfordelingstest. Disse vigtige begreber kan derfor blive grundigt indarbejdet, selv om der kun er kort tid til rådighed. Er det af tidsmæssige grunde svært at nå hele notatet kan man uden skade for helheden udover supplementafsnittene overspringe kapitlerne 8 og 9, ligesom man eventuelt kan tage kapitlerne 3, 6 og 7 mere oversigtsagtigt.

I et længere kursusforløb er denne bog tænkt at skulle efterfølges af M. Oddershede Larsen: “Videregående Statistik”, som kan rekvireres ved henvendelse til e-mailadressen www.larsen-net.dk

Regnemidler. Det er hensigtsmæssigt, at man har adgang til en lommeregner eller en PC med normalfordeling, t - fordeling, binomialfordeling og Poissonfordeling indbygget.

I eksemplerne angives således, hvorledes beregningerne kan foretages med den i øjeblikket mest populære lommeregner TI-89, men også (med petit) hvorledes beregningerne udføres ved benyttelse af de statistiske tabeller, som findes bagerst i bogen.

Endvidere er der i et appendix også angivet hvorledes beregningerne kan udføres med TI-83, HP 48G, Excel og Maple.

Denne 8 udgave er indholdsmæssigt uændret i forhold til 7. udgave, idet dog eksemplerne nu også beregnes ved anvendelse af regnemaskine og ikke kun ved opslag i tabel.

Da denne udgave ikke mere trykkes i bogform, men udelukkende kan findes på adressen:

www.larsen-net.dk, har det også givet anledning til visse typografiske ændringer.

Jeg vil gerne takke ingeniørdocent L. Brøndum og J. D. Monrad for de mange gode råd gennem årene.

En særlig tak til lektor Bjarne Hellesen, som dels har skrevet afsnit 1.7 og kapitel 9, dels er kommet med mange værdifulde kommentarer og bidrag til forbedringer.

juni 2007

Mogens Oddershede Larsen

INDHOLD

1 STATISTISKE GRUNDBEGREBER

1.1	INDLEDNING	1
1.2	HÆNDELSER OG SANDSYNLIGHED	1
1.3	STATISTISK VARIABEL	3
1.4	DISKRET STATISTISK VARIABEL	4
1.5	KONTINUERT STATISTISK VARIABEL	6
1.6	LINEARKOMBINATION AF STATISTISKE VARIABLE	16
1.7	USIKKERHEDSBEREGNING	18
	OPGAVER	21

2 NORMALFORDELINGEN

2.1	INDLEDNING	26
2.2	DEFINITION OG BEREKNING	28
	OPGAVER	35

3 STATISTISKE TESTFUNKTIONER

3.1	INDLEDNING	38
3.2	χ^2 - FORDELINGEN	38
3.3	t - FORDELINGEN	40
3.4	F - FORDELINGEN	41
	OPGAVER	43

4 ESTIMATER OG KONFIDENSINTERVALLER

4.1	PUNKTESTIMATION	44
4.2	KONFIDENSINTERVAL	45
	OPGAVER	51

5 HYPOTSETESTNING (1 STATISTISK VARIABEL)

5.1 INDLEDNING	53
5.2 GRUNDLÆGGENDE BEGREBER	53
5.3 EKSEMPLER PÅ HYPOTSETEST	61
5.4 PARVISE OBSERVATIONER	65
OPGAVER	67

6 SANDSYNLIGHEDSREGNING

6.1 INDLEDNING	75
6.2 REGNEREGLER FOR SANDSYNLIGHED	75
6.3 SYMMETRISK SANDSYNLIGHEDSFELT	79
OPGAVER	81

7 VIGTIGE DISKRETE FORDELINGER

7.1 INDLEDNING	85
7.2 HYPERGEOMETRISK FORDELING	85
7.3 BINOMIALFORDELING	88
7.4 POISSONFORDELINGEN	93
7.5 DEN GENERALISEREDE HYPERGEOMETRISKE FORDELING	97
7.6 POLYNOMIALFORDELINGEN	97
7.7 APPROKSIMATIONER	98
OPGAVER	102

8 ANDRE KONTINUERTE FORDELINGER

8.1 INDLEDNING	112
8.2 DEN REKTANGULÆRE FORDELING	112
8.3 EKSPONENTIALFORDELINGEN	114
8.4 WEIBULLFORDELINGEN	117
8.5 DEN LOGARITMISKE FORDELING	118
8.6 DEN TODIMENSIONALE NORMALFORDELING	119
OPGAVER	120

9 FLERDIMENSIONAL STATISTISK VARIABEL

ESSENS	122
9.1 INDLEDNING	123
9.2 KOVARIANS OG KORRELATIONSKOEFFICIENT	127
9.3 LINEARKOMBINATION	130
OPGAVER	132

APPENDIX

2. STATISTISKE BEREGNINGER UDFØRT PÅ LOMMEREGER OG PC	135
TI-89	136
TI-83	138
Excel	142
MAPLE	144
4.1. OVERSIGT OVER KONFIDENSINTERVALLER	143
5.1. OVERSIGT OVER TEST AF MIDDELVÆRDI μ FOR NORMALFORDELT VARIABLE (σ KENDT)	144
5.2. OVERSIGT OVER TEST AF MIDDELVÆRDI μ FOR NORMALFORDELT VARIABLE (σ UKENDT)	145
5.3. OVERSIGT OVER TEST AF VARIANS σ^2 FOR NORMALFORDELT VARIABLE (μ KENDT)	146
5.4. OVERSIGT OVER TEST AF VARIANS σ^2 FOR NORMALFORDELT VARIABLE (μ UKENDT)	147
5.5. OVERSIGT OVER TEST AF PARAMETER p FOR BINOMIALFORDELT VARIABLE	148
5.6. OVERSIGT OVER TEST AF PARAMETER μ FOR POISSONFORDELT VARIABLE	149
7.1. OVERSIGT OVER APPROKSIMATIONER	150

STATISTISKE TABELLER

TABEL 1 FORDELINGSFUNKTIONEN FOR U-FORDELINGEN	152
TABEL 2. FRAKTILER I U-FORDELINGEN	154
TABEL 3. FRAKTILER I χ^2 - FORDELINGEN	155
TABEL 4. FRAKTILER I t - FORDELINGEN	157
TABEL 5. FRAKTILER I F - FORDELINGEN	158
TABEL 6. KUMULEREDE SANDSYNLIGHEDER I BINOMIALFORDELINGEN ...	159
TABEL 7. KUMULEREDE SANDSYNLIGHEDER I POISSONFORDELINGEN	164
TABEL 8. DIMENSIONERINGSTABEL	169

FACITLISTE	170
-------------------------	-----

STIKORD	174
----------------------	-----

1 STATISTISKE GRUNDBEGREBER

1.1 INDLEDNING

Ved næsten alle ingeniørmæssige problemer vil de indsamlede data udvise **variation**. Måler man således gentagne gange indholdet (i %) af et bestemt stof i et levnedsmiddel, vil det procentvis indhold ikke blive præcis samme tal for hver gang man foretager en måling. Dette kunne naturligvis være en usikkerhed ved målemetoden, men det vil sjældent være den væsentligste årsag.

Ved mange industrielle processer vil en række ukontrollable forhold indvirke på det endelige resultat. Eksempelvis vil udbyttet af en kemisk proces variere fra dag til dag, fordi man ikke har fuldstændig kontrol over **forsøgsbetingelser** som temperatur, omrørings tid, tidspunkt for tilsætning af råmaterialer, fugtighed osv. Endvidere er **forsøgsmaterialerne** muligvis ikke homogene nok. Råmaterialerne kan f.eks. være af varierende kvalitet, der må bruges forskelligt apparatur under produktionsprocessen, forskelligt personale deltager i arbejdet osv.

Statistik drejer sig om at samle, præsentere og analysere data med henblik på at foretage beslutninger og løse problemer.

1.2 HÆNDELSER OG SANDSYNLIGHED

Statistik bygger på sandsynlighedsteorien, som giver metoder til at finde, hvor stor chancen (sandsynligheden) er for at et bestemt resultat af et eksperiment forekommer.

DEFINITION af tilfældigt eksperiment. *Et eksperiment som kan resultere i forskellige udfald, selv om eksperimentet gentages på samme måde hver gang, kaldes et tilfældigt eksperiment (engelsk : random experiment).*

Det er karakteristisk for tilfældige eksperimenter, at man kan afgrænse en mængde kaldet eksperimentets **udfaldsrum** U , der indeholder de mulige **udfald**. Derimod kan man ikke forudsige, hvilket udfald der vil indtræffe ved udførelsen af eksperimentet. Består eksperimentet eksempelvis i kast med en terning er udfaldsrummet $U = \{1, 2, 3, 4, 5, 6\}$, men man kan ikke forudsige udfaldet af næste kast (eksperiment). Selv om man 4 gange i træk har fået udfaldet "øjental 1", kan man ikke forudsige, hvilket udfald der indtræffer næste gang. Resultatet af 5. kast afhænger ikke af resultaterne af de foregående 4 spil. Man siger, at eksperimenterne er "**statistisk uafhængige**" (en præcis

1. Statistiske grundbegreber

definition ses i kapitel 9).

Som eksempler på tilfældige eksperimenter kan nævnes:

- Ét kast med en mønt. Udfaldsrum $U = \{\text{Plat, Krone}\}$.
- Fremstilling af et parti levnedsmiddel og måling af det procentvise indhold af protein.
 $U =$ mængden af reelle tal fra 0 til 100.
- Udtage en stikprøve på 400 elektroniske komponenter af en dagsproduktion og optælling af antallet af defekte komponenter. $U = \{0, 1, 2, 3, 4, 5, \dots, 400\}$
- Udtagning af et tilfældigt TV-apparat fra en dagsproduktion af TV-apparater og optælling af antallet af loddefejl. $U =$ mængden af positive hele tal.

DEFINITION af hændelse. En hændelse er en delmængde af et eksperiments udfaldsrum.

Eksempelvis er A : "At få et lige øjental" en hændelse ved kast med en terning.

Hændelsen A siges at indtræffe, hvis et udfald fra A forekommer.

Sandsynlighedsbegrebet tager udgangspunkt i begrebet relativ hyppighed.

DEFINITION af relativ hyppighed. Gentages et eksperiment n gange, og forekommer hændelsen A netop n_A gange af de n gange, er A 's relative hyppighed $h(A) = \frac{n_A}{n}$

Lad eksempelvis eksperimentet være kast med en terning og hændelsen A være at få et lige øjental. Kastes terningen 100 gange og bliver resultatet et lige øjental 45 af de 100 gange er $h(A) = 0.45$.

Det er en erfaring, at øges antallet af gentagelser af eksperimentet, vil den relative hyppighed af hændelsen A stabilisere sig. Når n går mod ∞ , vil den relative hyppighed erfaringsmæssigt nærme sig til en grænseværdi ("de store tals lov"). Ved sandsynligheden for A som benævnes $P(A)$ forstås denne grænseværdi.

Eksempel 1.1. De relative hyppigheders stabilitet

Et eksperiment består i at kaste en terning, og hændelsen A består i at få et lige øjental. Terningen kastes nu 100 gange, og man får et lige øjental 55 gange. Eksperimentet udføres igen 100 gange, og man får A 47 gange. Igen kastes 100 gange, og man får nu A 57 gange. Til sidst kastes 100 gange, og man får A 40 gange.

Eksperimentet foretages nu i serier på 1000 gange, hvor man hver gang optæller antal gange A forekommer. Resultaterne vises i følgende tabel:

	Serier på 100 gentagelser				Serier med 1000 gentagelser			
	1	2	3	4	1	2	3	4
Antal gange A : et lige øjental	55	47	51	40	486	508	488	509
Relativ hyppighed	0.55	0.47	0.51	0.40	0.486	0.508	0.488	0.509

Det ses, at med 1000 gentagelser er de relative hyppigheder tættere samlet (ligger mellem 48,6% og 50,9%) end hvis man kun kastede 100 gange (mellem 40% og 55%). Hvis terningen var en ægte terning (fuldstændig homogen og symmetrisk), måtte man på forhånd forvente, at det tal, som de relative hyppigheder grupperede sig omkring, var tallet 0.5. Man vil derfor sige, at sandsynligheden for at få et lige øjental er 0.5, eller kort $P(A) = 0.5$.



Da definitionen af sandsynlighed bygger på relativ hyppighed, er det naturligt, at det for ethvert par af hændelser A og B i udfaldsrummet U skal gælde :

$0 \leq P(A) \leq 1$, $P(U) = 1$ og $P(A \cup B) = P(A) + P(B)$ forudsat A og B ingen elementer har fælles (er disjunkte).

De 3 regler kaldes sandsynlighedsregningens aksiomer. I kapitel 6 udledes på dette grundlag en række regler for regning med sandsynligheder.

1.3 Stokastisk VARIABEL

Ethvert statistisk problem må det på en eller anden måde være muligt at behandle talmæssigt.

Betragtes det tidligere eksempel med kast med en mønt, kunne man til udfaldet plat tilordne tallet 0 og til udfaldet krone tilordne tallet 1 og på den måde få problemet overført til noget, hvor man kan foretage beregninger. Man siger, man har indført en stokastisk (eller statistisk) variabel X , som er 0, når udfaldet er plat, og 1 når udfaldet er krone.

Generelt gælder følgende definition:

DEFINITION af stokastisk variabel (engelsk: random variable). *En stokastisk variabel (også kaldet statistisk variabel) er en funktion, som tilordner et reelt tal til hvert udfald i udfaldsrummet for et tilfældigt eksperiment.*

En stokastisk variabel betegnes med et stort bogstav såsom X , mens det tilsvarende lille bogstav x betegner en mulig værdi af X .

Er eksempelvis eksperimentet “udtagning af en kasse med 100 møtrikker, ud af en løbende produktion af kasser”, kunne den stokastiske variabel X være defineret som “antal defekte møtrikker i kassen”. Et andet eksempel kunne være eksperimentet “anvendelse af en ny metode til fremstilling

1. Statistiske grundbegreber

af et produkt”. Her kunne den stokastiske variabel Y være det målte procentvise udbytte ved forsøget.

Ved en **diskret** variabel forstås en variabel, hvis mulige værdier udgør en endelig eller tællelig mængde. I eksemplet hvor X er antal defekte møtrikker, er X en diskret variabel, da den kun kan antage heltallige værdier fra 0 til 100. Diskrete variable kaldes ofte for tællevARIABLE, da de ofte optræder ved optællinger, for eksempel som ovenfor i forbindelse med kvalitetskontrol.

Ved en **kontinuert** variabel forstås en variabel, hvis mulige værdier er alle reelle tal i et vist interval. I eksemplet, hvor Y er det målte procentvise udbytte, er Y en kontinuert variabel, da den kan antage alle værdier fra 0% til 100%.

1.4 DISKRET Stokastisk VARIABLE

Vi vil i dette afsnit benytte følgende eksempel til illustration af definitioner og begreber.

Eksempel 1.2. Diskret variabel.

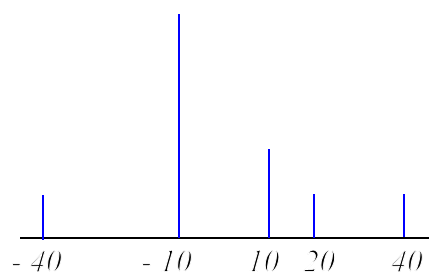
Ved et roulettespil vil resultatet “rød” give en gevinst på 40 kr. Tilsvarende vil “grøn” og “blå” give gevinster på henholdsvis 20 og 10 kr, mens “gul” og “sort” giver tab på henholdsvis 10 kr og 40 kr.

Lad X være gevinsten ved et spil. X får da værdierne 40, 20, 10, -10 og -40.

Ejeren af spillet ved, at $P(X = 40) = 0.1$, $P(X = 20) = 0.1$, $P(X = 10) = 0.2$, $P(X = -10) = 0.5$ og $P(X = -40) = 0.1$.

Man siger, at den diskrete variabel X har en **tæthedsfunktionen** $f(x) = P(X = x)$ bestemt ved

$$f(x) = \begin{cases} 0.1 & \text{for } x = -40 \\ 0.5 & \text{for } x = -10 \\ 0.2 & \text{for } x = 10 \\ 0.1 & \text{for } x = 20 \\ 0.1 & \text{for } x = 40 \\ 0 & \text{ellers} \end{cases}$$



Stolpediagram for tæthedsfunktion

En tæthedsfunktion illustreres grafisk ved at tegne et stolpediagram som vist på figuren.



Ejeren ønsker at finde ud af, hvor stor en gevinst pr. spil han i middel kan forvente. Hertil må middelværdien beregnes.

DEFINITION af middelværdi for diskret variabel. *Middelværdi for en diskret variabel X benævnes μ eller $E(X)$ og er defineret som $\mu = E(X) = \sum_x x \cdot P(X = x)$ (E står for *expected value*)*

I eksempel 1.2 fås middelværdien $\mu = -40 \cdot 0.1 + (-10) \cdot 0.5 + 10 \cdot 0.2 + 20 \cdot 0.1 + 40 \cdot 0.1 = -1$, dvs. i middel vil ejeren vinde 1 kr pr. spil.

Udover at angive en middelværdi, er man ofte også interesseret i at angive et mål for om værdierne ligger tæt omkring middelværdien, eller om de spredt sig meget (varierer meget). Som et mål herfor beregnes varians og spredning for variabelen X .

DEFINITION af varians og spredning for diskret variabel. *Variansen for en diskret variabel X benævnes σ^2 eller $V(X)$ og er defineret som*

$$\sigma^2 = V(X) = E((X - \mu)^2) = \sum_x (x - \mu)^2 \cdot P(X = x).$$

Spredningen (engelsk: standard deviation) for en diskret variabel X benævnes σ og er defineret som $\sigma = \sqrt{V(X)}$.

For fordelingen i eksempel 1.2, som havde middelværdien -1 fås således:

$$V(X) = (-40 - (-1))^2 \cdot 0.1 + (-10 - (-1))^2 \cdot 0.5 + (10 - (-1))^2 \cdot 0.2 + (20 - (-1))^2 \cdot 0.1 + (40 - (-1))^2 \cdot 0.1 = 429$$

$$\sigma(X) = \sqrt{429} = 20.71.$$

I afsnit 1.6 vises, at $V(X) = E(X^2) - \mu^2$ hvilket sædvanligvis giver en lettelse ved beregningerne.

For fordelingen i eksempel 1.2 fås således

$$V(X) = (-40)^2 \cdot 0.1 + (-10)^2 \cdot 0.5 + 10^2 \cdot 0.2 + 20^2 \cdot 0.1 + 40^2 \cdot 0.1 - (-1)^2 = 429$$

Det ses, at alle x -værdier i eksempel 1.2 ligger indenfor en afstand af $2 \cdot \sigma$ fra middelværdien -1.

For de fleste fordelinger vil mindst 99% af alle værdier ligge indenfor $[\mu - 3 \cdot \sigma; \mu + 3 \cdot \sigma]$, og sædvanligvis vil mindst 95% af værdierne ligge indenfor $[\mu - 2 \cdot \sigma; \mu + 2 \cdot \sigma]$.

1. Statistiske grundbegreber

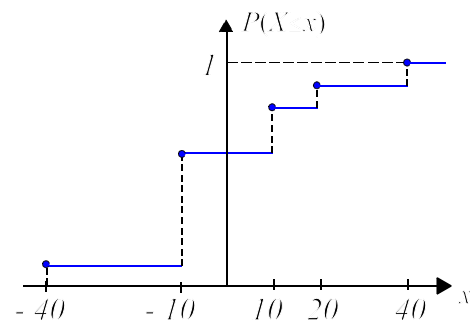
I visse situationer er det en fordel at betragte den stokastiske variabels fordelingsfunktion $F(x) = P(X \leq x)$.

DEFINITION af fordelingsfunktion $F(x)$ for diskret variabel. Fordelingsfunktionen for en diskret variabel X er defineret ved $F(x) = P(X \leq x)$.

Eksempel 1.2 (fortsat). Fordelingsfunktion for diskret variabel.

Fordelingsfunktionen for den i eksempel 1.2 angivne diskrete variabel er

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{for } x < -40 \\ 0.1 & \text{for } -40 \leq x < -10 \\ 0.6 & \text{for } -10 \leq x < 10 \\ 0.8 & \text{for } 10 \leq x < 20 \\ 0.9 & \text{for } 20 \leq x < 40 \\ 1 & \text{for } x \geq 40 \end{cases}$$



Trappekurve for fordelingsfunktion

Vi vil i kapitel 7 angive nogle vigtige diskrete fordelinger.

1.5 KONTINUERT Stokastisk VARIABEL

Vi vil i dette afsnit benytte det følgende eksempel til illustration.

Eksempel 1.3. Kontinuert stokastisk variabel

I menneskers led udskiller den inderste hinde en "ledvæske" som "smører" leddet. For visse ledsygdomme kan brintionkoncentrationen (pH) i denne væske tænkes at have betydning. Som led i en nordisk medicinsk undersøgelse af en bestemt ledsygdom udtog man blandt samtlige patienter der led af denne sygdom tilfældigt 75 patienter og målte pH i ledvæsken i knæet. Resultaterne var følgende:

7.02 7.26 7.31 7.16 7.45 7.32 7.21 7.35 7.25 7.24 7.20 7.21 7.27 7.28 7.19
7.39 7.40 7.33 7.32 7.35 7.34 7.41 7.28 7.27 7.28 7.33 7.20 7.15 7.42 7.35
7.38 7.32 7.71 7.34 7.10 7.35 7.15 7.19 7.44 7.12 7.22 7.12 7.37 7.51 7.19
7.30 7.24 7.36 7.09 7.32 6.95 7.35 7.36 7.52 7.29 7.31 7.35 7.40 7.23 7.16
7.26 7.47 7.61 7.23 7.26 7.37 7.16 7.43 7.08 7.56 7.07 7.08 7.17 7.29 7.20

Population og stikprøve. Samtlige indbyggere i Norden med denne sygdom udgør populationen. Da det er ganske uoverkommeligt at undersøge alle, udtages en **stikprøve** på 75 patienter. Det er målet ved hjælp af statistiske metoder på basis af en stikprøve at sige noget generelt om populationen.

Histogram. For at få et overblik over et større datamateriale, vil man sædvanligvis starte med at tegne et histogram. Først findes det største tal x_{max} og det mindste tal x_{min} i materialet og derefter beregne **variationsbredden** $x_{max} - x_{min}$. Vi ser, at største tal er 7.71 og mindste tal er 6.95 og variationsbredden derfor $7.71 - 6.95 = 0.76$.

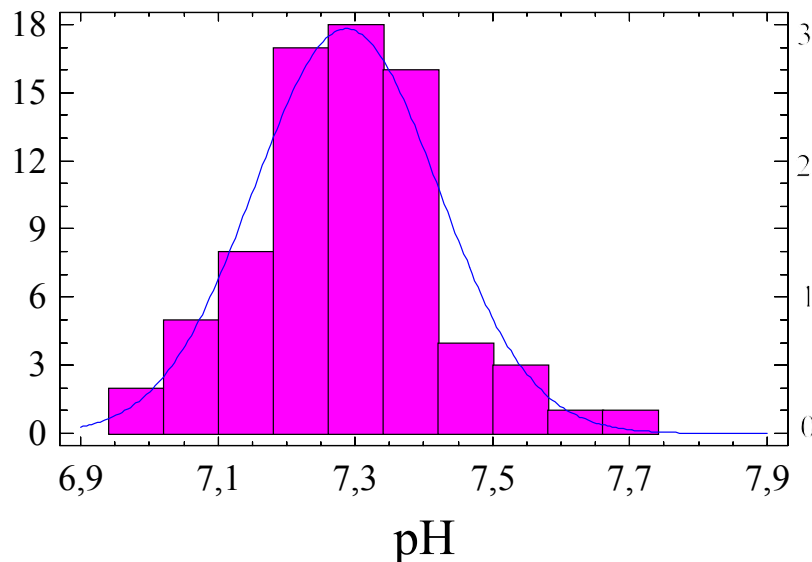
Dernæst deles tallene op i et passende antal intervaller (klasser). Som det første bud vælges ofte et antal nær $\sqrt{n}$. Da $\sqrt{75} \approx 9$ vælges ca. 9 klasser. Da $\frac{0.76}{9} \approx 0.08$ deler vi op i de klasser, der ses af tabellen. Dette giver 10 intervaller. Vi tæller op hvor mange tal der ligger i hvert interval (gøres nemmest ved at starte forfra og sæt en streg i det interval som tallet tilhører). I skemaet er dette sket, og endvidere er der af hensyn til det følgende foretaget en skalering ved at dividere den relative hyppighed med intervallængden.

Klasser		Antal n	Relativ hyppighed $\frac{n}{75}$	Skalering $\frac{n}{75 \cdot 0.08}$
]6.94 - 7.02]	//	2	0.0267	0.3333
]7.02 - 7.10]	////	5	0.0667	0.8333
]7.10 - 7.18]	////////	8	0.1067	1.3333
]7.18 - 7.26]	////////////////	17	0.2267	2.8333
]7.26 - 7.34]	////////////////	18	0.2400	3.0000
]7.34 - 7.42]	////////////////	16	0.2133	2.6667
]7.42 - 7.50]	////	4	0.0533	0.6667
]7.50 - 7.58]	///	3	0.0400	0.5000
]7.58 - 7.66]	/	1	0.0133	0.1667
]7.66 - 7.74]	/	1	0.0133	0.1667

Vi får det på næste side tegnede histogram (tegnet af Statgraphics).

Dette viser et "klokkeformet histogram", hvor der er flest tal fra 7.19 til 7.42, og derefter falder antallet til begge sider.

Histogram for pH



Man regner normalt med, at resultaterne af forsøg hvor man har foretaget målinger (hvis man lavede nok af dem) har et sådant klokkeformet histogram. Hvis man tænker sig antallet af forsøg stiger (for eksempel undersøger hele populationen på måske 1 million nordiske knæ), samtidig med at man øger antallet af klasser tilsvarende (til for eksempel $\sqrt{10^6} \approx 1000$), vil histogrammet blive mere og mere fintakket, og til sidst nærme sig til en kontinuert klokkeformet kurve (indtegnet på grafen). Hvis man benytter den skalerede skala fra skemaet, som også er afsat på højre side af tegningen, vil arealet af hver søjle være den relative hyppighed, og for den idealiserede kontinuerte kurve, vil arealet under kurven i et bestemt interval fra a til b være sandsynligheden for at få en værdi mellem a og b . Det samlede areal under kurven er naturligvis 1. Man siger, at den kontinuerte stokastiske variabel X (pH værdien) har en **tæthedsfunktion $f(x)$** hvis graf er den ovenfor nævnte kontinuerte kurve.

DEFINITION af tæthedsfunktion $f(x)$ for kontinuert stokastisk variabel. En tæthedsfunktion $f(x)$ skal tilfredsstille følgende betingelser:

1. $f(x) \geq 0$ for alle x
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P(a \leq X \leq b) = \int_a^b f(x)dx$ for ethvert interval af reelle tal $[a; b]$

Bemærk, at for kontinuerte variable er

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b).$$

Et eksempel på en tæthedsfunktion for en kontinuert variabel er den i næste kapitel beskrevne normalfordeling. Måleresultater vil sædvanligvis være værdier af normalfordelte variable, så en rimelig hypotese for den i eksempel 1.3 angivne kontinuerte stokastiske variabel $X = \text{pH}$ er således, at den er normalfordelt. Dette bestyrkes af at grafen for sådanne netop er klokkeformede. Som i det diskrete tilfælde, er det væsentlig at finde en central værdi i populationen.

DEFINITION af middelværdi for kontinuert variabel. *Middelværdi for en kontinuert variabel X med tæthedsfunktion $f(x)$ benævnes μ eller $E(X)$ og er defineret som*

$$\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

Som for den diskrete variabel, er det ikke nok at angive en middelværdi for populationen. Man må også angive et mål for hvor meget tallene spredt sig.

DEFINITION af varians og spredning for kontinuert variabel. *Variansen for en kontinuert variabel X med tæthedsfunktion $f(x)$ benævnes σ^2 eller $V(X)$ og er defineret som*

$$\sigma^2 = V(X) = E((X - \mu)^2) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx$$

Spredningen (engelsk: standard deviation) for en diskret variabel X med tæthedsfunktion $f(x)$ benævnes σ og er defineret som $\sigma = \sqrt{V(X)}$

I afsnit 1.6 vises, at $V(X) = E(X^2) - \mu^2$ hvilket sædvanligvis giver en lettelse ved beregningerne.

Eksempel 1.4 Kontinuert stokastisk variabel.

Lad der være givet følgende funktion: $f(x) = \begin{cases} \frac{3}{8} \cdot x^2 & \text{for } 0 \leq x < 2 \\ 0 & \text{ellers} \end{cases}$.

- 1) Vis, at $f(x)$ opfylder de to første betingelser i definitionen på en tæthedsfunktion. I det følgende antages, at $f(x)$ er tæthedsfunktion for en kontinuert stokastisk variabel X .
- 2) Skitser grafen for f .
- 3) Beregn middelværdi og spredning for X .

1. Statistiske grundbegreber

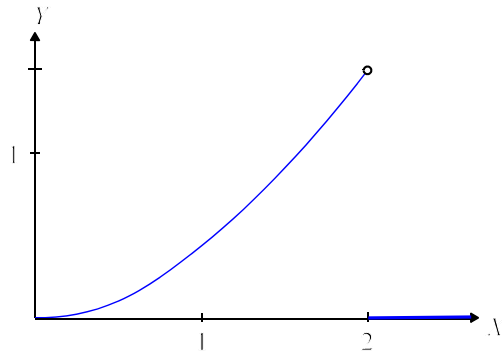
LØSNING:

Spørgsmål 1:

Det ses umiddelbart,

a) at $f(x) \geq 0$ for alle x .

$$b) \int_{-\infty}^{\infty} f(x) dx = \int_0^2 \frac{3}{8} x^2 dx = \left[\frac{x^3}{8} \right]_0^2 = 1.$$



Spørgsmål 2:

Grafen, som er en del af en parabel, ses på fig 1.1.

Spørgsmål 3:

Fig.1.1 Tæthedsfunktion

$$\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_0^2 x \cdot \frac{3}{8} x^2 dx = \left[3 \frac{x^4}{32} \right]_0^2 = \underline{\underline{\frac{3}{2}}}.$$

$$V(X) = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx - \mu^2 = \int_0^2 x^2 \cdot \frac{3}{8} x^2 dx - \left(\frac{3}{2} \right)^2 = \left[3 \frac{x^5}{40} \right]_0^2 - 2.25 = \underline{\underline{0.15}}.$$

$$\sigma(X) = \sqrt{0.15} = \underline{\underline{0.387}}.$$



Stikprøves gennemsnit og spredning.

Sædvanligvis kendes populationen ikke (det er ganske uoverkommeligt at undersøge for eksempel en million dårlige knæ). I stedet udtages en stikprøve på for eksempel 75 patienter, og man søger ud fra disse at give **estimer** (skøn) for middelværdi og spredning.

DEFINITION af en stikprøves gennemsnit, varians og spredning.

Hvis observationerne i en stikprøve af størrelsen n er $x_1, x_2, \dots, x_n$,

så er et estimat for middelværdien μ gennemsnittet $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$

og et estimat for variansen σ^2 er stikprøvens varians $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$.

Ved stikprøvens **spredning** forstås s .

Stikprøvens varians siges at have $f = n - 1$ **frihedsgrader**.

Bemærk, at man skelner mellem populationens middelværdi μ og stikprøvens middelværdi $\bar{x}$ (som kaldes gennemsnittet og læses som x strek), ligesom man skelner mellem populationens spredning σ og stikprøvens spredning s . $\bar{x}$ og s er estimater for henholdsvis μ og σ (græske bogstaver for populationens parametre).

En begrundelse for ovennævnte formler kunne være, at hvis man kender **hele populationen**, f.eks. antager at der kun findes n patienter med ovennævnte ledsygdom, og man har målt pH på dem alle, så vil der være en sandsynlighed på $\frac{1}{n}$ for at udvælge en bestemt patient. Er de målte pH værdier $x_1, x_2, \dots, x_n$ fås

$$E(X) = \sum_{i=1}^n x_i \cdot f(x) = \frac{\sum_{i=1}^n x_i}{n} = \bar{x} \quad \text{og} \quad V(X) = \sum_{i=1}^n (x_i - \mu)^2 \cdot f(x_i) = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

Eksempel 1.5. Beregning af $\bar{x}$ og s .

Lad os for en forsøgsvariabel X have udført en stikprøve på 4 forsøg. Resultaterne var:

35.9, 33.3, 34.7, 34.1 .

Beregn et estimat for X 's middelværdi μ og spredning σ

LØSNING:

Et estimat for μ er $\bar{x} = \frac{x_1 + x_2 + x_3 + x_4}{4} = \frac{35.9 + 33.3 + 34.7 + 34.1}{4} = \underline{\underline{34.5}}.$

Et estimat for σ^2 er

$$s^2 = \frac{(35.9 - 34.5)^2 + (33.3 - 34.5)^2 + (34.7 - 34.5)^2 + (34.1 - 34.5)^2}{4 - 1} = \frac{3.60}{3} = 1.2.$$

Et estimat for σ^2 er $s = \sqrt{1.2} = \underline{\underline{1.1}}.$

TI 89: MATH\Statistics\mean({35.9,33.3,34.7,34.1}) Resultat: 34.5

MATH\Statistics\stdDev({35.9,33.3,34.7,34.1}) Resultat: 1.095



Anskuelig forklaring på formlen for s .

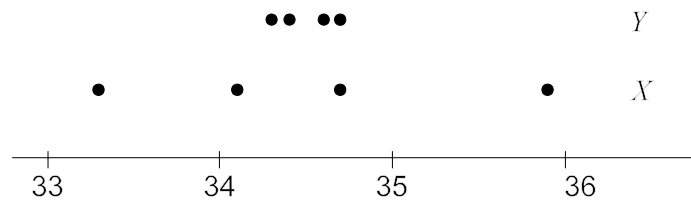
At formlen for s skulle være særlig velegnet til at angive, hvor meget resultaterne "spredt sig" (hvor megen støj der er på forsøget) er ikke umiddelbart indlysende. I det følgende gives en anskuelig forklaring.

Lad os betragte 2 forsøgsvariable X og Y , hvorpå der for hver er udført en stikprøve på 4 forsøg.

Resultaterne var: X : 35.9, 33.3, 34.7, 34.1 med gennemsnittet $\bar{x} = 34.5$, og

Y : 34.3, 34.6, 34.7, 34.4 med gennemsnittet $\bar{y} = 34.5$.

1. Statistiske grundbegreber



De to forsøgsvariable har samme gennemsnit, men det er klart, at Y -resultaterne grupperer sig meget tættere om gennemsnittet end X -resultaterne, dvs. Y -stikprøven har mindre spredning (der er mindre støj på Y - forsøget) end X -stikprøven.

For at få et mål for stikprøvens spredning beregnes resultaternes afvigelser fra gennemsnittet.

$x_i - \bar{x}$	$y_i - \bar{y}$
$35.9 - 34.5 = 1.4$	$34.3 - 34.5 = -0.2$
$33.3 - 34.5 = -1.2$	$34.6 - 34.5 = 0.1$
$34.7 - 34.5 = 0.2$	$34.7 - 34.5 = 0.2$
$34.1 - 34.5 = -0.4$	$34.4 - 34.5 = -0.1$

Summen af disse afvigelser er naturligvis altid 0 og kan derfor ikke bruges som et mål for stikprøvens spredning.

I stedet betragtes summen af kvadraterne på afvigelserne (forkortet SS: **S**um of **S**quares eller SAK: **S**um af **a**fvigelsernes **K**vadrat).

$$SAK_x = \sum_{i=1}^n (x_i - \bar{x})^2 = 1.4^2 + (-1.2)^2 + 0.2^2 + (-0.4)^2 = 3.60$$

$$SAK_y = \sum_{i=1}^n (y_i - \bar{y})^2 = (-0.2)^2 + 0.1^2 + 0.2^2 + (-0.1)^2 = 0.10$$

Da et mål for variansen ikke må være afhængig af antallet af forsøg, divideres med $n - 1$.

Umiddelbart ville det være mere rimeligt at dividere med n . Imidlertid kan det vises (se eventuelt supplement 4A), at $E(s^2) = \sigma^2$, mens $E\left(\frac{SAK}{n}\right) = \frac{n-1}{n}\sigma^2$. I middel bliver et skøn for variansen altså for lille, hvis man dividerer med n , mens den "rammer" præcist, hvis man dividerer med $n - 1$. Det kan forklares ved, at tallene x_i har en tendens til at ligge tættere ved deres gennemsnit $\bar{x}$ end ved middelværdien μ .

Stikprøvens spredning s (engelsk: standard deviation) fås nu ved at tage kvadratroden af resultatet (for at opnå samme enhed, da vi jo tidligere har kvadreret). Vi har altså

$$s_x = \sqrt{\frac{3.60}{4-1}} = 1.095 \text{ og } s_y = \sqrt{\frac{0.10}{4-1}} = 0.183$$

Som vi forudså, er stikprøvens spredning betydelig større for X -resultaterne end for Y -resultaterne. Man siger, at stikprøvens varians er baseret på $f = n - 1$ **frihedsgrader**. Navnet skyldes, at kun $n - 1$ af de n led $x_i - \bar{x}$ kan vælges frit, idet summen af de n led er nul. Eksempelvis ser vi af ovenstående eksempel, at der er 3 frihedsgrader, da kendskab til de første 3 led på 1.4, -1.2 og 0.2 er nok til at bestemme det fjerde led, da summen er nul.

Vurdering af størrelsen af stikprøvens spredning. De 75 pH værdier i eksempel 1.3 har et gennemsnit på 7.287 og stikprøvens spredning $s = 0.134$.

For at få en anskuelig fornemmelse af størrelsen af s (er det et stort eller lille tal?) foretager vi en optælling og finder, at

mellem $\bar{x} - s = 7.16$ og $\bar{x} + s = 7.42$ ligger ca. 53 tal eller 70% af resultaterne.

mellem $\bar{x} - 2 \cdot s = 7.03$ og $\bar{x} + 2 \cdot s = 7.55$ ligger ca. 69 tal eller 92% af resultaterne.

mellem $\bar{x} - 3 \cdot s = 6.90$ og $\bar{x} + 3 \cdot s = 7.68$ ligger ca. 74 tal eller 99% af resultaterne.

Dette stemmer meget godt overens med forestillingen om at resultaterne er normalfordelte, idet det for disse (som vi vil se i næste kapitel) gælder det på figurerne 1.2 og 1.3 nævnte.

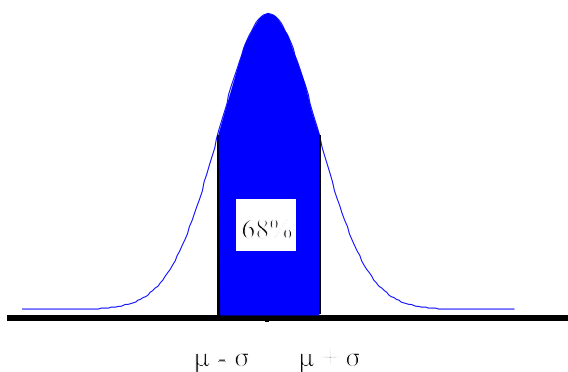


Fig. 1.2 Mellem $\mu - \sigma$ og $\mu + \sigma$ ligger ca. 68% af resultaterne.

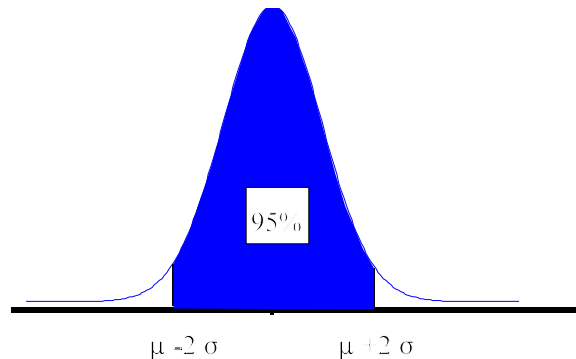


Fig 1.3. Mellem $\mu - 2 \cdot \sigma$ og $\mu + 2 \cdot \sigma$ ligger ca. 95% af resultaterne.

Endvidere gælder, at mellem $\mu - 3 \cdot \sigma$ og $\mu + 3 \cdot \sigma$ ligger ca. 99.7% af resultaterne. Dette benyttes bl.a. i statistisk kvalitetskontrol, hvor man løbende udtager stikprøver af produktionen. Eksempelvis kan man om den måling, der giver en pH-værdi i ledvæsken på 7.71 sige, at enten er der en fejl ved målingen eller også er der her et interessant medicinsk særtilfælde.

1. Statistiske grundbegreber

Fordelingsfunktion. I visse situationer er det en fordel at betragte den stokastiske variabels fordelingsfunktion $F(x)$

DEFINITION af fordelingsfunktion $F(x)$ for kontinuert variabel.

Fordelingsfunktionen for en kontinuert variabel X er defineret ved

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x) dx$$

DEFINITION af p-fraktil x_p . Lad p være et vilkårligt tal mellem 0 og 1. Ved p -fraktilen eller $100 p$ % fraktilen forstås det tal x_p , for hvilket det gælder, at

$$F(x_p) = P(X \leq x_p) = p \quad \left(= \int_0^{x_p} f(x) dx \right)$$

Særlig ofte benyttede fraktiler er 50% fraktilen, som kaldes **medianen** (eller 2. kvartil), 25 % fraktilen, som kaldes **nedre kvartil** (eller 1. kvartil) og 75% fraktilen, som kaldes **øvre kvartil** (eller 3. kvartil).

Eksempel 1.6 Fordelingsfunktion for kontinuert variabel.

For den i eksempel 1.4 angivne kontinuerte variabel X med tæthedsfunktion $f(x)$ ønskes fundet:

- 1) Fordelingsfunktionen $F(x)$.
- 2) Medianen .

LØSNING:

$$1) \quad F(x) = \int_{-\infty}^x f(x) dx = \begin{cases} \int_{-\infty}^x 0 dx = & 0 & \text{for } x < 0 \\ 0 + \int_0^x \frac{3}{8} x^2 dx = \left[\frac{x^3}{8} \right]_0^x = \frac{x^3}{8} & \text{for } 0 \leq x \leq 2 \\ 0 + \frac{2^3}{8} + \int_2^x 0 dx = & 1 & \text{for } x > 2 \end{cases} .$$

$$2) \text{ Medianen er bestemt ved } F(x) = 0.5 \Leftrightarrow \frac{x^3}{8} = 0.5 \Leftrightarrow x^3 = 4 \Leftrightarrow \underline{\underline{x = 1.59}} .$$



Anvendes den kontinuerte tæthedsfunktion, som angav pH i patienter med en bestemt ledsygdom (se eksempel 1.3), så fås den i fig. 1.4 angivne fordelingsfunktion (tegnet af Statgraphics). På grafen for fordelingsfunktionen kan aflæses, at for $p = 25\%$ er x_p ca. 7.2, dvs. 25% af alle med den pågældende ledsygdom har et pH i knæet på under 7.2.

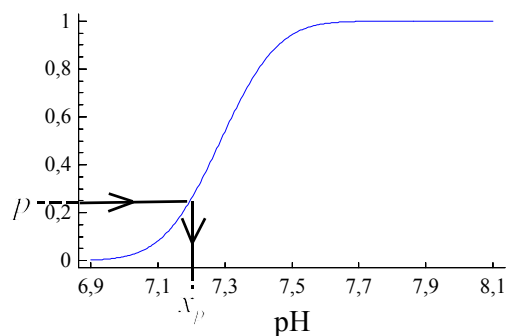


Fig 1.4. Fordelingsfunktion

Sumpolygon: Baseret på en stikprøve bliver en tilnærmelse til fordelingsfunktionen en sumpolygon. Baseret på de 75 pH værdier i eksempel 1.3 foretages den i den følgende tabel angivne opsummering(kumulering).

Afsættes punkterne (7.02 , 0.0267), (7.10, 0.0933) . . . (7.74, 1.00) (bemærk at x-værdierne er værdierne i højre intervalendepunkt), og forbindes de enkelte punkter med rette linier, fås den i figur 1.5 angivne sumpolygon, hvoraf man analogt med før kan aflæse, at 25% fraktilen er ca. 7.2.

Klasser	Antal	Opsum- mering	Kumuleret relativ hyppighed i %
]6.94 - 7.02]	2	2	2.67
]7.02 - 7.10]	5	7	9.33
]7.10 - 7.18]	8	15	20.00
]7.18 - 7.26]	17	32	42.67
]7.26 - 7.34]	18	50	66.67
]7.34 - 7.42]	16	66	88.00
]7.42 - 7.50]	4	70	93.33
]7.50 - 7.58]	3	73	97.33
]7.58 - 7.66]	1	74	98.67
]7.66 - 7.74]	1	75	100.00

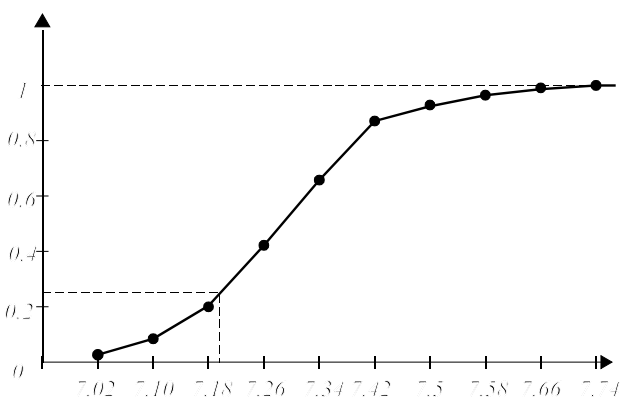


Fig 1.5. Sumpolygon

1.6 LINEARKOMBINATION AF STOKASTISKE VARIABLE

Vi betragter i dette afsnit flere stokastiske variable. Begreberne bliver behandlet mere udførligt i kapitel 9. mens vi her kun vil se på et par af de vigtigste begreber. Eksempel 1.7 vil blive benyttet som gennemgående eksempel

Eksempel 1.7. To variable.

Insektpulver sælges i papkartoner. Lad den stokastiske variable X_1 være vægten af pulveret, mens X_2 er vægten af papkartonen. I middel fyldes der 500 gram insektpulver i hver karton med en spredning på 5 gram. Kartonen vejer i middel 10 gram med en spredning på 1.0 gram.

$Y = X_1 + X_2$ er da bruttovægten.

1) Find middelværdien af Y

2) Find spredning af Y .



Mere generelt haves:

Lad $X_1, X_2, \dots, X_n$ være n stokastiske variable.

Ved en **linearkombination** af disse forstås

$$Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_n \cdot X_n, \text{ hvor } a_0, a_1, a_2, \dots, a_n \text{ er konstanter.}$$

Som vist i kapitel 9 gælder følgende :

Linearitetsregel: $E(Y) = a_0 + a_1 \cdot E(X_1) + a_2 \cdot E(X_2) + \dots + a_n \cdot E(X_n)$.

Svaret på spørgsmål 1 i eksempel 1.7 er derfor $E(Y) = E(X_1) + E(X_2) = 500 + 10 = \underline{\underline{510 \text{ gram}}}$.

I eksempel 1.7 synes det rimeligt at antage, at vægten af pulveret og vægten af papkartonen er **uafhængige** (påfyldningen kan tænkes at ske maskinelt, uden at den er afhængig på nogen måde af hvilken vægt, kartonen tilfældigvis har).

Generelt defineres statistisk uafhængighed på følgende måde:

DEFINITION af statistisk uafhængighed. De stokastiske variable $X_1, X_2, \dots, X_n$ kaldes **statistisk uafhængige**, såfremt de for alle værdier af $x_1, x_2, \dots, x_n$ opfylder betingelsen ($\wedge$ betyder både .. og)

$$P(X_1 \leq x_1 \wedge X_2 \leq x_2 \wedge \dots \wedge X_n \leq x_n) = P(X_1 \leq x_1) \cdot P(X_2 \leq x_2) \cdot \dots \cdot P(X_n \leq x_n)$$

Til bestemmelse af $V(Y)$ vises i kapitel 9 den såkaldte kvadratrege, som gælder generelt, dvs. der er ingen krav om, at de variable er statistisk uafhængige.

En variant af denne som kun gælder, hvis $X_1, X_2, \dots, X_n$ er statistisk uafhængige, er følgende

Kvadratregel for statistisk uafhængige variable:

$$V(Y) = a_1^2 \cdot V(X_1) + a_2^2 \cdot V(X_2) + \dots + a_n^2 \cdot V(X_n).$$

Svaret på spørgsmål 2 i eksempel 1.7 er derfor

$$V(Y) = V(X_1) + V(X_2) = 5^2 + 1^2 = 26. \quad \sigma(Y) = \sqrt{26} = \underline{\underline{5.1}} \text{ gram.}$$

Stikprøve $\vec{X} = (X_1, X_2, \dots, X_n)$. Lad os antage, at vi uafhængigt af hinanden og under de samme betingelser udtager n elementer fra en population med middelværdi μ og spredning σ . Lad X_1 være den stokastiske variabel, der er resultatet af første udtagning af et element i stikprøven, X_2 være den stokastiske variabel, der er resultatet af anden udtagning, o.s.v. $X_1, X_2, \dots, X_n$ vil da være **ensfordelte** uafhængige stokastiske variable, d.v.s. have samme fordeling med middelværdi μ og spredning σ .

Man kalder talsættet $\vec{X} = (X_1, X_2, \dots, X_n)$ for en stikprøve af størrelsen n .

For ensfordelte uafhængige stokastiske variable gælder:

SÆTNING 1.1 (*middelværdi og spredning for stikprøves gennemsnit*)

$$E(\bar{X}) = \mu \text{ og } \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}, \text{ hvor } \bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Bevis: Af linearitetsreglen fås $E(\bar{X}) = E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n}(E(X_1) + E(X_2) + \dots + E(X_n)) = \mu$

Af kvadratreglen fås $V(\bar{X}) = V\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n^2}(V(X_1) + V(X_2) + \dots + V(X_n)) = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}$.



Eksempel 1.7 fortsat

Hvis der udtages 5 kartoner insektpulver, hvad vil da være spredningen på gennemsnittet af vægten af insektpulveret.

LØSNING:

Da spredningen på 1 karton er 5.1 gram, vil spredningen på gennemsnittet af 5 kartoner være

$$\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{5.1}{\sqrt{5}} = \underline{\underline{2.28}}$$



Vi vil her til sidst bevise den i afsnit 1.4 og 1.5 angivne påstand: $V(X) = E(X^2) - \mu^2$.

Bevis: For såvel diskrete som kontinuerte variable er variansen defineret ved $\sigma^2 = V(X) = E((X - \mu)^2)$. Af linearitetsreglen fås da

$$\begin{aligned} V(X) &= E((X - \mu)^2) = E(X^2 - 2 \cdot X \cdot \mu + \mu^2) = E(X^2) - 2 \cdot \mu \cdot E(X) + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$



1.7 USIKKERHEDSBEREGNING (skrevet af Bjarne Hellesten)

Måles trykket P , volumenet V og temperaturen T af en ideal gas, optræder der tilfældige målefejl, som gør værdierne usikre. Beregnes molantallet n ud fra ligningen $P \cdot V = n \cdot R \cdot T$, bliver værdien af n derfor også usikker. Vi ønsker at kunne beregne usikkerheden på n ud fra usikkerhederne på P , V og T .

DEFINITION af absolut og relativ usikkerhed.

Absolutte usikkerheder $\sigma(X_1), \sigma(X_2), \dots, \sigma(X_k)$ er spredningerne hidrørende fra tilfældige målefejl.

Relative usikkerheder: $rel(X_1) \equiv \frac{\sigma(X_1)}{E(X_1)}, rel(X_2) \equiv \frac{\sigma(X_2)}{E(X_2)}, \dots, rel(X_k) \equiv \frac{\sigma(X_k)}{E(X_k)}$.

SÆTNING 1.2 (ophobningsloven for usikkerheder)

Lad $X_1, X_2, \dots, X_k$ være **stokastisk uafhængige** variable, som hidrører fra målinger.

Beregnes en ny størrelse $Y(X_1, X_2, \dots, X_k)$ ud fra $X_1, X_2, \dots, X_k$, får Y også en usikkerhed $\sigma(Y)$, som kan beregnes tilnærmet af

$$\sigma(Y) \approx \sqrt{\left(\frac{\partial Y}{\partial X_1}\right)^2 \cdot V(X_1) + \left(\frac{\partial Y}{\partial X_2}\right)^2 \cdot V(X_2) + \dots + \left(\frac{\partial Y}{\partial X_k}\right)^2 \cdot V(X_k)}$$

Ved den praktiske brug af formlerne for relative fejl og den ophobede fejl kan man erstatte alle middelværdier med de faktisk målte værdier.

Bevis for ophobningsloven, Funktionen $Y = Y(\vec{X})$ tilnærmes ved sit 1. Taylorpolynomium med udviklingspunkt i middelværdien $\vec{\mu}$:

$$Y(\vec{X}) \approx Y(\vec{\mu}) + \frac{\partial Y}{\partial X_1} \cdot (X_1 - \mu_1) + \frac{\partial Y}{\partial X_2} \cdot (X_2 - \mu_2) + \dots + \frac{\partial Y}{\partial X_k} \cdot (X_k - \mu_k),$$

hvor de partielle afledede er beregnet i middelværdien $\vec{\mu}$. Vi finder derfor

$$Y(\vec{X}) \approx \text{konstant} + \frac{\partial Y}{\partial X_1} \cdot X_1 + \frac{\partial Y}{\partial X_2} \cdot X_2 + \dots + \frac{\partial Y}{\partial X_k} \cdot X_k,$$

$$V(Y) \approx V\left(\text{konstant} + \frac{\partial Y}{\partial X_1} \cdot X_1 + \frac{\partial Y}{\partial X_2} \cdot X_2 + \dots + \frac{\partial Y}{\partial X_k} \cdot X_k\right)$$

Da $X_1, X_2, \dots, X_k$ er forudsat statistisk uafhængige, får vi ifølge kvadratreglen

$$V(Y) \approx \left(\frac{\partial Y}{\partial X_1}\right)^2 \cdot V(X_1) + \left(\frac{\partial Y}{\partial X_2}\right)^2 \cdot V(X_2) + \dots + \left(\frac{\partial Y}{\partial X_k}\right)^2 \cdot V(X_k)$$

$$\sigma(Y) \approx \sqrt{\left(\frac{\partial Y}{\partial X_1}\right)^2 \cdot V(X_1) + \left(\frac{\partial Y}{\partial X_2}\right)^2 \cdot V(X_2) + \dots + \left(\frac{\partial Y}{\partial X_k}\right)^2 \cdot V(X_k)}.$$



Eksempel 1.8. Usikkerhedsberegning.

En ideal gas opfylder ligningen $P \cdot V = n \cdot R \cdot T$, hvor $R = 8.314 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$.

Man har målt $P = 123400 \text{ Pa}$, $V = 5.67 \text{ m}^3$, $T = 678 \text{ K}$

med usikkerheder $\sigma(P) = 1000 \text{ Pa}$, $\sigma(V) = 0.06 \text{ m}^3$, $\sigma(T) = 3 \text{ K}$.

Det kan antages, at måleresultaterne for P , V og T er statistisk uafhængige.

Find molantallet n , usikkerheden $\sigma(n)$, samt den relative usikkerhed $\text{rel}(n)$.

LØSNING:

$$\text{Vi finder } n = \frac{P \cdot V}{R \cdot T} = \frac{123400 \cdot 5.67}{8.314 \cdot 678} = \underline{\underline{104.12}} \text{ mol}$$

Af ophobningsloven for usikkerheder fås usikkerheden på n :

$$\sigma(n) = \sigma(n(P, V, T)) = \sqrt{\left(\frac{\partial n}{\partial P}\right)^2 \cdot \sigma^2(P) + \left(\frac{\partial n}{\partial V}\right)^2 \cdot \sigma^2(V) + \left(\frac{\partial n}{\partial T}\right)^2 \cdot \sigma^2(T)}$$

Statistiske grundbegreber

$$\begin{aligned} &= \sqrt{\left(\frac{V}{R \cdot T}\right)^2 \cdot \sigma^2(P) + \left(\frac{P}{R \cdot T}\right)^2 \cdot \sigma^2(V) + \left(\frac{-P \cdot V}{R \cdot T^2}\right)^2 \cdot \sigma^2(T)} \\ &= \sqrt{\left(\frac{5.67}{8.314 \cdot 678}\right)^2 \cdot 1000^2 + \left(\frac{123400}{8.314 \cdot 678}\right)^2 \cdot 0.06^2 + \left(\frac{-123400 \cdot 5.67}{8.314 \cdot 678^2}\right)^2 \cdot 3^2} \\ &= \sqrt{1.01178 + 1.72526 + 0.30165} = 1.74318 = \underline{\underline{1.74 \text{ mol}}} , \end{aligned}$$

og dermed den relative usikkerhed $rel(n) = \frac{\sigma(n)}{n} = \frac{1.74318}{124.12} = 0.0140443 \approx \underline{\underline{1.40\%}}$.



OPGAVER

Opgave 1.1 (tæthedsfunktion)

Givet følgende to funktioner:

x	-3	-2	-1	3	6	ellers
$f(x)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	0
$g(x)$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	0

- 1) Hvilke af de to funktioner kan opfattes som en tæthedsfunktion for en stokastisk variabel X .
- 2) For den i spørgsmål 1 fundne tæthedsfunktion, skal man finde den tilsvarende fordelingsfunktion, og tegne henholdsvis stolpediagram og sumkurve.
- 3) Beregn middelværdien $E(X)$, variansen $V(X)$ og spredningen $\sigma(X)$.

Opgave 1.2 (tæthedsfunktion)

Givet følgende funktion:

x	-3	-2	0	1	ellers
$f(x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{3}$	0

- 1) Idet $f(x)$ opfattes som en tæthedsfunktion for en stokastisk variabel X skal man tegne det tilhørende stolpediagram.
- 2) Find den tilsvarende fordelingsfunktion $F(x)$ og tegn grafen.
- 3) Beregn middelværdien $E(X)$, variansen $V(X)$ og spredningen $\sigma(X)$.

Opgave 1.3 (fordelingsfunktion)

Givet følgende fordelingsfunktion $F(x)$

x	$x < -5$	$-5 \leq x < -3$	$-3 \leq x < -1$	$-1 \leq x < 1$	$1 \leq x < 3$	$3 \leq x$
$F(x)$	0	$\frac{1}{12}$	$\frac{4}{12}$	$\frac{8}{12}$	$\frac{11}{12}$	1

- 1) Tegn grafen for $F(x)$.
- 2) Find den tilsvarende tæthedsfunktion $f(x)$ og tegn grafen.
- 3) Beregn middelværdien $E(X)$, variansen $V(X)$ og spredningen $\sigma(X)$.

Opgave 1.4 (fordelingsfunktion)

Givet følgende fordelingsfunktion $F(x)$:

x	$x < 0$	$0 \leq x < 4$	$4 \leq x < 6$	$6 \leq x < 12$	$12 \leq x < 18$	$18 \leq x$
$F(x)$	0	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{3}{4}$	1

- 1) Tegn grafen for $F(x)$.
- 2) Find den tilsvarende tæthedsfunktion $f(x)$ og tegn grafen.
- 3) Beregn middelværdien $E(X)$, variansen $V(X)$ og spredningen $\sigma(X)$.

Opgave 1.5 (tæthedsfunktion)

Givet følgende funktion:

$$f(x) = \begin{cases} \frac{1}{4}(3-x) & \text{for } 0 < x < 2 \\ 0 & \text{ellers} \end{cases}$$

- 1) Vis, at $f(x)$ opfylder de to første betingelser i definitionen på en tæthedsfunktion. Idet følgende antages, at $f(x)$ er tæthedsfunktion for en stokastisk variabel X .
- 2) Skitser grafen for $f(x)$, find den tilsvarende fordelingsfunktion $F(x)$ og skitser også dennes graf.
- 3) Beregn middelværdien $E(X)$, variansen $V(X)$ og spredningen $\sigma(X)$.
- 4) Bestem $P(0.25 \leq X < 0.5)$
- 5) Bestem medianen for X .
- 6) Lad der være givet følgende stikprøve på X : 0.3 1.8 0.6 0.1 0.9 1.9 0.4 1.2 0.2 1.5
Find på dette grundlag et estimat for middelværdi $E(X)$ og spredning $\sigma(X)$.

Opgave 1.6 (tæthedsfunktion)

Givet følgende funktion:

$$f(x) = \begin{cases} \frac{3}{x^4} & \text{for } 1 \leq x \\ 0 & \text{ellers} \end{cases}$$

- 1) Vis, at $f(x)$ opfylder de to første betingelser i definitionen på en tæthedsfunktion. Idet følgende antages, at $f(x)$ er tæthedsfunktion for en stokastisk variabel X .
- 2) Skitser grafen for $f(x)$, find den tilsvarende fordelingsfunktion $F(x)$ og skitser dens graf.
- 3) Beregn middelværdien $E(X)$, variansen $V(X)$ og spredningen $\sigma(X)$.
- 4) Bestem $P(2 \leq X \leq 3)$.

Opgave 1.7 (fordelingsfunktion)

Givet følgende fordelingsfunktion

$$F(x) = \begin{cases} 0 & \text{for } x < 2 \\ \frac{1}{8}x - \frac{1}{4} & \text{for } 2 \leq x < 10 \\ 1 & \text{ellers} \end{cases}$$

- 1) Find den tilsvarende tæthedsfunktion $f(x)$ og tegn grafen for såvel $f(x)$ som $F(x)$.
- 2) Beregn middelværdien $E(X)$, variansen $V(X)$ og spredningen $\sigma(X)$.
- 3) Bestem $P(3 \leq X < 6)$
- 4) Find nedre kvartil.
- 5) Lad der være givet følgende stikprøve på X : 4.1 2.1 5.8 3.0 8.2 3.1 6.1 4.3 7.9
Find på dette grundlag et estimat for middelværdi $E(X)$ og spredning $\sigma(X)$.

Opgave 1.8 (fordelingsfunktion)

Givet følgende fordelingsfunktion:

$$F(x) = \begin{cases} 0 & \text{for } x < 4 \\ 1 - \left(\frac{4}{x}\right)^5 & \text{for } 4 \leq x \end{cases}$$

- 1) Find den tilsvarende tæthedsfunktion $f(x)$ og tegn grafen for såvel $f(x)$ som $F(x)$.
- 2) Beregn middelværdien $E(X)$, variansen $V(X)$ og spredningen $\sigma(X)$.
- 3) Bestem $P \leq X < 6)$
- 4) Find øvre kvartil.

Opgave 1.9 (histogram)

Færdselspolitiet overvejede, om der burde indføres en fartgrænse på 70 km/h på en bestemt landevejsstrækning, hvor der hidtil havde været fartgrænsen 80 km/h.

Som et led i analysen af hensigtsmæssigheden af den overvejede ændring observeredes inden for et bestemt tidsrum ved hjælp af radarkontrol de forbipasserende bilers fart. Resultatet af målingerne var:

50 observationer									
64	72	82	52	60	95	86	70	63	48
50	63	35	60	77	41	47	88	62	66
59	49	55	99	65	76	76	68	51	80
75	74	64	74	62	70	85	73	93	65
98	55	85	80	78	53	96	71	84	103

Statistiske grundbegreber

- 1) Gruppér de observerede data ved opdeling i et antal passende intervaller, og tegn stikprøvens histogram (opgaven kan eventuelt løses ved anvendelse af et statistikprogram som Statgraphics).
- 2) Beregn stikprøvens gennemsnit og spredning.
- 3) Tegn det tilsvarende sumpolygon, og giv ud fra figuren et skøn for fordelings median.

Opgave 1.10 (histogram)

Til fabrikation af herreskjorter benyttes et råmateriale, som indeholder en vis procentdel uld. For nærmere at undersøge uldprocenten, måles denne i 64 tilfældigt udvalgte batch. Resultatet var (i %):

34.2	33.1	34.5	35.6	36.3	35.1	34.7	33.6	33.6	34.7	35.0	35.4	36.2	36.8	35.1	35.3
33.8	34.2	33.4	34.7	34.6	35.2	35.0	34.9	34.7	33.6	32.5	34.1	35.1	36.8	37.9	36.4
37.8	36.6	35.4	34.6	33.8	37.1	34.0	34.1	32.6	33.1	34.6	35.9	34.7	33.6	32.9	33.5
35.8	37.6	37.3	34.6	35.5	32.8	32.1	34.5	34.6	33.6	34.1	34.7	35.7	36.8	34.3	32.7

- 1) Gruppér de observerede data ved opdeling i et antal passende intervaller, og tegn stikprøvens histogram (opgaven kan eventuelt løses ved anvendelse af et statistikprogram som Statgraphics).
- 2) Beregn stikprøvens gennemsnit og spredning.
- 3) Tegn det tilsvarende sumpolygon, og giv ud fra figuren et skøn for fordelings 1, 2 og 3 kvartil.

Opgave 1.11 (usikkerhedsberegning)

En mængde råmateriale til en produktion ligger i kegleformet bunke. En kegle med radius R og højde

H har volumenet $V = \frac{\pi}{3} R^2 H$.

Man har målt $R = 12.0$ m, $H = 11.0$ m,

med usikkerheder $\sigma(R) = 0.2$ m, $\sigma(H) = 0.1$ m.

Det kan antages, at måleresultaterne for R og H er statistisk uafhængige.

Find volumenet V , usikkerheden $\sigma(V)$, samt den relative usikkerhed $rel(V)$.

Opgave 1.12 (usikkerhedsberegning)

For en rektangulær flade har man målt længden L og bredden B :

$L = 12.3$ m, $B = 8.4$ m

med usikkerheder

$\sigma(L) = 0.1$ m, $\sigma(B) = 0.2$ m.

Det kan antages, at måleresultaterne for L og B er statistisk uafhængige.

Find fladens areal A , usikkerheden $\sigma(A)$, samt den relative usikkerhed $rel(A)$.

Opgave 1.13 (usikkerhedsberegning)

For et bassin af form som en retvinklet kasse har man målt længden L , bredden B og højden H :

$$L = 18.0 \text{ m} , \quad B = 12.3 \text{ m} \quad H = 4.5 \text{ m}$$

med usikkerheder

$$\sigma(L) = 0.2 \text{ m} , \quad \sigma(B) = 0.1 \text{ m} , \sigma(H) = 0.2 \text{ m} .$$

Det kan antages, at måleresultaterne for L , B og H er statistisk uafhængige.

Find bassinets volumen V , usikkerheden $\sigma(V)$, samt den relative usikkerhed $rel(V)$.

2 NORMALFORDELINGEN

2.1 INDLEDNING

Lad os som eksempel tænke os et kemisk forsøg, hvor vi måler udbyttet af et stof A. Selv om vi gentager forsøget ved anvendelse af den samme metode og i øvrigt søger at gøre forsøgsbetingelserne så ensartet som muligt, varierer udbyttet dog fra forsøg til forsøg. Disse variationer fra den ene forsøg til det næste må skyldes forhold vi ikke kan styre. Det kan skyldes små ændringer i temperaturen, i luftens relative fugtighed, vibrationer under fremstillingen, små forskelle i de anvendte råmaterialer (kornstørrelse, renhed), forskelle i menneskelig reaktionsevne osv. Hvis ingen af disse variationsårsager er dominerende, der er et stort antal af dem, de er uafhængige og lige så godt kan have en positiv som en negativ indvirkning på resultatet, så vil den totale fejl sædvanligvis approksimativt være fordelt efter den såkaldte **normalfordeling**. (også kaldet Gauss-fordelingen)

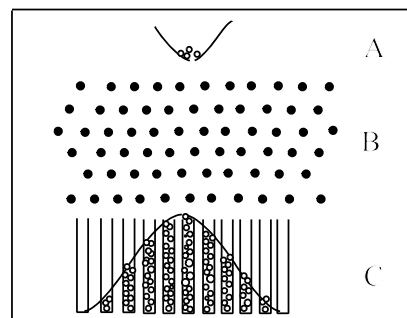
Som illustration af dette kan anvendes Galtons apparat.

Eksempel 2.1. Eksperiment med et Galton-apparat.

På den anførte figur er skitseret et Galton-apparat.

A er en tragt; B er sømrækker, hvor sømmene i en underliggende række er anbragt midt ud for mellemrummene mellem sømmene i den overliggende række; C er opsamlingskanaler.

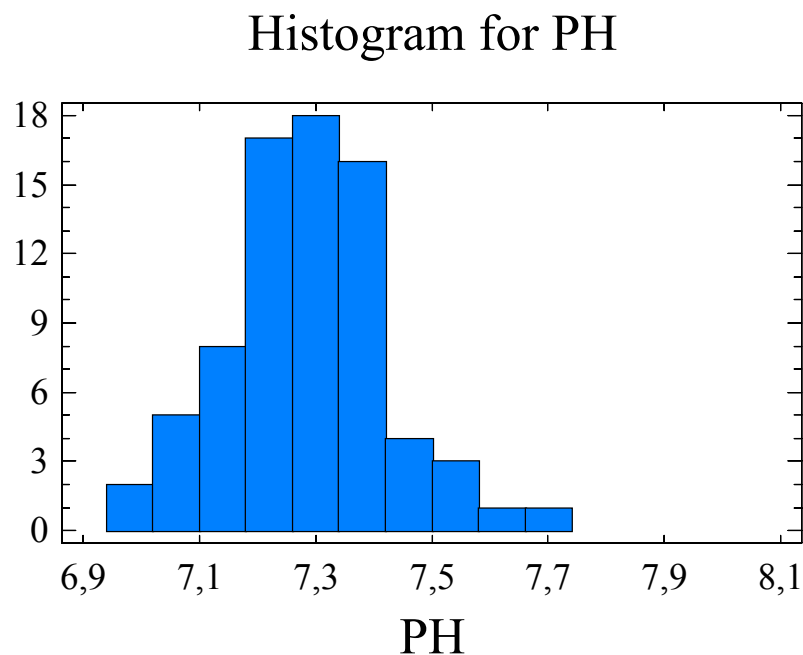
Lader man mange kugler passere gennem tragten A ned gennem sømrækkerne B til opsamlingskanalerne C, vil man konstatere, at de enkelte kugler nok bliver tilfældigt fordelt i opsamlingskanalerne, men at kuglernes samlede fordeling giver et mønster, som gentages, hver gang man udfører eksperimentet. Fordelingen er hver gang med tilnærmelse en klokkeformet symmetrisk fordeling som skitseret på tegningen, noget som er karakteristisk for normalfordelingen.



Galton-apparatet illustrerer, hvorfor man så ofte antager, at måleresultater er værdier af en normalfordelt variabel: Hver sømrække repræsenterer en faktor, hvis niveau det ikke er muligt at holde konstant fra måling til måling, og sømrækkernes påvirkning af kuglens bane symboliserer den samlede virkning, som de ukontrollerede faktorer har på størrelsen af den målte egenskab.



En anden illustration af under hvilke omstændigheder en normalfordelt variabel kan forekomme i praksis så vi i kapitel 1 eksempel 1.3 hvor man på 75 mennesker med en bestemt ledsygdom målte pH i knæleddet. Histogrammet som er gentaget nedenfor har et klokkeformet udseende, som kraftigt antyder, at den kontinuerte stokastiske variabel $X = \text{pH}$ er normalfordelt.



I den teoretiske statistik giver den centrale grænseværdisætning (se sætning 2.3) en forklaring på, hvorfor normalfordelingen er en god model i de ovenfor nævnte situationer.

2.2 DEFINITION OG BEREGNING

DEFINITION af normalfordeling Lad μ være et reelt tal og σ et positivt tal.

Sandsynlighedsfordelingen for en kontinuert stokastisk variabel X med tæthedsfunktionen $f(x)$

bestemt ved $f(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \cdot e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2}$ for ethvert x

kaldes **normalfordelingen** $n(\mu, \sigma)$ med parametrene μ og σ .

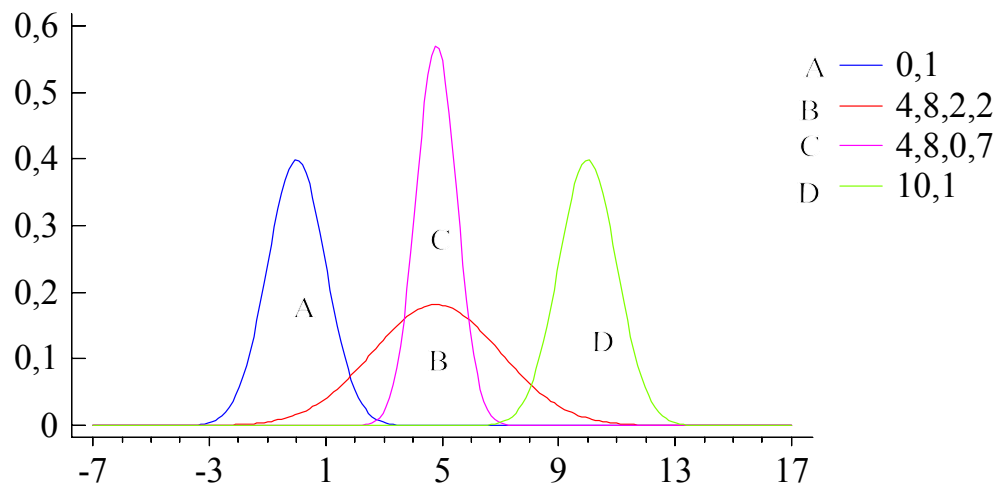
At $f(x)$ virkelig er en tæthedsfunktion vises i “Supplement til statistiske grundbegreber afsnit 2.A”

SÆTNING 2.1 (Middelværdi og spredning for normalfordeling).

Normalfordelingen $n(\mu, \sigma)$ har middelværdien μ og spredningen σ . Grafen for tæthedsfunktionen $f(x)$ er symmetrisk om linien $x = \mu$

Sætningen bevises i “Supplement til statistiske grundbegreber afsnit 2.A”

For at få et overblik over betydningen af μ og σ er der nedenfor afbildet tæthedsfunktionerne for normalfordelingerne $n(0, 1)$, $n(4,8, 2,2)$, $n(4,8, 0,7)$ og $n(10, 1)$.



Det ses, at tæthedsfunktionerne er klokkeformede, og at et interval på $[\mu - 3 \cdot \sigma; \mu + 3 \cdot \sigma]$ indeholder stort set hele sandsynlighedsmassen.

Stikprøves gennemsnit og spredning. Ofte er middelværdien μ og spredningen σ ukendt i en foreliggende normalfordeling. I så fald erstattes fordelingen $n(\mu, \sigma)$ i praksis med en approksimerende fordeling $n(\bar{x}, s)$, såfremt der foreligger et rimelig stort antal observationer fra den givne fordeling.

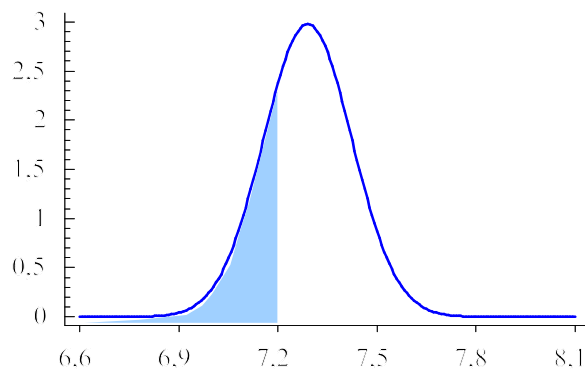
På basis af den i eksempel 1.3 angivne stikprøve på 75 patienter beregnes et gennemsnit af pH værdierne på $\bar{x} = \frac{546.52}{75} = 7.2868$ og en s værdi på $s = \sqrt{\frac{SAK}{n-1}} = 0.134355$.

(udregnet på lommeregner).

Vi vil altså antage, at pH værdierne er approksimativt normalfordelt $n(7.29, 0.134)$.

Beregning af $P(X \leq x)$ for en normalfordelt variabel X

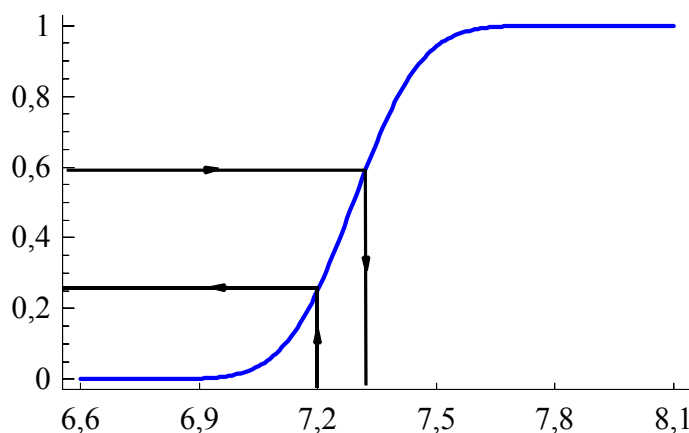
Ønsker vi at benytte ovenstående normalfordeling $n(7.29, 0.134)$ til at finde sandsynligheden for, at pH er mindre end 7.2, er denne sandsynlighed lig med arealet af det skraverede areal under tæthedsfunktionen.



Har man omvendt givet en sandsynlighed $p = 0.6$ og ønsker at finde den tilsvarende værdi x_p for hvilken $P(X \leq x_p) = 0.6$ betyder det, at man kender arealet 0.6 og skal finde x -værdien.

Nedenfor er tegnet den tilsvarende fordelingsfunktionen $F(x) = P(X \leq x)$.

Af figuren aflæses at $P(X \leq 7.2) \approx 0.25$ og $x_{0.6} \approx 7.3$



Normalfordelingen.

På større lommeregnere eller ved hjælp af regneark som Excel eller statistikprogrammer som Statgraphics kan man efter indtastning af middelværdi og spredning få beregnet den ønskede sandsynlighed (jævnfør eventuelt appendix 2).

På TI 89 findes således de ønskede funktioner ved tryk på CATALOG

$$P(X \leq 7.2) = \text{normCdf}(-\infty, 7.2, 7.29, 0.134) = 0.2509$$

$$x_{0.6} = \text{invnorm}(0.6, 7.29, 0.134) = 7.324$$

Har man ikke disse hjælpemidler til rådighed, må man benytte tabel. Den normalfordeling, hvis fordelingsfunktion er tabellagt, er den såkaldte **normerede normalfordeling**. Den er bestemt ved at have middelværdien 0 og spredningen 1. En stokastisk variabel, der er normalfordelt $n(0,1)$, kaldes sædvanligvis U og dens fordeling **U-fordelingen**¹⁾. Dens tæthedsfunktion² benævnes φ og dens fordelingsfunktion Φ ³.

I tabel 1 sidst i dette notat, er anført værdier af funktionen Φ , og i tabel 2 er anført ofte benyttede værdier af U -fordelingens p-fraktiler u_p .

Eksempel 2.2. Beregning af sandsynligheder i U-fordelingen ved benyttelse af tabel

Beregn sandsynlighederne $P(U \leq 1.24)$, $P(U > 0.83)$ og $P(-1.12 < U \leq 1.45)$.

LØSNING:

Vi får ved hjælp af tabel 1:

$$P(U \leq 1.24) = \Phi(1.24) = \underline{\underline{89.25\%}}$$

$$P(U > 0.83) = 1 - P(U \leq 0.83) = 1 - \Phi(0.83) = 1 - 0.7967 = \underline{\underline{20.33\%}}$$

$$P(-1.12 < U \leq 1.45) = P(U \leq 1.45) - P(U \leq -1.12) = \Phi(1.45) - \Phi(-1.12) = 0.9265 - 0.1314 = 0.7951 = \underline{\underline{79.51\%}}$$



Funktionen Φ benyttes også ved beregning af sandsynligheder vedrørende en stokastisk variabel X med normalfordelingen $n(\mu, \sigma)$. Der gælder nemlig følgende sætning:

SÆTNING 2.2. (normering af normalfordeling). Når X er normalfordelt $n(\mu, \sigma)$ er

den variable $U = \frac{X - \mu}{\sigma}$ normalfordelt $n(0,1)$, og der gælder

$$P(X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) \quad \text{og} \quad P(a < X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$

Bemærk, at det for de to formler er ligegyldigt, om ulighederne er med eller uden lighedstegn.

¹⁾ I angelsaksiske lande ofte Z og Z -fordelingen.

²⁾ $\varphi(u) = \frac{1}{\sqrt{2 \cdot \pi}} e^{-\frac{u^2}{2}}$ for ethvert u

³⁾ $\Phi(u) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^u e^{-\frac{t^2}{2}} dt$

Bevis: At U også er normalfordelt vises ikke her.

$$E(U) = E\left(\frac{X - \mu}{\sigma}\right) = \int_{-\infty}^{\infty} \frac{x - \mu}{\sigma} f(x) dx = \frac{1}{\sigma} \int_{-\infty}^{\infty} x \cdot f(x) dx - \frac{\mu}{\sigma} \int_{-\infty}^{\infty} f(x) dx = \frac{1}{\sigma} E(X) - \frac{\mu}{\sigma} = 0$$

$$V(U) = V\left(\frac{X - \mu}{\sigma}\right) = \int_{-\infty}^{\infty} \left(\frac{x - \mu}{\sigma}\right)^2 f(x) dx = \frac{1}{\sigma^2} \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx = \frac{V(X)}{\sigma^2} = 1$$

U har derfor middelværdi 0 og spredning 1.

$$\text{Endvidere fås} \quad P(X \leq b) = P\left(\frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) = P\left(U \leq \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) \text{ og}$$

$$P(a < X \leq b) = P\left(\frac{a - \mu}{\sigma} < U \leq \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$



Der gælder følgende sammenhæng mellem fraktiler for X og fraktiler for U :

$$\boxed{x_p = u_p \cdot \sigma + \mu} \quad (2.1)$$

Bevis: $P(X \leq x_p) = p \Leftrightarrow \Phi\left(\frac{x_p - \mu}{\sigma}\right) = p \Leftrightarrow \frac{x_p - \mu}{\sigma} = u_p \Leftrightarrow x_p = u_p \cdot \sigma + \mu$



2 og 3 σ -grænser ved normalfordeling.

I afsnit 1 forudsagde vi, at for normalfordelte variable vil sandsynligheden for at ligge indenfor “2 σ grænserne” $\mu - 2 \cdot \sigma$ og $\mu + 2 \cdot \sigma$ være ca. 95%, men sandsynligheden for at ligge indenfor “3 σ grænserne” $\mu - 3 \cdot \sigma$ og $\mu + 3 \cdot \sigma$ vil være ca. 99%.

Dette kan vi nu bevise:

$$P(\mu - 2 \cdot \sigma < X \leq \mu + 2 \cdot \sigma) = \Phi\left(\frac{\mu + 2 \cdot \sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - 2 \cdot \sigma - \mu}{\sigma}\right) = \Phi(2) - \Phi(-2) = 95.45\%$$

$$P(\mu - 3 \cdot \sigma < X \leq \mu + 3 \cdot \sigma) = \Phi\left(\frac{\mu + 3 \cdot \sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - 3 \cdot \sigma - \mu}{\sigma}\right) = \Phi(3) - \Phi(-3) = 99.73\%$$

Dette benyttes, som vi vil se i Statistik II kapitel 14, blandt andet ved opstillingen af kontrollkort i forbindelse med den statistiske proceskontrol (jævnfør også det følgende eksempel 2.3).

Eksempel 2.3. Normalfordeling.

En fabrik støber plastikkasser. Fabrikken får en ordre på kasser, som blandt andet har den specifikation, at kasserne skal have en længde på 90 cm. Kasser, hvis længder ikke ligger mellem 89.2 og 90.8 cm bliver kasseret.

Det vides, at fabrikken producerer kasserne med en længde X , som er normalfordelt med en spredning på 0.5 cm.

- 1) Hvis X har en middelværdi på 89.6, hvad er så sandsynligheden for, at en kasse har en længde, der ligger indenfor specifikationsgrænserne.
- 2) Hvor stor er sandsynligheden for at en kasse bliver kasseret, hvis man justerer støbningen, så middelværdien bliver den der giver den mindste procentdel kasseret (spredningen kan man ikke ændre).

Fabrikanten finder, at selv efter den i spørgsmål 2 foretagne justering kasseres for stor en procentdel af kasserne. Der ønskes højst 5% af kasserne kasseret.

- 3) Hvad skal spredningen σ formindskes til, for at dette er opfyldt?
- 4) Hvis det er umuligt at ændre σ , kan man prøve at få ændret specifikationsgrænserne.

Find de nye specifikationsgrænser (placeret symmetrisk omkring middelværdien 90,0) idet spredningen stadig er 0.5, og højst 5% må kasseres.

En ny maskine indkøbes, og som et led i en undersøgelse af, om der dermed er sket ændringer i middelværdi og spredning produceres 12 kasser ved anvendelse af denne maskine.

Man fandt følgende længder:

89.2 90.2 89.4 90.0 90.3 89.7 89.6 89.9 90.5 90.3 89.9 90.6.

- 5) Angiv på dette grundlag et estimat for middelværdi og spredning.

LØSNING:

På TI89 findes de benyttede sandsynlighedsfordelinger ved at trykke på CATALOG\

- 1) $P(89.2 \leq X \leq 90.8) = \text{normCdf}(89.2, 90.8, 89.6, 0.5) = 0.7700 = \underline{\underline{77.99\%}}$

$$\begin{aligned}\text{Tabel 1: } P(89.2 < X \leq 90.8) &= P(X \leq 90.8) - P(X \leq 89.2) = \Phi\left(\frac{90.8 - 89.6}{0.5}\right) - \Phi\left(\frac{89.2 - 89.6}{0.5}\right) \\ &= \Phi(2.4) - \Phi(-0.8) = 0.9918 - 0.2119 = \underline{\underline{77.99\%}}\end{aligned}$$

- 2) Middelværdien justeres til midtpunktet 90.0

$$P(X > 90.8) + P(X < 89.2) = \text{normCdf}(90.8, \infty, 90, 0.5) + \text{normCdf}(-\infty, 89.2, 90, 0.5) = 0.1096 = \underline{\underline{10.96\%}}$$

$$\text{Tabel 1: } P(X > 90.8) + P(X < 89.2) = 1 - P(X \leq 90.8) + P(X < 89.2)$$

$$= 1 - \Phi\left(\frac{90.8 - 90.0}{0.5}\right) + \Phi\left(\frac{89.2 - 90.0}{0.5}\right) = 1 - \Phi(1.6) + \Phi(-1.6) = 1 - 0.9452 + 0.0548 = \underline{\underline{10.96\%}}$$

- 3) Da der ligger 5% udenfor intervallet, så må af symmetri grunde 2,5% ligge på hver sin side af intervallet. Vi har følgelig, at vi skal finde spredningen σ så $P(X \leq 89.2) = 0.025$

$$\text{Af relationen (2.1) fås } 89.2 = u_{0.025} \cdot \sigma + 90 \Leftrightarrow \sigma = \frac{-0.8}{u_{0.025}}.$$

$$\text{Da } u_{0.025} = -1.96 \text{ (findes af tabel 2 eller ved invNorm(0.025,0,1)) fås } \underline{\underline{\sigma = 0.408}}$$

- 4) Kaldes den nedre specifikationsgrænse for a fås med samme begrundelse som under punkt 3 : $P(X \leq a) = 0.025$. Vi kan her benytte den "inverse" normalfordeling
 $a = \text{invNorm}(0.025, 90, 0.05) = \underline{89.02}$
 Øvre grænse $b = 90 + (90 - 89.02) = \underline{90.08}$
- 5) Ved indtastning af de 12 tal i en lommeregner (se eventuelt eksempel 1.5) findes $\bar{x} = \underline{89.97}$ og $s = \underline{0.435}$



Den næste sætning er af stor betydning for anvendelserne, idet den løst sagt siger, at selvom man ikke kender fordelingen for de n ensfordelte stikprøvevariable $X_1, X_2, \dots, X_n$, så vil gennemsnittet $\bar{X}$ være approksimativt normalfordelt blot n er tilstrækkelig stor (i praksis over 30).¹

Sætning 2.3 (den centrale grænseværdi sætning). Lad som ovenfor $\bar{X}$ være gennemsnittet for en stikprøve af størrelsen n taget fra en population med middelværdi μ og spredning σ .

Størrelsen $U = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ vil da for n gående mod ∞ være normeret normalfordelt.

Eksempel 2.4. Fordeling af gennemsnit

Den tid, en kunde venter i lufthavnen ved en check-in disk, er givet at være en stokastisk variabel med en middelværdi på 8.2 minutter og en spredning på 3 minutter. Udtages en stikprøve på 49 kunder, ønskes fundet sandsynligheden for at den gennemsnitlige ventetid for disse kunder er mellem 7 og 9 minutter.

LØSNING:

Da antallet i stikprøven $n = 49$ er større end 30, kan vi antage at gennemsnittet er approksimativt

normalfordelt med en middelværdi på 8.2 og en spredning på $\sigma_x = \frac{\sigma}{\sqrt{49}} = \frac{3}{7} = 0.429$.

$$P(7 < \bar{X} < 9) = \text{normCdf}(7, 9, 8.2, 3/7) = 0.9665 = \underline{96.65\%}$$

$$\text{Tabel: } P(7 < \bar{X} < 9) = \Phi\left(\frac{9 - 8.2}{0.429}\right) - \Phi\left(\frac{7 - 8.2}{0.429}\right) = \Phi(1.87) - \Phi(-2.80) = 0.9693 - 0.0026 = 0.9667 = \underline{96.67\%} \quad \blacklozenge$$

¹ $P\left(\frac{(\bar{X} - \mu)\sqrt{n}}{\sigma}\right) \rightarrow \phi(u)$ for $n \rightarrow \infty$

Normalfordelingen.

Vi nævner uden bevis følgende sætning:

SÆTNING 2.4 (*Additionssætning for linearkombination af normalfordelte variable*).

Er Y en linearkombination af n stokastisk uafhængige, normalfordelte variable, vil Y også være normalfordelt.

Eksempel 2.5. Additionssætning.

En boreproces fremstiller huller med en diameter X_1 , der er normalfordelt med en middelværdi μ_1 og en spredning på 0.04. En anden proces fremstiller aksler med en diameter X_2 , der er normalfordelt med en middelværdi μ_2 og en spredning på 0.03.

Antag, at $\mu_1 = 10.00$, og at $\mu_2 = 9.94$.

Find sandsynligheden for, at en tilfældig valgt aksel har en mindre diameter end en tilfældig valgt borehul.

LØSNING:

$$P(X_2 < X_1) = P(X_2 - X_1 < 0).$$

Sættes $Y = X_2 - X_1$ er Y normalfordelt.

$$E(Y) = E(X_2) - E(X_1) = 9.94 - 10.00 = -0.06.$$

$$V(Y) = 1^2 V(X_2) + (-1)^2 V(X_1) = 0.04^2 + 0.03^2 = 0.025$$

$$\sigma(Y) = \sqrt{0.025} = 0.05$$

$$P(X_2 < X_1) = P(Y < 0) = \text{normCdf}(-\infty, 0, -0.06, 0.05) = 0.8849 = \underline{\underline{88.49\%}}$$

Tabel: $P(X_2 < X_1) = P(Y < 0) = \Phi\left(\frac{0 - (-0.06)}{0.05}\right) = \Phi(1.2) = \underline{\underline{0.8849}}$



OPGAVER

Opgave 2.1 (normalfordeling)

- 1) En stokastisk variabel X er normalfordelt med $\mu = 0$ og $\sigma = 1$.
Find $P(X \leq 0.75)$, $P(X > 1.6)$ og $P(0.75 < X < 1.6)$.
- 2) En stokastisk variabel X er normalfordelt med $\mu = 25.1$ og $\sigma = 2.4$.
Find $P(22.3 < X \leq 27.8)$.

Opgave 2.2 (normalfordeling)

Maksimumstemperaturen, der opnås ved en bestemt opvarmningsproces, har en stokastisk fordeling med en middelværdi på 113.3° og en spredning på 5.6°C . Det antages, at maksimumstemperaturens variation er tilfældig og kan beskrives ved en normalfordeling.

- 1) Find procenten af maksimumstemperaturer, der er mindre end 116.1°C .
- 2) Find procenten af maksimumstemperaturer, der ligger mellem 115°C og 116.7°C .
- 3) Find den værdi, som overskrides af 57.8% af maksimumstemperaturerne.

Man overvejer at gå over til en anden opvarmningsproces. Man udfører derfor 16 gange i løbet af en periode forsøg, hvor man måler maksimumstemperaturen, der opnås ved denne nye proces. Resultaterne var 116.6 , 116.6 , 117.0 , 124.5 , 122.2 , 128.6 , 109.9 , 114.8 , 106.4 , 110.7 , 110.7 , 113.7 , 128.1 , 118.8 , 115.4 , 123.1

- 4) Giv et estimat for middelværdien og spredningen.

Opgave 2.3 (normalfordeling)

En fabrik planlægger at starte en produktion af rør, hvis diametre skal opfylde specifikationerne $2,500 \text{ cm} \pm 0,015 \text{ cm}$.

Ud fra erfaringer med tilsvarende produktioner vides, at de producerede rør vil have diametre, der er normalfordelte med en middelværdi på $2,500 \text{ cm}$ og en spredning på $0,010 \text{ cm}$. Man ønsker i forbindelse med planlægningen svar på følgende spørgsmål:

- 1) Hvor stor en del af produktionen holder sig indenfor specifikationsgrænserne.
- 2) Hvor meget skal spredningen σ ned på, for, at 95% af produktionen holder sig indenfor specifikationsgrænserne (middelværdien er uændret på $2,500 \text{ cm}$).
- 3) Fabrikken overvejer, om det er muligt at få indført nogle specifikationsgrænser (symmetrisk omkring $2,500$), som bevirker, at 95% af dets produktion falder indenfor grænserne. Find disse grænser, idet det stadig antages at middelværdien er 2.500 og spredningen 0.010 cm .

Normalfordelingen.

Opgave 2.4 (normalfordeling)

En automatisk dåsepåfyldningsmaskine fylder hønskødssuppe i dåser. Rumfanget er normalfordelt med en middelværdi på 800 ml og en spredning på 6,4 ml.

- 1) Hvad er sandsynligheden for, at en dåse indeholder mindre end 790 ml?
- 2) Hvis alle dåser, som indeholder mindre end 790 ml og mere end 805 ml bliver kasseret, hvor stor en procentdel af dåserne bliver så kasseret?
- 3) Bestem de specifikationsgrænser der ligger symmetrisk omkring middelværdien på 800 ml, og som indeholde 99% af alle dåser.

Opgave 2.5 (additionssætning)

I et laboratorium lægges et nyt gulv. Det forudsættes, at vægten Y der hviler på gulvet, er summen af vægten X_1 af maskiner og apparater og vægten X_2 af varer og personale. Da både X_1 og X_2 er sum af mange relativt små vægte, antages det, at de er normalfordelte. Det antages endvidere at X_1 og X_2 er statistisk uafhængige. Erfaringer fra tidligere gør det rimeligt at antage, at der gælder følgende middelværdier og spredninger (målt i tons): $E(X_1) = 6.0$, $\sigma(X_1) = 1.2$, $E(X_2) = 3.5$, $\sigma(X_2) = 0.4$.

- 1) Beregn $E(Y)$ og $\sigma(Y)$.
- 2) Beregn det tal y_0 , som vægten Y med de ovennævnte forudsætninger kun har en sandsynlighed på 1% for at overskride.
- 3) Beregn sandsynligheden for, at vægten af varer og personale en tilfældig dag, efter at det nye gulv er lagt, er større end vægten af maskiner og apparater. (Vink: se på differensen $X_2 - X_1$)

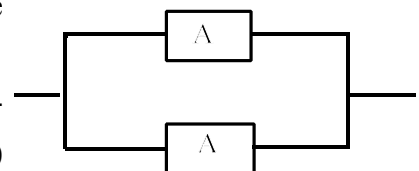
Opgave 2.6 (additionssætning)

Ved fabrikation af et bestemt mærke opvaskemiddel fyldes vaskepulver i papkartoner. I middel fyldes 4020 g pulver i hver karton, idet der herved er en spredning på 12 g. Pulverfyldningen kan forudsættes ikke at afhænge af kartonernes vægt, der i middel er 250 g med en spredning på 5g. Beregn sandsynligheden p for, at en tilfældig pakke opvaskemiddel har en bruttovægt mellem 4250 g og 4300 g.

Opgave 2.7 (additionssætning)

Et system er af sikkerhedsmæssige grunde opbygget af to apparater A, der er parallelforbundne (se figur) således, at systemet virker, så længe blot et af apparaterne virker.

Svifter et af apparaterne, startes reparation. Det antages, at reparationstiden er normalfordelt med middelværdien $\mu_{rep} = 10$ timer og spredning $\sigma_{rep} = 3$ timer.



I reparationstiden overbelastes den anden komponent, og det antages, at dens levetid fra reparationens start (approksimativt) er normalfordelt med middelværdi μ og spredning $\sigma = 4$ timer.

- 1) Find sandsynligheden for, at reparationen er afsluttet, inden den anden komponent fejler, hvis $\mu = 20$ timer.
- 2) Hvor stor skal μ være, for at sandsynligheden for, at reparationen kan afsluttes før den anden komponent fejler, er mere end 99.9%?

Opgave 2.8 (additionssætning)

Vægten af en (tilfældig udvalgt) tablet af en vis type mod hovedpine har middelværdien 0.65 g og spredningen 0.04 g.

- 1) Beregn middelværdi og spredning af den sammenlagte vægt af 100 (tilfældigt udvalgte) tabletter.
- 2) Antag, at man benytter følgende metode til at fylde tabletter i et glas. Man placerer glasset på en vægt og fylder tabletter på, indtil vægten af tabletterne i glasset overstiger 65,3 g. Beregn sandsynligheden for, at glasset kommer til at indeholde mere end 100 tabletter (se bort fra vægtens fejlvisning).

3 STATISTISKE TESTFUNKTIONER

3.1 INDLEDNING

I de følgende kapitler vil et hovedemne være test af forskellige såkaldte hypoteser. I den forbindelse er det nødvendigt udover de allerede beskrevne fordelinger yderligere at kende følgende 3 fordelinger: χ^2 -fordelingen, t -fordelingen og F -fordelingen. Det er til vort brug ikke nødvendigt at have et detaljeret kendskab til disse funktioner, idet det fortrinsvis er nok at vide i hvilke forbindelser de anvendes, og enten ved hjælp af lommeregner eller tabel kunne finde værdier af fordelingsfunktionerne samt udvalgte fraktiler. Ønskes en mere dybtgående behandling henvises til det supplements-tillæg der findes på nettet benævnt "Supplement til statistiske grundbegreber".

3.2 χ^2 -FORDELINGEN

χ^2 -fordelingen (udtales ki i anden) benyttes i de næste kapitler ved testning af hypoteser vedrørende middelværdien σ^2 af en s^2 -størrelse, af hypoteser om statistisk uafhængighed, samt om hypoteser vedrørende fordelingstypen for en given stokastisk variabel X (er X normalfordelt?).

Definition af χ^2 -fordelingen. Lad $U_1, U_2, \dots, U_f$ være uafhængige normerede normalfordelte variable.

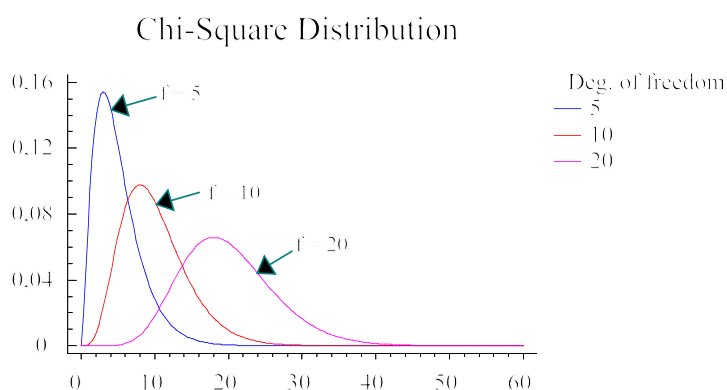
Sandsynlighedsfordelingen for den stokastiske variabel $\chi^2 = U_1^2 + U_2^2 + \dots + U_f^2$ kaldes χ^2 -fordelingen med frihedsgradstallet f og betegnes $\chi^2(f)$.

I "Supplement til statistiske grundbegreber" afsnit 3A er vist en række sætninger om χ^2 -fordelingen, og blandt andet at χ^2 -fordeling $\chi^2(f)$ har middelværdien $E(\chi^2) = f$ og spredningen $\sigma(\chi^2) = \sqrt{2f}$.

På figuren er afbildet tæthedsfunktionen for χ^2 - fordelingerne $\chi^2(5)$, $\chi^2(10)$ og $\chi^2(20)$.

Det ses, at χ^2 kun er defineret for tal større end eller lig nul, og at χ^2 -fordelinger ikke er symmetriske om middelværdien. Jo større frihedsgradstallet bliver jo mere symmetriske bliver de dog, og for store f - værdier - i praksis

$f > 30$ - kan en χ^2 -fordeling $\chi^2(f)$ approksimeres med normalfordelingen $n(\mu, \sigma)$, hvor $\mu = f$ og $\sigma = \sqrt{2 \cdot f}$.



TI89 har en kumuleret χ^2 - fordeling ligesom naturligvis programmer som Statgraphics eller Maple har det. Ved anvendelserne kan man dog udmærket klare sig med fraktiltabeller som angivet i tabel 3.

Eksempel 3.1. Bergning af χ^2 - værdier.

1) Find $\chi^2_{0.025}(8)$ og $\chi^2_{0.975}(8)$.

2) Find $P(X \leq 5)$,

hvor X er χ^2 - fordelt med 8 frihedsgrader.

LØSNING:

1) TI89: $\chi^2_{0.025}(8) = \text{invChi2}(0.025, 8) = \underline{2.18}$
 $\chi^2_{0.975}(8) = \text{invChi2}(0.975, 8) = \underline{17.5}$
 (se det skraverede areal på figuren)

Tabel: Ved opslag i tabel 3 under $f=8$ fås

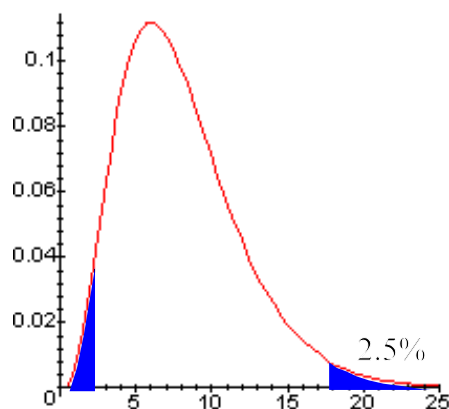
for $p=2.5\%$, at $\chi^2_{0.025}(8) = \underline{2.18}$,

og for $p=97.5\%$, at $\chi^2_{0.975}(8) = \underline{17.5}$

2) TI89: $P(X \leq 5) = \text{chi2Cdf}(0, 5, 8) = \underline{0.242}$

Tabel: Af tabel 3 fremgår, at $P(X \leq 5)$ ligger mellem 20% og 50%, men tættest ved de 20 %.

Ved interpolation mellem 4.59 og 7.34 fås ca. 24 %.



3.3 t -fordelingen.

t -fordelingen benyttes i de følgende kapitler ved testning af hypoteser vedrørende middelværdier μ på basis af stikprøver med gennemsnit $\bar{x}$, når de tilhørende spredninger er ukendte og derfor erstattet af s .

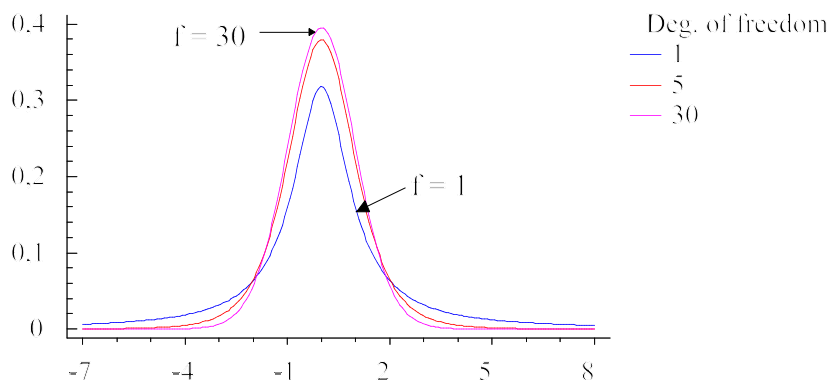
Til enhver t -fordeling er knyttet et helt positivt tal f , som kaldes frihedsgradstallet. En sådan fordeling betegnes $t(f)$.

I "Supplement til statistiske grundbegreber" afsnit 3 B er givet en præcis definition af denne fordeling, og her vises også, at størrelsen

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

er t -fordelt $t(n - 1)$.

Grafen nedenfor viser tæthedsfunktionen for t -fordelingerne $t(1)$, $t(5)$ og $t(30)$.



Alle t -fordelinger har 2-aksen som symmetriakse og har samme "klokkeformede" udseende som normalfordelingen $n(0,1)$. t -fordelingerne er dog "bredere" end U -fordelingen, men nærmer sig mere og mere til U -fordelingen jo større frihedsgradstallet f er. For store f -værdier - i praksis $f > 30$ - kan en t -fordeling $t(f)$ approksimeres med U -fordelingen $n(0,1)$.

Ti 89 har en kumuleret t -fordeling ligesom naturligvis programmer som Statgraphics eller Maple har det. Ved anvendelserne kan man dog udmærket klare sig med fraktiltabeller som angivet i tabel 4.

Eksempel 3.2. Beregning af t-værdier.

- 1) Find $t_{0.975}(12)$ og $t_{0.025}(12)$.
- 2) Find $P(X \geq 1)$, hvor X er t -fordelt med 12 frihedsgrader.

LØSNING:

$$1) \text{ TI-89: } t_{0.975}(12) = \text{inv_t}(0.975, 12) = \underline{\underline{2.18}}$$

$$t_{0.025}(12) = \text{inv_t}(0.025, 12) = \underline{\underline{-2.18}}$$

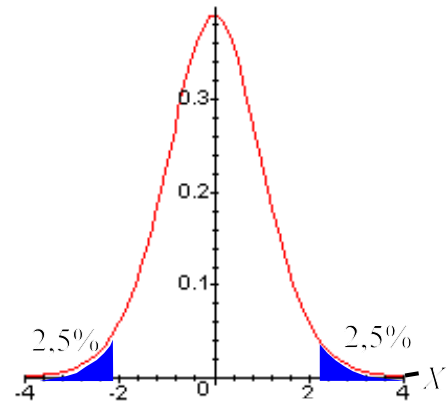
Tabel: Ved opslag i tabel 4 under $f = 12$ fås for $p = 97.5\%$,

$$\text{at } t_{0.975}(12) = \underline{\underline{2.18}}$$

Idet man udnytter, at t -fordelingen er symmetrisk omkring 0, (se figuren) fås, at

$$t_{0.025}(12) = -t_{0.975}(12) = \underline{\underline{-2.18}}$$

$$2) P(X \geq 1) = \text{tCdf}(1, \infty, 12) = 0.1685 = \underline{\underline{16.85\%}}$$

Tabel: Af tabel 4 fremgår, at $P(X \geq 1)$ ligger ca midt mellem 10% og 25%, dvs. er ca. 17%**3.4 F - fordelingen**

Til enhver F -fordeling er knyttet to hele positive tal f_T og f_N kaldet henholdsvis tællerfrihedsgradstallet og nævnerfrihedsgradstallet. En sådan fordeling benævnes $F(f_T, f_N)$.

I “afsnit til statistiske grundbegreber” 3C er givet en præcis definition af denne fordeling, og her

vises også, at den variable $F = \frac{\frac{s_1^2}{\sigma_1^2}}{\frac{s_2^2}{\sigma_2^2}}$ er F -fordelt $F(f_1, f_2)$, hvor f_1 og f_2 er de til s_1^2 og s_2^2 hørende

frihedsgradstal.

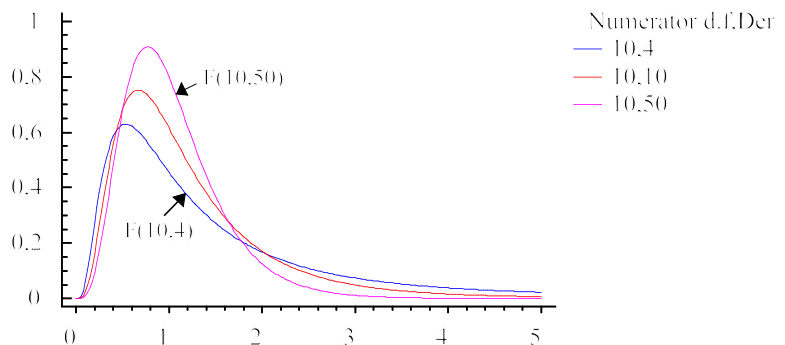
Dette benyttes i statistikken bl. a. til testning af hypoteser vedrørende forholdet mellem to varianser.

På figuren er afbildet tæthedsfunktion

nen for F -fordelingerne

$F(10,4)$, $F(10,10)$ og $F(10,50)$.

Det ses, at F -fordelinger kun er defineret for tal større end eller lig nul, og at F -fordelinger ikke er symmetriske.



TI-89 har en kumuleret F-fordeling ligesom naturligvis programmer som Statgraphics eller Maple har det. Ved anvendelserne kan man dog udmærket klare sig med fraktiltabeller.

I tabel 5 sidst i notatet er anført 95% fraktiler for forskellige værdier af frihedsgraderne. Der er dog kun givet fraktiltabeller for store p-værdier, idet der som vist i "Supplement til statistiske grundbegreber" afsnit 3C gælder, at

$$F_p(f_1, f_2) = \frac{1}{F_{1-p}(f_2, f_1)} \quad \text{f.eks.} \quad F_{0.025}(5, 7) = \frac{1}{F_{0.975}(7, 5)}.$$

Det skal bemærkes, at vi ikke kommer til at anvende F -fordelingen i dette notat.

Eksempel 3.3. Beregning af F -værdier.

1) Find $F_{0.95}(8, 12)$ og $F_{0.05}(8, 12)$

3) Find $P(X \geq 2)$, hvor X er F -fordelt $F(8, 12)$.

LØSNING:

1) TI-89: $F_{0.95}(8, 12) = \text{invF}(0.95, 8, 12) = \underline{\underline{2.85}}$

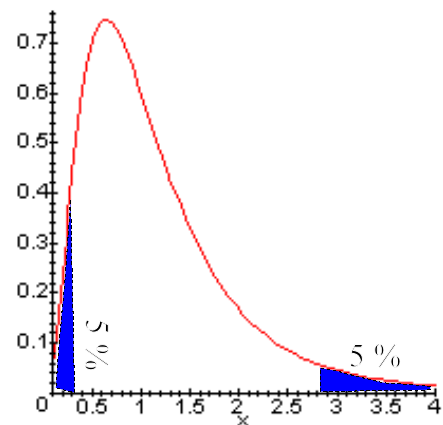
$$F_{0.05}(8, 12) = \text{invF}(0.05, 8, 12) = \underline{\underline{0.305}}$$

Tabel: Ved opslag i tabel 5 under $f_T = 8$ og

$$f_N = 12 \text{ fås } F_{0.95}(8, 12) = \underline{\underline{2.85}}.$$

$$F_{0.05}(8, 12) = \frac{1}{F(0.95)} = \frac{1}{3.28} = \underline{\underline{0.305}}$$

3) $P(X \geq 2) = \text{FCdf}(2, \infty, 8, 12) = \underline{\underline{0.1347}}.$



OPGAVER

Opgave 3.1. Opslag i tabel.

- 1) Find $\chi^2_{0.025}(10)$ og $\chi^2_{0.975}(10)$
- 2) Find $P(X \leq 24)$, hvor X er χ^2 - fordelt med 8 frihedsgrader.

Opgave 3.2. Opslag i tabel.

- 1) Find $t_{0.025}(15)$ og $t_{0.995}(15)$
- 2) Find $P(X \leq 2.7)$, hvor X er t - fordelt med 15 frihedsgrader.

Opgave 3.3. Opslag i tabel.

- 1) Find $F_{0.95}(10,14)$ og $F_{0.05}(10,14)$.

4 ESTIMATER OG KONFIDENSINTERVALLER

4.1 Punktestimation.

Vi har i kapitel 2 ofte udtaget en stikprøve, og ud fra denne og en forudsætning om at populationen er normalfordelt søgt at give et skøn (er estimat) for populationens “parametre” middelværdi og spredning. For en normalfordelt variabel X har vi således som estimat for middelværdien μ sat stikprøvens gennemsnit $\bar{X}$, og som et estimat for spredningen σ sat stikprøvens spredning s .

Generelt vil vi sige, at hvis en stokastisk variabel X har en tæthedsfunktion $f(x)$ som er karakteriseret ved en ukendt parameter θ , og hvis $X_1, X_2, \dots, X_n$, er en stikprøve af størrelsen n fra X , så kaldes funktionen $\Theta = h(X_1, X_2, \dots, X_n)$ en **punktestimator** for θ . Eksempelvis er $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ en punktestimator for middelværdien μ .

Kriterier for gode estimators. For en given parameter kan der tænkes foreslået flere punktestimatorer. Eksempelvis kunne både $S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}$ og $S_1^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n}$ tænkes som estimators

for variansen σ^2 . Der er derfor behov for at have kriterier for hvilke estimators der er gode, og blandt de gode estimators kunne afgøre hvilken der er den bedste.

I “Supplement til statistiske grundbegreber” afsnit 4A er angivet sådanne kriterier, og der bevises her, at de estimators vi hidtil har benyttet er gode estimators.

Der gælder således:

$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ er en god estimator for middelværdien μ .

$S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}$ er en god estimator for variansen σ^2 , og bedre end $S_1^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n}$.

$S_\mu^2 = \sum_{i=1}^n \frac{(X_i - \mu)^2}{n}$ er en god estimator for variansen σ^2 , og forudsat X er normalfordelt, og μ

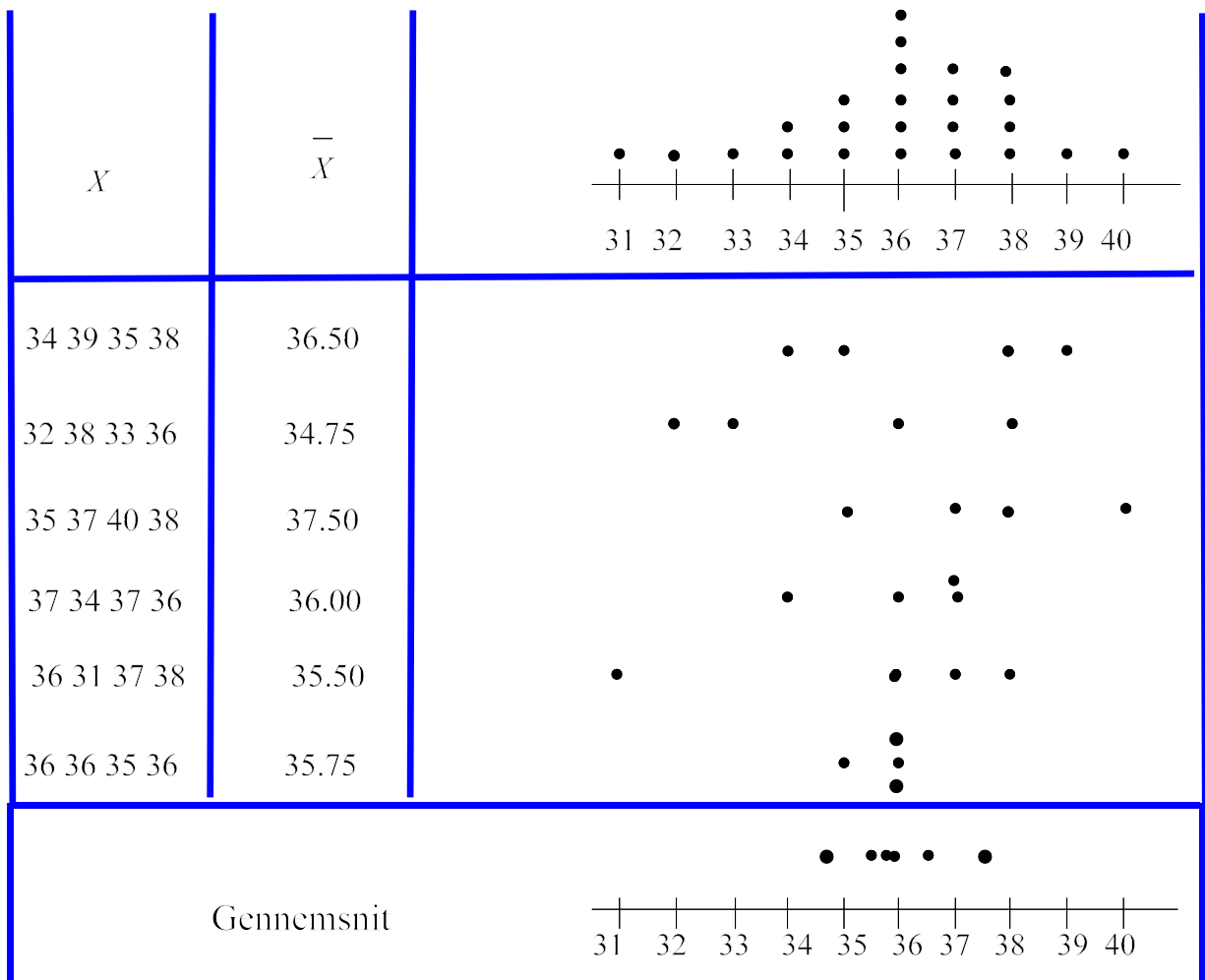
er kendt, vil S_μ^2 være en bedre estimator end S^2 .

4.2 Konfidensinterval

Et estimat for en parameter er behæftet med en vis usikkerhed. Det følgende eksempel 4.1 skal illustrere dette.

Eksempel 4.1 Variation af $\bar{x}$.

En fabrikant af stjernekastere ønsker at få et skøn for brændtiden for et nyt fabrikat af stjernekastere. Seks hold udtager hver en stikprøve på 4 stjernekastere af produktionen, og måler under ensartede forhold brændtiden. Man fandt følgende brændtider (i sekunder)



Af figuren ses tydeligt, at holdene har fået forskellige værdier for gennemsnittet, ligesom brændtiderne “spredt” sig forskelligt. Endvidere ses, at gennemsnitsbrændtiderne for hvert af de 6 hold (se nederst på figuren, hvor de 6 gennemsnitsbrændtider er indtegnet) spredt sig mindre, end det er tilfældet for de enkelte brændtider (se øverst på figuren, hvor alle brændtiderne er indtegnet).



Som resultaterne i eksempel 4.1 illustrerer, er et estimat som gennemsnittet selv en stokastisk variabel med en vis spredning. Af afsnit 1.6 fremgår, at spredningen på gennemsnittet er spredningen på den enkelte brændtid divideret med kvadratroden af antal observationer i stikprøven.

Derfor kan det ofte være nødvendigt at estimatet ledsages af angivelsen af et interval indenfor hvilken den sande middelværdi μ ligger med "f.eks 95% sikkerhed". Et sådant interval kaldes et **95% konfidensinterval**.¹ Mere præcist gælder det, at hvis man for et stort antal stikprøver på den samme stokastiske variabel angav 95% konfidensintervaller, så ville den sande middelværdi tilhøre 95% af disse intervaller. Havde man således i eksempel 4.1 ladet 100 personer hver måle brændtiden for 4 stjernekestere, kunne man forvente, at ca. 5 af 95% konfidensintervallerne ikke ville indeholde den sande brændtid μ .

Lad i det følgende X være en normalfordelt stokastisk variabel med middelværdi μ og spredning σ , og lad $X_1, X_2, \dots, X_n$ være en stikprøve af størrelsen n fra X . Lad endvidere $\bar{x}$ og s være de sædvanlige estimater for μ og σ .

Af sætning 2.4 fremgår, at hvis X er normalfordelt er også $\bar{X}$ normalfordelt. Af den centrale grænseværdisætning (se eventuelt sætning 2.3) fremgår, at for n stor (i praksis blot over 30) vil $\bar{X}$ være approksimativt normalfordelt, uanset X 's fordelingstype. Dette betyder, at når vi i det følgende forudsætter $\bar{X}$ normalfordelt, så er dette i praksis kun et nødvendigt krav for stikprøvestørrelser under 30.

Anskuelig betragtning om 95% konfidensinterval.

Vi har tidligere set, at ca. 95% af arealet under en normalfordelingskurve ligger mellem $\mu - 2 \cdot \sigma$ og $\mu + 2 \cdot \sigma$. På figur 4.1 er de resterende $\alpha = 5\%$ fordelt med 2,5% til hver side (det skraverede areal)

En rimelig antagelse er derfor, at et 95% konfidensinterval er ca. $\bar{x} - u_{0,975} \cdot \sigma(\bar{X}) \leq \mu \leq \bar{x} + u_{0,975} \cdot \sigma(\bar{X})$

Af sætning 1.1 fremgår, at $\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$.

Dette giver følgende 95% konfidensinterval:

$$\bar{x} - u_{0,975} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + u_{0,975} \cdot \frac{\sigma}{\sqrt{n}}.$$

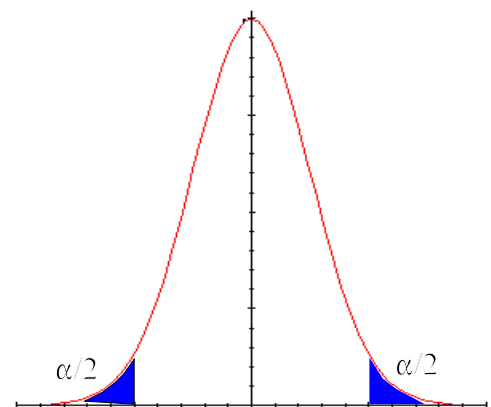


Fig 4.1. U-fordeling

¹**Præcis definition af konfidensinterval.** Lad være givet en stikprøve for en stokastisk variabel X , lad β være et tal mellem 0 og 1. Lad endvidere Θ være en punktestimator for parameteren θ og lad L og U være stokastiske variable, for hvilke det gælder, at $P(L \leq \theta \leq U) = \beta$. På basis af den givne stikprøve findes tal l og u som bestemmer det ønskede interval $l \leq \theta \leq u$. Dette kaldes et $100 \cdot \beta$ procent konfidensinterval for den ukendte parameter θ .

Vi vil sædvanligvis beregne 95%, 99% eller 99.9% konfidensintervaller svarende til $\alpha = 0.05, 0.01$ og 0.001 . Andre værdier kan naturligvis tænkes, men det kræver mere omfattende tabeller, eller anvendelse af edb.

I de fleste tilfælde er σ ikke kendt, og må erstattes af s . For $n \geq 30$ vil ovenstående formel dog stadig kunne anvendes, men for $n < 30$ bliver usikkerheden på s så stor, at man, som det vil fremgå af sætning 4.1, må erstatte normalfordelingskurven med den bredere t -fordeling. Denne er gennemgået i kapitel 3.

Ved $t_{1-\frac{\alpha}{2}}(f)$ forstås $1 - \frac{\alpha}{2}$ fraktilen i t -fordelingen med $f = n - 1$ frihedsgrader.

SÆTNING 4.1 ($100 \cdot (1 - \alpha)$ % konfidensinterval for μ). Et $100 \cdot (1 - \alpha)$ % konfidensinterval

for μ er bestemt ved
$$\bar{x} - t_{1-\frac{\alpha}{2}}(f) \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{1-\frac{\alpha}{2}}(f) \cdot \frac{s}{\sqrt{n}}, \quad f = n - 1.$$

Bevis: I kapitel 3 anførte vi, at $T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$ er t -fordelt med $f = n - 1$ frihedsgrader.

Da t -fordelingen er symmetrisk omkring 0 ligesom u -fordelingen, har vi følgelig, (jævnfør figur 4.1)

$$\text{at } P\left(-t_{1-\frac{\alpha}{2}}(f) \leq T \leq t_{1-\frac{\alpha}{2}}(f)\right) = 1 - \frac{\alpha}{2} \text{ eller } P\left(-t_{1-\frac{\alpha}{2}}(f) \leq \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \leq t_{1-\frac{\alpha}{2}}(f)\right) = 1 - \frac{\alpha}{2}$$

$$\text{Idet } -t_{1-\frac{\alpha}{2}}(f) \leq \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \leq t_{1-\frac{\alpha}{2}}(f) \Leftrightarrow -t_{1-\frac{\alpha}{2}}(f) \cdot \frac{s}{\sqrt{n}} \leq \bar{x} - \mu \leq t_{1-\frac{\alpha}{2}}(f) \cdot \frac{s}{\sqrt{n}}$$

$$\Leftrightarrow \bar{x} - t_{1-\frac{\alpha}{2}}(f) \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{1-\frac{\alpha}{2}}(f) \cdot \frac{s}{\sqrt{n}} \text{ er sætningen bevist.}$$



I appendix 4.1 findes en oversigt over 6 forskellige typer konfidensintervaller, hvoraf de to sidste først vil blive omtalt i kapitel 7. De formler der indgår i beregningerne vil blive bevist i "Supplement til statistiske grundbegreber" afsnit 4B.

Prædiktionsinterval. Ved mange anvendelser ønsker man at **forudsige**, hvor værdien af en **komende** observation af den variable med 95% "sikkerhed" vil falde, snarere end at give et 95% konfidensinterval for middelværdien af den variable. Man siger, at man ønsker at bestemme et 95% prædiktionsinterval (forudsigelsesinterval).

SÆTNING 4.2 ($100 \cdot (1 - \alpha)$ % prædiktionsinterval for en enkelt observation).

Et $100 \cdot (1 - \alpha)$ % prædiktionsinterval for en enkelt fremtidig observation X_{n+1} er bestemt ved

$$\bar{x} - t_{1-\frac{\alpha}{2}}(n-1) \cdot s \cdot \sqrt{1 + \frac{1}{n}} \leq \mu \leq \bar{x} + t_{1-\frac{\alpha}{2}}(n-1) \cdot s \cdot \sqrt{1 + \frac{1}{n}}.$$

Bevis: Lad X_{n+1} være en enkelt fremtidig observation. Eftersom X_{n+1} er uafhængig af de øvrige X 'er, er X_{n+1} også uafhængig af $\bar{X}$.

Variansen af differensen $\bar{X} - X_{n+1}$ er følgelig $V(\bar{X} - X_{n+1}) = V(\bar{X}) + V(X_{n+1}) = \frac{\sigma^2}{n} + \sigma^2 = \sigma^2 \left(1 + \frac{1}{n}\right)$.

Idet $E(\bar{X} - X_{n+1}) = E(\bar{X}) - E(X_{n+1}) = \mu - \mu = 0$, og en linearkombination af normalfordelte variable igen er

normalfordelt, er som anført i kapitel 3, $T = \frac{\bar{X} - X_{n+1} - 0}{S \cdot \sqrt{1 + \frac{1}{n}}}$ t -fordelt med $f = n - 1$ frihedsgrader.

Ved at foretage tilsvarende omskrivninger som i beviset for sætning 4.1, er sætningen bevist.

**Eksempel 4.2. Konfidens- og prædiktions-interval for middelværdi af normalfordeling.**

Ved fremstilling af et bestemt levnedsmiddel er det vigtigt, at et tilsætningsstof findes i levnedsmidlet i en koncentration på 8.50 (g/l). For at kontrollere dette udtager levnedsmiddelkontrollen 6 prøver af levnedsmidlet. Resultaterne var:

Måling nr	1	2	3	4	5	6
koncentration x (g/l)	8.54	7.89	8.50	8.21	8.15	8.32

Idet man antager, på baggrund af tidligere lignende målinger, at resultaterne er normalfordelte, skal man besvare følgende spørgsmål:.

- Angiv et estimat for koncentrationens middelværdi og spredning.
- Angiv et 95% konfidensinterval for koncentrationen, og vurder herudfra om kravet på 8.50 er opfyldt.
- Angiv et 95% konfidensinterval for koncentrationen, idet det antages, at man fra de tidligere målinger ved, at $\sigma = 0.25$. Vurder herudfra om kravet på 8.50 er opfyldt.
- Angiv et 95% prædiktionsinterval for en enkelt ny måling af koncentrationen.

LØSNING:

- a) TI-89: APPS, STAT/LIST\data indtastes i list1\F7, 2: T-Interval

Vælg Data\ Udfyld menuen: List = list1, C Level = 0.95

Resultater: $\bar{x} = \underline{8.268}$ og $s = \underline{0.241}$.

- b) C Int :
- $\underline{[8.02 ; 8.52]}$

Da intervallet indeholder 8.50, er kravet opfyldt, men da intervallet kun lige netop indeholder tallet 8.50, så det vil nok være rimeligt, at foretage en ny vurdering på basis af nogle flere målinger.

- c) TI-89: F7, 2: T-Interval\ i menu vælg
- $\sigma = 0.25$
- . C Int :
- $\underline{[8.07 ; 8.47]}$

Da intervallet ikke indeholder 8.50, må kravet siges ikke at være opfyldt. Det ses, at intervallet i spørgsmål b) er bredere end i spørgsmål c), i overensstemmelse med, at s er en mere usikker størrelse end det eksakte σ (t -værdien 2.57 er større end u -værdien 1.96).

- d) TI 89 har intet direkte program hertil, så man må indsætte i formelen (se nedenfor)

- a) Tabel: Vi finder ved lommeregner
- $\bar{x} = \underline{8.268}$
- og
- $s = \underline{0.241}$
- .

- b) Af tabel 4 findes
- $t_{1-\frac{\alpha}{2}}(n-1) = t_{0.975}(5) = 2.57$
- . Indsættes dette i formel 2 appendix 4.1 fås

$$\left[\bar{x} - t_{0.975}(n-1) \cdot \frac{s}{\sqrt{n}}; \bar{x} + t_{0.975}(n-1) \cdot \frac{s}{\sqrt{n}} \right] = \left[8.27 - 2.57 \cdot \frac{0.24}{\sqrt{6}}; 8.27 + 2.57 \cdot \frac{0.24}{\sqrt{6}} \right]$$

$$= [8.27 - 0.25; 8.27 + 0.25] = \underline{[8.02 ; 8.52]}.$$

- c) Af tabel 1 findes
- $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$
- . Indsættes dette i formel 1 appendix 4.1 fås

$$\left[\bar{x} - u_{0.975} \cdot \frac{\sigma}{\sqrt{n}}; \bar{x} + u_{0.975} \cdot \frac{\sigma}{\sqrt{n}} \right] = \left[8.27 - 1.96 \cdot \frac{0.25}{\sqrt{6}}; 8.27 + 1.96 \cdot \frac{0.25}{\sqrt{6}} \right]$$

$$= [8.27 - 0.20; 8.27 + 0.20] = \underline{[8.07 ; 8.47]}.$$

- d) Indsættes i formelen fra sætning 4.2 fås

$$\left[\bar{x} - t_{0.975}(n-1) \cdot s \cdot \sqrt{1 + \frac{1}{n}}; \bar{x} + t_{0.975}(n-1) \cdot s \cdot \sqrt{1 + \frac{1}{n}} \right]$$

$$= \left[8.27 - 2.57 \cdot 0.24 \cdot \sqrt{1 + \frac{1}{6}}; 8.27 + 2.57 \cdot 0.24 \cdot \sqrt{1 + \frac{1}{6}} \right] = [8.27 - 0.67; 8.27 + 0.67] = \underline{[7.60 ; 8.94]}.$$



Eksempel 4.3. Konfidensinterval for varians og spredning af normalfordeling.

En virksomhed ønsker at kontrollere med hvilken spredning en bestemt målemetode angiver saltindholdet i en opløsning. Der foretages følgende 12 målinger af en opløsning af det pågældende salt. Resultaterne var:

Måling nr	1	2	3	4	5	6	7	8	9	10	11	12
% opløsning	6.8	6.0	6.4	6.6	6.8	6.1	6.4	6.3	6.0	6.2	5.8	6.2

a) Angiv på basis af måleresultaterne et estimat for opløsningens middelværdi og spredning.

b) Angiv et 95% konfidensinterval for variansen og for spredningen.

Det antages nu, at man på forhånd havde fremstillet en 6.2 % saltopløsning som man benyttede ved forsøget. Middelværdien er derfor nu kendt .

c) Hvilket estimat kan på denne baggrund angives for opløsningens spredning.

d) Angiv på grundlag af spørgsmål c) et 95% konfidensinterval for spredningen.

LØSNING:

TI-89 har intet færdigt program.

a) Vi finder på lommeregneren $\bar{x} = 6.30$ og $s = 0.316$.

b) Formel 4 i appendix 4.1 benyttes:

$$\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \Leftrightarrow \frac{11 \cdot 0.316^2}{21.9} \leq \sigma^2 \leq \frac{11 \cdot 0.316^2}{3.82}$$

$$\Leftrightarrow 0.0502 \leq \sigma^2 \leq 0.288 .$$

For spredningen haves da $\sqrt{0.0502} \leq \sigma \leq \sqrt{0.2880} \Leftrightarrow \underline{0.2241 \leq \sigma \leq 0.5366}$.

c) Da $\mu = 6.2$ er kendt anvendes formel 3 i appendix 4.1

$$s_{\mu}^2 = \frac{(n-1)s^2 + n \cdot (\bar{x} - \mu)^2}{n} = \frac{(12-1) \cdot 0.316^2 + 12 \cdot (6.30 - 6.2)^2}{12} = \underline{0.1017} .$$

$$\text{d) } \frac{(n-1)s^2 + n \cdot (\bar{x} - \mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)} \leq \sigma^2 \leq \frac{(n-1)s^2 + n \cdot (\bar{x} - \mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)}$$

$$\Leftrightarrow \frac{11 \cdot 0.316^2 + 12(6.30 - 6.2)^2}{23.3} \leq \sigma^2 \leq \frac{11 \cdot 0.316^2 + 12(6.30 - 6.2)^2}{4.40}$$

$$\Leftrightarrow 0.0524 \leq \sigma^2 \leq 0.2773 .$$

For spredningen haves da $\sqrt{0.0524} \leq \sigma \leq \sqrt{0.2773} \Leftrightarrow \underline{0.2288 \leq \sigma \leq 0.5266}$.



OPGAVER

Opgave 4.1a (normalfordeling)

Lad der være givet 10 uafhængige observationer af en syres koncentration (i %).

12.4	10.8	12.1	12.0	13.2	12.6	11.5	11.9	12.8	12.0
------	------	------	------	------	------	------	------	------	------

- 1) Find et estimat for koncentrationens middelværdi μ og spredning σ .
- 2) Angiv et 95% konfidensinterval for μ .
- 3) Angiv et 95% prædiktionsinterval for en enkelt ny måling af koncentrationen..
- 4) Angiv et 95% konfidensinterval for μ , idet det antages, at man fra tidligere målinger ved, at $\sigma = 0.65$.

Opgave 4.1b = opgave 4.1a fortsat.

- 5) Find ud fra stikprøven 95% konfidensintervaller for koncentrationens varians og spredning.

Opgave 4.2a (normalfordeling)

Trykstyrken i beton blev kontrolleret ved at man støbte 12 betonklodser og testede dem. Resultatet var:

2216	2225	2318	2237	2301	2255	2249	2281	2275	2204	2263	2295
------	------	------	------	------	------	------	------	------	------	------	------

- 1) Find et estimat for trykstyrkens middelværdi μ og spredning σ .
- 2) Angiv et 95% konfidensinterval for μ .
- 3) Angiv et 95% prædiktionsinterval for en enkelt måling af trykstyrken på en ny betonklods.

Opgave 4.2b = opgave 4.2a fortsat.

- 4) Find et 95% konfidensinterval for trykprøvens spredning.

Opgave 4.3a (normalfordeling)

En fabrik producerer stempelringe til en bilmotor. Det vides, at stempelringenes diameter er approksimativt normalfordelt. Stempelringene bør have en diameter på 74.036 mm og en spredning på 0.001 mm. For at kontrollere dette udtog man tilfældigt 15 stempelringe af produktionen og målte diameteren. I resultaterne har man for simpelheds skyld, kun angivet de 3 sidste cifre, altså 74.0365 angives som 365. Man fandt følgende resultater

342	364	370	361	351	368	357	374	340	362	378	384	354	356	369
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

- 1) Find et estimat for ringenes diameter μ og spredning σ .
- 2) Angiv et 99% konfidensinterval for μ .
- 3) Angiv et 99% konfidensinterval for μ , når man fra tidligere målinger ved, at $\sigma = 0.001$.

Opgave 4.3b = opgave 4.3a fortsat.

4) Find ud fra stikprøven et 99% konfidensinterval for diameterens spredning.

Opgave 4.4 (normalfordeling)

En polymer produceres i batch. Viskositetsmålinger udført på hver batch gennem et stykke tid har vist, at variationen i processen er meget stabil med spredning $\sigma = 20$.

På 15 batch gav viskositetsmålingerne følgende resultater:

724	718	776	760	745	759	795	756	742	740	761	749	739	747	742
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

- 1) Find et estimat for viskositetens middelværdi μ .
- 2) Angiv et 95% konfidensinterval for μ idet man antager spredningen er 20.
- 3) Find et estimat for viskositetens spredning σ .
- 4) Angiv et 95% konfidensinterval for σ , for at kontrollere påstanden om, at $\sigma = 20$.

Opgave 4.5 (normalfordeling)

Ved en fabrikation af et bestemt sprængstof er det vigtigt, at en reaktoropløsning har en pH-værdi omkring 8.50. Der foretages 6 målinger på en bestemt reaktantopløsning. Resultaterne var:

pH	8.54	7.89	8.50	8.21	8.15	8.32
----	------	------	------	------	------	------

Den benyttede pH-målemetode antages på baggrund af tidligere lignende målinger at give normalfordelte resultater.

- 1) Angiv et estimat for opløsningens middelværdi og spredning.
- 2) Angiv et 95% konfidensinterval for pH.
- 3) Angiv et 95% konfidensinterval for pH, idet det antages, at man ved, at $\sigma = 0.25$.

Opgave 4.6 (E 80) (normalfordeling)

De 10 øverste ark papir i en pakke med printerpapir har følgende vægt

4.21	4.33	4.26	4.27	4.19	4.30	4.24	4.24	4.28	4.24
------	------	------	------	------	------	------	------	------	------

Angiv 95%-konfidensintervaller for middelværdi og spredning af papirets vægt.

Opgave 4.7 (E 80) (normalfordeling)

Til undersøgelse af alkoholprocenten i en persons blod foretages 4 uafhængige målinger, som gav følgende resultater (i %):

108	102	107	98
-----	-----	-----	----

- 1) Opstil et 95% konfidensinterval for personens alkoholkoncentration.
- 2) Opstil et 95% konfidensinterval for målemetodens spredning.

5 HYPOTSETESTNING

(ÉN STOKASTISK VARIABEL)

5.1 INDLEDNING

I afsnit 5.2 gennemgås de grundlæggende begreber vedrørende hypotesetest for én stokastisk variabel. I afsnit 5.3 gennemregnes, på basis af de oversigter der findes i appendix 5.1 til 5.4, en række eksempler. De test, som de pågældende appendix bygger på, er udledt i “Supplement til statistiske grundbegreber” afsnit 5A.

5.2 GRUNDLÆGGENDE BEGREBER

De forskellige begreber der indgår i en hypotesetest vil blive gennemgået i forbindelse med følgende eksempel.

Eksempel 5.1. Hypotesetest.

En fabrik har gennem mange år benyttet en metode, der på basis af en given mængde råmateriale gav et middeludbytte af et produceret stof på $\mu_0 = 69.2$ kg og spredningen $\sigma = 1.0$ kg. En nyansat ingeniør får til opgave at søge at forøge middeludbyttet ved en passende (billig) modifikation af procesbetingelserne. Efter en række lovende eksperimenter i laboratoriet synes opgaven at være lykkedes, men det endelige bevis herfor er, ud fra et passende antal driftsforsøg statistisk at kunne “bevise”, at middeludbyttet er blevet forøget. Ud fra kendskab til de forskellige mulige støjfaktorer antages spredningen at være uændret på 1.0 kg. Da driftsforsøgene er meget ressourcekrævende, bevilges der kun 12 delforsøg.

Der foretages 12 uafhængige delforsøg og udbyttet x målt:

Forsøg nr	1	2	3	4	5	6	7	8	9	10	11	12
x	68.8	70.7	70.3	70.1	70.7	68.7	69.2	68.9	70.0	69.6	71.0	69.1

Spørgsmål 1) Kan man ud fra disse data bevise på signifikansniveau $\alpha = 0.05$, at middeludbyttet er blevet forøget ?

Spørgsmål 2) Hvis svaret i spørgsmål 1 er bekræftende, så angiv et estimat for det nye middeludbytte, og angiv et 95% konfidensinterval herfor.

LØSNING:

Spørgsmål 1 Løsningen opdeles for overskuelighedens skyld i en række trin

1a) Definition af stokastisk variabel X .

Lad X = udbyttet ved den modificerede proces.

1b) Valg af X 's fordelingstype.

X antages at være approksimativt normalfordelt $n(\mu, 1.0)$.

1c) Opstilling af nulhypotese og alternativ hypotese.

Der opstilles en såkaldt **Nulhypotesen** $H_0 : \mu = 69.2 \text{ kg}$. Nulhypotesen **skal** indeholde en konkret påstand (her et lighedstegn). Påstanden er, at modifikationen ingen (nul) virkning har. Der opstilles endvidere en **alternativ hypotese** $H : \mu > 69.2 \text{ kg}$. Den alternative hypotese skal så vidt muligt indeholde det, der ønskes bevist. I dette tilfælde ønskes vist, at middeludbyttet er vokset, dvs. $\mu > 69.2 \text{ kg}$.

Testen kaldes en **ensidet test** i modsætning til en tosidet test :

$$H_0 : \mu = 78.8 \text{ kg} \text{ contra } H : \mu \neq 78.8 \text{ kg},$$

hvor vi blot ønsker at vise, at middeludbyttet har ændret sig.

1d) Angivelse af testens signifikansniveau.

Hvis stikprøvens gennemsnit $\bar{x}$ er meget større end 69.2 kg (måske helt op mod 100 kg), så er der stor sandsynlighed for at udbyttet er steget. Man siger så, at **nulhypotesen forkastes**, eller at $\bar{x}$ ligger i forkastelsesområdet (se figur 5.1). Hvis derimod $\bar{x}$ kun ligger lidt over 69.2 kg, så kan det skyldes tilfældige udsving, og man kan ikke med nogen stor sikkerhed konkludere, at udbyttet er steget. Man siger, at **nulhypotesen accepteres**, eller at $\bar{x}$ ligger i acceptområdet.

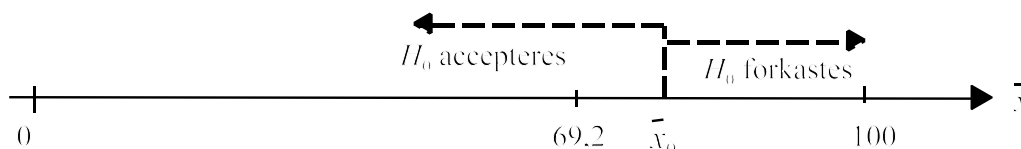


Fig. 5.1 Accept- og forkastelsesområde

Lad $\bar{x}_0$ være grænsen mellem acceptområdet og forkastelsesområdet. $\bar{x}_0$ skal bestemmes sådan, at **forudsat** $H_0 : \mu = 69.2 \text{ kg}$ er sand, så er det yderst usandsynligt, at en stikprøves gennemsnit $\bar{x}$ vil komme til at ligge i forkastelsesområdet. Hvis stikprøvens gennemsnit alligevel ligger i forkastelsesområdet, må det være forudsætningen H_0 der er forkert, d.v.s. middeludbyttet må være blevet større. Det er naturligvis ikke entydigt bestemt, hvad det vil sige, at noget er yderst usandsynligt. Generelt siges, at den såkaldte **P - værdi** (probability value) = $P(\bar{X} \geq \bar{x}_0)$ skal ligge under et tal α , som kaldes testens **signifikansniveau**. Signifikansniveauet kan vælges vilkårligt, men sættes sædvanligvis til enten 5%, 1% eller 0.1%.

I dette eksempel er det sat til 5%.

1e) **Beregning af P - værdi og konklusion**

Metode 1:

Gennemsnittet af de 12 resultater giver $\bar{x} = 69.76$ kg.

Antages som før at nulhypotesen $H_0: \mu = 69.2$ kg er sand, dvs. $\bar{X}$ antages normalfordelt med middelværdi $\mu_0 = 69.2$ og spredning $\frac{\sigma}{\sqrt{n}} = \frac{1.0}{\sqrt{12}} = 0.2887$, så fås

P -værdi = $P(\bar{X} \geq 69.76) = \text{normCdf}(69.76, \infty, 69.2, 1/\sqrt{12}) = 0.0262$ (se evt. appendix 5.2)

Da P -værdi = $2.62\% < 5\%$ forkastes H_0 ,

Vor konklusion er derfor, at vi har et (svagt) statistisk bevis for, at den modificerede proces giver et større middeludbytte.

Metode 2:

Alternativt kunne vi have benyttet TI-89 testfunktion:

APPS\STAT\LIST\ data indtastes i list1 \ F6, 1: Z-Test

Menu udfyldes: $\mu_0 = 69.2$, $\sigma = 1$, list=list1, Alternate Hyp: $\mu > \mu_0$, Calculate

Vi får P -værdi = 0.0265 , dvs. samme konklusion som før.

Tabel: Ved at foretage transformationen $U = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ føres problemet over til den normerede normalfordeling.

I tabel 2 ses de relevante fraktiler for U -fordelingen. Heraf ses, at forkastelsesområdet for $\alpha = 5\%$ er $U > u_{0.95} = 1.645$ (se figur 5.2). Valgte vi $\alpha = 1\%$ er forkastelsesområdet bestemt ved $U > u_{0.99} = 2.326$. (se eventuelt appendix 5.2)

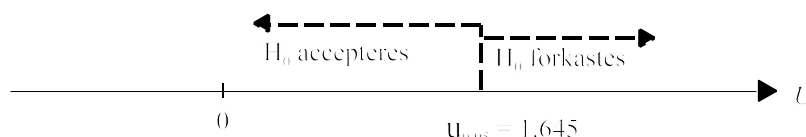


Fig. 5.2 Accept- og forkastelsesområde.

Gennemsnittet af de 12 resultater giver $\bar{x} = 69.76$ kg.

Vi får $U = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{69.76 - 69.2}{\frac{1.0}{\sqrt{12}}} = 1.94$.

Heraf ses, at H_0 forkastes på 5% niveau, mens vi ikke kan forkaste på 1% niveau.

Spørgsmål 3. Udbyttet kan i middel forventes at være ca. $\bar{x} = 69.76$ kg .

TI-89: APPS, STAT/LIST\data indtastes i list1\ F7, 2: T-Interval

C Int : [69.19;70.32]

Tabel: Et 95% konfidensinterval fås af formel 1 appendix 4.1

$$\bar{x} \pm u_{0.975} \cdot \frac{\sigma}{\sqrt{n}} = 69.76 \pm 1.96 \cdot \frac{1.0}{\sqrt{12}} = 69.76 \pm 0.566 \quad \underline{\underline{[69.19;70.32]}}.$$

Da testen er ensidet, ville en anden mulighed være at sige, at 95% konfidensintervallet var

$$\left[\bar{x} - u_{0.95} \cdot \frac{\sigma}{\sqrt{n}}; \infty \right] = \left[69.76 - 1.65 \cdot \frac{1.0}{\sqrt{12}}; \infty \right] = [69.28; \infty].$$

Graden af forkastelse.

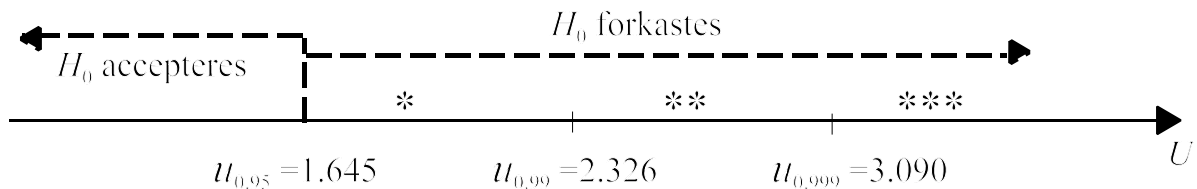


Fig. 5.3 Illustration af signifikans

Som nævnt ovenfor siger man, at man forkaster nulhypotesen H_0 på 5% niveau, hvis man kan forkaste på 5% niveau, men ikke på 1% niveau. Andre vendinger er “svagt eksperimentelt bevis mod H_0 ”, eller “1-stjernet signifikans mod H_0 ” (se figur 5.3).

Tilsvarende siges, at man forkaster nulhypotesen H_0 på 1% niveau, hvis man kan forkaste på 1% niveau, men ikke på 0.1% niveau. Andre vendinger er “eksperimentelt bevis mod H_0 ”, eller “2-stjernet signifikans mod H_0 ”.

Endelig siger man, at man forkaster nulhypotesen H_0 på 0.1% niveau. Andre vendinger er “stærkt eksperimentelt bevis mod H_0 ”, eller “3-stjernet signifikans mod H_0 ”.

Fejl af type I og type II: Ved enhver test kan der være to typer fejl, hvoraf vi hidtil kun har taget hensyn til den ene type. For bedre at forstå problemstillingen vil vi se på følgende skema.

		Beslutning	
		H_0 accepteres	H_0 forkastes
Forudsætning	H_0 er sand	Rigtig beslutning	Forkert beslutning Type I fejl
	H_0 er falsk	Forkert beslutning Type II fejl	Rigtig beslutning

Det må være et krav til en god test, at der kun er en lille sandsynlighed for at begå en fejl af type I eller type II.

I eksempel 5.1 ville en type I fejl være, hvis man konkluderer, at den modificerede proces giver et større udbytte, selv om det ikke er tilfældet. Virksomheden bruger måske millionbeløb på at omlægge produktionen, og det er ganske forgæves.

En type II fejl ville være, at man ikke opdager, at den modificerede proces giver et større udbytte. Dette er naturligvis uheldigt, men hvis det skyldes, at forbedringen ikke blev opdaget, fordi den er ganske ringe, har det muligvis ingen praktisk betydning.

Hvis en test har signifikansniveau α så vides, at $P(\text{type I fejl}) \leq \alpha$, dvs. **forkastes H_0** , så er vi rimelig sikre på, at have foretaget en korrekt beslutning.

Hvis vi **accepterer H_0** er det blot udtryk for, at vi ikke kan forkaste (svag konklusion: " H_0 frikendes på grund af bevisets stilling"). Man kan have begået en type II fejl, dvs. accepteret en falsk nulhypotese.

Dimensionering af forsøg (vælg stikprøvestørrelse n).

Lad os antage, at virksomheden i eksempel 5.1 finder, at hvis stigningen i udbyttet ved den modificerede proces er mindre end $\Delta = 0.5$ kg, så har det ingen praktisk interesse ($\Delta = 0.5$ kg er bagatelgrænsen), og derfor gør det intet, hvis man ikke opdager det (begår en type II fejl). Hvis derimod stigningen Δ er større end 0.5 kg, så har det stor betydning, og sandsynligheden for at begå en type II fejl må derfor være lille. Lad os sætte den til højst $\beta = 10\%$.

Problemet er nu, hvor stor en stikprøvestørrelse n (antallet af delforsøg) der skal udføres, for at ovennævnte krav er opfyldt. En sådan vurdering kaldes en **dimensionering** af forsøget. Udfører man det ud fra en dimensionering nødvendige antal forsøg, vil en accept af nulhypotesen nu betyde, at nok kan udbyttet være steget, men ikke så meget, at det har praktisk interesse.

SÆTNING 5.1 (dimensionering). Lad X være en normalfordelt variabel med ukendt middelværdi μ og kendt spredning σ . Lad en ensidet test med signifikansniveau α have nulhypotesen $H_0: \mu \leq \mu_0$ med den alternative hypotese $H: \mu > \mu_0$, og lad Δ være den mindste stigning i middelværdien, som har praktisk interesse. I tilfælde af en stigning i udbyttet på mindst Δ skal sandsynligheden for at acceptere H_0 være mindre end et opgivet tal β ($P(\text{type II fejl}) \leq \beta$).

Stikprøvestørrelsen n bestemmes da som det mindste hele tal n , for hvilket

$$n \geq \left(\frac{u_{1-\alpha} + u_{1-\beta}}{\frac{\Delta}{\sigma}} \right)^2.$$

Bevis: H_0 accepteres på signifikansniveau α , hvis $\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq u_{1-\alpha} \Leftrightarrow \bar{x} \leq \mu_0 + u_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$.

Er den sande middelværdi $\mu_0 + \Delta$, er $U = \frac{\bar{x} - (\mu_0 + \Delta)}{\frac{\sigma}{\sqrt{n}}}$ normalfordelt med middelværdi 0 og

spredning 1. Vi har derfor, at forudsat $\mu = \mu_0 + \Delta$ er

$$P(\text{type II fejl}) = P\left(\bar{x} \leq \mu_0 + u_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}\right) = \Phi\left(\frac{\mu_0 + u_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} - (\mu_0 + \Delta)}{\frac{\sigma}{\sqrt{n}}}\right) = \Phi\left(u_{1-\alpha} - \frac{\Delta}{\frac{\sigma}{\sqrt{n}}}\right) \leq \beta$$

Heraf fås $u_{1-\alpha} - \frac{\Delta}{\frac{\sigma}{\sqrt{n}}} \leq u_\beta$. Da $u_\beta = -u_{1-\beta}$ fås

$$u_{1-\alpha} - \frac{\Delta\sqrt{n}}{\sigma} \leq -u_{1-\beta} \Leftrightarrow u_{1-\alpha} + u_{1-\beta} \leq \frac{\Delta\sqrt{n}}{\sigma} \Leftrightarrow \sqrt{n} \geq \frac{(u_{1-\alpha} + u_{1-\beta})\sigma}{\Delta}, \text{ hvoraf formelen fremkommer.}$$



Eksempel 5.2. Dimensionering.

Inden man i eksempel 5.1 begyndte at lave de dyre delforsøg, vil ingeniøren gerne have en vurdering af, hvor mange driftsforsøg der er nødvendige, når det vides, at det først er økonomisk rentabelt at gå over til den nye metode, hvis middeludbyttet er steget med mindst 0.5 kg.

- 1) Find stikprøvestørrelsen n , i det tilfælde, hvor $\Delta = 0.5$ kg og $\beta = 10\%$. Det antages stadig, at $\sigma = 1.0$ kg og signifikansniveauet er $\alpha = 5\%$.

Lad n være den i spørgsmål 1 fundne stikprøvestørrelse.

2) Idet der udføres n delforsøg skal man besvare følgende spørgsmål:

- Hvilken konklusion kan drages, hvis man finder, at nulhypotesen $H_0: \mu \leq 69.2$ kg kan forkastes? Er der i så tilfælde behov for yderligere beregninger?
- Hvilken konklusion kan drages, hvis man finder, at nulhypotesen $H_0: \mu \leq 69.2$ kg kan accepteres? Er der i så tilfælde behov for yderligere beregninger?

LØSNING:

1) Af sætning 5.1 (eller formel i appendix 5.2) fås:

$$n \geq \left(\frac{u_{0.95} + u_{0.90}}{\frac{0.5}{1.0}} \right)^2 = \left(\frac{1.645 + 1.282}{0.5} \right)^2 = 34.26, \text{ dvs. } \underline{\underline{n = 35}}.$$

- Vi er (afhængig af P -værdien) nogenlunde sikre på, at udbyttet er steget. Hvor meget må afklares ved at lave et 95% konfidensinterval (og eventuelt flere forsøg).
- Udbyttet kan være steget (igen kan P -værdien sige lidt om muligheden herfor), men vi er rimeligt sikre på, at en eventuel stigning ikke har praktisk interesse.



OC-kurve. For at få et indtryk af, hvor stor sandsynligheden er for fejl af type II tegnes ofte en såkaldt OC kurve (Operating Characteristic).

SÆTNING 5.2 (funktionsudtryk for OC kurve). Lad X være en normalfordelt variabel med ukendt middelværdi μ og kendt spredning σ . Lad en ensidet test med signifikansniveau α have nulhypotesen $H_0: \mu \leq \mu_0$, og lad $a(\mu)$ være sandsynligheden for at acceptere H_0 hvis den sande middelværdi er $\mu > \mu_0$ (altså $a(\mu) = P(\text{type II fejl})$).

$$\text{Der gælder da } a(\mu) = \Phi \left(u_{1-\alpha} + \frac{\mu_0 - \mu}{\frac{\sigma}{\sqrt{n}}} \right).$$

Bevis:

H_0 accepteres på signifikansniveau α , hvis

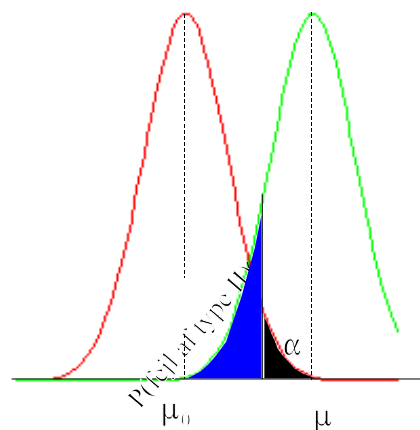
$$\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq u_{1-\alpha} \Leftrightarrow \bar{x} \leq \mu_0 + u_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}.$$

Er den sande middelværdi μ , er $U = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ normalfordelt med

middelværdi 0 og spredning 1.

Vi har derfor, at

$$P(\text{type II fejl}) = P\left(\bar{X} \leq \mu_0 + u_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}\right)$$



$$= \Phi \left(\frac{\mu_0 + u_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} - \mu}{\frac{\sigma}{\sqrt{n}}} \right) = \Phi \left(u_{1-\alpha} + \frac{\mu_0 - \mu}{\frac{\sigma}{\sqrt{n}}} \right) \quad \text{d.v.s.} \quad a(\mu) = \Phi \left(u_{1-\alpha} + \frac{\mu_0 - \mu}{\frac{\sigma}{\sqrt{n}}} \right).$$



Eksempel 5.3. OC-kurve.

Inden man i eksempel 5.1 begyndte at lave de dyre delforsøg vil ingeniøren gerne have en vurdering af, om 12 tildelte driftsforsøg giver mulighed for sikre konklusioner, når det vides, at det først er økonomisk rentabelt at gå over til den nye metode, hvis middeludbyttet er steget med mindst 0.5 kg.

- 1) Der ønskes tegnet en OC kurve, når signifikansniveauet er valgt til 5%.
- 2) Hvis testen giver en accept af nulhypotesen, er der så en rimelig grund til at tro, at man ikke skal gå over til den nye metode?
- 3) Hvis testen giver en forkastelse af nulhypotesen, er der så en rimelig grund til at tro, at man skal gå over til den nye metode?
- 4) Hvad bliver svarene på spørgsmål 2 og 3, hvis kravet til stigning i middeludbyttet stiger fra 0.5 kg til 1 kg

LØSNING:

- 1) Idet $\mu_0 = 69.2$ kg, $\sigma = 1.0$ kg og $n = 12$ fås af sætning 5.2, at

$$a(\mu) = \Phi \left(1.645 + \frac{69.2 - \mu}{\frac{1.0}{\sqrt{12}}} \right) = \Phi(1.645 + 3.4641 \cdot (69.2 - \mu)).$$

Ved indsættelse fås følgende skema

μ	69.2 kg	69.4 kg	69.6 kg	69.8 kg	70.0 kg	70.2 kg	70.4 kg
$a(\mu)$	95%	83.0%	60.2%	33.2%	13.0%	3.4%	0.6%

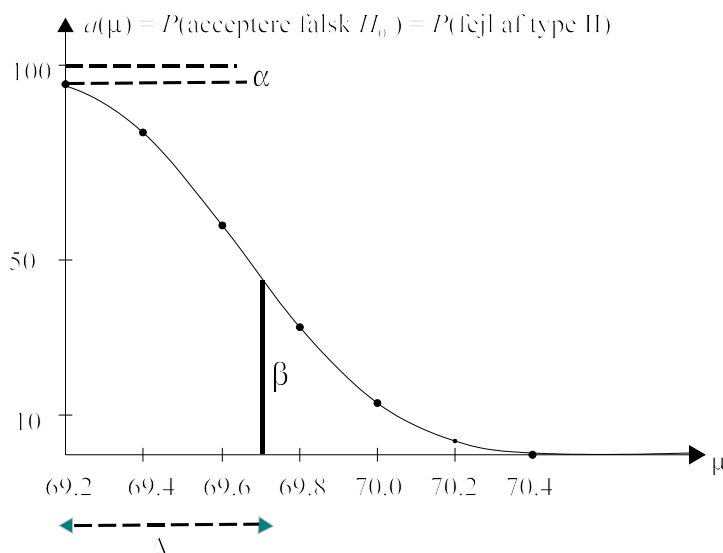


Fig.5.4 OC kurve

- 2) Af kurven ses, at hvis μ er 69.7 kg, vil sandsynligheden for at begå en type II fejl være ca. 50%. En accept af nulhypotesen giver derfor ingen mulighed for at konkludere, at middeludbyttet ikke er steget mere end 0.5 kg.
- 3) Der kan ikke svares umiddelbart herpå. En beregning af konfidensinterval er nødvendig (jævnfør også

- besvarelsen i eksempel 5.1)
- 4a) Af kurven ses, at hvis μ er 70.2 kg, vil sandsynligheden for at begå en type II fejl være ca. 3.4%. En accept af nulhypotesen giver derfor en rimelig grund til at konkludere, at middeludbyttet ikke er steget mere end 1 kg.
- 4b) Samme konklusion som i spørgsmål 3).



5.3 Eksempler på hypotesetest

De grundlæggende begreber for hypotesetest blev grundigt gennemgået i forrige afsnit. Det er derfor i dette afsnit kun nødvendigt at give nogle få eksempler på andre tilsvarende testtyper. Beregningerne bygger på de formler, der er angivet i de oversigter over hypotesetest, der findes i appendix 5.1 til 5.4.

Eksempel 5.4. Tosidet test for hypotese om middelværdi.

Ved fremstilling af et bestemt levnedsmiddel er det vigtigt, at et tilsætningsstof findes i levnedsmidler i en koncentration på 8.40 (g/l). En afvigelse i koncentrationens værdi på mere end 0.35 skal opdages med en sandsynlighed på 0.90, dvs. $\beta = P(\text{type II fejl}) = 0.10$ for $\Delta = 0.35$. Signifikansniveau sættes til $\alpha = 0.05$.

På baggrund af tidligere lignende målinger antages resultaterne at være normalfordelte med et kendt $\sigma = 0.25$.

- 1) Vis at ovenfor nævnte specifikationer kan opnås med 6 målinger af koncentrationen.
- 2) For at kontrollere om tilsætningsstoffet har en koncentration på ca. 8.40, udtager levnedsmiddelkontrollen 6 prøver af levnedsmidler. Resultaterne var:

Måling nr	1	2	3	4	5	6
Koncentration x (g/l)	8.54	7.89	8.50	8.21	8.15	8.32

Det ønskes på denne baggrund undersøgt om koncentrationen har den ønskede værdi.

LØSNING:

- 1) Lad X være koncentrationen af tilsætningsstoffet i levnedsmidlet.

Det antages, at X er normalfordelt $n(\mu, \sigma)$ hvor μ er ukendt, og $\sigma = 0.25$.

Da det både er uønsket, at koncentrationen er for lille og at den er for stor, bliver nulhypotesen $H_0: \mu = 8.4$ mod $H: \mu \neq 8.4$, dvs. vi har en tosidet test.

Bemærk, at selv om man vel egentlig hellere ville bevise, at koncentrationen er 8.4 og derfor helst ville have denne påstand i den alternative hypotese, er dette ikke muligt, da nulhypotesen skal indeholde et lighedstegn.

$$\text{Af appendix 5.2 fås } n \geq \left(\frac{u_{0.975} + u_{0.90}}{\frac{\Delta}{\sigma}} \right)^2 = \left(\frac{1.960 + 1.282}{\frac{0.35}{0.25}} \right)^2 = 5.36, \quad \text{dvs. } \underline{\underline{n=6}}.$$

2) Metode 1:

Vi finder $\bar{x} = 8.268$.

I følge appendix 5.2 fås nu

$$P\text{-værdien} = P(\bar{X} \leq 8.286) = \text{normCdf}(-\infty, 8.268, 8.40, 0.25/\sqrt{6}) = 0.0979 = 9.8\%$$

Da $P\text{-værdi} > 0.025$ accepteres nulhypotesen, dvs. vi må konkludere, at selv om det ikke er bevist, at koncentrationen er 8.4 g/l, så er en eventuel afvigelse ikke af praktisk betydning.

Metode 2:

Alternativt kunne vi have benyttet TI-89 testfunktion:

APPS\STAT\LIST\ data indtastes i list1 \ F6, 2: Z-Test

Menu udfyldes : $\mu_0 = 8.4$, $\sigma = 0.25$ list=list1, Alternate Hyp: $\mu \neq \mu_0$, Calculate

Vi får $P\text{-værdi} = 0.1970 = 19.7\%$.

Denne værdi er den dobbelte af $P\text{-værdien}$ ved metode 1, da TI-89 beregner begge "haler"

Vi skal altså her sammenligne med 5%. Da $P\text{-værdi} > 5\%$ fås samme konklusion som før.

$$\text{Tabel: Ifølge appendix 5.2 fås } u = \frac{(\bar{x} - \mu_0) \cdot \sqrt{n}}{\sigma} = \frac{(8.268 - 8.40) \cdot \sqrt{6}}{0.25} = -1.29.$$

Da $u_{0.975} = 1.96$ accepteres nulhypotesen



Eksempel 5.5. Ensided t-test .

En virksomhed bliver af miljøkontrollen pålagt at formindske indholdet i sit spildevand af et stof A, der mistænkes for at kunne forurene grundvandet. Indholdet af stoffet A i spildevandet skal under 1.7 mg/l, og miljøkontrollen henviser til en ny metode, som burde kunne formindske indholdet til det ønskede niveau. For at vurdere den nye metode ønskes foretaget en række delforsøg.

- 1) Hvor mange forsøg skal der mindst foretages, hvis $\alpha = 5\%$, $\beta = 10\%$, $\Delta = 0.10$ mg/l og et overslag over hvor stor σ er sætter denne til 0.15 mg/l.
- 2) Uanset resultatet af dimensioneringen i spørgsmål 1), er der kun praktiske muligheder for at lave 15 delforsøg (ét forsøg om dagen i 3 uger). Følgende værdier af indholdet af A fandtes (i mg/l).

1.55	1.66	1.98	1.37	1.60	1.62	1.53	1.79	1.72	1.57	1.70	1.40	1.77	1.58	1.84
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

Kan man ud fra disse data bevise på signifikansniveau $\alpha = 0.05$, at indholdet af A ved benyttelse af den ny metode er under 1.7 mg/l.

- 3) Angiv et estimat for middelinholdet af stoffet A ved den nye metode, og et 95% tosidet konfidensinterval for det forventede indhold af A .

LØSNING:

Lad X = indhold af A (i mg/l) efter benyttelse af den ny metode.

X antages normalfordelt $n(\mu, \sigma)$, hvor såvel μ som σ er ukendte.

Da indholdet af stoffet A ønskes formindsket, bliver

nulhypotesen $H_0: \mu = 1.7$ mg/l mod $H: \mu < 1.7$ mg/l, dvs. vi har en ensidet test.

1) Da σ ikke er kendt (kun et løst skøn kendes), er testen en t -test.

Idet $\frac{\Delta}{\sigma} \approx \frac{0.10}{0.15} = 0.667$, fås af dimensioneringstabel 8 for $\alpha = 5\%$ og $\beta = 10\%$, at $n = 22$.

(man finder det tal i tabellen, som er tæt ved 0.667, idet man dog altid runder n opad).

Den ønskede dimensionering kræver altså 22 forsøg.

2) Metode 1:

Vi benytter TI-89 testfunktion:

APPS\STAT\LIST\ data indtastes i list1 \ F6, 2: T-Test

Menu udfyldes : $\mu_0 = 1.7$, list=list1, Alternate Hyp: $\mu < \mu_0$, Calculate

Vi får P -værdi = 0.1058 = 10.58%.

Da P -værdi $> 5\%$ accepteres nulhypotesen, dvs. man kan ikke vise, at den nye metode bringer middelkoncentrationen ned på det ønskede niveau.

Da dimensioneringen ikke er korrekt, og vi trods alt er tæt på forkastelse, er det dog muligt, at vi begår en type II fejl.

Metode 2:

Vi finder $\bar{x} = 1.645$ mg/l og $s = 0.162$ mg/l.

Af appendix 5.1 fås $t = \frac{(1.645 - 1.70)\sqrt{15}}{0.162} = -1.31$.

P -værdien = $P(T \leq -1.31) = tCdf(-\infty, -1.31, 14) = 0.1051$

Da P -værdi $> 5\%$ fås samme konklusion som før.

3) Af appendix 4.1 formel 2 fås, at et estimat for middeludbyttet er $\bar{x} = 1.645$ mg/l, og et 95% konfidensinterval:

$$\bar{x} \pm t_{1-\frac{\alpha}{2}}(n-1) \cdot \frac{s}{\sqrt{n}} = 1.645 \pm t_{0.975}(14) \cdot \frac{0.1620}{\sqrt{15}} = 1.645 \pm 0.0899,$$

dvs. [1.56; 1.73] mg/l.

**Eksempel 5.6. Ensided χ^2 -test.**

En fabrikant af læskedrikke har købt en automatisk "påfyldningsmaskine". Ved købet af maskinen har man betinget sig, at rumfanget af den påfyldte væske i middel skal have en spredning, der ikke overstiger 0.20 ml. Efter kort tids anvendelse får man mistanke om, at spredningen er for stor. Mange klager over underfyldte flasker. Derfor foretages en kontrol, hvor man tilfældigt udtager 20 flasker med læskedrik, og måler rumfanget af væsken i flasken. Det viser sig, at stikprøvens spredning er $s = 0.24$ ml.

Med et signifikansniveau på 5% er det da et statistisk bevis for, at den nye maskine ikke opfylder det stillede krav?.

LØSNING:

Lad X = rumfang af drik i flaske.

X antages normalfordelt $n(\mu, \sigma)$, hvor såvel μ som σ er ukendte.

Man ønsker at teste hypoteserne $H_0: \sigma = 0.2$ imod $H: \sigma > 0.2$, eller udtrykt ved variansen σ^2 :
 $H_0: \sigma^2 = 0.2^2$ mod $H: \sigma^2 > 0.2^2$.

Ifølge appendix 5.3 ses, at vi skal beregne teststørrelsen χ^2 , hvor

$$\chi^2 = \frac{(n-1) \cdot s^2}{\sigma_0^2} \text{ dvs. i det foreliggende tilfælde } \chi^2 = \frac{(20-1) \cdot 0.24^2}{0.2^2} = 27.36.$$

$$P\text{-værdi} = P(Q \geq 27.36) = \text{chiCdf}(27.36, \infty, 19) = 0.0965 = 9.65\%$$

Da P -værdien er over 5 %, accepteres H_0 , dvs. det er ikke påvist, at spredningen ved påfyldningen er for stor, men der er dog nær ved at være signifikans.

Hvis man inden forsøgets udførelse havde ønsket en dimensionering af forsøget, således at for en forøgelse af spredningen på $\Delta = 0.03$ ville $P(\text{fejl af type II}) = \beta \leq 10\%$, er dette kun muligt ved hjælp af et passende statistikprogram. Benyttes Statgraphics fås, at man skulle have målt på 218 flasker.



5.4 Parvise Observationer .

I eksempelvis medicinske forsøg, hvor man anvender mennesker eller dyr til undersøgelse af om en behandling har virkning, er der stor forskel på den måde den enkelte “forsøgsenhed” (menneske) reagerer på behandlingen. Stikprøvens spredning bliver derved stor, hvilket gør det nødvendigt at have en stor stikprøvestørrelse for at kunne drage en sikker konklusion. I disse “før og efter” situationer, hvor dataene for hver forsøgsenhed er naturligt “parrede”, sker den statistiske analyse ved, at vi for hver forsøgsenhed (menneske) ser på differencen $d = x_{\text{før}} - x_{\text{efter}}$

Lad os illustrere det ved følgende eksempel.

Eksempel 5.6. Parvise observationer.

Et skofirma vil undersøge om der er forskel på slidstyrken af 2 typer skosåler, lavet af forskellige syntetiske materialer A og B.

Først overvejes følgende plan:

Efter lodtrækning blandt 10 drenge vil 5 af drengene få et par sko, som har såler af type A og de 5 andre af type B. Efter et bestemt antal uger måles så sliddet på sålerne.

Imidlertid står det firmaet klart, at det er meget forskelligt, hvor meget drenge bevæger sig, så sliddet fra dreng til dreng bliver alene af den grund meget forskelligt. Man må derfor forvente en stor spredning på resultaterne. Denne store "støj" vil sandsynligvis bevirke, at man med kun 10 gentagelser ikke vil kunne konstatere nogen forskel, også selv om der måske er en ret betydelig forskel.

En forsøgsplan, der kan fjerne denne støj, er følgende: Hver dreng får en sko med den ene type sål og en sko med den anden type sål. Ved lodtrækning fastlægges, om det er højre eller venstre fod der skal have type A. Man siger, at man laver et randomiseret blokforsøg, hvor drengene opfattes som blokke og vi har randomiseret indenfor blokkene. Ved for hver blok (dreng), at betragte differensen mellem sliddet på de to sko, har vi fjerner variationen mellem blokkene (drenge) og tilbage er variationen indenfor blokkene.

Efter en passende tid målte man sliddet x på sålerne. Resultaterne var følgende:

Dreng nr	1	2	3	4	5	6	7	8	9	10
x_A	13.2	8.1	10.8	14.2	10.7	6.7	9.6	10.9	8.9	13.0
x_B	14.0	8.8	11.2	14.1	11.8	6.4	10.0	11.4	9.4	13.6

Undersøg på grundlag af disse tal, om der er forskel på slidstyrken.

LØSNING:

Da det er et blokforsøg, med drenge som blokke, skal man betragte forskellene mellem sliddet på sålerne som *samme* dreng bærer. Disse differenser er jo rensset for den støj, der stammer fra, at drengene slider forskelligt på sålerne.

Vi får følgende tabel:

Dreng nr	1	2	3	4	5	6	7	8	9	10
x_A	13.2	8.1	10.8	14.2	10.7	6.7	9.6	10.9	8.9	13.0
x_B	14.0	8.8	11.2	14.1	11.8	6.4	10.0	11.4	9.4	13.6
$x_B - x_A$	0.8	0.7	0.4	-0.1	1.1	-0.3	0.4	0.5	0.5	0.6

Det ses af differenserne, at der tilsyneladende er et større slid på såler af typen B end på såler af typen A.

Lad $D = x_B - x_A$.

D antages normalfordelt $n(\mu, \sigma)$, hvor såvel μ som σ er ukendte.

Da vi ønsker at teste om der er forskel, bliver testen en tosidet t -test.

Nulhypotese $H_0: \mu = 0$. Alternativ hypotese $H: \mu \neq 0$.

Metode 1:

Vi benytter TI-89 testfunktion:

APPS\STAT\LIST\ differens-data indtastes i list1 \ F6, 2: T-Test

Menu udfyldes : $\mu_0 = 0$, list=list1, Alternate Hyp: $\mu \neq \mu_0$, Calculate

Vi får P -værdi = 0.006 = 0.6%.

Da P -værdi < 5% forkastes H_0 (2 stjernet), dvs. der er forskel på slidstyrken.

Metode 2:

Vi får, at gennemsnittet af differenserne er $d = 0.46$ og spredningen er $s_d = 0.409$.

Af appendix 5.1 fås $t = \frac{(\bar{d} - 0)\sqrt{n}}{s_d} = \frac{0.460\sqrt{10}}{0.409} = 3.567$

P -værdien = $P(T \geq 3.567) = tCdf(3.567, \infty, 9) = 0.0030 = 0.3\%$

Da P -værdi < 2.5% fås samme konklusion som før.

Et 95% konfidensinterval for differensen er

$$\bar{d} \pm t_{1-\frac{\alpha}{2}}(n-1) \cdot \frac{s_d}{\sqrt{n}} = 0.46 \pm t_{0.975}(9) \cdot \frac{0.409}{\sqrt{10}} = 0.46 \pm 0.29 \quad \text{dvs.} \quad \underline{\underline{[0.17; 0.75]}}.$$



OPGAVER

Opgave 5.1a (*U*-test for middelværdi).

Et levnedsmiddel ("corned beef") forhandles i pakker på 100 g. Ved fabrikationen tilsættes traditionelt et konserveringsmiddel B (nitrit). Da man har mistanke om, at B anvendt i større mængder kan have uønskede bivirkninger, må der højst tilsættes 2.5 mg B pr. 100 g.

Fabrikanten reklamerer med, at der i middel højst er 2 mg B pr. pakke. En konkurrent tvivler herpå, og vil teste påstanden. Der købes i forskellige butikker i alt 36 pakker, og indholdet af B blev målt. Man fandt et gennemsnit af B på $\bar{x} = 2.10$ mg med et estimat på spredningen på $s = 0.30$ mg. Idet stikprøvestørrelsen er over 30, kan man anvende en *U*-test, dvs. antage, at $\sigma = s = 0.30$.

- 1) Kan man ud fra disse data bevise på signifikansniveau $\alpha = 0.01$, at reklamen lyver.
- 2) Angiv et estimat for middelindholdet μ .
- 3) Angiv et 95% konfidensinterval for middelindholdet μ .
- 4) Find det mindste antal pakker, der skal købes, for at man ved en stigning af middelindholdet af nitrat på $\Delta = 0.1$ højst har, at $P(\text{type II fejl}) = \beta = 5\%$. ($\alpha = 0.01$ og $\sigma = 0.30$).

Opgave 5.1b^T (type II fejl) Opgaven er en fortsættelse af opgave 5.1a.

Antages pakkernes middelindhold af B at være $\mu = 2.10$ mg, hvor stor er så sandsynligheden for fejlagtigt at acceptere fabrikantens påstand om, at middelindholdet er 2.00 mg ($P(\text{type II fejl})$). α er stadig 0.01.

Opgave 5.1c (*t* - test)

Opgaven er en fortsættelse af opgave 5.1a.

Svar på spørgsmål 1), 2) og 3) i opgave 5.1a, idet der nu benyttes en *t* - test fremfor en *U* - test.

Opgave 5.2a (*U* - test for middelværdi).

Et flyselskab overvejer at lukke en flyrute, såfremt μ = "middelværdien af antal solgte pladser pr. afgang" er under 60. På de sidste $n = 100$ afgang er der i gennemsnit solgt $\bar{x} = 58.0$ pladser med en standardafvigelse på $s = 11.0$ pladser. Da stikprøvestørrelsen n er over 30, kan man anvende en *U*-test, dvs. antage at $\sigma = s = 11.0$ pladser.

- 1) Kan man ud fra disse data bevise på signifikansniveau $\alpha = 0.05$, at der i middel er solgt under 60 pladser pr. afgang? (Husk at anføre: Hvad X er. Antagelser. Nulhypotese. Beregninger. Konklusion.).
- 2) Angiv et estimat $\tilde{\mu}$ for middelværdien μ .
- 3) Angiv et 95% konfidensinterval for middelværdien μ .
- 4) Dimensionering. Bestem den nødvendige stikprøvestørrelse n , for at man ved en forskel i antal solgte pladser på $\Delta = 2.0$ højst har, at $P(\text{type I fejl}) = \alpha = 0.05$ og $P(\text{type II fejl}) = \beta = 0.10$ ($\sigma = 11.0$ pladser).

Hypotesetestning (1 variabel)

Opgave 5.2b (type II fejl)

Opgaven er en fortsættelse af opgave 5.2a.

Antag, at antal solgte pladser er 58. Hvor stor er så sandsynligheden for fejlagtigt at acceptere, at flyruten er rentabel (i middel mindst 60 besat). α er stadig 0.05.

Opgave 5.2c (t - test)

Opgaven er en fortsættelse af opgave 5.2a.

Svar på spørgsmål 1) i opgave 5.2a, idet der nu benyttes en t - test fremfor en U - test.

Opgave 5.3 (t - test for middelværdi).

Under produktionen forekommer blandt en fabriks affaldsprodukter 1,5 mg/l af et stof A, som i større mængder kan være kræftfremkaldende. Man håber ved en ny og mere kostbar metode, at formindske indholdet af det pågældende stof.

1) Ved en række kontrolmålinger efter tilsætning af additivet fandtes følgende resultater (i mg/l)

1.12	1.47	1.35	1.27	1.17	1.26	1.83	1.10	1.39	1.25	1.44	1.14
------	------	------	------	------	------	------	------	------	------	------	------

Test på 5% niveau, om målingerne beviser, at der er sket en formindskelse af middelindholdet af stoffet A. (Husk altid at anføre: Hvad X er. Antagelser. Nulhypotese. Beregninger. Konklusion.).

2) Inden man lavede forsøgene, foretog man en dimensionering. Hvis formindskelsen er under 0.2 mg/l, er det ikke rimeligt at gå over til den nye metode. Man ønsker derfor at finde det mindste antal målinger, der skal indgå i undersøgelsen, for at man ved en ændring i indholdet af A på $\Delta = 0.2$ mg/l højst har, at $P(\text{type II fejl}) = \beta = 10\%$. Man har en begrundet formodning om, at spredningen i resultaterne højst kan være 0.22 mg/l ($\alpha = 0.05$).

Er det på basis af disse oplysninger, samt resultaterne i spørgsmål 1) muligt at vurdere, om den fundne formindskelse er stor nok til, at man vil gå over til den nye metode?

Opgave 5.4 (test for middelværdi).

På pakken af en iscreme står, at portionen indeholder 14 gram fedt. For at kontrollere dette købes 10 pakker is, og fedtindholdet måles. Man finder et gennemsnit på 13.1 gram og et estimat s for spredningen på 0.42 gram.

- 1) Kan man ud fra disse data bevise på signifikansniveau $\alpha = 0.01$, at middelindholdet afviger fra 14 gram? (Husk altid at anføre: Hvad X er. Antagelser. Nulhypotese. Beregninger. Konklusion.).
- 2) Angiv et estimat for middelindholdet.
- 3) Angiv et 95% konfidensinterval for middelindholdet.
- 4) Dimensionering. Bestem den nødvendige stikprøvestørrelse n , for at man ved en forskel i fedtindhold på $\Delta = 0.50$ gram højst har, at $P(\text{type I fejl}) = \alpha = 0.05$ og $P(\text{type II fejl}) = \beta = 0.05$. ($\sigma \approx 0.42$ gram).

Opgave 5.5 (test for middelværdi).

En fabrikation er baseret på en kemisk reaktion, hvor processen forudsætter tilstedeværelse af en katalysator. Med den hidtil benyttede katalysatortype C_1 udnyttes i middel kun ca. 70% af den dyreste råvare. Firmaet overvejer at gå over til en mere effektiv katalysatortype C_2 ved produktionen. Omlægning hertil vil imidlertid kræve betydelige etableringsomkostninger, hvorfor firmaet kun vil lægge produktionen om, såfremt i middel mindst 80% af den dyreste råvare udnyttes, når C_2 benyttes. Til vurdering heraf foretoges en række forsøg med benyttelse af C_2 . Følgende udnyttelsesprocenter fandtes:

68.3	87.7	80.0	84.2	84.0	83.6	76.4	79.9	89.3	75.8
96.1	88.0	79.8	83.7	84.4	95.5	84.2	92.1	92.4	83.9

- 1) Beregn estimer $\bar{x}$ og s for middelværdi $E(X)$ og spredning $\sigma(X)$, idet X betegner udnyttelsesprocenten, når C_2 benyttes.
- 2) Vurder, om de opnåede forsøgsresultater kan opfattes som et eksperimentelt bevis for, at i middel over 80% af den dyreste råvare udnyttes, når C_2 benyttes.
- 3) Opstil et (tosidet) 95% konfidensinterval for $E(X)$.

Vi antager i det følgende, at X (approksimativt) er normalfordelt $n(\bar{x}, s)$.

- 4) Beregn sandsynligheden for, at udnyttelsesprocenten X er mindre end 80%, når C_2 benyttes.

Opgave 5.6 (test for middelværdi).

Et kemikalium fremstilles industrielt ved inddampning af en bestemt opløsning. Det var vigtigt, at denne opløsning var svagt basisk med $\text{pH} = 8.0$. Man foretog derfor kontrolmæssigt nogle pH-bestemmelser for den benyttede opløsning. Følgende værdier fandtes:

8.2	8.3	7.9	8.2	7.8	8.6	8.9	7.8	8.2
-----	-----	-----	-----	-----	-----	-----	-----	-----

Foretag en testning af om opløsningen kan antages at opfylde kravet til pH-værdi, og opstil et 95% konfidensinterval for pH-værdien.

Opgave 5.7 (test for middelværdi).

Man frygter, at den såkaldte "syreregn" er årsag til, at en bestemt skov er stærkt medtaget. Man måler SO_2 -koncentrationen forskellige steder i skovbunden (i $\mu\text{g}/\text{m}^3$) og finder:

32.7	23.9	21.7	18.6	27.6	35.1	42.2	36.5	13.4	41.8	34.3	30.0
------	------	------	------	------	------	------	------	------	------	------	------

I ubeskadede skove er SO_2 -koncentrationen $20 \mu\text{g}/\text{m}^3$. Giver forsøgene et bevis for, at middelskoncentrationen af SO_2 i den beskadigede skov er større end normalt?

Angiv et tosidet 95%-konfidensinterval for SO_2 -koncentrationen.

Hypotesetestning (1 variabel)

Opgave 5.8 (test for middelværdi).

På et kraftvarmeværk mener man, at en ny metode vil kunne formindske svovlindholdet i de slagger, der bliver tilbage efter kulfyringen. Med en bestemt kvalitet kul, har det hidtidige svovlindhold været 2.70 %.

For at vurdere den nye metode ønsker ingeniøren at foretage en række forsøg.

- 1) Hvor mange forsøg skal der mindst foretages, hvis $\alpha = 5\%$, $\beta = 10\%$, $\Delta = 0.05$ og et overslag over spredningens størrelse sætter den til højst 0.08%.
- 2) Uanset resultatet af dimensioneringen i spørgsmål 1), er der kun praktiske muligheder for at lave 16 forsøg. Følgende værdier af svovlindholdet fandtes (%).

2.58	2.64	2.80	2.50	2.52	2.69	2.60	2.73	2.61	2.62	2.65	2.58	2.70	2.67	2.62	2.64
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

Test om disse måleresultater beviser, at svovlindholdet ved den nye metode i middel er blevet mindre.

- 3) Hvis resultatet af spørgsmål 2 er, at svovlindholdet er mindre, skal angives et 95% tosidet konfidensinterval for det forventede svovlindhold.

Opgave 5.9 (dimensionering).

Ved en batchproduktion er under sædvanlige betingelser middeludbyttet lig med 83.2 kg og spredningen på udbytteerne lig med 1.5 kg.

Laboratorieforsøg har antydnet, at visse ændringer i produktionsbetingelserne medfører en forøgelse i middeludbyttet på 1.2 kg, uden at spredningen ændres.

Ved et driftsforsøg bestående af en produktion af n batches under de ændrede produktionsbetingelser ønskes derfor testet, om middeludbyttet beviseligt er blevet forbedret.

- 1) Opstil nulhypotesen for en sådan testning.
- 2) Bestem n således, at et U -test af denne nulhypotese med sandsynlighed 90% udviser 1-stjernet signifikans, såfremt forøgelsen i middeludbyttet virkelig er 1.2 kg.

Antag, at udbytteerne i et driftsforsøg med samme n som bestemt i spørgsmål 2) udviser et gennemsnit på 84.5 kg, og et estimat for spredningen på $s = 2.4$ kg.

- 3) Test formodningen om, at spredningen er uforandret under de ændrede produktionsbetingelser.
- 4) Test under hensyntagen til svaret på spørgsmål 3) nulhypotesen i spørgsmål 1)
- 5) Opstil 95%-konfidensintervaller (tosidede) for middelværdien og spredningen af udbytteerne under de ændrede produktionsbetingelser.

Opgave 5.10 (test for varians).

Et nyt måleapparat påstås at give måleresultater med spredningen $\sigma = 1.8$ mg/l ved måling af saltindholdet i en opløsning. Da dette er mindre end det sædvanlige, køber et laboratorium et eksemplar af apparatet for at kontrollere påstanden. Der foretages 15 målinger med følgende resultater:

3.4	7.7	6.0	8.1	8.4	2.7	4.9	1.2	2.1	5.4	3.5	1.5	5.2	4.1	3.9
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

- 1) Test på basis af disse resultater, om spredningen afviger fra 1.8 mg/l.

(Husk altid at anføre: Hvad X er. Antagelser. Nulhypotese. Beregninger. Konklusion.).

- 2) Det vides nu, at den saltopløsning der måles på har en koncentration på 4 mg/l.

Test på basis af denne yderligere oplysning om spredningen afviger fra 1.8 mg/l.

Opgave 5.11 (test for varians).

Kunststoffet PARAPHYMOSE oparbejdes af gamle kartofler. Det er væsentligt, at kartoflerne, som indgår i et produktionsparti, er nogenlunde homogene med hensyn til stivelsesprocent, mens værdien af denne er mindre væsentlig. Man sikrer homogenitet ved at udtage et antal prøver og undersøge deres variation. Ved en sådan prøveudtagning har man udtaget 64 prøver og målt stivelsesprocenten X . Man fandt, at $\bar{x} = 19.22$ og $s = 0.6539$. Af teknologiske grunde bør homogeniteten være sådan, at $V(X) \leq 0.300$. Det kan antages, at X er normalfordelt.

Skal man anvende det foreliggende parti? (Idet det antages, at X er normalfordelt beviser målingerne så, at partiet er uegnet?).

Opgave 5.12 (varians).

En medicinalvarefabrik overvejer at indføre en ny analysemetode. Det formodes, at spredningen er mindre end 2.0 mg/l. Man ved, at den nye metode er uden systematiske fejl. Der fremstilles ved afvejning et præparat med nøjagtig 40.5 mg/l, dvs. middelværdien er kendt. Følgende måleresultater (i mg/l) findes med den nye metode:

42.8	39.3	41.2	40.9	40.2	40.7	40.6	40.0	41.5
------	------	------	------	------	------	------	------	------

- 1) Bekræfter de foretagne observationer forhåndsformodningen om spredningen. (Husk altid at anføre: Hvad X er. Antagelser. Nulhypotese. Beregninger. Konklusion.).
- 2) Angiv et estimat for spredningen.
- 3) Angiv et 95% konfidensinterval for spredningen.

Opgave 5.13 (test for varians).

Et nyt analyseapparat formodes at give måleresultater med en spredning på under 5%. På en 48% standardopløsning foretoges 30 målinger med det nye apparat. Man fik herved et gennemsnit på 47.5% og et estimat for spredningen på 6.9%. Test om dette er i modstrid med det forventede.

Hypotesetestning (1 variabel)

Opgave 5.14 (test for varians).

Ved indkøbet af et nyt måleapparat oplystes det, at apparatet målte med en spredning på 2.8 enheder. Efter at have brugt apparatet et stykke tid nærede køberen mistanke om, at apparatet målte med større spredning end oplyst. For at få spørgsmålet undersøgt lod køberen en bestemt måling udføre et antal gange. Følgende resultater fandtes:

18.8	15.5	12.2	14.8	4.80	1.20	1.43	9.60	1.39	1.17	5.60	1.27	1.35
8.70	1.23	1.40	1.02	1.65	1.91	1.14	1.46	1.59	1.54	1.01	1.80	

Hvilke konklusioner kan køberen drage ud fra en statistisk analyse af de fundne forsøgsresultater?

Opgave 5.15 (test for middelværdi og varians).

En kemisk virksomhed kører en produktion, der har middeludbyttet 76.5% og spredningen 1.5%. Virksomheden overvejer at gå over til et andet beholdermateriale for sine reaktionskar. Det vides ikke på forhånd, om processens middeludbytte påvirkes af en sådan ændring. Processens spredning forventes at være upåvirket.

Til afklaring af ovennævnte uvished udføres 9 driftsforsøg med anvendelse af det nye beholdermateriale. Følgende resultater fandtes:

79.5	76.5	77.5	76.3	79.5	76.7	76.8	78.3	77.3
------	------	------	------	------	------	------	------	------

- 1) Test, om processens spredning kan antages at være uforandret.
- 2) Test under hensyntagen til svaret i 1), om processens middeludbytte kan antages at være uforandret.
- 3) Opstil et 95%-konfidensinterval for processens middeludbytte.

Opgave 5.16 (test for middelværdi og varians).

En kemiker har udtænkt en ny analysemetode til bestemmelse af den procentvise koncentration af et stof A i en stofblanding. Han er overbevist om, at metoden ikke er behæftet med systematiske fejl, og mener, at metoden har større reproducerbarhed end den sædvanligt benyttede metode, hvis spredning σ er 0.10%. På en stofblanding, der indeholder 1.30% af A, udførtes ved den nye metode 6 uafhængige analyser. Følgende resultater fandtes:

1.20	1.27	1.33	1.19	1.09	1.24
------	------	------	------	------	------

- 1) Kan disse resultater fortolkes som et "eksperimentelt bevis" for, at den nye metodes reproducerbarhed er større end den sædvanlige metodes?
- 2) Tyder resultaterne på, at den nye metode er behæftet med systematiske fejl?
- 3) Opstil, under hensyntagen til de udførte testninger, det bedst mulige tosidede 95%-konfidensinterval for den nye metodes spredning. Angiv forudsætningerne for det opstillede konfidensinterval.

Opgave 5.17 (test for middelværdi og varians).

En sukkerfabrik leverer sukkeret i 1 kg-posere og 2 kg-posere. Vægten af de fyldte posere varierer.

- 1) For 1 kg-posernes vedkommende antages vægten at have en middelværdi på 1000 gram. En række forsøg har vist, at sandsynligheden for, at en tilfældig udtaget 1 kg-pose vejer mere end 1025 gram, er 10%. Giv på det grundlag en vurdering af spredningen.
- 2) For 2 kg-posernes vedkommende bør middelværdien være 200 gram, og spredningen må ikke overstige 25 gram. For at kontrollere, om en ny pakkemaskine overholder disse normer, udtages tilfældigt 100 pakker af denne maskines produktion. Gennemsnittet beregnes til $\bar{x} = 2008$ gram, og et estimat for spredningen til $s = 25$ gram.

Det formodes på forhånd, at den nye maskine overholder de ovennævnte normer. Foretag en statistisk vurdering af, om dette kan antages at være tilfældet.

Opgave 5.18 (parvise observationer).

Et bestemt medikament ønskes testet for dets effekt på blodtrykket. 12 mænd fik deres blodtryk målt før og efter indtagelse af medikamentet. Resultaterne var:

mand nr	1	2	3	4	5	6	7	8	9	10	11	12
Før	120	124	130	118	140	128	140	135	126	130	126	127
Efter	128	131	131	127	132	125	141	137	118	132	129	135

Udfør en testning af, om disse tal tyder på, at medikamentet påvirker blodtrykket.

Opgave 5.19 (parvise observationer).

Et diætprodukt påstår i en reklame, at brug af produktet i en måned vil resultere i et vægttab på 3 kg. En forbrugerorganisation ønsker at teste denne påstand. 8 personer bruger produktet i en måned, og det resultat fremgår af nedenstående tabel:

person nr	1	2	3	4	5	6	7	8
Startvægt	81	101	98	99	78	71	75	93
Slutvægt	78	95	95	97	73	69	70	89

- 1) Undersøg på grundlag af disse tal, om det på basis af disse tal på et signifikansniveau på 5% kan vises, at reklamens påstand er fejlagtig?
- 2) Opstil et 95% tosidet konfidensinterval for middelværdien af vægttabet, og giv på grundlag heraf en vurdering af virkningen af diætproduktet.

Hypotesetestning (1 variabel)

Opgave 5.20 (parvise observationer).

En producent af malerverer har laboratorieresultater, der tyder på, at en ny lak A, har en større slidstyrke end den sædvanlige lak B. Han ønsker en afprøvning i praksis og aftaler med ejerne af 6 bygninger med mange trapper, at han må lakere deres trapper. Efter 3 måneders forløb måles graden af slid (i %) i hver bygning.

1) Angiv hvorledes du ville foretage forsøget.

2) De målte værdier af slid efter valg af plan var

Bygning nr	1	2	3	4	5	6
Ny lak	20.3	25.1	21.8	19.6	18.9	23.5
Sædvanlig lak	19.5	28.4	21.6	22.0	20.9	25.8

Undersøg om observationerne leverer et eksperimentelt bevis for, at den nye lak er mere slidstærk end den sædvanlige lak.

6. SANDSYNLIGHEDSREGNING

6.1 INDLEDNING

Har man gået i det matematiske gymnasium, er begreberne i dette kapitel velkendte, og kapitlet kan tjene som en repetition. For andre er begreberne beskrevet tilstrækkelig udførligt til, at det kan tjene som en første indførelse.

6.2 Regneregler for sandsynligheder

I dette afsnit vil følgende eksempel blive benyttet til illustration af definitioner og begreber.

Eksempel 6.1. Gennemgående eksempel.

En fabrik har erfaring for, at den daglige produktion af glasfigurer indeholder 10 % misfarvede, 20% har ridser, og 1 % af produktionen er både ridsede og misfarvede.

Et eksperiment består i tilfældigt at udtage en glasfigur af produktionen. Lad A være hændelsen at få en misfarvet og lad B være hændelsen at få en ridset.



DEFINITION af “ikke A ” eller komplementærmængden til A . Komplementærmængden til A benævnes $\bar{A}$ og er mængden af alle udfald i udfaldsrummet U , der ikke er i A (den blå mængde på figur 6.1).

Eksempelvis er $\bar{A}$ i eksempel 6.1 mængden af alle glasfigurer, der ikke er misfarvet.

Idet $P(A) = 0.1$ ses umiddelbart, at $P(\bar{A}) = 1 - P(A) = 0.9$.

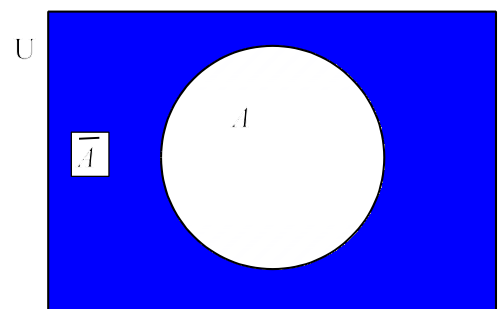
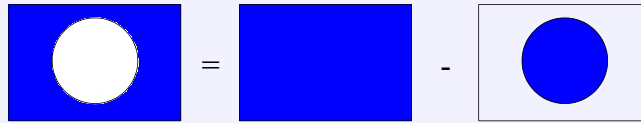


Fig. 6.1. Komplementærmængde

Vi har derfor klart følgende sætning:

SÆTNING 6.1 (om sandsynlighed). Idet A og B er vilkårlige hændelser gælder:

$$P(\bar{A}) = 1 - P(A)$$



DEFINITION af “både A og B ” eller fællesmængden af A og B . Fællesmængden til A og B benævnes $A \cap B$ og er mængden af alle udfald i udfaldsrummet U , der tilhører både A og B (Den blå mængde i fig 6.2).

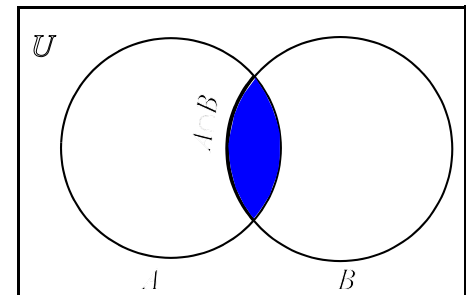


Fig. 6.2. Fællesmængde

Eksempelvis er $A \cap B$ i eksempel 6.1 mængden af alle glasfigurer, der både er misfarvede og ridsede.

Af eksemplet følger, at $P(A \cap B) = 0.01$.

DEFINITION af “ A men ikke B ” eller differensmængden A minus B . Differensmængden A minus B benævnes $A \setminus B$ og er mængden af alle udfald i udfaldsrummet U , der tilhører A men ikke B .

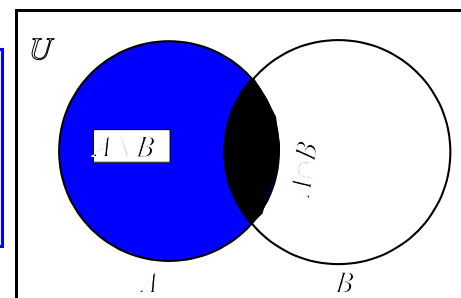


Fig. 6.3. Differensmængde

Eksempelvis er $A \setminus B$ i eksempel 6.1 mængden af alle glasfigurer, der er misfarvede men ikke ridsede.

Det ses umiddelbart af figur 6.3, at

$$P(A \setminus B) = P(A) - P(A \cap B).$$

I eksempel 6.1 er $P(A \setminus B) = 0.1 - 0.01 = 0.09$.

DEFINITION af “enten A eller B ” eller foreningsmængden af A og B . Foreningsmængden af A og B benævnes $A \cup B$ og er mængden af alle udfald i udfaldsrummet U , der enten tilhører A eller B eventuelt dem begge (den blå mængde på fig. 6.4).

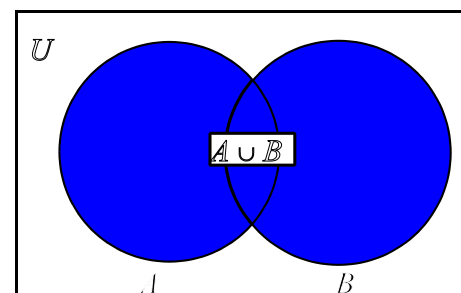
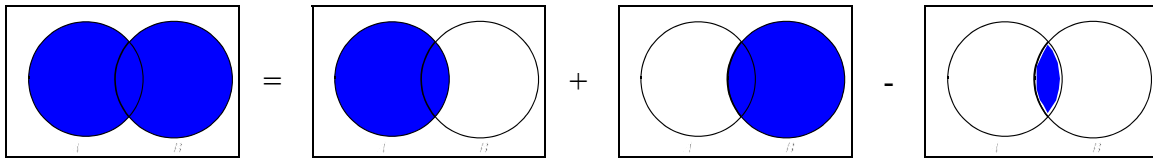


Fig. 6.4 Foreningsmængde

Eksempelvis er $A \cup B$ i eksempel 6.1 mængden af alle glasfigurer, der enten er misfarvede eller ridsede eventuelt begge dele.

Ved betragtning af de nedenfor anførte grafiske afbildninger (arealbetragtning) ses umiddelbart, at der gælder følgende

Additionssætning: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.



I eksempel 6.1 er $P(A \cup B) = 0.1 + 0.2 - 0.01 = 0.29$.

U kaldes for **den sikre hændelse**, dens komplementære hændelse for **den umulige hændelse** $\emptyset$.

$P(\emptyset) = 0$

En hændelse, som kun indeholder ét udfald, kaldes en **elementarhændelse**.

To mængder hvis fællesmængde er tom kaldes **disjunkte**.

For nogle hændelser A og B gælder, at $P(A \cap B) = P(A) \cdot P(B)$, men denne formel gælder ikke generelt. Eksempelvis i eksempel 6.1 er $P(A) \cdot P(B) = 0.1 \cdot 0.2 = 0.02 \neq P(A \cap B)$. For at få en mere generel regel indføres $P(B|A)$ som kaldes sandsynligheden for, at B indtræffer, når A er indtruffet (den af A **betingede sandsynlighed** for B).

For at forklare den følgende definition, vil vi simplificere eksempel 6.1, idet vi antager, at den daglige produktion er 100 glasfigurer. I så fald er der 10 misfarvede figurer, 20 ridsede figurer, og 1 figur der er både misfarvet og ridset. Hvis vi begrænser vort udfaldsrum til A , så er

$$P(B|A) = \frac{1}{10} = \frac{\frac{1}{100}}{\frac{10}{100}} = \frac{P(A \cap B)}{P(A)}.$$

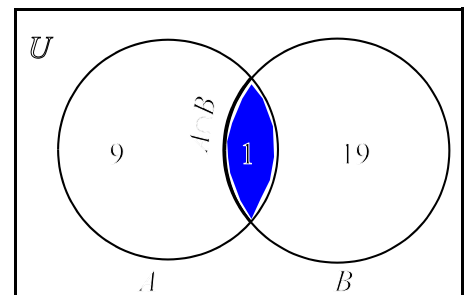


Fig. 6.2

Denne beregning begrundes rimeligheden i følgende definition:

DEFINITION af betinget sandsynlighed $P(B|A)$. som kaldes sandsynligheden for, at B indtræffer, når A er indtruffet (den af A **betingede sandsynlighed** for B) defineres ved

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Ved multiplikation fås

Produktsætningen: $P(A \cap B) = P(A) \cdot P(B|A) (= P(B) \cdot P(A|B))$.

Benyttes produktsætningen på eksempel 6.1 fås $P(A \cap B) = P(A) \cdot P(B|A) = 0.1 \cdot 0.1 = 0.01$.

Lad os illustrere anvendelsen af betinget sandsynlighed med et andet eksempel.

Eksempel 6.2: Betinget sandsynlighed.

En urne indeholder 3 røde og 3 hvide kugler. Vi udtrækker successivt 2 kugler fra urnen.

Vi betragter følgende 2 hændelser:

A : Den først udtrukne kugle er rød.

B : Den anden udtrukne kugle er rød.

Beregn $P(A \cap B)$ hvis

- 1) kugleudtrækningen foregår, ved at den først udtrukne kugle lægges tilbage før den anden udtrækkes.
- 2) kugleudtrækningen foregår, ved at den først udtrukne kugle **ikke** lægges tilbage før den anden udtrækkes.

LØSNING

1) Her er $P(B|A) = \frac{3}{6}$ og derfor ifølge produktsætningen $P(A \cap B) = P(A) \cdot P(B|A) = \frac{1}{4}$

2) Her er $P(B|A) = \frac{2}{5}$ og derfor $P(A \cap B) = \frac{3}{6} \cdot \frac{2}{5} = \frac{1}{5}$



DEFINITION af statistisk uafhængighed. To hændelser A og B siges at være statistisk uafhængige, såfremt $P(A \cap B) = P(A) \cdot P(B)$.

Navnet skyldes, at vi i dette tilfælde har $P(B|A) = P(B)$ og $P(A|B) = P(A)$, således at sandsynligheden for, at den ene hændelse indtræffer, ikke afhænger af, om den anden hændelse indtræffer.

Eksempelvis ved kast med en terning, så vil sandsynligheden for at få en sekser i andet kast være uafhængigt af udfaldet i første kast, således at sandsynligheden for at få 2 seksere i de første 2 kast er $\frac{1}{6} \cdot \frac{1}{6}$.

Definitionen af statistisk uafhængighed generaliseres til flere hændelser end 2, således at der i tilfælde af 3 uafhængige hændelser A_1, A_2 og A_3 også gælder:

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3).$$

6.3 Symmetrisk sandsynlighedsfelt.

Såfremt et udfaldsrum U indeholder n udfald som alle er lige sandsynlige (**symmetrisk sandsynlighedsfelt**), vil sandsynligheden for hvert udfald være $P(u) = \frac{1}{n}$

En hændelse A som indeholder a udfald vil da have sandsynligheden $P(A) = \frac{a}{n}$.

Dette udtrykkes ofte kort ved at sige, at sandsynligheden for A er antal gunstige udfald i A divideret med det totale antal udfald i udfaldsrummet.

I symmetriske sandsynlighedsfelter bliver problemet derfor, hvorledes man let kan optælle antal udfald. Dette kan ofte gøres ved benyttelse af **kombinatorik**.

Eksempel 6.3. Optælle antal rækkefølger.

En pentapeptide består af en kæde af følgende 5 aminosyrer: alanine, valine, glycine, cysteine, tryptophan. Den har forskellige egenskaber afhængig af den rækkefølge de 5 aminosyrer sidder i kæden. Hvor mange forskellige rækkefølger er der mulighed for i en pentapeptide?

LØSNING.

- | | |
|---|-----------------------------|
| 1) På førstepladsen i kæden placeres en aminosyre. | Der er 5 muligheder herfor. |
| 2) På den næste plads i kæden placeres en ny aminosyre | Der er 4 muligheder herfor |
| 3) På den tredje plads i kæden placeres en ny aminosyre | Der er 3 muligheder herfor |
| 4) På den fjerde plads i kæden placeres en ny aminosyre | Der er 2 muligheder herfor |
| 5) På den femte plads i kæden placeres den sidste aminosyre | Der er 1 mulighed herfor. |

I alt er der $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$ rækkefølger.



DEFINITION af $n!$ Produktet af de n første positive, hele tal kaldes $n!$ d.v.s.

$n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 2 \cdot 1$ ($n!$ udtales: n udråbstegn eller n fakultet).

Endvidere defineres $0! = 1$.

DEFINITION af kombination. En kombination er et udvalg af elementer udtaget af en mængde uden at tage hensyn til rækkefølgen af elementerne.

SÆTNING 6.1 (om antallet af kombinationer). Antal måder hvorpå man - uden mellemliggende tilbagelægning og uanset rækkefølgen - kan udtage i alt p elementer fra en mængde med n elementer, hvor $n \geq p$ kaldes $K(n, p)$ eller $\binom{n}{p}$ (n over p).

$$\text{Der gælder } \binom{n}{p} = \frac{n!}{p! \cdot (n-p)!}$$

Bevis: Beviset knyttes for enkelheds skyld til et taleksempel, som let kan generaliseres.

Lad os antage, vi på tilfældig måde udtager 3 kugler af en kasse, der indeholder 5 kugler med numrene k_1, k_2, k_3, k_4, k_5 .

Lad os først gå ud fra, at rækkefølgen hvori kuglerne trækkes ud er af betydning. Der er altså eksempelvis forskel på k_1, k_3, k_4 og k_3, k_1, k_4 . Første kugle kan da udtages på 5 forskellige måder, og for hver af disse 5 kugler kan den næste kugle udtages på 4 måder. I alt kan de to kugler udtages på $5 \cdot 4$ måder. Den sidste kugle skal udtages blandt de resterende 3 kugler, så det kan gøres på 3 måder. Hvis de 3 kugler udtages, så rækkefølgen spiller en rolle, kan det følgelig gøres på $5 \cdot 4 \cdot 3$ måder. Hvis de 3 kugler udtages, så rækkefølgen **ikke** spiller en rolle, har vi vedtaget, det kan gøres på $\binom{5}{3}$ måder. Lad en af disse måder være k_1, k_3, k_4 . Disse 3 elementer kan ordnes i rækkefølge på $3! = 3 \cdot 2 \cdot 1$ måder.

$$\text{Vi har følgelig, at } 5 \cdot 4 \cdot 3 = \binom{5}{3} \cdot 3! \text{ eller } \binom{5}{3} = \frac{5 \cdot 4 \cdot 3}{3!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3! \cdot 2 \cdot 1} = \frac{5!}{3! \cdot 2!}$$



TI - 89 : Beregning af tallet $\binom{n}{p}$: MATH\Probability\nCr(n,p)

$$\text{Eksempel: } \binom{5}{3} = nCr(5,3) = 10$$

OPGAVER

Opgave 6.1 (regneregler)

I en mindre by viser en undersøgelse, at 60% af alle husstande holder en lokal avis, mens 30% holder en landsdækkende avis. Endvidere holder 10% af husstandene begge aviser.

Lad en husstand være tilfældig udvalgt, og lad A være den hændelse, at husstanden holder en lokal avis, og B den hændelse, at husstanden holder en landsdækkende avis.

1) Angiv symbolsk udtrykt ved A og B hændelserne:

C : Husstanden holder begge aviser.

D : Husstanden holder kun den lokale avis.

E : Husstanden holder mindst én af aviserne.

F : Husstanden holder ingen avis

G : Husstanden holder netop én avis.

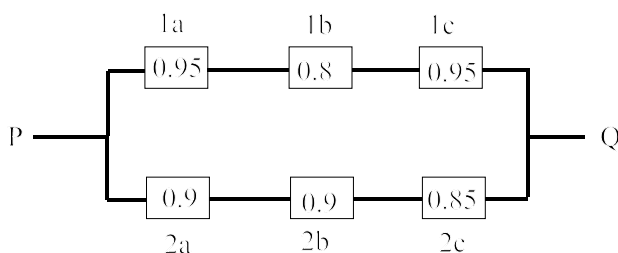
Det vides nu, at den tilfældigt valgte husstand holder den landsdækkende avis.

H : Husstanden holder den lokale avis.

2) Beregn sandsynlighederne for ovennævnte hændelser.

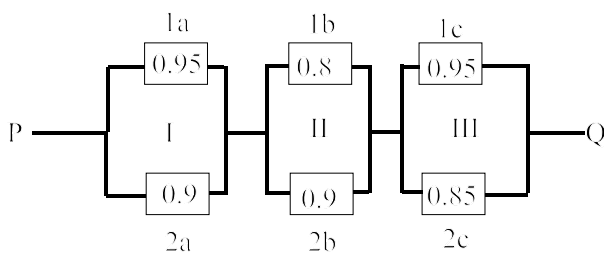
Opgave 6.2 (regneregler)

1) I figur 1 er vist et elektrisk apparat, som kun fungerer, hvis enten alle komponenter 1a, 1b og 1c i den øverste ledning eller alle komponenter 2a, 2b og 2c i den nederste ledning fungerer. Sandsynligheden for at hver komponent fungerer er vist på tegningen, og det antages, at sandsynligheden for at en komponent fungerer er uafhængig af om de øvrige komponenter fungerer.



Figur 1

1) Hvad er sandsynligheden for at apparatet i figur 1 fungerer.



Figur 2

2) I figur 2 er vist et andet elektrisk apparat, som tilsvarende kun fungerer, hvis alle de tre kredsløb I, II og III fungerer, og det er kun tilfældet hvis enten den øverste eller den nederste komponent fungerer.

Hvad er sandsynligheden for at apparatet i figur 2 fungerer.

Opgave 6.3 (regneregler)

Tre skytter skyder hver ét skud mod en skydeskive. De har træffesandsynligheder 0.75, 0.50 og 0.30. Beregn sandsynligheden for

- 1) ingen træffere, 2) én træffer, 3) to træffere, 4) tre træffere.

Opgave 6.4 (regneregler)

En “terning” har form som et regulært polyeder med 20 sideflader. På 4 sideflader er der skrevet 1, på 8 sideflader er der skrevet 6 mens der er skrevet 2, 3, 4 og 5 på hver 2 sideflader.

Find sandsynligheden for i tre kast med denne terning at få

- 1) tre seksere
2) mindst én sekser
3) enten tre seksere eller tre enere

Opgave 6.5 (regneregler)

En levnedsmiddelvirkomhed vil gerne forsøge at fremstille to nye produkter k_1 og k_2 . Idet r_1 og r_2 betegner råmaterialer, og m_1 og m_2 betegner visse mellemprodukter, påtænker virksomheden at udforske processerne

$$(1) \quad r_1 \longrightarrow m_1 + m_2$$

$$(2) \quad r_2 \longrightarrow m_2$$

$$(3) \quad m_1 \longrightarrow k_1$$

$$(4) \quad m_2 \longrightarrow k_2$$

Følgende hændelser betragtes:

H_1 : Det lykkes at gennemføre processen (1).

H_2 : Det lykkes at gennemføre processen (2).

H_3 : Det lykkes at gennemføre processen (3), hvis m_1 er til rådighed.

H_4 : Det lykkes at gennemføre processen (4), hvis m_2 er til rådighed.

H_5 : Det lykkes at fremstille k_1 ud fra råmaterialerne.

H_6 : Det lykkes at fremstille k_2 ud fra råmaterialerne.

Lad der på basis af tidligere erfaringer med lignende processer være givet sandsynlighederne:

$$P(H_1) = \frac{1}{2}, \quad P(H_2) = \frac{3}{4}, \quad P(H_3) = \frac{2}{3}, \quad P(H_4) = \frac{1}{3}.$$

og lad H_1 , H_2 , H_3 og H_4 være statistisk uafhængige hændelser.

- 1) Beregn sandsynlighederne $P(H_1 \cap H_2)$, $P(H_1 \cup H_2)$, $P(H_5)$, $P(H_6)$.
2) Beregn sandsynligheden $P(H_5 \cap H_6)$.

Opgave 6.6 (betinget sandsynlighed)

En virksomhed producerer regelmæssigt et levnedsmiddel i charger. I 5% af chargerne findes mikroorganismer af typen a, og i 20% af chargerne findes mikroorganismer af typen b. Såfremt disse mikroorganismer er til stede i en charge, forekommer de altid i en bestemt mængde (mætningsgraden). Lad en charge (nr 27) blive udtaget på tilfældig måde af produktionen, og lad for denne charge følgende hændelser være defineret:

A: Tilstedeværelse af type a

B: Tilstedeværelse af type b

C: Tilstedeværelse af type a alene

D: Tilstedeværelse af type b alene

E: Tilstedeværelse af både type a og type b

F: Ingen tilstedeværelse af hverken type a eller type b.

Det antages, at *A* og *B* er statistisk uafhængige hændelser.

1) Udtryk hændelserne *C*, *D*, *E* og *F* ved *A* og *B* og angiv sandsynlighederne for disse hændelser.

Hver charge kontrolleres af et prøvehold. Undersøgelser har vist, at prøveholdet kasserer

- 60% af de charger, der indeholder type a alene.
- 80% af de charger, der indeholder type b alene
- 90% af de charger, der indeholder både type a og type b.
- 0% af de charger, der hverken indeholder type a eller type b.

Lad *K* være hændelsen: Charge 27 kasseres af prøveholdet.

2) Udtryk de givne oplysninger ved *K* og en eller flere af hændelserne *A* til *F*.

3) Find sandsynligheden for, at prøveholdet kasserer charge nr 27.

Opgave 6.7 (betinget sandsynlighed)

En virksomhed fremstiller en bestemt slags apparater. Hvert apparat er sammensat af 5 komponenter. Heraf er 3 tilfældigt udvalgt blandt komponenter af typen a og 2 blandt komponenter af typen b. Det vides, at 10% af a-komponenterne er defekte og 20% af b-komponenterne er defekte. Et apparat fungerer hvis og kun hvis det ikke indeholder nogen defekt komponent.

Der udtages på tilfældig måde et apparat fra produktionen. Lad os betragte hændelserne:

A: Det udtagne apparat indeholder mindst 1 defekt a-komponent.

B: Det udtagne apparat indeholder mindst 1 defekt b-komponent.

- Find $P(A)$, $P(B)$ og $P(A \cap B)$.
- Find sandsynligheden for, at et apparat, der på tilfældig måde udtages af produktionen ikke fungerer.
- Et apparat udtages på tilfældig måde fra produktionen og det konstateres ved afprøvning at det ikke fungerer. Find sandsynligheden for, at apparatet ikke indeholder nogen defekt a-komponent.

Opgave 6.8 (sandsynlighed)

To skytter konkurrerer ved en turnering. De har hver én patron og skyder mod en skive som giver 10 point, hvis et centralt område af skiven rammes og ellers 5 point. Rammes skiven ikke noteres 0 point.

Skytte A's dygtighed kan beskrives ved, at han i et skud har samme sandsynlighed for at få 10 points, 5 points eller 0 points.

Skytte B er dygtigere, idet hans sandsynligheder for at ramme er givet ved

Points y	10	5	0
$P(y)$	0.6	0.3	0.1

B har imidlertid fået en defekt patron med, der har sandsynligheden 50% for at fungere.

- 1) Idet X betegner det af A opnåede antal points og Y det af B opnåede antal points, ønskes tæthedsfunktionen for X og Y beregnet.
- 2) Find $E(X)$, $E(Y)$, $\sigma(X)$ og $\sigma(Y)$.
- 3) Beregn sandsynligheden for, at A vinder.
- 4) Det oplyses, at A vandt konkurrencen. Beregn sandsynligheden for, at B opnåede 5 points.

7. VIGTIGE DISKRETE FORDELINGER

7.1 INDLEDNING

I dette kapitel behandles de ved anvendelserne vigtige diskrete fordelinger: Den hypergeometriske fordeling, binomialfordelingen og Poissonfordelingen, samt de tilsvarende hypotesetest for en variabel.

7.2 HYPERGEOMETRISK FORDELING

Ved mange beregninger blandt andet indenfor kvalitetskontrol, må man benytte et ræsonnement som vist i eksempel 7.1 . Denne er en anvendelse af den såkaldte hypergeometriske fordeling.

Eksempel 7.1: Stikprøveudtagning (kvalitetskontrol)

En producent fabrikkerer komponenter, som sælges i æsker med 600 komponenter i hver. Som led i en kvalitetskontrol udtages hvert kvarter tilfældigt en æske produceret indenfor de sidste 15 minutter, og 25 tilfældigt udvalgte komponenter i denne undersøges, hvorefter det foregående kvarters produktion godkendes, såfremt der højst er én defekt komponent i stikprøven.

Hvor stor er acceptsandsynligheden p , hvis æsken indeholder i alt 10 defekte komponenter, såfremt udtrækningen sker **uden** mellemliggende tilbagelægnings ?

LØSNING:

Lad X være antallet af defekte blandt de 25 komponenter

Vi har: $p = P(X=0) + P(X=1)$.

Hændelsen " $X=0$ " forudsætter, at vi i alt udtager 0 af de 10 defekte og 25 forskellige af de 590 ikke-defekte. Dette kan gøres på $\binom{10}{0} \cdot \binom{590}{25}$ måder.

Idet vi i alt kan udtage 25 komponenter ud af de 600 på $\binom{600}{25}$ måder, fås

7. Vigtige diskrete fordelinger

$$P(X=0) = \frac{\binom{10}{0} \cdot \binom{590}{25}}{\binom{600}{25}} = 0.6512.$$

Hændelsen " $X = 1$ " forudsætter, at vi i alt udtager 1 af de 10 defekte og 24 forskellige af de

$$590 \text{ ikke-defekte, dvs. } P(X=1) = \frac{\binom{10}{1} \cdot \binom{590}{24}}{\binom{600}{25}} = 0.2876.$$

Vi har altså $p = 0.6512 + 0.2876 = 0.9388 = \underline{\underline{93.88\%}}$.

TI-89: Vælg MATH\Probability\nCr

$$(\text{nCr}(10,0) \cdot \text{nCr}(590,25) + \text{nCr}(10,1) \cdot \text{nCr}(590,24)) / \text{nCr}(600,25) = 0.9388$$



I eksemplet har vi "udledt" den såkaldte **hypergeometriske fordeling**, som er defineret på følgende måde:

DEFINITION af hypergeometrisk fordeling.

Lad der i en urne befinde sig N kugler, hvoraf M er sorte og $N - M$ er hvide.

Vi betragter det tilfældige eksperiment: Udtrækning af en kugle og observation af farve.

Eksperimentet gentages n gange uden mellemliggende tilbagelægning.

Lad X være antallet af sorte kugler, som udtrækkes.

X er en diskret stokastisk variabel med tæthedsfunktionen:

$$f(x) = \begin{cases} \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}} & \text{for } x \in \{0, 1, 2, \dots, n\} \\ 0 & \text{ellers} \end{cases}$$

X siges at være hypergeometrisk fordelt $h(N, M, n)$.

I eksempel 7.1 benyttede vi den hypergeometriske fordeling $h(600, 10, 25)$ for $x=0$ og $x=1$.

I "Supplement til statistiske grundbegreber" afsnit 7A bevises, at den hypergeometriske fordeling har middelværdien $E(X) = n \cdot p$

og spredningen $\sigma(X) = \sqrt{n \cdot p \cdot (1-p) \cdot \frac{N-n}{N-1}}$, hvor $p = \frac{M}{N}$.

Eksempel 7.2: Hypergeometrisk fordeling $h(10, 6, 3)$.

I en urne findes 10 kugler, hvoraf 6 er sorte, 4 er hvide.

Vi betragter det tilfældige eksperiment: "Udtrækning af en kugle og observation af farven på kuglen". Eksperimentet gentages 3 gange, idet den udtrukne kugle ikke lægges tilbage mellem hver udtrækning.

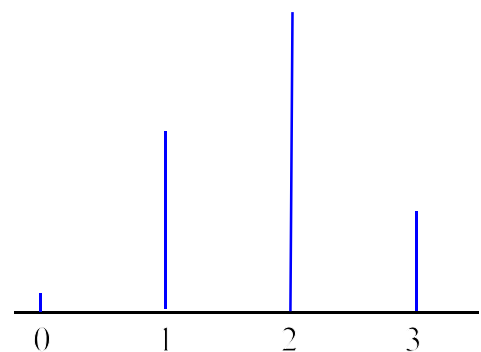
Lad X betegne antallet af udtrukne sorte kugler.

Find og skitser tæthedsfunktionen for X , og beregn middelværdi og spredning for X .

LØSNING:

X er en diskret stokastisk variabel, der som er hypergeometrisk fordelt $h(10, 6, 3)$ med tæthedsfunktionen $f(x) = P(X=x)$:

$$f(x) = \begin{cases} \frac{\binom{6}{0} \cdot \binom{4}{3}}{\binom{10}{3}} = \frac{4}{120} = 0.033 & \text{for } x = 0 \\ \frac{\binom{6}{1} \cdot \binom{4}{2}}{\binom{10}{3}} = \frac{36}{120} = 0.300 & \text{for } x = 1 \\ \frac{\binom{6}{2} \cdot \binom{4}{1}}{\binom{10}{3}} = \frac{60}{120} = 0.500 & \text{for } x = 2 \\ \frac{\binom{6}{3} \cdot \binom{4}{0}}{\binom{10}{3}} = \frac{20}{120} = 0.167 & \text{for } x = 3 \\ 0 & \text{ellers} \end{cases}$$



Stolpediagram for $h(10, 6, 3)$.

Sættes $p = \frac{M}{N} = \frac{6}{10}$ er middelværdien $E(X) = n \cdot p = 3 \cdot \frac{6}{10} = \underline{\underline{1.8}}$ og

spredningen $\sigma(X) = \sqrt{n \cdot p \cdot (1-p) \cdot \frac{N-n}{N-1}} = \sqrt{3 \cdot \frac{6}{10} \cdot \left(1 - \frac{6}{10}\right) \cdot \frac{10-3}{10-1}} = \underline{\underline{0.748}}$



7.3 BINOMIALFORDELING

Binomialfordelingen benyttes som model for antallet af "succeser" ved n uafhængige gentagelser af et eksperiment, som hver gang har samme sandsynlighed p for "succes". Problemstillingen fremgår af følgende eksempel.

Eksempel 7.3. En binomialfordelt variabel.

En drejebænk producerer 1 % defekte emner. Lad X være antallet af defekte blandt de næste 6 emner der produceres. Vi ønsker at finde sandsynligheden for at finde netop 2 defekte blandt disse 6, det vil sige $P(X = 2)$.

LØSNING:

Lad et eksperiment være at udtage et emne fra produktionen.

Resultatet af eksperimentet har to udfald: defekt, ikke defekt.

Eksperimentet gentages 6 gange uafhængigt af hinanden.

Der er en bestemt sandsynlighed for at få en defekt, nemlig $p = 0.01$.

Lad d være det udfald at få en defekt, og $\bar{d}$ være det udfald at få en fejlfri.

Et af de ønskede forløb med 2 defekte vil eksempelvis være $d, \bar{d}, d, \bar{d}, \bar{d}, \bar{d}$.

Dette forløb må have sandsynligheden

$$0.01 \cdot (1 - 0.01) \cdot 0.01 \cdot (1 - 0.01) \cdot (1 - 0.01) \cdot (1 - 0.01) = 0.01^2 \cdot (1 - 0.01)^4.$$

Et andet gunstigt forløb kunne være $\bar{d}, \bar{d}, d, \bar{d}, d, \bar{d}$ med sandsynligheden

$$(1 - 0.01) \cdot (1 - 0.01) \cdot 0.01 \cdot (1 - 0.01) \cdot 0.01 \cdot (1 - 0.01) = 0.01^2 \cdot (1 - 0.01)^4$$

Vi ser, at alle gunstige forløb har samme sandsynlighed. Antal forløb må være lig antal måder man kan placere 2 d 'er på 6 tomme pladser (eller antal måder man kan tage 2 kugler ud af en mængde på 6). Dette ved vi kan gøres på $\binom{6}{2}$ måder.

$$\text{Vi får følgelig, at } p = \binom{6}{2} \cdot 0.01^2 \cdot (1 - 0.01)^4 = 0.00144$$

$$\text{TI-89: CATALOG}\backslash\text{F3}\backslash\text{binomPdf}(6, 0.01, 2) = 0.00144$$



I eksemplet har vi "udledt" den såkaldte **binomialfordeling**, som er defineret på følgende måde:

DEFINITION af binomialfordeling.

- 1) Lad et tilfældigt eksperiment have 2 udfald "succes" og "fiasko"
- 2) Lad eksperimentet blive gentaget n gange uafhængigt af hinanden, og lad sandsynligheden for succes være en konstant p

Lad X være antallet af succeser blandt de n gentagelser

X er en diskret stokastisk variabel med tæthedsfunktionen

$$f(x) = \begin{cases} \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} & \text{for } x \in \{0, 1, 2, \dots, n\} \\ 0 & \text{ellers} \end{cases}$$

X siges at være binomialfordelt $b(n, p)$.

SÆTNING 7.1. (middelværdi og spredning for binomialfordeling).

Lad X være binomialfordelt $b(n, p)$.

Der gælder da $E(X) = n \cdot p$ og $\sigma(X) = \sqrt{n \cdot p \cdot (1-p)}$.

Bevis: Lad os betragte et eksperiment, hvor resultatet "succes" har sandsynligheden p for at ske.

Lad os foretage n **uafhængige gentagelser af eksperimentet**. At gentagelserne er uafhængige betyder, at udfaldet af et eksperiment ikke afhænger af udfaldet af de forrige eksperimenter.

Lad os betragte n stokastiske variable $X_1, X_2, \dots, X_n$,

$$\text{hvor } X_i = \begin{cases} 1 & \text{hvis i'te gentagelse af eksperimentet giver succes.} \\ 0 & \text{ellers} \end{cases}$$

Vi har $E(X_i) = \sum_i x_i f(x_i) = 1 \cdot p + 0 \cdot (1-p) = p$, og

$$V(X_i) = \sum_i (x_i - \mu)^2 f(x_i) = (1-p)^2 \cdot p + (0-p)^2 \cdot (1-p) = p - p^2 = p \cdot (1-p)$$

Idet $X = X_1 + X_2 + \dots + X_n$ er binomialfordelt $b(n, p)$ fås af linearitetsreglen (kapitel 1 afsnit

5), at $E(X) = E(X_1) + E(X_2) + E(X_3) + \dots + E(X_n) = p + p + p + \dots + p = n \cdot p$.

Endvidere fås af kvadratregele i kapitel 1 afsnit 5, idet vi har uafhængige gentagelser, at

$$V(X) = V(X_1) + V(X_2) + \dots + V(X_n) = p \cdot (1-p) + p \cdot (1-p) + \dots + p \cdot (1-p),$$

eller $V(X) = n \cdot p \cdot (1-p)$.



7. Vigtige diskrete fordelinger

Eksempel 7.4: Binomialfordelt variabel

Lad der på to af sidefladerne på en terning være skrevet tallet 1, på to andre sideflader være skrevet tallet 2 og på de sidste to sideflader være skrevet tallet 3. Vi betragter det tilfældige eksperiment: "7 kast med en terningen og observation af det fremkomne tal.

Lad X betegne antallet af toere ved de 7 kast. X antages at være binomialfordelt $b\left(7, \frac{1}{3}\right)$

- 1) Angiv tæthedsfunktionen $f(x)$ for X (3 betydende cifre), og tegn et stolpediagram for $f(x)$.
- 2) Find middelværdi og spredning for X

En person foretager eksperimentet 11 gange, d.v.s. foretager 11 gange en serie på 7 kast med terningen. Stikprøven gav følgende resultat

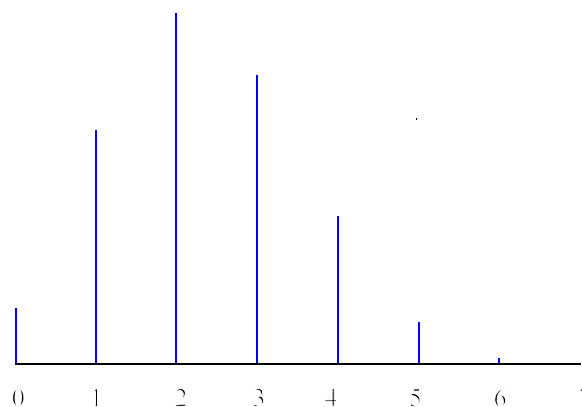
Antal toere i en serie	0	1	2	3	4	5	6	7
Antal gange dette skete	1	2	4	3	1	0	0	0

- 3) Giv på grundlag af stikprøven et estimat for p i binomialfordelingen.
- 4) Giv på grundlag af stikprøven et estimat for middelværdi og spredning

LØSNING:

- 1) Idet $f(x) = P(X = x) = \binom{7}{x} \cdot \left(\frac{1}{3}\right)^x \cdot \left(1 - \frac{1}{3}\right)^{7-x}$ fås

$$f(x) = \begin{cases} \left(\frac{2}{3}\right)^7 = 0.059 & \text{for } x = 0 \\ \binom{7}{1} \cdot \frac{1}{3} \cdot \left(\frac{2}{3}\right)^6 = 0.205 & \text{for } x = 1 \\ \binom{7}{2} \cdot \left(\frac{1}{3}\right)^2 \cdot \left(\frac{2}{3}\right)^5 = 0.307 & \text{for } x = 2 \\ \binom{7}{3} \cdot \left(\frac{1}{3}\right)^3 \cdot \left(\frac{2}{3}\right)^4 = 0.256 & \text{for } x = 3 \\ \binom{7}{4} \cdot \left(\frac{1}{3}\right)^4 \cdot \left(\frac{2}{3}\right)^3 = 0.128 & \text{for } x = 4 \\ \binom{7}{5} \cdot \left(\frac{1}{3}\right)^5 \cdot \left(\frac{2}{3}\right)^2 = 0.038 & \text{for } x = 5 \\ \binom{7}{6} \cdot \left(\frac{1}{3}\right)^6 \cdot \left(\frac{2}{3}\right) = 0.006 & \text{for } x = 6 \\ \left(\frac{1}{3}\right)^7 = 0.000 & \text{for } x = 7 \\ 0 & \text{ellers} \end{cases}$$



Stolpediagram for binomialfordelingen

$$2) E(X) = n \cdot p = 7 \cdot \frac{1}{3} = \underline{\underline{2.33}} \text{ og } \sigma(X) = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{7 \cdot \frac{1}{3} \cdot \left(1 - \frac{1}{3}\right)} = \underline{\underline{1.25}}$$

3) Der er i alt $1 \cdot 0 + 2 \cdot 1 + 4 \cdot 2 + 3 \cdot 3 + 1 \cdot 4 = 23$ toere i 77 kast.

$$\text{Et estimat for } p \text{ er } \hat{p} = \frac{23}{77} = \underline{\underline{0.299}}$$

4) Stikprøvens middelværdi er $\bar{x} = \frac{23}{11} = \underline{\underline{2.09}}$, og

$$\text{stikprøvens spredning er } \sigma(X) = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{7 \cdot \frac{23}{77} \cdot \left(1 - \frac{23}{77}\right)} = \underline{\underline{1.21}}$$



Hypotesetest for binomialfordelt variabel. I kapitel 5 gennemgik vi ved en række eksempler de grundlæggende begreber for hypotesetestning for én normalfordelt variabel. Disse begreber kan uændret overføres til hypotesetestning for binomialfordelt variabel. I tabel 6 findes en tabel over kumulerede sandsynligheder i binomialfordelingen for udvalgte værdier af p og n . Har man rådighed over mere omfattende tabeller eller over en lommeregner med kumuleret binomialfordeling kunne testene gennemføres eksakt. Som beskrevet i næste afsnit er det dog ofte muligt at approksimere med en normalfordeling, og i så fald er formlerne en simpel oversættelse fra de normalfordelte variable. Disse formler ses i appendix 5.5 og de tilsvarende konfidensintervaller kan findes i appendix 4.1, punkt 5. De følgende to eksempler viser anvendelser heraf.

Eksempel 7.5. Ensidet binomialfordelingstest.

En levnedsmiddelproducent fremstiller et levnedsmiddel A, som imidlertid har en ret ringe holdbarhed. Efter en række eksperimenter lykkedes det at frembringe et produkt B, som i alt væsentligt er identisk med A, men som har en bedre holdbarhed. Af markedsmæssige grunde er det vigtigt, at der ikke er forskel på smagen af B og af det velkendte produkt A. For at undersøge dette, lader producenten et panel af 24 eksperter smage vurdere, om man kan smage forskel. Man foretog derfor følgende smagsprøvningseksperiment.

Hver eksperter fik 3 ens udseende portioner, hvoraf en portion var af det ene levnedsmiddel og de to andre portioner var af det andet levnedsmiddel.

Hvilket af de 3 portioner der skulle indeholde et andet levnedsmiddel end de to andre, og om det skulle være levnedsmiddel A eller B, afgjordes hver gang ved lodtrækning. Kun forsøgslederen havde kendskab til resultatet.

Hver eksperter fik besked på, at de skulle fortælle forsøgslederen hvilken af de tre portioner der smagte anderledes. Hvis man ikke kunne smage forskel, skulle man gætte.

Resultatet viste, at af de 24 svar var 13 svar rigtige.

7. Vigtige diskrete fordelinger

Ved ren gætning kunne man forvente ca. $\frac{1}{3}$ dvs. ca. 8 rigtige svar. 13 rigtige svar er betydeligt flere, men kan det alligevel tilskrives tilfældigheder ved gætning?

Er der på et signifikansniveau på 5% statistisk påvist, at ekspertsmagerne kan smage forskel på smagen af A og B?

LØSNING:

Lad X = antallet af rigtige svar.

X er binomialfordelt $b(n, p)$, hvor $n = 24$ og p er ukendt.

Nulhypotese $H_0: p = \frac{1}{3}$ mod den alternative hypotese $H: p > \frac{1}{3}$

P -værdi = $P(X \geq 13) = \text{binomCdf}(24, 1/3, 13, 24) = 0.0284 = 2.84\%$

Da P -værdi $< 5\%$ forkastes nulhypotesen (enstjernet), dvs. der må konkluderes, at der er en smagsforskel mellem produkt A og B.

Tabel:

Ifølge appendix 5.5 kan man approksimere med en normalfordeling, hvis $n \cdot p_0 = 24 \cdot \frac{1}{3} = 8 > 5$ og $\frac{1}{10} \leq p_0 \leq \frac{9}{10}$.

Da dette er tilfældet dannes teststørrelsen $u = \frac{x - \frac{1}{2} - n \cdot p_0}{\sqrt{n \cdot p_0 \cdot (1 - p_0)}} = \frac{13 - \frac{1}{2} - \frac{24}{3}}{\sqrt{24 \cdot \frac{1}{3} \cdot \frac{2}{3}}} = 1.95$.

Idet $u_{0.95} = 1.645$ og $u_{0.99} = 2.326$ forkastes nulhypotesen enstjernet.



Eksempel 7.6. Konfidensinterval for parameteren p i binomialfordeling.

En plastikfabrik har udviklet en ny type affaldsbeholdere. Man overvejer at give en 6 års garanti for holdbarheden. For at få et skøn over om det er økonomisk rentabelt, bliver 100 beholdere udsat for et accelereret livstidstest som simulerer 6 års brug af beholderne. Det viste sig, at af de 100 beholdere overlevede de 85 testen.

Idet antallet af overlevende beholdere antages at være binomialfordelt, skal man

- 1) Angive et estimat for sandsynligheden p for at en beholder “overlever” i 6 år.
- 2) Angive et 95% konfidensinterval for p .

LØSNING:

1) Lad X være antallet af “overlevende” beholdere.

X forudsættes binomialfordelt $b(100, p)$.

Appendix 4.1 punkt 5 anvendes. Et estimat for p er $\tilde{p} = \frac{x}{n} = \frac{85}{100} = \underline{0.85}$

2) Da $10 \leq x \leq n - 10$ er forudsætningerne for at benytte normalfordelingsapproksimation opfyldt.

Vi får:

$$\tilde{p} \pm u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\tilde{p} \cdot (1 - \tilde{p})}{n}} = 0.85 \pm 1.96 \cdot \sqrt{\frac{0.85 \cdot (1 - 0.85)}{100}} = 0.85 \pm 0.07$$

dvs. $0.78 \leq p \leq 0.92$

TI 89 benytter netop denne formel i sin beregning, dvs. man skal altid her først undersøge om forudsætningen er opfyldt.

APPS/STATS/List\F7\5:1-PropZInt\ENTER

Menuen udfyldes med $x: 85$ $n: 100$ $C\text{-level}: 0.95$ ENTER

Resultat: C Int : [0.78 ; 0.92]



Direkte metode til beregning.

Hvis betingelserne for approksimation med normalfordeling ikke er opfyldt, må man benytte binomialfordelingen direkte.

Eksempel 7.7. Beregning af konfidensinterval hvis betingelserne ikke er opfyldt

I forbindelse med et reklamefremstød ønskede man at undersøge om borgerne i en mindre by havde set en bestemt reklame. Man spurgte et antal tilfældigt udvalgte husstande, og af 45 svar havde 9 set reklamen.

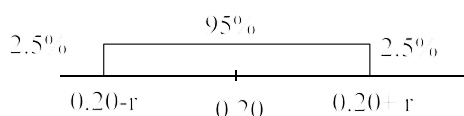
Opstil et 95% konfidensinterval for sandsynligheden p for at man har set reklamen.

Løsning:

Vi har, at $\hat{p} = \frac{9}{45} = 20\%$

Da $n \cdot \hat{p} = 9 < 10$ er betingelsen ikke opfyldt.

Udenfor et 95% konfidensinterval ligger 5%, og af symmetri Grunde ligger der 2,5% på hver side. (jævnfør figuren)



Jo større den sande værdi p er i forhold til 0.20 jo mindre bliver sandsynligheden for at få 9 svar eller færre. Vi leder derfor i grænsen efter et $p > 0.20$, så $P(X \leq 9) = 0.025$.

$\text{solve}(\text{binomCdf}(45, p, 9) = 0.025, p) | x > 0$ Resultatet blev $p = 0.346$.

Dernæst findes nedre grænse ved at lade p falde, indtil $P(X \geq 9) \approx 0.025$

Heraf fås $\text{solve}(\text{binomCdf}(45, p, 9, 45) = 0.025, p) | x > 0$ Resultatet blev $p = 0.0958$.

Konfidensinterval: [0.096; 0.346]

Bemærk, at konfidensintervallet her ikke ligger symmetrisk omkring 0.20, da binomialfordelingen i netop disse situationer netop ikke er symmetrisk omkring 0.20



7.4 POISSONFORDELINGEN

Ved udførelsen af Binomial-eksperimenter med lille sandsynlighed p , men stort n , er n undertiden ukendt: Man kan registrere antal gange en hændelse forekommer, men ikke eller kun meget approksimativt antal gange hændelsen ikke forekommer. Eksempler: Antal α partikler et radioaktivt præparat udsender pr. tidsenhed, antal trafikuheld på en bestemt vejstrækning i løbet af et år, antal støvpartikler i et bestemt lille volumen, antal defekte pr. km kabel, og antal varevogne der ankommer pr time til et stort varehus.

Under sådanne omstændigheder kan man ofte benytte den i det følgende omtalte Poissonfordeling som statistisk model for antallet af "impulser" pr. tidsenhed eller volumenenhed eller længdeenhed o.s.v.

Vigtige diskrete fordelinger

*) Præcis formulering: Følgende 3 betingelser skal være opfyldt:

- 1) Sandsynligheden for netop én impuls i et meget lille tidsrum Δt er med tilnærmelse proportional med Δt .

SÆTNING 7.2 (Poissonfordeling). Lad X være en stokastisk variabel, som angiver antallet af impulser i et givet tidsrum (eller areal, volumen, produktionsenhed osv.), idet ethvert tidspunkt i tidsrummet har samme mulighed for at være impulstidspunkt som ethvert andet tidspunkt. Endvidere skal impulserne indtræffe tilfældigt og uafhængigt af hinanden *).

Hvis det gennemsnitlige antal impulser i tidsrummet er $\mu > 0$, så siges X at være Poissonfordelt $p(\mu)$ med sandsynlighedsfordelingen (tæthedsfunktionen) $f(x) = P(X = x)$ bestemt ved

$$f(x) = \begin{cases} \frac{\mu^x}{x!} \cdot e^{-\mu} & \text{for } x \in \{0, 1, 2, \dots\} \\ 0 & \text{ellers} \end{cases}$$

Middelværdien for $p(\mu)$ er $E(X) = \mu$ og spredningen er $\sigma(X) = \sqrt{\mu}$.

(Matematisk formulering $\lim_{\Delta t \rightarrow 0} \frac{P(X=1)}{\Delta t} = \lambda$ (λ er en positiv konstant))

- 2) Sandsynligheden for 2 eller flere impulser i det meget lille tidsrum Δt er lille sammenlignet med Δt .

(Matematisk formulering $\lim_{\Delta t \rightarrow 0} \frac{P(X > 1)}{\Delta t} = 0$)

- 3) Antal impulser i forskellige, ikke overlappende tidsrum er statistisk uafhængige.

En bevisskitse for sætningen kan ses i “Supplement til statistiske grundbegreber” afsnit 7.C.

Eksempel 7.8: Antal revner p. meter i et tyndt kobberkabel.

På en fabrik fremstilles kobberkabler af en bestemt tykkelse. Mikroskopiske revner forekommer tilfældigt langs disse kabler. Man har erfaring for, at der i gennemsnit er 12.3 af den type revner p. 10 meter kabel.

Beregn sandsynligheden for, at der

- 1) ingen ridser er i 1 meter tilfældigt udvalgt kabel.
- 2) er mindst 2 ridser i 1 meter tilfældigt udvalgt kabel.
- 3) er højst 4 ridser i 2 meter tilfældigt udvalgt kabel

Fabrikken går nu over til en anden og billigere produktionsmetode. For at få et estimat for middelværdien ved den nye metode målt antallet af revner på 12 kabelstykker på hver 10 meter.

Resultaterne var

Kabel nr.	1	2	3	4	5	6	7	8	9	10	11	12
Antal revner	8	4	14	6	8	10	10	16	2	2	6	8

4) Angiv på basis heraf et estimat for middelværdien af antal revner p. 10 m kabel.

LØSNING:

Lad X betegne antallet af revner i 1 meter kabel. Idet vi med tilnærmelse kan antage, at betingelserne i sætning 7.2 er opfyldt (impuls er her ridser), er X Poissonfordelt $p(\mu)$. Da det gennemsnitlige antal revner p. 1m kabel er $\mu = \frac{12.3}{10} = 1.23$ fås:

$$1) P(X=0) = \frac{1.23^0}{0!} \cdot e^{-1.23} = \underline{0.292}.$$

$$2) P(X \geq 2) = 1 - P(X \leq 1) = 1 - \frac{1.23^0}{0!} \cdot e^{-1.23} - \frac{1.23^1}{1!} \cdot e^{-1.23} = \underline{0.3482}$$

$$\text{TI-89: } P(X \geq 2) = 1 - P(X \leq 1) = 1 - \text{PoissonCDF}(1.23, 0, 1) = \underline{0.3482}$$

3) Lad Y være antal revner i 2 meter kabel. Da der i gennemsnit er 2,46 revner i 2 meter kabel, er 2.46 et estimat for μ . Vi har derfor

$$P(X \leq 4) = \text{poissonCDF}(2.46, 0, 4) = \underline{0.8965}$$

$$\text{eller } P(X \leq 4) = \frac{2.46^0}{0!} \cdot e^{-2.46} + \frac{2.46^1}{1!} \cdot e^{-2.46} + \dots + \frac{2.46^4}{4!} \cdot e^{-2.46} = \underline{0.8965}$$

4) Der er i alt 94 revner i 12 kabelstykker på hver 10 meter. Et estimat for μ er derfor

$$\tilde{\mu} = \frac{94}{12} = \underline{7.83}.$$



Hypotesetest for Poissonfordelt variabel. Som ved binomialfordelt variabel kan test involverende Poissonfordelt variabel direkte udføres, hvis man har rådighed over en tabel med kumuleret Poissonfordeling (den i tabel 7 angivne er sædvanligvis ikke omfattende nok) eller over en lommeregner med kumuleret Poissonfordeling. Som beskrevet i næste afsnit er det dog ofte muligt at approksimere med en normalfordeling, og i så fald er formlerne en simpel oversættelse fra de normalfordelte variable. Disse formler ses i appendix 5.6 og de tilsvarende konfidensintervaller kan findes i appendix 4.1, punkt 6. Det følgende eksempel viser hvordan.

Eksempel 7.9. Ensidede Poisson-test.

I eksempel 7.8 betragtede vi mikroskopiske revner i et kobberkabel. Fabrikken gik over til en anden og billigere produktionsmetode.

- 1) Test, om den nye metode giver færre revner end den gamle metode.
- 2) Angiv et 95% konfidensinterval for middelværdien μ af antal revner p. 10 meter kabel.
- 3) Angiv et 95% konfidensinterval for middelværdien μ_1 af antal revner p. 120 meter kabel.

Vigtige diskrete fordelinger

LØSNING:

- 1) Lad X betegne antallet af revner i 120 meter kabel ved ny metode

Med tilnærmelse kan antages, at X er Poissonfordelt $p(\mu)$, hvor vi i eksempel 7.8 fandt at et estimat for μ var $\tilde{\mu} = 94$.

Ved gammel metode er antal revner i 120 m kabel i middel $12.3 \cdot 12 = 147.6$

Nulhypotese $H_0: \mu = 147.6$ mod den alternative hypotese $H: \mu < 147.6$.

P -værdi $= P(Y \leq 94) = \text{PoissonCdf}(147.6, 0, 94) = 0.000002$

Da P -værdi < 0.05 forkastes nulhypotesen (trestjernet), dvs. vi er sikre på, at middelantallet af revner er blevet formindsket ved at anvende den nye metode

Tabel:

Da $n \cdot \mu_0 = 12 \cdot 12.3 = 147.6 > 10$ kan der approksimeres med normalfordelingen.

Ifølge appendix 5.6 dannes

$$u = \frac{n \cdot \bar{x} + \frac{1}{2} - n \cdot \mu_0}{\sqrt{n \cdot \mu_0}} = \frac{12 \cdot 7.83 + \frac{1}{2} - 12 \cdot 12.3}{\sqrt{12 \cdot 12.3}} = -4.37.$$

Da $u_{0.9995} = 3.291$ forkastes nulhypotesen trestjernet, dvs. vi er sikre på, at middelantallet af revner er blevet formindsket ved at anvende den nye metode

- 2) Idet $m = 94 > 10$ kan formel 6 i appendix 4.1 anvendes. Et 95% konfidensinterval for μ er

$$\bar{x} \pm u_{0.975} \cdot \sqrt{\frac{\bar{x}}{n}} = 7.83 \pm 1.96 \cdot \sqrt{\frac{7.83}{12}}. \quad \underline{\underline{[6.25 ; 9.41]}}.$$

- 3) Et 95% konfidensinterval for μ_1 er

$$\bar{x} \pm u_{0.975} \cdot \sqrt{\frac{\bar{x}}{n}} = 94 \pm 1.96 \cdot \sqrt{\frac{94}{1}}. \quad \underline{\underline{[75 ; 113]}}.$$



7.5 Den generaliserede hypergeometriske fordeling.

Den hypergeometriske fordeling benyttes som model ved stikprøveudtagning uden tilbagelægning, hvor hvert element har enten en bestemt egenskab (defekt) eller ikke har denne egenskab (ikke defekt). Hvis der foreligger flere end to egenskaber, f.eks. udtagning af møtrikker, hvis diameter enten tilhører et givet toleranceinterval eller er for stor eller for lille, kan man generalisere den hypergeometriske fordeling. Dette illustreres ved følgende eksempel:

Eksempel 7.10. Generaliseret hypergeometrisk fordeling.

I en urne findes 12 kugler, hvoraf 5 er sorte, 4 er hvide og 3 er røde.

Vi betragter det tilfældige eksperiment: "Udtrækning af 6 kugler uden tilbagelægning og observation af farven på kuglerne". Beregn sandsynligheden for at få 2 sorte, 3 hvide og 1 rød kugle.

LØSNING:

Lad X_1 være antallet af sorte kugler, X_2 være antallet af hvide kugler og X_3 være antallet af røde kugler.

Analogt med begrundelsen for den hypergeometriske fordeling fås:

$$P(X_1 = 2, X_2 = 3, X_3 = 1) = \frac{\binom{5}{2} \cdot \binom{4}{3} \cdot \binom{3}{1}}{\binom{12}{6}} = \frac{10 \cdot 4 \cdot 3}{924} = \underline{0.13}$$



7.6 Polynomialfordelingen.

Binomialfordelingen benyttes som model ved uafhængige gentagelser af samme eksperiment. Eksperimentet har to udfald succes eller ikke succes og der er en konstant sandsynlighed for succes. Hvis der foreligger flere end to udfald, f.eks. udtagning af møtrikker fra en løbende produktion, hvor diameter enten tilhører et givet toleranceinterval eller er for stor eller for lille, kan man generalisere til polynomialfordelingen. Idet formelen for binomialfordelingen kan skrives

$$f(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} = \frac{n!}{x! \cdot (n-x)!} \cdot p^x \cdot (1-p)^{n-x} = \frac{n!}{x_1! \cdot x_2!} \cdot p_1^{x_1} \cdot p_2^{x_2}, \text{ hvor}$$

$p_1 + p_2 = 1$ og $x_1 + x_2 = n$ fås analogt

DEFINITION af polynomialfordeling.

Lad n være et positivt helt tal, og lad $p_1 + p_2 + \dots + p_k = 1$ og $x_1 + x_2 + \dots + x_k = n$ hvor alle p er positive tal og alle x er hele tal.

Sandsynlighedsfordelingen for en polynomialfordelt stokastisk variabel $(X_1, X_2, \dots, X_k)$ er

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{x_1! \cdot x_2! \cdot \dots \cdot x_k!} p_1^{x_1} \cdot p_2^{x_2} \cdot \dots \cdot p_k^{x_k}$$

Dette illustreres ved følgende eksempel:

Eksempel 7.11. Polynomialfordelingen

En stor produktion af glaskugler indeholder 40% sorte, 35% hvide og 25% røde kugler.

Vi betragter det tilfældige eksperiment: "Udtrækning af 6 kugler observation af farven på kuglerne".

Beregn sandsynligheden for at få 2 sorte, 3 hvide og 1 rød kugle.

LØSNING:

Lad X_1 være antallet af sorte kugler, X_2 være antallet af hvide kugler og X_3 være antallet af røde kugler.

Vi får nu
$$P(X_1 = 2, X_2 = 3, X_3 = 1) = \frac{6!}{2! \cdot 3! \cdot 1!} 0.4^2 \cdot 0.35^3 \cdot 0.25^1 = \underline{0.1029}$$



7.7 APPROKSIMATIONER

Vi har undertiden benyttet os af, at det under visse forudsætninger er muligt med en rimelig nøjagtighed, at foretage approksimationer, f.eks. at approksimere en binomialfordeling med en normalfordeling. I appendix 7.1 er angivet en samlet oversigt over de mulige approksimationer.

Approksimation af hypergeometrisk fordeling med binomialfordeling. At erstatte den hypergeometriske fordeling $h(N, M, n)$ med binomialfordelingen $b(n, p)$ vil for de fleste anvendelser kunne gøres med en passende nøjagtighed, hvis stikprøvestørrelsen n er mindre end eller lig 10% af partistørrelsen N ($n \leq \frac{N}{10}$). I så fald sættes i binomialfordelingen $p = \frac{M}{N}$.

Eksempel 7.12. Approksimation af hypergeometrisk fordeling til binomialfordeling.

I eksempel 7.1, hvor man udtog 25 komponenter fra æsker på 600 komponenter, skete udtagningen logisk nok uden tilbagelægning. Imidlertid er det klart, at da æskerne indeholder mange komponenter vil sandsynligheden for at få en defekt ikke ændrer sig meget, hvis man i stedet havde foretaget udtagningen med tilbagelægning. Der blev antaget, at der var 10 defekte i en sådan æske med 600, og dette antal defekte vil så være konstant, under hver udtrækning.

Vi har derfor, at $P(\text{at få en defekt}) = \frac{10}{600} = \frac{1}{60}$. Betingelserne for at benytte binomialfordelingen er nu til stede.

Løsningen af problemet i eksempel 7.1 vil derfor nu være:

$$\begin{aligned} p_a &= P(X \leq 1) = P(X = 0) + P(X = 1) \\ &= \binom{25}{0} \cdot \left(\frac{1}{60}\right)^0 \cdot \left(1 - \frac{1}{60}\right)^{25} + \binom{25}{1} \cdot \left(\frac{1}{60}\right)^1 \cdot \left(1 - \frac{1}{60}\right)^{24} = 0.6569 + 0.2784 = \underline{\underline{0.9353}} \end{aligned}$$

Det ses, at vi får praktisk samme resultat som i eksempel 7.1.



Approksimation af binomialfordeling til Poissonfordeling. Som det fremgår af det i “Supplement til statistiske grundbegreber” afsnit 7C angivne bevis for sætning 7.3 kan en binomialfordeling $b(n, p)$ approksimeres med en Poissonfordeling $p(\mu)$, hvor $\mu = n \cdot p$, hvis p er lille (og dermed n stor)¹⁾. I praksis vil en approksimation være tilstrækkelig god, hvis blot $p \leq \frac{1}{10}$.

Eksempel 7.13. Approksimation af hypergeometrisk fordeling til binomialfordeling videre til Poissonfordeling

I eksempel 7.12 betragtede vi følgende kvalitetskontrolproblem, hvor vi beregnede acceptsandsynligheden $p_a = P(X \leq 1)$ dels ud fra den hypergeometriske fordeling $h(600, 10, 25)$, dels ved approksimation til $b(25, p)$, hvor $p = \frac{10}{600} = \frac{1}{60}$.

Denne gang ønskes acceptsandsynligheden beregnet ved anvendelse af Poissonfordelingen.

¹⁾ Matematisk formulering: $P_{b(n,p)}(X = x) \xrightarrow{n \rightarrow \infty} P_{p(\mu)}(X = x)$, hvor $\mu = n \cdot p$

LØSNING:

Da $p = \frac{1}{60} \leq \frac{1}{10}$ kan approksimeres med Poissonfordelingen $p(n \cdot p) = P\left(\frac{25}{60}\right)$

$$p_a = P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{\left(\frac{25}{60}\right)^0}{0!} \cdot e^{-\left(\frac{25}{60}\right)} + \frac{\left(\frac{25}{60}\right)^1}{1!} \cdot e^{-\left(\frac{25}{60}\right)} = \underline{\underline{93.39\%}}.$$

Det ses, at vi får praktisk taget samme resultat som i eksempel 7.1.

**Approximation af binomialfordeling til normalfordeling.**

Det kan vises, at tæthedsfunktionen for binomialfordelingen $b(n, p)$ nærmer sig ubegrænset til normalfordelingen $n(\mu, \sigma)$, hvor $\mu = n \cdot p$ og $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$, når n vokser ubegrænset²⁾.

Approximation af en binomialfordeling med en normalfordeling anses, når $\frac{1}{10} < p < \frac{9}{10}$ i praksis for at være tilfredsstillende, såfremt $n \cdot p \geq 5$ (og $n \cdot (1 - p) \geq 5$), jævnfør også appendix 7.1).

Da binomialfordelingen kun antager heltalsværdier, medens en normalfordeling kan antage alle værdier på talaksen, svarer hvert helt tal ved binomialfordelingen til et interval af længden 1 ved normalfordelingen. På figur 7.1 er derfor tegnet en firkant, der har bredden 1, og hvis højde er $P(X = 4)$ udregnet ved binomialfordelingen. Arealet under normalfordelingskurven fra $x = 3.5$ til $x = 4.5$ er med tilnærmelse lig firkantens areal. Man siger, at man ved approksimationen må heltalskorrigere (korrigeres for kontinuitet).

Ved approksimationen benyttes derfor følgende anførte formler, gældende for en binomialfordelt variabel X fordelt $b(n, p)$, hvor $\frac{1}{10} \leq p \leq \frac{9}{10}$ og $5 \leq n \cdot p \leq n - 5$.

$$P(X \leq x) = \Phi\left(\frac{x + \frac{1}{2} - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}}\right),$$

$$P(X < x) = \Phi\left(\frac{x - \frac{1}{2} - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}}\right)$$

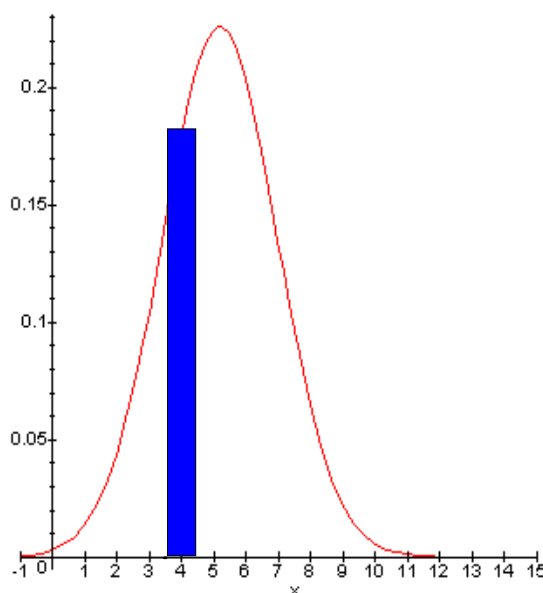


Fig. 7.1. Heltalskorrektion

²⁾ Matematisk formulering: Når X er fordelt $b(n, p)$, vil for den tilsvarende normerede variabel

$$Y = \frac{X - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}} \text{ gælde, at } P(Y \leq y) \xrightarrow{n \rightarrow \infty} \Phi(y) \text{ for ethvert tal } y.$$

Eksempel 7.14: Approximation af binomialfordeling med normalfordeling.

En kunde til de i eksempel 2.3 producerede plastikkasser køber kasserne i partier på 2000. Kunden godkender et parti efter en stikprøvekontrol, hvor der udtages 100 kasser. Hvis antallet af defekte kasser i stikprøven højst er 14 godkendes hele partiet. I modsat fald kasseres partiet.

Hvor stor er sandsynligheden for at et parti bliver godkendt, hvis der er 300 defekte kasser i hele partiet på de 2000.

LØSNING:

Lad X være antallet af defekte kasser i stikprøven. Vi ønsker at udregne $P(X \leq 14)$.

Umiddelbart er X hypergeometrisk fordelt med $N = 2000$, $M = 300$, og $n = 100$.

Da stikprøvestørrelsen er lille $\left(\frac{n}{N} = \frac{100}{2000} < \frac{1}{10}\right)$ kan fordelingen af X umiddelbart approksimeres

med binomialfordelingen $b(100, p)$, hvor $p = \frac{M}{N} = \frac{300}{2000} = 0.15$ (se appendiks 7.1). Dette giver ved benyttelse af en lommeregner som TI-89 at $P(X \leq 14) = 45.72\%$.

Idet $n \cdot p = 15 > 5$, kan i stedet for approksimeres med normalfordelingen med $\mu = 15$ og $\sigma = \sqrt{15 \cdot 0.85} = 3.57$.

Ved hjælp af denne approksimation kan vi beregne:

$$P(X \leq 14) = \text{normCdf}(-\infty, 14.5, 15, 3.57) = 44.43\%$$

Det ses, at der er ca. 1.5 % afvigelse, hvilket normalt ingen betydning har.



Approximation af Poissonfordeling til normalfordeling.

Approximation af en Poissonfordeling med en normalfordeling baseres på, at det gælder, at tæthedsfunktionen for Poissonfordelingen $p(\mu)$ nærmer sig ubegrænset til normalfordelingen $n(\mu, \sigma)$ med samme middelværdi μ og $\sigma = \sqrt{\mu}$ når $\mu \rightarrow \infty$ ³⁾.

Approximation af en Poissonfordeling med en normalfordeling anses i praksis for at være tilfredsstillende, når $\mu \geq 10$, jævnfør den skematiske oversigt i appendix 7.1.

Da Poissonfordelingen kun antager heltalsværdier, medens en normalfordeling kan antage alle værdier på talaksen, må der ved approksimationen benyttes *heltalskorrektion* (korrektur for kontinuitet) helt analogt med den tilsvarende situation ved approksimation af en binomialfordeling med en normalfordeling.

$$\text{Vi får derfor for } \mu \geq 10 \quad P(X \leq x) = \Phi\left(\frac{x + \frac{1}{2} - \mu}{\sqrt{\mu}}\right), \quad P(X < x) = \Phi\left(\frac{x - \frac{1}{2} - \mu}{\sqrt{\mu}}\right)$$

³⁾ Matematisk formulering: Lad X være fordelt $p(\mu)$ For den normerede variabel $Y = \frac{X - \mu}{\sqrt{\mu}}$

gælder, at $P(Y \leq y) \xrightarrow{\mu \rightarrow \infty} \Phi(y)$ for ethvert tal y .

Eksempel 7.15: Approksimation af Poissonfordeling med normalfordeling.

Er X binomialfordelt $b(600, 0.05)$, kan denne erstattes med $p(\mu)$, hvor $\mu = 600 \cdot 0.05 = 30$. Da tabel 7 kun når til $\mu = 16$, kan den ikke anvendes. Da $\mu \geq 10$ kan (se appendix 7.1) denne fordeling dog igen approksimeres nu med normalfordelingen $n(30, \sqrt{30}) = n(30, 5.477)$.

Ønskes beregnet $P(X \leq 24)$ fås $\text{normCdf}(-\infty, 24.5, 30, 5.477) = 15.87\%$

Ved direkte beregning med binomialfordeling bliver resultatet 15.07%, dvs. en forskel på 0.8%.



OPGAVER

Opgave 7.1 (sandsynlighed)

Ved en lodtrækning fordeles 3 gevinster blandt 25 lodsedler. En spiller har købt 5 lodsedler. Beregn sandsynligheden for hver af følgende hændelser:

- 1) Spilleren vinder alle tre gevinster.
- 2) Spilleren vinder ingen gevinster.
- 3) Spilleren vinder netop én gevinst.

Opgave 7.2 (sandsynlighed)

I en urne findes 2 blå, 3 røde og 5 hvide kugler. 3 gange efter hinanden optages tilfældigt en kugle fra urnen uden mellemliggende tilbagelægning.

- 1) Find sandsynligheden for hændelsen A , at der højst optages 2 hvide kugler,
- 2) Find sandsynligheden for hændelsen B , at de optagne kugler har hver sin farve.
- 3) Find sandsynligheden for, at de tre kugler har samme farve,

Opgave 7.3 (diskret fordeling)

En fabrikant fremstiller en bestemt type radiokomponenter. Disse leveres i æsker med 30 komponenter i hver æske. En køber har den aftale med fabrikanten, at hvis en æske indeholder 4 defekte komponenter eller derover, kan køberen returnere æsken, i modsat fald skal den godkendes. Køberen kontrollerer hver æske ved en stikprøve, idet han af æsken udtager 10 komponenter tilfældigt. Lad X være antal defekte i stikprøven. Der overvejes nu to planer:

- 1) Hvis $X = 0$, så godkendes æsken, ellers undersøges æsken nærmere.
- 2) Hvis $X \leq 1$, så godkendes æsken, ellers undersøges æsken nærmere.

Hvad er sandsynligheden for, at en æske, der indeholder netop 4 defekte komponenter, bliver godkendt af køberen ved metode 1 og ved metode 2.

Opgave 7.4 (diskret fordeling)

En sædvanlig benyttet undervisningsmetode A ønskedes sammenlignet med en ny undervisningsmetode B , som formodes at være mere effektiv.

En klasse indeholdende 20 studerende deltes tilfældigt i 2 lige store grupper. Gruppe I undervistes efter metode A og gruppe II efter metode B . Ved en fællesprøve for de 20 studerende bestod 12, mens 8 ikke bestod. I gruppe I bestod kun 3 af de 10 studerende, mens der var 9 personer, der bestod i gruppe II. Lad p være sandsynligheden for, at højst 3 af de 10 studerende bestod i gruppe I, såfremt det forudsættes, at de to undervisningsmetoder er lige effektive. Hvis p er under 1%, vil man påstå, at de to undervisningsmetoder ikke er lige effektive, dvs. at metode B er mere effektiv end metode A . Er metode B mere effektiv end metode A ?

Opgave 7.5 (fordeling)

En urmager sælger sædvanligvis et parti ure for 100 kr stykket. I forbindelse med et udsalg arrangerer han følgende lotteri:

Fra en tromle, hvori der befinder sig 20 sedler, skal kunderne trække én seddel. På hver seddel står salgssummen (x kr) anført. Følgende skema viser sedlernes fordeling:

Antal sedler	9	5	3	2	1
x kr	100	75	50	25	0

- 1) Idet X angiver den tilsvarende statistiske variabel, skal man angive fordelingsfunktionen for X , middelværdien $E(X)$ og spredningen $\sigma(X)$.
- 2) En kunde køber 3 ure, idet hver udtrukket seddel lægges tilbage i tromlen, inden ny udtrækning foretages.
 - a) Hvad må kunden forventes i gennemsnit at skulle betale for de tre ure?
 - b) Hvad er sandsynligheden for, at kunden får rabat (betaler mindre end 100 kr) på mindst 2 af urene?

Opgave 7.6 (fordeling)

En tipskupon har 13 kampe med 3 mulige tegn - 1, x og 2 - for hver kamp. En person bestemmer tegnet, der skal sættes for hver kamp, ved tilfældig udtrækning af en seddel fra 3 sedler med tegnene henholdsvis 1, x og 2. Angiv sandsynligheden for, at personen opnår netop 8 rigtige tippede kampe på sin kupon.

Opgave 7.7 (fordeling)

I en urne er der et meget stort antal kugler, hvoraf de 70% er sorte. Fra urnen tages en stikprøve på 10 kugler. Find sandsynligheden for, at der i stikprøven er:

- 1) 10 sorte kugler
- 2) 6, 7 eller 8 sorte kugler
- 3) Mindst 7 sorte kugler

Opgave 7.8 (fordeling)

Ved et køb af 100000 plastikbægre aftales med leverandøren, at det skal være en forudsætning for købet, at partiet godkendes ved en stikprøvekontrol.

Kontrollen udøves ved, at 100 bægre udtages tilfældigt af partiet og kontrolleres. Partiet godkendes, såfremt ingen af de 100 bægre er defekte.

Beregn sandsynligheden for, at partiet godkendes, hvis det i alt indeholder 250 defekte bægre.

Opgave 7.9 (fordeling)

I et elektrisk specialapparat indgår 30 komponenter, som hver er indkapslet i et heliumfyldt hylster. Beregn, idet sandsynligheden for, at et komponenthylster lækker, er 0.2%, sandsynligheden for, at mindst ét af de 30 komponenthylstre lækker.

Opgave 7.10 (fordeling)

En fabrikant får halvfabrikata hjem i partier på 200000 enheder. Fra hvert parti udtages en stikprøve på 100 enheder og antallet af fejlagtige blandt disse noteres.

Hvis dette antal er mindre end eller lig med 2, accepteres hele partiet; i modsat fald undersøges partiet yderligere.

- 1) Hvad er sandsynligheden for, at et parti med en fejlprocent på 1 vil blive yderligere undersøgt.
- 2) Hvor stor er sandsynligheden for, at et parti med en fejlprocent på 5 vil blive accepteret.

Opgave 7.11 (fordeling).

Ved en fabrikation af plastikposer leveres disse i æsker med 100 poser i hver. Ved en godkendelseskontrol af et parti plastikposer udtages og undersøges en tilfældigt udtaget æske, og partiet godkendes, såfremt æsken højst indeholder én defekt pose.

Vi antager, at den løbende produktion af poser er således, at hver produktion med sandsynligheden 2% giver en pose, der er defekt; vi vil senere formulere dette således, at produktionen er i **statistisk kontrol** med fejlsandsynligheden $p = 2\%$.

Hvor stor er sandsynligheden for, at partiet under disse omstændigheder accepteres?

Opgave 7.12 (fordeling).

En maskinfabrik fremstiller “ens” emner i masseproduktion; nogle af de producerede emner er defekte. En køber af et større parti emner gør handelen betinget af resultatet af følgende kontrol af partiets kvalitet: Der udtages tilfældigt ét emne ad gangen fra partiet; dette godkendes, såfremt man må udtage mere end 100 emner, før 2 defekte emner er udtaget (d.v.s. hvis der er højst 1 fejlemne blandt de 100 først udtagne emner); i modsat fald kasseres partiet.

Beregn sandsynligheden p for, at partiet godkendes, når det har en defektprocent på 4.

Opgave 7.13 (fordeling).

En kæde består af et antal ens led. Trækstyrken X (i kg) for ethvert led i kæden er en stokastisk variabel, som har fordelingsfunktionen:

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{x^2}{25} & \text{for } 0 \leq x < 5 \\ 1 & \text{ellers} \end{cases}$$

- 1) Angiv sandsynligheden for, at trækstyrken for en kæde med 10 led mindst er 1 kg.
- 2) Hvor mange led må kæden højst bestå af, hvis der skal være en sandsynlighed på mindst 80% for, at kædens trækstyrke er større end 1 kg.

Opgave 7.14 (fordeling).

Sandsynligheden for, at en vis frøsort spirer, er 0.6. I en pose sås 3 frø og i en anden potte 2 frø. Beregn sandsynligheden for, at der spirer flere planter i potten med 2 frø end med potten med 3 frø.

Opgave 7.15 (binomialfordeling)

En producent af billigt plastiklegetøj får mange klager over at en bestemt type legetøj er defekt ved salget. Legetøjet sælges til butikkerne i kasser på 10 stk, og som et led i en kvalitetetskontrol udtages 100 kasser og antallet x af defekt legetøj optaltes. Følgende resultater fandtes:

x	0	1	2	3	4	5	6
Antal kasser	34	38	19	6	2	0	1

Lad p være sandsynligheden for at få et defekt stykke legetøj.

- 1) Find et estimat $\tilde{p}$ for p .
- 2) Angiv et 95% konfidensinterval for p .
- 3) Lad X være antal defekte i en kasse på 10 stykker legetøj, og antag at X er binomialfordelt $b(10, \tilde{p})$. Beregn hvor mange af de 100 kasser, der kan forventes at have $x = 2$ defekte.

Opgave 7.16 (binomialfordeling)

En ny behandling af cancer forventes at give bedre overlevelseschancer end den hidtidige behandling. 120 patienter prøvede den nye behandling, og af disse overlevede 82 i mere end 5 år.

Idet antallet af overlevende patienter antages at være binomialfordelt, skal man

- 1) Angive et estimat for sandsynligheden p for at overleve i 5 år ved den nye behandling.
- 2) Angive et 95% konfidensinterval for p .

Opgave 7.17 (binomialfordeling)

Af 1000 tilfældigt udvalgte patienter, der led af lungekræft, var 823 døde senest 5 år efter sygdommen blev opdaget.

Angiv på dette grundlag et 95% konfidensinterval for sandsynligheden for at dø af denne sygdom senest 5 år efter at sygdommen bliver opdaget.

Opgave 7.18 (binomialfordeling)

En fabrikant af lommeregnere er interesseret i at få et skøn over hvor stor en procentdel p af de producerede lommeregnere, der er defekte. En stikprøve på 800 lommeregnere indeholder 10 defekte. Beregn et 95% konfidensinterval for p .

Opgave 7.19 (binomialtest).

En ny vaccine formodes med en sandsynlighed på mindst 85% at have en forebyggende virkning over for en bestemt influenzatype. Før en truende influenzaepidemi vaccineres et hospitalspersonale på 600 personer med den pågældende vaccine. 125 af disse bliver smittet af sygdommen. Kan dette opfattes som en eksperimentel påvisning af, at vaccinen er mindre virksom end ventet?

Opgave 7.20 (binomialtest).

- 1) Antag, at en vis type af fostermisdannelse normalt forekommer med hyppigheden 164 tilfælde p. 100000 fødsler. Beregn sandsynligheden for 3 eller flere fostermisdannelser blandt 256 fødsler.
- 2) For at undersøge om forholdene i et bestemt arbejdsmiljø forøger hyppigheden af denne type misdannelse, undersøgte man hyppigheden af misdannelser for mødre, som under graviditeten havde haft den aktuelle type af arbejde, og fandt 3 misdannelser blandt 256 fødsler. Kan den forøgede relative hyppighed i dette materiale skyldes tilfældigheder?

Opgave 7.21 (binomialtest).

Udsættes planterne af en bestemt sort roser for meldugssmitte, bliver i middel brøkdelen p angrebet, hvor p er mindst 0.20. En rosengartner fremavler en rosenstamme, som han påstår er mere modstandsdygtig over for meldugssmitte. For at kontrollere denne påstand bliver 100 roser af den nye stamme udsat for meldugssmitte. Det viser sig, at 12 roser bliver angrebet.

- 1) Bekræfter dette resultat rosengartnerens påstand? (Husk altid at anføre: Hvad X er. Antagelser. Nulhypotese. Beregninger. Konklusion.).
- 2) Angiv et estimat $\tilde{p}$ for den nye stammes p .
- 3) Angiv et 95% konfidensinterval for den nye stammes p .

Opgave 7.22 (Poissonfordeling)

På en fabrik fremstilles gulvtæpper, som har størrelsen 20 m^2 . Ved fabrikationen er der gennemsnitlig 6 vævefejl p. 100 m^2 klæde.

- 1) Beregn sandsynligheden for, at et tilfældigt gulvtæppe ingen vævefejl har.
- 2) Beregn sandsynligheden for, at et tilfældigt gulvtæppe højst har 2 vævefejl.

Fabrikken køber en ny væv. For at få et estimat for middelværdien målttes antallet af vævefejl i 12 gulvtæpper hver på 20 m^2 . Resultaterne var

Gulvtæppe nr	1	2	3	4	5	6	7	8	9	10	11	12
Antal vævefejl	4	2	7	3	4	5	5	8	1	1	3	5

- 3) Find et estimat for middelværdien af antal vævefejl p. 20 m^2 klæde.

Opgave 7.23 (Poissonfordeling)

Et radioaktivt præparat undergår gennemsnitligt 100 desintegrationer (sønderdelinger) p. minut. Lad X betegne antal desintegrationer i et sekund (som er lille i forhold til præparatets halveringstid).

Find $P(X \leq 1)$.

Opgave 7.24 (Poissonfordeling)

Ved en TV-fabrikation optælles som led i en godkendelseskontrol antal loddefejl p. 5 TV-apparater. Fabrikanten ønsker at få et overblik over antal loddefejl, og optalte derfor antal loddefejl på 24 tilfældigt udtagne TV apparater. Resultatet fremgår af skemaet:

Antal loddefejl	0	1	2	3	4	5	6	7	8	9
Antal TV apparater	3	2	4	6	5	2	1	0	1	0

Lad X være antallet af loddefejl i 5 TV apparater.

- 1) Angiv den sandsynlighedsfordeling X approksimativt kan antages at følge, og giv et estimat for parameteren i fordelingen.
- 2) Beregn på basis af svaret i spørgsmål 1 sandsynligheden for, at der på 5 tilfældigt udtagne TV-apparater højst er i alt 18 loddefejl?

Opgave 7.25 (Poissonfordeling)

På et teknisk universitet er et centralt edb-anlæg i konstant brug. Man har erfaring for, at anlægget i løbet af en 20 ugers periode har gennemsnitligt 7 maskinstop. Beregn sandsynligheden p for, at anlægget i en 4 ugers periode har mindst ét maskinstop.

Opgave 7.26 (Poissonfordeling)

På en fabrik indtræffer i gennemsnit 72 ulykker om året. Antag, at de forskellige ulykker indtræffer uafhængigt af hinanden, og at de er nogenlunde jævnt fordelt over året. Beregn, idet et arbejdsår sættes lig med 48 uger, sandsynligheden for at der i en uge indtræffer flere end 3 ulykker.

Opgave 7.27 (Poissonfordeling)

Til et bestemt telefonnummer er der i løbet af aftenen i middel 300 opkald i timen. Beregn sandsynligheden for, at der i løbet af et minut er højst 8 opkald.

Opgave 7.28 (Poissonfordeling)

I en kantine fremstilles 5 kg rejesalat. Hertil medgår 400 rejer. Salaten fordeles på 100 stk. smørrebrød med 50 g salat på hvert stykke. Find sandsynligheden for, at der på et stykke med rejesalat er over 8 rejer.

Opgave 7.29 (fordeling)

Et bestemt, meget lille volumen blod udtaget fra et menneske, som ikke lider af mangel på røde blodlegemer, vides i middel at indeholde 5 røde blodlegemer. Beregn sandsynligheden for, at en blodprøve udtaget fra et menneske kun indeholder 3 eller færre røde blodlegemer.

Opgave 7.30 (fordeling)

En fabrikation af fortinnede plader finder sted ved en kontinuerlig elektrolytisk proces. Umiddelbart efter produktionen kontrolleres for pladefejl. Man har erfaring for, at der i middel er 1 pladefejl hvert 5'te minut.

Beregn sandsynligheden for, at der højst er 5 pladefejl ved en halv times produktion.

Opgave 7.31 (Poissonfordeling)

Ved inspektion af en produktion med isolering af kobberledning taltes der i løbet af 50 minutter i alt 11 isoleringsfejl.

Idet antallet af isoleringsfejl p. 50 minutter antages at være Poissonfordelt $p(\mu_1)$, skal man

1a) angive et estimat for μ_1 .

1b) angive et 95% konfidensinterval for μ_1 .

Det oplyses nu, at man i hver 5 minutters periode i den ovenfor omtalte 50 minutters periode havde observeret følgende antal isoleringsfejl:

Periode	1	2	3	4	5	6	7	8	9	10
Antal fejl	1	0	2	2	1	1	3	0	1	0

Idet antallet af isoleringsfejl p. 5 minutter antages at være Poissonfordelt $p(\mu_2)$, skal man

- 2a) angive et estimat for μ_2 .
- 2b) angive et 95% konfidensinterval for μ_2 .

Opgave 7.32 (Poissonfordeling)

Nedenstående tabel viser fordelingen af 400 volumenenheder med hensyn til antal gærceller p. volumenenhed.

Antal gærceller	0	1	2	3	4	5	6	7	8	9	10	11	12
Antal volumenenheder	0	20	43	53	86	70	54	37	18	10	5	2	2

Lad X være antal gærceller p. volumenenhed. Det antages, at X er en stokastisk variabel der er Poissonfordelt $p(\mu)$.

- 1) Find et estimat $\tilde{\mu}$ for μ .
- 2) Angiv et 95% konfidensinterval for μ .
- 3) Forudsat at X er Poissonfordelt $p(\tilde{\mu})$ ønskes beregnet det forventede antal volumenenheder, hvori der forekommer 5 gærceller (for $x = 5$).

Opgave 7.33 (Poisson-test).

Et radioaktivt præparat formodes at udsende 500 α -partikler p. minut. Ved en undersøgelse af præparatet optaltes med en Geigertæller i løbet af et minut 475 counts. Test, om dette er i modstrid med det på forhånd forventede?

Opgave 7.34 (Poisson-test).

Lastbiler med affald ankommer tilfældigt og indbyrdes uafhængigt til en losseplads. Lossepladsens maksimale kapacitet er beregnet til, at der i middel ankommer 90 lastbiler p. time. Ledelsen af pladsen føler, at travlheden er blevet større i den sidste tid, således at antallet af lastbiler overskrider den maksimale kapacitet. For at undersøge dette, foretages en optælling af lastbiler i perioder à 10 minutter. Følgende resultater fremkom:

13	16	17	15	18	12	22	16	21	18
----	----	----	----	----	----	----	----	----	----

- 1) Bekræfter disse resultater ledelsens formodning? (Husk altid at anføre: Hvad X er. Antagelser. Nulhypotese. Beregninger. Konklusion.).
- 2) Angiv et estimat $\tilde{\mu}$ for middelværdien μ [lastbiler/time].
- 3) Angiv et 95% konfidensinterval for middelværdien μ [lastbiler/time].

Opgave 7.35 (approksimation)

En "sympigetipper" (M/K) deltog i tipning 42 gange i løbet af et år. På hver tipskupon var der 13 kampe, ved hver af hvilke tipperen ved systematisk gætning satte et af de 3 tegn: 1, x, 2. Beregn sandsynligheden p for, at tipperen det pågældende år tippede mindst 200 kampe rigtigt.

Opgave 7.36 (approksimation)

Blandt familier med 3 børn udvælges 50 familier tilfældigt. Angiv sandsynligheden for, at der i mindst 8 af disse familier udelukkede er børn af samme køn.

Opgave 7.37 (approksimation)

En maskinfabrikant påtænker at købe 100000 møtrikker af en bestemt type. Man beslutter sig til at købe et tilbudt parti af den nævnte størrelse, såfremt en stikprøve på 150 møtrikker højst indeholder 4% defekte møtrikker.

- 1) Beregn sandsynligheden for, at partiet bliver godkendt af maskinfabrikken, såfremt det indeholder
 - a) 4% defekte møtrikker,
 - a) 2,5% defekte møtrikker,
 - a) 7,5% defekte møtrikker,
- 2) Bestem, for hvilken procentdel defekte møtrikker det ovennævnte parti (approksimativt) har 50% sandsynlighed for at blive godkendt af maskinfabrikken.

Opgave 7.38 (approksimation)

Det er oplyst, at der for en given vaccine er 80% sandsynlighed for, at den ved anvendelse har den ønskede virkning.

På et hospital foretoges vaccination af 100 personer med den pågældende vaccine.

Beregn sandsynligheden for, at 15 eller færre af de foretagne vaccinationer er uden virkning.

Opgave 7.39 (approksimation af fordeling)

En virksomhed fremstiller chargevis et pulverprodukt indeholdende en bakteriekultur. For en bestemt charge udførtes 4 gange følgende eksperiment:

I laboratoriet opløstes 5g pulver fra chargen i 50 ml væske, som fortyndedes stærkt (faktor 10^8); dernæst "spredtes" 2 ml af fortyndingen ud i et næringssubstrat i en petriskål. Fortyndingen er så stor, at bakteriecellerne forekommer enkeltvis i fortyndingen, og hver danner udgangspunkt for en bakteriekoloni. Efter en vis henstandstid (inkubationstid) optaltes antal bakteriekolonier i petriskålen.

Resultaterne var:

Prøve	1	2	3	4
Antal	85	62	111	78

Lad X betegne antal bakteriekolonier p. 2 ml fortynding i petriskålen.

- 1) Angiv den sandsynlighedsfordeling X approksimativt kan antages at følge, og giv et estimat for parameteren i fordelingen.
- 2) Beregn på basis af svaret i spørgsmål 1 sandsynligheden for, at der er mindst 100 bakteriekolonier i skålen.

Opgave 7.40 (approximation af fordeling)

En bestemt persons blod indeholder $6 \cdot 10^6$ hvide blodlegemer p. liter.

- a) Antag, at man udtager en prøve på 1 ml blod. Hvad er sandsynligheden for, at prøven indeholder færre end 5900 hvide blodlegemer.
- b) Antag, at man først udtager en prøve på 1 ml blod og fortynder denne til 1 liter med en kogsaltopløsning. Derefter udtages fra denne blanding en prøve på 1 ml. Hvad er sandsynligheden for, at denne prøve indeholder færre end 4 hvide blodlegemer?

Opgave 7.41 (binomialtest).

En fabrikant af chip til computere reklamerer med, at højst 2% af en bestemt type chip, som fabrikken sender ud på markedet er defekte.

Et stort computerfirma, vil købe et meget stort parti af disse chip, hvis påstanden er rigtig. For at teste påstanden købes 1000 af dem. Det viser sig, at 33 ud af de 1000 er defekte.

Kan fabrikantens påstand på denne baggrund forkastes på signifikansniveau 5% ?

Opgave 7.42 (fordeling)

I en urne findes 10 røde kugler, 5 hvide kugler og 3 sorte kugler. 6 gange efter hinanden optages tilfældigt en kugle fra urnen.

Bestem sandsynligheden for, at der i alt er optaget 1 rød, 2 hvide og 3 sorte kugler, når

- 1) kuglerne optages uden tilbagelægning
- 2) kuglerne optages med tilbagelægning.

Opgave 7.43 (fordeling)

En virksomhed fabrikkerer farvede glasklodser til dekorationsbrug. Defekte glasklodser frasorteres. Man har erfaring for, at af de frasorterede klodser har i middel 50% kun revner, 35% kun farvefejl, medens resten har begge disse fejl.

Beregn sandsynligheden for, at af 12 tilfældige defekte klodser har 6 kun revner, 4 kun farvefejl og 2 begge disse fejl.

Opgave 7.44 (fordeling)

I et kortspil med de sædvanlige 52 spillekort har en spiller modtaget 13 kort. Angiv i procent med 2 decimaler sandsynligheden for, at 3 af disse er esser og 5 er billedkort.

8 ANDRE KONTINUERTE FORDELINGER

8.1 INDLEDNING

Vi vil i dette kapitel kort orientere om en række fordelinger, som er vigtige i specielle sammenhænge, men som ikke spiller nogen afgørende rolle for de statistiske problemstillinger vi vil beskæftige os med i det følgende.

8.2 DEN REKTANGULÆRE FORDELING

Eksempel 8.1 Kontinuert variabel.

Lad randen af en roulette være ækvidistant inddelt efter en skala fra 0 til 12, jævnfør figuren.

Ved et roulettespil bringes roulettens viser til at rotere, hvorefter den standser ud for et tilfældigt punkt på skalaen. Lad X være det tal som roulettens viser peger på. Idet X må kunne antage ethvert tal mellem 0 og 12, må X være en kontinuert variabel.

Angiv tæthedsfunktion og fordelingsfunktion for X og skitser disse.

LØSNING:

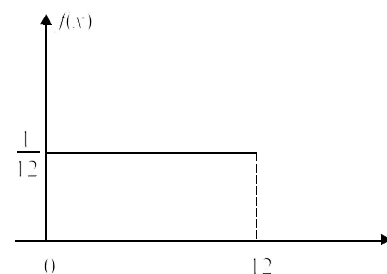
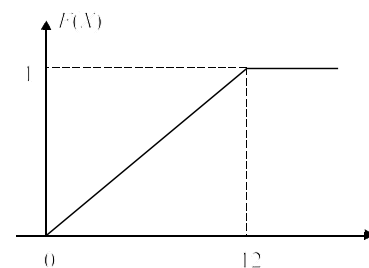
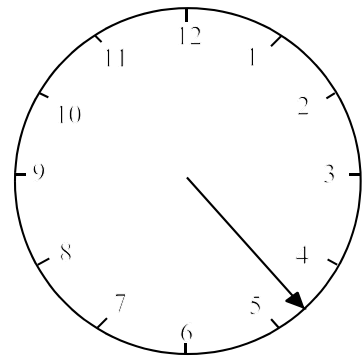
Da $P(0 \leq X \leq x) = \frac{x}{12}$ for $0 \leq x \leq 12$

er fordelingsfunktionen for X

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ \frac{x}{12} & \text{for } 0 \leq x \leq 12 \\ 1 & \text{for } x \geq 12 \end{cases}$$

Ved differentiation fås tæthedsfunktionen

$$f(x) = \begin{cases} \frac{1}{12} & \text{for } 0 \leq x \leq 12 \\ 0 & \text{ellers} \end{cases}$$



Eksempel 8.1 giver et eksempel på den rektangulære fordeling.

DEFINITION af rektangulær fordeling med parametrene a og b . Lad a og b være to reelle tal, hvor $a < b$. Sandsynlighedsfordelingen for en kontinuert stokastisk variabel X med tæthedsfunktionen $f(x)$ bestemt ved

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{ellers} \end{cases} \quad \text{siges at være}$$

rektangulært fordelt $\text{rekt}(a, b)$.

SÆTNING 8.1. (Middelværdi og spredning for rektangulær fordeling).

Den rektangulære fordeling har $E(X) = \frac{a+b}{2}$ og $\sigma(X) = \frac{b-a}{2\sqrt{3}}$ ($a < b$)

Bevis:

$$E(X) = \int_a^b \frac{x}{b-a} dx = \left[\frac{x^2}{2 \cdot (b-a)} \right]_a^b = \frac{b+a}{2}$$

$$V(X) = \int_a^b \frac{\left(x - \frac{a+b}{2}\right)^2}{b-a} dx = \left[\frac{\left(x - \frac{a+b}{2}\right)^3}{3(b-a)} \right]_a^b = \frac{(b-a)^2}{12}$$



8.3 EKSPONENTIALFORDELINGEN

I kapitel 7 betragtede vi antallet N af revner pr. meter langs et kobberkabel. Vi antog, at N var Poissonfordelt. Hvis vi i stedet havde betragtet afstanden X mellem revnerne, havde vi fået en ny stokastisk variabel, som må være kontinuert. Som det fremgår af følgende sætning er X eksponentialfordelt.

SÆTNING 8.2 (Eksponentialfordeling). Lad W være en Poissonfordelt stokastisk variabel. Lad det gennemsnitlige antal impulser i en tidsenhed være λ . Lad X være tiden indtil næste impuls. X er da en kontinuert stokastisk variabel med sandsynlighedsfordelingen (tæthedsfunktionen) $f(x) = P(X=x)$ bestemt ved

$$f(x) = \begin{cases} \frac{1}{\mu} \cdot e^{-\frac{x}{\mu}} & \text{for } x > 0 \\ 0 & \text{ellers} \end{cases} \quad \text{hvor } \mu = \frac{1}{\lambda}$$

X siges at være eksponentialfordelt $\exp(\mu)$ med parameteren μ .

Middelværdien for $\exp(\mu)$ er $E(X) = \mu$ og spredningen er $\sigma(X) = \mu$.

Bevis:

I tidsrummet fra x_0 til $x_0 + x$ er der i gennemsnit $\lambda \cdot x$ impulser. Lad W være det aktuelle antal impulser i tidsrummet $[x_0; x_0 + x]$. W er da Poissonfordelt $p(\lambda \cdot x)$.

Idet X er tiden fra én impuls til den næste, er $P(X > x) = P(W = 0)$, da der ingen impulser er i tidsrummet $[x_0; x_0 + x]$.

Da $P(W = 0) = \frac{(\lambda \cdot x)^0}{0!} \cdot e^{-\lambda \cdot x} = e^{-\lambda \cdot x}$, er $P(X > x) = e^{-\lambda \cdot x}$.

Vi har derfor $F(x) = P(X \leq x) = 1 - P(X > x) = 1 - e^{-\lambda \cdot x}$. Ved differentiation fås tæthedsfunktionen: $f(x) = F'(x) = \lambda \cdot e^{-\lambda \cdot x}$. Sættes $\lambda = \frac{1}{\mu}$ fås formelen.



Bevis for middelværdi og spredning:

$$E(X) = \int_0^{\infty} \lambda \cdot x \cdot e^{-\lambda \cdot x} dx = \left[-e^{-\lambda \cdot x} \left(x - \frac{1}{\lambda} \right) \right]_0^{\infty} = \frac{1}{\lambda} = \mu \quad (\text{brugt Schaum 14.510})$$

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 = \int_0^{\infty} \lambda \cdot x^2 \cdot e^{-\lambda \cdot x} dx - \mu^2 \\ &= \left[-e^{-\lambda \cdot x} \left(x^2 - \frac{2x}{\lambda} + \frac{2}{\lambda^2} \right) \right]_0^{\infty} - \mu^2 = \frac{2}{\lambda^2} - \mu^2 = \mu^2. \end{aligned}$$



Som det fremgår af beviset for sætning 8.2, er fordelingsfunktionen for en eksponentialfordelt variabel bestemt ved udtrykket

$$F(x) = P(X \leq x) = \begin{cases} 1 - e^{-\frac{x}{\mu}} & \text{for } x > 0 \\ 0 & \text{ellers} \end{cases}$$

På nedenstående graf er afbildet tæthedsfunktionen for eksponentialfordelingerne $\exp(1.0)$ og $\exp(2.0)$

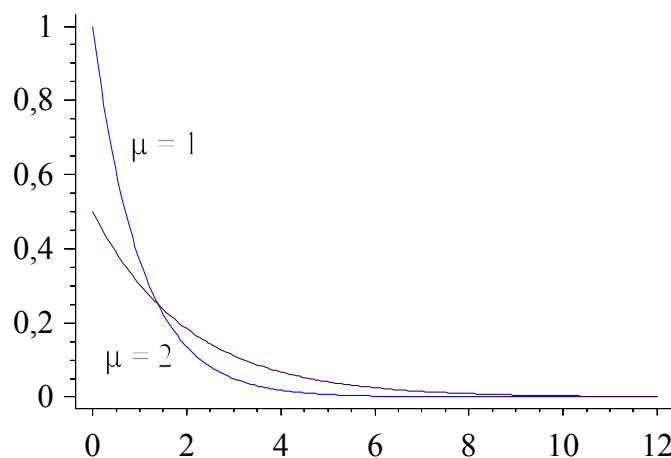


Fig 8.1 Eksponentialfordelingerne $\exp(1)$ og $\exp(2)$

Eksempel 8.2. Afstanden mellem successive revner i kabel.

Vi betragter det i eksempel 5.1 omtalte problem, hvor man fandt, at antallet N af mikroskopiske revner i et kobberkabel er Poissonfordelt. Der var i gennemsnit 12.3 af den type revner pr. 10 meter. Lad X være afstanden mellem to på hinanden følgende revner.

Beregn sandsynligheden for, at der er mere end 1 meter mellem to revner.

LØSNING:

Da der i gennemsnit er 1.23 revner pr. meter, må der i gennemsnit være $\frac{1}{1.23} = 0.812$ meter mellem to revner. Vi har derfor at X er eksponentialfordelt med $\mu = 0.813$.

$$P(X > 1) = 1 - P(X \leq 1) = 1 - \left(1 - e^{-\frac{1}{0.813}}\right) = e^{-1.23} = \underline{\underline{0.2923}}$$



Levetider. I apparater, som består af elektroniske komponenter (eksempelvis lommeregner), er der et meget ringe mekanisk slid. Apparatets fremtidige levetid vil derfor (næsten ikke) afhænge af, hvor længe det har fungeret indtil nu. I sådanne tilfælde vil eksponentialfordelingen erfaringsmæssigt være en god approksimativ model for apparatets levetid. Det kan nemlig vises, at eksponentialfordelingen er den eneste kontinuerte fordeling, som har ovennævnte egenskab (er uden hukommelse)

Bevis: Lad X være eksponentialfordelt med middelværdi μ og lad $b > a > 0$ være vilkårlige konstanter. Der gælder da:

$$P(X > a + b | X > a) = \frac{P((X > a + b) \wedge (X > a))}{P(X > a)} = \frac{P(X > a + b)}{P(X > a)} = \frac{e^{-\frac{a+b}{\mu}}}{e^{-\frac{a}{\mu}}} = e^{-\frac{b}{\mu}} = P(X > b)$$



Eksempel 8.3. Levetid for elektriske pærer.

Man har erfaring for, at en bestemt type elektriske pærer har en "brændtid" T (målt i timer), som approksimativt er eksponentialfordelt. På basis af et stort antal målinger ved man, at middellevetiden er $\mu = 1500$ timer.

- 1) Hvor stor er sandsynligheden for, at en tilfældig pære brænder over, inden den har været tændt i 1200 timer?
- 2) Find sandsynligheden for, at en tilfældig pære brænder i mere end 1800 timer.
- 3) En pære har brændt i 800 timer. Hvad er sandsynligheden for, at den brænder i mindst 1800 timer mere.

LØSNING:

1) $P(T < 1200) = F(1200) = 1 - e^{-\frac{1200}{1500}} = 1 - 0.449 = \underline{\underline{55.1\%}}$.

2) $P(T > 1800) = 1 - F(1800) = e^{-\frac{1800}{1500}} = \underline{\underline{30.12\%}}$

- 3) Da eksponentialfordelingen ingen hukommelse har, vil svaret blive som i spørgsmål 2, dvs. 30.12%.



8.4 WEIBULLFORDELINGEN

Hvis komponenterne i et elektronisk apparat ikke “slides”, dvs. den fremtidige levetid ikke afhænger af den foregående tid, er som nævnt i afsnit 8.3 eksponentialfordelingen velegnet som model for apparatets levetid. Hvis derimod de pågældende komponenters eventuelle svigten afhænger af den forløbne tid, kan man ofte med fordel benytte den i det følgende nævnte **Weibullfordeling** som approksimativ model for apparatets levetid (model for apparatets pålidelighed).

DEFINITION af Weibullfordeling. Lad k og μ være positive tal. Sandsynlighedsfordelingen for en kontinuert stokastisk variabel X med tæthedsfunktionen $f(x)$ bestemt ved

$$f(x) = \begin{cases} \frac{k}{\mu^k} \cdot x^{k-1} \cdot e^{-\left(\frac{x}{\mu}\right)^k} & \text{for } x > 0 \\ 0 & \text{ellers} \end{cases}$$

siges at være Weibullfordelingen $wei(k, \mu)$.

Det kan vises, at Weibullfordelingen $wei(k, \mu)$ har middelværdien $E(X) = \mu \cdot \Gamma\left(\frac{k+1}{k}\right)$ ¹⁾

og spredningen $\sigma(X) = \mu \cdot \sqrt{\Gamma\left(\frac{k+2}{k}\right) - \left(\Gamma\left(\frac{k+1}{k}\right)\right)^2}$

Det ses, at Weibullfordelingen kan opfattes som en generalisation af eksponentialfordelingen, idet $wei(1, \mu) = \exp(\mu)$.

Såfremt levetiderne for komponenter i et apparat aftager jo længere tid apparatet har været i funktion (på grund af slid), kan man benytte en Weibullfordeling med $k > 1$ som approksimativ model for apparatets levetid.

¹⁾ Gammafunktionen $\Gamma(x)$ er defineret i “Supplement til statistiske grundbegreber” 3A

8.5 DEN LOGARITMISKE NORMALFORDELING

Indenfor det biokemiske eller biologiske område (forsøgsdyrs reaktionstid, cellevækst m.v.) er den stokastiske variabel X ikke normalfordelt, men hvis man foretager en logaritmisk transformation $Y = \ln X$ er Y (approksimativt) normalfordelt.

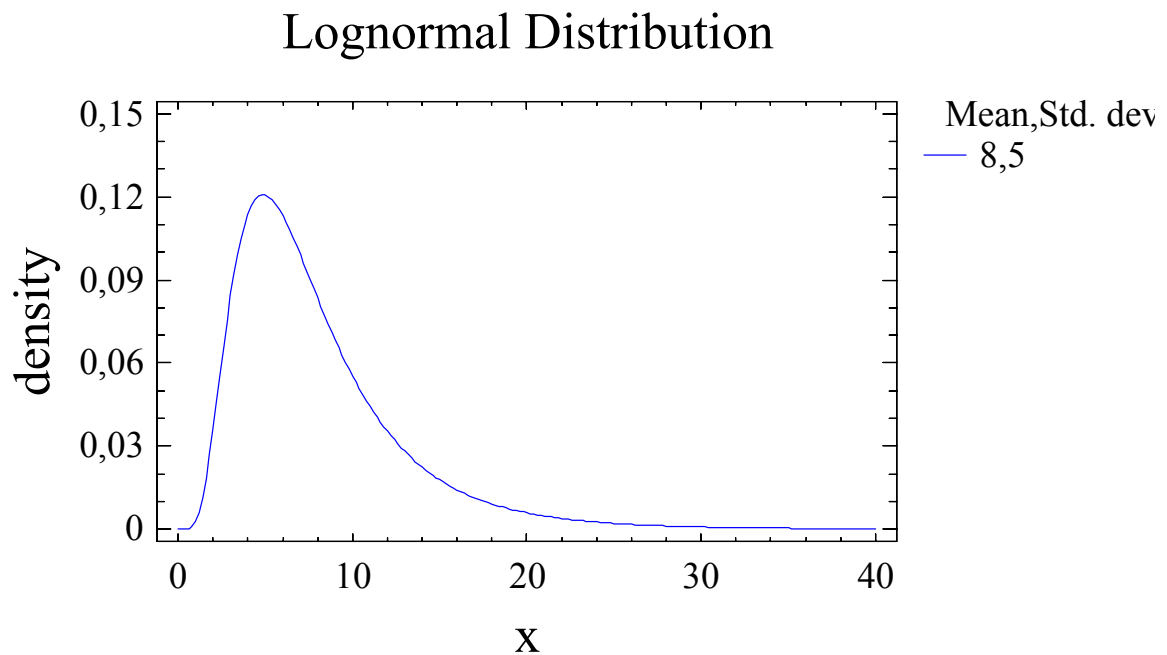
Man siger så, at X er **logaritmisk normalfordelt**.

Tæthedsfunktionen for X er bestemt ved $f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma x} e^{-\frac{1}{2} \left(\frac{\ln x - \mu}{\sigma} \right)^2}$ for $x > 0$.

Det kan vises, at mens $Y = \ln X$ har middelværdi μ og spredning σ har X middelværdi

$$E(X) = e^{\mu} \cdot e^{\frac{1}{2} \sigma^2} \text{ og } V(X) = e^{2\mu} \cdot e^{\sigma^2} \cdot (e^{\sigma^2} - 1).$$

Nedenfor er tegnet en logaritmisk normalfordeling med middelværdi 8 og spredning 5.



8.6 DEN 2-DIMENSIONALE NORMALFORDELING

Flerdimensionale fordelinger vil blive omtalt nærmere i kapitel 9. Her nævnes uden forklaring et eksempel herpå.

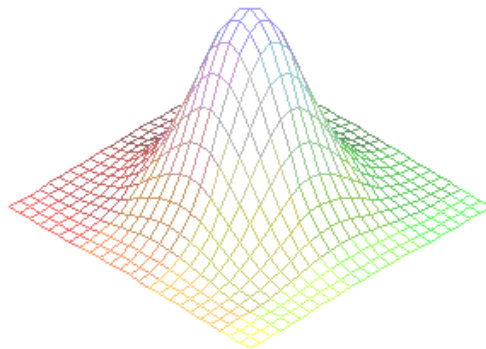
DEFINITION af 2-dimensional normalfordeling Lad μ_1, μ_2 være reelle tal og σ_1, σ_2 være positive tal. Sandsynlighedsfordelingen for 2-dimensional kontinuert stokastisk variabel (X_1, X_2) med tæthedsfunktion bestemt ved

$$f(x) = \frac{1}{2\pi \cdot \sigma_1 \cdot \sigma_2 \sqrt{1 - \rho^2}} \cdot e^{-\frac{1}{2(1-\rho^2)} \left(\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \frac{x_1 - \mu_1}{\sigma_1} \cdot \frac{x_2 - \mu_2}{\sigma_2} + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right)}$$

kaldes den 2-dimensionale **normalfordeling** med parametrene μ_1, μ_2, σ_1 og σ_2 .

Det kan vises, at $E(X_1) = \mu_1$, $E(X_2) = \mu_2$, $\sigma(X_1) = \sigma_1$, $\sigma(X_2) = \sigma_2$ og $\rho(X_1, X_2) = \rho$ (defineres i kapitel 9).

Grafen ses nedenfor.



OPGAVER

Opgave 8.1 (eksponentialfordeling)

På et betalingsnummer målt man i tidsrummet fra kl 20 til 22 tiden t (antal minutter) mellem på hinanden følgende telefonopkald. Følgende resultater fandtes:

Beliggenhed af t	$]0;1]$	$]1;2]$	$]2;3]$	$]3;4]$	$]4;5]$	$]5;6]$	$]6;7]$	$]7;8]$	$]8;9]$	$]9;10]$	$]10; \infty [$
Antal observationer.	36	21	16	13	7	9	6	1	2	6	0

Det antages, at antallet N af telefonopkald til nummeret er Poissonfordelt. Lad T være tiden mellem to opkald.

- 1) Angiv fordelingsfunktionen for T , og giv et estimat for middelværdien μ .

Vink: Antage, at for alle observationer i et interval er tidsrummet mellem observationerne intervallets midterværdis.

- 2) På baggrund af den i spørgsmål 1 fundne estimat for μ , ønskes bestemt $P(2 < T \leq 3)$.
- 3) Af tabellen ses, at i intervallet $]2; 3]$ forekommer i alt 16 observationer. Angiv hvor mange observationer man må forvente, ud fra resultatet i spørgsmål 2.

Opgave 8.2 (eksponentialfordeling)

Om en bestemt type elektriske komponenter vides, at deres levetider er eksponentialfordelte med en middellevetid på 800 timer.

- Find sandsynligheden for, at en komponent holder mindst 200 timer.
- Find sandsynligheden for, at en komponent holder mellem 600 og 800 timer.
- En komponent har holdt i 900 timer. Find sandsynligheden for, at den kan holde i mindst 200 timer mere.
- I et elektrisk system indgår netop én komponent af denne type. Hver gang komponenten svigter, udskiftes den øjeblikkeligt med en ny komponent af samme type. Find sandsynligheden for, at komponenten udskiftes 12 gange i løbet af 8000 timer.

Opgave 8.3 (eksponentialfordeling)

Antag, at levetiderne for en bestemt slags elektroniske komponenter er uafhængige og alle er eksponentialfordelt med en middellevetid på 3 (år). Betragt et delsystem bestående af 3 sådanne komponenter i seriekobling:



(en seriekobling ophører at fungere, når én af komponenterne ophører at fungere).

Bestem middellevetiden for et sådant system.

Opgave 8.4 (fordeling)

Nedbrydningstiden i den menneskelige organisme for et givet kvantum af et bestemt stof antages at være eksponentialfordelt med middelværdien 5 timer.

Ved et forsøg indsprøjtes stoffet samtidig i 10 patienter.

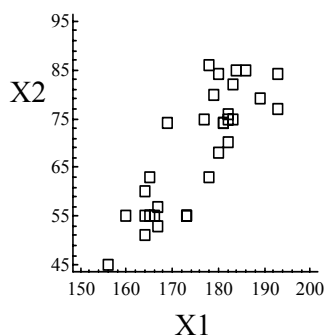
- 1) Beregn sandsynligheden (afrundet til et helt antal procent) for, at stoffet hos en tilfældig valgt patient vil være nedbrudt efter 8 timers forløb.
- 2) Beregn sandsynligheden for, at stoffet efter 8 timers forløb vil være nedbrudt hos mindst 5 af patienterne.
- 3) Efter hvor mange timers forløb vil der være ca. 90% sandsynlighed for, at stoffet er nedbrudt hos samtlige 10 patienter?
- 4) Hvor mange patienter skal indgå i en ny undersøgelse, hvis der skal være ca. 95% sandsynlighed for, at der er mindst en patient, hvis organisme efter 8 timers forløb endnu ikke har nedbrudt stoffet?

9 FLERDIMENSIONAL STOKASTISK VARIABEL

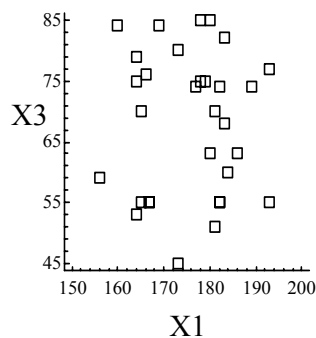
ESSENS

Kovariansen $V(X_i, X_j) = E((X_i - \mu_i) \cdot (X_j - \mu_j))$ er et mål for to variables tendens til at variere i takt med hinanden (samvarians). Kovariansen er f.eks. positiv(negativ), når afvigelsen $X_i - \mu_i$ har en tendens til at være positivt (negativt) proportional med afvigelsen $X_j - \mu_j$. Er X_i og X_j statistisk uafhængige, bliver kovariansen 0 (men man kan ikke slutte den anden vej).

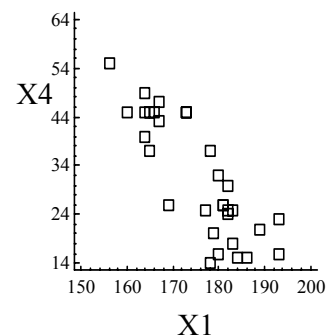
Korrelationskoefficienten $\rho(X_i, X_j) \equiv \frac{V(X_i, X_j)}{\sigma_i \cdot \sigma_j}$ er normeret, så $-1 \leq \rho(X_i, X_j) \leq 1$.



Stikprøve viser positiv
korrelation: $\rho(X_1, X_2) \approx 0.84$.



Stikprøve viser ingen
korrelation: $\rho(X_1, X_2) = 0.00$



Stikprøve viser negativ
korrelation: $\rho(X_1, X_2) \approx -0.84$.

Poolet estimat $s_{pool}^2 = \frac{f_1 s_1^2 + f_2 s_2^2 + \dots + f_k s_k^2}{f_1 + f_2 + \dots + f_k}$ med $f_{pool} = f_1 + f_2 + \dots + f_k$ frihedsgrader benyttes, når

man har k uafhængige estimater for den samme varians σ^2 :

$$s_1^2 = \frac{SAK_1}{f_1}, s_2^2 = \frac{SAK_2}{f_2}, \dots, s_k^2 = \frac{SAK_k}{f_k},$$

Har to stikprøver givet estimaterne $s_1^2 = 2.345, s_2^2 = 3.456$ med $f_1 = 6, f_2 = 4$ frihedsgrader,

bliver det poolede estimat $s_{pool}^2 = \frac{f_1 s_1^2 + f_2 s_2^2}{f_1 + f_2} = \frac{6 \cdot 2.345 + 4 \cdot 3.456}{6 + 4} = 2.7894 \approx \underline{2.789}$

med $f_{pool} = f_1 + f_2 = 6 + 4 = 10$ frihedsgrader.

Linearitetsreglen $E(a_0 + a_1 X_1 + a_2 X_2 + \dots + a_k X_k) = a_0 + a_1 E(X_1) + a_2 E(X_2) + \dots + a_k E(X_k),$

(a 'erne er konstanter).

Er $E(X_1) = 2, E(X_2) = 3$ fås

$$E(4 + 5X_1 + 6X_2) = 4 + 5E(X_1) + 6E(X_2) = 4 + 5 \cdot 2 + 6 \cdot 3 = \underline{32}.$$

Kvadratreglen $V(a_0 + a_1 X_1 + a_2 X_2 + \dots + a_k X_k)$

$$= a_1^2 V(X_1) + a_2^2 V(X_2) + \dots + a_k^2 V(X_k) + 2 \sum_{i=1}^k \sum_{j=i+1}^k a_i a_j V(X_i, X_j).$$

Er $V(X_1) = 2, V(X_2) = 3, V(X_1, X_2) = 1.5$, fås

$$\begin{aligned} V(4 + 5X_1 + 6X_2) &= 5^2 V(X_1) + 6^2 V(X_2) + 2 \cdot 5 \cdot 6 \cdot V(X_1, X_2) \\ &= 5^2 \cdot 2 + 6^2 \cdot 3 + 2 \cdot 5 \cdot 6 \cdot 1.5 = \underline{248} \end{aligned}$$

9.1 INDLEDNING

Ved anvendelserne optræder der ofte flere stokastiske variable $X_1, X_2, \dots, X_k$ ad gangen. Det kan da være naturligt at samle dem i et ordnet sæt $\vec{X} = (X_1, X_2, \dots, X_k)$, som kaldes en **k -dimensional stokastisk variabel**. Eksempelvis:

- * Et levnedsmiddel kan af en tilfældig udtaget forbruger bedømmes ved en karakter X_1 for smagen og en karakter X_2 for lugten. Så er $\vec{X} = (X_1, X_2) = (\text{Smag}, \text{Lugt})$ en 2-dimensional stokastisk variabel.
- * Et tilfældigt eksperiment går ud på at udtage en tilfældig person og måle vedkommendes højde X_1 og masse X_2 . Så er $\vec{X} = (X_1, X_2) = (\text{Højde}, \text{Masse})$ en 2-dimensional stokastisk variabel.
- * Ugens 7 lottotal udgør en 7-dimensional stokastisk variabel $\vec{X} = (X_1, X_2, \dots, X_7)$.
- * Et tilfældigt eksperiment går ud på at kaste en rød og en hvid terning. Så er $\vec{X} = (X_1, X_2) = (\text{Antal øjne op på rød terning}, \text{Antal øjne op på hvid terning})$ en 2-dimensional stokastisk variabel.

For hver af de 1-dimensionale stokastiske variable $X_1, X_2, \dots, X_k$ har vi tidligere defineret:

* **Fordelingsfunktioner** $F_1, F_2, \dots, F_k$:

$$F_1(X_1) \equiv P(X_1 \leq x_1), F_2(X_2) \equiv P(X_2 \leq x_2), \dots, F_k(X_k) \equiv P(X_k \leq x_k).$$

* **Tæthedsfunktioner** $f_1, f_2, \dots, f_k$, når $X_1, X_2, \dots, X_k$ er diskrete variable:

$$f_1(x_1) \equiv P(X_1 = x_1), f_2(x_2) \equiv P(X_2 = x_2), \dots, f_k(x_k) \equiv P(X_k = x_k),$$

og når de er kontinuerte variable:

$$f_1(x_1) \equiv \frac{dF_1(x_1)}{dx_1}, f_2(x_2) \equiv \frac{dF_2(x_2)}{dx_2}, \dots, f_k(x_k) \equiv \frac{dF_k(x_k)}{dx_k}.$$

* **Middelværdier**, når $X_1, X_2, \dots, X_k$ er diskrete variable:

$$E(g(X_1)) \equiv \sum_{x_1} g(x_1) \cdot f_1(x_1), \dots, E(g(X_k)) \equiv \sum_{x_k} g(x_k) \cdot f_k(x_k),$$

og når de er kontinuerte variable:

$$E(g(X_1)) \equiv \int_{-\infty}^{\infty} g(x_1) \cdot f_1(x_1) dx_1, \dots, E(g(X_k)) \equiv \int_{-\infty}^{\infty} g(x_k) \cdot f_k(x_k) dx_k,$$

specielt

$$\mu_1 \equiv E(X_1) \equiv \sum_{x_1} x_1 \cdot f_1(x_1), \dots, \mu_k \equiv E(X_k) \equiv \sum_{x_k} x_k \cdot f_k(x_k),$$

og

$$\mu_1 \equiv E(X_1) \equiv \int_{-\infty}^{\infty} x_1 \cdot f_1(x_1) dx_1, \dots, \mu_k \equiv E(X_k) \equiv \int_{-\infty}^{\infty} x_k \cdot f_k(x_k) dx_k.$$

Af definitionen på middelværdi følger linearitetsreglen:

$$E(a \cdot g(X_i) + b \cdot h(X_i)) = a \cdot E(g(X_i)) + b \cdot E(h(X_i)).$$

For en k -dimensional stokastisk variabel $\vec{X} = (X_1, X_2, \dots, X_k)$ definerer vi analogt:

* **Fordelingsfunktionen** F : ($\wedge$ betyder "både og")

$$F(x_1, x_2, \dots, x_k) \equiv P(X_1 \leq x_1) \wedge P(X_2 \leq x_2) \wedge \dots \wedge P(X_k \leq x_k).$$

* **Tæthedsfunktionen** f , når $X_1, X_2, \dots, X_k$ er diskrete variable:

$$f(x_1, x_2, \dots, x_k) \equiv P(X_1 = x_1) \wedge P(X_2 = x_2) \wedge \dots \wedge P(X_k = x_k)$$

og når de er kontinuerte variable: $f(x_1, x_2, \dots, x_k) \equiv \frac{\partial^k F(x_1, x_2, \dots, x_k)}{\partial x_1 \partial x_2 \dots \partial x_k}.$

* **Middelværdier**, når $X_1, X_2, \dots, X_k$ er diskrete variable:

$$E(g(X_1, X_2, \dots, X_k)) = \sum_{x_1} \sum_{x_2} \dots \sum_{x_k} g(x_1, x_2, \dots, x_k) \cdot f(x_1, x_2, \dots, x_k)$$

og når de er kontinuerte variable:

$$E(g(X_1, X_2, \dots, X_k)) = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \dots \int_{-\infty}^{\infty} g(x_1, x_2, \dots, x_k) \cdot f(x_1, x_2, \dots, x_k) dx_k$$

Af definitionen på middelværdi følger **linearitetsreglen**: $E(a \cdot g(\vec{X}) + b \cdot h(\vec{X})) = a \cdot E(g(\vec{X})) + b \cdot E(h(\vec{X}))$

De variable $X_1, X_2, \dots, X_k$ kaldes **stokastisk uafhængige**, såfremt de for alle værdier af $x_1, x_2, \dots, x_k$ opfylder betingelsen:

$$f(x_1, x_2, \dots, x_k) = f_1(x_1) \cdot f_2(x_2) \cdot \dots \cdot f_k(x_k),$$

der kan vises at være ækvivalent med betingelsen: $F(x_1, x_2, \dots, x_k) = F_1(x_1) \cdot F_2(x_2) \cdot \dots \cdot F_k(x_k).$

En **stikprøve** af størrelsen n på en stokastisk variabel $\vec{X} = (X_1, X_2, \dots, X_k)$ defineres som

$$(\vec{X}_1, \vec{X}_2, \dots, \vec{X}_n) = ((X_{11}, X_{21}, \dots, X_{k1}), (X_{12}, X_{22}, \dots, X_{k2}), \dots, (X_{1n}, X_{2n}, \dots, X_{kn}))$$

hvor $\vec{X}_1, \vec{X}_2, \dots, \vec{X}_n$ er statistisk uafhængige variable, der hver har samme fordeling som $\vec{X}$.

Eksempel 9.1. 2-dimensional stokastisk variabel.

Et levnedsmiddel kan af en tilfældig forbruger bedømmes ved en karakter X_1 for smagen og en karakter X_2 for lugten. Karakteren X_1 kan antage værdierne 0, 1 og 2, mens X_2 kun kan antage værdierne 0 og 2.

a) Antag, at man teoretisk kender tæthedsfunktionen $f(x_1, x_2)$:

$f(x_1, x_2)$		x_1		
		0	1	2
x_2	0	0.2	0.1	0.1
	2	0.1	0.2	0.3

a1) Find de 1-dimensionale tæthedsfunktioner $f_1(x_1)$ og $f_2(x_2)$.

a2) Er X_1 og X_2 statistisk uafhængige?

a3) Find middelværdierne $\mu_1 = E(X_1)$ og $\mu_2 = E(X_2)$ samt spredningerne $\sigma_1 = \sigma(X_1)$ og $\sigma_2 = \sigma(X_2)$.

a4) Find middelværdien $E(X_1, X_2)$.

b) Antag, at man i stedet kender en stikprøve på (X_1, X_2) :

(1,2), (0,0), (2,2), (2,2), (1,0), (2,2), (0,2), (2,2), (0,2), (2,2).

b1) Benyt stikprøven til at finde estimater for størrelserne i spørgsmål a3).

LØSNING:

a1) Ved summation ned gennem de lodrette søjler i tabellen for tæthedsfunktionen

$f(x_1, x_2) \equiv P(X_1 = x_1 \wedge X_2 = x_2)$ fås den 1-dimensionale tæthedsfunktion $f_1(x_1) = P(X_1 = x_1)$:

$$f_1(0) = 0.2 + 0.1 = \underline{0.3}, \quad f_1(1) = 0.1 + 0.2 = \underline{0.3}, \quad f_1(2) = 0.1 + 0.3 = \underline{0.4}.$$

Ved summation hen gennem de vandrette rækker i tabellen for tæthedsfunktionen $f(x_1, x_2)$ fås analogt den 1-dimensionale tæthedsfunktion $f_2(x_2)$:

$$f_2(0) = 0.2 + 0.1 + 0.1 = \underline{0.4}, \quad f_2(2) = 0.1 + 0.2 + 0.3 = \underline{0.6}.$$

a2) De variable X_1 og X_2 er statistisk uafhængige, hvis og kun hvis $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$ for alle værdier af (x_1, x_2) i definitionsområdet. Men da f.eks. $f_1(0) \cdot f_2(0) = 0.3 \cdot 0.4 = 0.12$ er forskellig fra $f(0,0) = 0.2$, er X_1 og X_2 ikke statistisk uafhængige.

a3) Vi finder

$$\mu_1 = E(X_1) = \sum x_1 \cdot f_1(x_1) = 0 \cdot f_1(0) + 1 \cdot f_1(1) + 2 \cdot f_1(2) = 0 \cdot 0.3 + 1 \cdot 0.3 + 2 \cdot 0.4 = \underline{1.1}$$

$$\mu_2 = E(X_2) = \sum x_2 \cdot f_2(x_2) = 0 \cdot f_2(0) + 2 \cdot f_2(2) = 0 \cdot 0.4 + 2 \cdot 0.6 = \underline{1.2}$$

$$\sigma_1 = \sigma(X_1) \equiv \sqrt{V(X_1)} \equiv \sqrt{E((X_1 - \mu_1)^2)} = \sqrt{\sum (x_1 - \mu_1)^2 \cdot f_1(x_1)}$$

$$= \sqrt{(0 - 1.1)^2 \cdot 0.3 + (1 - 1.1)^2 \cdot 0.3 + (2 - 1.1)^2 \cdot 0.4} = \sqrt{0.69} = \underline{0.83067}$$

$$\sigma_2 = \sigma(X_2) \equiv \sqrt{V(X_2)} \equiv \sqrt{E((X_2 - \mu_2)^2)} = \sqrt{\sum (x_2 - \mu_2)^2 \cdot f_2(x_2)}$$

$$= \sqrt{(0-1.2)^2 \cdot 0.4 + (2-1.2)^2 \cdot 0.6} = \sqrt{0.96} = \underline{\underline{0.97980}}$$

a4) Vi finder

$$\begin{aligned} E(X_1 \cdot X_2) &\equiv \sum_{x_1} \sum_{x_2} x_1 \cdot x_2 \cdot f(x_1, x_2) = 0 \cdot 0 \cdot f(0,0) + 1 \cdot 0 \cdot f(1,0) + \\ &\quad 2 \cdot 0 \cdot f(2,0) + 0 \cdot 2 \cdot f(0,2) + 1 \cdot 2 \cdot f(1,2) + 2 \cdot 2 \cdot f(2,2) \\ &= 0 \cdot 0.2 + 0 \cdot 0.1 + 0 \cdot 0.1 + 0 \cdot 0.1 + 2 \cdot 0.2 + 4 \cdot 0.3 = \underline{\underline{1.6}} . \end{aligned}$$

b1) Stikprøvens x_1 -værdier 1, 0, 2, 2, 1, 2, 0, 2, 0, 2 kan indtastes på en lommeregner, der finder gennemsnittet $\bar{x}_1$ og standardafvigelsen S_1 som tilnærmelser til middelværdi μ_1 og spredning σ_1 for X_1 . Man finder:

$$\mu_1 \equiv E(X_1) \approx \bar{x}_1 = \underline{\underline{1.2}}, \quad \sigma_1 = \sigma(X_1) \approx s_1 = \underline{\underline{0.9189}} .$$

Analogt indtastes stikprøvens x_2 -værdier 2, 0, 2, 2, 0, 2, 2, 2, 2, 2, og man finder

$$\mu_2 \equiv E(X_2) \approx \bar{x}_2 = \underline{\underline{1.6}}, \quad \sigma_2 = \sigma(X_2) \approx s_2 = \underline{\underline{0.8433}}$$

Det ses, at estimaterne har en vis lighed med de eksakte værdier i spørgsmål a3).



9.2 KOVARIANS OG KORRELATIONSKOEFFICIENT

Vi har omtalt, at hver stokastisk variabel har en varians. Men et par variable X_1 og X_2 kan have en tendens til at variere i overensstemmelse med hinanden (samvarians), således at afvigelse $X_1 - \mu_1$ og $X_2 - \mu_2$ overvejende har samme fortegn (positiv korrelation) eller overvejende har modsat fortegn (negativ korrelation). Eksempelvis kan en høj forekomst af ét vitamin i et levnedsmiddel ofte være ledsaget af en høj forekomst af et andet vitamin (positiv korrelation). Og studerendes højde og masse kan også have en positiv korrelation.

Vi betragter igen en k -dimensional stokastisk variabel $\vec{X} = (X_1, X_2, \dots, X_k)$. For et par af variable X_i og X_j defineres **kovariansen** ("samvariansen")

$$V(X_i, X_j) \equiv E((X_i - \mu_i) \cdot (X_j - \mu_j))$$

(den giver jo et vist mål for, om afvigelserne $X_i - \mu_i$ og $X_j - \mu_j$ i middel har samme fortegn eller modsat fortegn). Sættes $i = j$, fås $V(X_i, X_i) = E((X_i - \mu_i)^2)$, som er identisk med variansen $V(X_i)$ for variabelen X_i .

Man kan vise (se nedenfor), at $V(X_i, X_j) = E(X_i \cdot X_j) - \mu_i \cdot \mu_j$, som for $i = j$ giver $V(X_i) = E(X_i^2) - \mu_i^2$.

Bevis:

$$V(X_i, X_j) \equiv E((X_i - \mu_i) \cdot (X_j - \mu_j)) = E(X_i X_j - \mu_i X_j - X_i \mu_j + \mu_i \mu_j).$$

Anvendes linearitetsreglen kan sidste led omformes:

$$= E(X_i X_j) - \mu_i E(X_j) - E(X_i) \mu_j + \mu_i \mu_j = E(X_i X_j) - \mu_i \mu_j - \mu_i \mu_j + \mu_i \mu_j = E(X_i X_j) - \mu_i \mu_j$$



For bedre at kunne vurdere hvor meget de variable varierer i "takt" med hinanden, divideres kovariansen med spredningerne, så man får den såkaldte **korrelationskoefficient**:

$$\rho(X_i, X_j) \equiv \frac{V(X_i, X_j)}{\sigma_1 \cdot \sigma_2}$$

Man kan vise (se nedenfor), at $-1 \leq \rho(X_i, X_j) \leq 1$.

Bevis: Vi har

$$\begin{aligned} 0 &\leq E\left(\left(\lambda \cdot (X_i - \mu_i) + (X_j - \mu_j)\right)^2\right) = E\left(\lambda^2 \cdot (X_i - \mu_i)^2 + 2\lambda(X_i - \mu_i)(X_j - \mu_j) + (X_j - \mu_j)^2\right) \\ &= \lambda^2 \cdot E((X_i - \mu_i)^2) + 2\lambda E((X_i - \mu_i)(X_j - \mu_j)) + E((X_j - \mu_j)^2) = \lambda^2 \cdot V(X_i) + 2V(X_i, X_j)\lambda + V(X_j) \end{aligned}$$

Da dette andengradspolynomium i λ aldrig er negativt, kan diskriminanten ikke være positiv, dvs.

$$4(V(X_i, X_j))^2 - 4 \cdot V(X_i) \cdot V(X_j) \leq 0 \Leftrightarrow \frac{(V(X_i, X_j))^2}{V(X_i) \cdot V(X_j)} \leq 1 \Leftrightarrow (\rho(X_i, X_j))^2 \leq 1 \Leftrightarrow -1 \leq \rho(X_i, X_j) \leq 1.$$



Man kan (som det ses nedenfor) vise, at

$$X_1 \text{ og } X_2 \text{ stat. uafhængige} \Rightarrow E(X_i \cdot X_j) = E(X_i) \cdot E(X_j) \Leftrightarrow V(X_i, X_j) = 0 \Leftrightarrow \rho(X_i, X_j) = 0$$

Bevis: Vi har

$$\begin{aligned} X_1 \text{ og } X_2 \text{ stat. uafhængige} &\Leftrightarrow f(x_i, x_j) = f_i(x_i) \cdot f_j(x_j) \\ \Rightarrow E(X_i \cdot X_j) &= \sum_{x_i} \sum_{x_j} x_i \cdot x_j \cdot f(x_i, x_j) = \sum_{x_i} \sum_{x_j} x_i \cdot x_j \cdot f_i(x_i) \cdot f_j(x_j) \\ &= \sum_{x_i} x_i \cdot f_i(x_i) \sum_{x_j} x_j \cdot f_j(x_j) = E(X_i) \cdot E(X_j) \\ \Leftrightarrow V(X_i, X_j) &= E(X_i \cdot X_j) - \mu_i \cdot \mu_j = E(X_i) \cdot E(X_j) - \mu_i \cdot \mu_j = 0 \Leftrightarrow \rho(X_i, X_j) = \frac{V(X_i, X_j)}{\sigma_i \cdot \sigma_j} = 0. \end{aligned}$$



Estimer for kovarians, varians og korrelationskoefficient

Ud fra en stikprøve $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ kan man beregne

$$SAP_{XY} \equiv \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}), \quad SAK_X \equiv \sum_{i=1}^n (x_i - \bar{x})^2, \quad SAK_Y \equiv \sum_{i=1}^n (y_i - \bar{y})^2$$

(SAP = “Sum af Afvigelsers Produkter”, SAK = “Sum af Afvigelsers Kvadrater”)

og heraf danne estimer for kovarians, varianser og korrelationskoefficient:

$$\text{kovarians: } V(X, Y) \approx \frac{SAP_{XY}}{n-1} \text{ og varianser: } V(X) \approx \frac{SAK_X}{n-1}, \quad V(Y) \approx \frac{SAK_Y}{n-1}$$

$$\text{korrelationskoefficient: } \rho(X, Y) \approx r \equiv \frac{SAP_{XY}}{\sqrt{SAK_X \cdot SAK_Y}}$$

Det kan således vises (for enhver fordelingstype), at

$$E\left(\frac{SAP_{XY}}{n-1}\right) = V(X, Y), \quad E\left(\frac{SAK_X}{n-1}\right) = V(X), \quad E\left(\frac{SAK_Y}{n-1}\right) = V(Y)$$

Bevis: Vi har

$$\begin{aligned} SAP_{XY} &= \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^n ((X_i - \mu_X) - (\bar{X} - \mu_X)) \cdot ((Y_i - \mu_Y) - (\bar{Y} - \mu_Y)) \\ &= \sum_{i=1}^n (X_i - \mu_X)(Y_i - \mu_Y) + \sum_{i=1}^n (\bar{X} - \mu_X)(\bar{Y} - \mu_Y) - \sum_{i=1}^n (X_i - \mu_X)(\bar{Y} - \mu_Y) - \sum_{i=1}^n (\bar{X} - \mu_X)(Y_i - \mu_Y) \\ &= \sum_{i=1}^n (X_i - \mu_X)(Y_i - \mu_Y) + n \cdot (\bar{X} - \mu_X)(\bar{Y} - \mu_Y) - (n\bar{X} - n \cdot \mu_X)(\bar{Y} - \mu_Y) - (\bar{X} - \mu_X)(n\bar{Y} - n \cdot \mu_Y) \\ &= \sum_{i=1}^n (X_i - \mu_X)(Y_i - \mu_Y) - n \cdot (\bar{X} - \mu_X)(\bar{Y} - \mu_Y). \end{aligned}$$

Altså fås ved hjælp af linearitetsreglen:

$$\begin{aligned} E(SAP_{XY}) &= \sum_{i=1}^n E((X_i - \mu_X)(Y_i - \mu_Y)) - n \cdot E((\bar{X} - \mu_X)(\bar{Y} - \mu_Y)) \\ &= \sum_{i=1}^n V(X_i, Y_i) - \frac{1}{n} \cdot E((X_1 + X_2 + \dots + X_n - n \cdot \mu_X)(Y_1 + Y_2 + \dots + Y_n - n \cdot \mu_Y)) \\ &= \sum_{i=1}^n V(X, Y) - \frac{1}{n} \cdot E\left(\sum_{i=1}^n (X_i - \mu_X) \sum_{j=1}^n (Y_j - \mu_Y)\right) \end{aligned}$$

$$\begin{aligned}
&= n \cdot V(X, Y) - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n E((X_i - \mu_x)(Y_j - \mu_y)) = n \cdot V(X, Y) - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n V(X_i, Y_j) \\
&= n \cdot V(X, Y) - \frac{1}{n} \sum_{i=1}^n V(X_i, Y_i) \quad (\text{idet } V(X_i, Y_j) = 0 \text{ for } i \neq j \text{ i en stikprøve}) \\
&= n \cdot V(X, Y) - \frac{1}{n} \cdot n \cdot V(X, Y) = (n-1) \cdot V(X, Y), \quad \text{dvs. } E\left(\frac{SAP_{XY}}{n-1}\right) = V(X, Y).
\end{aligned}$$

Erstattes Y med X i beviset, bliver SAP_{XY} erstattet med SAK_X , og $V(X, Y)$ bliver erstattet med $V(X)$, hvorved vi også får bevist, at $E\left(\frac{SAK_X}{n-1}\right) = V(X)$. Erstattes X med Y , fås endelig $E\left(\frac{SAK_Y}{n-1}\right) = V(Y)$



Poolet estimat

Som nævnt er SAK en forkortelse for "Sum af Afvigelsers Kvadrater". De afvigelser der tænkes på er de n differenser $X_1 - \bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X}$. De har summen 0, så når $n-1$ af dem er kendt, er den sidste fastlagt. Da SAK således kun er baseret på $n-1$ uafhængige differenser, siger man, at SAK har $f = n-1$ **frihedsgrader**. Det er også antallet af frihedsgrader der optræder i estimatet for varians: $s^2 = \frac{SAK_X}{f}$

Ofte har man taget k stikprøver på variable med samme varians σ^2 , så vi får k uafhængige estimater for den samme varians σ^2 :

$$s_1^2 = \frac{SAK_1}{f_1}, \quad s_2^2 = \frac{SAK_2}{f_2}, \dots, s_k^2 = \frac{SAK_k}{f_k}$$

og det er da fordelagtigt at forene dem i et såkaldt fællesestimat eller **poolet estimat**:

$$s_{pool}^2 = \frac{f_1 s_1^2 + f_2 s_2^2 + \dots + f_k s_k^2}{f_1 + f_2 + \dots + f_k} \quad \text{med} \quad f_{pool} = f_1 + f_2 + \dots + f_k \quad \text{frihedsgrader}.$$

Dette kan også skrives

$$s_{pool}^2 = \frac{SAK_1 + SAK_2 + \dots + SAK_k}{f_1 + f_2 + \dots + f_k}, \quad \text{med} \quad f_{pool} = f_1 + f_2 + \dots + f_k \quad \text{frihedsgrader}.$$

Det ses, at s_{pool}^2 har den rigtige middelværdi σ^2 , idet linearitetsreglen giver

$$E(s_{pool}^2) = \frac{f_1 E(s_1^2) + f_2 E(s_2^2) + \dots + f_k E(s_k^2)}{f_1 + f_2 + \dots + f_k} = \frac{f_1 \sigma^2 + f_2 \sigma^2 + \dots + f_k \sigma^2}{f_1 + f_2 + \dots + f_k} = \sigma^2.$$

Eksempel 9.2. Kovarians. Korrelationskoefficient.

Vi betragter igen den 2-dimensionale fordeling fra eksempel 9.1.

a5) Find kovariansen og korrelationskoefficienten.

b2) Benyt stikprøven til at finde estimater for kovariansen og korrelationskoefficienten.

LØSNING:

a5) Idet vi i eksempel 9.1 har fundet $\mu_1 = 1.1$, $\mu_2 = 1.2$, $\sigma = \sqrt{0.69}$ og $\sigma = \sqrt{0.96}$, finder vi nu kovariansen $V(X_1, X_2)$ og korrelationskoefficienten $\rho(X_1, X_2)$:

$$\begin{aligned}
V(X_i, X_j) &\equiv E((X_i - \mu_i) \cdot (X_j - \mu_j)) = \sum_{x_1} \sum_{x_2} (x_1 - \mu_1)(x_2 - \mu_2) f(x_1, x_2) \\
&= (0 - \mu_1) \cdot (0 - \mu_2) \cdot f(0,0) + (1 - \mu_1) \cdot (0 - \mu_2) \cdot f(1,0) + (2 - \mu_1) \cdot (0 - \mu_2) \cdot f(2,0)
\end{aligned}$$

$$\begin{aligned}
 &+ (0 - \mu_1) \cdot (2 - \mu_2) \cdot f(0,2) + (1 - \mu_1) \cdot (2 - \mu_2) \cdot f(1,2) + (2 - \mu_1) \cdot (2 - \mu_2) \cdot f(2,2) \\
 &= (0 - 1.1) \cdot (0 - 1.2) \cdot 0.2 + (1 - 1.1) \cdot (0 - 1.2) \cdot 0.1 + (2 - 1.1) \cdot (0 - 1.2) \cdot 0.1 \\
 &\quad + (0 - 1.1) \cdot (2 - 1.2) \cdot 0.1 + (1 - 1.1) \cdot (2 - 1.2) \cdot 0.2 + (2 - 1.1) \cdot (2 - 1.2) \cdot 0.3 = \underline{\underline{0.28}} \\
 \rho(X_1, X_2) &\equiv \frac{V(X_1, X_2)}{\sigma(X_1) \cdot \sigma(X_2)} = \frac{0.28}{\sqrt{0.69} \cdot \sqrt{0.96}} = \underline{\underline{0.3440}}.
 \end{aligned}$$

b2) Stikprøvens værdier (1,2), (0,0), (2,2), (2,2), (1,0), (2,2), (0,2), (2,2), (0,2), (2,2) kan indtastes på en lommeregner, der finder tilnærmelser (estimer) til kovariansen $V(X_1, X_2)$ og korrelationskoefficienten $\rho(X_1, X_2)$:

$$V(X_1, X_2) \approx \frac{SAP_{12}}{n-1} = \frac{\sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{n-1} = \underline{\underline{0.3111}}$$

$$\rho(X_1, X_2) \approx r \equiv \frac{SAP_{12}}{\sqrt{SAK_1 \cdot SAK_2}} = \frac{\sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sqrt{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 \cdot \sum_{j=1}^n (x_{2j} - \bar{x}_2)^2}} = \underline{\underline{0.4015}}.$$

Det ses, at estimerne har en vis lighed med de eksakte værdier i spørgsmål a5).



9.3 LINEARKOMBINATION

Når vi skal tage en stikprøve $(X_1, X_2, \dots, X_n)$ på en 1-dimensional stokastisk variabel X , så skal vi n gange skaffe et tal X fra et tilfældigt eksperiment. Derfor kan vi opfatte stikprøven som en n -dimensional stokastisk variabel $\vec{X} = (X_1, X_2, \dots, X_n)$.

Vi bruger ofte stikprøven til at danne gennemsnittet $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} X_1 + \frac{1}{n} X_2 + \dots + \frac{1}{n} X_n$

som er en speciel linearkombination af $X_1, X_2, \dots, X_n$.

Ved en **linearkombination** L for en k -dimensional stokastisk variabel $\vec{X} = (X_1, X_2, \dots, X_k)$ forstås et udtryk af formen $L = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_k X_k$, hvor $a_0, a_2, a_3, \dots, a_k$ er konstanter.

For middelværdien af L giver **linearitetsreglen**:

$$E(L) = a_0 + a_1 E(X_1) + a_2 E(X_2) + \dots + a_k E(X_k).$$

Eksempelvis kan vi se, at et gennemsnit $\bar{X}$ altid har den "rigtige" middelværdi μ :

$$E(\bar{X}) = \frac{1}{n} E(X_1) + \frac{1}{n} E(X_2) + \dots + \frac{1}{n} E(X_n) = \frac{1}{n} \mu + \frac{1}{n} \mu + \dots + \frac{1}{n} \mu = \mu.$$

For variansen af en linearkombination L gælder **kvadratreglen**:

$$V(L) = a_1^2 \cdot V(X_1) + a_2^2 \cdot V(X_2) + \dots + a_k^2 \cdot V(X_k) + 2 \sum_{i=1}^k \sum_{j=i+1}^k a_i a_j V(X_i, X_j)$$

Eksempelvis:

- a) $V(a + bX + cY) = b^2V(X) + c^2V(Y) + 2bcV(X, Y)$
 b) $V(a + bX + cY + dZ)$
 $= b^2V(X) + c^2V(Y) + d^2V(Z) + 2bcV(X, Y) + 2cdV(Y, Z) + 2dbV(Z, X)$.
 c) $V(\bar{X}) = \frac{1}{n^2}V(X_1) + \frac{1}{n^2}V(X_2) + \dots + \frac{1}{n^2}V(X_n)$ (X 'erne statistisk uafhængige)
 $= \frac{1}{n^2}\sigma^2 + \frac{1}{n^2}\sigma^2 + \dots + \frac{1}{n^2}\sigma^2 = \frac{\sigma^2}{n}$, dvs. $\sigma(\bar{X}) = \frac{\sigma(X)}{\sqrt{n}}$.

Den sidste ligning viser, at spredningen på et gennemsnit kun er omvendt proportional med kvadratroden på stikprøvestørrelsen n . For at få et gennemsnit med en 10 gange mindre spredning, skal stikprøven altså gøres 100 gange større!

Bevis for kvadratreglen. Vi finder

$$\begin{aligned} V(L) &\equiv E((L - E(L))^2) = E((a_0 + a_1X_1 + \dots + a_kX_k - a_0 - a_1\mu_1 - \dots - a_k\mu_k)^2) \\ &= E((a_1(X_1 - \mu_1) + \dots + a_k(X_k - \mu_k))^2) \\ &= E\left(a_1^2(X_1 - \mu_1)^2 + \dots + a_k^2(X_k - \mu_k)^2 + 2\sum_{i=1}^k \sum_{j=1}^k a_i a_j (X_i - \mu_i)(X_j - \mu_j)\right) \\ &= a_1^2 E((X_1 - \mu_1)^2) + \dots + a_k^2 E((X_k - \mu_k)^2) + 2\sum_{i=1}^k \sum_{j=1}^k a_i a_j E((X_i - \mu_i)(X_j - \mu_j)) \\ &= a_1^2 V(X_1) + \dots + a_k^2 V(X_k) + 2\sum_{i=1}^k \sum_{j=1}^k a_i a_j V(X_i, X_j). \end{aligned}$$



Eksempel 9.3. Linearkombination af stokastiske variable.

Et levnedsmiddel leveres i poser. Lad X_1 og X_2 [mg/kg] betegne koncentrationerne af to stoffer A og B i en tilfældig udvalgt pose. Det vides, at $E(X_1) = 20.0$, $E(X_2) = 30.0$, $\sigma(X_1) = 2.0$, $\sigma(X_2) = 4.0$ og $V(X_1, X_2) = -4.0$. Holdbarheden Y er teoretisk givet ved $Y = 5 + 4X_1 + 3X_2$ [dage].

Find holdbarhedens middelværdi $E(Y)$ og spredning $\sigma(Y)$.

LØSNING:

Vi finder

$$E(Y) = E(5 + 4X_1 + 3X_2) = 5 + 4 \cdot E(X_1) + 3 \cdot E(X_2) = 5 + 4 \cdot 20.0 + 3 \cdot 30.0 = \underline{175.0}$$

$$V(Y) = V(5 + 4X_1 + 3X_2) = 4^2 \cdot V(X_1) + 3^2 \cdot V(X_2) + 2 \cdot 4 \cdot 3 \cdot V(X_1, X_2)$$

$$= 4^2 \cdot 2.0^2 + 3^2 \cdot 4.0^2 + 2 \cdot 4 \cdot 3 \cdot (-4.0) = 112$$

$$\sigma(Y) \equiv \sqrt{V(Y)} = \sqrt{112} = \underline{10.5830}.$$



OPGAVER

Opgave 9.1.1 (2-dimensional stokastisk variabel)

Et spil i et casino går ud på at trække en tilfældig seddel fra en urne (og lægge sedlen tilbage igen). Urnen indeholder 10 sedler, og på hver seddel står 2 tal (X_1, X_2) :

(1,0)	(3,0)	(4,0)
(1,0)	(3,3)	(4,3)
(1,3)		(4,3)
(1,3)		
(1,3)		

- a1) Find den 2-dimensionale tæthedsfunktion $f(x_1, x_2)$:
 a2) Find de 1-dimensionale tæthedsfunktioner $f_1(x_1)$ og $f_2(x_2)$.
 a3) Er X_1 og X_2 statistisk uafhængige ?
 a4) Find middelværdierne $\mu_1 = E(X_1)$ og $\mu_2 = E(X_2)$ samt spredningerne $\sigma_1(X_1)$ og $\sigma_2(X_2)$.
 a5) Find middelværdien $E(X_1 \cdot X_2^2)$.
 b) Antag, at man i stedet kender en stikprøve på (X_1, X_2) :
 (1,3), (1,0), (1,0), (4,3), (3,0), (4,3), (1,0), (3,0), (3,3), (1,3).
 b1) Benyt stikprøven til at finde estimater for størrelserne i spørgsmål a4).

Opgave 9.1.2 (kovarians, korrelationskoefficient)

- a6) Find kovariansen $V(X_1, X_2)$ og korrelationskoefficienten $\rho(X_1, X_2)$.
 b2) Benyt stikprøven til at finde estimater for størrelserne i spørgsmål a6).

Opgave 9.1.3 (linearkombination)

For det i opgave 9.1.1 og 9.1.2 omtalte casino aftales et spil, hvor gevinsten er

$$G = 20 + 10X_1 + 5X_2.$$

- a7) Find gevinstens middelværdi $E(G)$ og spredning $\sigma(G)$.
 b3) Benyt stikprøven til at finde estimater for størrelserne i spørgsmål a7).

Opgave 9.2.1 (2-dimensional stokastisk variabel)

Under en produktion kan der optræde fejl. Lad (X_1, X_2) = (Antal gange der optræder fejl af type 1, Antal gange der optræder fejl af type 2) i en tilfældig produktion. Variablen X_1 kan antage værdierne 0, 1 og 2, mens X_2 kun kan antage værdierne 0 og 1.

- a) Antag, at man teoretisk kender tæthedsfunktionen $f(x_1, x_2)$:

$f(x_1, x_2)$		x_1		
		0	1	2
x_2	0	0.3	0.1	0.1
	1	0.1	0.2	0.2

- a1) Find de 1-dimensionale tæthedsfunktioner $f_1(x_1)$ og $f_2(x_2)$.
 a2) Er X_1 og X_2 statistisk uafhængige ?
 a3) Find middelværdierne $\mu_1 = E(X_1)$ og $\mu_2 = E(X_2)$ samt spredningerne $\sigma_1(X_1)$ og $\sigma_2(X_2)$.

a4) Find middelværdien $E(\sqrt{X_1 + X_2})$.

b) Antag, at man i stedet kender en stikprøve på (X_1, X_2) :

(0,1), (0,0), (1,1), (1,1), (0,0), (0,0), (0,1), (2,1), (0,0), (2,1).

b1) Benyt stikprøven til at finde estimater for størrelserne i spørgsmål a3).

Opgave 9.2.2 (kovarians, korrelationskoefficient)

Vi betragter igen produktionsprocessen fra opgave 9.2.1.

a5) Find kovariansen $V(X_1, X_2)$ og korrelationskoefficienten $\rho(X_1, X_2)$.

b2) Benyt stikprøven til at finde estimater for størrelserne i spørgsmål a5).

Opgave 9.2.3 (linearkombination)

For den i opgave 9.2.1 og 9.2.2 omtalte produktionsproces er fortjenesten

$$F = 20000 - 3000X_1 - 4000X_2.$$

a6) Find fortjenestens middelværdi $E(F)$ og spredning $\sigma(F)$.

b3) Benyt stikprøven til at finde estimater for størrelserne i spørgsmål a6).

Opgave 9.3.1 (2-dimensional stokastisk variabel)

År 4001. En sonde er vendt hjem med oplysninger om individer på en fremmed planet. De kan have 2, 4 eller 6 øjne, og 2 eller 4 ører. Lad $(X_1, X_2) = (\text{Antal øjne}, \text{Antal ører})$ for et tilfældigt udtaget individ på planeten.

a) Professor Cosmussen har teoretisk opstillet tæthedsfunktionen $f(x_1, x_2)$:

$f(x_1, x_2)$		x_1		
		2	4	6
x_2	2	0.1	0.2	0.1
	4	0.1	0.3	0.2

a1) Find de 1-dimensionale tæthedsfunktioner $f_1(x_1)$ og $f_2(x_2)$.

a2) Er X_1 og X_2 statistisk uafhængige?

a3) Find middelværdierne $\mu_1 = E(X_1)$ og $\mu_2 = E(X_2)$ samt spredningerne $\sigma_1(X_1)$ og $\sigma_2(X_2)$.

a4) Find middelværdien $E\left(\frac{1}{X_1} + \frac{1}{X_2}\right)$.

b) Antag, at man i stedet kender en stikprøve på (X_1, X_2) :

(6,2), (2,4), (6,4), (2,2), (6,4), (4,4), (2,2), (4,2), (4,4), (4,4).

b1) Benyt stikprøven til at finde estimater for størrelserne i spørgsmål a3).

Opgave 9.3.2 (kovarians, korrelationskoefficient)

Vi betragter igen individerne fra opgave 9.2.1.

a5) Find kovariansen $V(X_1, X_2)$ og korrelationskoefficienten $\rho(X_1, X_2)$.

b2) Benyt stikprøven til at finde estimater for størrelserne i spørgsmål a5).

Opgave 9.3.3 (linearkombination)

For de i opgave 9.3.1 og 9.3.2 omtalte individer har professor Cosmussen opstillet en formel for deres masse:

$$M = 200 + 20X_1 + 10X_2 \text{ kg.}$$

a6) Find massens middelværdi $E(M)$ og spredning $\sigma(M)$.

b3) Benyt stikprøven til at finde estimater for størrelserne i spørgsmål a6).

Opgave 9.4.1 (2-dimensional stokastisk variabel)

Lad $(X_1, X_2) = (\text{Højde [cm]}, \text{Masse [kg]})$ af en tilfældigt udtaget studerende på 3. halvår.

Der foreligger følgende stikprøve:

(178,63), (186,85), (180,68), (183,75), (164,55), (193,77),
(193,84), (160,55), (165,63), (180,84), (169,74), (189,79).

Find estimater for middelværdierne $\mu_1 = E(X_1)$ og $\mu_2 = E(X_2)$ samt spredningerne $\sigma_1(X_1)$ og $\sigma_2(X_2)$.

Opgave 9.4.2 (kovarians, korrelationskoefficient)

Benyt stikprøven i opgave 9.4.1 til at finde estimater for kovariansen $V(X_1, X_2)$ og korrelationskoefficienten $\rho(X_1, X_2)$.

Opgave 9.4.3 (linearkombination)

Vi betragter igen de i opgave 9.4.1 og 9.4.2 omtalte studerende. En frugtavlser har opstillet en formel for den timeløn, han vil give dem som frugtplukkere:

$$L = 100 + 0.3X_1 - 0.2X_2 \text{ kroner/time.}$$

Benyt stikprøven til at finde estimater for timelønnens middelværdi $E(L)$ og spredning $\sigma(L)$.

Opgave 9.5 (poolet estimat)

Koncentrationen af et stof A blev målt i 3 partier råvarer:

Råvare 1: 56, 60, 54, 49, 61

Råvare 2: 78, 73, 80

Råvare 3: 66, 62, 70, 72, 60

Det antages, at der er samme spredning i de 3 tilfælde. Find et estimat s_{pool} for spredningen σ .

Opgave 9.6 (poolet estimat)

Koncentrationen af et stof A blev målt i 2 levnedsmidler:

Levnedsmiddel 1: 87, 89, 94, 86, 89, 95 Levnedsmiddel 2: 93, 99, 94, 91, 98.

Det antages, at der er samme spredning i de 2 tilfælde. Find et estimat s_{pool} for spredningen σ .

Opgave 9.7 (poolet estimat)

Koncentrationen af et stof A blev målt i mælken fra 5 køer:

Ko 1: 44, 48, 46, 43, 45

Ko 2: 40, 38, 41

Ko 3: 43, 45, 42, 42

Ko 4: 36, 32

Ko 5: 50

Det antages, at der er samme spredning i de 5 tilfælde. Find et estimat s_{pool} for spredningen σ .

APPENDIX

Appendix 2. Statistiske beregninger udført ved hjælp af udvalgte lommeregnere og PC-programmer.

I dette appendix gennemgås hvorledes de i notatet mest almindelige statistiske beregninger kan udføres på

1) **TI-89:** En “stor” matematik/statistik-lommeregner

2) **TI-83:** En “billig” god “statistik-lommeregner

3) **Excel:** Et udbredt regneark

4) **Maple:** Et udbredt matematikprogram til PC-er.

Appendix 4.1. Oversigt over formler til beregning af konfidensintervaller

Appendix 5.1-5.6. Oversigt over de forskellige typer hypotesetest af 1 variabel.

Appendix 7.1. Oversigt over approksimationer af:

Hypergeometrisk-fordeling til binomial-fordeling

Binomial-fordeling til Poisson- eller normal-fordeling

Poisson-fordeling til normalfordeling.

TI 89

1) Generelt:

Metode 1: Vælg HOME\ CATALOG,, F3\ vælg den ønskede fordeling\ENTER

(benyt evt. ALPHA,+ for bogstav for hurtigt at komme til det ønskede navn).

Fordel: Hurtig ved beregning af sandsynligheder, såsom $P(X < 0.87)$ da resultatet straks indsættes på HOME-linien.

Ulempe: Man skal huske parametrene rækkefølge (de kan dog ses nederst på skærmen)

Metode 2: Vælg APPS\ Stats/List\ indtast eventuel data i eksempelvis "list1" \ vælg en relevant "F- knap".

Fordel: Der fremkommer nu en menu, som er næsten selvforklarende.

Ulempe: Skal resultatet ned på HOME-linien (man vil regne videre), bliver det lidt besværligt: HOME, Var-Link\I StatsVar mappen markeres den ønskede størrelse, ENTER

Oprette en "Folder": VAR-Link\ F1\ 5: Create Folder\ Skriv navn på folder.

Vælg en mappe som den aktuelle mappe: MODE\ Current Folder\navn

Formål: Det kan være praktisk ikke at gemme alle sine resultater i MAIN.

2) Sandsynlighedsfordelinger.

Normalfordeling $n(\mu, \sigma)$

a) Find $p = P(a \leq X \leq b)$, hvor a, b, μ, σ er givne konstanter. $p = \text{normcdf}(a, b, \mu, \sigma)$

Eksempel: $p = P(X \leq 11.6)$, hvor $\mu = 11.3, \sigma = 5$ $p = \text{normcdf}(-\infty, 11.6, 11.3, 5) = 0.524$

b) Find $x_p : P(X \leq x_p) = p$, hvor p, μ, σ er givne konstanter. $x_p = \text{invNorm}(p, \mu, \sigma)$

Eksempel: $P(X \leq x_p) = 0.7$, hvor $\mu = 11.4, \sigma = 6$ $x_p = \text{invNorm}(0.7, 11.4, 6) = 14.55$

t - fordeling.

Lad T være t - fordelt med frihedsgradstallet f .

a) Find $p = P(a \leq T \leq b)$, hvor a og b er givne konstanter. $t\text{Cdf}(a, b, f)$

Eksempel: $p = P(T \leq -1.3)$, med $f = 14$ $p = t\text{Cdf}(-\infty, -1.3, 14) = 0.1073$

b) Find $t_\alpha(f) : P(T \leq t_\alpha(f)) = \alpha$ (α given konstant). $\text{invT}(\alpha, f)$

Eksempel: $t_{0.975}(12) = \text{invT}(0.975, 12) = 2.179$

χ^2 - fordeling.

Lad Q være χ^2 - fordelt med frihedsgradstallet f .

a) Find $p = P(a \leq Q \leq b)$, hvor a og b er givne konstanter. $p = \text{chi2Cdf}(a, b, f)$

Eksempel: Find $p = P(Q \geq 27.3)$, med $f = 19$ $p = \text{chi2Cdf}(27.3, \infty, 19) = 0.0979$

b) Find fraktilen $\chi_\alpha^2(f) : P(Q \leq \chi_\alpha^2(f)) = \alpha$ (α given konstant). $\text{invChi2}(\alpha, f)$

Eksempel: $\chi_{0.025}^2(8) = \text{invChi2}(0.025, 8) = 2.18$

Binomialfordeling.

Lad X være binomialfordelt $b(n, p)$

Find $P(l \leq X \leq m)$, hvor $0 \leq l < m \wedge m \leq n$ og l og m er hele tal. $\text{binomtCdf}(n, p, l, m)$

Eksempel: $p = P(3 \leq X \leq 6)$ hvor $n = 10$ og $p = 0.3$ $p = \text{binomtCdf}(10, 0.3, 3, 6) = 0.6066$

Poissonfordeling.

Lad X være Poissonfordelt $p(\mu)$

Find $P(l \leq X \leq m)$, hvor $0 \leq l < m$ og l og m er hele tal. $\text{poissCdf}(\mu, l, m)$

Eksempel: $p = P(X \leq 94)$, hvor $\mu = 147.6$ $p = \text{poissCdf}(147.6, 0, 94) = 0.00002$

3) Gennemsnit, varians og spredning

Find gennemsnit, varians og spredning af tallene 1, 3, 4, 8

Metode 1. HOME\ MATH\6.Statistics\

Anvendes hvis man har få tal og kun ønsker beregning af en enkelt størrelse.

Gennemsnit: Mean (liste)

Eksempel :

$$\text{Mean}(\{1,3,4,8\}) = 4$$

Varians: Variance(liste)

Eksempel :

$$\text{Variance}(\{1,3,4,8\}) = 8.667$$

Spredning: stdDev (liste)

Eksempel :

$$\text{stdDev}(\{1,3,4,8\}) = 2.944$$

Metode 2. APPS\ Stats/List\ Data indtastes i "list1" F4\ 1: 1-Var Stats

I menu sættes "List" til "List1" (Benyt evt. Var-Link til at finde List1)

Udskriften består af en række statistiske størrelser. Man finder

$$\bar{x} = 4, \sigma_x = 2.944$$

4) Hypotesetest og konfidensintervaller

APPS\ STAT\LIST hvorefter eventuelle data indtastes i list1

4.1. Normalfordeling.

a1) **Hypotesetest;** σ kendt: F6\ 1: Z-Test\

Eksempel: Lad data være gemt i list1: $\{1,3,4,8\}$, $\sigma = 3$ og $H: \mu < 5$

Vælg Data: I menu: $\mu_0 = 5$, $\sigma = 3$, list = list1, Altern. Hyp $\mu < \mu_0$

P-værdi = 0.25

Eksempel: Opgivet $\bar{x} = 4$, $\sigma = 3$, $n = 10$, $H: \mu < 5$

Vælg Stats: I menu: $\mu_0 = 5$, $\sigma = 3$, $\bar{x} = 4$, $n = 10$, Altern. Hyp $\mu < \mu_0$

P-værdi = 0.145

a2) **Konfidensinterval** σ kendt: F7\ 1: Z-Interval

Eksempel: Lad data være gemt i list1: $\{1,3,4,8\}$, $\sigma = 3$

Vælg Data: I menu: $\sigma = 3$, list = list1, C Level = .95

C Int = [1.05; 6.54]

Eksempel: Opgivet $\bar{x} = 5$, $\sigma = 3$, $n = 10$,

Vælg Stats: I menu: $\sigma = 3$, $\bar{x} = 5$, $n = 10$

C Int = [3.14; 6.85]

b1) **Hypotesetest;** σ ukendt: F6\ 2: T - Test\

Eksempel: Lad data være gemt i list1: $\{1,3,4,8\}$ $H: \mu < 5$

Vælg Data: I menu: $\mu_0 = 5$, list = list1, Altern. Hyp $\mu < \mu_0$

P-værdi = 0.27

Eksempel: Opgivet $\bar{x} = 4$, $s = 3$, $n = 10$, $H: \mu < 5$

Vælg Stats: I menu: $\mu_0 = 5$, $\bar{x} = 4$, $s = 3$, $n = 10$, Altern. Hyp $\mu < \mu_0$

P-værdi = 0.160

b2) **Konfidensinterval** σ ukendt: F7\ 1: T-Interval

Eksempel: Lad data være gemt i list1: $\{1,3,4,8\}$

Vælg Data: I menu: list = list1, C Level = .95

C Int = [-0.684; 8.684]

Eksempel: Opgivet $\bar{x} = 4$, $s = 3$, $n = 10$,

Vælg Stats: I menu: $\bar{x} = 4$, $s = 3$, $n = 10$

C Int = [1.85; 6.15]

4.2. Binomialfordeling.

a1) Hypotesetest: F6, 5: 1-Prop-ZTest

(Kræver der kan approksimeres til normalfordeling)

Eksempel: X er binomialfordelt $b(24, p)$, $x = 13$, $H: p > 0.3$

Vælg: $p_0 = 0.3$, Successes = 13, $n = 24$, Alternate Hyp : $\text{prop} > p_0$

P -værdi = 0.00489

a2) Konfidensinterval: F7, 5: 1-Prop-ZInt

(Kræver der kan approksimeres til normalfordeling)

Eksempel: Af 24 forsøg de 13 en succes:

Vælg: Successes = 13, $n = 24$,

C Int = [0.34; 0.74]

Poissonfordeling: findes ikke, så her må formel for konfidensinterval benyttes

TI 83

1) Sandsynlighedsfordelinger.

Man vælger **2nd DISTR**

Normalfordeling $n(\mu, \sigma)$

- a) Find $p = P(a \leq X \leq b)$, hvor a, b, μ, σ er givne konstanter.

$$p = \text{normcdf}(a, b, \mu, \sigma)$$

Eksempel: $p = P(X \leq 11.6)$, hvor $\mu = 11.3$, $\sigma = 5$

$$p = \text{normcdf}(-1000, 11.6, 11.3, 5) = 0.524$$

Bemærk: nedre grænse er sat til -1000

- b) Find $x_p : P(X \leq x_p) = p$, hvor p, μ, σ er givne konstanter.

$$x_p = \text{invNorm}(p, \mu, \sigma)$$

Eksempel: $P(X \leq x_p) = 0.7$, hvor $\mu = 11.4$, $\sigma = 6$

$$x_p = \text{invNorm}(0.7, 11.4, 6) = 14.55$$

t - fordeling.

Lad T være t - fordelt med frihedsgradstallet f .

- a) Find $p = P(a \leq T \leq b)$, hvor a og b er givne konstanter.

$$t\text{Cdf}(a, b, f)$$

Eksempel: $p = P(T \leq -1.3)$, med $f = 14$

$$p = t\text{Cdf}(-\infty, -1.3, 14) = 0.1073$$

- b) Find $t_\alpha(f) : P(T \leq t_\alpha(f)) = \alpha$ (α given konstant).

Kan ikke findes direkte (mærkværdigt) så lettest at slå op i tabel.

Kan dog besværligt findes ved at løse den tilsvarende ligning:

Eksempel: Find $t_{0.975}(12)$

Vi løser ligningen $t\text{Cdf}(-100, x, 12) - 0.975 = 0$ med hensyn til x , idet vi som startgæt vælger 1

CATALOG, solve($t\text{Cdf}(-100, x, 12) - 0.975, x, 1$)

Resultat: 2.179

χ^2 fordeling.

- a) Find $p = P(a \leq Q \leq b)$, hvor a og b er givne konstanter.

$$p = \text{chi2Cdf}(a, b, f)$$

Eksempel: Find $p = P(Q \geq 27.3)$, med $f = 19$

$$p = \text{chi2Cdf}(27.3, 1000, 19) = 0.0979$$

- b) Find fraktilen $\chi_\alpha^2(f) : P(Q \leq \chi_\alpha^2(f)) = \alpha$ (α given konstant). Findes ikke, så enten tabelopslag eller løse ligning som ovenfor.

Binomialfordeling.

Lad X være binomialfordelt $b(n, p)$

Find $P(l \leq X \leq m)$, hvor $0 \leq l < m \leq n$ og l og m er hele tal.

$$\text{binomtCdf}(n, p, l, m)$$

Eksempel: $p = P(3 \leq X \leq 6)$ hvor $n = 10$ og $p = 0.3$

$$p = \text{binomtCdf}(10, 0.3, 6) - \text{binomtCdf}(10, 0.3, 2) = 0.6066$$

Poissonfordeling.

Lad X være Poissonfordelt $p(\mu)$

Find $P(l \leq X \leq m)$, hvor $0 \leq l < m$ og l og m er hele tal.

$$\text{poissCdf}(\mu, l, m)$$

Eksempel: $p = P(X \leq 94)$, hvor $\mu = 147.6$

$$p = \text{poissCdf}(147.6, 94) = 0.00002$$

Find gennemsnit, varians og spredning af tallene 1, 3, 4, 8

Metode 1. List \ MATH \

Anvendes hvis man har få tal og kun ønsker beregning af en enkelt størrelse.

Gennemsnit: Mean (liste)

Eksempel :

$$\text{Mean}(\{1, 3, 4, 8\}) = 4$$

Varians: Variance(liste)

Eksempel :

$$\text{Variance}(\{1, 3, 4, 8\}) = 8.667$$

Spredning: stdDev (liste)

Eksempel :

$$\text{stdDev}(\{1, 3, 4, 8\}) = 2.944$$

Metode 2. STAT \ Edit

I listen L1 skrives 1 ENTER 3 ENTER 4 ENTER 8 ENTER (slet eventuelle allerede indtastede tal med DEL STAT ► hvorved cursor står på CALC. ENTER vælg 1 -Var STATS ENTER

2nd L1 ENTER

Resultat blandt meget andet: $\bar{x} = 4$ og $s_x = 2.9439$

Beregn **binomialkoefficient** $K(8,3) = \binom{8}{3}$: 8 MATH, PROB, nCr 3. ENTER

Resultat 560

2. Hypotesetest og konfidensintervaller

Eventuelle data indtastes i list1. Derefter vælges STAT, TESTS

2.1. Normalfordeling.

a1) Hypotesetest\ σ kendt: 1: Z-Test

Eksempel: Lad data være gemt i list1: {1,3,4,8}, $\sigma = 3$ og H: $\mu < 5$

Vælg Data: I menu: $\mu_0 = 5$, $\sigma = 3$, List = L1, $\mu < \mu_0$ CALCULATE

P-værdi = 0.25

Eksempel: Opgivet $\bar{x} = 4$, $\sigma = 3$, $n = 10$, H: $\mu < 5$

Vælg Stats: I menu: $\mu_0 = 5$, $\sigma = 3$, $\bar{x} = 4$, $n = 10$, $\mu < \mu_0$

P-værdi = 0.145

a2) Konfidensinterval σ kendt: 7: Z-Interval

Eksempel: Lad data være gemt i list1: {1,3,4,8}, $\sigma = 3$

Vælg Data: I menu: $\sigma = 3$, list = L1, C Level = .95

C Int = [1.05; 6.54]

Eksempel: Opgivet $\bar{x} = 5$, $\sigma = 3$, $n = 10$,

Vælg Stats: I menu: $\sigma = 3$, $\bar{x} = 5$, $n = 10$

C Int = [3.14; 6.86]

b1) **Hypotesetest;** σ ukendt: 2: T - Test\

Eksempel: Lad data være gemt i list1: {1,3,4,8} H: $\mu < 5$

Vælg Data: I menu: $\mu_0 = 5$, list = L1, $\mu < \mu_0$

P-værdi = 0.27

Eksempel: Opgivet $\bar{x} = 4$, $s = 3$, $n = 10$, H: $\mu < 5$

Vælg Stats: I menu: $\mu_0 = 5$, $\bar{x} = 4$, $s = 3$, $n = 10$, $\mu < \mu_0$

P-værdi = 0.160

b2) **Konfidensinterval** σ ukendt: 8: T-Interval

Eksempel: Lad data være gemt i list1: {1,3,4,8}

Vælg Data: I menu: list = list1, C Level = .95

C Int = [-0.684; 8.684]

Eksempel: Opgivet $\bar{x} = 4$, $s = 3$, $n = 10$,

Vælg Stats: I menu: $\bar{x} = 4$, $s = 3$, $n = 10$

C Int = [1.85; 6.15]

2.2. Binomialfordeling.

a1) Hypotesetest: 5: 1-Prop-ZTest (Kræver der kan approksimeres til normalfordeling)

Eksempel: X er binomialfordelt $b(24, p)$, $x = 13$, H: $p > 0.3$

Vælg: $p_0 = 0.3$, $x = 13$, $n = 24$, $\text{prop} > p_0$

P -værdi = 0.00489

a2) Konfidensinterval: A: 1-Prop-ZInt (Kræver der kan approksimeres til normalfordeling)

Eksempel: Af 24 forsøg er de 13 en succes:

Vælg: $x = 13$, $n = 24$,

C Int = [0.34; 0.74]

Poissonfordeling: findes ikke, så her må formel for konfidensinterval benyttes

Excel

1: Forudsætninger.

Excel har en statistikpakke "Analysis ToolPak", som ikke automatisk bliver installeret. Dette "tilføjelsesprogram" skal man ofte særskilt "bede" om. Hvorledes dette gøres, kan ses ved at trykke på hjælpetasten F1 og søge "Analysis ToolPak"+Analysis ToolPak VBA"

2). Sandsynlighedsfunktioner

- 1) Vælg den celle hvor resultatet skal stå (eksempelvis A1).
- 2) På værktøjslinien foroven:
 - 2a) Tryk på f_x , hvorved der fremkommer en rullemenu
 - 2c) På denne vælges den ønskede funktion eksempelvis "NORMALFORDELING"
 - 2d) Der fremkommer en tabel med anvisning på, hvordan den skal udfyldes.

a2.1. Normalfordeling.

Normalfordeling $n(\mu, \sigma)$

- a) Find $p = P(a \leq X \leq b)$, hvor a, b, μ, σ er givne konstanter.

$$p = P(X \leq b) - P(X \leq a) = \text{NORMFORDELING}(b; \mu; \sigma; \text{SAND}) - \text{NORMFORDELING}(a; \mu; \sigma; \text{SAND})$$

Eksempel: $p = P(X \leq 11.6)$, hvor $\mu = 11.3$, $\sigma = 5$ $p = \text{NORMFORDELING}(11.6; 11.3; 5; \text{SAND}) = 0.524$

- b) Find $x_p : P(X \leq x_p) = p$, hvor p, μ, σ er givne konstanter.

$$x_p = \text{NORMINV}(p; \mu; \sigma)$$

Eksempel: $P(X \leq x_p) = 0.7$, hvor $\mu = 11.4$, $\sigma = 6$ $x_p = \text{NORMINV}(0.7; 11.4; 6) = 14.55$

t - fordeling.

Lad T være t - fordelt med frihedsgradstallet f .

- a) Find $p = P(T \leq a)$ hvor a er en given konstant.

$$p = 1 - \text{TFORDELING}(\text{abs}(a); f; 1)$$

(bemærk: $\text{TFORDELING}(\text{abs}(a); f; 1)$ udregner "øvre hale" af fordelingen)

Eksempel: $p = P(T \leq -1.3)$, med $f = 14$

$$p = 1 - \text{TFORDELING}(\text{abs}(-1.3); 14; 1) = 0.1073$$

- b) Find $t_\alpha(f) : P(T \leq t_\alpha(f)) = \alpha$ (α given konstant).

$$t_\alpha(f) = \text{TINV}(2(1 - \alpha); f), \alpha > 0.5$$

$$t_\alpha(f) = -\text{TINV}(2\alpha; f), \alpha < 0.5$$

(bemærk: $\text{TINV}(\alpha; f)$ udregner "øvre hale", svarende til $1 - \frac{\alpha}{2}$)

Eksempel:

$$t_{0.975}(12) = \text{TINV}(2(1-0.975); 12) = 2.179$$

$$t_{0.025}(12) = -\text{TINV}(0.05; 12) = -2.179$$

χ^2 - fordeling.

Lad Q være χ^2 - fordelt med frihedsgradstallet f .

- a) Find $p = P(Q \leq a)$, hvor a er en positiv konstant.

$$p = 1 - \text{CHIFORDELING}(a; f)$$

(bemærk: $\text{CHIFORDELING}(x; f)$ udregner "øvre hale" af fordelingen)

Eksempel: Find $p = P(Q \geq 27.3)$, med $f = 19$

$$p = \text{CHIFORDELING}(27.3; 19) = 0.0979$$

- b) Find fraktilen $\chi_\alpha^2(f) : P(Q \leq \chi_\alpha^2(f)) = \alpha$ (α given konstant).

$$\chi_\alpha^2(f) = \text{CHIINV}(1 - \alpha; f)$$

(bemærk: $\text{CHIINV}(\alpha; f)$ udregner "øvre hale")

Eksempel

$$\chi_{0.025}^2(8) = \text{CHIINV}(1-0.025; 8) = 2.18$$

Binomialfordeling.Lad X være binomialfordelt $b(n, p)$

$$P(X = x) = \text{BINOMIALFORDELING}(x; n; p; \text{FALSK})$$

$$P(X \leq x) = \text{BINOMIALFORDELING}(x; n; p; \text{SAND})$$

Eksempel: $p = P(3 \leq X \leq 6)$ hvor $n = 10$ og $p = 0.3$

$$p = P(X \leq 6) - P(X \leq 2) = \text{BINOMIALFORDELING}(6; 10; 0.3; \text{SAND}) - \text{BINOMIALFORDELING}(2; 10; 0.3; \text{SAND}) = 0.6066$$

Poissonfordeling.Lad X være Poissonfordelt $p(\mu)$

$$P(X = x) = \text{POISSON}(x; \mu; \text{FALSK})$$

$$P(X \leq x) = \text{POISSON}(x; \mu; \text{SAND})$$

Eksempel: $p = P(X \leq 94)$, hvor $\mu = 147.6$

$$p = \text{POISSON}(94; 147.6; \text{SAND}) = 0.00002$$

3. Gennemsnit, varians og spredning**Placering:** Vi vil i det følgende for kortheds skyld antage, at stikprøvens værdier står i cellerne A1, A2, A3 ... An.Gennemsnit: $\bar{x} = \text{MIDDEL}(A1:An)$,

$$\text{MIDDEL}(1; 3; 4; 8) = 4$$

Varians: $V(X) = \text{VARIANS}(A1:An)$

$$\text{VARIANS}(1; 3; 4; 8) = 8.667$$

Spredning $s = \text{STDAFV}(A1:An)$,

$$s = \text{STDAFV}(1; 3; 4; 8) = 2.944$$

4. Hypotesetest og konfidensintervaller**4.1. Normalfordeling.**Lad data være gemt i en søjle med navnet "tal". Lad tal være $\{1, 3, 4, 8\}$ a1) **Hypotesetest;** σ kendt:**Eksempel:** $\sigma = 3$ og $H: \mu < 5$

$$P\text{-værdi} = \text{ZTEST}(\text{tal}; 5; 3) = 0.25$$

 $\sigma = 3$ og $H: \mu > 5$

$$P\text{-værdi} = 1 - \text{ZTEST}(\text{tal}; 5; 3)$$

a2) **Konfidensinterval** σ kendt. $\text{middel}(\text{tal}) \pm \text{KONFIDENSINTERVAL}(0.05; \sigma; n)$ **Eksempel:** $\sigma = 3$ $\text{middel}(\text{tal}) \pm \text{KONFIDENSINTERVAL}(0.05; 3; 4) = 4 \pm 2.94$ b1) **Hypotesetest;** σ ukendt: $F6 \setminus 2$: T - Test

$$\text{Der findes ikke en t-test for 1 variabel, så man må beregne P - værdi ud fra } t = \frac{(\bar{x} - \mu_0) \cdot \sqrt{n}}{s}$$

Eksempel: $H: \mu < 5$

Da t-fordelingen i Excel ikke accepterer negative værdier må man tage den numeriske værdi (abs) af resultatet.

$$P\text{-værdi} = \text{TFORDELING}((\text{ABS}(\text{MIDDEL}(\text{tal}) - 5) / \text{STDAFV}(\text{tal}) * 4^{0.5}); 4; 1) = 0.27$$

b2) **Konfidensinterval** σ ukendt:

$$\text{Der beregnes radius } r \text{ i et 95\% konfidensinterval for } \mu: \bar{x} \pm r = \bar{x} \pm t_{0.975}(f) \frac{s}{\sqrt{n}}$$

Eksempel: Vælg "Funktioner" \ Dataanalyse \ Beskrivende statistik \ Udfylder menu, bl.a. sæt kryds ved konfidensniveau for middelværdi.

Kolonnel	
Konfidensniveau(95,0%)	4,684434

$$r = 4.684 \quad 95\% \text{ konfidensinterval: } 4 \pm 4.684$$

4.2. Binomialfordeling.

Der findes ikke en binomialtest

Appendix 2: Statistiske beregninger på lommeregner og PC'er

MAPLE

Beregn gennemsnit og spredning af tallene 1 3 4 8

```
> with(stats):
data:=[1,3,4,8];
data := [1, 3, 4, 8]
> describe[mean](data);
4
> describe[standarddeviation[1]](data);

$$\frac{\sqrt{78}}{3}$$

```

Beregn **korrelationskoefficient** for den i eksempel 9.2 nævnte stikprøve
(1,2), (0,0), (2,2), (2,2), (1,0), (2,2), (0,2), (2,2), (0,2), (2,2) .

Programudførelse:

```
> data1:=[1,0,2,2,1,2,0,2,0,2];
data1 := [1, 0, 2, 2, 1, 2, 0, 2, 0, 2]
> data2:=[2,0,2,2,0,2,2,2,2,2];
data2 := [2, 0, 2, 2, 0, 2, 2, 2, 2, 2]
> describe[linearcorrelation](data1,data2): evalf("");
.4014775343
```

x- værdier
udskrift
y-værdier
udskrift
resultat

Normalfordeling.

Find for $n(113.3, 5.6)$ $P(X \leq 116.1)$.

Programudførelse:

```
> with(stats):
> with(statevalf):
> cdf[normald[113.3,5.6]](116.1);
Facit .6914624613
```

χ^2 fordeling.

Find en tests P -værdi: $P(Q \geq 27.26)$ idet frihedsgradstallet er 19 (jævnfør eksempel 5.6)

Programudførelse:

```
> with(stats):
> with(statevalf):
> 1-cdf[chisquare[19]](27.36);
Facit: .0965431211
```

t - fordeling.

Find en tests P -værdi: $P(T \leq -1.31)$ idet frihedsgradstallet er 14 (jævnfør eksempel 5.5)

Programudførelse:

```
> with(stats):
> with(statevalf):
> cdf[studentst[14]](-1.31);
Facit: .1056420798
```

Find for **binomialfordelingen** $b(100, 0.3)$ $P(X \leq 35)$

Programudførelse:

```
> with(stats):
> with(statevalf):
> dcdff[binomiald[100,0.3]](35);
Facit: .8839213940
```

APPENDIX 4.1. Oversigt over konfidensintervaller

nr	Forudsætninger	Estimat for parameter	100 (1 - α) % konfidensinterval for parameter
X er normalfordelt $n(\mu, \sigma)$. Givet stikprøve af størrelsen n med gennemsnit $\bar{x}$ og spredning s			
1	μ ukendt. σ kendt	For μ : $\bar{x}$	$\bar{x} - u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$
2	μ ukendt. σ ukendt	For μ : $\bar{x}$	$\bar{x} - t_{1-\frac{\alpha}{2}}(n-1) \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{1-\frac{\alpha}{2}}(n-1) \cdot \frac{s}{\sqrt{n}}$
3	μ kendt σ ukendt.	For σ^2 : $s_\mu^2 = \sum_{i=1}^n \frac{(X_i - \mu)^2}{n} = \frac{(n-1)s^2 + n(\bar{x} - \mu)^2}{n}$	$\frac{(n-1)s^2 + n(\bar{x} - \mu)^2}{\chi_{1-\frac{\alpha}{2}}^2(n)} \leq \sigma^2 \leq \frac{(n-1)s^2 + n(\bar{x} - \mu)^2}{\chi_{\frac{\alpha}{2}}^2(n)}$
4	μ ukendt σ ukendt.	For σ^2 : S^2	$\frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}}^2(n-1)}$
X er binomialfordelt $b(n, p)$, hvor n er kendt og p ukendt. Givet stikprøveværdi x .			
5	$10 \leq x \wedge x \leq n-10$	For p : $\tilde{p} = \frac{x}{n}$	$\tilde{p} - u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}} \leq p \leq \tilde{p} + u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}$
X er Poissonfordelt $p(\mu)$, hvor μ ukendt. Der udtages en stikprøve af størrelsen n , og der optælles i alt m impulser.			
6	$m \geq 10$	For μ : $\bar{x} = \frac{m}{n}$	$\bar{x} - u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\bar{x}}{n}} \leq \mu \leq \bar{x} + u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\bar{x}}{n}}$

APPENDIX 5.1. Test af middelværdi μ for normalfordelt variabel

Der foreligger en stikprøve på X af størrelsen n med gennemsnit $\bar{x}$ og spredning s . T er en stokastisk variabel der er t -fordelt med $f = n - 1$.

Forudsætninger	Alternativ hypotese H	P -værdi	Beregning på TI-89	H_0 forkastes	Dimensionering *)
σ ukendt. Signifikansniveau: α . μ_0 er en given konstant $t = \frac{(\bar{x} - \mu_0) \cdot \sqrt{n}}{s}$	$H: \mu > \mu_0$	$P(T \geq t)$	$tCdf(t, \infty, n-1)$	P -værdi $< \alpha$	$\Delta = \mu - \mu_0 $ er den mindste ændring i μ der har praktisk interesse. $\beta = P(\text{type II fejl})$ Benyt dimensioneringstabel tabel 8.
	$H: \mu < \mu_0$	$P(T \leq t)$	$tCdf(-\infty, t, n-1)$		
	$H: \mu \neq \mu_0$	$P(T \geq t)$ for $\bar{x} > \mu_0$ $P(T \leq t)$ for $\bar{x} \leq \mu_0$	som ovenfor	P -værdi $< \frac{1}{2}\alpha$	
	$H: \mu > \mu_0$	Tabelmetode		$t > t_{1-\alpha}(n-1)$	
	$H: \mu < \mu_0$	Tabelmetode		$t < -t_{1-\alpha}(n-1)$	
	$H: \mu \neq \mu_0$	Tabelmetode		$ t > t_{1-\frac{\alpha}{2}}(n-1)$	

*) Se eventuelt kommentar i supplement 5B

APPENDIX 5.2. Test af middelværdi μ for normalfordelt variabel .

Der foreligger en stikprøve på X af størrelsen n med gennemsnit $\bar{x}$ og spredning s .

Forudsætninger	Alternativ hypotese H	P - værdi	Beregning på TI-89	H ₀ forkastes	Dimensionering
σ kendt eksakt Signifikansniveau : α μ_0 er en given konstant Y er normalfordelt $n\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$.	$H: \mu > \mu_0$	$P(Y \geq \bar{x})$	$normCdf(\bar{x}, \infty, \mu_0, \sigma)$	$P\text{-værdi} < \alpha$	$\Delta = \mu - \mu_0 $ er den mindste ændring i μ der har praktisk interesse. $\beta = P(\text{type II fejl})$ $n \geq \left(\frac{u_{1-\alpha} + u_{1-\beta}}{\frac{\Delta}{\sigma}} \right)^2$
	$H: \mu < \mu_0$	$P(Y \leq \bar{x})$	$normCdf(-\infty, \bar{x}, \mu_0, \sigma)$		
	$H: \mu \neq \mu_0$	$P(Y \geq \bar{x})$ for $\bar{x} > \mu_0$ $P(Y \leq \bar{x})$ for $\bar{x} \leq \mu_0$	som ovenfor	$P\text{-værdi} < \frac{1}{2}\alpha$	Betegnelser som ovenfor: $n \geq \left(\frac{u_{1-\frac{\alpha}{2}} + u_{1-\beta}}{\frac{\Delta}{\sigma}} \right)^2$
$u = \frac{(\bar{x} - \mu_0) \cdot \sqrt{n}}{\sigma}$	$H: \mu > \mu_0$	Tabelmetode		$u > u_{1-\alpha}$	Som ovenfor
	$H: \mu < \mu_0$	Tabelmetode		$u < -u_{1-\alpha}$	
	$H: \mu \neq \mu_0$	Tabelmetode		$ u > u_{1-\frac{\alpha}{2}}$	Som ovenfor

Hvis $n > 30$ kan σ approksimativt erstattes af s .

APPENDIX 5.3. Test af varians σ^2 for normalfordelt variabel .

Der foreligger en stikprøve på X af størrelsen n med gennemsnit $\bar{x}$ og spredning s . Q er χ^2 fordelt med $f = n - 1$.

Forudsætninger	Alternativ hypotese	P - værdi	Beregning på TI-89	H_0 forkastes
μ ukendt Signifikansniveau: α σ_0 er en given konstant $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$H: \sigma^2 > \sigma_0^2$	$P(Q \geq \chi^2) 0$	$chi2Cdf(\chi^2, \infty, n-1)$	$P\text{-værdi} < \alpha$
	$H: \sigma^2 < \sigma_0^2$	$P(Q \leq \chi^2)$	$chi2Cdf(-\infty, \chi^2, n-1)$	
	$H: \sigma^2 \neq \sigma_0^2$	$P(Q \geq \chi^2)$ for $\chi^2 \geq n-1$ $P(Q \leq \chi^2)$ for $\chi^2 < n-1$	som ovenfor	$P\text{-værdi} < \frac{1}{2} \alpha$
	$H: \sigma^2 > \sigma_0^2$	Tabelmetode		$\chi^2 > \chi_{1-\alpha}^2(n-1)$
	$H: \sigma^2 < \sigma_0^2$	Tabelmetode		$\chi^2 < \chi_{\alpha}^2(n-1)$
	$H: \sigma^2 \neq \sigma_0^2$	Tabelmetode		$\chi^2 > \chi_{1-\frac{\alpha}{2}}^2(n-1)$ eller $\chi^2 < \chi_{\frac{\alpha}{2}}^2(n-1)$

APPENDIX 5.4 Oversigt over test af varians σ^2 for normalfordelt variabel .

Der foreligger en stikprøve på X af størrelsen n med gennemsnit $\bar{x}$ og spredning s . Q er χ^2 fordelt med $f=n$.

Forudsætninger	Alternativ hypotese:	Beregning	Beregning på TI-89	H_0 forkastes
μ kendt Signifikansniveau er α σ_0 er en given konstant $\chi^2 = \frac{(n-1)s^2 + n(\bar{x} - \mu)^2}{\sigma_0^2}$	$H: \sigma^2 > \sigma_0^2$	$P(Q \geq \chi^2)$	$chi2Cdf(\chi^2, \infty, n)$	$P\text{-værdi} < \alpha$
	$H: \sigma^2 < \sigma_0^2$	$P(Q \leq \chi^2)$	$chi2Cdf(-\infty, \chi^2, n)$	
	$H: \sigma^2 \neq \sigma_0^2$	$P(Q \geq \chi^2)$ for $\chi^2 \geq n$ $P(Q \leq \chi^2)$ for $\chi^2 < n$,	som ovenfor	$P\text{-værdi} < \frac{1}{2} \alpha$
	$H: \sigma^2 > \sigma_0^2$	Tabelmetode		$\chi^2 > \chi_{1-\alpha}^2(n)$
	$H: \sigma^2 < \sigma_0^2$	Tabelmetode		$\chi^2 < \chi_{\alpha}^2(n)$
	$H: \sigma^2 \neq \sigma_0^2$	Tabelmetode		$\chi^2 > \chi_{1-\frac{\alpha}{2}}^2(n)$ eller $\chi^2 < \chi_{\frac{\alpha}{2}}^2(n)$

APPENDIX 5.5. Test af parameter p for binomialfordelt variabel

X er binomialfordelt $b(n, p)$, hvor n er kendt og p er ukendt. Der foreligger en stikprøve på X . Observeret stikprøveværdi x . Signifikansniveau er α . Y er binomialfordelt $b(n, p_0)$, hvor p_0 er en given konstant.

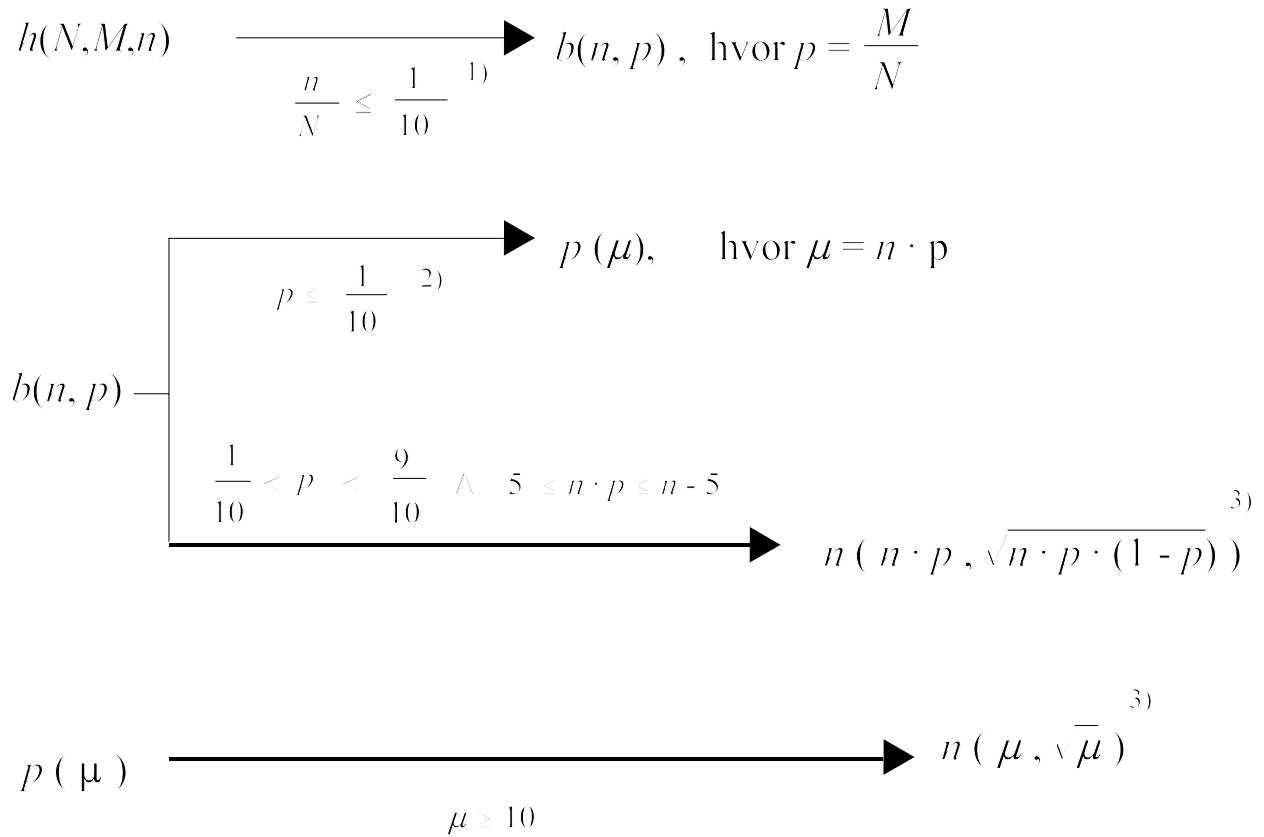
Alternativ hypotese	P - værdi	Beregning på TI 89	H_0 forkastes
$H: p > p_0$	$P(Y \geq x)$	$\text{binomCdf}(n, p_0, x, \infty)$	$P\text{-værdi} < \alpha$
$H: p < p_0$	$P(Y \leq x)$	$\text{binomCdf}(n, p_0, -\infty, x)$	
$H: p = p_0$	$P(Y \geq x)$ for $x > n \cdot p_0$ $P(Y \leq x)$ for $x \leq n \cdot p_0$	som ovenfor	$P\text{-værdi} < \frac{1}{2} \alpha$
Forudsættes $\frac{1}{10} \leq p_0 \leq \frac{9}{10}$ og $5 \leq n \cdot p_0 \leq n - 5$ kan approksimeres med normalfordeling			
Alternativ hypotese	Beregning		H_0 forkastes
$H: p > p_0$	$u = \frac{x - \frac{1}{2} - n \cdot p_0}{\sqrt{n \cdot p_0 \cdot (1 - p_0)}}$	Tabelmetode	$u > u_{1-\alpha}$
$H: p < p_0$		Tabelmetode	$u < -u_{1-\alpha}$
$H: p = p_0$		Tabelmetode	$ u > u_{1-\frac{\alpha}{2}}$

APPENDIX 5.6 Test af parameter μ for Poissonfordelt variabel .

X er Poissonfordelt $p(\mu)$, hvor μ er ukendt. Der foreligger en stikprøve på X af størrelsen n med gennemsnit $\bar{x}$. Signifikansniveau er α .
 Y er Poissonfordelt $p(n \cdot \mu_0)$, hvor μ_0 er en given konstant.

Alternativ hypotese.	P - værdi	Beregning på TI 89	H_0 forkastes
$H: \mu > \mu_0$	$P(Y \geq n \cdot \bar{x})$	$possCdf(n \cdot \mu_0, n \cdot \bar{x}, 1000)$	P - værdi $< \alpha$
$H: \mu < \mu_0$	$P(Y \leq n \cdot \bar{x})$	$possCdf(n \cdot \mu_0, 0, n \cdot \bar{x})$	
$H: \mu \neq \mu_0$	$P(Y \geq n \cdot \bar{x})$ for $x \geq n \cdot \mu_0$ $P(Y \leq n \cdot \bar{x})$ for $x < n \cdot \mu_0$	som ovenfor	P -værdi $< \frac{1}{2} \alpha$
Er $n \cdot \mu_0 \geq 10$ kan approksimeres med normalfordeling			
Alternativ hypotese	Beregning		H_0 forkastes
$H: \mu > \mu_0$	$u = \frac{n \cdot \bar{x} - \frac{1}{2} - n \cdot \mu_0}{\sqrt{n \cdot \mu_0}}$	Tabelmetode	$u > u_{1-\alpha}$
$H: \mu < \mu_0$		Tabelmetode	$u < -u_{1-\alpha}$
$H: \mu \neq \mu_0$		Tabelmetode	$ u > u_{1-\frac{\alpha}{2}}$

APPENDIX 7.1. Oversigt over approksimationer.



1) Når $\frac{n}{N} > \frac{1}{10}$ og $\frac{M}{N} \leq \frac{1}{10}$ benyttes, at $h(N, M, n) = h(N, n, M)$.

2) For $p \geq \frac{9}{10}$ benyttes, i stedet for at tælle X_{gammel} = "antal af succeser", så at tælle X = antal fiaskoer dvs. $p = 1 - p_{\text{gammel}}$

og $X = n - X_{\text{gammel}}$.

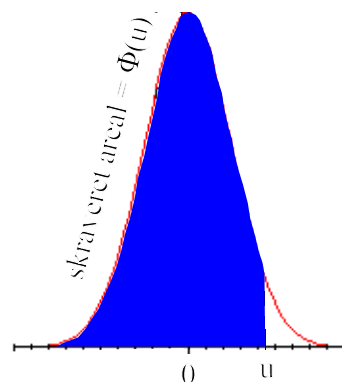
3) Husk heltalskorrektion ved approksimation med normalfordeling.

STATISTISKE TABELLER

Tabel 1 Fordelingsfunksjonen Φ for U -fordelingen $n(0,1)$.

$$\Phi(u) = P(U \leq u) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^u e^{-\frac{t^2}{2}} dt$$

Eksempler: $\Phi(-1.97)=0.0244$ og $\Phi(1.97)=0.9756$



$\Phi(u)$										
u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
P-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0014	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0108	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0126	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0352	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0722	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2297	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Tabel 1 (fortsat) **Fordelingsfunksjonen Φ for U -fordelingen $n(0,1)$.**

$$\Phi(u) = P(U \leq u) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^u e^{-\frac{t^2}{2}} dt$$

u	$\Phi(u)$									
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9441
1.6	0.9452	0.9463	0.9474	0.9485	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9700	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

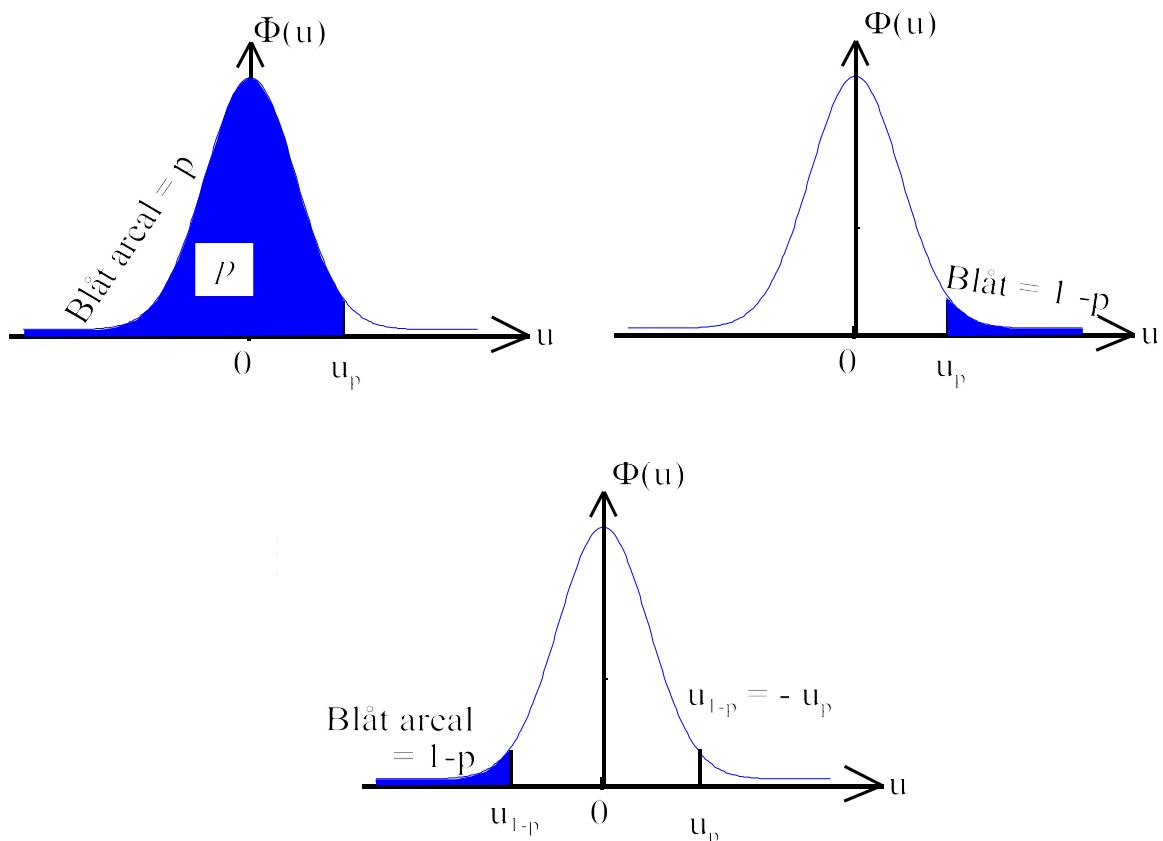
Tabel 2 Fraktiler u_p i U-fordelingen $n(0,1)$. $P(U \leq u_p) = p$.

Bemærk: $u_p = -u_{1-p}$

p	0.0005	0.001	0.005	0.01	0.025	0.05	0.10
u_p	-3.291	-3.090	-2.576	-2.326	-1.960	-1.645	-1.282

p	0.90	0.95	0.975	0.99	0.995	0.999	0.9995
u_p	1.282	1.645	1.960	2.326	2.576	3.090	3.291

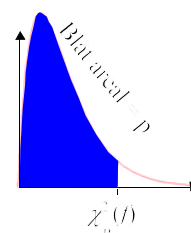
Eksempler: $u_{0.975} = 1.960$



Tabel 3 Fraktiler i χ^2_p -fordelingen $\chi^2(f)$. $P(\chi^2 \leq \chi^2_p) = p$

$\chi^2 = U_1^2 + U_2^2 + \dots + U_f^2$, hvor $U_1, U_2, \dots, U_f$.

Approximativ formel: $\chi^2_p(f) = \frac{1}{2}(\sqrt{2f-1} + u_p)^2$ for $f > 30$



f	p	0.0005	0.001	0.005	0.01	0.025	0.05	0.10	0.20	0.50
1		0.000	0.000	0.000	0.000	0.001	0.004	0.016	0.064	0.455
2		0.001	0.002	0.010	0.020	0.051	0.103	0.211	0.446	1.39
3		0.015	0.024	0.072	0.115	0.216	0.352	0.584	1.01	2.37
4		0.064	0.091	0.207	0.297	0.484	0.711	1.06	1.65	3.36
5		0.158	0.210	0.412	0.554	0.831	1.15	1.61	2.34	4.35
6		0.299	0.381	0.676	0.872	1.24	1.64	2.20	3.07	5.35
7		0.485	0.598	0.989	1.24	1.69	2.17	2.83	3.82	6.35
8		0.710	0.857	1.34	1.65	2.18	2.73	3.49	4.59	7.34
9		0.972	1.15	1.73	2.09	2.70	3.33	4.17	5.38	8.34
10		1.26	1.48	2.16	2.56	3.25	3.94	4.87	6.18	9.34
11		1.59	1.83	2.60	3.05	3.82	4.57	5.58	6.99	10.3
12		1.93	2.21	3.07	3.57	4.40	5.23	6.30	7.81	11.3
13		2.31	2.62	3.57	4.11	5.01	5.89	7.04	8.63	12.3
14		2.70	3.04	4.07	4.66	5.63	6.57	7.79	9.47	13.3
15		3.11	3.48	4.60	5.23	6.26	7.26	8.55	10.3	14.3
16		3.54	3.94	5.14	5.81	6.91	7.96	9.31	11.2	15.3
17		3.98	4.42	5.70	6.41	7.56	8.67	10.1	12.0	16.3
18		4.44	4.90	6.26	7.01	8.23	9.39	10.9	12.9	17.3
19		4.91	5.41	6.84	7.63	8.91	10.1	11.7	13.7	18.3
20		5.40	5.92	7.43	8.26	9.59	10.9	12.4	14.6	19.3
22		6.40	6.98	8.64	9.54	11.0	12.3	14.0	16.3	21.3
24		7.45	8.08	9.89	10.9	12.4	13.8	15.7	18.1	23.3
26		8.54	9.22	11.2	12.2	13.8	15.4	17.3	19.8	25.3
28		9.66	10.4	12.5	13.6	15.3	16.9	18.9	21.6	27.3
30		10.8	11.6	13.8	15.0	16.8	18.5	20.6	23.4	29.3
32		12.0	12.8	15.1	16.4	18.3	20.1	22.3	25.1	31.3
34		13.2	14.1	16.5	17.8	19.8	21.7	24.0	26.9	33.3
36		14.4	15.3	17.9	19.2	21.3	23.3	25.6	28.7	35.3
38		15.6	16.6	19.3	20.7	22.9	24.9	27.3	30.5	37.3
40		16.9	17.9	20.7	22.2	24.4	26.5	29.1	32.3	39.3
45		20.1	21.3	24.3	25.9	28.4	30.6	33.4	36.9	44.3
50		23.5	24.7	28.0	29.7	32.4	34.8	37.7	41.4	49.3
60		30.3	31.7	35.5	37.5	40.5	43.2	46.5	50.6	59.3
70		37.5	39.0	43.3	45.4	48.8	51.7	55.3	59.9	69.3
80		44.8	46.5	51.2	53.5	57.2	60.4	64.3	69.2	79.3
90		52.3	54.2	59.2	61.8	65.6	69.1	73.3	78.6	89.3
100		59.9	61.9	67.3	70.1	74.2	77.9	82.4	87.9	99.3
110		67.6	69.8	75.5	78.5	82.9	86.8	91.4	97.4	109
120		75.5	77.8	83.9	86.9	91.6	95.7	101	107	119
130		83.4	85.8	92.2	95.4	100	105	110	116	129
140		91.4	93.9	101	104	109	114	119	126	139
150		99.5	102	109	113	118	123	128	135	149
160		108	110	118	121	127	132	137	145	159
180		124	127	135	139	145	150	156	164	179
200		141	144	152	156	163	168	175	183	199
300		226	230	241	246	254	261	269	279	299
400		313	318	331	337	346	355	364	376	399
500		403	408	422	429	440	449	460	473	499
1000		859	868	889	899	914	928	943	962	999

Tabel 3 (fortsat). Fraktiler i χ^2 -fordelingen $\chi^2(f)$. $P(\chi^2 \leq \chi_p^2) = p$

Eksempler: For $\chi^2(10)$ er $P(X \leq 20,5) = 0.975$ og $\chi_{0.05}^2(30) = 18.5$

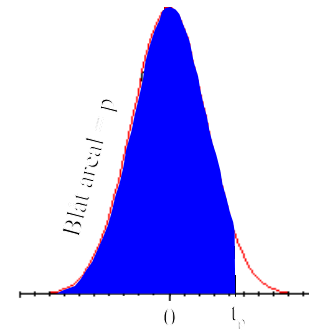
f	p	0.80	0.90	0.95	0.975	0.99	0.995	0.999	0.9995
1		1.64	2.71	3.84	5.02	6.63	7.88	10.8	12.1
2		3.22	4.61	5.99	7.38	9.21	10.6	13.8	15.2
3		4.64	6.25	7.81	9.35	11.3	12.8	16.3	17.7
4		5.99	7.78	9.49	11.1	13.3	14.9	18.5	20.0
5		7.29	9.24	11.1	12.8	15.1	16.7	20.5	22.1
6		8.56	10.6	12.6	14.4	16.8	18.5	22.5	24.1
7		9.80	12.0	14.1	16.0	18.5	20.3	24.3	26.0
8		11.0	13.4	15.5	17.5	20.1	22.0	26.1	27.9
9		12.2	14.7	16.9	19.0	21.7	23.6	27.9	29.7
10		13.4	16.0	18.3	20.5	23.2	25.2	29.6	31.4
11		14.6	17.3	19.7	21.9	24.7	26.8	31.3	33.1
12		15.8	18.5	21.0	23.3	26.2	28.3	32.9	34.8
13		17.0	19.8	22.4	24.7	27.7	29.8	34.5	36.5
14		18.2	21.1	23.7	26.1	29.1	31.3	36.1	38.1
15		19.3	22.3	25.0	27.5	30.6	32.8	37.7	39.7
16		20.5	23.5	26.3	28.8	32.0	34.3	39.3	41.3
17		21.6	24.8	27.6	30.2	33.4	35.7	40.8	42.9
18		22.8	26.0	28.9	31.5	34.8	37.2	42.3	44.4
19		23.9	27.2	30.1	32.9	36.2	38.6	43.8	46.0
20		25.0	28.4	31.4	34.2	37.6	40.0	45.3	47.5
22		27.3	30.8	33.9	36.8	40.3	42.8	48.3	50.5
24		29.6	33.2	36.4	39.4	43.0	45.6	51.2	53.5
26		31.8	35.6	38.9	41.9	45.6	48.3	54.1	56.4
28		34.0	37.9	41.3	44.5	48.3	51.0	56.9	59.3
30		36.3	40.3	43.8	47.0	50.9	53.7	59.7	62.2
32		38.5	42.6	46.2	49.5	53.5	56.3	62.5	65.0
34		40.7	44.9	48.6	52.0	56.1	59.0	65.2	67.8
36		42.9	47.2	51.0	54.4	58.6	61.6	68.0	70.6
38		45.1	49.5	53.4	56.9	61.2	64.2	70.7	73.4
40		47.3	51.8	55.8	59.3	63.7	66.8	73.4	76.1
45		52.7	57.5	61.7	65.4	70.0	73.2	80.1	82.9
50		58.2	63.2	67.5	71.4	76.2	79.5	86.7	89.6
60		69.0	74.4	79.1	83.3	88.4	92.0	99.6	103
70		79.7	85.5	90.5	95.0	100	104	112	116
80		90.4	96.6	102	107	112	116	125	128
90		101	108	113	118	124	128	137	141
100		112	119	124	130	136	140	149	153
110		122	129	135	141	147	152	162	165
120		133	140	147	152	159	164	174	178
130		143	151	158	163	170	175	186	190
140		154	162	169	175	182	187	197	202
150		164	172	180	186	193	198	209	214
160		175	183	191	197	205	210	221	225
180		196	205	212	219	227	233	244	249
200		217	226	234	241	249	255	268	272
300		320	332	341	350	360	367	381	387
400		424	437	448	457	469	477	493	500
500		527	541	553	564	577	585	604	611
1000		1038	1058	1075	1090	1107	1119	1144	1154

Tabel 4 Fraktiler t_p i t -fordelingen $t(f)$.

$$P(T \leq t_p) = p, \text{ hvor } T = \frac{(\bar{X} - \mu)\sqrt{n}}{s}$$

Eksempler: For $t(27)$ er $P(X \leq 3.42) = 0.999$.

$$t_{0.95}(1)=6.31. \quad t_{0.05}(10) = -t_{0.95}(10) = -1.81.$$



f \ p	0.60	0.75	0.90	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.33	1.00	3.08	6.31	12.7	31.8	63.7	318	637
2	0.29	0.82	1.89	2.92	4.30	6.97	9.93	22.3	31.6
3	0.28	0.74	1.64	2.35	3.18	4.54	5.84	10.2	12.9
4	0.27	0.74	1.53	2.13	2.78	3.75	4.60	7.17	8.61
5	0.27	0.72	1.48	2.02	2.57	3.37	4.03	5.89	6.86
6	0.27	0.72	1.44	1.94	2.45	3.14	3.71	5.21	5.96
7	0.26	0.71	1.42	1.90	2.37	3.00	3.50	4.79	5.41
8	0.26	0.71	1.40	1.86	2.31	2.90	3.36	4.50	5.04
9	0.26	0.70	1.38	1.83	2.26	2.82	3.25	4.30	4.78
10	0.26	0.70	1.37	1.81	2.23	2.76	3.17	4.14	4.59
11	0.26	0.70	1.36	1.80	2.20	2.72	3.11	4.03	4.44
12	0.26	0.70	1.36	1.78	2.18	2.68	3.06	3.93	4.32
13	0.26	0.69	1.35	1.77	2.16	2.65	3.01	3.85	4.22
14	0.26	0.69	1.35	1.76	2.15	2.62	2.98	3.79	4.14
15	0.26	0.69	1.34	1.75	2.13	2.60	2.95	3.73	4.07
16	0.26	0.69	1.34	1.75	2.12	2.58	2.92	3.69	4.02
17	0.26	0.69	1.33	1.74	2.11	2.57	2.90	3.65	3.97
18	0.26	0.69	1.33	1.73	2.10	2.55	2.88	3.61	3.92
19	0.26	0.69	1.33	1.73	2.09	2.54	2.86	3.58	3.88
20	0.26	0.69	1.33	1.73	2.09	2.53	2.85	3.55	3.85
21	0.26	0.69	1.32	1.72	2.08	2.52	2.83	3.53	3.82
22	0.26	0.69	1.32	1.72	2.07	2.51	2.82	3.51	3.79
23	0.26	0.69	1.32	1.71	2.07	2.50	2.81	3.49	3.77
24	0.26	0.69	1.32	1.71	2.06	2.49	2.80	3.47	3.75
25	0.26	0.68	1.32	1.71	2.06	2.49	2.79	3.45	3.73
26	0.26	0.68	1.32	1.71	2.06	2.48	2.78	3.44	3.71
27	0.26	0.68	1.31	1.70	2.05	2.47	2.77	3.42	3.69
28	0.26	0.68	1.31	1.70	2.05	2.47	2.76	3.41	3.67
29	0.26	0.68	1.31	1.70	2.05	2.46	2.76	3.40	3.66
30	0.26	0.68	1.31	1.70	2.04	2.46	2.75	3.39	3.65
40	0.26	0.68	1.30	1.68	2.02	2.42	2.70	3.31	3.55
50	0.26	0.68	1.30	1.68	2.01	2.40	2.68	3.26	3.50
60	0.25	0.68	1.30	1.67	2.00	2.39	2.66	3.23	3.46
80	0.25	0.68	1.29	1.66	1.99	2.37	2.64	3.20	3.42
100	0.25	0.68	1.29	1.66	1.98	2.37	2.63	3.17	3.39
120	0.25	0.68	1.29	1.66	1.98	2.36	2.62	3.16	3.37
200	0.25	0.68	1.29	1.65	1.97	2.35	2.60	3.13	3.34
500	0.25	0.68	1.28	1.65	1.97	2.33	2.59	3.11	3.31
∞	0.25	0.67	1.28	1.65	1.96	2.33	2.58	3.09	3.29

Tabel 5 95%-fraktil $F_{0.95}$ i F -fordelingen $F(f_T, f_N)$. $P(F \leq F_{0.95}) = 0.95$

$f_N \backslash f_T$	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	40	60	120	∞
1	161.5	199.5	215.7	224.6	230.5	234.0	236.8	238.9	240.5	241.9	243.9	245.4	246.5	247.3	248.0	249.3	250.1	251.1	252.2	253.3	253.7
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.42	19.43	19.44	19.45	19.46	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.71	8.69	8.67	8.66	8.63	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.87	5.84	5.82	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.64	4.60	4.58	4.56	4.52	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.96	3.92	3.90	3.87	3.83	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.53	3.49	3.47	3.44	3.40	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.24	3.20	3.17	3.15	3.11	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.03	2.99	2.96	2.94	2.89	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.86	2.83	2.80	2.77	2.73	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.74	2.70	2.67	2.65	2.60	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.64	2.60	2.57	2.54	2.50	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.55	2.52	2.48	2.46	2.41	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.48	2.44	2.41	2.39	2.34	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.42	2.38	2.35	2.33	2.28	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.37	2.33	2.30	2.28	2.23	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.33	2.29	2.26	2.23	2.18	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.29	2.25	2.22	2.19	2.14	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.26	2.21	2.18	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.23	2.18	2.15	2.12	2.07	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.20	2.16	2.12	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.17	2.13	2.10	2.07	2.02	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.15	2.11	2.08	2.05	2.00	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.13	2.09	2.05	2.03	1.97	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.11	2.07	2.04	2.01	1.96	1.92	1.87	1.82	1.77	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.04	1.99	1.96	1.93	1.88	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.95	1.90	1.87	1.84	1.78	1.74	1.69	1.64	1.58	1.51
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.89	1.85	1.81	1.78	1.73	1.69	1.63	1.58	1.51	1.44
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.86	1.82	1.78	1.75	1.69	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.78	1.73	1.69	1.66	1.60	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.69	1.64	1.60	1.57	1.51	1.46	1.39	1.32	1.22	1.00

Tabel 6 Kumulerede sandsynligheder i binomialfordelingen $b(n, p)$. $P(X \leq x)$ Eksempel: For $b(6, 0.4)$ er $P(X \leq 3) = 0.821$.

		$P(X \leq x)$ i binomialfordelingen $b(n, p)$																		
		p																		
n	x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
2	0	0.902	0.810	0.722	0.640	0.562	0.490	0.422	0.360	0.302	0.250	0.203	0.160	0.123	0.090	0.063	0.040	0.023	0.010	0.003
	1	0.998	0.990	0.978	0.960	0.938	0.910	0.878	0.840	0.798	0.750	0.697	0.640	0.578	0.510	0.438	0.360	0.278	0.190	0.097
	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0	0.857	0.729	0.614	0.512	0.422	0.343	0.275	0.216	0.166	0.125	0.091	0.064	0.043	0.027	0.016	0.008	0.003	0.001	0.000
	1	0.993	0.972	0.939	0.896	0.844	0.784	0.718	0.648	0.575	0.500	0.425	0.352	0.282	0.216	0.156	0.104	0.061	0.028	0.007
	2	1.000	0.999	0.997	0.992	0.984	0.973	0.957	0.936	0.909	0.875	0.834	0.784	0.725	0.657	0.578	0.488	0.386	0.271	0.143
	3		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0	0.815	0.656	0.522	0.410	0.316	0.240	0.179	0.130	0.092	0.062	0.041	0.026	0.015	0.008	0.004	0.002	0.001	0.000	0.000
	1	0.986	0.948	0.890	0.819	0.738	0.652	0.563	0.475	0.391	0.312	0.242	0.179	0.126	0.084	0.051	0.027	0.012	0.004	0.000
	2	1.000	0.996	0.988	0.973	0.949	0.916	0.874	0.821	0.759	0.688	0.609	0.525	0.437	0.348	0.262	0.181	0.110	0.052	0.014
	3		1.000	0.999	0.998	0.996	0.992	0.985	0.974	0.959	0.938	0.908	0.870	0.821	0.760	0.684	0.590	0.478	0.344	0.185
	4			1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	0	0.774	0.590	0.444	0.328	0.237	0.168	0.116	0.078	0.050	0.031	0.018	0.010	0.005	0.002	0.001	0.000	0.000	0.000	0.000
	1	0.977	0.919	0.835	0.737	0.633	0.528	0.428	0.337	0.256	0.188	0.131	0.087	0.054	0.031	0.016	0.007	0.002	0.000	0.000
	2	0.999	0.991	0.973	0.942	0.896	0.837	0.765	0.683	0.593	0.500	0.407	0.317	0.235	0.163	0.104	0.058	0.027	0.009	0.001
	3	1.000	1.000	0.998	0.993	0.984	0.969	0.946	0.913	0.869	0.812	0.744	0.663	0.572	0.472	0.367	0.263	0.165	0.081	0.023
	4			1.000	1.000	0.999	0.998	0.995	0.990	0.982	0.969	0.950	0.922	0.884	0.832	0.763	0.672	0.556	0.410	0.226
	5				1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6	0	0.735	0.531	0.377	0.262	0.178	0.118	0.075	0.047	0.028	0.016	0.008	0.004	0.002	0.001	0.000	0.000	0.000	0.000	0.000
	1	0.967	0.886	0.776	0.655	0.534	0.420	0.319	0.233	0.164	0.109	0.069	0.041	0.022	0.011	0.004	0.002	0.000	0.000	0.000
	2	0.998	0.984	0.953	0.901	0.831	0.744	0.647	0.544	0.442	0.344	0.255	0.179	0.117	0.070	0.038	0.017	0.006	0.001	0.000
	3	1.000	0.999	0.994	0.983	0.962	0.930	0.883	0.821	0.745	0.656	0.558	0.456	0.353	0.256	0.169	0.099	0.047	0.016	0.002
	4		1.000	1.000	0.998	0.995	0.989	0.978	0.959	0.931	0.891	0.836	0.767	0.681	0.580	0.466	0.345	0.224	0.114	0.033
	5				1.000	1.000	0.999	0.998	0.996	0.992	0.984	0.972	0.953	0.925	0.882	0.822	0.738	0.623	0.469	0.265
	6						1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Statistiske tabeller

		$P(X \leq x)$ i binomialfordelingen $b(n, p)$																		
		p																		
n	x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
7	0	0.698	0.478	0.321	0.210	0.133	0.082	0.049	0.028	0.015	0.008	0.004	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.956	0.850	0.717	0.577	0.445	0.329	0.234	0.159	0.102	0.062	0.036	0.019	0.009	0.004	0.001	0.000	0.000	0.000	0.000
	2	0.996	0.974	0.926	0.852	0.756	0.647	0.532	0.420	0.316	0.227	0.156	0.096	0.056	0.029	0.013	0.005	0.001	0.000	0.000
	3	1.000	0.997	0.988	0.967	0.929	0.874	0.800	0.710	0.608	0.500	0.392	0.290	0.200	0.126	0.071	0.033	0.012	0.003	0.000
	4		1.000	0.999	0.995	0.987	0.971	0.944	0.904	0.847	0.773	0.684	0.580	0.468	0.353	0.244	0.148	0.074	0.026	0.004
	5			1.000	1.000	0.999	0.996	0.991	0.981	0.964	0.938	0.898	0.841	0.766	0.671	0.555	0.423	0.283	0.150	0.044
	6					1.000	1.000	0.999	0.998	0.996	0.992	0.985	0.972	0.951	0.918	0.867	0.790	0.679	0.522	0.302
	7							1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
8	0	0.663	0.430	0.272	0.168	0.100	0.058	0.032	0.017	0.008	0.004	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.943	0.813	0.657	0.503	0.367	0.255	0.169	0.106	0.063	0.035	0.018	0.009	0.004	0.001	0.000	0.000	0.000	0.000	0.000
	2	0.994	0.962	0.895	0.797	0.679	0.552	0.428	0.315	0.220	0.145	0.088	0.050	0.025	0.011	0.004	0.001	0.000	0.000	0.000
	3	1.000	0.995	0.979	0.944	0.886	0.806	0.706	0.594	0.477	0.363	0.260	0.174	0.106	0.058	0.027	0.010	0.003	0.000	0.000
	4		1.000	0.997	0.990	0.973	0.942	0.894	0.826	0.740	0.637	0.523	0.406	0.294	0.194	0.114	0.056	0.021	0.005	0.000
	5			1.000	0.999	0.996	0.989	0.975	0.950	0.912	0.855	0.780	0.685	0.572	0.448	0.321	0.203	0.105	0.038	0.006
	6				1.000	1.000	0.999	0.996	0.991	0.982	0.965	0.937	0.894	0.831	0.745	0.633	0.497	0.343	0.187	0.057
	7					1.000	1.000	0.999	0.998	0.996	0.992	0.983	0.968	0.942	0.901	0.832	0.728	0.570	0.337	
	8							1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
9	0	0.630	0.387	0.232	0.134	0.075	0.040	0.021	0.010	0.005	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.929	0.775	0.599	0.436	0.300	0.196	0.121	0.071	0.039	0.020	0.009	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.992	0.947	0.859	0.738	0.601	0.463	0.337	0.232	0.150	0.090	0.050	0.025	0.011	0.004	0.001	0.000	0.000	0.000	0.000
	3	0.999	0.992	0.966	0.914	0.834	0.730	0.609	0.483	0.361	0.254	0.166	0.099	0.054	0.025	0.010	0.003	0.000	0.000	0.000
	4	1.000	0.999	0.994	0.980	0.951	0.901	0.828	0.733	0.621	0.500	0.379	0.267	0.172	0.099	0.049	0.020	0.006	0.001	0.000
	5		1.000	0.999	0.997	0.990	0.975	0.946	0.901	0.834	0.746	0.639	0.517	0.391	0.270	0.166	0.086	0.034	0.008	0.001
	6			1.000	1.000	0.999	0.996	0.989	0.975	0.950	0.910	0.850	0.768	0.663	0.537	0.399	0.262	0.141	0.053	0.008
	7				1.000	1.000	0.999	0.996	0.991	0.980	0.961	0.929	0.879	0.804	0.700	0.564	0.401	0.225	0.071	
	8					1.000	1.000	0.999	0.998	0.995	0.990	0.979	0.960	0.925	0.866	0.768	0.613	0.370		
	9							1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Statistiske tabeller

		$P(X \leq x)$ i binomialfordelingen $b(n, p)$																			
		p																			
n	x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	
10	0	0.599	0.349	0.197	0.107	0.056	0.028	0.013	0.006	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	1	0.914	0.736	0.544	0.376	0.244	0.149	0.086	0.046	0.023	0.011	0.005	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.988	0.930	0.820	0.678	0.526	0.383	0.262	0.167	0.100	0.055	0.027	0.012	0.005	0.002	0.000	0.000	0.000	0.000	0.000	
	3	0.999	0.987	0.950	0.879	0.776	0.650	0.514	0.382	0.266	0.172	0.102	0.055	0.026	0.011	0.004	0.000	0.000	0.000	0.000	
	4	1.000	0.998	0.990	0.967	0.922	0.850	0.751	0.633	0.504	0.377	0.262	0.166	0.095	0.047	0.020	0.006	0.001	0.000	0.000	
	5		1.000	0.999	0.994	0.980	0.953	0.905	0.834	0.738	0.623	0.496	0.367	0.249	0.150	0.078	0.033	0.010	0.002	0.000	
	6			1.000	0.999	0.996	0.989	0.974	0.945	0.898	0.828	0.734	0.618	0.486	0.350	0.224	0.121	0.050	0.013	0.001	
	7				1.000	1.000	0.998	0.995	0.988	0.973	0.945	0.900	0.833	0.738	0.617	0.474	0.322	0.180	0.070	0.012	
	8					1.000	0.999	0.998	0.995	0.989	0.977	0.954	0.914	0.851	0.756	0.624	0.456	0.264	0.086		
	9						1.000	1.000	1.000	0.999	0.997	0.994	0.987	0.972	0.944	0.893	0.803	0.651	0.401		
10									1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		
11	0	0.569	0.314	0.167	0.086	0.042	0.020	0.009	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	1	0.898	0.697	0.492	0.322	0.197	0.113	0.061	0.030	0.014	0.006	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.985	0.910	0.779	0.617	0.455	0.313	0.200	0.119	0.065	0.033	0.015	0.006	0.002	0.001	0.000	0.000	0.000	0.000	0.000	
	3	0.998	0.981	0.931	0.839	0.713	0.570	0.426	0.296	0.191	0.113	0.061	0.029	0.012	0.004	0.001	0.000	0.000	0.000	0.000	
	4	1.000	0.997	0.984	0.950	0.885	0.790	0.668	0.533	0.397	0.274	0.174	0.099	0.050	0.022	0.008	0.002	0.000	0.000	0.000	
	5		1.000	0.997	0.988	0.966	0.922	0.851	0.753	0.633	0.500	0.367	0.247	0.149	0.078	0.034	0.012	0.003	0.000	0.000	
	6			1.000	0.998	0.992	0.978	0.950	0.901	0.826	0.726	0.603	0.467	0.332	0.210	0.115	0.050	0.016	0.003	0.000	
	7				1.000	0.999	0.996	0.988	0.971	0.939	0.887	0.809	0.704	0.574	0.430	0.287	0.161	0.069	0.019	0.002	
	8					1.000	0.999	0.998	0.994	0.985	0.967	0.935	0.881	0.800	0.687	0.545	0.383	0.221	0.090	0.015	
	9						1.000	1.000	0.999	0.998	0.994	0.986	0.970	0.939	0.887	0.803	0.678	0.508	0.303	0.102	
	10							1.000	1.000	1.000	0.999	0.996	0.991	0.980	0.958	0.914	0.833	0.686	0.431		
	11										1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	

Statistiske tabeller

		$P(X \leq x)$ i binomialfordelingen $b(n, p)$																			
		p																			
n	x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	
12	0	0.540	0.282	0.142	0.069	0.032	0.014	0.006	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	1	0.882	0.659	0.443	0.275	0.158	0.085	0.042	0.020	0.008	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.980	0.889	0.736	0.558	0.391	0.253	0.151	0.083	0.042	0.019	0.008	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000	
	3	0.998	0.974	0.908	0.795	0.649	0.493	0.347	0.225	0.134	0.073	0.036	0.015	0.006	0.002	0.000	0.000	0.000	0.000	0.000	
	4	1.000	0.996	0.976	0.927	0.842	0.724	0.583	0.438	0.304	0.194	0.112	0.057	0.026	0.009	0.003	0.000	0.000	0.000	0.000	
	5		0.999	0.995	0.981	0.946	0.882	0.787	0.665	0.527	0.387	0.261	0.158	0.085	0.039	0.015	0.004	0.001	0.000	0.000	
	6		1.000	0.999	0.996	0.986	0.961	0.915	0.842	0.739	0.613	0.473	0.335	0.213	0.118	0.054	0.019	0.005	0.001	0.000	
	7			1.000	0.999	0.997	0.991	0.974	0.943	0.888	0.806	0.696	0.562	0.417	0.276	0.158	0.073	0.024	0.004	0.001	
	8				1.000	1.000	0.998	0.994	0.985	0.964	0.927	0.865	0.775	0.653	0.507	0.351	0.205	0.092	0.026	0.002	
	9					1.000	0.999	0.997	0.992	0.981	0.958	0.917	0.849	0.747	0.609	0.442	0.264	0.111	0.020		
	10						1.000	1.000	0.999	0.997	0.992	0.980	0.958	0.915	0.842	0.725	0.557	0.341	0.118		
	11							1.000	1.000	0.999	0.998	0.994	0.986	0.968	0.931	0.858	0.718	0.460			
12											1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		
13	0	0.513	0.254	0.121	0.055	0.024	0.010	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	1	0.865	0.621	0.398	0.234	0.127	0.064	0.030	0.013	0.005	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.975	0.866	0.692	0.502	0.333	0.202	0.113	0.058	0.027	0.011	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	3	0.997	0.966	0.882	0.747	0.584	0.421	0.278	0.169	0.093	0.046	0.020	0.008	0.003	0.001	0.000	0.000	0.000	0.000	0.000	
	4	1.000	0.994	0.966	0.901	0.794	0.654	0.501	0.353	0.228	0.133	0.070	0.032	0.013	0.004	0.001	0.000	0.000	0.000	0.000	
	5		0.999	0.992	0.970	0.920	0.835	0.716	0.574	0.427	0.291	0.179	0.098	0.046	0.018	0.006	0.001	0.000	0.000	0.000	
	6		1.000	0.999	0.993	0.976	0.938	0.871	0.771	0.644	0.500	0.356	0.229	0.129	0.062	0.024	0.007	0.001	0.000	0.000	
	7			1.000	0.999	0.994	0.982	0.954	0.902	0.821	0.709	0.573	0.426	0.284	0.165	0.080	0.030	0.008	0.001	0.000	
	8				1.000	0.999	0.996	0.987	0.968	0.930	0.867	0.772	0.647	0.499	0.346	0.206	0.099	0.034	0.006	0.000	
	9					1.000	0.999	0.997	0.992	0.980	0.954	0.907	0.831	0.722	0.579	0.416	0.253	0.118	0.034	0.003	
	10						1.000	1.000	0.999	0.996	0.989	0.973	0.942	0.887	0.798	0.667	0.498	0.308	0.134	0.025	
	11							1.000	0.999	0.998	0.995	0.987	0.970	0.936	0.873	0.766	0.602	0.379	0.135		
	12								1.000	1.000	1.000	0.999	0.996	0.990	0.976	0.945	0.879	0.746	0.487		
	13												1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	

Statistiske tabeller

		$P(X \leq x)$ i binomialfordelingen $b(n, p)$																			
		p																			
n	x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	
14	0	0.488	0.229	0.103	0.044	0.018	0.007	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	1	0.847	0.585	0.357	0.198	0.101	0.047	0.021	0.008	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.970	0.842	0.648	0.448	0.281	0.161	0.084	0.040	0.017	0.006	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	3	0.996	0.956	0.853	0.698	0.521	0.355	0.220	0.124	0.063	0.029	0.011	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000	
	4	1.000	0.991	0.953	0.870	0.742	0.584	0.423	0.279	0.167	0.090	0.043	0.018	0.006	0.002	0.000	0.000	0.000	0.000	0.000	
	5		0.999	0.988	0.956	0.888	0.781	0.641	0.486	0.337	0.212	0.119	0.058	0.024	0.008	0.002	0.000	0.000	0.000	0.000	
	6		1.000	0.998	0.988	0.962	0.907	0.816	0.692	0.546	0.395	0.259	0.150	0.075	0.031	0.010	0.002	0.000	0.000	0.000	
	7			1.000	0.998	0.990	0.969	0.925	0.850	0.741	0.605	0.454	0.308	0.184	0.093	0.038	0.012	0.002	0.000	0.000	
	8				1.000	0.998	0.992	0.976	0.942	0.881	0.788	0.663	0.514	0.359	0.219	0.112	0.044	0.012	0.001	0.000	
	9					1.000	0.998	0.994	0.982	0.957	0.910	0.833	0.721	0.577	0.416	0.258	0.130	0.047	0.009	0.000	
	10						1.000	0.999	0.996	0.989	0.971	0.937	0.876	0.780	0.645	0.479	0.302	0.147	0.044	0.004	
	11							1.000	0.999	0.998	0.994	0.983	0.960	0.916	0.839	0.719	0.552	0.352	0.158	0.030	
	12								1.000	1.000	0.999	0.997	0.992	0.979	0.953	0.899	0.802	0.643	0.415	0.153	
	13									1.000	1.000	0.999	0.998	0.993	0.982	0.956	0.897	0.771	0.512		
14												1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		
15	0	0.463	0.206	0.087	0.035	0.013	0.005	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	1	0.829	0.549	0.319	0.167	0.080	0.035	0.014	0.005	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.964	0.816	0.604	0.398	0.236	0.127	0.062	0.027	0.011	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	3	0.995	0.944	0.823	0.648	0.461	0.297	0.173	0.091	0.042	0.018	0.006	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	4	0.999	0.987	0.938	0.836	0.686	0.515	0.352	0.217	0.120	0.059	0.025	0.009	0.003	0.000	0.000	0.000	0.000	0.000	0.000	
	5	1.000	0.998	0.983	0.939	0.852	0.722	0.564	0.403	0.261	0.151	0.077	0.034	0.012	0.004	0.001	0.000	0.000	0.000	0.000	
	6		1.000	0.996	0.982	0.943	0.869	0.755	0.610	0.452	0.305	0.182	0.095	0.042	0.015	0.004	0.001	0.000	0.000	0.000	
	7			0.999	0.996	0.983	0.950	0.887	0.787	0.654	0.500	0.346	0.213	0.113	0.050	0.017	0.004	0.001	0.000	0.000	
	8				1.000	0.999	0.996	0.985	0.958	0.905	0.818	0.696	0.548	0.390	0.245	0.131	0.057	0.018	0.004	0.000	
	9					1.000	0.999	0.996	0.988	0.966	0.923	0.849	0.739	0.597	0.436	0.278	0.148	0.061	0.017	0.002	
	10						1.000	0.999	0.997	0.991	0.975	0.941	0.880	0.783	0.648	0.485	0.314	0.164	0.062	0.013	
	11							1.000	1.000	0.998	0.994	0.982	0.958	0.909	0.827	0.703	0.539	0.352	0.177	0.056	
	12								1.000	0.999	0.996	0.989	0.973	0.938	0.873	0.764	0.602	0.396	0.184	0.036	
	13									1.000	1.000	0.998	0.995	0.986	0.965	0.920	0.833	0.681	0.451	0.171	
	14										1.000	1.000	0.998	0.995	0.987	0.965	0.913	0.794	0.537		
	15													1.000	1.000	1.000	1.000	1.000	1.000	1.000	

Tabel 7 Kumulerede sandsynligheder i Poissonfordelingen $p(\mu)$. $P(X \leq x)$

$P(X \leq x)$ i Poissonfordelingen $p(\mu)$																					
x	μ	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7
0		0.951	0.905	0.861	0.819	0.779	0.741	0.670	0.607	0.549	0.497	0.449	0.407	0,368	0.333	0.301	0.273	0.247	0.223	0.202	0.183
1		0.999	0.995	0.990	0.982	0.974	0.963	0.938	0.910	0.878	0.844	0.809	0.772	0.736	0.699	0.663	0.627	0.592	0.558	0.525	0.493
2		1.000	1.000	0.999	0.999	0.998	0.996	0.992	0.986	0.977	0.966	0.953	0.937	0.920	0.900	0.879	0.857	0.833	0.809	0.783	0.757
3				1.000	1.000	1.000	1.000	0.999	0.998	0.997	0.994	0.991	0.987	0.981	0.974	0.966	0.957	0.946	0.934	0.921	0.907
4								1.000	1.000	1.000	0.999	0.999	0.998	0.996	0.995	0.992	0.989	0.986	0.981	0.976	0.970
5											1.000	1.000	1.000	0.999	0.999	0.999	0.998	0.997	0.996	0.994	0.992
6														1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.998
7																		1.000	1.000	1.000	1.000
-																					

$P(X \leq x)$ i Poissonfordelingen $p(\mu)$																					
x	μ	1.8	1.9	2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7
0		0.165	0.150	0.135	0.123	0.111	0.100	0.091	0.082	0.074	0.067	0.061	0.055	0.050	0.045	0.041	0.037	0.033	0.030	0.027	0.025
1		0.463	0.434	0.406	0.380	0.355	0.331	0.308	0.287	0.267	0.249	0.231	0.215	0.199	0.185	0.171	0.159	0.147	0.136	0.126	0.116
2		0.731	0.704	0.677	0.650	0.623	0.596	0.570	0.544	0.518	0.494	0.469	0.446	0.423	0.401	0.380	0.359	0.340	0.321	0.303	0.285
3		0.891	0.875	0.857	0.839	0.819	0.799	0.779	0.758	0.736	0.714	0.692	0.670	0.647	0.625	0.603	0.580	0.558	0.537	0.515	0.494
4		0.964	0.956	0.947	0.938	0.928	0.916	0.904	0.891	0.877	0.863	0.848	0.832	0.815	0.798	0.781	0.763	0.744	0.725	0.706	0.687
5		0.990	0.987	0.983	0.980	0.975	0.970	0.964	0.958	0.951	0.943	0.935	0.926	0.916	0.906	0.895	0.883	0.871	0.858	0.844	0.830
6		0.997	0.997	0.995	0.994	0.993	0.991	0.988	0.986	0.983	0.979	0.976	0.971	0.966	0.961	0.955	0.949	0.942	0.935	0.927	0.918
7		0.999	0.999	0.999	0.999	0.998	0.997	0.997	0.996	0.995	0.993	0.992	0.990	0.988	0.986	0.983	0.980	0.977	0.973	0.969	0.965
8		1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999	0.998	0.998	0.997	0.996	0.995	0.994	0.993	0.992	0.990	0.988	0.986
9							1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999	0.999	0.998	0.998	0.997	0.997	0.996	0.995
10											1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999	0.998
11																	1.000	1.000	1.000	1.000	1.000
-																					

Statistiske tabeller

$P(X \leq x)$ i Poissonfordelingen $p(\mu)$																					
x	μ	3.8	3.9	4.0	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0	5.1	5.2	5.3	5.4	5.5	5.6	5.7
0		0.022	0.020	0.018	0.017	0.015	0.014	0.012	0.011	0.010	0.009	0.008	0.007	0.007	0.006	0.006	0.005	0.004	0.004	0.004	0.003
1		0.107	0.099	0.092	0.085	0.078	0.072	0.066	0.061	0.056	0.052	0.048	0.044	0.040	0.037	0.034	0.031	0.029	0.027	0.024	0.022
2		0.269	0.253	0.238	0.224	0.210	0.197	0.185	0.174	0.163	0.152	0.143	0.133	0.125	0.116	0.109	0.102	0.095	0.088	0.082	0.077
3		0.473	0.453	0.433	0.414	0.395	0.377	0.359	0.342	0.326	0.310	0.294	0.279	0.265	0.251	0.238	0.225	0.213	0.202	0.191	0.180
4		0.668	0.648	0.629	0.609	0.590	0.570	0.551	0.532	0.513	0.495	0.476	0.458	0.440	0.423	0.406	0.390	0.373	0.358	0.342	0.327
5		0.816	0.801	0.785	0.769	0.753	0.737	0.720	0.703	0.686	0.668	0.651	0.634	0.616	0.598	0.581	0.563	0.546	0.529	0.512	0.495
6		0.909	0.899	0.889	0.879	0.867	0.856	0.844	0.831	0.818	0.805	0.791	0.777	0.762	0.747	0.732	0.717	0.702	0.686	0.670	0.654
7		0.960	0.955	0.949	0.943	0.936	0.929	0.921	0.913	0.905	0.896	0.887	0.877	0.867	0.856	0.845	0.833	0.822	0.809	0.797	0.784
8		0.984	0.981	0.979	0.976	0.972	0.968	0.964	0.960	0.955	0.950	0.944	0.938	0.932	0.925	0.918	0.911	0.903	0.894	0.886	0.877
9		0.994	0.993	0.992	0.990	0.989	0.987	0.985	0.983	0.980	0.978	0.975	0.972	0.968	0.964	0.960	0.956	0.951	0.946	0.941	0.935
10		0.998	0.998	0.997	0.997	0.996	0.995	0.994	0.993	0.992	0.991	0.990	0.988	0.986	0.984	0.982	0.980	0.977	0.975	0.972	0.969
11		0.999	0.999	0.999	0.999	0.999	0.998	0.998	0.998	0.997	0.997	0.996	0.995	0.995	0.994	0.993	0.992	0.990	0.989	0.988	0.986
12		1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.998	0.998	0.997	0.997	0.996	0.996	0.995	0.994
13							1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.998	0.998
14													1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999
15																			1.000	1.000	1.000
-																					

Statistiske tabeller

$P(X \leq x)$ i Poissonfordelingen $p(\mu)$																					
x	μ	5.8	5.9	6.0	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	7.0	7.1	7.2	7.3	7.4	7.5	7.6	7.7
0		0.003	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.000
1		0.021	0.019	0.017	0.016	0.015	0.013	0.012	0.011	0.010	0.009	0.009	0.008	0.007	0.007	0.006	0.006	0.005	0.005	0.004	0.004
2		0.072	0.067	0.062	0.058	0.054	0.050	0.046	0.043	0.040	0.037	0.034	0.032	0.030	0.027	0.025	0.024	0.022	0.020	0.019	0.017
3		0.170	0.160	0.151	0.143	0.134	0.126	0.119	0.112	0.105	0.099	0.093	0.087	0.082	0.077	0.072	0.067	0.063	0.059	0.055	0.052
4		0.313	0.299	0.285	0.272	0.259	0.247	0.235	0.224	0.213	0.202	0.192	0.182	0.173	0.164	0.156	0.147	0.140	0.132	0.125	0.118
5		0.478	0.462	0.446	0.430	0.414	0.399	0.384	0.369	0.355	0.341	0.327	0.314	0.301	0.288	0.276	0.264	0.253	0.241	0.231	0.220
6		0.638	0.622	0.606	0.590	0.574	0.558	0.542	0.527	0.511	0.495	0.480	0.465	0.450	0.435	0.420	0.406	0.392	0.378	0.365	0.351
7		0.771	0.758	0.744	0.730	0.716	0.702	0.687	0.673	0.658	0.643	0.628	0.614	0.599	0.584	0.569	0.554	0.539	0.525	0.510	0.496
8		0.867	0.857	0.847	0.837	0.826	0.815	0.803	0.792	0.780	0.767	0.755	0.742	0.729	0.716	0.703	0.689	0.676	0.662	0.648	0.634
9		0.929	0.923	0.916	0.909	0.902	0.894	0.886	0.877	0.869	0.860	0.850	0.840	0.830	0.820	0.810	0.799	0.788	0.776	0.765	0.753
10		0.965	0.961	0.957	0.953	0.949	0.944	0.939	0.933	0.927	0.921	0.915	0.908	0.901	0.894	0.887	0.879	0.871	0.862	0.854	0.845
11		0.984	0.982	0.980	0.978	0.975	0.972	0.969	0.966	0.963	0.959	0.955	0.951	0.947	0.942	0.937	0.932	0.926	0.921	0.915	0.909
12		0.993	0.992	0.991	0.990	0.989	0.987	0.986	0.984	0.982	0.980	0.978	0.976	0.973	0.970	0.967	0.964	0.961	0.957	0.954	0.950
13		0.997	0.997	0.996	0.996	0.995	0.995	0.994	0.993	0.992	0.991	0.990	0.989	0.987	0.986	0.984	0.982	0.980	0.978	0.976	0.974
14		0.999	0.999	0.999	0.998	0.998	0.998	0.997	0.997	0.997	0.996	0.996	0.995	0.994	0.994	0.993	0.992	0.991	0.990	0.989	0.987
15		1.000	1.000	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.998	0.998	0.998	0.997	0.997	0.996	0.996	0.995	0.995	0.994
16				1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.998	0.998	0.997
17										1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999	0.999
18																	1.000	1.000	1.000	1.000	1.000
-																					

Statistiske tabeller

$P(X \leq x)$ i Poissonfordelingen $p(\mu)$																					
x	μ	7.8	7.9	8.0	8.1	8.2	8.3	8.4	8.5	8.6	8.7	8.8	8.9	9.0	9.1	9.2	9.3	9.4	9.5	9.6	9.7
0		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1		0.004	0.003	0.003	0.003	0.003	0.002	0.002	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
2		0.016	0.015	0.014	0.013	0.012	0.011	0.010	0.009	0.009	0.008	0.007	0.007	0.006	0.006	0.005	0.005	0.005	0.004	0.004	0.004
3		0.048	0.045	0.042	0.040	0.037	0.035	0.032	0.030	0.028	0.026	0.024	0.023	0.021	0.020	0.018	0.017	0.016	0.015	0.014	0.013
4		0.112	0.106	0.100	0.094	0.089	0.084	0.079	0.074	0.070	0.066	0.062	0.058	0.055	0.052	0.049	0.046	0.043	0.040	0.038	0.035
5		0.210	0.201	0.191	0.182	0.174	0.165	0.157	0.150	0.142	0.135	0.128	0.122	0.116	0.110	0.104	0.099	0.093	0.089	0.084	0.079
6		0.338	0.326	0.313	0.301	0.290	0.278	0.267	0.256	0.246	0.235	0.226	0.216	0.207	0.198	0.189	0.181	0.173	0.165	0.157	0.150
7		0.481	0.467	0.453	0.439	0.425	0.412	0.399	0.386	0.373	0.360	0.348	0.336	0.324	0.312	0.301	0.290	0.279	0.269	0.258	0.248
8		0.620	0.607	0.593	0.579	0.565	0.551	0.537	0.523	0.509	0.496	0.482	0.469	0.456	0.443	0.430	0.417	0.404	0.392	0.380	0.368
9		0.741	0.729	0.717	0.704	0.692	0.679	0.666	0.653	0.640	0.627	0.614	0.601	0.587	0.574	0.561	0.548	0.535	0.522	0.509	0.496
10		0.835	0.826	0.816	0.806	0.796	0.785	0.774	0.763	0.752	0.741	0.729	0.718	0.706	0.694	0.682	0.670	0.658	0.645	0.633	0.621
11		0.902	0.895	0.888	0.881	0.873	0.865	0.857	0.849	0.840	0.831	0.822	0.813	0.803	0.793	0.783	0.773	0.763	0.752	0.741	0.730
12		0.945	0.941	0.936	0.931	0.926	0.921	0.915	0.909	0.903	0.897	0.890	0.883	0.876	0.868	0.861	0.853	0.845	0.836	0.828	0.819
13		0.971	0.969	0.966	0.963	0.960	0.956	0.952	0.949	0.945	0.940	0.936	0.931	0.926	0.921	0.916	0.910	0.904	0.898	0.892	0.885
14		0.986	0.984	0.983	0.981	0.979	0.977	0.975	0.973	0.970	0.967	0.965	0.962	0.959	0.955	0.952	0.948	0.944	0.940	0.936	0.931
15		0.993	0.993	0.992	0.991	0.990	0.989	0.987	0.986	0.985	0.983	0.982	0.980	0.978	0.976	0.974	0.972	0.969	0.967	0.964	0.961
16		0.997	0.997	0.996	0.996	0.995	0.995	0.994	0.993	0.993	0.992	0.991	0.990	0.989	0.988	0.987	0.985	0.984	0.982	0.981	0.979
17		0.999	0.999	0.998	0.998	0.998	0.998	0.997	0.997	0.997	0.996	0.996	0.995	0.995	0.994	0.993	0.993	0.992	0.991	0.990	0.989
18		1.000	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.998	0.998	0.998	0.997	0.997	0.997	0.996	0.996	0.995	0.995
19			1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.998	0.998	0.998	0.998
20									1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999	0.999	0.999	0.999
21															1.000	1.000	1.000	1.000	1.000	1.000	1.000
-																					

Statistiske tabeller

$P(X \leq x)$ i Poissonfordelingen $p(\mu)$																					
x	μ	9.8	9.9	10.0	10.2	10.4	10.6	10.8	11.0	11.2	11.4	11.6	11.8	12.0	12.5	13.0	13.5	14.0	14.5	15.0	16.0
0		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
1		0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
2		0.003	0.003	0.003	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	
3		0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.005	0.004	0.004	0.003	0.003	0.002	0.002	0.001	0.000	0.000	0.000	0.000	
4		0.033	0.031	0.029	0.026	0.023	0.020	0.017	0.015	0.013	0.012	0.010	0.009	0.008	0.005	0.004	0.003	0.002	0.001	0.001	
5		0.075	0.071	0.067	0.060	0.053	0.048	0.042	0.038	0.033	0.029	0.026	0.023	0.020	0.015	0.011	0.008	0.006	0.004	0.003	
6		0.143	0.137	0.130	0.118	0.107	0.097	0.087	0.079	0.071	0.064	0.057	0.051	0.046	0.035	0.026	0.019	0.014	0.010	0.008	
7		0.239	0.229	0.220	0.203	0.186	0.171	0.157	0.143	0.131	0.119	0.108	0.099	0.090	0.070	0.054	0.041	0.032	0.024	0.018	
8		0.356	0.344	0.333	0.311	0.290	0.269	0.250	0.232	0.215	0.198	0.183	0.169	0.155	0.125	0.100	0.079	0.062	0.048	0.037	
9		0.483	0.471	0.458	0.433	0.409	0.385	0.363	0.341	0.319	0.299	0.279	0.260	0.242	0.201	0.166	0.135	0.109	0.088	0.070	
10		0.608	0.596	0.583	0.558	0.533	0.508	0.484	0.460	0.436	0.413	0.391	0.369	0.347	0.297	0.252	0.211	0.176	0.145	0.118	
11		0.719	0.708	0.697	0.674	0.650	0.627	0.603	0.579	0.555	0.532	0.508	0.485	0.462	0.406	0.353	0.304	0.260	0.220	0.185	
12		0.810	0.801	0.792	0.772	0.752	0.732	0.710	0.689	0.667	0.644	0.622	0.599	0.576	0.519	0.463	0.409	0.358	0.311	0.268	
13		0.879	0.872	0.864	0.849	0.834	0.817	0.799	0.781	0.762	0.743	0.723	0.702	0.682	0.628	0.573	0.518	0.464	0.413	0.363	
14		0.927	0.922	0.917	0.906	0.894	0.882	0.868	0.854	0.839	0.823	0.807	0.790	0.772	0.725	0.675	0.623	0.570	0.518	0.466	
15		0.958	0.955	0.951	0.944	0.936	0.927	0.918	0.907	0.896	0.885	0.872	0.859	0.844	0.806	0.764	0.718	0.669	0.619	0.568	
16		0.977	0.975	0.973	0.968	0.963	0.957	0.951	0.944	0.936	0.928	0.919	0.909	0.899	0.869	0.835	0.798	0.756	0.711	0.664	
17		0.988	0.987	0.986	0.983	0.980	0.976	0.972	0.968	0.963	0.957	0.951	0.944	0.937	0.916	0.890	0.861	0.827	0.790	0.749	
18		0.994	0.993	0.993	0.991	0.989	0.987	0.985	0.982	0.979	0.976	0.972	0.967	0.963	0.948	0.930	0.908	0.883	0.853	0.819	
19		0.997	0.997	0.997	0.996	0.995	0.994	0.992	0.991	0.989	0.987	0.984	0.982	0.979	0.969	0.957	0.942	0.923	0.901	0.875	
20		0.999	0.999	0.998	0.998	0.997	0.997	0.996	0.995	0.994	0.993	0.992	0.990	0.988	0.983	0.975	0.965	0.952	0.936	0.917	
21		0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.998	0.997	0.997	0.996	0.995	0.994	0.991	0.986	0.980	0.971	0.960	0.947	
22		1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999	0.999	0.998	0.998	0.998	0.997	0.995	0.992	0.989	0.983	0.976	0.967	
23						1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999	0.999	0.998	0.996	0.994	0.991	0.986	0.981	
24										1.000	1.000	1.000	0.999	0.999	0.999	0.998	0.997	0.995	0.992	0.989	
25													1.000	1.000	0.999	0.999	0.998	0.997	0.996	0.994	
26															1.000	1.000	0.999	0.999	0.998	0.997	
27																	1.000	0.999	0.999	0.998	
28																		1.000	0.999	0.998	
29																			1.000	0.999	
30																				0.999	
31																				1.000	

TABEL 8. Dimensioneringstabel.

Dimensionering med henblik på udførelse af **t-test vedrørende 1 statistisk variabel X** som er normalfordelt $n(\mu, \sigma)$, hvor μ og σ er ukendte. (kopi af Statistik II side 689)

Størrelse n af stikprøve på X :	Værdier af $\frac{\Delta}{\sigma}$ for $\alpha = 5\%$.			
	Ensided test $H_0: \mu \leq \mu_0$ eller $H_0: \mu \geq \mu_0$		Tosidet test $H_0: \mu = \mu_0$	
	$\beta = 5\%$	$\beta = 10\%$	$\beta = 5\%$	$\beta = 10\%$
3	3.19	2.78	4.49	3.92
4	2.23	1.97	2.84	2.51
5	1.82	1.61	2.20	1.97
6	1.58	1.40	1.87	1.67
7	1.42	1.26	1.65	1.47
8	1.30	1.15	1.49	1.33
9	1.21	1.07	1.38	1.24
10	1.13	1.01	1.29	1.16
11	1.07	0.95	1.21	1.09
12	1.02	0.90	1.14	1.02
13	0.97	0.86	1.09	0.98
14	0.93	0.83	1.04	0.94
15	0.89	0.80	1.00	0.90
16	0.86	0.77	0.97	0.86
17	0.83	0.74	0.94	0.83
18	0.81	0.72	0.91	0.81
19	0.78	0.70	0.88	0.79
20	0.76	0.68	0.86	0.77
22	0.72	0.65	0.82	0.73
24	0.69	0.62	0.78	0.70
26	0.66	0.59	0.75	0.67
28	0.64	0.57	0.72	0.64
30	0.62	0.55	0.69	0.61
40	0.53	0.47	0.59	0.53
60	0.43	0.38	0.48	0.43

FACITLISTE

KAPITEL 1

1.1	(1) f	(2) -	(3) 1.7	9.41	3.068				
1.2	(1) -	(2) -	(3) - 0.9167	2.743	1.656				
1.3	(1) -	(2) -	(3) -1	4.67	2.16				
1.4	(1) -	(2) -	(3) 8.833	46.31	6.805				
1.5	(1) -	(2) -	(3) 0.833	0.3056	0.5528	(4) 0.1641	(5) 0.764	(6) 0.89	0.674
1.6	(1) -	(2) -	(3) 1.5	0.75	0.866	(4) 0.088			
1.7	(1) -	(2) 6	5.33	2.31	(3) 0.375	(4) 4	(5) 4.956	2.173	
1.8	(1) -	(2) 5	1.667	1.291	(3) 0.196	(4) 5.278			
1.9	(1) -	(2) 69.94	15.91	(3) 70					
1.10	(1) -	(2) 34.8	1.36	(3) 33.8	34.7	35.6			
1.11		1658.76	57.31	3.46%					
1.12		103.32	2.6	2.52%					
1.13		996.3	46.36	4.65%					

KAPITEL 2

2.1	(1) 0.7734	0.0548	0.1718	(2) 0.7480					
2.2	(1) 69.15%	(2) 10.87%	(3) 112.2	(4) 117.3	6.535				
2.3	(1) 86.64%	(2) 0.008	(3) 0.020						
2.4	(1) 5.94%	(2) 27.71%	(3) [783.51; 816.49]						
2.5	(1) 9.5	1.265	(2) 12.45	(3) 2.41%					
2.6	(1) 92.8%								
2.7	(1) 97.71%	(2) 25.45							
2.8	(1) 65	0.4	(2) 77.34%						

KAPITEL 3

3.1	(1) 3.25	20.5	(2) 99%						
3.2	(1) -2.13	2.95	(3) 99%						
3.3	(1) 2.60	0.35							

KAPITEL 4

4.1a	(1) 12.13	0.6783	(2) [11.65 ; 12.61]	(3) [10.52 ; 13.74]	(4) [11.73 ; 12.53]				
4.1b	(5) [0.218 ; 1.534]	[0.467 ; 1.238]							
4.2a	(1) 2259.92	35.569	(2) [2237 ; 2283]	(3) [2178 ; 2341]					
4.2b	(4) [25.21 ; 60.36]								
4.3a	(1) 74.0362	0.00124	(2) [74.035; 74.037]	(3) [74.036 ; 74.037]					
4.3b	(4) [0.00083 ; 0.00231]								
4.4	(1) 750.2	(2) [740.1 ; 760.3]	(3) 19.13	(4) [14.0 ; 30.2]					
4.5	(1) 8.268	0.241	(2) [8.02 ; 8.52]	(3) [8.08 ; 8.46]					
4.6	[4.23 ; 4.29]	[0.028 ; 0.076]							
4.7	(1) [0.965 ; 1.111]	(2) [0.0263; 0.1714]							

KAPITEL 5

- 5.1a** (1) nej (2) 2.27% (3) 2.10 (4) [2.00; 2.20] (5) 142
5.1b 62.7% **5.1c** (1) nej $t = 2.0$ (2) ca 2,5% (3) 2.10 (4) 2.00;2.20] (5) 145
5.2a ja (2) 3.46% (3) 58.0 (4) [55.84; 60.16] (5) 260 **5.2b** 43.2% **5.2c** ja
5.3 (1) ja $t = -3.13^{**}$ (2) 12 nej
5.4 (1) ja $t = -6.78^{***}$ (2) ca 0.004% (3) 13.1 (4) [12.80 ; 13.40] (5) 12
5.5 (1) 84.47 6.85 (2) ja $t = 2.92^{**}$ (3) [81.27 ; 87.67] (4) 25.7%
5.6 $t = 1.72$ [7.93 ; 8.49]
5.7 ja $t = 3.76^{**}$ [24.07 ; 35.56]
5.8 (1) 24 (2) ja $t = -3.49^{**}$ (3) [2.59 ; 2.67]
5.9 (1) - (2) 14 (3) $\chi^2 = 33.3^{**}$ (4) $t = 2.03^*$ (5) [83.1; 85.9] [1.74 ; 3.87]
5.10 (1) $\chi^2 = 22.75$ accept (2) $\chi^2 = 24.10$ accept
5.11 (1) $\chi^2 = 89.79^*$
5.12 $\chi^2 = 2.19^*$ (2) $p \approx 0.012$ (3) 0.987 (4) [0.68 ; 1.80]
5.13 $\chi^2 = 55.5^{**}$
5.14 $\chi^2 = 46.7^{**}$
5.15 (1) nej $\chi^2 = 5.40$ (2) ja $u = 2.20^*$ (3) [76.6 ; 78.6]
5.16 (1) $\chi^2 = 7.16$ (2) $t = -2.40$ (3) [0.07 ; 0.24]
5.17 (1) 19.5 (2) $\chi^2 = 107.1$ $t = 3.08^{**}$
5.18 (1) nej $t = 1.09$
5.19 (1) $t = 1.426$ (2) [0.22 ; 7.28]
5.20 (1) - (2) $t = 2.268^*$

KAPITEL 6

- 6.1** (1) - (2) 0.1 0.8 0.2 0.7
6.2 (1) 0.9134 (2) 0.9678
6.3 (1) 8.75% (2) 38.75% (3) 41.25% (4) 11.25%
6.4 (1) 6.4% (2) 78.4% (3) 7.2%
6.5 (1) 37.5% 87.5% 33.3% 29.17% (2) 11.11%
6.6 (1) - 0.05 0.20 0.04 0.19 0.01 0.76 (2) - (3) 0.185
6.7 (1) 27.1% 36.0% 9.756% (2) 53.34% (3) 49.20%
6.8 (1) - (2) 5 3.75 4.082 4.437 (3) 41.67% (4) 12%

KAPITEL 7

- 7.1** (1) 0.435% (2) 49.57% (3) 41.30%
7.2 (1) 91.67% (2) 25.00% (3) 9.167%
7.3 (1) 17.68% (2) 59.28%
7.4 (1) 0.988%
7.5 (1) - 73.75 30.08 (2) 221.25 kr 57.48%
7.6 2.58%
7.7 (1) 2.83% (2) 70.04% (3) 64.97%
7.8 77.86%
7.9 5.83%
7.10 (1) 7.94% (2) 11.8%
7.11 40.33%

Facitliste

- 7.12** 8.72%
7.13 (1) 66.48% (2) 5 led
7.14 15.74%
7.15 (1) 0.108 (2) [0.089 ; 0.127] (3) 21.04
7.16 (1) 0.683 (2) [0.600 ; 0.767]
7.17 [0.799 ; 0.847]
7.18 [0.0048 ; 0.0202]
7.19 $u = -3.94^{***}$
7.20 (1) 0.9% (2) nej
7.21 (1) $u = -1.875^*$ (2) 0.030 (3) 0.12 (4) [0.056 ; 0.184]
7.22 (1) 30.1% (2) 87.9% (3) 4
7.23 50.37%
7.24 (1) 15 (2) 81.9%
7.25 75.3%
7.26 6.56%
7.27 76.3%
7.28 2.1%
7.29 26.5%
7.30 44.6%
7.31 (1a) 11 (1b) [4.5 ; 17.5] (2a) 1.1 (2b) [0.45 ; 1.75]
7.32 (1) 4.68 (2) [4.47 ; 4.89] (3) 69.44
7.33 nej $u = -1.10$
7.34 (1) nej $u = 1.43$ (2) 7.7% (3) 100.8 (4) [85.56 ; 116.04]
7.35 5.6%
7.36 94.9%
7.37 (1) 60.7% 91.4% 6.9% (2) 4.5%
7.38 13%
7.39 (1) 84 (2) 4.55%
7.40 (1) 9.68% (2) 15.12%
7.41 ja $p = 1.7\%$
7.42 (1) 0.539% (2) 0.119%
7.43 7.31%
7.44 0.188%

KAPITEL 8

- 8.1** (1) 2.90 (2) 14.6% (3) 16.98
8.2 (1) 77.88% (2) 10.45% (3) 77.88% (4) 9.48%
8.3 1
8.4 (1) 79.8% (2) 99.33% (3) 22.8 (4) 14

KAPITEL 9

- 9.1.1** (a1) - (a2) - (a3) nej (a4) 2.3 1.8 1.345 1.470 (a5) 12.6 (b1) 2.2 1.5 1.316
1.5811
- 9.1.2** (a6) 0.06 0.0304 (b2) 0.6667 0.3203
- 9.1.3** (a7) 52 15.52 (b3) 53.33 17.51
- 9.2.1** (a1) - (a2) nej (a3) 0.9 0.5 0.830 0.5 (a4) 0.9707 (b1) 0.6 0.6 0.843
0.516
- 9.2.2** (a5) 0.15 0.36 (b2) 0.2667 0.6124
- 9.2.3** (a6) 15300 3716.18 (b3) 15800 4131.18
- 9.3.1** (a1) - (a2) nej a3) 4.2 3.2 1.4 0.9798 (a4) 0.625 (b1) 4 3.2 1.633 1.033
- 9.3.2** (a5) 0.16 0.0342 (b2) 0.4444 0.2635
- 9.3.3** (a6) 316 44.36 (b3) 300 28.28
- 9.4.1** 178.33 71.833 11.428 10,870
- 9.4.2** 100.97 0.8128
- 9.4.3** 139.13 2.089
- 9.5** 4.6344
- 9.6** 3.559
- 9.7** 3.3466

STIKORDSREGISTER

A

absolut usikkerhed 18
acceptområde 54
additionssætning
 for sandsynligheder 77
 for linearkomb. af normalf. variable 34
alternativ hypotese 54
approksimation
 binomial til normalfordeling 99
 binomial til Poissonfordeling 98
 hypergeometrisk til binomialford. 98
 hypergeometrisk til binomialford. 98
 Poisson til normalfordeling 100

B

bagatelgrænse 57
betinget sandsynlighed 77
binomialfordeling 89,
 tabel 159
binomialfordelingstest 91
 oversigt 148
både A og B 76

C

centrale grænseværdisætning 33
chi i anden fordeling 38

D

differensmængde 76
dimensionering 57
dimensionerinstabel 169
disjunkte mængder 3, 77
diskret variabel 4, 85

E

eksperiment, tilfældigt 1
eksponentialfordeling 114
elementarhændelse 77
ensfordelte variable 16
ensidet
 binomialtest 91
 chi test 63
 Poisson test 95
 t-test 62
 test 54
enstjernet signifikans 54

enten A eller B 76

estimat 10, 44
estimator 44

F

fakultet 79
fejl af type I 57
fejl af type II 57
F-fordeling 41
flerdimensional variabel 122
fordeling
 binomial- 89
 chi i anden- 38
 eksponential- 114
 F-fordeling 41
 hypergeometrisk- 85
 kontinuert 6, 112
 logaritmisk normalfordeling 118
 rektangulær 112
 t-fordeling 40
 To-dimensional normalfordeling 119
 Weibull 117
fordelingsfunktion 6, 14, 124
foreningsmængde 76
forkastelsesområde 54
fraktil 14
fraktiltabel for
 normalfordeling 154
 chi i anden-fordeling 155
 t - fordeling 157
 F - fordeling 158
frihedsgrad 13, 123, 128
fællesmængde 76

G

Galton apparat 26
Gauss fordeling 26
generaliseret hypergeometrisk ford. 97
gennemsnit 10

H

heltalskorrektion 99
histogram 7
hypergeometrisk fordeling 85
hypotesetest 53, 91, 95
hyppighed, relativ 2

hændelse 2

A men ikke B 76

additionssætning 77

både A og B 76

den sikre 77

den umulige 77

differensmængde 76

disjunkte 77

enten A eller B 76

foreningsmængde 76

fællesmængde 76

ikke A 75

komplementærmængden 76

uafhængige 78

hypotesetest 53, 91, 95

hyppighed, relativ 2

hændelse 2

A men ikke B 76

additionssætning 77

både A og B 76

den sikre 77

den umulige 77

differensmængde 76

disjunkte 77

enten A eller B 76

foreningsmængde 76

fællesmængde 76

ikke A 75

komplementærmængden 76

uafhængige 78

I,J

ikke A 75

K

kombinatorik 79

komplementærmængde 75

konfidensinterval 45

kontinuert variabel 4, 6

korrelationskoefficient 122, 127

kovarians 122, 127

kumuleret binomialfordeling 101

kumuleret Poissonfordeling 101

kvadratsregul 17, 123, 131

kvartil 14

L

levetid 116

linearitetsregel 16, 123, 132

linearkombination 15, 132

logaritmisk normalfordeling 118

lommeregner

TI - 83 138

TI - 89 136

M

Maple 144

median 14

middelværdi 5, 9, 124

N

nedre kvartil 14

n fakultet 79

normalfordeling 26, 28

logaritmisk 118

normeret 30

tabel 152

test, oversigt 144-147

todimensional 119

nulhypotese 54

O

observationer, parvise 64

OC - kurve 59

ophobningslov 18

oversigt

binomialtest 148

konfidensintervaller 143

normalfordelingstest 144 -147

Poissonstest 149

P

parvise observationer 64

Poissonfordeling 93

tabel 164 - 169

Poissonfordelingstest 95, 149

polynomialfordeling 97

poolet estimat 122, 129

population 7, 11

produksætning 78

prædestinationsinterval 47

punkttestimation 44

punkttestimator 44

P-værdi 55

R

random experiment 1

Stikord

random variable 1
rektangulær fordeling 112
relativ hyppighed 2
relativ usikkerhed 18

S

SAK 12, 128
sandsynlighed 75
 additionssætning 77
 betinget 77
 produktsætning 78
SAP 129
signifikansniveau 55
sikre hændelse 77
spredning 5, 9, 28
SS 12
standard deviation 5, 9
statistisk uafhængige 2, 16, 78, 125
statistisk variabel 3
stikprøve 7, 17, 124
 gennemsnit 10, 29
 spredning 10, 29
 varians 10, 129
stikprøvestørrelse 57
stokastisk variabel 3
sumpolygon 15
symmetrisk sandsynlighedsfelt 79

T

tabeller over
 binomialfordeling 159 - 163
 dimensionering 169
 normalfordeling 152, 153
 Poissonfordeling 164
tabeller over fraktiler i
 normalfordeling 154
 chi i anden-fordeling 155
 t - fordeling 157
 F - fordeling 158
test
 ensidet 54, 62, 63, 91, 95
 fejl af type I 57
 fejl af type II 57
 P-værdi 55
 tosidet 54, 61
testfunktioner
 χ^2 - fordeling 38
 F - fordeling 41

t - fordelingen 40
t-fordeling 40
to-dimensional normalfordeling 119
tosidet test 54, 61
Ti-83 138
TI - 89 136
t-test, ensidet 62
tæthedsfunktion 4, 8, 124

U

uafhængige hændelser 78
uafhængige statistiske variable 2, 16, 78, 125
udfald 1
udfaldsrum 1
umulige hændelse 77
usikkerhed
 absolut 18
 relativ 18
usikkerhedsberegning 18
U-fordeling 30

V

variabel
 binomialfordelt 89
 diskret 4, 85
 kontinuert 4, 6
 statistisk 3
varians 5, 9, 129
variationsbredde 7

W

Weibullfordeling 117

Ø

øvre kvartil 14