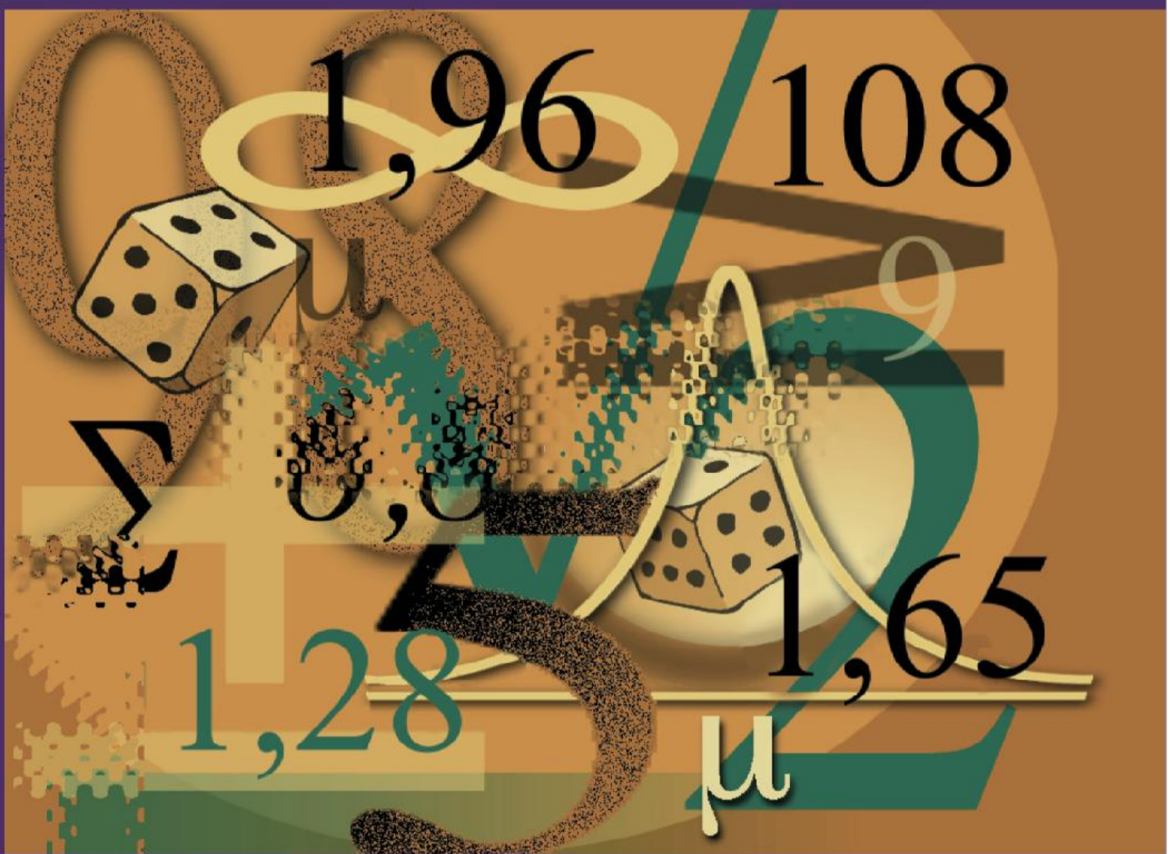
ALF HARBITZ

Statistikk og sannsynlighetsregning

– oppgaver og løsningsforslag



Forord

Denne oppgavesamling inneholder alle oppgavene fra læreboka «Statistikk og sannsynlighetsregning» (samme forfatter), og inneholder komplette løsningsforslag til alle oppgavene. Totalt er det ca. 200 oppgaver. Ca. 1/3 av disse er eksamensoppgaver i statistikk gitt ved 3-årig ingeniørutdanning ved ulike høgskoler i Norge. Eksamensoppgavene er merket med (E).

Oppgavene er av generell karakter og dekker det meste av klassiske emner innen innføringskurs i statistikk (se innholdsfortegnelse) og sannsynlighetsregning. Jevnt over ligger oppgavene på et typisk «innføringsnivå», og oppgavesamlingen forventes å ha interesse langt utover ingeniørstudenter.

Det er lagt vekt på å benytte det samme antall siffer som vises i mellomregningene, slik at man ved hjelp av tabeller fra læreboka og det samme antall siffer som er vist skal kunne få eksakt samme tallsvar. Dette vil i enkelte tilfelle føre til at siste siffer kan avvike noe fra de svar man får ved bruk av dataverktøy som f.eks. Minitab eller Matlab, da avrundingsfeilen ved bruk av dataverktøy som regel er neglisjerbar. I praksis er dette et lite problem. Forøvrig er fasitsvarene i læreboka beregnet med dataverktøy, så den interesserte kan studere avrundingseffekter ved å sammenligne svarene i lærebokas fasit med svarene i oppgavesamlingen.

Det er i praksis svært vanskelig å unngå feil, særlig i et førsteopplag av en bok. Forfatteren har imidlertid en god følelse av at feilraten er betryggende lav, etter at oppgavene har vært brukt og bearbeidet i en årrekke. Bjørn Davidsen ved Høgskolen i Tromsø, avdeling for ingeniør- og økonomifag, har vært en flittig hjelper ved at han har brukt de første utkastene til oppgavesamlingen overfor sine studenter. Den lave feilraten Davidsen og hans studenter har funnet er en god indikasjon på at det ikke er alarmerende mange feil i oppgavesamlingen.

Belønning på kr. 20 pr. feil som først oppdages utloves. Ikke nøl med å ta kontakt med forfatteren, også i tilfelle du har annet på hjertet tilknyttet oppgavesamlingen.

Tromsø, 5. september 2021

Alf Harbitz

mobiltlf: 93432615, E-mail: alf.harbitz@hi.no

Da Vårherre skapte verden og menneskene som skulle leve i den – et tiltak som ifølge moderne vitenskap tok svært langt tid – kan jeg forestille meg at Han resonnerte med seg selv som følger: «Hvis jeg lager alt forutsigelig, så vil disse menneskene, som jeg har utstyrt med ganske gode hjerner, utvilsomt lære å forutsi alt. De vil derfor ikke ha noen grunn til å gjøre noe som helst, fordi de vil oppdage at framtida er fullstendig forutbestemt, uten å kunne bli påvirket av noen menneskelig handling. På den andre siden, hvis jeg lager alt uforutsigelig, så vil de gradvis oppdage at alt vil skje tilfeldig uavhengig av menneskelig påvirkning. Som i det første tilfellet vil de ikke finne noen grunn til å gjøre noe som helst da heller. Ingen av de to oppskriftene vil være fornuftig. Jeg må derfor lage en blanding: La noe være forutsigelig, og la andre ting være uforutsigelig. Menneskene vil da, blant mye annet, få den svært så viktige oppgaven å finne ut av hva som er hva.»

E.F. Shcumacher[1]

[1] Fra boka *Small is beautiful*, fritt oversatt av Alf Harbitz

Symbolliste og forkortelser

Bino(n,p)	binomisk fordeling med parametre n og p
Corr(X,Y)	korrelasjonskoeffisient mellom X og Y
CV	variasjonskoeffisient
d.f.	«degrees of freedom», antall frihetsgrader
E($\cdot$)	forventnings-operator
expo(b)	eksponensial-fordeling med skalaparameter b
$F(m,n)$	F-fordeling med m og n frihetsgrader
f.eks.	for eksempel
gamma(b,c)	gammafordeling med skalaparameter b og formparameter c
h.k.	halvkorreksjon
hyp(n,D,N)	hypergeometrisk fordeling med parametre n, D og N
KI	konfidensintervall
Kji2(n)	kjikvadratfordeling med n frihetsgrader
$N(\mu,\sigma)$	normalfordeling med parametre μ og σ
$N_2(\mu_1,\sigma_1;\mu_2,\sigma_2;\rho)$	binormal fordeling med parametre μ_1, σ_1, μ_2, σ_2 og ρ
P^*	signifikanssansynlighet, også kalt p-verdi
Po(λ)	Poisson-fordeling med parameter λ
r	empirisk korrelasjonskoeffisient
Rayl(b)	Rayleigh-fordeling med skalaparameter b
S, s	empirisk standardavvik
SSE	feilkvadratsum
SS_T	behandlingskvadratsum
std($\cdot$)	standardavvik-operator
stud(n)	Students-t fordeling med n frihetsgrader
t_α	øvre α-fraktil i t-fordelingen
tn.	tilnærmet
U[a,b]	uniform fordeling definert på [a,b]
uif, u.i.f.	uavhengige og identisk fordelte
Var($\cdot$)	varians-operator
Weib(b,c)	Weibull-fordeling med skalaparameter b og formparameter c
$\bar{X}$, $\bar{x}$	Overstrekning betyr empirisk middelverdi
z_α	øvre α-fraktil i $N(0,1)$-fordelingen
μ	populasjons-forventning
σ	populasjons-standardavvik
ρ	korrelasjonskoeffisient
θ	Generelt symbol for ukjent parameter
Φ, ϕ	$N(0,1)$-funksjoner: Φ: kumulativ, ϕ: sannsynlighetstetthet
$\Gamma(x)$	Gammafunksjonen ($\Gamma(x+1)=x!$)

Innhold

Kapittel 1

Beskrivende statistikk

Oppgave 1.1 Beregn empirisk middelverdi og standardavvik av følgende tall:

a) −1, 3, 8, 4, 13
b) −8, −9, −14, −11, 0

Løsningsforslag:

a) $\sum_{i=1}^{5} x_i = -1+3+8+4+13 = 27$, middelverdi: $\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = 27/5 = \underline{5,4}$

$\sum_{i=1}^{5} x_i^2 = (-1)^2 + 3^2 + 8^2 + 4^2 + 13^2 = 1+9+64+16+169 = 259$

standardavvik: $s = \sqrt{\frac{1}{5-1}\left(\Sigma\, x_i^2 - \frac{1}{5}(\Sigma\, x_i)^2\right)} = \sqrt{\frac{1}{4}(259 - \frac{1}{5}\cdot 27^2)} = \frac{1}{2}\sqrt{113,2} = \underline{5,32}$

b) $\sum_{i=1}^{5} x_i = (-8)+(-9)+(-14)+(-11)+0 = -42$,
middelverdi: $\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = -42/5 = \underline{-8,4}$

standardavvik: $s = \sqrt{\frac{1}{5-1}\left(\Sigma\, x_i^2 - \frac{1}{5}(\Sigma\, x_i)^2\right)} = \sqrt{\frac{1}{4}(462 - \frac{1}{5}\cdot(-42)^2)} = \frac{1}{2}\sqrt{109,2} = \underline{5,22}$

Oppgave 1.2 Gitt følgende tallmateriale:

1,43438 1,43443 1,43442 1,4344

a) La tallene ovenfor være x-verdier, og finn de tilsvarende z-verdier ved transformasjonen

$z = (x-1,4344)/0,00001$

b) Beregn empirisk middelverdi, $\bar{z}$, og standardavvik, s_z, til z-verdiene.

c) Beregn middelverdi, $\bar{x}$, og standardavvik, s_x, til x-verdiene ved transformasjonen $s_x = 0,00001 s_z$ og

$\bar{x} = 1,4344 + 0,00001\cdot\bar{z}$

Løsningsforslag:

a) $z_1 = (x_1 - 1,4344)/0,00001 = (1,43438 - 1,4344)/0,00001 = \underline{-2}$

Tilsvarende finner vi $z_2 = \underline{3}$, $z_3 = \underline{2}$ og $z_4 = \underline{0}$

b) $\bar{z} = (-2+3+2+0)/4 = \underline{3/4}$, $\Sigma z^2 = 17$, $s_z = \sqrt{\frac{1}{3}(17 - 4\cdot(\frac{3}{4})^2)} = \sqrt{\frac{59}{12}} = \underline{2,22}$

c) $\bar{x} = 1,4344 + 0,00001\cdot 0,75 = \underline{1,4344075}$, $s_x = 0,00001\cdot 2,22 = \underline{0,0000222}$

Oppgave 1.3 Følg fremgangsmåten i oppgave 1.2 og beregn middelverdi og standardavvik til tallene

1221,3 1220,9 1221,7 1220,8 1221

Løsningsforslag: Vi søker først fornuftige verdier for a og b slik at transformasjonen $z = (x-a)/b$ fører til at $z_{max}-z_{min}$ verken blir for liten eller for stor (størrelsesorden lik 1). a er en såkalt lokaliseringsparameter som karakteriserer senter i tallmaterialet, og b er en såkalt skalaparameter som karakteriserer spredningen i tallmaterialet (rundt middelverdien). Vi velger skjønnsmessig $a = 1221$ og $b = 1$ og får:

$z_1 = (1221{,}3-1221)/1 = 0{,}3$, $z_2 = -0{,}1$, $z_3 = 0{,}7$, $z_4 = -0{,}2$ og $z_5 = 0$.

$$\bar{z} = (0{,}3-0{,}1+0{,}7-0{,}2+0)/5 = 0{,}7/5 = 0{,}14$$

$$\Sigma z^2 = 0{,}3^2 + (-0{,}1)^2 + 0{,}7^2 + (-0{,}2)^2 + 0^2 = 0{,}63,\ s_z = \sqrt{\tfrac{1}{4}(0{,}63-\tfrac{1}{5}\cdot 0{,}7^2)} = 0{,}365$$

$$\bar{x} = a + b\cdot\bar{z} = 1221 + 0{,}14 = \underline{1221{,}14},\ s_x = b\cdot s_z = \underline{0{,}365}$$

Oppgave 1.4 I en bedrift med 14 ansatte er gjennomsnittsinntekten kr 158 000 pr. ansatt. Hvor store lønnsutgifter har bedriften?

Løsningsforslag: $\frac{1}{14}\Sigma x = \text{kr } 158000 \Rightarrow \Sigma x = 158000\cdot 14 = \underline{\text{kr } 2212000}$

Oppgave 1.5

a) Finn empirisk middelverdi og median for følgende tilfeldig målte utetemperaturer (°C) som en person har notert i sin dagbok i løpet av ett år:

x: 14, −12, 0, 16, 7

b) En annen person har følgende måleresultater fra sin dagbok (samme sted og samme år):

y: 14, −12, 0, 16, 7, 112

Beregn empirisk middelverdi og median også i dette tilfellet. Ville du stole mest på middelverdien eller medianen?

Løsningsforslag:

a) $\bar{x} = (14-12+0+16+7)/5 = \underline{5\ ^\circ\text{C}}$. For å finne medianen ordner vi først x'ene i stigende rekkefølge: $x_{(1)} = -12, x_{(2)} = 0, x_{(3)} = 7, x_{(4)} = 14, x_{(5)} = 16$. Siden n =5 er et oddetall, blir medianen lik den midterste verdien:

$$m_x = x_{(\frac{5+1}{2})} = x_{(3)} = \underline{7\ ^\circ\text{C}}$$

b) $\bar{y} = (14-12+0+16+7+112)/6 = \underline{22{,}8\ ^\circ\text{C}}$.

For å finne medianen ordner vi første y'ene i stigende rekkefølge:

$y_{(1)} = -12, y_{(2)} = 0, y_{(3)} = 7, y_{(4)} = 14, y_{(5)} = 16, y_{(6)} = 112$

Siden $n = 6$ er et liketall, får vi to verdier på midten, og vi definerer medianen som middelverdien av disse: $m_y = \frac{1}{2}(y_{(6/2)} + y_{(6/2+1)}) = \frac{1}{2}(7+14) = \underline{10{,}5\ ^{o}\text{C}}$. Jeg ville stolt mer på medianen enn middelverdien, da målingen på 112 ^{o}C er tvilsom.

Oppgave 1.6 Gitt følgende x-verdier:

1, –3, 7, 12, 6

Beregn 20- og 70-prosentilen.

Løsningsforslag: $n = 5,\ x_{(1)} = -3,\ x_{(2)} = 1,\ x_{(3)} = 6,\ x_{(4)} = 7,\ x_{(5)} = 12$

20-prosentil: $np = 5 \cdot 0{,}2 = 1$ (heltall) $\Rightarrow x_{0,2} = \frac{1}{2}(x_{(1)} + x_{(2)}) = \frac{1}{2}(-3+1) = \underline{-1}$

70-prosentil: $np = 5 \cdot 0{,}7 = 3{,}5$ (ikke heltall, forhøyes til 4) $\Rightarrow x_{0,7} = x_{(4)} = \underline{7}$

Oppgave 1.7 Gitt x-verdiene:

21, 17, 18, 17, 22, 21, 78, 22, 24

Beregn standardavvik og interkvartilbredde. Hvilket av de to målene på spredning ville du stole mest på, dersom du i ettertid fikk opplyst at en av verdiene var feil?

Løsningsforslag:

$\Sigma x = 21 + 17 + \cdots + 24 = 240, \quad \Sigma x^2 = 21^2 + 17^2 + \cdots + 24^2 = 9412$

standardavvik: $s = \sqrt{\frac{1}{9-1}(9412 - \frac{1}{9} \cdot 240^2)} = \underline{19{,}4}$

For å beregne interkvartilbredde ordner vi først dataene i stigende rekkefølge:

17, 17, 18, 21, 21, 22, 22, 24, 78

Nedre kvartil: $np = 9 \cdot 0{,}25 = 2{,}25 \Rightarrow Q_1 = x_{(3)} = 18$

Øvre kvartil: $np = 9 \cdot 0{,}75 = 6{,}75 \Rightarrow Q_{(3)} = x_{(7)} = 22$

Interkvartilbredden blir: $Q_3 - Q_1 = 22 - 18 = \underline{4}$.

Det er større grunn til å stole på interkvartilbredden enn på standardavviket, da «ekstrem»-verdien 78 i datamaterialet trolig er feil og bidrar sterkt til å øke standardavviket. Interkvartilbredden er imidlertid robust (lar seg lite påvirke) av tallet som er feil.

Oppgave 1.8 Tabellutsnittet til høyre viser 3 vektklasser. Angi nedre og øvre klassegrense, samt klassemidtpunkt og klassebredde.

Vekt [kg]
⋮
[10–20>
[20–30>
[30–40>
⋮

Løsningsforslag: Siden det er snakk om vekter er det fornuftig å tenke seg at disse er forhøyet. Nedre klassegrenser for de tre klassene som er angitt med tall blir da 9,5, 19,5, og 29,5. De tilsvarende øvre klassegrensene blir 19,499.., 29,499... og 39,499..., mens midtverdiene blir 14,5, 24,5 og 34,5 (midt mellom øvre og nedre klassegrense).

Oppgave 1.9 Neste tabell viser daglig inntekt av CD-salg i en musikkforretning, fordelt på 4 inntektgrupper. Anta at et dagssalg på kr. 9999 kommer i den første klassen.

a) Bestem klassemidtpunkt i hver klasse.
b) Bestem gruppert middelverdi.
c) Bestem gruppert standardavvik.
d) Bestem gruppert median.
e) Bestem gruppert interkvartilbredde.

Videosalg (kr 1000)	Frekvens (dager)
0–9	37
10–19	148
20–29	123
30–39	57

Løsningsforslag:

klasse:	frekvens	relativ frekvens	klasse-midtpkt			kumulativ frekvens	relativ kum.frekv.
	f_i	f_i/n	m_i	$m_i \cdot f_i$	$m_i^2 \cdot f_i$	F_i	F_i/n
0-9	37	0,1014	5	185	925	37	0,1014
10-19	148	0,4055	15	2220	33300	185	0,5068
20-29	123	0,3370	25	3075	76875	308	0,8438
30-39	57	0,1562	35	1995	69825	365	1,0000
sum:	365			7475	180925		

Kronebeløpene som inngår i tabellen (f.eks. m_i-verdiene) er angitt i enheter på kr 1000.

a) Siden det er pengebeløp forhøyer vi ikke, og klassemidtpunktene blir som angitt i tabellen.

b) Gruppert middelverdi: $\bar{x}_g = \frac{1}{n}\Sigma m_i f_i = \frac{7475}{365} = 20{,}480$, dvs. <u>kr 20480</u>

c) Gruppert standardavvik:

$$s_g = \sqrt{\tfrac{1}{n}\Sigma m_i^2 f_i - \bar{x}_g^{\,2}} = \sqrt{\tfrac{180925}{365} - 20{,}480^2} = 8{,}73, \quad \text{dvs. } \underline{\text{kr } 8730}$$

d) Medianklassen er første klasse med relativ kumulativ frekvens større enn 0,5. Fra tabellen ser vi at dette er klassen 10-19 med nedre klassefrekvens x_1=10, klassefrekvens f_m =148, klassebredde Δx_m = 10 og med kumulativ frekvens F_1 =37 i klassen forut for medianklassen. Vi får da:

$$m_g = x_1 + \frac{n/2 - F_1}{f_m} \cdot \Delta x_m = 10 + \frac{365/2 - 37}{148} \cdot 10 = 19{,}83, \quad \text{dvs. } \underline{\text{kr } 19\ 830}$$

e) Nedre og øvre kvartilklasser er første klasse der relativ kumulativ frekvens overstiger henholdsvis 0,25 og 0,75. Klassen 10-19 er følgelig nedre kvartilklasse og klassen 20-29 er øvre kvartilklasse. Analogt med utregning av medianen får vi:

$$Q_{1g} = 10 + \frac{365 \cdot 0{,}25 - 37}{148} \cdot 10 = 13{,}666, \quad Q_{3g} = 20 + \frac{365 \cdot 0{,}75 - 185}{123} \cdot 10 = 27{,}215$$

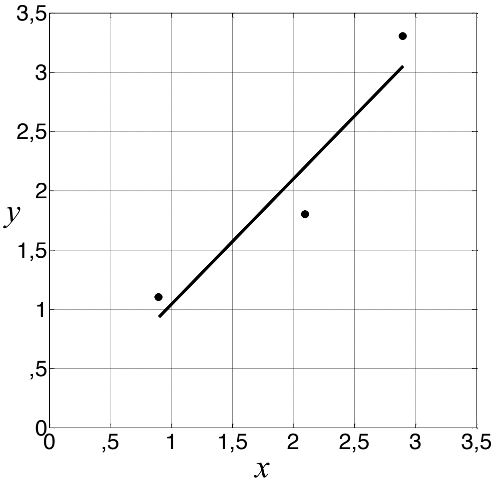
Gruppert interkvartilbredde: Q_{3g}–Q_{1g} = 27,215 −13,666 =13,55, dvs. <u>kr 13 550</u>

Oppgave 1.10 Gitt tallparene

(0,9 , 1,1), (2,1 , 1,8) og (2,9 , 3,3)

a) Velg 1 cm/enhet langs *x*- og *y*-aksen, og lag et spredningsdiagram.
b) Beregn korrelasjonskoeffisienten, *r*.
c) Beregn en regresjonslinje tilpasset dataene, og tegn den inn i spredningsdiagrammet.

<u>Løsningsforslag</u>: a) se figur til høyre



Hjelpestørrelser for b) og c):

$n = 3$

$\Sigma\, x = ,9 + 2{,}1 + 2{,}9 = 5{,}9$

$\Sigma\, y = 1{,}1 + 1{,}8 + 3{,}3 = 6{,}2$

$\Sigma\, x^2 = ,9^2 + 2{,}1^2 + 2{,}9^2 = 13{,}63$

$\Sigma\, y^2 = 1{,}1^2 + 1{,}8^2 + 3{,}3^2 = 15{,}34$

$\Sigma\, xy = ,9 \cdot 1{,}1 + 2{,}1 \cdot 1{,}8 + 2{,}9 \cdot 3{,}3 = 14{,}34$

b) korrelasjonskoeffisient: $r = \dfrac{\Sigma\, xy - \frac{1}{n}\Sigma x \Sigma y / n}{\sqrt{(\Sigma x^2 - \frac{1}{n}(\Sigma x)^2) \cdot (\Sigma y^2 - \frac{1}{n}(\Sigma y)^2)}} = \underline{0{,}949}$

c) regresjonslinje $y_r^* = a^* + b^* x,\ b^* = \dfrac{\Sigma\, xy - \frac{1}{n}\Sigma x \Sigma y}{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2} = 1{,}06,\ a^* = \bar{y} - b^* \cdot \bar{x} = -0{,}016$

Regresjonslinja $\underline{y^* = -{,}016 + 1{,}06 \cdot x}$ er tegnet inn i figuren mellom punktene

(0,9, y_r^*(0,9)) og (2,9, y_r^*(2,9)).

Oppgave 1.11 Gitt følgende samhørende x- og y-verdier:

x	,68	,52	,58	,72	,62
y	11	5	1	4	6

a) Beregn korrelasjonskoeffisienten, r.

b) Beregn en regresjonslinje tilpasset dataene, $y^* = a^* + b^* x$.

c) Tegn (x,y)-verdiene i et spredningsdiagram med 1 cm pr. enhet ***både*** langs x- og y-aksen, og tegn inn regresjonslinja. Gir figuren inntrykk av sterk korrelasjon?

d) Gjør c) på nytt, men velg selv en fornuftig skalering langs aksene.

<u>Løsningsforslag</u>: Hjelpestørrelser:

$\Sigma\, x = ,68 + ,52 + \cdots + ,62 = 3{,}12, \quad \Sigma\, y = 11 + 5 + \cdots + 6 = 27$

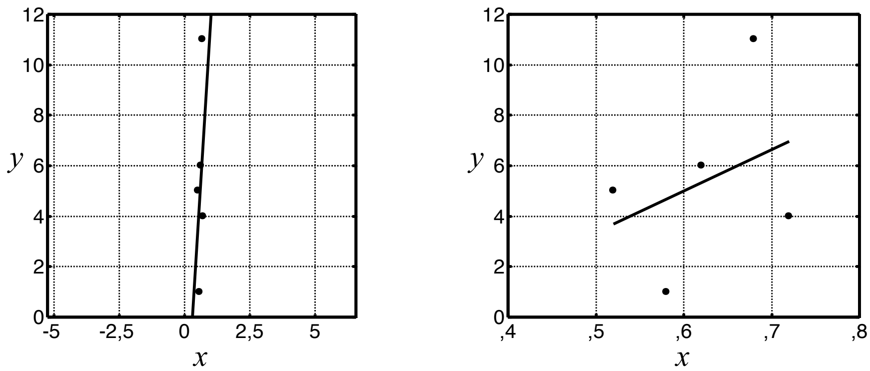
$\Sigma\, x^2 = ,68^2 + ,52^2 + \cdots + ,62^2 = 1{,}972, \quad \Sigma\, y^2 = 11^2 + 5^2 + \cdots + 6^2 = 199$

$\Sigma\, xy = ,68 \cdot 11 + ,52 \cdot 5 + \cdots + ,62 \cdot 6 = 17{,}26$

a) korrelasjonskoeffisient, $r = \dfrac{\Sigma xy - \frac{1}{5}\Sigma x \Sigma y}{\sqrt{(\Sigma x^2 - \frac{1}{5}(\Sigma x)^2)\cdot(\Sigma y^2 - \frac{1}{5}(\Sigma y)^2)}} = \underline{,356}$

b) $b^* = \dfrac{\Sigma xy - \frac{1}{5}\Sigma x \Sigma y}{\Sigma x^2 - \frac{1}{5}(\Sigma x)^2} = 16{,}40, \quad a^* = \bar{y} - b^* \cdot \bar{x} = -4{,}834, \quad \underline{y^* = -4{,}834 + 16{,}40 \cdot x}$

c), d) Spredningsplot og regresjonslinjer tegnet inn i figuren nedenfor. Skaleringen bør velges slik at utstrekningen av data i hver retning på figuren blir ca. like stor. I figuren til venstre har dataene for liten utsrekning i x-retning, og vi kan lett bli narret til å tro at vi har en sterk korrelasjon karakterisert ved en tilnærmet loddrett linje.



Oppgave 1.12 Størrelsen av en dyrebestand måles en gang i året. De 4 siste årene ble følgende observert:

år:	87	88	89	90
bestand:	123	237	471	982

La x = 1, 2, 3 og 4 betegne årstallene 87, 88, 89 og 90, og la y betegne de tilsvarende bestandene.

a) Beregn korrelasjonskoeffisienten.
b) Utfør transformasjonen $z = \ln y$, og beregn z-verdiene.
c) Bestem korrelasjonskoeffisienten for (x,z)-verdiene.
d) Tilpass en rett linje til (x,z)-dataene.
e) Bestem på basis av d) en regresjonsfunksjon, $y = ce^{dx}$, tilpasset (x,y)-dataene.

Løsningsforslag: Hjelpestørrelser:

z-verdier: {ln(123), ln(237), ln(471), ln(982)} = {4,8122, 5,4681, 6,1549, 6,8896}

$$\Sigma x = 1 + \cdots + 4 = 10,\ \Sigma y = 123 + \cdots + 982 = 1813,\ \Sigma z = 4{,}8122 + \cdots + 6{,}8896 = 23{,}3247$$

$$\Sigma x^2 = 30,\ \ \Sigma y^2 = 1257463,\ \ \Sigma z^2 = 138{,}4055,\ \ \Sigma xy = 5938,\ \ \Sigma xz = 61{,}7712,\ n = 4.$$

a) Korrelasjonskoeffisient $r(x,y) = \dfrac{\Sigma xy - \frac{1}{n}\Sigma x \Sigma y}{\sqrt{(\Sigma x^2 - \frac{1}{n}(\Sigma x)^2)\cdot(\Sigma y^2 - \frac{1}{n}(\Sigma y)^2)}} = \underline{,9522}$

b) z-verdiene er beregnet først i løsningsforslaget.

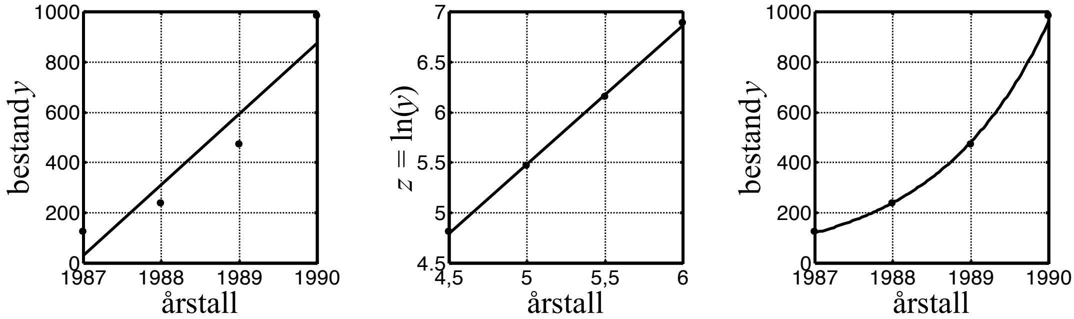
c) Korrelasjonskoeffisient $r(x,z) = \dfrac{\Sigma xy - \frac{1}{n}\Sigma x \Sigma z}{\sqrt{(\Sigma x^2 - \frac{1}{n}(\Sigma x)^2)\cdot(\Sigma z^2 - \frac{1}{n}(\Sigma z)^2)}} = \underline{,9997}$

Sammenlignet med $r(x,y)$ ser vi at en rettlinjet tilpasning til (x,z)-dataene later til å være en bedre tilpasning enn en rettlinjet tilpasning til (x,y)-dataene.

d) Regresjonslinje,

$$z^* = a^* + b^* \cdot x,\ \ b^* = \frac{\Sigma xz - \frac{1}{4}\Sigma x \Sigma z}{\Sigma x^2 - \frac{1}{4}(\Sigma x)^2} = \underline{,6919},\ \ a^* = \bar{z} - b^* \cdot \bar{x} = \underline{4{,}1014}$$

e)
$$y = ce^{dx} \Rightarrow z = \ln(y) = \ln(c) + d \cdot x = a^* + b^* \cdot x$$
$$\Rightarrow a^* = \ln(c),\ \ c = e^{a^*} = 60{,}426,\ d = b^* = ,6919,\ \ \underline{y = 60{,}426 \cdot e^{,6919 \cdot x}}$$



Oppgave 1.13 (Vekt-data i tabell lenger ned).

a) Finn middelverdi og standardavvik til vekt-dataene på basis av rådataene.
b) Rangordne vekt-dataene og bestem median og interkvartilbredde.
c) Gruppér vekt-dataene og bestem gruppert middelverdi, gruppert standardavvik og gruppert median.
d) Fremstill de grupperte vekt-dataene i et relativ frekvens-histogram.

Løsningsforslag:

a) vekt: $\Sigma y = 55 + 59 + \cdots + 85 = 2386$, middelverdi: $\bar{y} = \frac{1}{35}\Sigma y = \underline{68{,}17 \text{ kg}}$

standardavvik: $s_y = \sqrt{\frac{1}{34}(\Sigma y^2 - \frac{1}{35}(\Sigma y)^2)} = \underline{10{,}92 \text{ kg}}$

b) Rangordna vektdata:

$x_{(1)} = 50,\ x_{(2)} = 52,\ x_{(9)} = 59,\ x_{(18)} = 67,\ x_{(27)} = 76,\ x_{(34)} = 85,\ x_{(35)} = 90$

Median: $n = 35 \Rightarrow n\cdot{,}5 = 17{,}5$. Forhøyer til 18 og får: $m = x_{(18)} = \underline{67 \text{ kg}}$

Interkvartilbredde: $n\cdot{,}25 = 8{,}75 \Rightarrow Q_1 = x_{(9)} = 59$, $n\cdot{,}75 = 26{,}26 \Rightarrow Q_3 = x_{(27)} = 76$

$Q_3 - Q_1 = 76 - 59 = 17$, dvs. interkvartilbredden er $\underline{17 \text{ kg}}$

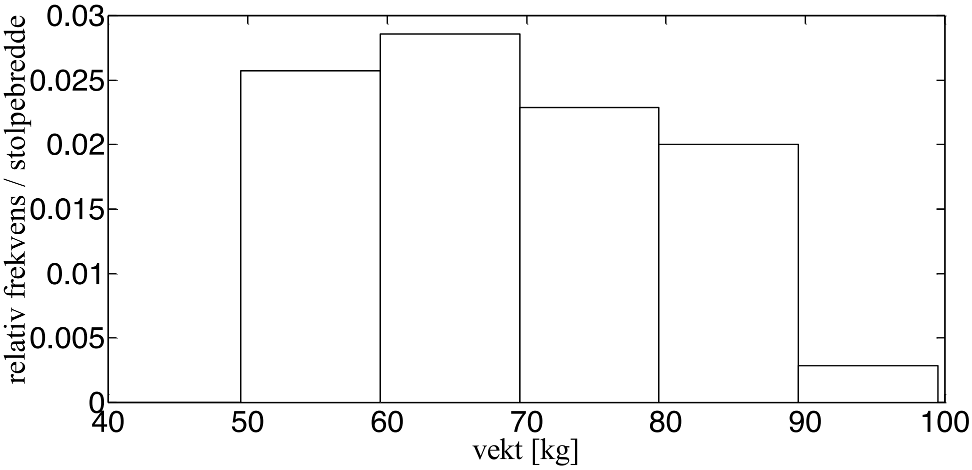
c) Minste vekt er på 50 kg og største på 90 kg. Vi velger klassebredde på 10 kg og får følgende tabell:

vekt-klasse: [kg]	frekvens f_i	relativ frekvens f_i/n	klasse-midtpkt m_i	$m_i \cdot f_i$	$m_i^2 \cdot f_i$	kumulativ frekvens F_i	relativ kum.frekv. F_i/n
[50-60>	9	0,257	54,5	490,5	26732	9	,257
[60-70>	10	,286	64,6	645	41602	19	,543
[70-80>	8	,229	74,5	596	44402	27	0,771
[80-90>	7	,200	84,5	591,5	49982	34	0,971
[90-100>	1	,029	94,5	94,5	8930	35	1,000
sum:	35	1,001		2417,5	171650		

Gruppert middelverdi: $\bar{y}_g = \frac{1}{35}\Sigma m_i \cdot f_i = \frac{2417{,}5}{35} = 69{,}07$, dvs. $\underline{69{,}07 \text{ kg}}$

Gruppert standardavvik: $s_y = \sqrt{\frac{1}{35}\Sigma m_i^2 \cdot f_i - (\bar{x}_g)^2} = \sqrt{\frac{171650}{35} - 69{,}07^2} = \underline{11{,}55 \text{ kg}}$

d) Relativ frekvenshistogram for de grupperte vektene:



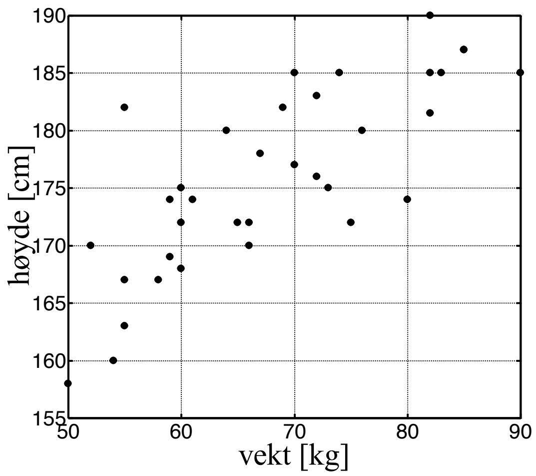
Oppgave 1.14 (data i tabell lenger ned).

a) Tegn et spredningsdiagram for samhørende vekt- og høyde-data.

b) Beregn korrelasjonskoeffisienten for vekt- og høyde-dataene, og kommenter spredningsdiagrammet i a).

c) Beregn korrelasjonskoeffisienten for høyde- og alders-dataene, og kommenter resultatet.

<u>Løsningsforslag</u>:



a) Se spredningsplot til venstre

b) Hjelpestørrelser:

$\Sigma x = 6160{,}5\,\text{cm},\ \Sigma y = 2386\ \text{kg}$

$\Sigma x^2 = 1{,}086500,\ \Sigma y^2 = 166700,$

$\Sigma xy = 422300,$

Korrelasjonskoeffisient: $r(x,y) =$

$$\frac{\Sigma xy - \Sigma x \Sigma y / 35}{\sqrt{(\Sigma x^2 - \frac{1}{35}(\Sigma x)^2)(\Sigma y^2 - \frac{1}{35}(\Sigma y)^2)}}$$

$= \underline{0{,}78}$

En korrelasjonskoeffisient på 0,78 indikerer en rimelig sterk positiv korrelasjon mellom vekt og høyde, hvilket stemmer overens med inntrykket fra figuren under pkt. a).

c) Ved tilsvarende regning som i b) finner vi korrelasjonskoeffisienten for (høyde,alders)- dataene: $r(x,z) = \underline{0{,}11}$. At denne indikerer en svak korrelasjon er intuitivt nokså logisk, da studenter stort sett er noskå fullt utvokste.

Tabell for oppgave 1.13 og 1.14

høyde [cm]	vekt [kg]	alder [år]	høyde [cm]	vekt [kg]	alder [år]	høyde [cm]	vekt [kg]	alder [år]
167	55	32	180	64	24	185	83	25
169	59	23	175	73	21	160	54	26
163	55	25	170	52	20	172	75	22
177	70	23	172	65	21	167	58	19
180	76	24	170	66	23	190	82	25
174	61	25	175	60	19	172	66	21
172	60	19	185	90	23	185	74	21
174	59	19	178	67	21	182	69	21
174	80	21	187	85	28	181,5	82	21
182	55	31	185	82	27	176	72	21
158	50	23	183	72	19	187	85	28
168	60	21	185	70	19			

Oppgave 1.15 (E) I et utvalg på 20 fire-barnsfamilier er antall gutter følgende:

3 2 3 2 3 1 3 3 2 3 1 3 0 3 2 1 4 2 3 2

a) Regn ut middelverdien for utvalget.
b) Regn ut utvalgsstandardavviket.
c) Sett opp en fordelingstabell for de absolutte og relative frekvensene.
d) Bestem median og interkvartilbredde.
e) Lag relativ frekvens stolpediagram.

Løsningsforslag:

a) middelverdi (antall gutter): $\bar{x} = \frac{1}{20}(3+2+\cdots+2) = \underline{2{,}30}$

b) $\Sigma x^2 = 3^2 + 2^2 + \cdots + 2^2 = 124,\ s_x = \sqrt{\frac{1}{19}(\Sigma x^2 - 20\bar{x}^2)} = \underline{0{,}979}$

c) # gutter	0	1	2	3	4
frekvens	1	3	6	9	1
relativ frekvens	,05	,15	,30	,60	,05

d) Vi ordner rådataene i stigende rekkefølge:

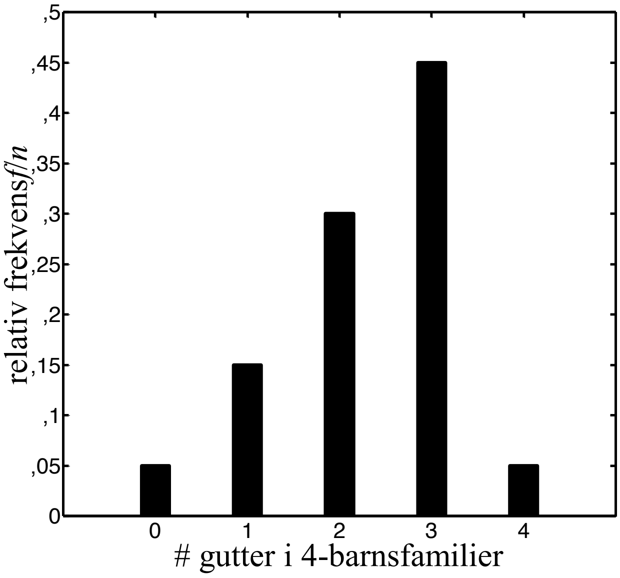
$x_{(1)} = 0, x_{(2)} = 1, \ldots, x_{(5)} = 2, x_{(6)} = 2, \ldots, x_{(10)} = 2, x_{(11)} = 3, \ldots, x_{(15)} = 3, x_{(16)} = 3, \ldots$

Median: $n \cdot 0{,}5 = 10 \Rightarrow m = \frac{1}{2}(x_{(10)} + x_{(11)}) = \underline{2{,}5}$

$n \cdot 0{,}25 = 5 \Rightarrow Q_1 = \frac{1}{2}(x_{(5)} + x_{(6)}) = 2,\ n \cdot 0{,}75 = 15 \Rightarrow Q_1 = \frac{1}{2}(x_{(15)} + x_{(16)}) = 3$

Interkvartilbredde: $Q_3 - Q_1 = 3 - 2 = \underline{\underline{1}}$

e) Relativ frekvens
stolpediagram:



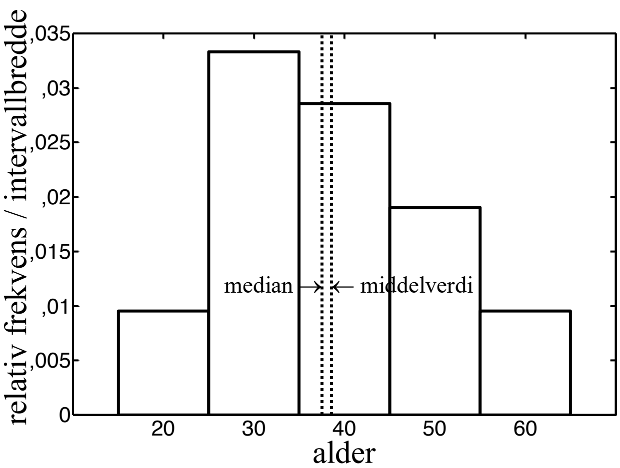
Oppgave 1.16 (E) Aldersfordelingen i en bedrift er:

Alder [år]	15-25	25-35	35-45	45-55	55-65
Frekvens	2	7	6	4	2

a) Tegn histogram over aldersfordelingen.
b) Beregn aritmetisk middelverdi, standardavvik og median.
c) Avmerk de beregnede størrelsene i histogrammet.

Løsningsforslag:

a)



b)

Klassemidtpunkt: 20, 30, 40, 50, 60

middelverdi: $\bar{x}_g = \frac{1}{21}\Sigma m_i f_i = \frac{1}{21}(20\cdot 2 + \cdots + 60\cdot 2) = 38{,}57$, dvs. $\underline{\underline{38{,}57 \text{ år}}}$

standardavvik: $s = \sqrt{\frac{1}{21}\Sigma m_i^2 f_i - \bar{x}_g^2} = \sqrt{\frac{33900}{21} - 38{,}57^2} = 11{,}25$, dvs. $\underline{11{,}25 \text{ år}}$

Klassen 35-45 år blir medianklassen fordi dette er første klasse der relativ kumulativ frekvens er større enn 0,5. Nedre klassegrense i medianklassen er x_1=35, f_m = 6 er frekvensen til medianklassen, Δx_m = 10 er bredden til medianklassen og F_1 =9 er kumulativ frekvens til klassen forut for medianklassen.

Gruppert median: $m_g = x_1 + \dfrac{n/2 - F_1}{f_m} \cdot \Delta x_m = 35 + \dfrac{10{,}5 - 9}{6} \cdot 10 \text{ år} = \underline{37{,}5 \text{ år}}$

c) Median og middelverdi er inntegnet i figuren under pkt. a).

Oppgave 1.17 (E) En måleserie resulterte i følgende klassedelte observasjonsmateriale:

intervall	frekvens
[195, 200>	2
[200, 205>	4
[205, 210>	10
[210, 215>	11
[215, 220>	9
[220, 225>	4

a) Utvid tabellen med relative frekvenser, kumulative frekvenser og relative kumulative frekvenser.

Lag et histogram for observasjonsmaterialet.

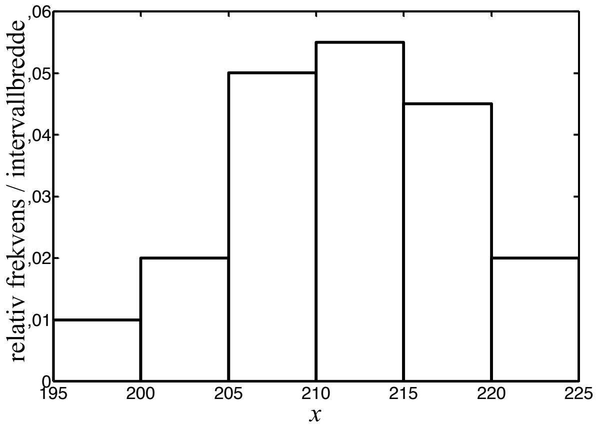
b) Beregn tilnærmet aritmetisk middelverdi og standardavvik for observasjonsmaterialet.

c) Beregn tilnærmet median og interkvartilbredde for observasjonsmaterialet.

Løsningsforslag:

a)

intervall	frekvens f_i	relativ frekvens	kumulativ frekvens	relativ kum. frekvens
[195, 200>	2	,050	2	,050
[200, 205>	4	,100	6	,150
[205, 210>	10	,250	16	,400
[210, 215>	11	,275	27	,675
[215, 220>	9	,225	36	,900
[220, 225>	4	,100	40	1,000
sum:	$n = 40$	1		



b) Klassemidtpunkter: m_i: 197,5, 202,5, 207,5, 212,5, 217,5, 222,5

middelverdi: $\bar{x}_g = \frac{1}{40}\Sigma m_i f_i = \frac{1}{40}(197{,}5\cdot 2 + \cdots + 222{,}5\cdot 2) = \underline{211{,}6}$

standardavvik: $s = \sqrt{\frac{1}{40}\Sigma m_i^2 f_i - \bar{x}_g^2} = \sqrt{\frac{1793100}{40} - 211{,}6^2} = \underline{6{,}5}$

c) Klassen 210-215 blir medianklassen fordi dette er første klasse der relativ kumulativ frekvens er større enn 0,5. Nedre klassegrense i medianklassen er x_1=210, f_m =11 er frekvensen til medianklassen, $\Delta x_m = 5$ er bredden til medianklassen og F_1 =16 er kumulativ frekvens til klassen forut for medianklassen.

Gruppert median: $m_g = x_1 + \frac{n/2 - F_1}{f_m}\cdot \Delta x_m = 210 + \frac{20-16}{11}\cdot 5 = \underline{211{,}8}$

Analogt til median er nedre og øvre interkvartilklasser de første klassene der relativ kumulativ frekvens overstiger henholdsvis 0,25 og 0,75, dvs. klassene 205-210 og 215-220. Vi får:

$Q_1 = 205 + \frac{40\cdot 0{,}25 - 6}{10}\cdot 5 = 207,\ Q_3 = 215 + \frac{40\cdot 0{,}75 - 27}{9}\cdot 5 = 216{,}7$

Interkvartilbredde: $Q_3 - Q_1 = 216{,}7 - 207 = \underline{9{,}7}$

Kapittel 2

Sannsynlighetsregning

Oppgave 2.1 Du trekker ett lodd fra en urne med 100 lodd. 3 lodd gir kakegevinst (K), 1 lodd gir bokgevinst (B) og 96 lodd gir ingen gevinst (I).

Skriv opp et passende utfallsrom, S, for det trukne loddet.

Løsningsforslag: Utfallsrommet består av alle mulige utfall av eksperimentet (loddet). Vi får da direkte at $S = \underline{\{K, B, I\}}$

Oppgave 2.2 Som i oppgave 2.1, men du trekker nå to lodd.

Løsningsforslag: Rekkefølgen er her uvesentlig og vi har trekning med tilbakelegging. Siden det kun er ett lodd som gir bokgevinst, utelukkes utfallet BB. Forøvrig må vi ta med alle kombinasjoner av to lodd: $S = \underline{\{KK, KB, KI, BI, II\}}$

Oppgave 2.3 En urne inneholder en rød (R), en blå (B) og en gul (G) kule. Hvor mange enkeltutfall består utfallsrommet S av i følgende tilfeller:

a) 2 trekninger uten tilbakelegging, rekkefølgen vesentlig.
b) 2 trekninger med tilbakelegging, rekkefølgen uvesentlig.
c) 3 trekninger uten tilbakelegging, rekkefølgen vesentlig.

Løsningsforslag:

a) uten tilbakelegging, ordnet rekkefølge. Ordningsregel: $P_2^3 = 3 \cdot 2 = \underline{6 \text{ kombinasjoner}}$

b) med tilbakelegging, uordnet rekkefølge. $S = \{RR, RB, RG, BB, BG, GG\}$,
$\underline{6 \text{ kombinasjoner}}$

c) uten tilbakelegging, ordnet rekkefølge. Kombinasjonsregel:
$\binom{3}{2} = \frac{3 \cdot 2}{1 \cdot 2} = \underline{6 \text{ kombinasjoner}}$

Oppgave 2.4 Et ruteark med 3·3 = 9 ruter skal fylles ut med 5 like kryss og 4 like ringer. Hvor mange kombinasjoner får vi?

Løsningsforslag: La posisjonen til hver rute være nummerert fra 1 til 9. Antall kombinasjoner tilsvarer antall kombinasjoner av posisjoner til de 5 kryssene. Vi har følgelig en situasjon som tilsvarer trekning uten tilbakelegging med uordnet rekkefølge. Kombinasjonsregelen gir: $\binom{9}{5} = \binom{9}{4} = \frac{9 \cdot 8 \cdot 7 \cdot 6}{1 \cdot 2 \cdot 3 \cdot 4} = \underline{126 \text{ kombinasjoner}}$

Oppgave 2.5 Et brett med 3·3 ruter skal dekkes med 9 forskjellige brikker. Hvor mange kombinasjoner får vi?

Løsningsforslag: Vi velger brikkene i vilkårlig rekkefølge. Den første kan dekke 9 forskjellige ruter. For hver av disse kan den andre dekke de resterende 8, osv. Dette tilsvarer en situasjon med trekning uten tilbakelegging med ordnet rekkefølge. Ordningsregelen gir: $P_9^9 = 9! = 9 \cdot 8 \cdots 1 = \underline{362880 \text{ kombinasjoner}}$

Oppgave 2.6 En terning kastes en gang. Hendelse A betyr at terningen viser en sekser, mens hendelse B betyr at terningen viser høyst 4 øyne. Forklar hva følgende hendelser betyr, og beregn sannsynligheten for hver av dem:

a) $(A \cup B)^C$, b) $A^C \cap B^C$, c) $(A \cap B)^C$, d) $A^C \cup B^C$

Løsningsforslag: $A = \{6\}, B = \{1,2,3,4\}$

a) $(A \cup B)^C = \{1,2,3,4,6\}^C = \{5\}$, $P(5'\text{er}) = \underline{1/6}$
b) $A^C \cap B^C = \{1,2,3,4,5\} \cap \{5,6\} = \{5\}$, $P(A^C \cup B^C) = \underline{1/6}$
c) $(A \cap B)^C = (\{6\} \cap \{1,2,3,4\})^C = (\emptyset)^C = S$, $P((A \cap B)^C) = \underline{1}$
d) $A^C \cup B^C = \{1,2,3,4,5\} \cup \{5,6\} = S$, $P(A^C \cup B^C) = \underline{1}$

Oppgave 2.7 Vis matematisk at

$P(A \cup B \cup C) = P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)$

Løsningsforslag:

$$P(A \cup B \cup C) = P((A \cup B) \cup C) = P(A \cup B) + P(C) - P((A \cup B)\, C) =$$
$$P(A) + P(B) - P(AB) + P(C) - P((AC) \cup (BC)) =$$
$$P(A) + P(B) - P(AB) + P(C) - P(AC) - P(BC) + P(ABC) =$$
$$P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC), \quad \underline{\text{q.e.d.}}$$

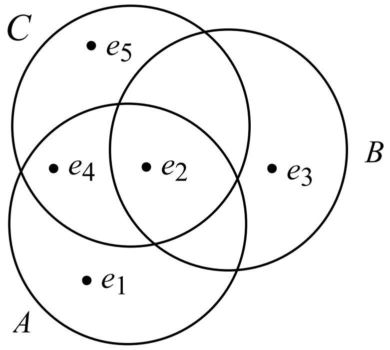
Oppgave 2.8 Gitt utfallsrommet $S = \{e_1, e_2, e_3, e_4, e_5\}$ der alle enkeltutfallene er like sannsynlige. La videre hendelsene A, B og C være definert som følger:

$A = \{e_1, e_2, e_4\}$, $B = \{e_2, e_3\}$ og $C = \{e_2, e_4, e_5\}$.

a) Tegn Venn-diagram og tegn inn hendelsene A, B og C.
b) Beregn $P(A)$, $P(A \cup B)$, $P(AC)$, $P(B^C)$, $P[(A \cup B)^C]$ og $P(A^C \cup B^C)$

Løsningsforslag:

a) Se Venn-diagram til høyre



b) Sannsynligheten for hendelsene kan vi finne ved å telle opp antall enkeltutfall hver hendelse inneholder, og så dele på 5:

$P(A) = \underline{,6}$, $P(A \cup B) = \underline{,8}$, $P(AC) = \underline{,4}$,

$P(B^C) = \underline{,6}$, $P[(A \cup B)^C] = \underline{,2}$ og $P(A^C \cup B^C) = \underline{,8}$

Oppgave 2.9 Angi telleregel og finn hvor mange enkeltutfall følgende eksperiment består av:

a) Ett myntkast med 5 pengestykker
b) Ett terningkast med 3 terninger
c) Trekning av 1., 2. og 3. premie blant 7 personer med like store vinnersjanser (trekning uten tilbakelegging).
d) Trekning uten tilbakelegging av 3 kuler fra urne med 2 gule og 4 røde kuler.

e) Antall kombinasjoner av 13 kort fra kortstokk med 52 forskjellige kort.

Løsningsforslag:

a) $m = 2$ utfall pr. mynt. (Med, uordna) situasjon. Potensregel: $2^5 = \underline{32 \text{ enkeltutfall}}$

b) $m = 6$ utfall pr. terning. (Med, uordna) situasjon. Potensregel: $6^3 = \underline{216 \text{ enkeltutfall}}$

c) (Uten, ordna) situasjon. Ordningsregel: $P_3^7 = 7 \cdot 6 \cdot 5 = \underline{210 \text{ enkeltutfall}}$

d) (Uten, uordna) situasjon. Vi kan ha 0, 1 eller 2 gule, og i hvert av tilfellene er resten av de 3 kulene røde. Neglisjeres fargerekkefølgen har vi derfor <u>3 enkeltutfall</u> (*RRR* ,*RRG*, *RGG*). Dersom fargerekkefølgen hadde vært vesentlig, ville vi hatt <u>7 enkeltutfall</u>, $S = \{RRR, RRG, RGR, GRR, GGR, GRG, RGG\}$.

e) (Uten, uordna) situasjon. $\binom{52}{13} = \dfrac{52!}{13! \cdot 39!} = \dfrac{8{,}1 \cdot 10^{67}}{6{,}2 \cdot 10^9 \cdot 2{,}0 \cdot 10^{46}} = \underline{6{,}35 \cdot 10^{11}}$

Oppgave 2.10 En oversikt over hvilken ut-danning ansatte ved en høyteknologisk bedrift har, er satt opp i følgende tabell:

	ingeniør-utdannet	annen utdanning
kvinner	10	30
menn	20	40

a) Hva er sannsynligheten for at en tilfeldig ansatt er kvinne?
b) Hva er sannsynligheten for at vedkommende er ingeniørutdannet og kvinne?
c) Hva er sannsynligheten for at vedkommende er ingeniørutdannet når vi vet at vedkommende er kvinne?
d) To tilfeldige ansatte snakker sammen. Hva er sannsynligheten for at ingen av dem er ingeniørutdannet?

Løsningsforslag: a) P(kvinne) = (10+30)/100 = 0,4

b) P(kvinne og ingeniørutdannet) = 10/100 = 0,1

c) P(ingeniørutdannet | kvinne) = P(ingeniørutdannet og kvinne)/P(kvinne) = 0,1/0,4 = 0,25

d) P(to ikke ingeniørutdannet) =

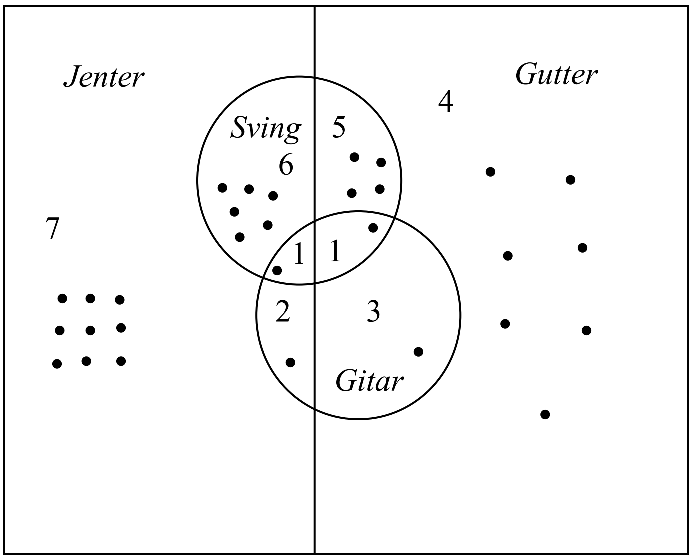
P(nr.1 ikke ingeniørutd.)·P(nr.2 ikke ingeniørutd. | nr. 1 ikke er ingeniørutd.) = (70/100)·(69/99) = 161/330

Oppgave 2.11 I en klasse på 30 elever kunne 12 danse sving (S), 4 kunne spille gitar (G) og det var 17 jenter (J) i klassen. 2 av jentene spilte gitar. En av guttene og en av jentene i klassen kunne både danse sving og spille gitar. 7 av guttene kunne verken spille gitar eller danse sving.

a) Hvor mange jenter kunne verken danse sving eller spille gitar?
b) Hvor mange gutter kunne danse sving?

(hint: Tegn Venn-diagram over alle elevene i klassen).

Løsningsforslag: Venn-diagrammet er utfylt slik forklart nedenfor.



1 Gutt og jente som både danser sving og spiller gitar

2 Siden 2 av jentene spiller gitar, må det også bli et punkt i område 2.

3 Siden 4 tilsammen spiller gitar, må det siste gitar-punktet bli i område 3.

4 De 7 guttene som verken kan spille gitar eller danse sving er merket i område 4.

5 Det er 13 gutter tilsammen, og 9 av disse er plottet. De 4 som mangler må ligge i område 5.

6 12 danser sving, og 6 av disse er plottet. De 6 som mangler må ligge i område 6.

7 Det er 17 jenter, og 8 er plottet. De 9 som mangler må ligge i område 7.

a) Antall jenter som verken kunne danse sving eller spille gitar er lik det antallet som ligger i område 7, dvs. 9 jenter.

b) Antall gutter som danser sving er lik det antallet som ligger i område 5 pluss det antall som ligger i høyre del av område 1, dvs. 5 gutter.

Oppgave 2.12 Anta at du har 20 ingredienser tilgjengelig for å lage pizzafyll.

a) Hvor mange forskjellige pizzaer kan bli laget av
   i) eksakt en ingrediens?
   ii) eksakt to ingredienser?
   iii) eksakt tre ingredienser?

b) Svar på ii) dersom to halvparter av en pizza kan være forskjellige.

Løsningsforslag: (Uten, uordna) situasjon. Kombinasjonsregelen gir:

a) i) $\binom{20}{1} = \underline{\underline{20 \text{ forskjellige pizza}}}$

ii) $\binom{20}{2} = \dfrac{20 \cdot 19}{1 \cdot 2} = \underline{190 \text{ forskjellige pizza}}$

iii) $\binom{20}{3} = \dfrac{20 \cdot 19 \cdot 18}{1 \cdot 2 \cdot 3} = \underline{1140 \text{ forskjellige pizza}}$

b) Vi kan nå trekke 2 forskjellige halvdeler uten tilbakelegging fra et utvalg på 190 forskjellige typer. Uordna rekkefølge gir $\binom{190}{2}$=17955 varianter. I tillegg kan de to halvdelene være like, så vi får totalt 17955 + 190 = <u>18145 forskjellige pizza</u>

Oppgave 2.13 Fra en kortstokk på 52 kort gjør vi to trekk etter hverandre uten tilbakelegging. Hva er sannsynligheten for å trekke

a) et svart og et rødt kort?
b) ikke mer enn én honnør (ess, konge, dame, knekt)?

<u>Løsningsforslag</u>: (uten, uordna) situasjon.

a) P(et svart og et rødt kort) = $P(SR)$ + $P(RS)$, der f.eks. SR står for hendelsen først svart, så rødt. Av symmetrigrunner er $P(SR) = P(RS)$.

$P(SR) = \frac{26}{52} \cdot \frac{26}{51} = \frac{13}{51}$, P(et svart og et rødt kort) = 2·(13/51) = <u>26/51</u>

b) P(maks én honnør) = P(0 honnør) + P(1 honnør)

P(0 honnør) = (36/52)·(35/51) = 105/221

P(1 honnør) = 2·(36/52)·(16/51) = 96/221

P(maks én honnør) = <u>201/221</u>

Oppgave 2.14 På en skole med 120 elever er det 70 jenter. 20 gutter og 10 jenter synes matematikk er gøy.

a) Hva er sannsynligheten for at en tilfeldig elev ikke syntes matematikk er gøy, når det viste seg at eleven var en gutt?
b) Hva er sannsynligheten for at to tilfeldige elever har forskjellig kjønn og synes matematikk er gøy?

<u>Løsningsforslag</u>:

a) Det er $120 - 70 = 50$ gutter. Av disse liker 20 matte og 30 liker ikke matte. Følgelig: P(matte ikke gøy | gutt) = 30/50 = <u>0,6</u>

b) A: Nr. 1 er gutt og liker matte, B: Nr. 2 er jente og liker matte

P(begge liker matte) = 2·$P(AB)$, der 2-tallet skyldes symmetri (like sannsynlig at nr. 1 er jente som at vedkommende er gutt).

$P(AB) = P(A) \cdot P(B \mid A)$ = (10/120)·(20/119) = 10/714

P(begge liker matte) = <u>10/357</u>

Oppgave 2.15 Du trekker 13 kort fra en kortstokk på 52 kort der 16 er honnørkort. Beregn sannsynligheten for at trekningen gir:

a) 13 kort med samme farge (hjerter, ruter, kløver eller spar).
b) Ingen honnørkort.
c) Bare honnørkort.

Løsningsforslag: (uten, uordna) situasjon, kombinasjonsregel.

a) Kun ett utfall med en gitt farge, dvs. kun 4 utfall der alle kort har samme farge.

$$P(13\text{ kort med samme farge}) = 4\cdot\frac{\binom{13}{13}\cdot\binom{39}{0}}{\binom{52}{13}} = \underline{6{,}3\cdot 10^{-12}}$$

b) Tilsammen 16 honnørkort og 36 ikke-honnørkort i kortstokken.

$$P(\text{ingen honnørkort}) = \frac{\binom{36}{13}\cdot\binom{16}{0}}{\binom{52}{13}} = \underline{0{,}0036}$$

c) $P(\text{bare honnørkort}) = \frac{\binom{16}{13}\cdot\binom{36}{0}}{\binom{52}{13}} = \underline{8{,}8\cdot 10^{-10}}$

Oppgave 2.16 En 6-kantet kubisk terning har to røde sider (R) og 4 blå sider (B). Terningen kastes 3 ganger.

a) Bruk den generelle addisjonssetningen til å beregne sannsynligheten for minst ett rødt kast.
b) Beregn også sannsynligheten på enklest mulig måte.

Løsningsforslag: R_i = rødt i i'te kast, i = 1, 2, 3. $P(R_i) = 2/3$, R_1, R_2 og R_3 uavhengige

a) P(minst ett rødt kast) = $P(R_1 \cup R_2 \cup R_3) =$

$$P(R_1)+P(R_2)+P(R_3)-P(R_1R_2)-P(R_1R_3)-P(R_2R_3)+P(R_1R_2R_3) =$$

$$3\cdot\tfrac{2}{3} - 3\cdot\left(\tfrac{2}{3}\right)^2 + \left(\tfrac{2}{3}\right)^3 = \underline{19/27}$$

b) P(minst ett rødt kast) = 1–P(ingen røde kast) =

$$1-P(R_1^C R_2^C R_3^C) = 1-(1-\tfrac{2}{3})^3 = \underline{19/27}$$

***Oppgave 2.17** En tipperekke inneholder 4 kamper, hver med 3 utfall (hjemmeseier (H), uavgjort (U) eller borteseier (B)). Hvor mange rekker må fylles ut for å være sikret minst 3 av 4 rette?

Løsningsforslag: Vi kan tenke som følger: Dersom vi fyller ut en tilfeldig rekke vil det være 9 vinnerrekker der den tilfeldige rekka har minst 3 rette. Eks: For rekka H H H H vil de 9 rekkene H H H H, H H H U, H H H B, H H U H, H H B H, H U H H, H B H H, U H H H, B H H H alle ha minst 3 enere. Siden det er $3^4 = 81$

mulige vinnerrekker, må vi derfor fylle ut minst 81/9 = 9 rekker for å være sikret minst 3 av 4 rette. Av symmetrigrunner er dette også tilstrekkelig. Svaret er altså 9 rekker. Nedenfor er vist et eksempel på 9 slike rekker:

H	B	B	B	H	U	U	U	H
H	U	H	B	U	U	H	B	B
H	U	B	H	B	H	U	B	U
U	U	B	H	H	B	H	U	B

Sjekk selv mot de 81 mulige kombinasjonene at en av rekkene ovenfor i hvert tilfelle gir minst 3 rette.

Oppgave 2.18 Sannsynligheten for at en mann er fargeblind er 0,08. Sannsynligheten for at en kvinne er fargeblind er 0,004.

a) Hva er sannsynligheten for at en tilfeldig utvalgt person er fargeblind?

Forutsett at halvparten av befolkningen er kvinner.

b) Det er begått en forbrytelse. Det er åpenbart at forbryteren er fargeblind. Hva er sannsynligheten for at forbryteren er en mann?

Løsningsforslag: F: fargeblind, M: mann, K: kvinne, $P(F \mid M) = ,08$, $P(F \mid K) = ,004$

a) $P(F) = P(F \mid M) \cdot P(M) + P(F \mid K) \cdot P(K) = ,08 \cdot ,5 + ,004 \cdot ,5 = \underline{,042}$

b) Bayes' formel: $P(M \mid F) = P(F \mid M) \cdot P(M)/P(F) = ,08 \cdot ,5/,042 = \underline{20/21}$

Oppgave 2.19 Bladet Motor slår fast i en omfattende artikkel 2/85 at bilparken i Norge er beheftet med betydelige feil. Anta at du besøker en bruktbilforretning som har 25 biler å tilby av en bestemt årsmodell. Av disse har to biler ingen feil, 5 har en feil, 10 har to feil, syv har tre feil og en har flere enn tre feil.

Du kjøper to biler. Hva er sannsynligheten for at du får:

a) begge biler uten feil,
b) én og bare én med høyst to feil,
c) to med minst tre feil hver,
d) to med minst tre feil tilsammen?

Løsningsforslag:

a) To biler uten feil å «trekke» fra: $P(\text{begge uten feil}) = \dfrac{\binom{2}{2}\binom{23}{0}}{\binom{25}{2}} = \underline{1/300}$

b) 17 biler med høyst to feil: $P(\text{én og kun én med høyst to feil}) = \dfrac{\binom{17}{1}\binom{8}{1}}{\binom{25}{2}} = \underline{34/75}$

c) 8 biler med minst 3 feil: $P(\text{to med minst 3 feil hver}) = \dfrac{\binom{8}{2}\binom{17}{0}}{\binom{25}{2}} = \underline{7\,/\,75}$

d) La (X,Y) betegne antall feil på hver bil: X feil på den ene, Y feil på den andre. $P(\text{minst 3 feil tilsammen}) = 1 - P(\text{høyst 2 feil tilsammen}) = 1 - P(X + Y \leq 2)$ $= \mathrm{P}(X + Y \in \{(0,0), (0,1), (0,2), (1,1)\})$

(0,0): Kun én kombinasjon

(0,1): $\binom{2}{1}\binom{5}{1} = 10$ kombinasjoner

(0,2): $\binom{2}{1}\binom{10}{1} = 20$ kombinasjoner

(1,1): $\binom{5}{2} = 10$ kombinasjoner

$P(\text{minst 3 feil tilsammen}) = 1 - (41/300) = \underline{259/300}$

Oppgave 2.20 Riktig eller galt?

a) Mulige utfall når vi kaster en rettferdig mynt to ganger er: ingen «kron», en «kron» og to «kron».
b) Hvis en rettferdig mynt har vært kastet 5 ganger med «kron» som utfall i alle 5 kast, er sannsynligheten for å få «mynt» i det 6. kastet større enn 0,5.
c) Det er mulig at vi for to hendelser, der $P(A) = 0{,}5$ og $P(\mathrm{B}) = 0{,}7$, kan ha $P(A \cup B) = 1{,}2$.
d) $P(A \text{¤} B)$ er alltid mindre enn $P(B)$
e) To ikke-tomme og gjensidig ekskluderende (disjunkte) hendelser er aldri uavhengige.
f) Hvis $P(A) = P(A \text{¤} B)$, så er A og B gjensidig ekskluderende hendelser.
g) Hvis $P(A) = 0{,}4$ og $P(B) = 0{,}3$ og $P(A \text{¤} B) = 0{,}8$, så er $P(AB) = 0{,}2$.

<u>Løsningsforslag</u>:

a) <u>Riktig</u>, forutsatt at vi neglisjerer at mynten lander på høykant.

b) <u>Galt</u>, $P(\text{mynt i } i\text{'te kast}) = 0{,}5$ uavhengig av tidligere utfall.

c) <u>Galt</u>, $0 \leq P(\text{hendelse}) \leq 1$ gjelder alltid.

d) <u>Galt</u>, hvis f.eks. A og B er uavhengige, så er $P(A \mid B) = P(A)$, og det er vilkårlig om $P(A)$ er større eller mindre enn $P(B)$.

e) <u>Riktig</u>, dersom den ene hendelsen har inntruffet vet vi jo med 100 % sikkerhet at den andre ikke kan ha inntruffet.

f) <u>Galt</u>, $P(A \mid B) = 0$ dersom A og B er gjensidig ekskluderende.

g) <u>Galt</u>, $P(AB) = P(A \mid B) \cdot P(B) = ,8 \cdot ,3 = ,24 \neq ,20$

Oppgave 2.21 En student skal opp til eksamen i 4 fag (*A*, *B*, *C* og *D*), og anslår de respektive sannsynligheter for å bestå eksamen til 0,8, 0,9, 0,7 og 0,5. Hvis vi forutsetter uavhengighet, hva er sannsynligheten for at studenten:

a) består alle 4 eksamener,
b) stryker i minst ett fag,
c) hvis studenten stryker i ett (og bare ett) fag, at han/hun stryker i fag *D*?

Løsningsforslag: $P(A) = ,8,\ P(B) = ,9,\ P(C) = ,7,\ P(D) = ,5$

a) $P(\text{består alle 4}) = P(ABCD) = ,8 \cdot ,9 \cdot ,7 \cdot ,5 = \underline{,252}$

b) $P(\text{stryke i minst ett fag}) = 1 - P(\text{bestå alle 4}) = 1 - ,252 = \underline{,748}$

c) $P(D^C \mid \text{én stryk}) = P(D^C ABC)/P(\text{én stryk}) =$

$$\frac{P(D^C ABC)}{P(A^C BCD) + P(AB^C CD) + P(ABC^C D) + P(ABCD^C)} =$$

$$\frac{(1-,5)\cdot,8\cdot,9\cdot,7}{(1-,8)\cdot,9\cdot,7\cdot,5+,8\cdot(1-,9)\cdot,7\cdot,5+,8\cdot,9\cdot(1-,7)\cdot,5+,8\cdot,9\cdot,7\cdot(1-,5)} = \frac{,252}{,451} = \underline{,559}$$

Oppgave 2.22 I en klubb med 10 medlemmer skal det utpekes et styre som består av i alt 3 medlemmer.

a) Hvor mange styrer har en å velge mellom?

Anta at et styre skal bestå av en formann, en nestformann og en sekretær.

b) Hvor mange styrer har en nå å velge mellom?

Løsningsforslag:

a) (uten, uordna) situasjon. Totalt: $\binom{10}{3} = \frac{10\cdot 9\cdot 8}{1\cdot 2\cdot 3} = \underline{120 \text{ styresammensetninger}}$

b) (uten, ordna) situasjon. Totalt: $P_3^{10} = 10\cdot 9\cdot 8 = \underline{720 \text{ styresammensetninger}}$

Oppgave 2.23 En medisin har 2 forskjellige «bivirkninger»: Kvalme (*K*) og hodepine (*H*). 20 % av brukerne blir kvalme ($P(K) = 0{,}2$), og blant de som blir kvalme er det 30 % som også får hodepine ($P(H¤K) = 0{,}3$). 10 % av brukerne får hodepine. De som verken blir kvalme eller får hodepine får ingen bivirkninger (*I*).

a) Hva er sannsynligheten for at en tilfeldig bruker både blir kvalm og får hodepine?
b) Er hendelsen bivirkning (*B*) det samme som unionen ($K \cup H$) eller snittet (*KH*) av *K* og *H*? Begrunn svaret og finn deretter $P(B)$ for en tilfeldig bruker.

Det viser seg at røykere (*R*) er betydelig mer utsatt for bivirkninger enn ikke- røykere (R^C). Av brukerne er det 75 % som ikke røyker, og blant disse er det 4 % som får bivirkninger.

c) Vis at sannsynligheten for å få bivirkning for en røyker er 84%.

d) Finn sannsynligheten for at en tilfeldig bruker uten bivirkninger er ikke-røyker.

Løsningsforslag: $P(K) = ,3$, $P(H \mid K) = ,3$, $P(H) = ,1$, $B =$ bivirkning

a) $P(KH) = P(K) \cdot P(H \mid K) = ,2 \cdot ,3 = \underline{,06}$

b) Bivirkning er enten hodepine, kvalme eller begge deler. Derfor får vi:

$P(B) = P(K \cup H) = P(K) + P(H) - P(KH) = ,2 + ,1 - ,06 = \underline{,24}$

c) $P(R^C) = ,75$, $P(B|R^C) = ,04$, $P(R^C|B) = \dfrac{P(B|R^C) \cdot P(R^C)}{P(B)} = \dfrac{,04 \cdot ,75}{,24} = ,125$

$P(B|R) = P(R|B) \cdot P(B) / P(R) = (1 - P(R^C|B)) \cdot P(B) / P(R) = ,875 \cdot ,24 / ,25 = \underline{,84}$

d) $P(R^C|B^C) = P(B^C|R^C) \cdot P(R^C) / P(B^C) = ,96 \cdot ,75 / ,76 = \underline{,947}$

Oppgave 2.24 (E) Sannsynligheten for at et tilfeldig besøk hos tannlegen vil resultere i tann-uttrekning er 0,06, sannsynligheten for at det vil resultere i plombering er 0,23 og sannsynligheten for at det vil resultere i både tannuttrekning og plombering er 0,02.

a) Hva er sannsynligheten for at et tilfeldig besøk hos tannlegen vil resultere i tannuttrekning, men ikke plombering?

b) Hva er sannsynligheten for at besøket vil resultere i tannuttrekning, plombering eller begge deler?

c) Hva er sannsynligheten for at besøket vil resultere i tannuttrekning eller plombering, men ikke begge deler?

d) Hva er sannsynligheten for at besøket verken vil resultere i tannuttrekning eller plombering?

Løsningsforslag: U: tannuttrekning, B:plombering, $P(U) = ,06$ $P(B) = ,23$ $P(UB) = ,02$

a) $P(U) = P(UB) + P(UB^C) \Rightarrow P(UB^C) = P(U) - P(UB) = ,06 - ,02 = \underline{,04}$

b) $P(U \cup B) = P(U) + P(B) - P(UB) = ,06 + ,23 - ,02 = \underline{,27}$

c) $P(U \cup B) - P(UB) = ,27 - ,02 = \underline{,25}$

d) $P((U \cup B)^C) = 1 - P(U \cup B) = 1 - ,27 = \underline{,73}$

Oppgave 2.25 (E) A og B er to kjennetegn slik at $P(A) = 0{,}25$, $P(B) = p$ og $P(A \cup B) = 0{,}5$.

a) Finn sannsynligheten $P(A \cap B)$ uttrykt ved p.

b) Hvilken verdi har p når A og B er uavhengige kjennetegn?

Løsningsforslag:

a) $P(AB) = P(A) + P(B) - P(A \cup B) = ,25 + p - ,5 = \underline{p - ,25}$

b) $P(AB) = P(A) \cdot P(B) = ,25p = p - ,25 \Rightarrow \underline{p = 1/3}$

Oppgave 2.26 (E) En ny medisin mot en bestemt sykdom blir innført. Forsøk har vist at 70 % av pasientene som får medisinen blir helbredet. Dessverre kan medisinen ha bivirkninger. For en pasient som blir helbredet er sannsynligheten for bivirkninger 0,25. For en pasient som ikke blir helbredet er sannsynligheten for bivirkninger 0,1.

a) Hva er sannsynligheten for at en tilfeldig pasient skal bli helbredet og dessuten ikke skal få bivirkninger?

b) Hva er sannsynligheten for bivirkninger hos en tilfeldig pasient?

c) Hva er sannsynligheten for at en pasient blir helbredet når vi vet at han ikke får bivirkninger av medisinen?

Løsningsforslag: B = bivirkning, H = helbredes, $P(H) =,7$ $P(B \mid H) =,25$ $P(B \mid H^C) =,1$

a) $P(HB^C) = P(H)\cdot P(B^C \mid H) = P(H)\cdot(1 - P(B \mid H)) = ,7\cdot(1 - ,25) = \underline{,525}$

b) $P(B) = P(B \mid H)\cdot P(H) + P(B \mid H^C)\cdot P(H^C) = ,25\cdot,7 + ,1\cdot,3 = ,175 + ,030 = \underline{,205}$

c) $P(H \mid B^C) = P(HB^C)/P(B^C) = ,0525/,795 = \underline{,660}$

Oppgave 2.27 (E) I en klubb er 60 % av medlemmene kvinner. 50 % av kvinnene røyker og 30 % av mennene røyker. Et medlem trekkes ut på måfå.

a) Hva er sannsynligheten for at medlemmet røyker?

b) Hva er sannsynligheten for at medlemmet er en kvinne, forutsatt at vedkommende røyker?

c) Hvor mange prosent av ikke-røykerne er menn?

Løsningsforslag: K: Kvinne, M: Mann, R: Røyker, $P(K) =,6$ $P(R \mid K) =,5$ $P(R \mid M) =,3$

a) $P(R) = P(R \mid K)\cdot P(K) + P(R \mid M)\cdot P(M) = ,5\cdot,6 + ,3\cdot,4 = ,3 + ,12 = \underline{,42}$

b) $P(K \mid R) = P(R \mid K)\cdot P(K)/P(R) = (,5\cdot,6)/,42 = ,3/,42 = \underline{,714}$

c) $P(M \mid R^C) = P(R^C \mid M)\cdot P(M)/P(R^C) = (1 - P(R \mid M))\cdot P(M)/(1 - P(R)) =$

$(1 -,3)\cdot,4/(1 -,42) = ,28/,58 = \underline{,483}$

Oppgave 2.28 (E) FBI bruker en løgndetektor som viser «skyldig» med sannsynlighet 0,92 for personer som virkelig er skyldige i en bestemt forbrytelse. Dersom løgndetektoren brukes på en uskyldig person vil den vise «uskyldig» med sannsynlighet 0,98.

Vi definerer følgende hendelser:

A: Personen er «skyldig».
B: Løgndetektoren viser «skyldig».

Det velges en tilfeldig person fra en (mistenkelig) gruppe der 10 % er «skyldige».

a) Skriv opp sannsynlighetene $P(A)$, $P(B \mid A)$, $P(B^C \mid A^C)$ og beregn

$P(A^C)$, $P(B^C \mid A)$, $P(B \mid A^C)$ og $P(AB)$.

Forklar hva den siste sannsynligheten står for.

b) Finn sannsynligheten for at løgndetektoren skal vise skyldig.

c) Anta at detektoren viser «uskyldig» for personen. Hva er sannsynligheten for at personen i virkeligheten er «uskyldig»?

Løsningsforslag:

a) $P(A) = \underline{,1}$, $P(B \mid A) = \underline{,92}$, $P(B^C \mid A^C) = \underline{,98}$,

$P(A^C) = 1 - P(A) = \underline{,9}$, $P(B^C \mid A) = 1 - P(B \mid A) = \underline{,08}$,

$P(B \mid A^C) = 1 - P(B^C \mid A^C) = \underline{,02}$, $P(AB) = P(A) \cdot P(B \mid A) = ,1 \cdot ,92 = \underline{,092}$

b) $P(B) = P(B \mid A) \cdot P(A) + P(B \mid A^C) \cdot P(A^C) = ,92 \cdot ,1 + ,02 \cdot ,9 = ,092 + ,018 = \underline{,110}$

c) $P(A^C \mid B^C) = P(B^C \mid A^C) \cdot P(A^C) / P(B^C) = ,98 \cdot ,9 / ,890 = \underline{,991}$

Oppgave 2.29 (E) I en urne er det 4 røde og 2 grønne kuler. Bortsett fra farge er de like. I denne oppgaven skal vi finne sannsynligheten for å trekke 2 røde og 1 grønn kule ved forskjellige trekningsprosedyrer.

a) Bestem sannsynligheten for å trekke 2 røde og 1 grønn kule dersom trekningen er *uten* tilbakelegging.

b) Hva blir sannsynligheten for å trekke 2 røde og 1 grønn kule dersom trekningen er *med* tilbakelegging?

c) Trekk 3 kuler på følgende måte: Trekk først 2 *uten* tilbakelegging, registrer farge og legg dem så oppi urna igjen før 3. trekk gjøres. Hva blir nå sannsynligheten for å få 2 røde og 1 grønn kule?

Løsningsforslag: *R*: Rød, *G*: Grønn

a) (uten, uordna) situasjon, $P(2R + 1G) = \dfrac{\binom{4}{2}\binom{2}{1}}{\binom{6}{3}} = \dfrac{6 \cdot 2}{20} = \underline{,6}$

b) (med, uordna) situasjon. Eks: hendelsen *RRG*, dvs. *R* i 1. og 2. trekning, *G* i 3. trekning, med sannsynlighet P(RRG) $= \frac{4}{6} \cdot \frac{4}{6} \cdot \frac{2}{6} = \frac{4}{27}$. I hvilken av de tre trekningene *G* trekkes er vilkårlig, og vi har $\binom{3}{1} = 3$ muligheter. For hver av disse er rekkefølgen til de to røde bestemt. Følgelig:

$P(2R + 1G) = 3 \cdot (4/27) = 12/27 = \underline{4/9}$

c) Blanda situasjon. Mulige utfall i 1. runde som gir $2R + 1G$ etter 2. runde: $\{\text{RR, RG}\}$. Disse utfall har sannsynligheter:

$$P(RR) = \frac{\binom{4}{2}\binom{2}{0}}{\binom{6}{2}} = \tfrac{2}{5}, \quad P(RG) = \frac{\binom{4}{1}\binom{2}{1}}{\binom{6}{2}} = \tfrac{8}{15}$$

$P(2R + 1G) = P(G$ i 2. runde | RR i 1.runde)·$P(RR$ i 1. runde) +

$P(R$ i 2. runde | RG i 1. runde)·$P(RG$ i 1. runde) =

(1/3)·(2/5) + (2/3)·(8/15) = <u>22/45</u>

Oppgave 2.30 (E) I et bygg på 6 etasjer kommer det 5 personer inn i heisen i 1. etasje. Personene har ingenting med hverandre å gjøre, og det er like stor sannsynlighet for at hver enkelt skal gå av i etasje 2-6 (5 stopp).

a) Hvor stor sannsynlighet er det for at alle går av i 2. etasje?
b) Hvor stor sannsynlighet er det for at alle går av i samme etasje?
c) Hvor stor sannsynlighet er det for at ingen skal til 6. etasje?
d) Hva er sannsynligheten for at det går av én i hver etasje?
e) Anta at det i starten bare var 4 personer i heisen. Hva er da sannsynligheten for at det går av høyst én i hver etasje?

<u>Løsningsforslag</u>: *A*-*E* tilsvarer person 1-5.

a) P(alle til 2) = $P(A$ til 2,…,E til 2) = $(1/5)^5$ = 1/3125 = <u>,00032</u>
b) P(alle til samme etasje) = 5·P(alle til etasje 2) = 5/3125 = 1/625 = <u>,0016</u>
c) P(ingen til 6) = $P(A$ ikke til 6,…,E ikke til 6) = $(1 - P(A \text{ til } 6))^5 = (4/5)^5$ = <u>,3277</u>
d) P(én til hver etasje) = $\dfrac{5!}{5^5}$ = <u>,0384</u>

Resonnement: Totalt $5^5 = 3125$ like sannsynlige utfall. *A* kan dra til 5 forskjellige etasjer. For hver av disse skal B til en av de 4 resterende etasjene, osv., som gir 5·4·3·2·1 = 5! = 120 like sannsynlige kombinasjoner.

e) P(høyst én i hver etasje) = P(alle til forskjellige etasjer) = $\dfrac{P_4^5}{5^4} = \dfrac{5\cdot4\cdot3\cdot2}{625} = \dfrac{120}{625}$ = <u>,1920</u>

Kapittel 3

Stokastisk variabel og sannsynlighetsfordeling

Oppgave 3.1 La X være en stokastisk variabel definert ved antall kron etter 2 kast med en rettferdig mynt. La K betegne kron og M betegne mynt i ett enkelt kast.

a) Skriv opp utfallsrommet, S.
b) Bestem fordelingen til x, $f(x)$.

<u>Løsningsforslag</u>: a) $S = \{\underline{KK, KM, MK, MM}\}$

b) Alle enkeltufall er like sannsynlige

$x = \underline{0}$: $f(0) = P(0 \text{ kron}) = P(MM) = \underline{1/4}$

$x = \underline{1}$: $f(1) = P(1 \text{ kron}) = P(KM) + P(MK) = \underline{2/4}$

$x = \underline{2}$: $f(2) = P(2 \text{ kron}) = P(KK) = \underline{1/4}$

Oppgave 3.2 La X være antall kast inntil første kron med en rettferdig mynt. Bestem fordelingen, $f(x)$.

<u>Løsningsforslag</u>:

$$f(x) = P(\underbrace{MMMK}_{x \text{ kast}}) = \underbrace{P(M)\cdot P(M)\cdots P(M)}_{x-1 \text{ ledd}}\cdot P(K) = \left(\tfrac{1}{2}\right)^{x-1}\cdot\left(\tfrac{1}{2}\right) = \underline{\left(\tfrac{1}{2}\right)^{x}},\ x = \underline{1,2,\ldots}$$

Oppgave 3.3 La X være antall seksere ved ett terningkast med 3 rettferdige terninger. Finn fordelingen, $f(x)$.

<u>Løsningsforslag</u>: A = 6'er, A^C = ikke-sekser

$S = \{AAA, AAA^C, AA^CA, AA^CA^C, A^CAA, A^CAA^C, A^CA^CA, A^CA^CA^C\}$, X = antall A'er

$x = \underline{0}$: $f(0) = P(A^CA^CA^C) = (5/6)^3 = \underline{125/216}$
$x = \underline{1}$: $f(1) = P(AA^CA^C) + P(A^CAA^C) + P(A^CA^CA) = 3\cdot(1/6)\cdot(5/6)^2 = \underline{75/216}$
$x = \underline{2}$: $f(2) = P(AAA^C) + P(AA^CA) + P(A^CAA) = 3\cdot(1/6)^2\cdot(5/6) = \underline{15/216}$
$x = \underline{3}$: $f(3) = P(AAA) = (1/6)^3 = \underline{1/216}$

Kontroll: $\Sigma f(x) = (125 + 75 + 15 + 1)/216 = 1$, OK!

Oppgave 3.4 La X være antall kast med en rettferdig terning inntil du får en sekser. Bestem fordelingen, $f(x)$.

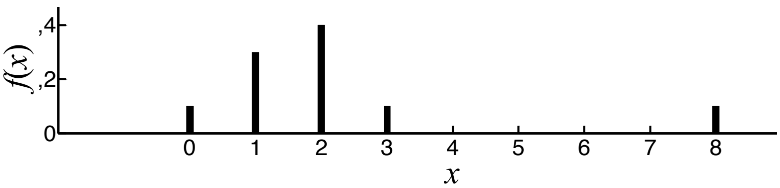
Løsningsforslag: A = 6'er i ett kast, B = ikke-sekser i ett kast.

$$f(x) = P(\underbrace{BBBA}_{x \text{ kast}}) = \underbrace{P(B)\cdot P(B)\cdots P(B)}_{x-1 \text{ ledd}}\cdot P(A) = \left(\tfrac{5}{6}\right)^{x-1}\cdot\left(\tfrac{1}{6}\right) = \underline{\left(\tfrac{5}{6}\right)^{x-1}\cdot\left(\tfrac{1}{6}\right)},\; x = \underline{1,2,\ldots}$$

Oppgave 3.5 Tegn stolpediagram for følgende fordeling, $f(x)$:

x:	0	1	2	3	8
$f(x)$:	,1	,3	,4	,1	,1

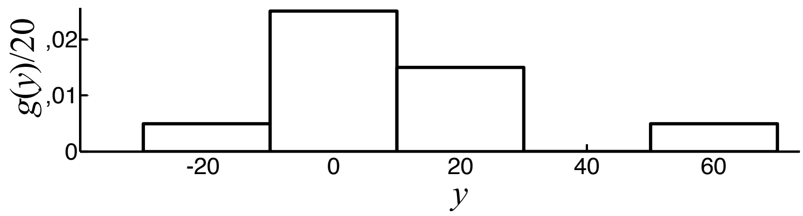
Løsningsforslag:



Oppgave 3.6 Tegn sannsynlighetshistogram for følgende fordeling, $g(y)$:

y:	–20	0	20	40	60
$g(y)$:	,1	,5	,3	0	,1

Løsningsforslag:



Oppgave 3.7 Finn forventningene $\mu_X = E(X)$ og $\mu_Y = E(Y)$, der de tilhørende fordelingene er definert i henholdsvis oppgave 3.5 og 3.6. Bestem deretter $E(2X-5Y)$.

Løsningsforslag: $\mu_X = \Sigma x\cdot f(x) = 0\cdot{,}1 + 1\cdot{,}3 + 2\cdot{,}4 + 3\cdot{,}1 + 8\cdot{,}1 = \underline{2{,}2}$

$\mu_Y = \Sigma y\cdot g(y) = -20\cdot{,}1 + 0\cdot{,}5 + 20\cdot{,}3 + 40\cdot 0 + 60\cdot{,}1 = \underline{10}$

$E(2X - 5Y) = 2\cdot E(X) - 5\cdot E(Y) = 2\cdot 2{,}2 - 5\cdot 10 = \underline{-45{,}6}$

Oppgave 3.8 Bestem standardavvikene, σ_X = std(X), og σ_Y = std(Y), der fordelingene til X og Y er gitt i henholdsvis oppgave 3.5 og 3.6. Bestem også E(X^3).

Løsningsforslag: Benytter resultatene E(X) = 2,2 og E(Y) = 10 fra oppgave 3.7.

$E(X^2) = \Sigma x^2 \cdot f(x) = 0^2 \cdot ,1 + 1^2 \cdot ,3 + 2^2 \cdot ,4 + 3^2 \cdot ,1 + 8^2 \cdot ,1 = 9{,}2$

$Var(X) = EX^2 - E^2X = 9{,}2 - 2{,}2^2 = 4{,}36, \qquad std(X) = \sqrt{Var(X)} = \sqrt{4{,}36} = \underline{2{,}09}$

$E(Y^2) = \Sigma y^2 \cdot g(y) = (-20)^2 \cdot ,1 + 0^2 \cdot ,5 + 20^2 \cdot ,3 + 40^2 \cdot 0 + 60^2 \cdot ,1 = 520$

$Var(Y) = EY^2 - E^2Y = 520 - 10^2 = 420, \qquad std(Y) = \sqrt{Var(Y)} = \sqrt{420} = \underline{20{,}49}$

$E(X^3) = \Sigma x^3 \cdot f(x) = 0^3 \cdot ,1 + 1^3 \cdot ,3 + 2^3 \cdot ,4 + 3^3 \cdot ,1 + 8^3 \cdot ,1 = \underline{57{,}4}$

Oppgave 3.9 (E) Vi skal betrakte TILs første 2 hjemmekamper neste sesong. La H betegne hjemmeseier for TIL, la U betegne uavgjort og la B betegne borteseier for TILs motstander (TIL-tap). Enkeltutfallet HU betyr for eksempel TIL-seier i 1. kamp og uavgjort i 2. kamp.

a) Bestem utfallsrommet S for utfallet av TILs første 2 hjemmekamper neste sesong.

Vi gjør nå følgende antagelser: Utfallene av de 2 kampene er uavhengige av hverandre, og sannsynligheten for seier, uavgjort og tap for TIL i de 2 kampene er som følger:

1. kamp: $P(H) = ,7 \quad P(U) = ,2 \quad P(B) = ,1$
2. kamp: $P(H) = ,5 \quad P(U) = ,2 \quad P(B) = ,3$

b) Finn sannsynligheten for hvert enkeltutfall i utfallsrommet S.
c) La X være antall TIL-seire i de første to hjemmekampene, og bestem fordelingen til X.

Løsningsforslag: 1. og 2. bokstav angir utfall av henholdsvis 1. og 2. kamp.

a) $S = \{\underline{HH, HU, HB, UH, UU, UB, BH, BU, BB}\}$

b) $P(HH) = P(H_1) \cdot P(H_2) = ,7 \cdot ,5 = \underline{,35}$
$P(HU) = P(H_1) \cdot P(U_2) = ,7 \cdot ,2 = \underline{,14}$
$P(HB) = P(H_1) \cdot P(B_2) = ,7 \cdot ,3 = \underline{,21}$
$P(UH) = P(U_1) \cdot P(H_2) = ,2 \cdot ,5 = \underline{,10}$
$P(UU) = P(U_1) \cdot P(U_2) = ,2 \cdot ,2 = \underline{,04}$
$P(UB) = P(U_1) \cdot P(B_2) = ,2 \cdot 3 = \underline{,06}$
$P(BH) = P(B_1) \cdot P(H_2) = ,1 \cdot ,5 = \underline{,05}$
$P(BU) = P(B_1) \cdot P(U_2) = ,1 \cdot ,2 = \underline{,02}$
$P(BB) = P(B_1) \cdot P(B_2) = ,1 \cdot ,3 = \underline{,03}$

c) $x = \underline{0}$: $\quad f(0) = P(\text{ingen hjemmeseier}) =$

$P(UU) + P(UB) + P(BU) + P(BB) = ,04 + ,06 + ,02 + ,03 = \underline{,15}$

$x = \underline{1}$: $f(1) = P(1 \text{ hjemmeseier}) =$

$P(HU) + P(HB) + P(UH) + P(BH) = ,14 + ,21 + ,10 + ,05 = \underline{,50}$

$x = \underline{2}$: $f(2) = P(2 \text{ hjemmeseire}) = P(HH) = \underline{,35}$

Oppgave 3.10 Gitt følgende simultanfordeling, $f(x,y)$:

$y \backslash x$	0	1	2
0	0,1	0	0,2
1	0,2	0,4	0,1

a) Bestem marginalfordelingene $f_X(x)$ og $f_Y(y)$.
b) Beregn forventning og varians til X og Y.
c) Er X og Y uavhengige? Begrunn svaret.
d) Beregn fordelingen til $Z = X - Y$.
e) Beregn forventning og varians til Z.

Løsningsforslag:

a)

$y \backslash x$	0	1	2	$f_Y(y)$
0	0,1	0	0,2	0,3
1	0,2	0,4	0,1	0,7
$f_X(x)$	0,3	0,4	0,3	1

b) $\mathrm{E}X = \Sigma x \cdot f_X(x) = 0 \cdot ,3 + 1 \cdot ,4 + 2 \cdot ,3 = \underline{1,0}$, $\mathrm{E}Y = \Sigma y \cdot f_Y(y) = 0 \cdot ,3 + 1 \cdot ,7 = \underline{,7}$

$\mathrm{E}X^2 = \Sigma x^2 \cdot f_X(x) = 0^2 \cdot ,3 + 1^2 \cdot ,4 + 2^2 \cdot ,3 = 1,6,$

$\mathrm{Var}(X) = \sigma_X^2 = \mathrm{E}X^2 - \mathrm{E}^2 X = 1,6 - 1,0^2 = \underline{,6}$

Tilsvarende finner vi at $\mathrm{Var}(Y) = \sigma_Y^2 = ,7 - ,7^2 = \underline{,21}$

c) $f_{X,Y}(0,0) = ,1 \neq f_X(0) \cdot f_Y(0) = ,3 \cdot ,3 = ,09$, dvs. X og Y ikke uavhengige

d) $Z = X - Y$ kan ha verdiene -1, 0, 1 og 2. Sannsynlighetene fås ved å summere langs diagonalene med negativ helningsvinkel i tabellen.

$f_Z(\underline{-1}) = f_{X,Y}(0,1) = \underline{,2}$, $f_Z(\underline{0}) = f_{X,Y}(0,0) + f_{X,Y}(1,1) = ,1 + ,4 = \underline{,5}$

$f_Z(\underline{1}) = f_{X,Y}(1,0) + f_{X,Y}(2,1) = 0 + ,1 = \underline{,1}$, $f_Z(\underline{2}) = f_{X,Y}(2,0) = \underline{,2}$

e) $\mathrm{E}Z = \Sigma z \cdot f_Z(z) = -1 \cdot ,2 + 0 \cdot ,5 + 1 \cdot ,1 + 2 \cdot ,2 = \underline{,3}$

$\mathrm{E}Z^2 = \Sigma z^2 \cdot f_Z(z) = (-1)^2 \cdot ,2 + 1^2 \cdot ,1 + 2^2 \cdot ,2 = 1,1$

$\mathrm{Var}Z = \mathrm{E}Z^2 - \mathrm{E}^2 Z = 1,1 - ,3^2 = 1,1 - ,09 = \underline{1,01}$

Oppgave 3.11 La Var(X) = 4, Var(Y) = 9 og Cov(X,Y) = −3.

a) Bestem korrelasjonskoeffisienten, $\rho(X,Y) = \mathrm{Corr}(X,Y)$.
b) Bestem Var(2X+3Y).
c) Er X og Y uavhengige ?

Løsningsforslag:

a) $\mathrm{Corr}(X, Y) = \mathrm{Cov}(X, Y)/(\mathrm{std}(X)\cdot\mathrm{std}(Y)) = \frac{-3}{\sqrt{4}\cdot\sqrt{9}} = \underline{-1/2}$

b) $\mathrm{Var}(2X + 3Y) = 2^2\cdot\mathrm{Var}(X) + 3^2\cdot\mathrm{Var}(Y) + 2\cdot2\cdot3\cdot\mathrm{Cov}(X, Y) =$

$4\cdot4 + 9\cdot9 + 12\cdot(-3) = \underline{61}$

c) $\mathrm{Corr}(X, Y) \neq 0 \Rightarrow$ X og Y ikke uavhengige.

Oppgave 3.12 I mangel av noe bedre å gjøre sitter noen studenter og kaster terning. Spillet er: En terning kastes til man får 1 eller 6.

a) Finn sannsynligheten for at spillet stopper etter første kast.
b) Vis at sannsynligheten for at spillet stopper etter 3 eller færre kast er 19/27.
c) Anta at man spiller dette spillet med en innsats på kr. 100. Man vinner kr 100 hvis antall kast som må utføres er forskjellig fra 2 og 3 og taper innsatsen ellers. Finn forventet gevinst.

Løsningsforslag: X = antall kast inntil 1'er eller 6'er. A = 1'er eller 6'er i ett kast.

a) $P(X = 1) = P$(1'er eller 6'er i ett kast) $= p = 2/6 = \underline{1/3}$

b) $P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3)$
$P(X = 2) = P(A^C A) = (1 - p)\cdot p = (2/3)\cdot(1/3) = 2/9$
$P(X = 3) = P(A^C A^C A) = (1 - p)^2\cdot p = (2/3)^2\cdot(1/3) = 4/27$
$P(X \leq 3) = 1/3 + 2/9 + 4/27 = \underline{19/27}$

c) Definerer $Z = -1$ når $x = 2,3$ og $Z = 1$ når $x \neq 2, 3$ $\Rightarrow$ Gevinst $G = 100\cdot Z$.

$P(Z = -1) = P(X = 2) + P(X = 3) = 2/9 + 4/27 = 10/27$
$P(Z = 1) = 1 - P(Z = -1) = 17/27$

Forventet gevinst: $E(G) = \mathrm{E}(100\cdot Z) = 100\cdot\mathrm{E}(Z) = 100\cdot\Sigma_z\, z\cdot P(Z = z)$
$= 100\cdot(-1\cdot P(Z = -1) + 1\cdot P(Z = 1)) = 100/27\cdot(-10 + 17) = \underline{700/27}$

Oppgave 3.13 (E) En sannsynlighetsfordeling er gitt ved:

x	0	1	2	3	4
$P(X=x)$	$2k$	$3k$	k	$3k$	$2k$

a) Bestem verdien av konstanten k.
b) Finn forventning og standardavvik.

Løsningsforslag:

a) $\Sigma_x P(X=x) = 2k + 3k + k + 3k + 2k = 1 \Rightarrow \underline{k = 1/11}$

b) $EX = \Sigma\, x \cdot P(X=x) = 1/11 \cdot (0 \cdot 2 + 1 \cdot 3 + 2 \cdot 1 + 3 \cdot 3 + 4 \cdot 2) = \underline{2}$

$EX^2 = \Sigma\, x^2 \cdot P(X=x) = 1/11 \cdot (0^2 \cdot 2 + 1^2 \cdot 3 + 2^2 \cdot 1 + 3^2 \cdot 3 + 4^2 \cdot 2) = 6$

$\mathrm{Var}(X) = EX^2 - E^2X = 6 - 2^2 = 2, \qquad \mathrm{Std}(X) = \sqrt{Var(X)} = \sqrt{2} = \underline{1{,}41}$

Oppgave 3.14 (E) I en modell for olje- og gass-reservoarer er det laget prognoser for oljemengde og gassmengde. Vi betegner oljemengde med X og gassmengde med Y. De stokastiske variablene X og Y antas å ha følgende fordelinger (i passende enheter):

x	12	15	18
$P(X=x)$	1/3	1/3	1/3

y	21	27	33
$P(Y=y)$	1/4	1/2	1/4

a) Finn forventet oljemengde $E(X)$ og finn forventningen $E(X/2-5)$.

b) Finn variansene $\mathrm{Var}(X)$ og $\mathrm{Var}(X/2-5)$.

c) Gitt at $P(X = 15 \cap Y = 27) = 1/4$. Er X og Y uavhengige?

Hva er betinget sannsynlighet $P(X =\square 15 \mid Y =\square 27)$?
Er det mulig å regne ut $E(X+Y)$ og $\mathrm{Var}(X+Y)$ med de gitte opplysninger?

d) Drivverdigheten av olje- og gassfeltet avhenger av en rekke andre faktorer. Hvis oljemengden er hhv 12,15 og 18, regner en sannsynligheten for at feltet er drivverdig er hhv ,6 ,8 og ,9. Bruk disse opplysningene og fordelingen til X for å beregne sannsynligheten for at feltet er drivverdig.

Løsningsforslag:

a) $E(X) = \Sigma_x\, x \cdot P(X=x) = 12 \cdot (1/3) + 15 \cdot (1/5) + 18 \cdot (1/3) = \underline{15}$

$E(X/2 - 5) = (1/2) \cdot E(X) - 5 = \underline{2{,}5}$

b) $E(X^2) = \Sigma_x\, x^2 \cdot P(X=x) = 12^2 \cdot (1/3) + 15^2 \cdot (1/3) + 18^2 \cdot (1/3) = 231$

$\mathrm{Var}(X) = EX^2 - E^2X = 231 - 15^2 = \underline{6}$

$\mathrm{Var}(X/2 - 5) = (1/2)^2 \cdot \mathrm{Var}(X) = 6/4 = \underline{1{,}5}$

c) $P(X = 15, Y = 27) = ¼ \neq P(X = 15) \cdot P(Y = 27) = (1/3) \cdot (1/2) = 1/6$
$\Rightarrow$ X og Y er <u>ikke uavhengige</u>.

$P(X = 15 \mid Y = 27) = P(X = 15, Y = 27)/P(Y = 27) = (1/4)/(1/2) = \underline{½}$

$E(X+Y) = EX + EY$ <u>kan</u> regnes ut.

$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$ kan <u>ikke</u> regnes ut, fordi $\text{Cov}(X,Y)$ er ukjent.

d) D = drivverdig.

$P(D) = \Sigma_x P(D \mid X) \cdot P(X = x) = ,6 \cdot (1/3) + ,8 \cdot (1/3) + ,9 \cdot (1/3) = 2,3/3 = \underline{,77}$

Kapittel 4

Diskrete fordelinger

Oppgave 4.1 Vi betrakter en tilfeldig valgt familie med 6 barn. Anta at sannsynligheten for «guttefødsel» er 1/2, og at hvilket kjønn hvert enkelt av de 6 barna har er uavhengig av hvilket kjønn de andre barna har. Beregn følgende sannsynligheter:

a) P(3 gutter og 3 jenter i familien)
b) P(flere jenter enn gutter i familien)

La X betegne antall gutter i familien.

c) Beregn E(X) og sd(X).

Løsningsforslag: X = antall gutter. Vi antar at X er binomisk fordelt: $X \sim$ Bino(6, ,5), dvs.:

$$f(x) = P(X = x) = \binom{6}{x} \cdot ,5^x \cdot (1 - ,5)^{6-x} = \binom{6}{x} \cdot ,5^6 = \tfrac{1}{64} \cdot \binom{6}{x}, \ x = 0,1,\ldots,6$$

a) $P(\text{3 gutter og 3 jenter}) = P(X = 3) = \tfrac{1}{64} \cdot \binom{6}{3} = \tfrac{20}{64} = \underline{5/16}$ (eksakt)

Alternativt, fra tabell: $P(X = 3) = P(X \leq 3) - P(X \leq 2) = ,656 - ,344 = \underline{,312}$

b) $P(\text{flere gutter enn jenter}) = P(X \geq 4) = 1 = P(X \leq 3) =$
$1 - (P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) =$
$1 - \tfrac{1}{64} \cdot \left(\binom{6}{0} + \binom{6}{1} + \binom{6}{2} + \binom{6}{3}\right) = 1 - \tfrac{1+6+15+20}{64} = \underline{11/32}$ (eksakt)

Alternativt, fra tabell: $1 - P(X \leq 3) = 1 - ,656 = \underline{,344}$

c) $E(X) = np = 6 \cdot (1/2) = \underline{3}$

$\text{Var}(X) = np(1 - p) = 6 \cdot (1/2) \cdot (1 - (1/2)) = 3/2$, $\ \text{Std}(X) = \sqrt{Var(X)} = \sqrt{3/2} = \underline{1{,}22}$

Oppgave 4.2 I et stort lotteri gav 5 % av loddene gevinst. Du kjøper 10 lodd. Hva er sannsynligheten for

a) 1 gevinst
b) Minst 2 gevinster

Løsningsforslag: stort lotteri $\Rightarrow$ tilnærmet binomisk situasjon med n = 10 og p = ,05.

X = antall gevinster, $f(x) = P(X = x) = \binom{10}{x} \cdot ,05^x \cdot (1 - ,05)^{10-x}$, $\ x = 0,1,\ldots,10$

a) $P(X = 1) = \binom{10}{1} \cdot ,05^1 \cdot (1 - ,05)^9 = \underline{,315}$

Alternativt, fra tabell: $P(X \leq 1) - P(X \leq 0) = ,914 - ,59 = \underline{,315}$

b) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) =$

$$1-\binom{10}{0}\cdot,05^0\cdot,95^{10}-,315 = 1-,599-,315 = \underline{,086}$$

Alternativt, fra tabell: $1 - P(X \leq 1) = 1 - ,914 = \underline{,086}$

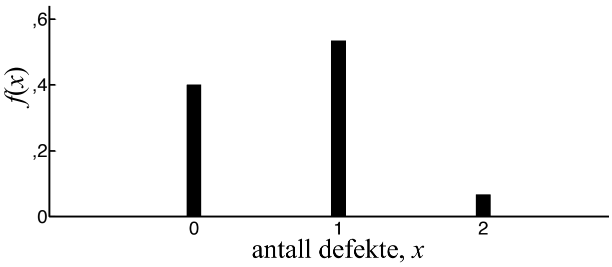
Oppgave 4.3 En bilforretning har 6 nye varevogner av en bestemt type på lager. 1/3 av disse bilene har fabrikasjonsfeil. Du kjøper 2 biler og de blir tatt ut tilfeldig blant de 6. La X være antall biler med fabrikasjonsfeil som du får utlevert. Sett opp fordelingen til x i tabellform og framstill den grafisk. Finn E(X) og std(X).

Løsningsforslag: Hypergeometrisk situasjon, $N = 6$, $D = 2$, $n = 2$, X = # defekte i utvalg er hyp(n,D,N).

$$f(0) = P(X=0) = \frac{\binom{D}{x}\cdot\binom{N-D}{n-x}}{\binom{N}{n}} = \frac{\binom{2}{0}\cdot\binom{6-2}{2-0}}{\binom{6}{2}} = \frac{1\cdot\binom{4}{2}}{\binom{6}{2}} = \frac{6}{15} = \underline{2/5}$$

$$f(1) = P(X=1) = \frac{\binom{2}{1}\cdot\binom{4}{1}}{\binom{6}{2}} = \underline{8/15},\quad f(2) = P(X=2) = \frac{\binom{2}{2}\cdot\binom{4}{0}}{\binom{6}{2}} = \underline{1/15}$$

x	0	1	2
$f(x)$	6/15	8/15	1/15



$p = D/N = 1/3$, $\mathrm{E}(X) = n\cdot p = 2\cdot(1/3) = \underline{2/3}$

$\mathrm{Var}(X) = np\cdot(1-p)\cdot(N-n)/(N-1) = (2/3)\cdot(1-(2/3))\cdot(6-4)/(6-1) = (4/3)^2/5$

$\Rightarrow \mathrm{std}(X) = \sqrt{Var(X)} = \frac{4}{3\sqrt{5}} = \underline{,596}$

Oppgave 4.4 En skole har 30 elever i videregående kurs II. De fordeler seg med 20 elever på regnskapslinjen og 10 elever på markedsføringslinjen. Blant disse elevene blir det gjort et tilfeldig utvalg på 8. La X være antall elever i utvalget fra markedsføringslinjen.

a) Begrunn at f(x) = $P(X=x)$

$$= \frac{\binom{10}{x}\binom{20}{8-x}}{\binom{30}{8}},\quad x = 0,1,2,\ldots,8$$

b) Finn E(X) og std(X)

c) Finn sannsynligheten for at 5 elever i utvalget kommer fra markedsføringslinjen.

Løsningsforslag:

a) Tilfeldig trekning uten tilbakelegging, to utfall per. trekning $\Rightarrow$ hypergeometrisk fordeling med parametre N = 30 (totalt antall i klassen), D = 20 (totalt antall fra markedsføringslinjen) og n = 8 (antall i utvalg).

b) $p = D/N = 10/30 = 1/3$, $\mathrm{E}(X) = np = 8\cdot(1/3) = \underline{8/3}$

$\mathrm{Var}(X) = np\cdot(1-p)\cdot(N-n)/(N-1) = (16/3)\cdot(1-(2/3))\cdot(30-8)/29 =$

$(4/3)^2\cdot(22/29) \qquad \Rightarrow \qquad \mathrm{std}(X) = \frac{4}{3}\sqrt{\frac{22}{29}} = \underline{1{,}16}$

c) $f(5) = P(X=5) = \dfrac{\binom{10}{5}\cdot\binom{20}{3}}{\binom{30}{8}} = \dfrac{252\cdot 1140}{5852925} = \underline{{,}049}$

Oppgave 4.5 Fødselsstatistikken forteller oss at i det lange løp er en av 60 fødsler tvillingfødsler. Av 500 fødsler hva er:

a) P(10 tvillingfødsler)
b) P(minst 10 tvillingfødsler)
c) I en liten bygd i Valdres har det i de siste årene tilsammen blitt født 9 gutter og en jente. Nå forventer folk i bygda seg relativt mange jenter i de nærmeste årene. Hvordan vil du som statistiker forholde deg til denne forventningen? Forklar.
d) Hva er sannsynligheten for at det i løpet av de 10 neste fødslene blir født minst 8 jenter?

Løsningsforslag: $p = 1/60 << 1$, $\lambda = np = 500/60 = 8{,}3$, X = # tvillingfødsler tn. Po(8,3)

a) $P(X=10) = \dfrac{\lambda^{10}}{10!}\cdot e^{-25/3} = \dfrac{(25/3)^{10}}{10!}\cdot e^{-25/3} = \underline{{,}107}$

b) Tabell: $P(X \geq 10) = 1 - P(X \leq 9) = 1 - {,}679 = \underline{{,}32}$ (Matlab: ,3255)

c) Falsk forventning hvis vi antar at sannsynlighet for kjønn til hver ny fødsel er uavhengig av tidligere fødsler (antagelse om uavhengighet).

d) Antar p = P(jente ved hver ny fødsel) = 0,5, samt uavhengige kjønn for fremtidige fødsler. Da vil X = antall jenter i løpet av de 10 neste fødslene være binomisk fordelt Bino(10, ,5).

$P(X \geq 8) = 1 - P(X \leq 7) = 1 - {,}945 = \underline{{,}055}$

Oppgave 4.6 Til en restaurant ankommer gjennomsnittlig 10 gjester mellom kl. 20.00 og 21.00. Forutsett at gjestene ankommer i henhold til Poisson-fordelingen.

a) Hva er sannsynligheten for at det ankommer akkurat 10 gjester i dette tidsrommet en tilfeldig kveld?

b) Hva er sannsynligheten for at det ankommer 12 eller flere gjester i dette tidsrommet?

Løsningsforslag: X = antall gjester som ankommer kl. 20-21, $X \sim \text{Po}(10)$.

a) $P(X = 10) = \dfrac{10^{10}}{10!} \cdot e^{-10} = \underline{,125}$

b) $P(X \geq 12) = 1 - P(X \leq 11) = 1 - ,697 = \underline{,303}$

Oppgave 4.7 I året 1919 utførte to engelske forskere (Greenwood og Woods) en studie over forekomsten av arbeidsulykker blant kvinner ansatt i ammuni

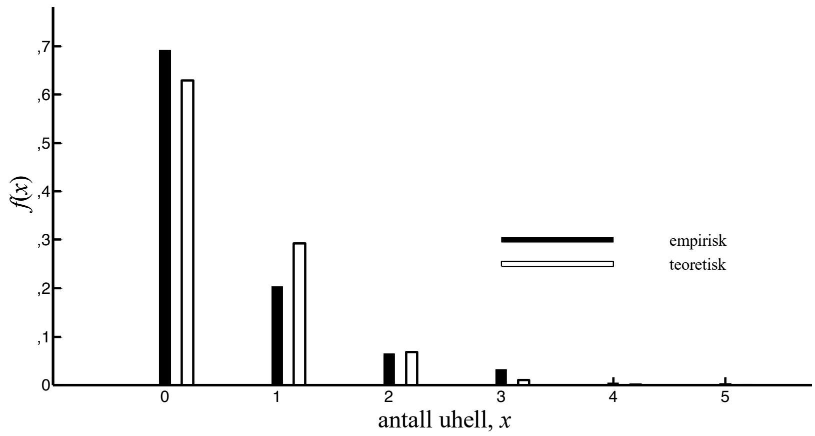
sjonsfabrikker. De tok for seg en gruppe på 648 kvinner og registrerte antall uhell hver kvinne var utsatt for i løpet av en periode på 13 måneder. De fikk følgende resultat:

Antall uhell	0	1	2	3	4	5
antall kvinner	448	132	42	21	3	2

a) Tegn stolpediagram over fordelingen og beregn gjennomsnitt og standardavvik.

b) Tilpass en Poisson-fordeling og tegn inn den beregnede fordeling i samme diagram som den observerte. Beskriv det avviket du ser mellom de to fordelingene. Hva slags forklaring kan et slikt avvik ha?

Løsningsforslag: X betegner antall uhell.



a) Stolpediagrammet over den empiriske fordelingen er vist i figuren ovenfor med svarte stolper.
Middelverdi: $\bar{x} = \frac{1}{n} \Sigma x_i \cdot f_i = \frac{1}{648} \cdot (0 \cdot 448 + 1 \cdot 132 + \cdots + 5 \cdot 2) = \underline{0,4645}$

$\overline{x^2} = \frac{1}{n}\Sigma x_i^2 \cdot f_i = \frac{1}{648}\cdot(0^2\cdot 448 + 1^2\cdot 132 + \cdots + 5^2\cdot 2) = ,9059$

Standardavvik: , $s = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{,9059 - ,4645^2} = \underline{,83}$

b) Bruker $\lambda^* = \bar{x} = ,4645$ som estimator for λ i Poisson-fordelingen Po(λ) som skal tilpasses:

$$f^*(x) = \frac{(\lambda^*)^x}{x!}\cdot e^{-\lambda^*}, \quad x = 0,1,\ldots$$

For x-verdiene 0,1,2,3,4 og 5 får vi henholdsvis f^*-verdiene 0,6284, 0,2919, 0,0678, 0,0105, 0,0012, og 0,0001. Den teoretiske fordelingen er vist i figuren ovenfor med hvite stolper. Den viser betydelig lavere sannsynlighet for tre uhell eller mer sammenlignet med empirisk stolpediagram. Avviket kan skyldes at vi ikke kan neglisjere ulykker der flere kvinner er involvert samtidig.

Oppgave 4.8 I en matforretning selges det gjennomsnittlig 7 litersglass med eksklusive grønne oliven pr. dag. Hvor mange slike glass bør butikken ha i hylla når den åpner om morgenen for at det høyst skal være 5 % sannsynlighet for at hylla blir tom i løpet av dagen?

Løsningsforslag: X = antall glass som etterspørres en tilfeldig dag, $X \sim$ Po(7). La k betegne antall glass i hylla.

$$P(X \geq k) \leq ,05 \;\Rightarrow\; 1 - P(X \leq k-1) \leq ,05 \;\Rightarrow\; P(X \leq k-1) \geq ,95$$

Fra Poisson-tabell med λ = 7 finner vi at $P(X \leq 11) = ,947$, $P(X \leq 12) = ,973$. Vi får derfor at $k - 1 = 12$, butikken bør ha k = 13 glass i hylla på morgenen.

Oppgave 4.9 Et bryggeri har funnet ut at gjennomsnittlig 0,2 % av pilsflaskene inneholder for sterk alkoholprosent. Hver dag tas det en stikkprøve på 100 flasker som kontrolleres. Dersom et visst antall k eller flere flasker viser for stor alkoholprosent stoppes og justeres prosessen der alkohol tilsettes.

a) Hva er sannsynligheten for stans i denne prosessen hvis $k = 1$?
b) Hva må k minst være for at sannsynligheten for stans i prosessen skal være høyst 0,3 %?

Løsningsforslag: $p = 0{,}002$, $n = 100$, $\lambda = np = ,2$, X = antall flasker en tilfeldig dag som inneholder for sterk alkolholprosent. Antar X ~Po(,2).

a) $P(X \geq k = 1) = 1 - P(X = 0) = 1 - \frac{,2^0}{0!}\cdot e^{-,2} = \underline{,181}$

b) $P(X \geq k) \leq ,003 \;\Rightarrow\; 1 - P(X \leq k-1) \leq ,003 \;\Rightarrow\; P(X \leq k-1) \geq ,997$

Fra Poisson-tabell med λ = ,2 finner vi at $P(X \leq 1) = ,982$, $P(X \leq 2) = ,999$. Vi får derfor at $k - 1 = 2$, dvs. k = 3.

Oppgave 4.10 På et sentralbord kommer det inn gjennomsnittlig 2,2 innringninger pr. kvarter.

a) Hva er sannsynligheten for at det kommer inn 10 innringninger i løpet av en time?
b) Hva er sannsynligheten for minst 8 innringninger i løpet av en time?
c) Hvilke forutsetninger må du gjøre for å løse a) og b)?

Løsningsforslag: X = antall innringninger pr. time, X er da tn. Po(2,2·4) = Po(8,8).

a) $P(X=10)=\frac{8{,}8^{10}}{10!}\cdot e^{-8{,}8}=\underline{,116}$

b) $P(X\geq 8)=1-P(X\leq 7)=1-{,}348=\underline{,652}$

c) Forutsetninger: Kun én oppringning av gangen. Forventet antall oppringninger pr. tidsintervall proporsjonal med tidsintervallets lengde. Uavhengige oppringninger.

Oppgave 4.11 a) La betegnelsen X ~ Bino(n,p) bety at X er binomisk fordelt med parametre n og p. Bruk utdelte tabeller og finn de tallene som mangler nedenfor:

$P(X \leq\ ) = 0{,}724$ når
$X \sim$ Bino(12, 0,3)

$P(X =\quad) = 0{,}273$
når $X \sim$ Bino(7, 0,5)

$P(X = 9) = ?$
når $X \sim$ Bino(16, 0,8)

b) I et lotteri er det bare 12 lodd igjen, og hele 3 av disse gir premie. Vis at sannsynligheten for å vinne er 75 % dersom du kjøper 4 lodd.

c) La X være binomisk fordelt med parametre n = 300 og p = 0,030. Finn en tilnærmet verdi for $P(X \geq 9)$.

Løsningsforslag:

a) Fra tabell for Bino(12, 0,3) finner vi: $P(X\leq \underline{4})={,}723$, tallet som mangler er $\underline{4}$.

Fra tabell for Bino(7, 0,5) finner vi: $P(X\leq \underline{3})-P(X\leq 2)={,}500-{,}227={,}273$, tallet som mangler er $\underline{3}$.

Fra tabell over Bino(16, 0,8) finner vi:
$P(X=9)=P(X\leq 9)-P(X\leq 8)={,}027-{,}007=\underline{,02}$, tallet som mangler er $\underline{,02}$.

b) X = antall lodd som gir gevinst av n = 4. X er da hypergeometrisk fordelt (uten tilbakelegging, uordna rekkefølge) hyp(4,3,12), og vi får:

P(gevnist) = $P(X\geq 0)=1-P(X=0)=\frac{\binom{3}{0}\cdot\binom{9}{4}}{\binom{12}{4}}=1-{,}25={,}75$, q.e.d.

c) X er tilnærmet Poisson-fordelt Po(np) = Po(300·,03) = Po(9).

Fra Po(9)-tabell: $P(X \geq 9) = 1 - P(X \leq 8) = 1 - ,456 = \underline{,544}$

Oppgave 4.12 Definer *X* som antall seksere etter 18 kast med rettferdig terning.

a) Begrunn hvorfor *X* er binomisk fordelt.

b) Sett opp et uttrykk for fordelingen, *f*(x) = *P*(*X* = *x*).

c) Bestem forventning, μ = E(*X*), og standardavvik, std(*X*).

d) Bestem sannsynlighetene for

- minst én sekser
- ingen seksere
- bare seksere
- 3 seksere
- minst 2 og høyst 4 seksere

Løsningsforslag:

a) To utfall pr. kast, $p = P(6\text{'er}) = 1/6$ i hvert kast, uavh. utfall $\Rightarrow X \sim \text{Bino}(18, 1/6)$.

b) $f(x) = P(X = x) = \binom{18}{x} \cdot \left(\frac{1}{6}\right)^x \cdot \left(\frac{5}{6}\right)^{18-x}, \ x = 0,1,\ldots,18$

c) $\mu = \text{E}(X) = np = 18 \cdot 1/6 = \underline{3}$,

$\text{Var}(X) = np(1 - p) = 3 \cdot 5/6 = 5/2, \quad \text{std}(X) = \sqrt{Var(X)} = \sqrt{\frac{5}{2}} = \underline{1{,}58}$

d) $P(\text{minst én sekser}) = 1 - P(X = 0) = 1 - (5/6)^{18} = \underline{,962}$

$P(\text{ingen seksere}) = 1 - P(\text{minst én sekser}) = 1 - ,962 = \underline{,038}$

$P(\text{bare seksere}) = P(X = 18) = \binom{18}{18} \cdot \left(\frac{1}{6}\right)^{18} \cdot \left(\frac{5}{6}\right)^0 = \underline{10^{-14}}$

$P(3 \text{ seksere}) = \binom{18}{3} \cdot \left(\frac{1}{6}\right)^3 \cdot \left(\frac{5}{6}\right)^{15} = \underline{,245}$

$P(\text{minst 2 og høyst 6 seksere}) = P(X = 2) + P(X = 3) + P(X = 4) =$

$\binom{18}{2} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^{16} + ,245 + \binom{18}{4} \cdot \left(\frac{1}{6}\right)^4 \cdot \left(\frac{5}{6}\right)^{14} = ,230 + ,245 + ,184 = \underline{,66}$

Oppgave 4.13 Ved hjelp av en Geigerteller kan en telle antall partikler som utsendes fra et radioaktivt materiale. En forsker har foretatt slike tellinger over en rekke tidsperioder, hver på 6 sekunder. Han har fått resultatene vist i neste tabell.

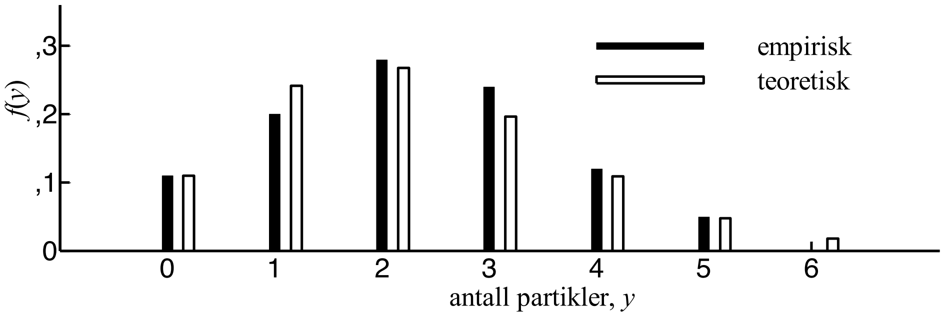
Antall partikler Y	Antall tidsperioder med Y partikler
0	11
1	20
2	28
3	24
4	12
5	5
6	0
	Totalt: 100

a) Lag et stolpediagram over fordelingen av antall partikler.

b) Finn gjennomsnittlig antall partikler pr. tidsperiode.

c) Anta at Y er Poisson-fordelt med parameter $\mu = 2{,}21$. Beregn forventet hyppighet for de enkelte Y-verdier fra 0 til 6. Tegn fordelingen inn i samme diagram som du brukte under punkt a). Synes du det ser ut som om Poisson-fordelingen passer rimelig bra?

<u>Løsningsforslag</u>:



a) Empirisk relativ frekvens stolpediagram vist i figuren ovenfor (svarte stolper).

b) $E(Y) = \Sigma\, y{\cdot}f(y) = 0{\cdot},11 + 1{\cdot},20 + 2{\cdot},28 + 3{\cdot}.24 + 4{\cdot},12 + 5{\cdot},05 + 6{\cdot}0 = \underline{2{,}21}$

c) Antar $f(y) = \dfrac{2{,}21^y}{y!} \cdot e^{-2{,}21}, \quad y = 0,1,2,\ldots$

y	0	1	2	3	4	5	6
$f(y)$	,110	,242	,268	,197	,109	,048	,018

De teoretiske verdiene fra tabellen er tegnet inn (hvite stolper) ved siden av de empiriske i figuren ovenfor. Som vi ser er det en rimelig bra tilpasning.

Oppgave 4.14 En person (A) hevdet at han kunne smake forskjell på viner fra forskjellige produsenter innen *samme* vindistrikt. For å teste dette ble A gitt 5 smaksprøver. Han fikk *hver gang* oppgitt at prøven kom fra *en* av *to* navngitte produsenter og ble bedt om å oppgi hvilken. La Y være antall riktige svar på de 5 forsøkene.

a) Anta at A bare gjetter. Forklar *kort* hvorfor punktsannsynligheten til Y da blir

$$P(Y = y) = \binom{5}{y} \cdot \left(\tfrac{1}{2}\right)^5$$

b) Hva er sannsynligheten for at han svarer riktig samtlige ganger dersom han gjetter?

c) Anta at A klarer 4 av 5 riktige. Kan vi ut fra dette være rimelig sikre på at han ikke gjetter? Begrunn svaret.

Løsningsforslag:

a) Gjetting gir at P(riktig svar) = ,5 i hvert forsøk. $n = 5$ uavhengige gjettinger $\Rightarrow$ Y = antall riktige svar ~ Bino(5, ,5). Vi får derfor:

$$f(y) = P(Y = y) = \binom{5}{y} ,5^y (1-,5)^{5-y} = \binom{5}{y} \cdot \left(\tfrac{1}{2}\right)^5, \; y = 0,1,\ldots,5$$

b) $P(Y = 5) = \binom{5}{5} \cdot \left(\tfrac{1}{2}\right)^5 = \tfrac{1}{32} = \underline{,0313}$

c) $P(Y \geq 4) = P(Y = 4) + P(Y = 5) = \binom{5}{4} \cdot \left(\tfrac{1}{2}\right)^5 + \tfrac{1}{32} = \tfrac{5+1}{32} = \tfrac{3}{16} = \underline{,188}$

Siden det er hele 18,8 % sannsynlighet for å få minst 4 rette svar ved ren gjetning, er det på grensen å si at vi er rimelig sikre på at han ikke gjetter.

Oppgave 4.15 Det er velkjent at krabber av og til mister ett eller flere av sine ben. I denne oppgaven skal vi studere et observasjonsmateriale som belyser dette fenomen. Det er observert n = 1344 tilfeldig utvalgte hankrabber. Tabellen nedenfor viser hvor mange av dem som hadde mistet henholdsvis 0,1,2,3,4, eller 5 av en på forhånd spesifisert gruppe på i alt 5 ben.

Antall ben som mangler: ↓

0	1	2	3	4	5	
1137	155	43	8	1	0	1344

Observert frekvens: ↑ Sum ↑

a) Tegn et stolpediagram av observasjonsmaterialet.

I det følgende (punktene b)-c)) skal vi undersøke om dataene støtter en hypotese om at observasjonene er binomisk fordelte med samme binomiske fordeling. La Y_i = antall ben som mangler hos i'te krabbe. Anta at $Y_1,\ldots,Y_n$ alle har sannsynlighetsfordelingen (punktsannsynligheten)

$$P(Y_i = y) = \binom{5}{y} p^y q^{5-y}, \; y = 0,1,2,3,4,5$$

der $q = 1 - p$

b) Hvilken tolkning har parameteren p?

Forklar hvorfor det ut fra dataene er rimelig å anslå p som tallet 269/6720 = 0,040?

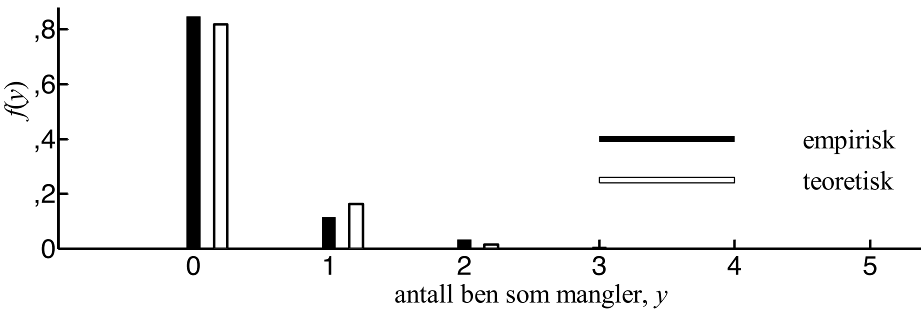
Hvilken tolkning har telleren og hvilken tolkning har nevneren i denne brøken?

c) Beregn en forventet (teoretisk) frekvenstabell med p fra b). Tegn frekvensene inn på samme diagram som i a).

Det oppgis at

$$\binom{5}{0}=\binom{5}{5}=1,\ \binom{5}{1}=\binom{5}{4}=5,\ \binom{5}{2}=\binom{5}{3}=10$$

Løsningsforslag:



a) Empirisk stolpediagram er vist i figuren ovenfor med svarte stolper.

b) p = forventet antall ben som mangler hos en tilfeldig krabbe i. Det rimelig å anslå p som gjennomsnittlig antall ben som mangler. Teller: totalt antall ben som mangler, nevner: totalt antall ben vi studerer.

c) $$p^* = \frac{155\cdot 1 + 43\cdot 2 + 8\cdot 3 + 1\cdot 4}{5\cdot 1344} = ,040$$

Vi antar videre at antall ben som mangler, Y, er binomsik fordelt, $Y \sim$ Bino(5, ,040), og vi kan da beregne teoretisk frekvenstabell fra formelen

$f(y) = \binom{5}{y} \cdot ,04^{y} \cdot ,96^{5-y}$, $y = 0,1,\ldots,5$, som gir følgende verdier:

y	0	1	2	3	4	5
$f(y)$	,815	,170	,014	$6{,}0\cdot 10^{-3}$	$1{,}2\cdot 10^{-5}$	10^{-7}

De teoretiske frekvensene er tegnet inn i diagrammet sammen med de empiriske.

Oppgave 4.16 Tabellen nedenfor angir for 200 større amerikanske byer antall dødsfall forårsaket av en bestemt sykdom i løpet av ett år.

# døde	0	1	2	3	4
# byer	93	70	26	8	3

Vi skal i denne oppgaven undersøke om en Poisson-modell passer til dataene. Det vil si at dersom Y er lik antall dødsfall i en gitt by, skal vi undersøke om Y følger en punktsannsynlighet av typen

$$P(Y=y)=\frac{\lambda^y}{y!}e^{-\lambda},\ y=0,1,\ldots \qquad (*)$$

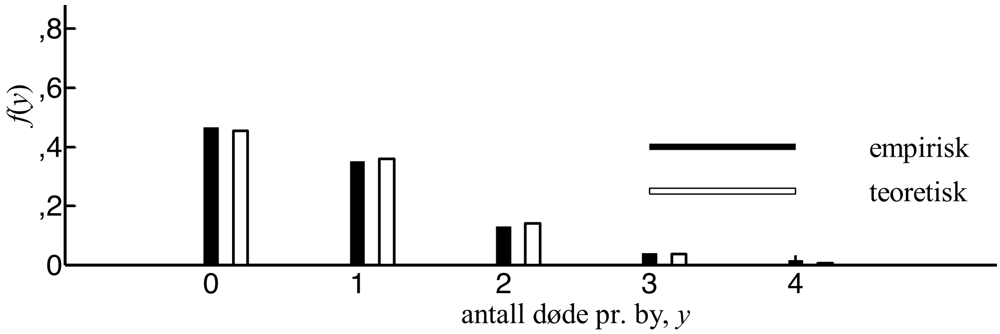
a) Lag et stolpediagram av dataene.

b) Hvilken tolkning har tallet 158/200 = ,79?

c) Estimer λ i punktsannsynligheten (*) ved å sette λ lik tallet i b). Beregn $P(Y=y)$ for y = 0,1,2,3,4 med denne verdi av λ.

d) Beregn det forventede (teoretiske) antall byer med 0,1,2,3 eller 4 dødsfall gitt at Y følger punktsannsynligheten (*) med den estimerte verdi av λ. Lag et stolpediagram av det forventede antall på samme figur som i a).

Løsningsforslag:



a) Stolpediagram over de empiriske frekvensene er vist i figuren ovenfor med svarte stolper.

b) Tolkning: gjennomsnittlig antall døde pr. by (teller: totalt antall døde, nevner: totalt antall byer).

c) $\lambda^* = \dfrac{0\cdot 93+1\cdot 70++2\cdot 26+3\cdot 8+4\cdot 3}{200}=\dfrac{158}{200}=\underline{,79}$

Tilpasset Poisson-fordeling $f(y)=P(Y=y)=\dfrac{,79^y}{y!}\cdot e^{-,79},\ y=0,1,\ldots$ gir verdiene

y	0	1	2	3	4
$f(y)$	,454	,359	,142	,037	,007

d) Forventet antall byer med i dødsfall med basis i fordelingen i tabellen finnes ved å multiplisere sannsynlighetene med 200: For stigende verdier av i fra 0 til 4 får vi da følgende resultat: 91, 72, 28, 7 og 1.

Den teoretiske fordelingen er tegnet inn sammen med den empiriske fordelingen i figuren under pkt. a). Som vi ser er det en god tilpasning.

Oppgave 4.17 Fangst-gjenfangst-metoden er mye brukt for å anslå ville dyrebestanders størrelse. Den går i korthet ut på å sette ut feller flere ganger (netter). De fangete dyrene merkes og settes fri, og det er gjenfangstandelen som brukes til å anslå bestandsstørrelsen. Vi skal i denne oppgaven ikke studere selve metoden, men snarere de forutsetningene den bygger på. I den enkleste modellen antar en

i) Et dyrs adferd blir ikke påvirket av å bli fanget.
ii) Alle dyrene har samme sannsynlighet for å bli fanget hver gang (natt).

a) Anta at fellene blir satt ut et stort antall netter og at sannsynligheten for at et bestemt dyr skal gå i fella en natt er liten. Forklar hvorfor en på dette grunnlag kan anta at antall ganger et bestemt dyr blir fanget, *X*, er (tilnærmet) Poisson-fordelt.

b) I et lukket område hvor det var 135 (gjenkjennbare) kaniner gjennom hele fangstperioden, fikk Edwards & Ebenhardt i 1967 følgende fangstresultat:

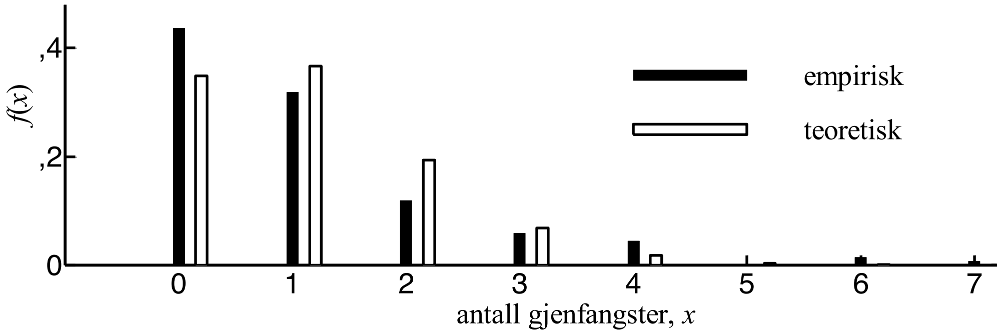
Antall ganger fangst: ↓

0	1	2	3	4	5	6	7	
59	43	16	8	6	0	2	1	135

Antall kaniner: ↑ Sum: ↑

Tilpass en Poisson-fordeling til dette materialet. Tegn den observerte fordeling og den tilpassede fordeling inn i samme stolpediagram. Er det tegn til systematiske avvik mellom observert og tilpasset modell? Hvordan vil du eventuelt karakterisere dette avviket? Diskuter kort mulige årsaker til dette avviket - hva kan være galt med antagelsene i) og ii)?

<u>Løsningsforslag</u>:



a) Begrunnelse for Poisson-tilnærming: 1) Under de gitte forutsetninger vil X være tilnærmet binomisk fordelt med parametre n = antall netter og $p = P$(ett og samme dyr blir fanget én natt). Siden n er stor og p liten, vil da X være tilnærmet Poisson-fordelt med parameter $\lambda = np$.

b) Vi skal tilpasse en Poisson-fordeling til data, og estimerer (anslår) da først parameteren l som er forventet antall gjenfangster, ved gruppert middelverdi med basis i data:

$$\lambda^* = \frac{0\cdot 59 + 1\cdot 43 + 2\cdot 16 + 3\cdot 8 + 4\cdot 6 + 5\cdot 0 + 6\cdot 2 + 7\cdot 1}{135} = 1{,}05$$

Tilnærmet Poisson-fordeling: $f(x) = \dfrac{1{,}05^x}{x!}\cdot e^{-1{,}05},\ \ x = 0{,}1{,}\ldots$ som gir verdiene

x	0	1	2	3	4	5	6	7
$f(x)$	,349	,367	,193	,068	,018	,004	,001	,0001

Empirisk og tilpasset modell er plottet i figuren nedenfor. Den teoretiske tilpasningen later til å underestimere sannsynligheten for ingen gjenfangst samt for de største gjenfangstene (x = 4-7). En forklaring på underestimeringen av sannsynligheten for ingen gjenfangst kan ligge i at dyrene lar seg påvirke av at de har gått i fella, og dermed viser mer forsiktighet (feil med antagelse i)). Dessuten kan det være at dyrene oppdager andre som er gått i fella, og av den grunn blir mer forsiktige, hvilket strider med antagelse ii). Vedrørende underestimeringen av mange gjenfangster, kan en forklaring være at når dyrene har blitt fanget noen ganger og skjønner at det ikke er farlig, viser de mindre forsiktighet enn normalt med å bli fanget på nytt.

Oppgave 4.18 Antallet tankbåter som kommer til en bestemt havn er Poisson-fordelt med parameter lik 2,0. Havnen kan maksimalt betjene 3 tankbåter pr. dag. De tre første fartøyene som kommer, blir betjent, mens de neste fartøyene blir omdirigert til en annen havn.

a) Hva er sannsynligheten for at fartøyer skal bli omdirigert en bestemt dag?

b) Hvor stor kapasitet må havnen ha om den med minst 95 % sannsynlighet skal kunne betjene samtlige tankbåter som kommer en bestemt dag?

Løsningsforslag:

a) $P(X > 3) = 1 - P(X \leq 3) = 1 - ,857 = \underline{,143}$

b) $P(X \leq k) \geq ,95$ Fra Poisson-tabell med $\lambda = 2$: $P(X \leq 4) = ,947$, $P(X \leq 5) = ,983$

Følgelig: kapasiteten må være k = <u>5</u>

Oppgave 4.19 I en bok er det 220 trykkfeil tilfeldig spredt på bokens 200 sider. Vi antar at antall trykkfeil pr. side i boken er Poisson-fordelt. Finn sannsynligheten for at det på en tilfeldig side er:

a) ingen trykkfeil,
b) akkurat en trykkfeil,
c) høyst to trykkfeil,
d) minst to trykkfeil.

Løsningsforslag: X = # trykkfeil pr. side, $X \sim \text{Po}(\lambda = 220/200) = \text{Po}(1{,}1)$.

$$f(x) = P(X = x) = \frac{1{,}1^x}{x!} \cdot e^{-1{,}1}, \quad x = 0{,}1{,}\ldots$$

a) $P(X = 0) = \frac{1{,}1^0}{0!} \cdot e^{-1{,}1} = \underline{,333}$

b) $P(X = 1) = \frac{1{,}1^1}{1!} \cdot e^{-1{,}1} = \underline{,366}$

c) $P(X \leq 2) = \underline{,900}$ (tabell)

d) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) = 1 - ,333 - ,366 = \underline{,301}$

Oppgave 4.20 Strålingen fra en radioaktiv kilde antas å være Poisson-fordelt. I en periode på 2 timer ble det registrert at kilden sendte ut 492 α-partikler.

a) Hva er sannsynligheten for at det blir registrert nøyaktig 3 partikler i det neste minuttet?

b) Hva er sannsynligheten for at det ikke blir registrert flere enn 3 partikler i det neste minuttet?

Løsningsforslag: X = antall partikler som blir registrert neste minutt. Siden kilden sender ut ca. 492 a-partikler iløpet av 2 timer, vil forventet antall partikler som sendes ut pr. minutt være ca. 492/120 = 4,1. Vi antar derfor at $X \sim \mathrm{Po}(4{,}1)$.

a) $P(X=3) = \dfrac{4{,}1^3}{3!} \cdot e^{-4{,}1} = \underline{,190}$

b) $P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3) =$

$e^{-4{,}1} \cdot (1 + 4{,}1 + \frac{1}{2} \cdot 4{,}1^2) + {,}190 = \underline{,414}$

Oppgave 4.21 Antall syklister som passerer et bestemt veikryss i løpet av en tilfeldig 2-minutters periode antas å være Poisson-fordelt med $\lambda = 3$.

a) Finn sannsynligheten for at
   i) ingen syklister passerer i løpet av 2-minutters-perioden,
   ii) minst 3 syklister passerer i løpet av 2-minutters-perioden.

b) Finn forventning og standardavvik til antall syklister som passerer i løpet av 2-minutters-perioden.

c) Finn sannsynligheten for at eksakt 3 syklister passerer i løpet av en 4-minutters periode. Vi antar at hendelser i ikke-overlappende tidsintervaller er uavhengige.

Løsningsforslag: X er antall syklister som passerer.

a) i) $P(X=0) = \dfrac{3^0}{0!} \cdot e^{-3} = \underline{,0498}$

ii) $P(X \geq 3) = 1 - P(X \leq 2) = 1 - {,}423 = \underline{,577}$

b) $\mathrm{E}(X) = \lambda = \underline{3}$, $\mathrm{std}(X) = \sqrt{\lambda} = \sqrt{3} = \underline{1{,}73}$

c) Y = antall syklister som passerer i løpet av en 4 minutters periode $\Rightarrow$ antar $Y \sim \mathrm{Po}(2\lambda) = \mathrm{Po}(6)$.

$P(Y=3) = \dfrac{6^3}{3!} \cdot e^{-6} = \underline{,0892}$

Oppgave 4.22 En komponent som benyttes i et elektronisk apparat har vist seg å være defekt med en sannsynlighet på 5,0 %. Vi kaller denne komponenten for «den kritiske komponenten». De øvrige komponentene i apparatet er alltid feilfrie, og ingen komponenter blir ødelagt i forbindelse med monteringen.

Bestem sannsynligheten for at firmaet skal kunne levere 75 apparater uten å bestille inn flere kritiske komponenter dersom antall kritiske komponenter på lager er

a) 75,
b) 77.

Produsenten av apparatet har funnet ut at det fra en spesiell by kommer gjennomsnittlig 1,8 ordrer pr. dag.

c) Bestem sannsynligheten for at det en dag vil komme inn eksakt 2 ordrer fra byen.

Det blir sendt varer til byen hver gang det har hopet seg opp 3 eller flere ubesørgede ordrer.

d) En morgen er ordrelisten fra byen tom. Bestem sannsynligheten for at det vil bli sendt varer til byen før det er gått 2 dager.

Produsenten går over til å pakke apparatene i en ny emballasje, hvor det ved hver forsendelse er en sannsynlighet på 0,10 for at det skal bli skade på emballasjen.

e) Bestem sannsynligheten for at første skade på emballasjen vil skje i den 4. forsendelsen.

f) Bestem sannsynligheten for at produsenten vil oppleve emballasjeskade i løpet av de fire første forsendelsene.

Løsningsforslag: X = antall etterpurte komponenter som er defekte, $X \sim$ Bino(n, ,05).

a) $n = 75,\ P(X=0) = \binom{75}{0} \cdot ,05^0 \cdot ,95^{75} = \underline{,0213}$

b) $n = 77,\ P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) =$

$\binom{77}{0} \cdot ,05^0 \cdot ,95^{77} + \binom{77}{1} \cdot ,05^1 \cdot ,95^{76} + \binom{77}{2} \cdot ,05^2 \cdot ,95^{75} = ,0193 + ,0781 + ,1561 = \underline{,2535}$

Antall ordrer pr. dag fra spesiell by antas Poisson-fordelt med parameter $\lambda = 1,8$.

c) $P(2 \text{ ordrer på en dag}) = \dfrac{1,8^2}{2!} \cdot e^{-1,8} = \underline{,268}$

d) La Y = antall ordrer fra byen på 2 dager, dvs. $Y \sim$ Po(2λ) = Po(3,6).
P(vareforsendelse til byen) = $P(Y \geq 3) = 1 - P(Y \leq 2) = 1 - ,303 = \underline{,697}$

e) $p = P$(emballasjeskade pr. forsendelse) = ,1, A = skade i tilfeldig forsendelse.
P(1. skade i 4. forsendelse) = $P(A^C A^C A^C A) = ,9^3 \cdot ,1 = \underline{,0729}$

f) La N = antall emballasjeskader i løpet av 4 forsendelser, N ~ Bino(4, ,1).
$P(N \geq 1) = 1 - P(N=0) = 1 - ,9^4 = \underline{,3439}$

Oppgave 4.23 Man ønsker å undersøke utbredelsen av en bestemt sykdom i en befolkning. Til dette formål er 1000 personer trukket ut tilfeldig fra befolkningen (som totalt er mye større enn 1000). Videre har man en diagnosemetode for å avgjøre om en person er syk eller frisk. De 1000 personene blir undersøkt med metoden. La X være antall personer som blir diagnostisert til å være syke.

a) Forklar kort hvorfor vi kan bruke binomisk fordeling for X.

Diagnosemetoden er ikke helt pålitelig: Metoden kan gi konklusjonen at en frisk person er syk og at en syk person er frisk med visse sannsynligheter. La begivenheten A bety at en tilfeldig person blir diagnostisert til å være syk. Man har ved grundige undersøkelser tidligere funnet at $P(A \mid \text{personen er syk}) = 0{,}9$, og at $P(A \mid \text{personen er frisk}) = 0{,}05$.

b) Dersom andelen π av befolkningen er syke (og andelen $1-\pi$ er friske), vis at sannsynligheten for at en tilfeldig person blir diagnostisert til å være syk er: $p = 0{,}85\pi + 0{,}05$.

Løsningsforslag:

a) Vi har en (uten tilbakelegging, uordna rekkefølge)-situasjon. Fordi befolkningen har mange flere innbyggere, N, enn utvalgets størrelse, n ($n << N$), vil det gi minimal forskjell å foreta analysen som om utvalget var trukket med tilbakelegging. Siden utvalget er tilfeldig, vil det også være rimelig å anta at vi kan behandle diagnosene som om de var uavhengige observasjoner. Det er også rimelig å anta at $p = P(A)$ er rimelig konstant for hvert individ fordi $n << N$. Det følger da at vi med rimelighet kan anta at X er binomisk fordelt med parametre $n = 1000$ og ukjent p.

b) $P(A) = P(A \mid \text{syk})\cdot P(\text{syk}) + P(A \mid \text{frisk})\cdot P(\text{frisk}) =,9\cdot\pi + ,05\cdot(1-\pi) =,85\pi + ,05$, q.e.d.

Kapittel 5

Kontinuerlige fordelinger

Oppgave 5.1 La X_1, X_2 og X_3 være 3 stokastisk uavhengige variabler der $E(X_i) = (-1)^i$ og $Var(X_i) = i$; $i = 1, 2, 3$. La Y være definert som $Y = 2X_1 - 3X_2 - X_3$. Bestem forventning og standardavvik til Y.

<u>Løsningsforslag</u>:

$E(Y) = 2 \cdot E(X_1) - 3 \cdot E(X_2) - E(X_3) = 2 \cdot (-1)^1 - 3 \cdot (-1)^2 - E(-1)^3 = -2 - 3 + 1 = \underline{-4}$

$Var(Y) = Var(2X_1 - 3X_2 - X_3) = 2^2 \cdot Var(X_1) + 3^2 \cdot Var(X_2) + Var(X_3) = 4 \cdot 1 + 9 \cdot 2 + 3 = 25$

$std(Y) = \sqrt{Var(Y)} = \sqrt{25} = \underline{5}$

Oppgave 5.2 La X_1 og X_2 være 2 stokastisk uvahengige $N(0,1)$-variabler. En tredje $N(0,1)$-variabel, X_3, er korrelert med både X_1 og X_2 ved at $Cov(X_1, X_3) = ,5$ og $Cov(X_2, X_3) = -0,5$.

Bestem forventning og varians til

$Y = X_1 + X_2 + X_3$, samt korrelajonskoeffisienten mellom X_1 og X_3.

<u>Løsningsforslag</u>:

$EY = E(X_1 + X_2 + X_3) = EX_1 + EX_2 + EX_3 = 0 + 0 + 0 = \underline{0}$

$Var(Y) = Var(X_1 + X_2 + X_3) = Var(X_1) + Var(X_2) + Var(X_3) + 2 \cdot Cov(X_1, X_3) + 2 \cdot Cov(X_2, X_3) = 1 + 1 + 1 + 2 \cdot ,5 + 2 \cdot (-,5) = \underline{3}$

$Corr(X_1, X_3) = Cov(X_1, X_3)/(std(X_1) \cdot std(X_3) = ,5/(1 \cdot 1) = \underline{,5}$

Oppgave 5.3 La Z være en $N(0,1)$-variabel, og finn ved hjelp av tabell følgende sannsynligheter:

a) $P(-3 < Z < 3)$ b) $P(-2 < Z < 2)$ c) $P(-1 < Z < 1)$
d) $P(-5 < Z < 5)$ e) $P(2,11 < Z < 3,38)$ f) $P(-1,74 < Z < -0,27)$

<u>Løsningsforslag</u>:

a) $P(-3 < Z < 3) = P(Z < 3) - P(Z < -3) = \Phi(3) - \Phi(-3) = \Phi(3) - (1 - \Phi(3)) = 2 \cdot \Phi(3) - 1 = 2 \cdot ,9987 - 1 = \underline{,997}$

b) $P(-2 < Z < 2) = P(Z < 2) - P(Z < -2) = 2 \cdot \Phi(2) - 1 = 2 \cdot ,9772 - 1 = \underline{,954}$

c) $P(-1 < Z < 1) = P(Z < 1) - P(Z < -1) = 2 \cdot \Phi(1) - 1 = 2 \cdot ,8413 - 1 = \underline{,683}$

d) $P(-5 < Z < 5) = P(Z < 5) - P(Z < -5) = 2 \cdot \Phi(5) - 1 \approx 2 \cdot 1 - 1 = \underline{1,000}$

e) $P(2{,}11 < Z < 3{,}38) = \Phi(3{,}38) - \Phi(2{,}11) = {,}9996 - {,}9826 = \underline{{,}0170}$

f) $P(-1{,}74 < Z < -{,}27) = \Phi(-{,}27) - \Phi(-1{,}74) = {,}3936 - {,}0409 = \underline{{,}353}$

Oppgave 5.4 La Z være en $N(0,1)$-variabel, og finn ved hjelp av tabell den z-verdi som er slik at følgende likheter blir tilfredsstilt:

a) $P(Z < z) = {,}95$ b) $P(-z < Z < z) = {,}95$ c) $P(Z > z) = 0{,}5$
d) $P(|Z| > z) = 0{,}05$ e) $P(-z < Z < z) = 0{,}01$

Løsningsforslag:

a) $P(Z < z) = {,}95$ Tabell: $P(Z < 1{,}64) = {,}9495$, $P(Z < 1{,}65) = {,}9505 \Rightarrow \underline{z \approx 1{,}645}$

b) $P(-z < Z < z) = {,}95 \Leftrightarrow 2 \cdot P(Z < z) - 1 = {,}95 \Leftrightarrow P(Z < z) = 1{,}95/2 = {,}975$
Tabell: $P(Z < 1{,}96) = {,}975$ $z = \underline{1{,}96}$

c) $P(Z > z) = {,}5 \Leftrightarrow 1 - P(Z < z) = {,}5 \Leftrightarrow P(Z < z) = {,}5 \Rightarrow z = \underline{0}$

d) $P(|Z| > z) = {,}05 \Leftrightarrow 1 - P(|Z| < z) = {,}05 \Leftrightarrow P(|Z| < z) = {,}95 \Leftrightarrow$
$P(-z < Z < z) = {,}95 \Rightarrow z = \underline{1{,}96}$ (se b))

e) $P(-z < Z < z) = {,}01 \Leftrightarrow 2P(Z < z) - 1 = {,}01 \Leftrightarrow P(Z < z) = 1{,}01/2 = {,}505$
Tabell: $P(Z < {,}01) = {,}5040$, $P(Z < {,}02) = {,}5080 \Rightarrow z \approx \underline{{,}013}$

Oppgave 5.5 La X være en $N(\mu,\sigma)$-variabel, og bestem følgende sannsynligheter:

a) $P(2 < X < 3)$ når $\mu = 2{,}5$ og $\sigma = 1$ b) $P(-5 < X < -4)$ når $\mu = 1$ og $\sigma = {,}5$
c) $P(X > 8)$ når $\mu = 0$ og $\sigma = 2$

Løsningsforslag:

a) $P(2 < X < 3) = P\left(\frac{2-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{3-\mu}{\sigma}\right) = P\left(\frac{2-2{,}5}{1} < Z < \frac{3-2{,}5}{1}\right) =$
$P(-{,}5 < Z < {,}5) = 2 \cdot \Phi({,}5) - 1 = 2 \cdot {,}6915 - 1 = \underline{{,}383}$

b) $P(-5 < X < -4) = \Phi\left(\frac{-4-1}{{,}5}\right) - \Phi\left(\frac{-5-1}{{,}5}\right) = \Phi(-10) - \Phi(-12) \approx \underline{0}$

c) $P(X > 8) = 1 - P(X < 8) = 1 - \Phi((8\text{-}0)/2) = 1 - \Phi(4) \approx 1 - 1{,}000 = \underline{0{,}000}$

Oppgave 5.6

a) Anta at X er $N(0,1)$. Finn $P(-0{,}47 \le X \le 0{,}94)$.
b) Anta at X er $N(0,1)$. Finn $P(X > 1{,}00)$.
c) Anta at X er $N(9,9)$. Finn $P(X > 2{,}00)$.
d) Anta at X er $N(9,9)$. Finn $P(5 \le X \le 11)$.

Løsningsforslag:

a) $P(-{,}47 \le X \le {,}94) = \Phi({,}94) - \Phi(-{,}47) = {,}8274 - {,}3192 = \underline{{,}507}$

b) $P(X > 1{,}00) = 1 - P(X < 1{,}00) = 1 - \Phi(1{,}00) = 1 - {,}8413 = \underline{{,}159}$

c) $P(X > 2{,}00) = 1 - P(X < 2{,}00) = 1 - P(Z < \frac{2{,}00-9}{9}) = 1 - \Phi(-{,}78) = 1 - {,}218 = \underline{{,}782}$

d) $P(5 \le X \le 11) = P(X \le 11) - P(X \le 5) = P(Z \le \frac{11-9}{9}) - P(Z \le \frac{5-9}{9}) =$
$P(Z \le {,}22) - P(Z \le -{,}44) = \Phi({,}22) - \Phi(-{,}44) = {,}5871 - {,}3300 \approx \underline{{,}257}$

Oppgave 5.7 En tappemaskin for kartonger med 1 liter melk er innstilt på 1,004 liter. Det er kjent at forventningen for tappevolumet da er lik 1,004 liter med et standardavvik lik 0,005 liter. Man går ut fra at tappevolumet er normalfordelt. Vis at sannsynligheten for at en melkekartong inneholder mindre enn 1 liter melk er 0,2119.

Løsningsforslag:

X = melkevolum i tilfeldig melkekartong, antar $X \sim N(\mu, \sigma)$ med $\mu = 1{,}004$ l, $\sigma = {,}005$ l. $P(X < 1) = P(Z < \frac{1-1{,}004}{{,}005}) = \Phi(-{,}8) = \underline{{,}2119}$

Oppgave 5.8 Anta at X er $N(5,10)$. La $k > 0$. Bestem k slik at $P(5 - k < X < 5 + k) = 0{,}90$.

Løsningsforslag:

$$P(5-k < X < 5+k) = P(5-k-5 < X-5 < 5+k-5) = P(-k < X-5 < k) =$$
$$P(-\tfrac{k}{10} < \tfrac{X-5}{10} < \tfrac{k}{10}) = P(-\tfrac{k}{10} < Z < \tfrac{k}{10}) = 2 \cdot \Phi(\tfrac{k}{10}) - 1 = {,}90 \Rightarrow$$
$$\Phi(\tfrac{k}{10}) = 1{,}90/2 = {,}95 \Rightarrow \tfrac{k}{10} = 1{,}645 \Rightarrow k = \underline{16{,}45}$$

Oppgave 5.9 En bestemt sort batterier har en levetid som er tilnærmet normalfordelt med forventning 1200 dager og standardavvik 100 dager. Hvor lang tid bør garantitiden være, hvis produsenten tar sikte på at 10 prosent av batteriene skal gi grunnlag for reklamasjon fordi levetiden er for kort?

Løsningsforslag:

X = levetid til tilfeldig batteri. $X \sim N(\mu, \sigma)$ med $\mu = 1200$ dager og $\sigma = 100$ dager. La k være garantitiden. Vi får da:

$$P(X < k) = 0{,}1 \Leftrightarrow P(Z < \tfrac{k-1200}{100}) = 0{,}1$$

Fra t-tabell med uendelig antall frihetsgrader finner vi at $P(Z < -1{,}282) = 0{,}1$. Følgelig: $\frac{k-1200}{100} = -1{,}282 \Rightarrow k = 1200 - 1{,}282 \cdot 100 = 1070$, dvs. $\underline{1070 \text{ timer}}$

Oppgave 5.10 Målefeilen til et instrument som måler blodsukkernivået er normalfordelt med forventning 0,05 og standardavvik 1,5. Dvs. ved gjentatte forsøk vil fordelingen til differansen (målt nivå minus sant nivå) være $N(0{,}05,\ 1{,}5)$.

a) Hvor stor prosentandel av målingene overestimerer det sanne nivået?

b) Anta at en målefeil anses som alvorlig når den målte verdien avviker fra den sanne verdien med mer enn 2,8. Hvor stor prosentandel av målingene vil være alvorlig feil?

Løsningsforslag: X = tilfeldig målefeil av blodsukkernivå, $X \sim N(,05,\ 1,5)$

a) P(overskride sant nivå) =
$P(X > 0) = 1 - P(X < 0) = 1 - P(Z = \frac{x-,05}{1,5} < \frac{0-,05}{1,5}) = 1 - \Phi(-,03) = 1 - ,488 = \underline{51,2\ \%}$

b) $P(|X| > 2,8) = 1 - P(|X| < 2,8) = 1 - P(-2,8 < X < 2,8) =$
$1 - (\Phi(\frac{2,8-,05}{1,5}) - \Phi(\frac{-2,8-,05}{1,5})) = 1 - (\Phi(1,83) - \Phi(-1,9)) = 1 - ,9664 + ,0287 = \underline{6,23\ \%}$

Oppgave 5.11 En bedrift framstiller en bestemt type brød. Vekten i gram av tilfeldig valgte brød antas å være uavhengige og normalfordelt $N(\mu,\sigma)$. Forventningen μ vil anhenge av deigen, innstillingen av maskinen som porsjonerer ut deigen, samt steketiden. Ifølge forskriftene skal denne type brød veie minst 750 gram.

Anta nå at en bestemt produksjonsplan innebærer at $\mu = 760$ og $\sigma = 10$, hvilket vil si at et tilfeldig valgt brød er $N(760,10)$.

a) Hva er sannsynligheten for at et tilfeldig valgt brød er undervektig?

b) Hvis 3 brød kjøpes, hva er da sannsynligheten for at minst ett er undervektig?

c) Hvis 3 brød kjøpes, hva er sannsynligheten for at det tyngste brødet veier mer enn 780 gram?

d) Finn sannsynligheten for at den gjennomsnittlige vekten av fem tilfeldig valgte brød er mindre enn 750 gram.

Løsningsforslag: X_i = vekt av tilfeldig brød, $X_1, X_2, \ldots$ antas uavhengige.

a) $P(X_i < 750) = \Phi\left(\frac{750-760}{10}\right) = \Phi(-1) = \underline{,159}$

b) P(minst ett av 3 brød undervektige) = 1 – P(ingen undervektige) =
$1 - P(X_1 > 750, X_2 > 750, X_3 > 750) =$
$1 - P(X_1 > 750) \cdot P(X_2 > 750) \cdot P(X_3 > 750) = 1 - (1 - ,159)^3 = \underline{,405}$

c) $P(\max(X_1, X_2, X_3) > 780) = 1 - P(\max(X_1, X_2, X_3) < 780) =$
$1 - P(X_1 < 780) \cdot P(X_2 < 780) \cdot P(X_3 < 780) = 1 - \Phi^3\left(\frac{780-760}{10}\right) = 1 - ,9772^3 = \underline{,067}$

d) $P(\overline{X} < 750) = P(\frac{\overline{X}-760}{10/\sqrt{5}} < \frac{750-760}{10/\sqrt{5}}) = P(Z < -\sqrt{5}) = \Phi(-2,24) = \underline{,013}$

Oppgave 5.12 X er binomisk fordelt med $n = 100$ og $p = 0,1$.

a) Sett opp uttrykket for beregning av $P(X \geq 8)$ (du skal ikke regne det ut).

Finn sannsynligheten under a) tilnærmet ved:

b) Poissonfordelingen.
c) Normalfordelingen.

Løsningsforslag:

a) $P(X \geq 8) = \sum_{x=8}^{100} \binom{100}{x} \cdot ,1^x \cdot ,9^{100-x}$

b) X tn. Po($\lambda = np = 10$). Vi benytter Poisson-tabell og finner:

$P(X \geq 8) = 1 - P(X \leq 7) = 1 - ,220 = \underline{,780}$

c) X tn. $N\left(np, \sqrt{np(1-p)}\right)$. Vi benytter halvkorreksjon (h.k.) og finner:

$1 - P(X \leq 7) \approx 1 - \Phi\left(\frac{7-np+,5}{\sqrt{np(1-p)}}\right) = 1 - \Phi\left(\frac{7-10+,5}{\sqrt{10 \cdot ,9}}\right) = 1 - \Phi(-,83) = 1 - ,203 = \underline{,797}$

Oppgave 5.13 La X = årsinntekten til en nyutdannet diplomøkonom i 1983. Anta at X er uavhengig av kjønn og tilnærmet normalfordelt med μ = 120000 kroner og σ = 20000 kroner. Finn

a) $P(X < 110000)$
b) $P(90000 < X < 150000)$

Bitten og Birger er to av de nyutdannete diplomøkonomer 1983, som nevnt ovenfor.

c) Hva er sannsynligheten for at minst en av dem tjener mer enn 120000?

d) Hva er sannsynligheten for at de til sammen tjener minst 220000?

Løsningsforslag: Lar alle beløp være i kr 1000.

a) $P(X < 110) = \Phi\left(\frac{110-120}{20}\right) = \Phi(-,5) = \underline{,309}$

b) $P(90 < X < 150) = \Phi\left(\frac{150-120}{20}\right) - \Phi\left(\frac{90-120}{20}\right) = \Phi(1,5) - \Phi(-1,5) = 2 \cdot \Phi(1,5) - 1 =$
$2 \cdot ,9332 - 1 = \underline{,866}$

c) P(minst én tjener mer enn 120) = 1 – P(ingen tjener mer enn 120) =
$1 - P(X_1 < 120) \cdot P(X_2 < 120) = 1 - \Phi^2\left(\frac{120-120}{20}\right) = 1 - \Phi^2(0) = 1 - ,5^2 = \underline{3/4}$

d) $P(X_1 + X_2 > 220) = P\left(\frac{(X_1+X_2)-2\mu}{\sqrt{2}\sigma} > \frac{220-2\mu}{\sqrt{2}\sigma}\right) = \Phi\left(-\frac{220-240}{\sqrt{2} \cdot 20}\right) = \Phi(,707) = \underline{,761}$

Oppgave 5.14 Et tilfeldig utvalg på n = 100 blir tatt fra en populasjon som har en forventning på 20 og et standardavvik på 5. Formen på populasjonfordelingen er ukjent.

a) Hva kan du si om fordelingen til utvalgsmiddelverdien, $\overline{X}$?

b) Finn tilnærmet sannsynlighet for at $\overline{X}$ skal overskride 20,75.

Løsningsforslag:

a) $\overline{X}$ tn. $N(20, \frac{5}{\sqrt{100}}) = \underline{N(20, ,5)}$

b) $P(\overline{X} > 20{,}75) = 1 - P(\overline{X} < 20{,}75) = 1 - \Phi\left(\frac{20{,}75-20}{,5}\right) = 1 - \Phi(1{,}5) = 1 - {,}9332 = \underline{{,}067}$

Oppgave 5.15 Fyll ut med det som mangler nedenfor:

a) $P(Z > \ ) = 0{,}975$, når Z er $N(0,1)$
b) $P(Z < \ ) = 0{,}050$, når Z er $N(0,1)$
c) $P(Z = 0) = ?$ når Z er $N(0,1)$
d) $P(-3{,}8 < X < 3{,}8) = ?$, X er $N(-3{,}1,\ 0{,}7)$

Løsningsforslag:

a) $P(Z > k) = {,}975 \Leftrightarrow 1 - P(Z < k) = {,}975 \Leftrightarrow P(Z < k) = {,}025$
Fra t-tabell med uendelig antall frihetsgrader: $P(Z < -1{,}96) = {,}025 \Rightarrow \underline{k = -1{,}96}$

b) $P(Z < k) = {,}050$. Utfra tabell som i a): $\underline{k = -1{,}645}$

c) $P(Z = 0) = 0$ fordi Z er en kontinuerlig variabel.

d) $P(-3{,}8 < X < 3{,}8) = \Phi\left(\frac{3{,}8-(-3{,}1)}{,7}\right) - \Phi\left(\frac{-3{,}8-(-3{,}1)}{,7}\right) = \Phi\left(\frac{6{,}9}{,7}\right) - \Phi(-1) \approx 1 - \Phi(-1) = \underline{{,}841}$

Oppgave 5.16 Anta at «persene» på 60-meteren til studentene i Norge er normalfordelt med forventning $\mu = 9{,}2$ sek og standardavvik $\sigma = 0{,}7$ sek.

a) Finn sannsynligheten for at en tilfeldig valgt student har «pers» på under 8,0 sek.

b) Hva er sannsynligheten for at det blant 5 tilfeldig valgte studenter er akkurat 3 som har «pers» over 8,0 sek?

Løsningsforslag: X_i: pers til tilfeldig student, $X_i \sim N(9{,}2,\ {,}7)$, $X_1, X_2, \ldots$ uavhengige.

a) $P(X_i < 8{,}0) = \Phi\left(\frac{8-9{,}2}{,7}\right) = \Phi\left(-\frac{1{,}2}{,7}\right) = \Phi(-1{,}71) = \underline{{,}044}$

b) Y = antall som har pers over 8,0 sek blant $n = 5$ tilfeldige studenter.
Y er da tilnærmet binomisk fordelt Bino(5, 1–,044) = Bino(5, ,956)
$P(Y = 3) = \binom{5}{3} \cdot {,}956^3 \cdot {,}044^2 = \underline{{,}017}$

Oppgave 5.17 Anta at karakterene i begynnerkurs i statistikk ved universitet og høyskoler er tilnærmet normalfordelt med forventning $\mu = 3{,}2$ og standardavvik $\sigma = 0{,}8$, og at 4,1 gir stryk.

a) Hvor stor er strykprosenten?

b) Hva er tilnærmet fordeling til gjennomsnittskarakteren til et tilfeldig utvalg på 35 besvarelser? Er fordelingen avhengig av at populasjons-fordelingen er normalfordelt?

c) Nevn 2 (prinsipielle) grunner til at karakterer kun kan være *tilnærmet* normalfordelt.

Løsningsforslag:

a) $P(\text{stryk}) = \Phi\left(-\frac{4{,}1-3{,}2}{,8}\right) = \Phi\left(-\frac{,85}{,8}\right) = \Phi(-1{,}06) \approx {,}1446$, dvs. 14 % stryk

b) $\overline{X}$ tn. $N(\mu, \sigma / \sqrt{35}) \approx N(3{,}2,\ {,}135)$, uavhengig av populasjonsfordeling.

c) 1) Karakterer er diskrete og ikke kontinuerlige, 2) karakterer er kun definert på et endelig intervall (0 til 6).

Oppgave 5.18 Anta at sannsynligheten for at en tilfeldig valgt student drikker mer enn 5 dl melk pr. dag er 0,6. Gitt en gruppe på 70 studenter. Finn sannsynligheten (tilnærmet) for at flere enn 30, men samtidig færre enn 50 av disse drikker mer enn 5 dl melk pr. dag.

Løsningsforslag: X = antall studenter blant n = 70 tilfeldige som drikker mer enn 5 dl. melk pr. dag. $X \sim$ Bino(70, ,6) med EX = np = 42 og Var(X) = $np(1-p)$ = 16,8. Bruker normaltilnærmelse og halvkorreksjon:

$$P(30 < X < 50) = P(X \leq 49) - P(X \leq 30) \approx \Phi\left(\tfrac{49-42+{,}5}{\sqrt{16{,}8}}\right) - \Phi\left(\tfrac{30-42+{,}5}{\sqrt{16{,}8}}\right) =$$

$$\Phi(1{,}83) - \Phi(-1{,}83) = 2 \cdot \Phi(1{,}83) - 1 = \underline{{,}933}$$ (eksakt, uten normaltiln.: ,965)

Oppgave 5.19 La X angi høyden (i cm) til en tilfeldig valgt norsk soldat. Vi skal anta at X er normalfordelt $N(\mu,\sigma)$, med forventning μ = 179 og standardavvik σ = 4,5.

a) Hva er sannsynligheten for at en tilfeldig valgt soldat er minst 190 cm?

b) Av et tilfeldig utvalg på 5 soldater, hva er sannsynligheten for at den høyeste av de 5 er minst 190 cm?

c) Hva er sannsynligheten for at gjennomsnittshøyden av 100 tilfeldige soldater er mindre enn 190 cm?

Vi ser nå på sammenhengen mellom vekt (i kg) og høyde (i cm) for soldatene illustrert ved følgende simultanfordeling $f(x,y)$:

Høyde [cm] → Vekt [kg] ↓	≤ 179	> 179
< 70	,25	,05
≥ 70	,25	,45

d) Finn sannsynligheten for at en tilfeldig valgt soldat skal være lettere enn 70 kg, gitt at han er høyere enn 179 cm.

Løsningsforslag: X_i = vekt av tilfeldig soldat, $X_1, X_2\ldots$ uavhengige.

a) $P(X_i > 190) = \Phi\left(-\frac{190-179}{4,5}\right) = \Phi\left(-\frac{11}{4,5}\right) = \Phi(-2{,}44) = \underline{,0073}$

b) $P(\max(X_1,\ldots,X_5) > 190) = 1 - P(\max(X_1,\ldots,X_5) < 190) =$
$1 - P(X_1 < 190,\ldots,X_5 < 190) = 1-(1-P(X_1 > 190))\cdots(1-P(X_5 > 190)) =$
$1-(1-{,}0073)^5 = \underline{,036}$

c) $P(\overline{X} < 190) = P(\frac{\overline{X}-179}{4,5/\sqrt{100}} < \frac{190-179}{4,5/\sqrt{100}}) = P(Z < 24{,}2) \approx \underline{1}$

d) $P(X_i < 70 | L_i > 179) = P(X_i < 70 \cap L_i > 179) / P(L_i > 179) = ,05/,50 = \underline{,1}$

Oppgave 5.20 Du har tilgjengelig slumptall mellom 0 og 1. Hvordan vil du gå fram for å benytte dette til å simulere et terningkast med rettferdig terning?

Løsningsforslag: La R betegne et tilfeldig slumptall mellom 0 og 1, og la X betegne antall øyne i simulert terningkast. Vi kan da simulere et terningkast ved å sette

$X = \text{int}(R\cdot 6) + 1$ der int betyr heltallsverdi.

Eks: $R = ,81 \Rightarrow x = \text{Int}(,81\cdot 6) + 1 = \text{int}(4{,}86) + 1 = 4 + 1 = 5$

Oppgave 5.21 Vis at $F(x) = P(X \le x) = x$ når X er $U(0,1)$-fordelt.

Løsningsforslag: $X \sim U(0,1) \Leftrightarrow f(x) = 1,\ 0 \le x \le 1$

$$F(x) = P(X \le x) = \int_0^x f(u)\,du = \int_0^x 1\,du = [x]_0^x = x, \ \underline{\text{q.e.d.}}$$

Oppgave 5.22 Vis at $E(X) = 1/2$ og at $\text{std}(X) = 1/\sqrt{12}$ når X er $U[0,1]$-fordelt.

Løsningsforslag:

$$EX = \int_0^1 x\cdot f(x)\,dx = \int_0^1 x\cdot 1\cdot dx = \left[\tfrac{1}{2}x^2\right]_0^1 = \tfrac{1}{2}\cdot 1^2 = 1/2, \ \underline{\text{q.e.d.}}$$

$$EX^2 = \int_0^1 x^2\cdot f(x)\,dx = \int_0^1 x^2\cdot 1\cdot dx = \left[\tfrac{1}{3}x^3\right]_0^1 = 1/3$$

$$\text{var}(X) = EX^2 - E^2X = 1/3 - (1/2)^2 = 1/3 - 1/4 = 1/12$$

$$\Rightarrow \text{std}(X) = \sqrt{\text{var}(X)} = 1/\sqrt{12}, \ \underline{\text{q.e.d.}}$$

Oppgave 5.23 Levetiden til et radioaktivt atom er eksponensialfordelt med parameter $b = 1{,}000$ min. Bestem halveringstida til det radioaktive atomet, dvs. den tida det tar før det er 50% sikkert at atomet vil disintegrere.

Løsningsforslag: Bruker sek som benevning, $T \sim \text{Expo}(60 \text{ sek})$.

$$P(T > t_{1/2}) = ,5 \Leftrightarrow 1 - P(T < t_{1/2}) = ,5 \Leftrightarrow P(T < t_{1/2}) = ,5 \Leftrightarrow 1 - e^{t_{1/2}/60} = ,5$$
$$\Rightarrow e^{-t_{1/2}/60} = ,5 \Rightarrow -t_{1/2}/60 = \ln(,5) = \ln(1/2) \Rightarrow t_{1/2} = 60 \cdot \ln(2) = 41{,}6, \text{ dvs. } \underline{41{,}6 \text{ sek}}$$

Oppgave 5.24 Et støvkorn inneholder 10^{11} atomer av et radioaktivt stoff med halveringstid $t = t_{1/2} = 1$ minutt.

a) Bestem parameteren b i eksponensialfordelingen til levetida til et vilkårlig av de radioaktive atomene.

b) Bestem variasjonskoeffisienten til antall radioaktive atomer som er igjen ved tidspunktet $t = t_{1/2}$.

Løsningsforslag: Bruker sek som benevning, $t_{1/2} = 60$ sek

a) $P(T < t_{1/2}) = ,5 \Leftrightarrow 1 - e^{-t_{1/2}/b} = ,5 \Leftrightarrow e^{-60/b} = ,5 \Leftrightarrow -60/b = \ln(\tfrac{1}{2}) = -\ln(2)$
$\Rightarrow b = 60/\ln(2) = 86{,}6$, dvs. $\underline{86{,}6 \text{ sek}}$

b) La X betegne antall atomer som er igjen ved tidspunktet $t = t_{1/2}$. X er da binomisk fordelt Bino(n,p) med $n = 10^{11}$ og $p = ,5$, dvs. X er tn. $N(\tfrac{1}{2}n, \tfrac{1}{2}\sqrt{n})$ som gir variasjonskoeffisient

$$\text{CV} = \frac{\text{std}(X)}{\text{E}(X)} \approx \frac{\sqrt{n}/2}{n/2} = 1/\sqrt{n} = 1/\sqrt{10^{11}} = \underline{3{,}2 \cdot 10^{-6}}$$

Oppgave 5.25 (E) En tappemaskin for kartonger med 1 liter melk er innstilt på 1,004 liter.Vi går ut fra at tappevolumet er normalfordelt med forventning lik 1,004 liter og standardavvik lik 0,005 liter.

a) Hva er sannsynligheten for at en tilfeldig melkekartong inneholder mindre enn 1,000 liter?

b) Vi kjøper 5 kartonger. Hva er sannsynligheten for at disse inneholder mindre enn 5,000 liter til sammen?

Løsningsforslag: X_i = volum i tilfeldig literpakke med melk, $X_1, X_2, \ldots$ uavhengige. $X_i \sim N(\mu, \sigma)$ med $\mu = 1{,}004$ og $\sigma = ,005$.

a) $P(X_i < 1{,}000) = \Phi\left(\frac{1{,}000 - 1{,}004}{,005}\right) = \Phi(-,8) = \underline{,212}$

b) $P(\Sigma_{i=1}^{5} X_i < 5) = P\left(\frac{\Sigma X_i - 5\mu}{\sqrt{5}\sigma} < \frac{5 - 5 \cdot 1{,}004}{\sqrt{5} \cdot ,005}\right) = P(Z < -1{,}79) = \underline{,037}$

Oppgave 5.26 (E) Sannsynligheten for å få bivirkninger ved bruk av en bestemt medisin er lik 0,20. Denne medisinen blir gitt til 30 pasienter, hvor forekomstene av bivirkningene er uavhengige.

a) Bestem forventningsverdien for antall pasienter som får bivirkninger.

b) Bestem variansen og standardavviket for antall pasienter som får bivirkninger.

c) Hva er sannsynligheten for at høyst 8 pasienter skal få bivirkninger? Bruk normaltilnærmelse.

Løsningsforslag: B = bivirkning, $p = P(B) = ,2$. X = antall av $n = 30$ som får bivirkning, $X \sim \text{Bino}(n,p) = \text{Bino}(30, ,2)$.

a) $\text{E}X = np = 30 \cdot ,2 = \underline{6}$

b) $\text{Var}(X) = np \cdot (1-p) = 6 \cdot ,8 = \underline{4,8}$, $\quad \text{Std(X)} = \sqrt{\text{Var}(X)} = \sqrt{4,8} = \underline{2,19}$

c) $P(X \le 8) = \Phi\left(\frac{8-6+,5}{2,19}\right) = \Phi(1,14) = \underline{,873}$ (,819 uten h.k.)

Oppgave 5.27 (E) Z er en tilfeldig kontinuerlig variabel som er normalfordelt med forventning $\mu = 0$ og varians $\sigma^2 = 1$.

a) Finn $P(Z \le 0,75)$, $P(Z > -0,75)$ og $P(-0,75 < Z \le 0,75)$.
b) Bestem k slik at $P(-k < Z \le k) = 0,9$.

Resistansen for en bestemt type elektriske motstander antas å være normalfordelt med forventning $\mu = 100\ \Omega$ og varians $\sigma^2 = (2\ \Omega)^2$. Motstandene er ubrukbare dersom resistansen er mindre enn 98 Ω eller større enn 103 Ω.

c) Hva er sannsynligheten for at en fritt valgt motstand er ubrukbar?

Motstandene pakkes i esker med 10 stykker i hver eske.

d) Hva er sannsynligheten for at en eske inneholder høyst en ubrukbar motstand?

Løsningsforslag:

a) $P(Z \le ,75) = \Phi(,75) = \underline{,773}$
$P(Z > -,75) = \Phi(-(-,75)) = \Phi(,75) = \underline{,773}$
$P(-,75 < Z < ,75) = 2 \cdot \Phi(,75) - 1 = 2 \cdot ,773 - 1 = \underline{,546}$

b) $P(-k < Z \le k) = ,9 \Leftrightarrow 2\Phi(k) - 1 = ,9 \Leftrightarrow \Phi(k) = 1,9/2 = ,95$
Fra t-tabell med d.f. $= \infty$: $\Phi(1,645) = ,95$, dvs. $\underline{k = 1,645}$

c) X = resistans i tilfeldig motstand, $X \sim N(\mu, \sigma)$.
$p = P(\text{ubrukbar motstand}) = P(X < 98) + P(X > 103) =$
$\Phi\left(\frac{98-100}{2}\right) + 1 - \Phi\left(\frac{103-100}{2}\right) = \Phi(-1) + 1 - \Phi(1,5) = ,1587 + 1 - ,9332 = \underline{,226}$

d) Y = antall ubrukbare motstander i tilfeldig eske med $n = 10$ motstander. $Y \sim \text{Bino}(n,p) = \text{Bino}(10, ,226)$.

$$P(Y \le 1) = P(X = 0) + P(X = 1) = \binom{10}{0} \cdot ,226^0 \cdot ,774^{10} + \binom{10}{1} \cdot ,226^1 \cdot ,774^9 = \underline{,302}$$

Oppgave 5.28 (E) Ved måling av alkoholinnholdet i blod kan måleresultatet X antas å være normalfordelt $N(\mu,\sigma)$. μ [promille] er den sanne verdien av alkoholinnholdet og σ [promille] er et mål for analysemetodens nøyaktighet. Anta at $\sigma = 0{,}04$ promille er kjent.

a) Hva er sannsynligheten for at et tilfeldig måleresultat overstiger 0,50 promille dersom den sanne verdien for alkoholinnholdet er 0,48 promille?

Ved en promilletest tas to uavhengige blodprøver av en person som er mistenkt for promillekjøring. Domstolen foreslår følgende kriterium: Personen dømmes for promillekjøring dersom det gjennomsnittlige alkoholinnholdet $\overline{X}$ i de to prøvene overstiger 0,53 promille.

b) Hva er sannsynligheten for at en person som har et alkoholinnhold på 0,48 promille dømmes for promillekjøring dersom dette kriteriet brukes?

Advokaten hevder at rettens fremgangsmåte gir altfor stor sannsynlighet for at en uskyldig blir dømt, og man blir derfor enige om å finne et nytt kriterium:

Personen skal dømmes hvis $\overline{X} > k$, der k skal bestemmes slik at sannsynligheten for at en person med alkoholinnhold på 0,49 promille blir dømt, er lik 0,01.

c) Bestem k.

<u>Løsningsforslag</u>:

a) $P(X > ,50) = \Phi\left(-\frac{,50-,48}{,04}\right) = \Phi(-,5) = \underline{,309}$

b) $\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{2}})$, $P(\overline{X} > ,53) = \Phi\left(-\frac{,53-,48}{,04/\sqrt{2}}\right) = \Phi(-1{,}77) = \underline{,038}$

c) $P(\overline{X} > k \,|\, \mu = ,49) = ,01 \Leftrightarrow P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{2}} > \frac{k-\mu}{\sigma/\sqrt{2}}\right) = ,01 \Leftrightarrow P\left(Z > \frac{k-\mu}{\sigma/\sqrt{2}}\right) = \Phi\left(\frac{k-\mu}{\sigma/\sqrt{2}}\right) = ,01$

Fra t-tabell: $P(Z > 2{,}326) = ,01$
$\Rightarrow \frac{k-\mu}{\sigma/\sqrt{2}} = 2{,}326 \Rightarrow k = \mu + \frac{\sigma}{\sqrt{2}} \cdot 2{,}326 = ,49 + \frac{,04}{\sqrt{2}} \cdot 2{,}326 = \underline{,556}$

Oppgave 5.29 (E) En bedrift produserer syltetøy som leveres på glass. Vekten X av innholdet (nettovekten) antas å være normalfordelt $N(\mu,\sigma)$, der μ avhenger av innstillingen på tappemaskinen.

Vekten Y av emballasjen (glass med lokk) antas å være N(0,20 kg, 0,01 kg). Et syltetøyglass med nettovekt (X) under 0,80 kg anses for å være undervektig. Anta at $\sigma = 0{,}02$ kg er kjent, og at maskinen er innstilt slik at $\mu = 0{,}82$ kg.

a) Hva er sannsynligheten for at et tilfeldig syltetøyglass er undervektig?

b) Hva er sannsynligheten for at nøyaktig 2 av 10 glass er undervektige?

c) La Z være bruttovekten (syltetøy + emballasje) av et tilfeldig glass. Hvilken sannsynlighetsfordeling får Z? Finn sannsynligheten for at bruttovekten er større enn 1,0 kg.

d) Man ønsker nå å justere tappemaskinen slik at 90 % av bruttovektene overstiger 1,00 kg. Hvilken verdi for μ må vi da innstille tappemaskinen på?

Løsningsforslag:

a) $P(X <,8) = \Phi\left(\frac{,80-,82}{,02}\right) = \Phi(-1) = \underline{,159}$

b) Y = antall undervektige glass, $Y \sim$ Bino(10, ,159).

$P(Y=2) = \binom{10}{2} \cdot ,159^2 \cdot (1-,159)^2 = \underline{,285}$

c) B = bruttovekt av tilfeldig syltetøyglass (i kg), $B \sim \underline{N}(\mu_B, \sigma_B)$ med $\mu_B = \mu_X + \mu =$ $,82 + ,20 = \underline{1,02}$ og $\sigma_B^2 = \sigma_X^2 + \sigma^2 = ,02^2 + ,01^2 = ,0005$, $\sigma_B = \sqrt{,0005} = \underline{,0224}$

$$P(Z > 1,0) = \Phi\left(-\frac{1-1,02}{,0224}\right) = \Phi(,89) = \underline{,813}$$

d) $\mu_B = \mu + ,20$, $\sigma_B = ,0224$

$P(B > 1) = 1 - P(B < 1) = 1 - \Phi\left(\frac{1-\mu-,2}{,0224}\right) = ,9 \Rightarrow \Phi\left(\frac{,8-\mu}{,0224}\right) = ,1$

Fra t-tabell: $P(Z < 1,282) = ,1 \Leftrightarrow P(Z < -1,282) = \Phi(-1,282) = ,1$

Vi får derfor: $\mu = ,8 + 1,282 \cdot ,0224 = ,829 \approx \underline{,83}$

Oppgave 5.30 (E) En mann trenger 165 kg kunstgjødsel. Kunstgjødselen leveres i sekker á 50 kg og 10 kg. Vekten av innholdet i en 50 kg-sekk antas å være normalfordelt $N(50, 2,0)$. Vekten av innholdet i en 10 kg-sekk antas å være normalfordelt $N(10, 0,4)$. Vektene av forskjellige sekker antas å være uavhengige. Mannen bestemmer seg for å kjøpe 3 stk. 50 kg-sekker og 2 stk. 10 kg-sekker.

Vi lar Z betegne totalvekten av de 5 sekkene.

a) Hvilken sannsynlighetsfordeling har Z?

b) Beregn sannsynligheten for at mannen får for lite kunstgjødsel (mindre enn 165 kg).

c) Beregn sannsynligheten for at den tyngste av de tre 50 kg-sekkene veier mer enn 53 kg.

d) Beregn sannsynligheten for at de tre 50 kg-sekkene i gjennomsnitt veier mer enn 51 kg.

e) Hvor mange 50 kg-sekker må kjøpes for at sannsynligheten for at minst en skal veie over 53 kg er minst 0,5?

Løsningsforslag: X_1, X_2 og X_3 er u.i.f. $N(50, 2,0)$, Y_1 og Y_2 er u.i.f. $N(10, ,4)$.

a) $Z = X_1+X_2+X_3+Y_1+Y_2 \sim \underline{N}(\mu_Z, \sigma_Z)$ der $\mu_Z = 3\cdot\mu_X + 2\cdot\mu_Y = 3\cdot 50 + 2\cdot 10 = \underline{170}$, og
$\sigma_Z^2 = 3\cdot\sigma_X^2 + 2\cdot\sigma_Y^2 = 3\cdot 2{,}0^2 + 2\cdot{,}4^2 = 12{,}32,\ \sigma_Z = \sqrt{\sigma_Z^2} = \sqrt{12{,}32} = \underline{3{,}51}$

b) $P(Z < 165) = \Phi\left(\frac{165-170}{3{,}51}\right) = \Phi(-1{,}42) = \underline{{,}078}$

c) $P(\max(X_1, X_2, X_3) > 53) = 1 - P(\max(X_1, X_2, X_3) < 53) =$
$1 - P(X_1 < 53 \cap X_2 < 53 \cap X_3 < 53) = 1 - \Phi^3\left(\frac{53-50}{2{,}0}\right) = 1 - {,}9332^3 = \underline{{,}187}$

d) $P(\overline{X} > 51) = P(Z = \frac{\overline{X}-50}{2/\sqrt{3}} > \frac{51-50}{2/\sqrt{3}}) = \Phi(-{,}87) = \underline{{,}192}$

e) $p = P(X_i > 53) = \Phi\left(-\frac{53-50}{2}\right) = \Phi(-1{,}5) = \underline{{,}067}$
Y = antall av n 50 kg-sekker som veier mer enn 53 kg. $Y \sim \text{Bino}(n,p)$.
$P(Y \geq 1) = 1 - P(Y = 0) > {,}5 \Leftrightarrow P(Y = 0) < {,}5 \Leftrightarrow \binom{n}{0}\cdot p^0 \cdot (1-p)^n < {,}5$
$\Rightarrow n\cdot\ln(1-p) < \ln({,}5) = -\ln(2) \Rightarrow n > -\ln(2)/\ln(1-p) = -\frac{\ln(2)}{\ln({,}923)} = 9{,}99 \Rightarrow n = \underline{10}$

Oppgave 5.31 (E) En fabrikk produserer vareenheter med vekt Y [gram] som antas å være normalfordelt $N(\mu,\sigma)$. En kasseringsautomat kasserer vareenheter som veier mindre enn 1168 g, og vareenheter som veier mer enn 1518 g. En har funnet ut at i det lange løp blir 12,3 % av enhetene kassert fordi de er for lette, mens 27,7 % av artiklene blir kassert fordi de er for tunge.

a) Vis at $\sigma = 200$ og $\mu = 1400$.

b) Hva er sannsynligheten for at en vareenhet har en vekt som er større enn 1755 g?

Vi tar ut et vareparti på 5 enheter.

c) Hva er sannsynligheten for at nøyaktig x av de 5 enhetene hver har en vekt som er større enn 1755 g?

d) Hva er sannsynligheten for at minst en av de fem vareenhetene har en vekt som er større enn 1755 g?

e) Hva er sannsynligheten for at de fem vareenhetene totalt veier mer enn 7,5 kg?

f) Bestem tilnærmet sannsynlighet for at det i et stort vareparti på 1000 enheter, høyst vil være 45 som veier over 1775 g.

<u>Løsningsforslag</u>:

a) $P(Y < 1168) = \Phi\left(\frac{1168-1400}{200}\right) = \Phi(-1{,}16) = {,}123 = 12{,}3\,\%$, <u>q.e.d.</u>
$P(Y > 1518) = \Phi\left(-\frac{1518-1400}{200}\right) = \Phi(-{,}59) = {,}278 = 27{,}8\,\%$, <u>q.e.d.</u>

b) $p = P(X > 1755) = \Phi\left(-\frac{1755-1400}{200}\right) = \Phi(-1{,}775) = \underline{{,}038}$

c) $P(X = x) = \binom{5}{x}\cdot p^x \cdot (1-p)^{5-x}$, $x = 0,1,\ldots,5$ med $p = {,}038$ fra b).

d) $P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{5}{0}\cdot p^0 \cdot (1-p)^5 = 1 - {,}962^5 = \underline{{,}176}$

e) $P(W = Y_1 + \cdots + Y_5 > 7{,}5) = P(Z = \frac{W-5\mu}{\sqrt{5}\sigma} > \frac{7{,}5-5\mu}{\sqrt{5}\sigma}) = \Phi\left(-\frac{7{,}5-5\cdot 1{,}4}{\sqrt{5}\cdot{,}2}\right) = \Phi(-1{,}12) = \underline{{,}131}$

f) $n = 1000$, X = antall som veier mer enn 1755 g. $X \sim$ Bino(n,p) $\approx N\left(np, \sqrt{np(1-p)}\right)$

$= N(38,\ 6{,}05)$.

$P(X \le 45) = \Phi\left(\frac{45-38+{,}5}{6{,}05}\right) = \Phi(1{,}24) = \underline{{,}89}$

Oppgave 5.32 (E) Anta at X har en standard normalfordeling og bestem:

a) $P(X \le 0{,}42)$ og $P(X \le -1{,}69)$
b) k slik at $P(-0{,}3 < X \le k) = 0{,}2$

Med et doseringsapparat fylles det automatisk fargepulver i poser på et transportbånd. Apparatet kan innstilles slik at vekten X av fargepulver i hver pose blir mellom 10 og 30 gram. Med apparatet innstilt på en bestemt vekt, kan X betraktes som en normalfordelt stokastisk (tilfeldig) variabel med forventning μ = innstilt vekt og standardavvik $\sigma = 0{,}5$ gram.

c) Anta at innstilt vekt er $\mu = 20{,}4$ gram, og at det for hver pose fylles to ganger etter hverandre for å gi samlet vekt på 40,8 gram. Hvor stor andel av posene vil en forvente har en samlet vekt mindre eller lik 40 gram?

d) Anta at innstilt vekt er $\mu = 15$ gram, og at det for hver pose bare fylles én gang. Hvor stor er da sannsynligheten for at vektforskjellen mellom to vilkårlig valgte poser er større enn 0,7 gram?

Løsningsforslag:

a) $P(X \le {,}42) = \Phi({,}42) = \underline{{,}663}$, $\qquad P(X \le -1{,}69) = \Phi(-1{,}69) = \underline{{,}046}$

b) $P(-{,}3 < X \le k) = {,}2 \Leftrightarrow P(X \le k) - P(X < -{,}3) = {,}2 \Leftrightarrow$
$P(X \le k) = P(X < -{,}3) + {,}2 = \Phi(-{,}3) + {,}2 = {,}382 + {,}2 = {,}582$
Fra $N(0,1)$-tabell: $\underline{k \approx {,}21}$

c) $Y = X_1 + X_2$, $Y \sim N(2\mu, \sqrt{2}\sigma) = N(40{,}8, \sqrt{2}\cdot{,}5)$
$P(Y < 40) = \Phi\left(\frac{40-40{,}8}{\sqrt{2}\cdot{,}5}\right) = \Phi(-1{,}13) = \underline{{,}129}$

d) $D = X_1 - X_2 \sim N(0, \sqrt{2}\cdot{,}5)$
$P(|D| > {,}7) = 1 - P(|D| < {,}7) = 1 - P(-{,}7 < D < {,}7) = 1 - P(D < {,}7) - P(D < -{,}7)$
$= 1 - \Phi\left(\frac{{,}7-0}{\sqrt{2}\cdot{,}5}\right) + \Phi\left(\frac{-{,}7-0}{\sqrt{2}\cdot{,}5}\right) = 1 - \Phi({,}990) + \Phi(-{,}990) = 2\cdot(1-\Phi({,}990)) =$
$= 2\cdot(1-{,}839) = 2\cdot{,}161 = \underline{{,}322}$

Oppgave 5.33 (E) Vi skal få anta at vekten X i kg for en voksen mann i en bestemt befolkningsgruppe er normalfordelt med forventning $\mu = 75$ kg og standardavvik $\sigma = 15$ kg.

a) Finn sannsynligheten for at en tilfeldig mann veier mer enn 90 kg. For en gruppe på 4 menn, finn sannsynligheten for at minst 1 veier mer enn 90 kg.

b) En varsellampe i en heis begynner å lyse når den samlede vekten av personene i heisen er mer enn 800 kg. Finn sannsynligheten for at lampa vil lyse når 10 tilfeldige menn har gått inn i heisen.

c) Hvis Y er vekten av en voksen kvinne i den samme befolkningsgruppen, har det vist seg at $Y = 0{,}8X' + 2$, der X' har samme fordeling som X. Finn E(Y), std(Y) og sannsynligheten for at en tilfeldig kvinne veier mindre enn 52 kg.

d) 8 kvinner og 4 menn går inn i den heisen vi omtalte ovenfor. Finn sannsynligheten for at varsellampa vil lyse denne gangen.

<u>Løsningsforslag</u>:

a) $P(X - < 90) = 1 - P(X < 90) = 1 - \Phi\left(\frac{90-75}{15}\right) = 1 - \Phi(1) = 1 - {,}841 = \underline{{,}159}$

Y = antall menn av $n = 4$ som er tyngre enn 90 kg. $Y \sim \text{Bino}(4, {,}159)$.

$P(Y \geq 1) = 1 - P(Y = 0) = 1 - (1 - {,}159)^4 = \underline{{,}450}$

b) $W = X_1 + \cdots + X_{10} \sim N(10\mu, \sqrt{10}\,\sigma)$

$P(W > 800) = \Phi\left(-\frac{800-10\cdot 75}{\sqrt{10}\cdot 15}\right) = \Phi(-1{,}05) = \underline{{,}147}$

c) $\text{E}Y = \text{E}({,}8\cdot X' + 2) = {,}8\cdot \text{E}X + {,}2 = {,}8\cdot\mu + 2 = {,}8\cdot 75 + 2 = \underline{62}$

$\text{Std}(Y) = {,}8\cdot\text{Std}(X) = {,}8\cdot 15 = \underline{12}$

$P(Y < 52) = \Phi\left(\frac{52-62}{12}\right) = \Phi(-{,}83) = \underline{{,}203}$

d) $\text{W} = X_1 + \cdots + X_4 + Y_1 + \cdots + Y_8,\ \text{W} \sim N(4\mu_X + 8\mu_Y, \sqrt{4\sigma_X^2 + 8\sigma_Y^2}) =$

$N(4\cdot 75 + 8\cdot 62, \sqrt{4\cdot 15^2 + 8\cdot 12^2} = N(45, 3).$

$P(W > 800) = \Phi\left(-\frac{800-796}{45}\right) = \Phi(-{,}09) \approx \underline{{,}464}$

Oppgave 5.34 (E) En maskin lager sylinderformede bolter. Det har vist seg at en bolts diameter D er normalfordelt med forventning $\mu = 2$ cm og standardavvik $\sigma = 0{,}1$ cm. En bolt anses å være ubrukbar for et bestemt formål dersom D avviker fra μ med mer enn 0,1 cm. Diameteren til en bolt antas å være uavhengig av diameteren til de andre boltene.

a) Hva er sannsynligheten for at en bolts diameter er større enn 2,1 cm?

b) Bestem sannsynligheten for at en tilfeldig bolt er brukbar.

Boltene pakkes i esker med 10 bolter i hver eske. Den gjennomsnittlige diameteren til disse 10 boltene er

$$\overline{D} = \tfrac{1}{10}\sum_{i=1}^{10} D_i$$

c) Hva er sannsynligheten for at den gjennomsnittlige diameteren er mindre enn 2,02 cm?

d) Bestem sannsynligheten for at minst 9 av boltene i en tilfeldig eske er brukbare.

Boltene skal ned i et hull hvis diameter Y er normalfordelt med $E(Y) = 2{,}2$ cm og $std(Y) = 0{,}2$ cm.

e) Hva er sannsynligheten for at en tilfeldig valgt bolt skal gå ned i et tilfeldig valgt hull?

Løsningsforslag:

a) $P(D > 2{,}1) = 1 - P(D < 2{,}1) = 1 - \Phi\left(\frac{2{,}1-2}{,1}\right) = 1 - \Phi(1) = 1 - {,}8413 = \underline{,159}$

b) $P(|D-\mu| > {,}1) = P(-{,}1 < D - \mu < {,}1) = P\left(-\frac{,1}{,1} < Z = \frac{D-\mu}{\sigma} < \frac{,1}{,1}\right) = 2 \cdot \Phi(1) - 1 = \underline{,683}$

c) $P(\overline{D} < 2{,}01) = P\left(Z = \frac{\overline{D}-\mu}{\sigma/\sqrt{10}} < \frac{2{,}02-2{,}00}{,1/\sqrt{10}}\right) = \Phi({,}632) = \underline{,736}$

d) X = antall brukbare bolter av $n = 10$ tilfeldige. $X \sim \text{Bino}(n, p) = \text{Bino}(10, {,}683)$.
$P(X \geq 9) = P(X = 9) + P(X = 10) = \binom{10}{9} \cdot {,}683^9 \cdot {,}317^1 + {,}683^{10} = {,}103 + {,}022 = \underline{,125}$

e) $W = Y - D$, $W \sim N(2{,}2 - 2, \sqrt{,1^2 + ,2^2}) = N({,}2, {,}224)$
$P(W > 0) = 1 - P(W < 0) = 1 - \Phi\left(\frac{-,2}{,224}\right) = \Phi({,}89) = \underline{,813}$

Oppgave 5.35 (E) Levetiden T (i timer) for en viss type lyspærer betraktes som en stokastisk variabel med sannsynlighetsfordeling gitt ved frekvensfunksjonen (tetthetsfunksjonen):

$$f(t) = \begin{cases} k \cdot e^{-0{,}001t}, & t > 0 \\ 0, & t \leq 0 \end{cases}$$

a) Bestem k (klarer du ikke det, bruk $k = 0{,}005$ i resten av oppgaven).

b) Beregn den forventede levetiden m for slike lyspærer.

c) Beregn sannsynligheten for at en lyspære skal ha kortere levetid enn m.

d) En lyspære har brent i 900 timer og er fremdeles i orden. Hvor stor er sannsynligheten for at den fortsetter å brenne i minst 200 timer til?

Løsningsforslag:

a) $\int_0^\infty f(t)\,dt = 1 \Leftrightarrow \int_0^\infty ke^{-,001t}\,dt = \left[-\frac{ke^{-,001t}}{,001}\right]_0^\infty = 1000k = 1 \Rightarrow k = \underline{,001}$

b) $m = \mathrm{E}X = \int_0^\infty t \cdot f(t)\,dt = k\int_0^\infty \underset{u}{t} \cdot \underbrace{e^{-kt}dt}_{dv} = k\left[\underset{u}{t} \cdot \underbrace{\left(-\frac{e^{-kt}}{k}\right)}_{v}\right]_0^\infty + k\int_0^\infty \underbrace{\tfrac{1}{k}e^{-kt}}_{v}\underbrace{dt}_{du} = \frac{1}{k} = \underline{1000}$

c) $P(T < m) = \int_0^m ke^{-kt}dt = 1 - e^{-mk} = 1 - e^{-1} = \underline{,632}$

d) $P(T > 1100 | T > 900) = P(T > 1100) / P(T > 900) = \frac{e^{-,001 \cdot 1100}}{e^{-,001 \cdot 900}} = e^{-,001 \cdot 200} = e^{-2} = \underline{,819}$

Oppgave 5.36 (E) La $X \sim N(0,1), Y = 2X+3$ og $Z = X^2$.

a) Bestem $P(X<2)$, $P(Y>2)$ og $P(Z<2)$.

b) Bestem $\mathrm{E}(Y)$, $\mathrm{Var}(Y)$ og $\mathrm{E}(Z)$. Finn også sannsynlighetstettheten til Z.

La $X_1, X_2, \ldots, X_{10}$ alle være uavhengige og $N(0,1)$-fordelte.

c) Beregn $P(X_1 + \cdots + X_{10} > 2)$ og $P(X_1 + \cdots + X_7 < 2X_8 + 2X_9 + 2X_{10})$

Løsningsforslag:

a) $P(X < 2) = \Phi(2) = \underline{,977}$
$P(Y > 2) = P(2X + 3 > 2) = P(2X > -1) = P(X > -,5) = \Phi(-(-,5)) = \Phi(,5) = \underline{,692}$
$P(Z < 2) = P(X^2 < 2) = P(-\sqrt{2} < X < \sqrt{2}) = 2\Phi(\sqrt{2}) - 1 = 2 \cdot ,921 - 1 = \underline{,842}$

b) $\mathrm{E}Y = \mathrm{E}(2X + 3) = 2\mathrm{E}X + 3 = \underline{3}$
$\mathrm{Var}Y = \mathrm{Var}(2X + 3) = 2^2 \cdot \mathrm{Var}X = 4 \cdot 1 = \underline{4}$, $\mathrm{Std}(Y) = \sqrt{\mathrm{Var}(Y)} = \sqrt{4} = 2$
$\mathrm{E}Z = \mathrm{E}X^2 = \mathrm{Var}X + \mathrm{E}^2 X = 1 + 0^2 = \underline{1}$

Finner først kumulativ fordeling, $F_Z(z)$ til Z:

$$F_Z(z) = P(Z < z) = P(X^2 < z) = P(|X| < \sqrt{z}) = P(-\sqrt{z} < X < \sqrt{z}) = 2\Phi(\sqrt{z}) - 1$$

Den søkte tetthetsfunksjon finnes så ved å derivere med hensyn på z:

$$f_Z(z) = \frac{dF_Z(z)}{dz} = 2\phi(\sqrt{z}) \cdot \frac{d\sqrt{z}}{dz} = \frac{\phi(\sqrt{z})}{\sqrt{z}} = \underline{\frac{1}{\sqrt{2\pi z}} e^{-z/2}, \ z \geq 0}$$

Fordelingen er identisk med Kji2(1).

c) $P(X_1 + \cdots + X_{10} > 2) = 1 - P(X_1 + \cdots + X_{10} < 2) = 1 - P(Z = \frac{X_1 + \cdots + X_{10} - 0}{\sqrt{10 \cdot 1}} < \frac{2-0}{\sqrt{10}}) =$
$1 - \Phi(,632) = 1 - ,736 = \underline{,264}$
$P(X_1 + \cdots + X_7 < 2X_8 + 2X_9 + 2X_{10}) = P(X_1 + \cdots + X_7 - 2X_8 - 2X_9 - 2X_{10} < 0)$
$W = X_1 + \cdots + X_7 - 2(X_8 + X_9 + X_{10}) \sim N(0, \sigma_W) \Rightarrow P(W < 0) = \underline{,5}$

Oppgave 5.37 (E) Anta at X er normalfordelt $N(0,1)$.

a) Beregn $P(X \leq 0,80)$, $P(X > -0,80)$ og $P(-1,35 < X \leq 0,95)$.

b) Bestem konstantene a, b, c slik at

$P(X \leq a) = 0{,}90,$
$P(X > b) = 0{,}025,$
$P(c < X \leq 0{,}35) = 0{,}50$

En bedrift produserer betongsylindere av en bestemt type. Sylindernes trykkfasthet [N/mm^2] er uavhengige normalfordelte variabler med forventning 29 og standardavvik 0,5 N/mm^2. Sylinderne er uegnet for sitt formål dersom de har trykkfasthet under 28,4 N/mm^2, og de blir i så fall kassert.

c) Hva er sannsynligheten for at en tilfeldig sylinder blir kassert?

Hva er sannsynligheten for at en tilfeldig sylinder har en trykkfasthet på minst 30,0 N/mm^2?

Bedriften har påtatt seg å levere 900 brukbare sylindere til en kunde. På grunn av faren for svinn produserer bedriften for sikkerhets skyld 1000 sylindere.

d) Hvor stor er (tilnærmet) sannsynligheten for at oppdraget kan utføres? Forklar hvordan du resonnerer for å løse denne oppgaven.

e) Hva er det minste antall sylindere bedriften må produsere hvis leveransen skal kunne gjennomføres med 99 % sannsynlighet?

Løsningsforslag:

a) $P(X \leq ,80) = \Phi(,80) = \underline{,788}, \qquad P(X > -,80) = P(X < ,80) = \underline{,788}$
$P(-1{,}35 < X \leq ,95) = \Phi(,95) - \Phi(-1{,}35) = ,829 - ,089 = \underline{,740}$

b) $P(X \leq a) = ,90 \Leftrightarrow P(X \geq a) = 1 - ,90 = ,10$
t-tabell, d.f. = ∞: $P(Z \geq 1{,}282) = ,10 \Rightarrow a = \underline{1{,}282}$
$P(X > b) = ,025$, t-tabell, d.f. = ∞: $P(Z \geq 1{,}96) = ,025 \Rightarrow b = \underline{1{,}96}$
$P(c < X \leq ,35) = ,50 \Rightarrow \Phi(,35) - \Phi(c) = ,50 \Rightarrow \Phi(c) = ,637 - ,50 = ,137$
N(0,1)-tabell: P(Z < -1,09) = ,1379 $\Rightarrow$ $c \approx \underline{-1{,}09}$

c) X = trykkfasthet til tilfeldig sylinder. $X \sim N(29, ,5)$.
$q = P(\text{tilfeldig sylinder kasseres}) = P(X < 28{,}4) = \Phi\left(\frac{28{,}4-29}{n,5}\right) = \Phi(-1{,}2) = \underline{,115}$

$P(X > 30{,}0) = 1 - P(X < 30{,}0) = 1 - \Phi\left(\frac{30{,}0-29}{,5}\right) = 1 - \Phi(2) = 1 - ,977 = \underline{,023}$

d) Y = antal brukbare sylindre av n = 1000, dvs. $Y \sim$ Bino(n,p) med n = 1000 og p = 1 − q = ,885 (fra a)). np(1−p) = 885·,115 >> 5, dvs. Y tn. $N(np, \sqrt{np(1-p)}) =$ N(885, 10,1).

$P(Y \geq 900) = 1 - P(Y \leq 899) \approx 1 - \Phi\left(\frac{899-885\cdot,115+,5}{10{,}1}\right) = 1 - \Phi(1{,}44) = 1 - ,925 = \underline{,075}$

e) $P(Y \geq 900) \geq ,99 \Rightarrow P(Y \leq 899) \leq ,01 \Rightarrow \Phi\left(\frac{899-np+,5}{\sqrt{np(1-p)}}\right) \leq ,01$

Fra t-tabell, d.f. = ∞: $P(Z \geq 2{,}326) = {,}01 \Rightarrow \dfrac{899{,}5 - np}{\sqrt{np(1-p)}} \leq -2{,}326$

Vi starter med å erstatte ulikhetstegnet i uttrykket til høyre med likhetstegn, og løse den ligningen som fremkommer med hensyn på n. Vi får da:

$$899{,}5 - np = -2{,}326\sqrt{np(1-p)} \Rightarrow np - 899{,}5 = 2{,}326\sqrt{np(1-p)} \Rightarrow$$
$$(np - 899{,}5)^2 = 2{,}326^2 \cdot np(1-p)$$

Uttrykket ovenfor gir en 2. grads ligning i n. Etter innsetting av p = ,115 og noe regning får vi:

$$n^2 - 2033{,}6 \cdot n + 1{,}033 \cdot 10^6 = 0$$

som har løsninger n_1 = 987 og n_2 = 1046. n_1 er mindre enn 1000 og er ingen aktuell løsning, siden vi fra før vet at 1000 var for lite antall. På grunn av avrundings-usikkerhet sjekker vi svaret n = 1046 og får:

$$\tfrac{899{,}5 - np}{\sqrt{np(1-p)}} = -2{,}540 < -2{,}326$$

så n = 1046 er tilstrekkelig. Vi sjekker også om vi kan nøye oss med en noe mindre n:

$$n = 1045: \quad \tfrac{899{,}5 - np}{\sqrt{np(1-p)}} = -2{,}456 < -2{,}326$$
$$n = 1044: \quad \tfrac{899{,}5 - np}{\sqrt{np(1-p)}} = -2{,}371 < -2{,}326$$
$$n = 1043: \quad \tfrac{899{,}5 - np}{\sqrt{np(1-p)}} = -2{,}286 > -2{,}326$$

Konklusjon: n = $\underline{1044}$

Oppgave 5.38 (E) Et supermarked selger melk i kartonger på ½, 1 og 1½ liter. Da tappingen ikke skjer med eksakt presisjon, vil vi betrakte melkevolumet i hver av kartongtypene som normalfordelte stokastiske variabler med følgende fordelinger: $N(0{,}5,\ 0{,}05)$, $N(1{,}0,\ 0{,}1)$ og $N(1{,}5,\ 0{,}20)$. Melkemengdene i de forskjellige kartongene antas å være uavhengige av hverandre.

a) Hvor stor er sannsynligheten for at en 1-liters kartong inneholder mer enn 1,03 l?

Kunden skal ha 2,5 liter melk.

b) Hva er sannsynligheten for at han får mindre enn 2,40 liter melk hvis han tar kartonger på 1 liter og 1½ liter?

c) Hadde det vært fornuftig av kunden å velge andre kombinasjoner? Svaret skal begrunnes ved å regne ut sannsynligheten for å få mindre enn 2,40 l melk ved andre måter å velge på.

d) Bestem a slik at sannsynligheten for at samlet melkevolum ligger innenfor $[6{,}0 - a,\ 6{,}0 + a]$ blir 0,95 hvis det kjøpes seks 1-liters melkekartonger.

Løsningsforslag: X, Y og Z er volum i tilfeldig pakning med henholdsvis ,5, 1 og 1,5 liter. $X \sim N(,5,\ ,05)$, $Y \sim N(1,0,\ ,1)$ og $Z \sim N(1,5,\ ,20)$.

a) $P(Y > 1{,}03) = 1 - P(Y < 1{,}03) = 1 - \Phi\left(\frac{1{,}03-1{,}0}{,1}\right) = 1 - \Phi(,3) = 1 - ,618 = \underline{,382}$

b) $Y + Z \sim N(2{,}5,\ \sqrt{,1^2 + ,2^2}\,) = N(2{,}5,\ \sqrt{,05}) = N(2{,}5,\ ,224)$.
$P(Y + Z < 2{,}40) = \Phi\left(\frac{2{,}40-2{,}5}{,224}\right) = \Phi(-,45) = \underline{,326}$

c) Alternativer:

i) $X_1 + \cdots + X_5 \sim N(2{,}5,\ \sqrt{5}\cdot,05) = N(2{,}5,\ \sqrt{,0125})$

ii) $X_1 + Y_1 + Y_2 \sim N(2{,}5,\ \sqrt{,05^2 + 2\cdot,1^2}\,) = N(2{,}5,\ \sqrt{,0225})$

iii) $X_1 + X_2 + X_3 + Y_1 \sim N(2{,}5,\ \sqrt{3\cdot,05^2 + 2\cdot,1^2}\,) = N(2{,}5,\ \sqrt{,0275})$

Siden alle alternativer har forventning 2,5 > 2,4, vil alternativet med mins varians være lurest, dvs. $\underline{5}$ halvlitere.

d) La Q betegne samlet melkevolum i seks tilfeldige liter-kartonger. Q er da normalfordelt $N(\mu, \sigma)$ med $\mu = 6{,}0$ og $\sigma = \sqrt{6\cdot,1^2} = ,245$. Vi får:

$P(6 - a < Q < 6 + a) = P(-a < Q - 6 < a) = P(-\frac{a}{\sigma} < \frac{Q-\mu}{\sigma} < \frac{a}{\sigma}) = ,95$

Siden $Z = (Q-\mu)/\sigma$ er $N(0,1)$-fordelt får vi:

$P(-\frac{a}{\sigma} < Z < \frac{a}{\sigma}) = \Phi(a/\sigma) - \Phi(-a/\sigma) = 2\Phi(a/\sigma) - 1 = ,95 \Rightarrow \Phi(a/\sigma) = ,975$

Fordi $\Phi(1{,}96) = ,975$ får vi at $a/\sigma = a/,245 = 1{,}96 \Rightarrow a = 1{,}96\cdot,245 = \underline{,48}$

Oppgave 5.39 (E) I langtids bølgestatistikk brukes ofte Weibull-fordelingen til å beskrive hvordan bølgehøyden varierer. La H betegne bølgehøyden (målt i meter) til en vilkårlig bølge. Da har H følgende fordelingsfunksjon:

$$F(h) = P(H \le h) = 1 - e^{-\frac{h^\beta}{\alpha}},\ h \ge 0$$

α og β er parametre som må estimeres ut fra de lokale forhold og observasjoner. I vårt tilfelle skal vi regne med $\alpha = 1{,}12$ og $\beta = 0{,}97$.

a) Hva er sannsynligheten for at en bølge har en høyde på over 4,5 m?

Hva er sannsynligheten for at høyst 2 av 100 bølger er over 4,5 m høye?

b) Bestem tilnærmet sannsynlighet for at høyst 200 av 10000 bølger er over 4,5 m høye.

c) La oss anta at det kommer en ny bølge hvert 7. sekund på det stedet vi betrakter. Hvor mange bølger over 19 meter vil en forvente å observere i løpet av en 10 års-periode (regn med 365 dager pr. år)? Hva er tilnærmet sannsynlighet for å observere minst 4 bølger på over 19 meter i løpet av en 10 års-periode?

Løsningsforslag:

a) $P(H > 4{,}5) = 1 - P(H < 4{,}5) = e^{-4{,}5^{,97}/1{,}12} = \underline{{,}0215}$

Y = antall bølger over 4,5 m av totalt n bølger. Antar at enketlbølgene kan behandles som om de var uavhengige. Da vil $Y \sim$ Bino(n,p) med n = 100 og p = ,0215, og vi får:

$P(Y \le 2) = P(Y = 0) + P(Y = 1) + P(Y = 2) =$
$(1-p)^{100} + 100 \cdot p \cdot (1-p)^{99} + \binom{100}{2} \cdot p^2 \cdot (1-p)^{98}$

Innsatt p = ,0215 får vi:

$P(Y \le 2) = {,}114 + {,}250 + {,}272 = \underline{{,}636}$

b) $Y \sim$ Bino(n,p) med n = 10000 og p = ,0215 $\Rightarrow np(1-p) \approx np = 215 >> 5$, dvs. Y tn. $N(np, \sqrt{np(1-p)})$.

$P(Y \le 200) \approx \Phi\left(\frac{200-215+,5}{\sqrt{215 \cdot ,9785}}\right) \approx \Phi(-1) = \underline{{,}159}$

c) $p = P(H > 19) = e^{-19^{,97}/1{,}12} = 1{,}8 \cdot 10^{-7}$
n = antall bølger iløpet av 10 år:
$n = \underbrace{10}_{\text{år}} \cdot \underbrace{365}_{\text{dager}} \cdot \underbrace{24}_{\text{timer}} \cdot \underbrace{60}_{\text{min}} \cdot \underbrace{60/7}_{\text{bølger/min}} = 4{,}5 \cdot 10^7$

M = antall bølger over 19 m pr. 10 år. $M \sim$ Bino(n,p)
E$M = np = 4{,}5 \cdot 10^7 \cdot 1{,}8 \cdot 10^{-7} = \underline{8}$
$P(M \ge 4) = 1 - P(N \le 3) \approx 1 - \Phi\left(\frac{3-8+,5}{\sqrt{8}}\right) = 1 - \Phi(-1{,}59) = \Phi(1{,}59) = \underline{{,}94}$

Oppgave 5.40 (E) Levetiden (i timer) til en bestemt type elektrisk komponent kan betraktes som en stokastisk variabel X med sannsynlighetstetthet $f_X(t) = \lambda e^{-\lambda t}$ for $t \ge 0$ og $f_X(t) = 0$ for $t < 0$. La $\lambda = 0{,}001$ i hele denne oppgaven.

a) Bestem $P(X>1000)$. Finn også L_{25} som er slik at $F_X(L_{25})$ = ¼ hvor F_X er kumulativ fordelingsfunksjon til X.

La X_1 og X_2 være levetiden til to elektriske komponenter av typen over. Vi forutsetter at X_1 og X_2 er stokastisk uavhengige. En kan da vise (det skal ikke du gjøre) at $Y = X_1+X_2$ har sannsynlighetstettheten $f_Y(t) = \lambda^2 \cdot te^{-\lambda t}$ for $t \ge 0$.

b) Vis at den kumulative fordelingsfunksjonen F_Y til Y er gitt ved $F_Y(y)$ = 1− $(1+\lambda y)e^{-\lambda y}$ for y ≥ 0. Bestem $P(Y > 2000)$.

La $T_{\min} = \min(X_1,X_2)$. $T_{\min}$ er altså den maksimale tiden der begge komponentene fungerer.

c) Vis at $P(T_{\min} < t) = 1 - e^{-2\lambda t}$ for $t \ge 0$.

Hva er sannsynligheten for at begge komponentene fungerer etter 800 timer?

Løsningsforslag:

a) $P(X > 1000) = 1 - P(X < 1000) = e^{-\lambda\cdot 1000} = e^{-,001\cdot 1000} = e^{-1} = \underline{,368}$

$F_X = 1 - e^{-\lambda t} = 1/4 \Rightarrow e^{-\lambda t} = 3/4 \Rightarrow t = L_{25} = -\ln(3/4)/\lambda = -1000\cdot\ln(3/4) = \underline{288}$

b) $F_Y(y) = \int_0^y \lambda^2 te^{-\lambda t}dt = \lambda^2 \int_0^y \underbrace{t}_{u} \cdot \underbrace{e^{-\lambda t}dt}_{dv} = \lambda^2 \left[\underbrace{t}_{u} \cdot \underbrace{\left(\tfrac{e^{-\lambda t}}{-\lambda}\right)}_{v} \right]_0^y - \lambda^2 \int_0^y \underbrace{\tfrac{e^{-\lambda t}}{-\lambda}}_{v} \cdot \underbrace{dt}_{du}$

Ved å integrere ut siste ledd og ordne ledd får vi svaret

$F_Y(y) = 1 - (1 + \lambda y)\cdot e^{-\lambda t}, \ y \geq 0,$ q.e.d.

$P(Y > 2000) = 1 - P(Y < 2000) = (1 - \lambda\cdot 2000)\cdot e^{-\lambda\cdot 2000} = 3e^{-2} = \underline{,406}$

c) $P(T_{\min} < t) = P(\min(X_1, X_2) < t) = 1 - P(\min(X_1, X_2) > t) =$

$1 - P(X_1 > t \cap X_2 > t) = 1 - (e^{-\lambda t})^2 = 1 - e^{-2\lambda t}, \ t \geq 0,$ q.e.d.

$P(T_{\min} > 800) = e^{-2\lambda\cdot 800} = e^{-1,6} = \underline{,202}$

Oppgave 5.41 En elektronisk bedrift trenger elektriske motstander i framstillingen av sine produkter. De motstandene som kjøpes inn har en resistans som er N(310, 10).

a) Beregn forventningsverdi og standardavvik for summen av resistansene for 5 slike motstander. Beregn sannsynligheten for at samlet resistans for de 5 motstandene overskrider 1600 ohm.

Ved en annen anledning trenger firmaet 80 slike motstander, men disse motstandene må hver ha en resistans mellom 300 og 330 ohm. Det kjøpes inn 100 motstander, som hver har resistans som er N(310, 10).

b) Hvor stor er sannsynligheten for å få minst 80 motstander med ønsket resistans?

c) Motstandene er pakket i esker á 10 stykker. Hva er sannsynligheten for at en eske skal inneholde eksakt 8 motstander med resistans mellom 300 og 330 ohm?

Løsningsforslag: X_i [ohm] = motstand til tilfeldig resistans, $X_i \sim N(310, 10)$, $X_1, X_2, \ldots$ uavhengige.

a) $Y = X_1 + \cdots + X_5$, $EY = 5\cdot EX_i = 5\cdot 310 = \underline{1550}$

Var(Y) = 5·Var(X_i) = $5\cdot 10^2$ = 500, Std(Y) = $\sqrt{500} = \underline{22,4}$

$P(Y > 1600) = 1 - P(Y < 1600) = 1 - \Phi\left(\tfrac{1600-1550}{22,4}\right) = 1 - \Phi(-2,23) = \underline{,013}$

b) Z = antall brukbare motstander av n = 100 tilfeldige. Z ~ Bino(n,p) med

$p = P(300 < X_i < 330) = P(X_i < 330) - P(X_i < 310) = \Phi\left(\tfrac{330-310}{10}\right) - \Phi\left(\tfrac{310-310}{10}\right) =$

$\Phi(2) - \Phi(-1) = ,977 - ,159 = ,818$

$np(1-p) = 14,9 > 5 \Rightarrow \ Z$ tn. $N(np, \sqrt{np(1-p)}) = N(81,8, 3,86)$

$$P(Z \geq 80) = 1 - P(Z \leq 79) \approx 1 - \Phi\left(\tfrac{79-81,8+,5}{3,86}\right) = 1 - ,274 = \underline{,73}$$

c) Q = antall brukbare av n = 10 tilfeldige reisistanser. $Q \sim$ Bino(n,p) med p = ,818.

$$P(Q = 8) = \binom{10}{8} \cdot ,818^8 \cdot ,182^2 = \underline{,30}$$

Oppgave 5.42 Tiden, Y, som går med for en dataterminal til å prosessere, editere og forandre på en storskjerm, er uniformt (rektangulært) fordelt mellom 0,75 og 2,5 sekunder.

a) Bestem sannsynlighetstetthetsfunksjonen for Y. Tegn den. Beregn forventning, μ, og standardavvik, σ, til Y.

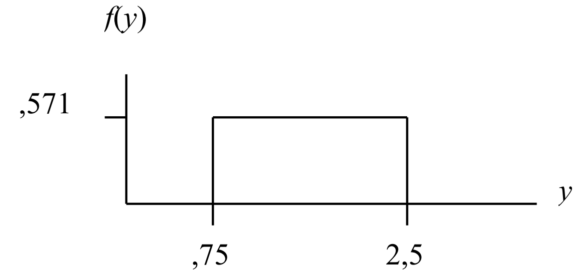
b) Bestem intervallet $(\mu \pm \sigma)$ på kurven. Beregn sannsynligheten for $(\mu - \sigma) < y < (\mu + \sigma)$.

<u>Løsningsforslag</u>:

a) $Y \sim$U(,75, 2,5), dvs.

$$f(y) = \frac{1}{b-a} = \frac{1}{2,5 - ,75} = \underline{,571,}$$

$\underline{,75 \leq y \leq 2,5}$



$$\text{E}Y = a + (b - a)/2 = ,75 + (2,5 - ,75)/2 = \underline{1,625}$$

$$\text{Std}(Y) = (b - a) / \sqrt{12} = (2,5 - ,75) / \sqrt{12} = \underline{,505}$$

b) $\mu \pm \sigma = 1,625 \pm ,505 = \underline{[1,12,\ 2,13]}$

$$P(\mu - \sigma < Y < \mu + \sigma) = (2,13 - 1,12) / 1,75 = \underline{,577}$$

Oppgave 5.43 En fabrikk produserer bolter. Boltenes diameter kan betraktes som en normalfordelt stokastisk variabel med forventning μ = 8,20 mm og standardavvik 0,14 mm.

a) Finn sannsynligheten for at en tilfeldig valgt bolt har diameter mindre enn 8,40 mm.

b) Beregn sannsynligheten for at eksakt 7 av 10 bolter har diameter mindre enn 8,40 mm.

c) Finn sannsynligheten for at en tilfeldig valgt bolt har diameter større enn 8,40 mm, gitt at diameteren er større enn 8,10 mm.

Boltene skal brukes sammen med muttere som produseres ved samme fabrikk. Mutrene har hull med diameter som antas normalfordelt med forventning 8,50 mm og standardavvik 0,12 mm. For at en mutter og en bolt skal passe sammen, må mutterens diameter være større enn boltens, men differansen må ikke overstige 0,60 mm.

d) Finn sannsynligheten for at en tilfeldig valgt bolt skal passe sammen med en tilfeldig valgt mutter.

Løsningsforslag: D = tilfeldig boltediameter. $D \sim N(8{,}20, {,}14)$.

a) $P(D < 8{,}40) = \Phi\left(\frac{8{,}40-8{,}20}{{,}14}\right) = \Phi(1{,}429) = \underline{{,}924}$

b) Y = antall bolter av n = 10 med D < 8,40. $Y \sim$ Bino(n,p) med n = 10 og p = ,924.
$P(Y = 7) = \binom{10}{7} \cdot {,}924^7 \cdot {,}076^3 = \underline{{,}030}$

c) $P(D > 8{,}40 \mid D > 8{,}10) = \frac{P(D>8{,}40 \cap D>8{,}10)}{P(D>8{,}10)} = \frac{P(D>8{,}40)}{P(D>8{,}10)}$
$P(D > 8{,}40) = 1 - P(D < 8{,}40) = 1 - {,}924 = {,}076$
$P(D > 8{,}10) = 1 - P(D < 8{,}10) = 1 - \Phi\left(\frac{8{,}10-8{,}20}{{,}14}\right) = 1 - \Phi(-{,}71) = \Phi({,}71) = {,}761$
$\Rightarrow P(D > 8{,}40 \mid D > 8{,}10) = {,}076/{,}761 = \underline{{,}10}$

d) H = hulldiameter, $H \sim$ N(8,50, ,12). $H - D \sim N(8{,}50 - 8{,}20, \sqrt{{,}14^2 + {,}12^2})$
$= N({,}30, {,}184)$.
$P(0 < H - D < {,}60) = \Phi\left(\frac{{,}60-{,}30}{{,}184}\right) - \Phi\left(\frac{0-{,}30}{{,}184}\right) = 2\Phi(1{,}63) - 1 = 2 \cdot {,}948 - 1 = \underline{{,}896}$

Oppgave 5.44 Et firma produserer betongelementer med lengder 3 m og 6 m. Lengdene kan betraktes som uavhengige og normalfordelte stokastiske variabler, henholdsvis

$X \sim N(3{,}00, 0{,}05)$ og $Y \sim N(6{,}00, 0{,}08)$.

a) Beregn sannsynligheten for at et 3-meterselement skal være kortere enn 2,90 m.

Hvor stor er sannsynligheten for at et 6-meterselement er lengre enn 6,15 m, dersom vi vet at lengden er større enn 6,10 m?

b) Et 6-meterselement settes sammen med et 3-meterselement. Beregn sannsynligheten for at den totale lengden blir mindre enn 8,90 m.

En husbygger skal dekke et spenn på 11,95 m. Hvilken av følgende kombinasjoner av elementer har størst sannsynlighet for å ha total lengde større enn 11,95 m:

i) (3+3+3+3) m,
ii) (3+3+6) m,
iii) (6+6) m?

Løsningsforslag:

a) $P(X < 2{,}90) = \Phi\left(\frac{2{,}90-3{,}00}{,05}\right) = \Phi(-2) = \underline{,023}$

$P(Y > 6{,}15 \mid Y > 6{,}10) = P(Y > 6{,}15) / P(Y > 6{,}10) =$

$\left(1-\Phi\left(\frac{6{,}15-6{,}00}{,08}\right)\right) / \left(1-\Phi\left(\frac{6{,}10-6{,}00}{,08}\right)\right) = (1-\Phi(1{,}875)) / (1-\Phi(1{,}25)) =$

$(1-{,}970) / (1-{,}894) = {,}03/{,}106 = \underline{,283}$

b) $Z = X + Y \sim N(9{,}0, \sqrt{{,}05^2 + {,}08^2}) = (9{,}0, {,}0943)$

$P(Z < 8{,}90) = \Phi\left(\frac{8{,}90-9{,}0}{,0943}\right) = \Phi(-1{,}06) = \underline{,145}$

Alternativene i), ii) og iii) har alle forventning 12 $> 11{,}95$. Det er størst sannsynlighet for total lengde $> 11{,}95$ for alternativet med minst varians, σ^2:

i) $\sigma^2 = 4 \cdot {,}05^2 = {,}0100$
ii) $\sigma^2 = 2 \cdot {,}05^2 + {,}08^2 = {,}0114$
iii) $\sigma^2 = 2 \cdot {,}08^2 = {,}0128$

Alternativ <u>i)</u> har minst varians.

Oppgave 5.45 Et meieri ønsker å framstille kartonger med kremfløte. Vektene i gram av tilfeldig valgte kartonger med kremfløte antas å være uif $N(\mu,\sigma)$. Ifølge forskriftene skal denne type kartonger veie minst 350 gram. Anta nå at en bestemt produksjonsplan innebærer at $\mu = 360$ og $\sigma = 8$, hvilket vil si at vekten X på kartongene har fordeling $N(360, 8)$.

a) Hva er sannsynligheten for at en tilfeldig valgt kartong er undervektig?

b) Hvis 3 kartonger kjøpes, hva er sannsynligheten for at minst en er undervektig?

c) Hvis 3 kartonger kjøpes, hva er sannsynligheten for at den tyngste kartongen veier mer enn 370 gram?

d) Bestem sannsynligheten for at den gjennomsnittlige vekten av fire tilfeldig valgte kartonger er mindre enn 350 gram.

<u>Løsningsforslag</u>:

a) $P(X < 350) = \Phi\left(\frac{350-360}{8}\right) = \Phi\left(-\frac{10}{8}\right) = \Phi(-1{,}25) = \underline{,106}$

b) Y = antall undervektige kartonger av n tilfeldige. $Y \sim$ Bino(n,p) med $n = 3$ og $p =$,106. $P(Y \geq 1) = 1 - P(Y = 0) = 1 - (1-{,}106)^3 = \underline{,285}$

c) $p = P(X > 370) = 1 - P(X < 370) = 1 - \Phi\left(\frac{370-360}{8}\right) = 1 - \Phi(1{,}25) = \Phi(-1{,}25) = {,}106$

Z = antall kartonger som veier mer enn 370 g av n tilfeldige.

$P(\max(X_1, X_2, X_3) > 370) = 1 - P(\max(X_1, X_2, X_3) < 370) =$

$1 - P(X_1 < 370 \cap X_2 < 370 \cap X_3 < 370) = 1 - (1-p)^3 = \underline{,285}$

d) $P(\overline{X} < 350) = P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{4}} < \frac{350-\mu}{\sigma/\sqrt{4}}\right) = \Phi\left(\frac{350-360}{8/2}\right) = \Phi(-2{,}5) = \underline{,0062}$

Kapittel 6

Estimering

Oppgave 6.1 La X_1 og X_2 være 2 stokastisk uavhengige variabler fra en og samme populasjon med populasjonsforventning μ og populasjons-standardavvik σ.

Du skal velge en av følgende 2 estimatorer for μ:

$$\mu_1^* = \frac{3X_1 + 7X_2}{10}, \quad \mu_2^* = \frac{X_1 + X_2}{2}$$

Vis at begge estimatorer er forventningsrette, og bestem hvilken som er best.

Løsningsforslag:

$$\mathrm{E}(\mu_1^*) = \mathrm{E}\left(\tfrac{1}{10}(3X_1 + 7X_2)\right) = \tfrac{1}{10}\mathrm{E}(3X_1 + 7X_2) = \tfrac{1}{10}(\mathrm{E}(3X_1) + \mathrm{E}(7X_2)) =$$
$$\tfrac{1}{10}\cdot(3\mathrm{E}X_1 + 7\mathrm{E}X_2) = \tfrac{1}{10}\cdot(3\cdot\mu + 7\cdot\mu) = \tfrac{10\mu}{10} = \mu, \quad \mathrm{E}(\mu_1^*) = \mu, \quad \underline{\text{q.e.d.}}$$

$$\mathrm{E}(\mu_2^*) = \mathrm{E}\left(\tfrac{1}{2}(X_1 + X_2)\right) = \tfrac{1}{2}\mathrm{E}(\mathrm{X}_1 + X_2) = \tfrac{1}{2}(\mathrm{E}X_1 + \mathrm{E}X_2) = \tfrac{1}{2}(\mu + \mu) = \mu, \quad \underline{\text{q.e.d.}}$$

$$\mathrm{Var}(\mu_1^*) = \mathrm{Var}\left(\tfrac{1}{10}(3X_1 + 7X_2)\right) = \left(\tfrac{1}{10}\right)^2 \cdot \mathrm{Var}(3X_1 + 7X_2) =$$
$$\tfrac{1}{100}\cdot(\mathrm{Var}(3X_1) + \mathrm{Var}(7X_2)) = \tfrac{1}{100}\cdot(3^2\,\mathrm{Var}X_1 + 7^2\,\mathrm{Var}X_2) = \tfrac{1}{100}\cdot(9\sigma^2 + 49\sigma^2) = \tfrac{29}{50}\sigma^2$$

$$\mathrm{Var}(\mu_2^*) = \mathrm{Var}\left(\tfrac{1}{2}(X_1 + X_2)\right) = \tfrac{1}{4}\mathrm{Var}(X_1 + X_2) = \tfrac{1}{4}\cdot(\mathrm{Var}X_1 + \mathrm{Var}X_2) = \tfrac{2\sigma^2}{4} = \tfrac{\sigma^2}{4} = \tfrac{25}{50}\sigma^2$$

$$\underline{\mu_2^* \text{ best}} \text{ fordi } \mathrm{Var}(\mu_2^*) = \tfrac{25}{50}\sigma^2 < \mathrm{Var}(\mu_1^*) = \tfrac{29}{50}\sigma^2.$$

Oppgave 6.2 La X_1, X_2, X_3, X_4, Y_1, Y_2 og Y_3 være 7 uavhengige stokastiske variabler med forventning μ og standardavvik σ. Betrakt følgende 2 estimatorer for μ:

$$\mu_1^* = \frac{1}{2}\left(\overline{X} + \overline{Y}\right), \quad \mu_2^* = \frac{1}{7}\left(4\overline{X} + 3\overline{Y}\right)$$

der $\overline{X}$ og $\overline{Y}$ er middelverdiene til henholdsvis X-ene og Y-ene. Vis at begge estimatorene er forventningsrette, og at μ_2^* er best.

Løsningsforslag:

$$\mathrm{E}\mu_1^* = \mathrm{E}\left(\tfrac{1}{2}(\overline{X} + \overline{Y})\right) = \tfrac{1}{2}\mathrm{E}(\overline{X} + \overline{Y}) = \tfrac{1}{2}(\mathrm{E}\overline{X} + \mathrm{E}\overline{Y}) = \tfrac{1}{2}\left(\mathrm{E}\left(\tfrac{1}{4}\Sigma X_i\right) + \mathrm{E}\left(\tfrac{1}{3}\Sigma Y_i\right)\right) =$$
$$\tfrac{1}{2}\left(\tfrac{1}{4}\mathrm{E}(\Sigma X_i) + \tfrac{1}{3}\mathrm{E}(\Sigma Y_i)\right) = \tfrac{1}{2}\left(\tfrac{1}{4}\Sigma(\mathrm{E}X_i) + \tfrac{1}{3}\Sigma(\mathrm{E}Y_i)\right) = \tfrac{1}{2}\left(\tfrac{1}{4}\cdot 4\mu + \tfrac{1}{3}\cdot 3\mu\right) = \mu, \quad \underline{\text{q.e.d.}}$$

$$\mathrm{E}(\mu_2^*) = \mathrm{E}\left(\tfrac{1}{7}\cdot(4\overline{X} + 3\overline{Y})\right) = \tfrac{1}{7}\cdot(4\mathrm{E}\overline{X} + 3\mathrm{E}\overline{Y}) = \tfrac{1}{7}\cdot(4\mu + 3\mu) = \mu, \quad \underline{\text{q.e.d.}}$$

$$\mathrm{Var}(\mu_1^*) = \mathrm{Var}\left(\tfrac{1}{2}(\overline{X} + \overline{Y})\right) = \tfrac{1}{4}(\mathrm{Var}(\overline{X}) + \mathrm{Var}(\overline{Y})) = \tfrac{1}{4}\left(\tfrac{\sigma^2}{4} + \tfrac{\sigma^2}{3}\right) = \tfrac{7}{48}\sigma^2$$

$$\mathrm{Var}(\mu_2^*) = \mathrm{Var}\left(\tfrac{1}{7}(4\overline{X} + 3\overline{Y})\right) = \tfrac{1}{49}\left(16\cdot\tfrac{\sigma^2}{4} + 9\cdot\tfrac{\sigma^2}{3}\right) = \tfrac{7}{49}\sigma^2$$

$\mathrm{Var}(\mu_2^*) = \frac{7}{49}\sigma^2 < \mathrm{Var}(\mu_1^*) = \frac{7}{48}\sigma^2 \Rightarrow \mu_2^*$ best, q.e.d.

Oppgave 6.3 La $\overline{X}$ være $N(\mu,\sigma/\sqrt{n})$.

a) Bestem et 95 % KI for μ når σ = 2,7 og n = 16 på basis av et utvalgsmiddel på $\bar{x} =$ 0,2.

b) Bestem et 99 % KI for μ når σ er ukjent, på basis av følgende tilfeldige x-verdier:

12,7	14,1	9,8	10,2	10,9
7,8	11,3	8,9	9,7	

c) En medisin har ukjent helbredelsesrate, p. Bestem et 95 % KI for p basert på en undersøkelse der 72 av 100 ble helbredet. Gjør rede for de forutsetninger du gjør.

Løsningsforslag:

a) $\bar{x} \mp z_{,025} \cdot \frac{\sigma}{\sqrt{16}} = 0{,}2 \mp 1{,}96 \cdot \frac{2{,}7}{4} = 0{,}2 \mp 1{,}32 \Rightarrow \underline{(-1{,}12,\ 1{,}52)}$ er et 95 % KI for μ.

b) $t_{,025}(n-1) = t_{,025}(8) = 2{,}306, \quad s = \mathrm{std}^*(x) = \sqrt{\frac{1}{8}\Sigma(x_i - \bar{x})^2} = 1{,}923$

$\bar{x} \mp t_{,025}(8) \cdot \frac{s}{\sqrt{9}} = 10{,}6 \mp 2{,}306 \cdot \frac{1{,}923}{\sqrt{9}} = 10{,}6 \mp 2{,}15 \Rightarrow \underline{(8{,}45,\ 12{,}75)}$ er et 99 % KI for μ.

c) La Y betegne antallet av et tilfeldig utvalg på n = 100 som blir helbredet, og la y = 72 være utfallet av et forsøk. Y vil da være tilnærmet binomisk fordelt med forventning $100p$ og varians $100p(1-p)$. $p^* = Y/100$ vil videre være en forventningsrett estimator for p med tilnærmet varians $p(1-p)/100$. Med basis i resultatet, vil det også være rimelig å anta at p^* er tilnærmet normalfordelt, fordi $np^*(1-p^*) = 72\cdot(1-{,}72) = 20 > 5$. Vi får derfor:

$p^* \mp z_{,025} \cdot \sqrt{p^*(1-p^*)/100} = {,}72 \mp 1{,}96 \cdot \sqrt{{,}72 \cdot {,}28/100} = {,}72 \mp 1{,}96 \cdot {,}088 \Rightarrow$

$\underline{({,}632,\ {,}808)} = (63{,}2\ \%,\ 80{,}8\ \%)$ er et tn. 95 % KI for p.

Oppgave 6.4

a) Ved en politisk gallup svarer x = 40 av n = 1600 at de vil stemme på et bestemt politisk parti. Beregn et tilnærmet 95 % konfidensintervall for partiets oppslutning.

b) Antall biler som passerer mellom kl. 12.00 og kl. 12.15 i en tett trafikkert envegskjøring antas å være Poisson- fordelt med ukjent parameter λ. Ved opptelling en dag passerer det 127 biler i det aktuelle tidsrommet. Bestem på basis av denne tellingen et tilnærmet 95 % KI for λ.

Løsningsforslag:

a) La X betegne antall som svarer de vil stemme på partiet av et tilfeldig utvalg på n = 1600. X er da tn. Bino(n, p). p^* = X/n er en forvetningsrett estimator for p. Setter vi inn resultatet x = 40, får vi at $np^*(1-p^*) = 40\cdot(1-40/1600) \approx 40 >> 5$. Det er derfor rimelig å anta at p^* er tn. $N(p, \sqrt{p^*(1-p^*)/n}) = N(p, ,00417)$. Et tn. 95 % KI for p er da gitt som følger:

$$p^* \mp z_{,025} \cdot ,00417 = p^* \mp 1{,}96 \cdot ,00417 = \underline{(1{,}7\ \%,\ \ ,3{,}3\ \%)}$$

b) La nå X betegne antall biler som passerer mellom kl. 12.00 og kl. 12.15 en tilfeldig dag. Vi forutsetter at X er tn. Po(λ). Utfra observasjonen x = 127 > 30 er det da rimelig å anta at X er tn. $N(\lambda, \sqrt{\lambda}) = N(\lambda, \sqrt{127})$. Et tn. 95 % KI for λ er da gitt som følger:

$$\lambda^* \mp z_{,025} \cdot \sqrt{\lambda^*} = 127 \mp 1{,}96 \cdot \sqrt{127} = \underline{(105, 149)}$$

Oppgave 6.5 Et skjenkested serverer halvlitere. La X betegne det nøyaktige innholdet (volumet) i en tilfeldig halvliter. Anta at std(X) = 1 cl (= 0,01 l) og at E(X) = μ er ukjent.

a) Hvor mange halvlitere må du undersøke for at bredden på et 95 % KI for μ skal være på 1 cl?

b) Bestem et 95 % KI for μ på basis av en undersøkelse med middelverdi $\bar{x}$ = 0,502 l, og der n = det antall du fant under a).

Løsningsforslag:

a) 95 % KI for μ gitt ved $\bar{x} \mp z \cdot \sigma / \sqrt{n}$, der σ = std(X), n er antall halvlitere og z = $z_{\alpha/2}$ = 1,96 er øvre 2,5 % fraktil i N(0,1)-fordelingen. Bredden på intervallet er da

$$(\bar{x} + z \cdot \sigma / \sqrt{n}) - (\bar{x} - z \cdot \sigma / \sqrt{n}) = 2z \cdot \sigma / \sqrt{n}$$
$$2z \cdot \sigma / \sqrt{n} = ,01 \Rightarrow n = (2z\sigma)^2 / ,01^2 = (2 \cdot 1{,}96 \cdot ,01)^2 / ,01^2 = 15{,}37$$

Følgelig: $\underline{n = 16}$ halvlitere må undersøkes

b) 95 % KI for μ: $\bar{x} \pm z\sigma / \sqrt{n} = 0{,}502 \mp 1{,}96 \cdot ,01 / \sqrt{16} = \underline{(,497\ ,\ ,507)\ \text{liter}}$

Oppgave 6.6 I en by er det N drosjer, og drosjene er nummerert fra 1 til N med nummerskilt på taket. En tilreisende står på byens flyplass og venter på drosje. Han bruker tida til å observere taknummeret til de ankommende drosjer, og registrerer følgende nummer:

1 58 21 5 14

På basis av tallene ovenfor anslår han antallet drosjer i byen til å være 69. Han forteller dette til drosjesjåføren, som blir mektig imponert, fordi det nettopp er 69 drosjer i byen.

a) La X betegne et tilfeldig drosjenummer. Angi populasjon og populasjonsfordeling, $f(x)$, i dette tilfellet.

Det kan vises at forventningen til den største av n tilfeldige nummer, $X_{(n)} = X_{\text{maks}}$, er gitt ved følgende uttrykk:

$$E\left(X_{(n)}\right) = \frac{n}{n+1}(N+1)$$

b) Benytt opplysningen ovenfor til å konstruere en forventningsrett estimator, N^*, for N, som funksjon av n og $X_{(n)}$. Undersøk om du får samme estimat som den tilreisende når du anvender din estimator på tallmaterialet til den tilreisende.

Løsningsforslag:

a) Populasjonen er her mengden bestående av alle drosjenumrene: $S = \{1,2,\ldots,N\}$.

$$f(x) = \frac{1}{N}, \quad x = 1,2,\ldots,N$$

b) $E\left(X_{(n)}\right) = \frac{n}{n+1}(N+1) \Rightarrow N = \frac{n+1}{n}E\left(X_{(n)}\right) - 1$

dvs. $\underline{N^* = \frac{n+1}{n}X_{(n)} - 1}$ er en forventningsrett estimator for N (vis dette selv).

$X_{(5)} = 58 \Rightarrow N^* = \frac{6}{5}\cdot 58 - 1 = 68{,}6 \approx 69$, dvs. samme estimat som tilreisende.

Oppgave 6.7 (E) Seks målinger ble utført for å bestemme en konstant m. Målingene kan antas å være observasjoner av en normalfordelt stokastisk variabel med forventning m og et kjent standardavvik s. Et 95 % konfidensintervall for m blir funnet til (6,15, 6,25).

a) Finn s.

Hvor mange flere målinger må utføres for å få et konfidensintervall som har:

b) konfidensgrad på minst 98 % og samme bredde?

c) konfidensgrad på minst 95 % og halvparten av bredden?

Løsningsforslag:

a) Bredden på konfidensintervallet er 6,25 – 6,15 = 0,10. Vi får derfor:

$$2\cdot z_{,025}\cdot s/\sqrt{6} = 0{,}10 \Rightarrow s = 0{,}10\cdot\sqrt{6}/(2\cdot 1{,}96) = \underline{,0625}$$

b) $2\cdot z_{,01}\cdot s/\sqrt{n} = ,10 \Rightarrow n = (2\cdot 2{,}326\cdot ,0625)^2/,10^2 = 8{,}45 \approx 9$, dvs. vi må foreta $9 - 6 = \underline{3}$ flere målinger for å få konfidensgrad på minst 98 %.

c) $2\cdot z_{,025}\cdot s/\sqrt{n} = ,05 \Rightarrow n = (2\cdot 1{,}96\cdot ,062)^2/,05^2 = 23{,}6 \approx 24$, dvs. vi må foreta $24 - 6 = \underline{18}$ flere målinger for å få konfidensgrad på minst 95 % og halv bredde.

Oppgave 6.8 (E) En potetmelfabrikk kjøper poteter levert i standardsekker. Av forskjellige grunner vil vekten av en sekk poteter variere fra sekk til sekk.

Vi antar at vekten av en tilfeldig valgt sekk poteter kan betraktes som normalfordelt $N(\mu,\sigma)$. Videre antar vi at vektene til forskjellige sekker er uavhengige.

Anta at $\mu = 50{,}5$ kg og $\sigma = 0{,}9$ kg.

a) Hvor stor er sannsynligheten for at en sekk veier mindre enn 50 kg?

b) Hvor stor er sannsynligheten for at totalvekten av 25 sekker overstiger 1255 kg?

c) Hvor stor er sannsynligheten for at minst en av tre tilfeldig valgte sekker skal veie mindre enn 50 kg?

Anta at μ er ukjent og at $\sigma = 0{,}9$ kg. Veiing av 9 tilfeldige sekker gav følgende vekter i kg:

50,9 48,8 50,6 51,2 51,4 50,7 49,8 49,2 51,4

d) Bestem et 95 % konfidensintervall for μ.

e) Bestem også intervallet vi får dersom σ er ukjent.

f) Hvor mange målinger måtte vi minst gjort dersom $\sigma = 0{,}9$ antas kjent og vi ville ha et 95 % konfidensintervall med en lengde mindre enn 0,9 kg?

Løsningsforslag:

La X_i betegne vektinnhold til en tilfeldig valgt potetsekk, $X_i \sim N(50{,}5\ ,\ {,}9)$.

a) $P(X_i < 50) = \Phi\left(\frac{50-50{,}50}{{,}9}\right) = \Phi(-{,}56) = \underline{{,}289}$

b) $S = X_1 + \cdots + X_{25} \sim N(25\mu, \sqrt{25}\sigma) = N(1262{,}5\ , 4{,}5)$.

$P(S > 1255) = \Phi\left(-\frac{1255-1262{,}5}{4{,}5}\right) = \Phi(1{,}67) = \underline{{,}952}$

c) La Y betegne antall av $n = 3$ tilfeldige 50 kg sekker med vekt mindre enn 50 kg. Y er da binomisk fordelt Bino(3, p) med $p = {,}289$ fra a).

$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{3}{0} \cdot {,}289^0 \cdot (1 - {,}289)^3 = \underline{{,}642}$

Empirisk middelverdi $\bar{x}$ og standardavvik s til de $n = 9$ oppgitte vektene er henholdsvis $\bar{x} = 50{,}44$ og $s = {,}959$. La $z = 1{,}96$ og $t = 2{,}306$ betegne øvre 2,5 % fraktil i henholdsvis $N(0,1)$-fordelingen og stud(8)-fordelingen.

d) 95 % KI for μ med kjent standardavvik $\sigma = 0{,}9$ er

$\bar{x} \mp z \cdot \sigma / \sqrt{n} = 50{,}44 \mp 1{,}96 \cdot {,}9 / 3 = \underline{(49{,}9\ , 51{,}0)\ \text{kg}}$

e) 95 % KI for μ med estimert standardavvik $s = 0{,}959$ er

$\bar{x} \mp t \cdot s / \sqrt{n} = 50{,}44 \mp 2{,}306 \cdot {,}959 / 3 = \underline{(49{,}7\ , 51{,}2)\ \text{kg}}$

f) $2z\sigma/\sqrt{n} <,9 \Rightarrow n > (2z\sigma)^2/,9^2 = (2\cdot 1{,}96\cdot,9)^2/,9^2 = 15{,}37$, dvs. vi måtte minst foretatt $\underline{n=16}$ målinger for å få et 95 % KI med mindre bredde enn 0,9 kg.

Oppgave 6.9 (E) Omfattende målinger av surhetsgraden i en innsjø en tid tilbake, gav grunnlag for å anta at pH-verdien X for en enkeltmåling etter denne målemetoden er normalfordelt $N(5,1)$.

a) Finn sannsynligheten for at gjennomsnittet $\overline{X}=\frac{1}{9}\Sigma x_i$ av 9 uavhengige målinger er mindre enn 4,5.

En tid etter ble det observert et større antall døde fisk, og det ble besluttet å måle pH-verdien på nytt etter en annen målemetode. Resultatet av en enkeltmåling antas å være normalfordelt $N(\mu,\sigma)$, hvor μ er «sann pH-verdi». 9 uavhengige målinger av pH-verdien ga verdiene:

4,3 3,7 4,6 3,5 3,9 4,0 3,8 4,5 3,7

b) Estimer både forventet pH-verdi og standardavvik for en enkeltmåling ved å ta utgangspunkt i forventningsrette estimatorer for μ og σ^2, og i observasjonsmaterialet ovenfor.

c) Bestem et 90 % konfindensintervall for «sann pH-verdi» på grunnlag av de estimerte verdiene i punkt b).

<u>Løsningsforslag</u>:

a) $\overline{X} \sim N(5{,}1/\sqrt{9}) = N(5{,}1/3)$. $P(\overline{X} < 4{,}5) = \Phi\left(\frac{4{,}5-5}{1/3}\right) = \Phi(-1{,}5) = \underline{,0668}$

b) $\bar{x} = \frac{1}{9}\Sigma x = \underline{4{,}00}$, $\quad s^2 = \frac{1}{8}(\Sigma x^2 - (\Sigma x)^2/9) = ,1475$, $s = \sqrt{s^2} = \underline{,384}$

c) $t = 1{,}860$ = øvre 5 % fraktil i stud(8)-fordelingen. 90 % KI for μ blir da:
$\bar{x} \mp t\cdot s/\sqrt{n} = 4{,}00 \mp 1{,}860\cdot,384/\sqrt{9} = \underline{(3{,}76\,,\,4{,}24)}$

Oppgave 6.10 (E) I en stor by blir det utført en meningsmåling for å undersøke om innbyggerne vil ha kristen formålsparagraf i byens barnehager. En velger tilfeldig $n = 1225$ innbyggere og lar X være antallet blant disse, betraktet som stokastisk variabel, som vil ha en slik formålsparagraf. Undersøkelsen resulterte i verdien x = 735. La p være andelen av innbyggerne i byen som ønsker kristen formålsparagraf og velg $p^* = X/n$ som estimator for p.

a) Vis at p^* er forventningsrett og bestem dens varians. Begrunn at p^* er tilnærmet normalfordelt.

b) Betrakt Var(p^*) som en funksjon i p og finn dens maksimum. Vis at std(p^*) $\leq \dfrac{1}{2\sqrt{n}}$

c) Vis at et 95 % konfidensintervall for p (tilnærmet) er gitt ved

$(\, p^* - 1{,}96 \cdot \mathrm{std}(p^*)\, ,\, p^* + 1{,}96 \cdot \mathrm{std}(p^*)\,)$

og bestem et tn. 95 % KI for p.

Løsningsforslag:

Det er her rimelig å anta at X er tn. binomisk fordelt Bino(n, p) med $n = 1225$ og ukjent p.

a) $\mathrm{E}(p^*) = \mathrm{E}(X/n) = \frac{1}{n}\mathrm{E}(X) = \frac{1}{n} \cdot np = \underline{p}$

$V(p) = \mathrm{Var}(p^*) = \mathrm{Var}(X/n) = \frac{1}{n}\mathrm{Var}(X) = \frac{1}{n} \cdot np(1-p) = \underline{p(1-p)/n}$

$np^*(1-p^*) = 735 \cdot (1 - 735/1225) >> 5 \Rightarrow X$ er tn. normalfordelt. Dermed er også $p^* = X/n$ tn. normalfordelt.

b) $\dfrac{\mathrm{d}V}{\mathrm{d}p} = \frac{1}{n} \cdot (1 \cdot (1-p) + p \cdot (-1)) = \dfrac{1-2p}{n}, \qquad \dfrac{\mathrm{d}V}{\mathrm{d}p} > 0 \Rightarrow 1-2p > 0 \Rightarrow p < 1/2$

dvs. V har maks. verdi for $p = ½$ og

$V(p) \le V(1/2) = \dfrac{\frac{1}{2} \cdot (1-\frac{1}{2})}{n} = \dfrac{1}{4n}, \ \mathrm{std}(p^*) \le \dfrac{1}{2\sqrt{n}}, \ \underline{\text{q.e.d.}}$

c) Siden p^* er tn. $N(\mu,\sigma)$ med μ = p og $\sigma = \mathrm{std}(p^*)$, så vil $p^* \mp 1{,}96 \cdot \mathrm{std}(p^*)$ være et tn. 95 % KI for p, der 1,96 er øvre 2,5 % fraktil i $N(0,1)$-fordelingen.

Et konservativt 95 % KI for p får ved å erstatte den ukjente $\mathrm{std}(p^*)$ med $1/(2\sqrt{n})$, og vi får da:

$p^* \mp 1{,}96 \cdot \frac{1}{2\sqrt{1225}} = \underline{(57{,}2 \ ,\ 62{,}8)\ \%}$

Et litt mindre pålitelig tn. 95 % KI for p med smalere bredde fås ved å erstatte $\mathrm{std}(p^*)$ med

$\mathrm{std}^*(p^*) = \sqrt{p^*(1-p^*)/n} = \sqrt{(735/1225) \cdot (1 - 735/1225)/1225} = {,}014$

og vi får da intervallet $p^* \mp 1{,}96 \cdot {,}014 = \underline{(57{,}3\, ,\, 62{,}7)\ \%}$

Oppgave 6.11 (E) I et fylke er det 3 videregående skoler. Det har vært en del langtidsfravær (14 dager eller mer) blant lærerne en tid, og ledelsen vil av budsjettmessige årsaker anslå sannsynligheten p for at en tilfeldig valgt lærer vil ha minst ett langtidsfravær i løpet av en 1-årsperiode. Sannsynligheten p antas å være den samme for alle lærerne, uansett skole, og fravær for en lærer er uavhengig av fravær for en annen lærer.

Anta at det i løpet av 1 år er n_i lærere ved skole nr i (i = 1,2,3), hvorav et antall på X_i lærere har minst ett langtidsfravær. $X = \Sigma X_i$ blir da antall lærere ved alle skolene tilsammen som har hatt slikt fravær i perioden.

a) Bestem fordelingen til X og til hver X_i. Skriv opp forventningen og variansen til disse.

To estimatorer er foreslått:

$$p_1^* = X/n, \text{ der } n = n_1 + n_2 + n_3 \text{ og } p_2^* = \frac{1}{3}\sum_{i=1}^{3}\frac{X_i}{n_i}$$

b) Vis at begge estimatorene er forventningsrette og bestem $\mathrm{Var}\left(p_1^*\right)$. Vis at

$$\mathrm{Var}\left(p_2^*\right) = \frac{p(1-p)}{9}\sum_{i=1}^{3}\frac{1}{n_i}$$

For et bestemt år har vi følgende data:

n_i	30	40	60
X_i	3	3	5

c) Hvilken estimator vil du foretrekke?

Rapporter tilbake estimat av p ± estimert standardavvik av estimator.

Løsningsforslag:

a) $X_i \sim \mathrm{Bino}(n_i, p),\ i = 1,2,3.$ $X \sim \mathrm{Bino}(n, p),\ n = n_1 + n_2 + n_3.$
$\mathrm{E}X_i = n_i p,\ \mathrm{Var}X_i = n_i p(1-p),\ \mathrm{E}X = np,\ \mathrm{Var}X = np(1-p)$

b) $\mathrm{E}p_1^* = \mathrm{E}(X/n) = \frac{1}{n}\mathrm{E}X = \frac{1}{n}\cdot np = p$

$\mathrm{E}p_2^* = \mathrm{E}\left(\frac{1}{3}\Sigma\frac{X_i}{n_i}\right) = \frac{1}{3}\mathrm{E}\left(\Sigma\frac{X_i}{n_i}\right) = \frac{1}{3}\Sigma\mathrm{E}\left(\frac{X_i}{n_i}\right) = \frac{1}{3}\Sigma\frac{1}{n_i}\mathrm{E}X_i = \frac{1}{3}\Sigma\frac{1}{n_i}\cdot n_i p = \frac{1}{3}\Sigma p = \frac{1}{3}\cdot 3p = p$

$\mathrm{Var}(p_1^*) = \mathrm{Var}(X/n) = 1/n^2\cdot\mathrm{Var}X = \frac{np(1-p)}{n^2} = p(1-p)/n$

$\mathrm{Var}(p_2^*) = \mathrm{Var}\left(\frac{1}{3}\Sigma\frac{X_i}{n_i}\right) = \frac{1}{9}\Sigma\mathrm{Var}\left(\frac{X_i}{n_i}\right) = \frac{1}{9}\Sigma\frac{\mathrm{Var}X_i}{n_i^2} = \frac{1}{9}\Sigma\frac{n_i p(1-p)}{n_i^2} = \frac{p(1-p)}{9}\Sigma\frac{1}{n_i}$

c) $\mathrm{Var}(p_1^*) = p(1-p)/130$

$\mathrm{Var}(p_2^*) = \frac{p(1-p)}{9}\cdot\left(\frac{1}{30}+\frac{1}{40}+\frac{1}{60}\right) = \frac{p(1-p)}{9}\cdot\left(\frac{4+3+2}{120}\right) = \frac{p(1-p)}{120} > \frac{p(1-p)}{130}$

Foretrekker p_1^* siden denne har minst varians.

$$p_1^* = X/n = 11/130 = 8{,}46\,\%,\ \ \mathrm{std}^*(p_1^*) = \sqrt{\frac{\frac{11}{130}\cdot\left(1-\frac{11}{130}\right)}{130}} = 2{,}4\,\%$$

$$p_1^* \mp \mathrm{std}^*(p_1^*) = \underline{(6{,}0\,,10{,}9)\,\%}$$

Oppgave 6.12 (E) Vi skal i denne oppgaven se på behandling med et bestemt medikament på en gruppe av n pasienter. La X_i være en stokastisk variabel, slik at $X_i = 1$ dersom pasient i blir helbredet og $X_i = 0$ dersom pasient i ikke blir helbredet ($i = 1,\ldots,n$). Anta videre at $P(X_i = 1) = p$, $P(X_i = 0) = 1-p$ og at pasientene helbredes uavhengig av hverandre.

a) Beregn forventning og varians for X_i.

b) Forklar hva den stokastiske variabelen $X = X_1 + X_2 + \cdots + X_n$ beskriver.
Hvilken fordeling har X?

c) Vis at $\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ er en forventningsrett estimator for p.

d) Begrunn at $\overline{X}$ tn. $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$
dersom n er «tilstrekkelig stor».

e) På grunnlag av tidligere erfaring vet vi at $p = 0{,}6$. 1000 pasienter blir behandlet. Hvor stor er sannsynligheten (tilnærmet) for at minst 630 personer blir helbredet?

Løsningsforslag:

a) $\mathrm{E}X_i = \Sigma x \cdot P(X_i = x) = 0 \cdot (1-p) + 1 \cdot p = \underline{p}$
$\mathrm{E}X_i^2 = \Sigma x^2 \cdot P(X_i = x) = 0^2 \cdot (1-p) + 1^2 \cdot p = p$
$\mathrm{Var}(X_i) = \mathrm{E}X_i^2 - \mathrm{E}^2 X_i = p - p^2 = \underline{p(1-p)}$

b) X = antall av n pasienter som helbredes, $X \sim \mathrm{Bino}(n, p)$.

c) $\mathrm{E}\overline{X} = \mathrm{E}\left(\frac{1}{n} \cdot (X_1 + \cdots + X_n)\right) = \frac{1}{n}\mathrm{E}(X_1 + \cdots + X_n) = \frac{1}{n} \cdot (\underbrace{p + \cdots + p}_{n}) = \frac{1}{n} \cdot np = p$

d) n tilstrekkelig stor $\Rightarrow$ X tn. $N(np, \sqrt{np(1-p)})$. Da vil også $\overline{X}$ være tn. normalfordelt med $\mathrm{Var}\overline{X} = \frac{1}{n^2}\mathrm{Var}X = \frac{1}{n^2} \cdot np(1-p) = p(1-p)/n$, og følgelig vil
$\overline{X}$ tn. $N(p, \sqrt{p(1-p)/n})$

e) $P(X \geq 630) = 1 - P(X \leq 629) \approx 1 - \Phi\left(\frac{629 - np + ,5}{\sqrt{np(1-p)}}\right) =$
$1 - \Phi\left(\frac{629 - 600 + ,5}{\sqrt{,6 \cdot ,4/1000}}\right) = 1 - \Phi(1{,}90) = 1 - ,971 = \underline{,029}$

Oppgave 6.13 (E) Ventetiden T (målt i minutter) etter tilfeldig ankomst til en viss bussholdeplass antas å være uniformt fordelt med sannsynlighetstetthet

$$f(t) = \begin{cases} 1/\theta\,, & 0 < t < \theta \\ 0, & \text{ellers} \end{cases}$$

der θ er en ukjent parameter (konstant).

a) Hva beskriver parameteren θ i ventetidssituasjonen?
Bestem $P(T < \theta/4)$ og $P(T = \theta/2)$.

b) Vis at forventning og varians til T er gitt ved $\mathrm{E}(T) = \theta/2$ og $\mathrm{Var}(T) = \theta^2/12$.

La $T_1, \ldots, T_n$ være n uavhengige observasjoner med samme fordeling som T.

c) Vis at $\theta^* = \dfrac{2}{n}\sum_{i=1}^{n} T_i$

er en forventningsrett estimator for θ. Bestem variansen til θ^*.

La $T_{\max}$ betegne den største av observasjonene $T_1, T_2, \ldots, T_n$. En kan da vise at

$$\mathrm{E}(T_{\max}) = \frac{n\theta}{n+1},$$

$$\mathrm{Var}(T_{\max}) = \frac{n\theta^2}{(n+2)(n+1)^2}$$

d) La $\theta^{**} = \frac{1}{n}(n+1)T_{\max}$. Beregn θ^* og θ^{**} når følgende data er gitt:

T_1	T_2	T_3	T_4	T_5
7,5	1,0	4,0	12,5	8,5

Hvilken av estimatorene θ^* og θ^{**} er best? Begrunn svaret.

Løsningsforslag:

a) θ er tid mellom to påfølgende bussavganger

$$P(T < \theta/4) = \int_0^{\theta/4} f(t)\,dt = \tfrac{1}{\theta}\int_0^{\theta/4} dt = \tfrac{1}{\theta}\left[t\right]_0^{\theta/4} = \underline{1/4}$$

$$P(T = \theta/2) = \underline{0} \quad (T \text{ kontinuerlig})$$

b) $$\mathrm{E}T = \int_0^{\theta} t \cdot f(t)\,dt = \tfrac{1}{\theta}\int_0^{\theta} t\,dt = \tfrac{1}{\theta}\cdot\left[\tfrac{t^2}{2}\right]_0^{\theta} = \underline{\theta/2}$$

$$\mathrm{E}T^2 = \int_0^{\theta} t^2 \cdot f(t)\,dt = \tfrac{1}{\theta}\int_0^{\theta} t^2\,dt = \tfrac{1}{\theta}\cdot\left[\tfrac{t^3}{3}\right]_0^{\theta} = \theta^2/3$$

$$\mathrm{Var}T = \mathrm{E}T^2 - \mathrm{E}^2T = \theta^2/3 - (\theta/2)^2 = \underline{\theta^2/12}$$

c) $$\mathrm{E}\left(\tfrac{2}{n}\Sigma T_i\right) = \tfrac{2}{n}\Sigma \mathrm{E}T_i = \tfrac{2}{n}\Sigma\tfrac{\theta}{2} = \tfrac{2}{n}\cdot n\cdot\tfrac{\theta}{2} = \underline{\theta}$$

$$\mathrm{Var}\theta^* = \mathrm{Var}\left(\tfrac{2}{n}\Sigma T_i\right) = \tfrac{4}{n^2}\Sigma\mathrm{Var}T_i = \tfrac{4}{n^2}\Sigma\tfrac{\theta^2}{12} = \tfrac{4}{n^2}\cdot n\cdot\tfrac{\theta^2}{12} = \underline{\theta^2/(3n)}$$

d) $$\theta^* = \tfrac{2}{5}\cdot(7{,}5 + \cdots + 8{,}5) = \underline{13{,}4}, \quad \theta^{**} = \tfrac{6}{5}\cdot 12{,}5 = \underline{15{,}0}$$

$$\mathrm{E}\theta^{**} = \mathrm{E}\left(\tfrac{1}{n}\cdot(n+1)\cdot T_{\max}\right) = \tfrac{n+1}{n}\cdot\mathrm{E}T_{\max} = \tfrac{n+1}{n}\cdot\tfrac{n\theta}{n+1} = \theta$$

θ^{**} er følgelig forventningsrett, dvs. $\mathrm{E}\theta^{**} = \theta$.

$$n = 5 \Rightarrow Var\theta^* = \theta^2/(3\cdot 5) = \theta^2/15, \quad Var\theta^{**} = 5\cdot\theta^2/(7\cdot 6^2) = \theta^2/50{,}4$$

$$\mathrm{Var}\theta^{**} < \mathrm{Var}\theta^* \Rightarrow \underline{\theta^{**}\text{ best}}$$

Oppgave 6.14 (E) En maskin produserer elektriske komponeneter der hver av komponentene har en viss motstand (målt i ohm). På grunn av tilfeldigheter under produksjonen, vil det kunne variere litt hvor stor motstand en komponent får.

a) Erfaring tyder på at motstandene har verdier som er normalfordelt $N(45,\ 0{,}70)$ og at motstandene til forskjellige komponenter er uavhengige av hverandre.

Hva er sannsynligheten for at en tilfeldig valgt komponent har en motstand på mellom 44,2 ohm og 45,7 ohm?

b) Anta i dette punkt at verdiene på motstandene er uavhengige realiseringer av den normalfordelte stokastiske variabel $X \sim N(\mu,\ 0{,}70)$, men at vi ikke kjenner størrelsen på μ. Vi tar en tilfeldig stikkprøve på 6 komponenter og måler motstanden i hver av dem.

Måleresultat i ohm:

45,4 45,7 46,1 44,9 45,2 44,6

Forklar hva vi forstår med *medianen*, m, til X.
Bestem et 97 % konfidensintervall for m.
Forklar hva dette konfidensintervallet forteller oss.

c) Anta nå at vi har fått nye opplysninger som har fått oss til å tvile på at X er normalfordelt, men at vi fortsatt har grunn til å tro at verdiene på motstandene kan ses som uavhengige realiseringer av samme stokastiske variabel X.

Bestem nå, ut fra måleresultatene ovenfor, et konfidensintervall for m med konfidensgrad så nær 97 % som mulig. Angi konfidensgraden så nøyaktig det lar seg gjøre.

Kommentér dernest forskjellen mellom lengdene av konfidensintervallene i punktene b) og c).

Løsningsforslag:

X = motstand i tilfeldig komponent, $[X] = \Omega$.

a) $X \sim N(45\ ,\ ,70)$
$P(44{,}2 < X < 45{,}7) = \Phi\left(\frac{45{,}7-45}{,7}\right) - \Phi\left(\frac{44{,}2-45}{,7}\right) = \Phi(1) - \Phi(-1) = 2\cdot{,}8413 - 1 = \underline{,683}$

b) $X \sim N(\mu\ ,\ ,70)$, μ ukjent. Median, m, definert ved at $P(X < m) = ½$. Siden normalfordelingen er symmetrisk så er $m = \mu$

$\bar{x} = \frac{1}{6}\cdot(45{,}4 + \cdots + 45{,}6) = 45{,}32$. $z = z_{,015} = 2{,}17$ = øvre 1,5 % fraktil i $N(0,1)$-fordelingen. 97 % KI for m (og μ) er da
$\bar{x} \mp z_{,015}\cdot\sigma/\sqrt{n} = 45{,}32 \mp 2{,}17\cdot{,}70/\sqrt{6} = \underline{(44{,}7\ ,\ 45{,}9)}$

Konfidensintervallet kan tolkes som følger: Dersom $m \leq 44{,}7\ \Omega$ ville det bare vært 1,5 % sannsynlighet for å få vår observerte middelverdi $\bar{x} = 45{,}32\ \Omega$ eller større verdier. Tilsvarende: Dersom $m \geq 45{,}9\ \Omega$ ville det bare vært 1,5 % sannsynlighet for å få vår observerte middelverdi eller mindre verdier. Litt grovt kan vi si at vi er «97 % sikre» på at den sanne og ukjente medianen ligger mellom 44,7 Ω og 45,9 Ω.

c) Vi antar at vi kan benytte sentralgrenseteoremet og at $\overline{X}$ fremdeles er normalfordelt, selv om X ikke er det. Videre antar vi at $(\overline{X}-\mu)/(S/\sqrt{n})$ er stud(n–1)-fordelt, der S er empirisk standardavvik. Vi må også anta at fordelingen til X er symmetrisk, slik at median og forventning faller sammen. Empirisk standardavvik til det oppgitte tallmaterialet er s = ,54. La $t = t_{,015}$ = 3,00 angi øvre 1,5 % fraktil i stud(5)-fordelingen. Et 97 % KI for m blir da under de angitte forutsetninger

$$\bar{x} \mp t_{,015} \cdot s/\sqrt{n} = 45{,}32 \mp 3{,}00 \cdot ,54/\sqrt{6} = \underline{(44{,}6\ ,\ 46{,}0)}$$

Konfidensintervallet er litt bredere enn i *b*), hvilket er forventet fordi usikkerheten øker når σ er ukjent.

Oppgave 6.15 (E) Ved en bestemt politisk meningsmåling velger en tilfeldig $n = 1600$ personer med stemmerett, og spør om hvilket parti de ville stemme på dersom det var valg i morgen. La X være antall Arbeiderparti-velgere blant de 1600 spurte, og la p være andelen av Ap-velgere i hele velgermassen.

a) Hvilken sannsynlighetsfordeling har X? Begrunn svaret.

b) Angi en estimator for p. Bestem estimatorens forventning og varians. Er estimatoren forventningsrett?

c) Ved undersøkelsen viste det seg at 37,2 % i utvalget ville stemme på Arbeiderpartiet. Er dette en svært sterk indikasjon på at oppslutningen om partiet i befolkningen er større enn 36 %?

Besvar spørsmålet ved først å beregne en tilnærmet verdi for $P(p^* > 0{,}372 \mid p = 0{,}36)$.

d) Bestem et tilnærmet 95 % konfidensintervall for p. Hvor stor måtte utvalgsstørrelsen n minst være for at et slikt tilnærmet 95 % konfidensintervall ikke skulle inneholde 0,36?

<u>Løsningsforslag</u>:

a) X antas binomisk fordelt Bino(n, p) fordi det antas tilfeldig utvalg fra en populasjon med størrelse $N >> n = 1600$. X er derfor tn. normalfordelt dersom $np(1-p) > 5$, hvilket er rimelig å anta fordi n er stor og p neppe er særlig nære 0 eller 1.

b) $p^* = X/n,\ \mathrm{E}(p^*) = E(X/n) = \frac{1}{n} \cdot EX = \frac{1}{n} \cdot np = p,$

$\mathrm{Var}(p^*) = \mathrm{Var}(X/n) = \frac{1}{n^2} \cdot \mathrm{Var}X = \frac{1}{n^2} \cdot np(1-p) = p(1-p)/n$

p^* er forventningsrett fordi $\mathrm{E}(p^*) = p$.

c) $P(p^* > ,372 \mid p = ,36) = 1 - P(p^* < ,372 \mid p = ,36) \approx$

$$1-\Phi\left(\frac{,372-,360}{\sqrt{,372\cdot,628/1600}}\right)=1-\Phi(,99)=1-,839=,161=16,1\ \%$$

Det er med andre ord 16,1 % sannsynlighet for å få resultatet 37,2 % eller høyere dersom p = 36,0 %, hvilket neppe kan karakteriseres som en sterk indikasjon på økt oppslutning.

d) La z = 1,96 betegne øvre 2,5 % fraktil i N(0,1)-fordelingen. Et 95 % KI for p er da $p^* \mp z\cdot\sqrt{p^*(1-p^*)/n} =,372\mp 1,96\cdot\sqrt{,372\cdot,628/1600} = \underline{(34,8\ ,\ 39,6)\ \%}$

For at intervallet ikke skulle dekke 36,0 % måtte intervallet vært en faktor (37,2-36)/(37,2-34,8) = 0,5 smalere. Siden bredden avtar omvendt proporsjonalt med kvadratrota av n, måtte følgelig n firedobles, dvs. $\underline{n=6400}$.

Oppgave 6.16 (E) En personaldirektør i et meget stort konsern ønsket å studere sykefraværet blant de ansatte det siste året. Han valgte tilfeldig 60 ansatte. Disse hadde tilsammen 918 sykedager, og 40 av de 60 hadde vært syke i minst 18 dager. Videre ble standardavviket for antall sykedager, X, for en tilfeldig ansatt estimert til s = 3,8 dager i den samme undersøkelsen. Vi antar at X er normalfordelt.

Bestem ut fra opplysningene et 95 % konfidensintervall (intervallet i b) får konfidensgrad tilnærmet 95 %) for disse to størrelsene (i a) og b)):

a) Det forventede antall sykedager for en tilfeldig ansatt.

b) Andelen av ansatte i konsernet som hadde vært syke i minst 18 dager.

Personaldirektøren ønsket å gjøre en grundigere tilsvarende undersøkelse hvor han fortsatt var minst 95 % sikker på konklusjonen, men hvor konfidensintervallet var smalere.

c) Hvor mange personer måtte han velge dersom konfidensintervallet for det forventede antall sykedager skulle ha en lengde på høyst 1 dag? (Han antar nå at std(X) = 4).

d) Hvor mange personer må velges dersom konfidensintervallet i b) ikke skal være bredere enn 0,1, uansett andelen av ansatte som hadde vært syke i minst 18 dager?

<u>Løsningsforslag:</u>

a) $\overline{X}$ tn. $N(\mu, s/\sqrt{n}) = N(\mu\ , 3{,}8/\sqrt{60})$. $\bar{x} = 918/60 = 15{,}3$. [$X$] = dager. Antar at én ansatts sykefravær er uavhengig av andre ansattes sykefravær (trolig noe urealistisk, blant annet grunnet smittefare). 95 % KI for μ er da gitt ved

$$\bar{x}\mp z_{,025}\cdot s/\sqrt{n} = 15{,}3\mp 1{,}96\cdot 3{,}8/\sqrt{60} = \underline{(14{,}3\ ,\ 16{,}3)}$$

b) La Y betegne antall syke av n = 60 i minst 18 dager. Under forutsetningen om uavhengighet er da Y tn. Bino(n, p) med ukjent andel p og $p^* = X/n$. Fordi

$np^*(1-p^*) = 60\cdot(1-40/60) >> 5$ så vil p^* være tn. $N(p, \sqrt{p(1-p)/n})$. Et tn. 95 % KI for p blir da

$$p^* \mp z_{,025}\sqrt{p^*(1-p^*)/n} = 2/3 \mp 1{,}96\cdot\sqrt{(\tfrac{2}{3}\cdot\tfrac{1}{3})/1600} = \underline{(54{,}7\ ,\ 78{,}6)\ \%}$$

c) $2\cdot z_{,025}\cdot \mathrm{std}(X)/\sqrt{n} = 1 \Rightarrow n = (2\cdot 1{,}96\cdot 4)^2 = \underline{246}$

d) For å være på den sikre siden benyttes den verdi for p som gir bredest intervall, dvs. $p = ½$ som gir $\mathrm{std}(p^*) = \sqrt{\left(\tfrac{1}{2}\cdot\tfrac{1}{2}\right)/n} = 1/(2\sqrt{n})$. Vi finner da

$$2\cdot z_{,025}\cdot 1/(2\sqrt{n}) = ,1 \Rightarrow n = (2\cdot 1{,}96/(2\cdot,1))^2 = 384{,}1 \text{ som forhøyes til } \underline{n = 385}.$$

Kapittel 7
Hypotesetesting

Oppgave 7.1 Gitt en normalfordeling med varians $\sigma^2 = 400$. Vi tar et tilfeldig utvalg bestående av 36 elementer, og beregner aritmetiske gjennomsnitt til $\bar{x} = 159$. Test hypotesen H_0: $\mu = 150$ mot alternativet H_1: $\mu > 150$ på 10 % nivået.

Løsningsforslag:

Følger 4-trinns skjemaet:

1) Sannsynlighetsmodell: $\bar{X} \sim N(\mu, 20/\sqrt{n}) = N(\mu, 20/\sqrt{36}) = N(\mu, 10/3)$

2) Tester: H_0: $\mu = 150$, H_1: $\mu > 150$

3) Testobservator: $T = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{\bar{X} - 150}{10} \cdot 3$

   Forkastingsområdet, R, følger ulikhetstegnet til H_1: R: $T > z_{,1} = 1{,}282$ der $z_{,1}$ er øvre 10 % fraktil i $N(0,1)$-fordelingen (se t-tabell med d.f. = ∞).

4) Gjennomfører testen: $T = \dfrac{159 - 150}{10} \cdot 3 = 2{,}7 > 1{,}282$

   Konklusjon: Det er på nivå 10 % grunnlag for å <u>forkaste</u> H_0, dvs. konkludere at μ er større enn 150.

Oppgave 7.2 Av et større vareparti kan det være feil med emballasjen på den enkelte artikkel. Vi antar at denne situasjonen kan beskrives ved hjelp av binomialfordelingen. Et tilfeldig utvalg bestående av 78 artikler gav at 35 hadde emballasjefeil. Test:

H_0: $p = 0{,}55$ mot H_1: $p < 0{,}55$, der p er sannsynligheten for emballasjefeil.

Løsningsforslag:

1) X = antall av tilfeldig utvalg på $n = 78$ som har emballasjefeil, $X \sim$ Bino(n, p). Velger $p^* = X/n$ som estimator for p. Med basis i resultatet $x = 35$ er $np^*(1 - p^*) >> 5$ og det er rimelig å anta at X er tn. $N(np, \sqrt{np(1-p)})$.

2) H_0: $p = p_0 = 0{,}55$ mot H_1: $p < 0{,}55$

3) $T = \dfrac{X - np_0 - {,}5}{\sqrt{np_0(1-p_0)}} = \dfrac{X - 42{,}4}{4{,}39}$

   R: $T < -z_{,05} = -1{,}645$ der $z_{,05}$ er øvre 5 % fraktil i $N(0,1)$-fordelingen.

4) Gjennomfører testen: $T = \dfrac{35-42{,}4}{4{,}39} = -1{,}69 < -1{,}645$

   Konklusjon: Forkaster H_0 på nivå 5 % og konkluderer at $p < ,55$.

Oppgave 7.3 Se oppgave 8.9.

Oppgave 7.4 *S* er Poisson-fordelt med parameter λ. Man skal teste nullhypotesen H_0: $\lambda = 5$ mot alternativet H_1: $\lambda > 5$.

a) Konstruer den testen som har signifikansnivå nærmest 0,06

b) Anta at $\lambda = 8$. Hva er sannsynligheten for feil av type II (ß) ved den funne test under punkt a?

c) Finn sannsynligheten for å forkaste nullhypotesen ved den testen du fant under punkt a), når den sanne verdi av λ er

   i) 6, ii) 7, iii) 8, iv) 9

   Skisser testens styrkekurve.

Løsningsforslag:

a) Testobservator: $T = S$, Forkastingsområde, R: $T > k$ der k skal bestemmes slik at

   $P(T > k \mid \lambda = 5) = 1 - P(T \le k \mid \lambda = 5) \approx ,06 \Leftrightarrow P(T \le k \mid \lambda = 5) \approx ,94$

   Bestemmer k fra Poisson-tabell over Po(5)-fordelingen:

   $P(T \le 8 \mid \lambda = 5) = ,932, \;\; P(T \le 9 \mid \lambda = 5) = ,968$

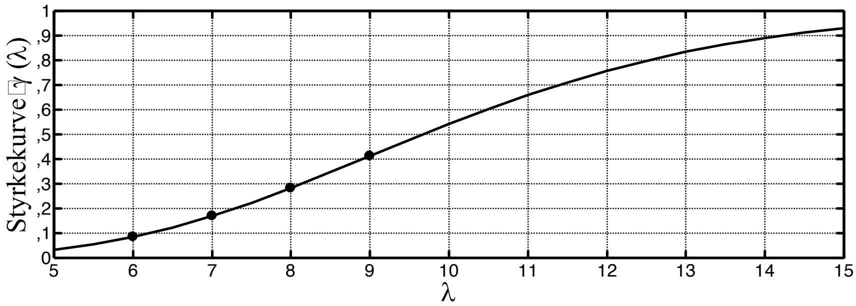
   Vi er konservative (velger den verdi som gir minst signifikansnvå) og velger $k = 9$. Testen blir da: $\underline{T = S, \;\; R: T > 9}$

b) $\beta(8) = P(R^C \mid \lambda = 8) = 1 - P(R \mid \lambda = 8) = 1 - P(T > 9 \mid \lambda = 8) = P(T \le 9 \mid \lambda = 8)$. Fra Poisson-tabell med $\lambda = 8$ finner vi at $\underline{\beta(8) = ,717}$

c) $\gamma(\lambda) = P(R \mid \lambda) = P(T > 9 \mid \lambda) = 1 - P(T \le 9 \mid \lambda)$. Fra Poisson-tabell finner vi

   $\lambda = 6, 7, 8, 9 \Rightarrow P(T \le 9 \mid \lambda) = ,916, ,831, ,717, ,587$. Følgelig:

   $\underline{\gamma(6) = ,084, \gamma(7) = ,170, \gamma(8) = ,283, \gamma(9) = ,413}$. Styrkekurven er tegnet i neste figur.



Bemerk at testen virker ganske svak, λ må være større enn 12 for at vi skal forkaste H_0 med sannsynlighet over 80 %.

Oppgave 7.5 En kjøttprodusent reklamerer med at de pølsevarer de selger inneholder høyst 12 % fett. Det er mistanke om at det faktiske fettinnholdet er høyere, og kjøttkontrollen foretar derfor en test av produsentens pølseprodukter. Sett

X: Fettinnholdet i en tilfeldig produksjon og anta at X er $N(\mu,2)$

Vi forutsetter at erfaring har vist at fettinnholdet har et standardavvik på 2 prosentenheter. En undersøkelse av ti produksjoner gav følgende data for fettinnholdet i prosent:

13 11 15 14 17 16 11 13 12 14
$\bar{x} = 13{,}6$

a) Test om fettinnholdet er 12 % mot at det er høyere. Velg signifikansnivå 0,05. Uttrykk i ord hva feil av type I i dette tilfellet innebærer.

b) Anta det er riktig at produsenten fusker med fettinnholdet og at forventet fettinnhold er 14 %. Hva er sannsynligheten for feil av type II? Hva er testens styrke?

c) Vi ønsker å utføre testen med et signifikansnivå på 0,01. Hva blir nå svarene på punkt b)? Sammenlign og kommenter forskjellene.

d) Kjøttkontrollens statistiker synes at svaret i c) gir en for høy sannsynlighet for feil av type II. Han ønsker å øke stikkprøvens størrelse n, slik at sannsynligheten for feil av type II høyst er 0,10, men under forutsetning av at signifikansnivået fremdeles er 0,01. Hvor stor må n være for at disse krav skal være oppfylt?

Løsningsforslag:

a) [X] = %. Sannsynlighetsmodell: $\overline{X} \sim N(\mu, 2/\sqrt{10})$. H_0: $\mu = 12$, H_1: $\mu > 12$.

$T = \dfrac{\overline{X} - 12}{2/\sqrt{10}}$, R: $T > z_{,05} = 1{,}645$. Setter inn $\bar{x} = 13{,}6$ og får:

$T = \dfrac{13{,}6 - 12}{2/\sqrt{10}} = 2{,}53 > 1{,}645$.

Konklusjon: Vi konkluderer på nivå 5 % at fettinnholdet er $\underline{\text{høyere enn 12 \%}}$.

Type I feil vil si feilaktig å konkludere at fettinnholdet er høyere enn 12 %.

b) $P(\text{type II feil}) = \beta(\mu) = P(R^C \mid \mu) = P(T < z_{,05} \mid \mu) =$

$P\left(\frac{\bar{X}-12}{2/\sqrt{10}} < z_{,05} \mid \mu\right) = P\left(\frac{\bar{X}-\mu+\mu-12}{2/\sqrt{10}} < z_{,05} \mid \mu\right) = P\left(\frac{\bar{X}-\mu}{2/\sqrt{10}} + \frac{\mu-12}{\sqrt{2/10}} < z_{,05} \mid \mu\right)$

Uansett verdi for m så er $Z = (\bar{X} - \mu)/(2/\sqrt{10})$ $N(0, 1)$-fordelt. Vi får derfor:

$\beta(\mu) = \Phi\left(z_{,05} - \frac{\mu-12}{2/\sqrt{10}}\right)$ og $\underline{\beta(14)} = \Phi\left(1{,}645 - \frac{14-12}{2/\sqrt{10}}\right) = \underline{{,}065}$

Styrkefunksjonen er $\gamma(\mu) = P(R \mid \mu) = 1 - \beta(\mu) = \Phi\left(-(z_{,05} - \frac{\mu-12}{2/\sqrt{10}})\right)$, og

$\underline{\gamma(14)} = 1 - \beta(14) = \underline{{,}935}$

c) Vi erstatter nå $z_{,05} = 1{,}645$ med $z_{,01} = 2{,}326$ og får:

$\underline{\beta(14)} = \Phi\left(2{,}326 - \frac{14-12}{2/\sqrt{10}}\right) = \underline{{,}202}$, $\underline{\gamma(14)} = 1 - \beta(14) = \underline{{,}798}$

Effekten av en strengere test er følgelig at den er blitt svakere og at vi har større sannsynlighet for feilaktig å unnlate å konkludere med et høyere fettinnhold enn 12 %.

d) Vi skal nå oppnå at $\beta(14) \leq {,}10$ ved å øke antall observasjoner, n. Dette gjør vi ved å erstatte 10 i uttrykket for $\beta(14)$ fra c) med n:

$\beta(14) = \Phi\left(2{,}326 - \frac{14-12}{2/\sqrt{n}}\right) \leq {,}10 \Rightarrow 2{,}326 - \sqrt{n} \leq -z_{,10} = -1{,}282$

dvs. $n > (2{,}326 + 1{,}282)^2 = 13{,}02$, som forhøyes til $\underline{n = 14}$

Oppgave 7.6 Støygrensen for tunge lastebiler er satt til 83 desibel. La μ være midlere støynivå for alle tunge lastebiler. Et tilfeldig utvalg på 6 tunge lastebiler ble observert med følgende støynivåer:

85,4 86,8 86,1 85,3 84,8 86,0

Det antas at desibelnivåene for de forskjellige lastebilene er uavhengige, normalfordelte med forventning μ og ukjent varians σ^2.

a) Gir resultatene grunnlag for å påstå at $\mu > 83$? Formuler nullhypotese og alternativhypotese for å besvare dette spørsmålet. Bruk en test med signifikansnivå $\alpha = 0{,}05$.

b) Konstruer et 90 % konfidensintervall for μ.

<u>Løsningsforslag</u>:

$\bar{X} \sim N(\mu,\sigma)$ med ukjent μ og σ. $[X]$ = dB (desibel).

a) H_0: $\mu = 83$, H_1: $\mu > 83$. Dataene gir $\bar{x}$ =85,73 og s = ,709.

t-test: $T = \frac{\bar{x} - 83}{s/\sqrt{n}} = \frac{85{,}73 - 83}{{,}709/\sqrt{6}} = 3{,}35 > t_{,05} = 2{,}015$ $(6 - 1 = 5$ frihetsgrader$)$

Konklusjon: Dataene gir på nivå 5 % <u>grunnlag for å påstå at μ > 83</u>.

b) 90 % KI for μ: $\bar{x} \mp t_{,05} \cdot s/\sqrt{n} = 85{,}73 \mp 2{,}015 \cdot {,}709/\sqrt{6} = \underline{(85{,}2\ ,\ 86{,}3)}$

Oppgave 7.7 Antall fellingstillatelser som blir gitt for elg hver høst i Nordskogen avhenger av størrelsen, N, på elgpopulasjonen i skogen. Ett år frykter man at populasjonen er så liten at den ikke tåler beskatning, og ingen fellingstillatelser vil bli gitt hvis $N \leq 12$.

For å vurdere populasjonsstørrelsen har en på et tidligere tidspunkt merket 6 elger. Nå vil man fange 6 elger på nytt og telle hvor mange av dem, X, som er merket. Merk at ingen elger slippes ut igjen før antall merkede er talt opp.

a) Nullhypotesen: H_0: $N \leq 12$ skal testes mot H_1: $N > 12$. Som testobservator skal vi bruke X. Skal vi forkaste for store eller små verdier av testobservatoren? Svaret skal begrunnes.

b) Hva slags sannsynlighetsfordeling har testobservatoren?

c) Bestem forkastingsområdet for en test med signifikansnivå ca. 4 %.

<u>Løsningsforslag</u>:

a) H_0 forkastes for små verdier av X, liten «gjenfangst» tyder på stor bestand.

b) $T = X$ er hypergeometrisk fordelt (trekning uten tilbakelegging) hyp(n,D,N) = hyp(6,6,N) med ukjent N:

$$f(x) = \frac{\binom{6}{x} \cdot \binom{N-6}{6-x}}{\binom{N}{6}}, \quad x = 0,1,\ldots,6, \qquad F(k) = P(X \leq k) = \sum_{x=0}^{k} f(x)$$

c) $R\colon T \leq k,\ P(T \leq k \mid N = 12) \leq ,04$. Setter inn for $x = 0,1,\ldots$ i uttrykket for $f(x)$ fra b) med $N = 12$ og summerer inntil $F(x) < ,04$: $f(0) = 1/924$, $f(1) = 36/924$ som tilsvarer $F(0) = 1/924 > ,04$, $F(1) = 37/924 = ,04$. Følgelig: $\underline{R\colon T \leq 2}$

Oppgave 7.8 En presidentkandidat i USA påstår at han har støtte fra 60 % av velgerne. Anta at kandidatens påstand er korrekt.

a) Et tilfeldig utvalg på 10 velgere tas. Hva er sannsynligheten for at minst 5 av disse støtter kandidaten?

b) Anta at 100 velgere trekkes tilfeldig. Hva er sannsynligheten for at minst 50 av disse støtter kandidaten?

c) For å fremlegge «statistisk bevis» for sin påstand, ber kandidaten et meningsmålingsinstitutt foreta en undersøkelse med et tilfeldig utvalg på 900 velgere. Det viser seg at av disse støtter 567 kandidaten. Er dette tilstrekkelig til å si at mer enn 60 % av velgerne støtter kandidaten? Bruk signifikansnivå 5 %.

Finn signifikanssannsynligheten P^*, og forklar hva den betyr.

<u>Løsningsforslag</u>:

La p betegne andel velgere av totalt N som støtter presidenten, og la $p^* = X/n$ betegne en estimator for p der X er antall som svarer de støtter presidenten i et tilfeldig utvalg på n. Det er rimelig å anta at $n << N$ og vi antar at X tn. Bino(n,p).

a) $n = 10, p = ,6.$ $P(X \geq 5 \mid p = ,6) = 1 - P(X \leq 4 \mid p = ,6) = 1 - 166 = \underline{,834}$

b) $n = 100, p = ,6, np(1-p) = 60 >> 5 \Rightarrow X$ tn. $N(np, \sqrt{np(1-p)})$,

$$P(X \geq 50) = 1 - P(X \leq 49) \approx \Phi\left(-\tfrac{49-60+,5}{\sqrt{60\cdot,4}}\right) = \Phi(2,14) = \underline{,984}$$

c) p^* tn. $N(np, \sqrt{np(1-p)/n})$. Sannsynligheten for resultatet $x = 537$ eller bedre er

$$P(X \geq 537 \mid p = ,6) = 1 - P(X \leq 567 \mid p = ,6) \approx 1 - \Phi\left(\tfrac{567-900\cdot,6+,5}{\sqrt{900\cdot,6\cdot,4}}\right) = \Phi(-1,87) = ,031$$

Følgelig: Signifikanssannsynligheten er $\underline{P^* = ,031}$ < ,05, følgelig er det på nivå 5 % grunnlag i data til å konkludere at mer enn 60 % av velgerne støtter presidenten.

Oppgave 7.9 En fabrikk produserer skruer der skruediametrene X (mm) antas å være normalfordelt med forventning μ og standardavvik σ. Skruene skal normalt være 2 mm i diameter, og et standardavvik på høyst 0,04 mm er akseptabelt. Gjennom en viss tid har en observert at skruediameteren er større enn 2 mm, og en ønsker å teste påstanden at $\mu > 2$. Det tas et tilfeldig utvalg på 9 skruer, der disse har følgende diametre (i mm):

2,00 2,08 2,10 2,12 2,15 2,15 2,10 2,05 2,15

a) Regn ut gjennomsnittet $\bar{x}$ og empirisk standardavvik (n−1 i nevner).

b) Formuler nullhypotese og alternativ- hypotese, og utfør testen med α = 5 % nivå. Testkonklusjonen formuleres med ord.

c) Forklar hva α betyr.

d) Hva er forskjellen på sannsynlighetsfordelingen til diameteren på en tilfeldig valgt skrue, og sannsynlighetsfordelingen til gjennomsnittet av 9 tilfeldig valgte skruer?

<u>Løsningsforslag</u>:

a) $n = 9, \Sigma x = 18{,}9, \Sigma x^2 = 39{,}7108, [x] = \text{mm}.$

$$\bar{x} = 18{,}9/9 = \underline{2{,}10}, s = \sqrt{(\Sigma x^2 - (\Sigma x)^2/n)/(n-1)} = \underline{,051}$$

b) H_0: $\mu = 2$, H_1: $\mu > 2$. $T = \dfrac{\bar{X} - 2}{S/\sqrt{n}}$, R: $T > t_{,05} = 1{,}86$ $(9 - 1 = 8$ frihetsgrader$)$.

Gjennomfører testen: $T = \dfrac{2{,}10 - 2}{,051/\sqrt{9}} = 5{,}88 > 1{,}86$

H_0 <u>forkastes</u>, dvs. det er på nivå 5 % grunnlag i data til å konkludere at $\mu > 2$.

c) Testnivået α = 5 % betyr at det skal være maksimalt 5 % sannsynlighet for feilaktig å konkludere at $\mu > 2$.

d) X og $\overline{X}$ er begge normalfordelte med forventning μ, men $\overline{X}$ har mindre standard-avvik: $\text{std}(X) = \sigma$, $\text{std}(\overline{X}) = \sigma / \sqrt{n} = \sigma / \sqrt{9} = \sigma / 3$.

Oppgave 7.10 Vi kaster en mynt 50 ganger. I 18 av kastene ble det krone og i 32 mynt. Hvis vi bruker et 5 % signifikansnivå, er det da noen grunn til å tvile på at mynten er i orden? (Formuler H_0 og H_1 og utfør en passende test).

Løsningsforslag:

La X være antall kast som gir kron av n = 50 tilfeldige kast. X er da binomisk fordelt Bino(n,p) med n = 50 og med p = 0,5 hvis mynten er iorden. Dersom dette er tilfellet er også np(1-p) = 25 > 5, slik at X er tn. $N(np, \sqrt{np(1-p)}) = N(25, 5/\sqrt{2})$.

H_0: p = 0,5, H_1: p ≠ 0,5 (tosidig test fordi det ikke er angitt noen grunn til å tro at p < 0,5 før forsøket er utført).

$$T = \frac{x-25}{5/\sqrt{2}} = \frac{18-25}{5/\sqrt{2}} = -1{,}98,\quad R{:}\,|T| > z_{,025} = 1{,}96,\quad |T| = 1{,}98 > 1{,}96$$

Konklusjon: Det er grunnlag i data på nivå 5 % å konkludere at mynten ikke er iorden.

Oppgave 7.11 En boks inneholder fire kuler. En kule kan være sort eller hvit. Boksen inneholder en av disse 5 kombinasjonene: Fire sorte, fire hvite, en sort og tre hvite, en hvit og tre sorte eller to sorte og to hvite. Du skal teste hypotesen at det er to av hver farge. Du velger ut to uten tilbakelegging og konkluderer med at det ikke er to av hver farge dersom begge de utvalgte er av samme farge.

Hva er i denne situasjonen:

a) Feil av type I?

b) Feil av type II?

c) Hva er sannsynligheten for feil av type I?

d) Hva er sannsynligheten for feil av type II hvis det er 3 sorte kuler og en hvit kule?

Løsningsforslag:

a) Type I feil: Feilaktig å konkludere at det er to av hver farge.

b) Type II feil: Feilaktig å konkludere at det ikke er to av hver farge.

Lar H og S betegne henoholdsvis hvit og svart, og lar subskript 1 og 2 betegne henholdsvis 1. og 2. trekning. Videre lar vi P_0 og P_1 betegne sannsynligheter betinget med hensyn på at henholdsvis H_0 og H_1 er sann.

c) $P(\text{Type I feil}) = P(R \mid H_0) = P_0(R) = P_0(2S) + P_0(2H)$, $\quad H_0: \quad 2S + 2H$
$P_0(2S) = P_0(S_1)\cdot P_0(S_2 \mid S_1) = (2/4)\cdot(1/3) = 1/6 = P_0(2H)$ p.g.a. symmetri.
Følgelig: $P(\text{Type I feil}) = \underline{1/3}$

d) $P(\text{Type II feil}) = P(R^C \mid H_1) = P_1(H_1S_2) + P_1(S_1H_2)$, $\quad H_1: \quad 3S + 1H$
$P_1(H_1S_2) = P_1(H_1)\cdot P_1(S_2 \mid H_1) = (1/4)\cdot(3/3) = 1/4$
$P_1(S_1H_2) = P_1(S_1)\cdot P_1(H_2 \mid S_1) = (3/4)\cdot(1/3) = 1/4$
Følgelig: $P(\text{Type II feil} \mid 3S + 1H) = 1/4 + 1/4 = \underline{1/2}$

Oppgave 7.12 En bedrift produserer fiskesnører. Kvaliteten på snørene måles ved bruddstyrken, som er den maksimale strekkraften snøret tåler uten å briste. Bedriften har en gjennomprøvet produksjonsprosess, hvor bruddstyrken er normalfordelt med forventning 20,00 kg og standardavvik 0,90 kg. Bedriften skal utprøve en ny og kostnadsbesparende produksjonsprosess. Den nye prosessen kan ikke produsere snører med en større forventet bruddstyrke enn 20,00 kg (som den burde være), men bedriften har mistanke om at forventningsverdien er lavere enn dette. Av produksjonstekniske grunner kan bedriften anta at bruddstyrken ved den nye produksjonsprosessen også er normalfordelt med standardavvik 0,90 kg. La μ betegne forventet bruddstyrke ved den nye produksjonsprosessen. Bedriften vil teste

H_0: $\mu = 20{,}00$ mot H_1: $\mu < 20{,}00$

a) Det innhentes en stikkprøve ved å måle bruddstyrken til 20 snører fra den nye produksjonsprosessen. Gjennomsnittlig bruddstyrke er 19,50. Utfør hypotesesprøvingen ovenfor på 5 % nivå. Hva er testens konklusjon? Finn signifikanssannsynligheten.

Anta at forventningen til bruddstyrken med den nye produksjonsprosessen er 19,50.

b) Beregn sannsynligheten ß for feil av type II.

c) Bedriften mener at sannsynligheten for feil av type II i b) er for høy. Den kan redusere sannsynligheten for feil av type II ved å øke signifikansnivået α. Hva blir α dersom testen tilpasses slik at ß = 0,10?

d) Illustrer sammenhengen mellom punktene a), b) og c) i en figur.

e) Dersom bedriften ønsker å redusere sannsynligheten for feil av type II uten å gå på bekostning av signifikansnivået, må den øke stikkprøvens størrelse. Hvor stor må stikkprøven være dersom sannsynligheten for feil av type II skal være høyst 0,10 og signifikansnivået høyst 0,05?

<u>Løsningsforslag</u>: Sannsynlighetsmodell: $\bar{X} \sim N(\mu, {,}90/\sqrt{n})$, $[X] = \text{kg}$

a) $T = \dfrac{\bar{X} - 20}{{,}90/\sqrt{n}}$, R: $T < -z_{,05} = -1{,}645$. $\bar{x} = 19{,}5 \Rightarrow T = \dfrac{19{,}5 - 20}{{,}90/\sqrt{20}} = -2{,}48 < -1{,}645$

Konklusjon: Grunnlag i data på nivå 5 % å konkludere at forventet bruddstyrke er mindre enn 20. Signifikanssannsynligheten, P^*, er sannsynligheten for å oppnå vårt eksperimentelle resultat eller mer ekstreme verdier (enda mindre middelverdi) når H_0 er sann. Siden T er $N(0,1)$-fordelt under H_0 får vi derfor:

$P^* = P(T < -2{,}48) = \Phi(-2{,}48) = \underline{0{,}0066}$ som sterkt underbygger konklusjonen at $\mu < 20$.

b) $P(\text{Type II feil}) = \beta(\mu) = P(R^C \mid \mu) = 1 - P(T < -z_\alpha \mid \mu) = P(T > -z_\alpha \mid \mu)$

$P\left(\frac{\bar{X}-20}{,9/\sqrt{n}} > -z_\alpha \mid \mu\right) = P\left(\frac{\bar{X}-\mu+\mu-20}{,9/\sqrt{n}} > -z_\alpha \mid \mu\right) = P\left(\frac{\bar{X}-\mu}{,9/\sqrt{n}} + \frac{\mu-20}{,9/\sqrt{n}} > -z_\alpha \mid \mu\right)$

Fordi $Z = (\bar{X} - \mu)/(,9/\sqrt{n})$ er $N(0,1)$ får vi derfor: $\beta(\mu) = \Phi(z_\alpha + \frac{\mu-20}{,9/\sqrt{n}})$

Innsatt $z_a = 1{,}96$, $\mu = 19{,}5$ og $n = 20$ får vi:

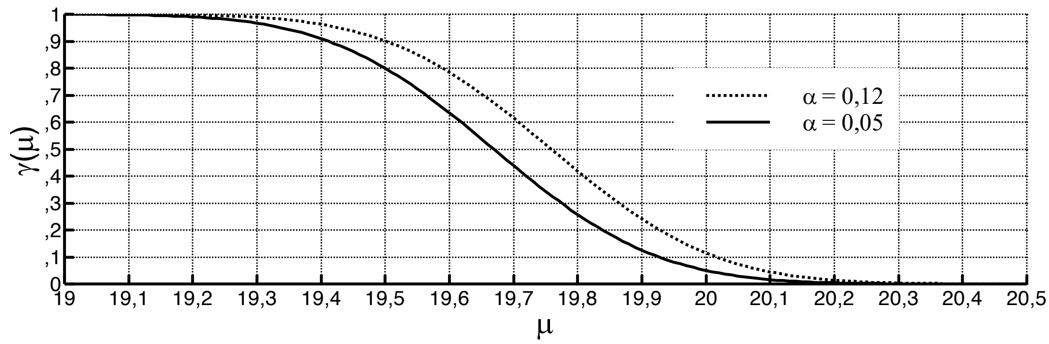
$\beta(19{,}5) = \Phi(1{,}645 + \frac{19{,}5-20}{,9/\sqrt{20}}) = \Phi(-{,}84) = \underline{{,}20}$

c) $\beta(19{,}5) = {,}1 \Rightarrow \Phi\left(z_\alpha + \frac{19{,}5-20}{,9/\sqrt{20}}\right) = {,}1 \Rightarrow z_\alpha + \frac{19{,}5-20}{,9/\sqrt{20}} = -z_{,1} \Rightarrow z_\alpha = -z_{,1} + 2{,}48 = 1{,}20$

$\alpha = P(Z > 1{,}2) = \Phi(-1{,}2) = \underline{{,}12}$, dvs. større sannsynlighet for å begå Type I feil, vi får ikke i pose og sekk uten å øke antall observasjoner.

d) Sammenhengen mellom a), b) og c) kan illustreres ved å tegne styrkekurvene $\gamma(\mu)$ for testene med testnivå 5 % og 12 %. Styrkefunksjonen $\gamma(\mu)$ er gitt som følger:

$\gamma(\mu) = P(R \mid \mu) = 1 - \beta(\mu) = \Phi\left(-\left(z_\alpha + \frac{20-\mu}{,9/\sqrt{20}}\right)\right)$



e) $\beta(19{,}5) = \Phi\left(z_{,05} + \frac{19{,}5-20}{,9\sqrt{n}}\right) \leq {,}10 \Rightarrow 1{,}645 + \frac{19{,}5-20}{,9\sqrt{n}} < -z_{,1} \Rightarrow$

$n > ((1{,}96 + 1{,}28) \cdot \frac{,9}{,5})^2 = 27{,}72$, dvs. $\underline{n = 28}$

Oppgave 7.13 En foreleser ved BI vil undersøke om 10 minutters-pausen mellom to forelesningstimer overholdes. Tiden på 15 tilfeldig utvalgte pauser var:

14, 14, 8, 12, 13, 10, 14, 10, 11, 13, 9, 10, 15, 12, 15.

a) Test H_0: $\mu = 10$ mot H_1: $\mu \neq 10$ på 5 %-nivået.

b) Kan du trekke noen konklusjoner av dette? I tilfelle hvilke?

Løsningsforslag:

a) Sannsynlighetsmodell: $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$ med ukjent μ og σ og med $n = 15$.

$T = \dfrac{\bar{X} - 10}{S/\sqrt{n}}$, $R\colon |T| > t_{n-1}$ der t_{n-1} er øvre 2,5 % fraktil i stud(n−1)-fordelingen.

Datamaterialet gir $\bar{x} = 12$, $s = 2{,}236$. Innsatt i uttrykket for T får vi:

$$T = \frac{12 - 10}{2{,}236/\sqrt{15}} = 3{,}46,\ |T| = 3{,}46 > t_{14} = 2{,}14$$

Konklusjon: Grunnlag til å konkludere på nivå 5 % at 10 minutters-pausen ikke overholdes.

b) Mulige konklusjoner: Travle forelesere som trenger mer enn 10 min., dårlig disiplin, «snille» lærere som venter til sene studenter er på plass.

Oppgave 7.14 Ifølge statistisk årbok i 1982, var gjennomsnittshøyden for de i alt 34 027 vernepliktige i 1982 179,4 cm, mens det i 1981 var 33 337 vernepliktige med en gjennomsnittshøyde på 179,5 cm. Begge år var standardavviket 6,6 cm. Det har i moderne tid aldri skjedd en nedgang i gjennomsnittlig rekrutthøyde. Undersøk om nedgangen denne gang er signifikant på 5 %-nivået. Diskuter mulige feilkilder.

Løsningsforslag:

La X betegne høyde (i cm) til tilfeldig rekrutt. Antar $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$ med $\sigma = 6{,}6$ og n = antall rekrutter. La subskript 1 og 2 betegne henholdsvis 1981 og 1982.

$H_0\colon \mu_2 - \mu_1 = 0$, $H_1\colon \mu_2 - \mu_1 \neq 0$, $T = \dfrac{\bar{X}_2 - \bar{X}_1}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, $R\colon |T| > z_{,025} = 1{,}96$

Gjennomfører testen: $T = \dfrac{179{,}4 - 179{,}5}{6{,}6\sqrt{\frac{1}{33337} + \frac{1}{34027}}} = -1{,}97$, $|T| = 1{,}97 > 1{,}96$.

Konklusjon: Det er grunnlag i data på nivå 5 % å hevde at gjennomsnittshøyden i 1982 er signifikant forskjellig fra gjennomsnittshøyden i 1981, men konklusjonen er ikke veldig sterk (såvidt forkasting). Testen er tosidig fordi man ikke på forhånd har forventet en nedgang, men tatt utgangspunkt i data i etterhånd. En mulig feilkilde kan være at gjennomsnittshøyden er oppgitt med for få siffer (179,43 og 179,47 ville f.eks. ikke gitt forkasting).

Oppgave 7.15 Fra legehold er det hevdet at 10 % av alle diagnoser er feilstilte. Anta det blir iverksatt tiltak rettet mot å forbedre diagnosestillingen. Etter at tiltakene er gjennomført, blir 200 diagnoser tilfeldig valgt ut, og det blir undersøkt om de var gale eller riktige. Resultatene viser at 13 av de 200 diagnosene var feilstilte. Test om det har vært en signifikant nedgang i antall feilstilte diagnoser. Velg signifikansnivå 0,05.

Løsningsforslag: La X betegne antall feilstilte diagnoser i tilfeldig utvalg på $n = 200$ diagnoser. Vi antar X er tn. Bino(n,p) med ukjent p med estimator $p^* = X/n$. Fordi $13\cdot(1-13/200) >> 5$ antar vi at X er tn. $N(np, \sqrt{np(1-p)})$.

$$H_0: p = ,1,\ \ H_1: \text{p} < ,1,\ \ T = \frac{X - n\cdot,1}{\sqrt{n\cdot,1\cdot,9}},\ \ R: T < -z_{,05} = -1{,}645$$

Gjennomfører testen: $T = \dfrac{13-20}{\sqrt{20\cdot,9}} = -1{,}650 > 1{,}645$

Konklusjon: Vi kan (såvidt) konkludere på nivå 5 % at det har vært en signifikant nedgang i antall feilstilte diagnoser.

Oppgave 7.16 (E) En bedrift fremstiller artikler med en kvalitet X som har forventning μ og standardavvik σ. Hvis maskinen er riktig innstilt, skal μ være lik 4,3. Ved en kontrollmåling av 9 tilfeldige artikler fant en følgende verdier på kvaliteten:

4,4 4,7 4,0 4,6 4,6 4,5 4,2 4,7 4,5

Anta at σ = 0,22 er kjent.

a) Lag et konfidensintervall med konfidensgrad 95 % for μ. Hvilke forutsetninger må du gjøre?

Fomannen på avdelingen har en tid hatt mistanke om at μ er større enn 4,3, og at maskinen dermed bør justeres.

b) Formuler mannens problemstilling som et hypoteseprøvingsproblem, og gjennomfør testen på grunnlag av kontrollmålingene. Bruk signifikansnivå 5 %.

Bestem testens styrkefunksjon, og beregn styrken når μ = 4,5.

Anta nå at vi ikke har tillit til at σ = 0,22 og anser σ som ukjent.

c) Test hypotesen i b) under denne forutsetningen.

d) Bestem til slutt på grunnlag av kontrollmålingene et 99 % konfidensintervall på formen [0,b] for σ (μ antas ukjent).

Løsningsforslag:

a) Antar $X_1,\ldots,X_n$ uavhengige, dvs. $\overline{X} \sim N(\mu, \sigma/\sqrt{n})$ med μ ukjent, $\sigma = ,22$ og $n = 9$. Et 95 % KI for μ er da gitt ved $\overline{X} \mp z_{,025}\cdot\sigma/\sqrt{n}$. Innsatt eksperimentelle verdier får vi: $4{,}467 \mp 1{,}96\cdot,22/\sqrt{9} = \underline{(4{,}31\,,\,4{,}62)}$

b) H_0: $\mu = 4{,}3$, H_1: $\mu > 4{,}3$. $T = \dfrac{\overline{X} - 4{,}3}{\sigma / \sqrt{n}}$, R: $T > z_{,05} = 1{,}645$. Gjennomfører testen:

$T = \dfrac{4{,}467 - 4{,}3}{{,}22 / \sqrt{9}} = 2{,}28 > 1{,}645$. Konklusjon: Det er grunnlag i data på nivå 5 % å konkludere at μ er signifikant større enn 4,3

Styrkefunksjon:

$\gamma(\mu) = P(R \mid \mu) = P(T > z_\alpha \mid \mu) = P\left(\frac{\overline{X} - \mu + \mu - 4{,}3}{\sigma/\sqrt{n}} > z_\alpha \mid \mu\right) = P(Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} > z_\alpha - \frac{\mu - 4{,}3}{\sigma/\sqrt{n}} \mid \mu)$

$= \underline{\Phi\left(-\left(z_\alpha - \frac{\mu - 4{,}3}{{,}22/\sqrt{n}}\right)\right)}$. Innsatt $\mu = 4{,}5$ får vi

$\gamma(4{,}5) = \Phi\left(-\left(1{,}645 - \frac{4{,}5 - 4{,}3}{{,}22/\sqrt{9}}\right)\right) = \Phi(1{,}08) = \underline{{,}86}$

c) σ ukjent og erstattes med empirisk standardavvik, S. Vi får da:

$T = \dfrac{\overline{X} - 4{,}3}{S / \sqrt{n}}$, R: $T > t_{,05} = 1{,}86$ = øvre 5 % fraktil i stud(8)-fordelingen.

Gjennomfører testen og får: $T = \dfrac{4{,}467 - 4{,}3}{{,}235 / \sqrt{9}} = 2{,}13 > 1{,}86$. Følgelig: Vi kan på nivå 5 % konkludere at μ er signifikant større enn 4,3.

d) $\dfrac{(n-1)S^2}{\sigma^2}$ ~ Kji2(n–1) med nedre 1 % fraktil $\chi^2_{,01} = 1{,}647$. Et 99 % KI for σ kan konstrueres som følger:

$P(\chi^2_{,01} < \frac{(n-1)S^2}{\sigma^2} < \infty) = P(0 < \frac{\sigma^2}{(n-1)S^2} < 1/\chi^2_{,01}) = P(0 < \sigma < \frac{\sqrt{(n-1)} \cdot S}{\chi_{,01}})$

Følgelig: $[0\,, \frac{\sqrt{n-1} \cdot S}{\chi_{,01}}] = [0, \frac{\sqrt{8} \cdot {,}235}{\sqrt{1{,}647}}] = \underline{[0\,, {,}52]}$ er et 99 % KI for σ.

Oppgave 7.17 (E) En metallurg skal bestemme smeltepunktet μ for en legering. Han vet av erfaring at gjentatte målinger av smeltepunktet kan oppfattes som uavhengige og normalfordelte stokastiske variabler med forventning μ og standardavvik $\sigma = 2{,}0$ °C.

Metallurgen utfører 8 målinger og får følgende resultater (i °C):

1468,5 1469,0 1471,0 1470,0 1469,5 1467,5 1469,0 1472,0

a) Bestem et 95 % konfidensintervall for μ basert på de 8 målingene.

b) Metallurgen synes at intervallets lengde er noe stor. Hvor mange målinger måtte han minst gjort for å få et intervall som var kortere enn 2,0 °C?

c) Metallurgen er interessert i om det er grunnlag for å hevde at $\mu > 1468{,}0$ °C. Formuler hans problem med en passende nullhypotese og alternativ.

Konstruer testen med 5 % signifikansnivå. Hva blir konklusjonen ut fra de målte verdiene?

d) Bestem styrkefunksjonen for testen i c) og lag en skisse av den. Finn styrken for $\mu = 1470{,}0\ ^{o}$C, og forklar med ord hva det betyr at styrken har denne verdien.

e) Forklar kort hvilke modifiseringer vi måtte gjøre i a) og c) dersom vi ikke hadde kjent σ, og gjennomfør testen i c) også i dette tilfelle.

Løsningsforslag: (benevning i ^{o}C).

a) 95 % KI for μ: $\bar{x} \mp z_{,025} \cdot \sigma / \sqrt{n} = 1469{,}56 \mp 1{,}96 \cdot 2 / \sqrt{8} = \underline{(1468{,}2\ , 1470{,}9)}$

b) Intervallbredde fra a): 1470,9 – 1468,2 = 2,700. Siden intervallbredden avtar omvendt proporsjonalt med $n^{1/2}$ må vi øke antall observasjoner med en faktor $(2{,}7/2)^2 = 1{,}82$, dvs. vi må foreta $n = 8 \cdot 1{,}82 = 14{,}56$, dvs. n = 15 målinger for å få en intervallbredde på høyst 2. Merk at resultatet er noe avhengig av hvor mange siffer vi har med i mellomregningene, ved bruk av dataverktøy med mange siffer ville svaret blitt $n = 16$.

c) H_0: $\mu = 1468{,}0$, H_1: $\mu > 1468{,}0$, $T = \dfrac{\bar{X} - 1468{,}0}{\sigma / \sqrt{n}}$, R: $T > z_{,05} = 1{,}645$.

Gjennomfører testen: $T = \dfrac{1469{,}56 - 1468{,}0}{2 / \sqrt{8}} = 2{,}21 > 1{,}645$

Konklusjon: Grunnlag i data på nivå 5 % å konkludere at $\mu > 1468{,}0$.

d) Styrkefunksjon: $\gamma(\mu) = P(R \mid \mu) = P\left(\frac{\bar{X} - \mu + \mu - 1468{,}0}{\sigma / \sqrt{n}} > z_\alpha \mid \mu\right) = \underline{\Phi\left(-\left(z_\alpha - \frac{\mu - 1468{,}0}{\sigma / \sqrt{n}}\right)\right)}$

$\gamma(1470{,}0) = \Phi\left(-\left(1{,}645 - \frac{1470{,}0 - 1468{,}0}{2 / \sqrt{8}}\right)\right) = \Phi(1{,}18) = \underline{0{,}88}$

hvilket betyr at det hele 88 % sannsynlighet for å «avsløre» (konkludere) at $\mu > 1468$ når $\mu = 1470$.

e) σ ukjent medfører at vi må erstatte σ med empirisk standardavvik, s, og erstatte normalfordelingsfraktilene med t-fraktiler basert på stud(7)-fordelingen. Vi ville da fått:

$T = \dfrac{\bar{x} - 1468{,}0}{s / \sqrt{n}} = \dfrac{1469{,}56 - 1468{,}0}{1{,}425 / \sqrt{8}} = 3{,}10 > t_{,05}(7) = 1{,}89$, forkast H_0 (som i c)).

Oppgave 7.18 (E) Ved en bedrift produseres en bestemt type artikler. Sannsynligheten p for at en tilfeldig artikkel er defekt ligger normalt på ca. 0,03. Dersom defektsannsynligheten øker til mer enn 0,05, må produksjonsutstyret justeres.

a) Ved en kontroll av produksjonsutstyret undersøkte man en dag 40 tilfeldig valgte artikler. La X være antall defekte i utvalget. Som estimator for p bruker vi $p^* = X/40$.

Vis at p^* er forventningsrett, og bestem variansen til p^*.

b) Kontrollen viste at 3 av de 40 artiklene var defekte. Estimer p og standardavviket for p.

c) Gir det observerte resultatet grunn til å hevde at $p > 0{,}05$, dvs. at produksjonsutstyret bør justeres? (Signifikansnivå 5 %).

Løsningsforslag:

a) Rimelig å anta at X er Bino(n,p) med $n = 40$ og ukjent p. Da er E$X = np$, og E$p^* = $ E$(X/n) = $ E$(X)/n = np/n = p$, dvs. p^* forventningsrett, $\underline{Ep^* = p}$
$\text{Var}(p^*) = \text{Var}(X/n) = (1/n^2)\cdot\text{Var}(X) = np(1-p)/n^2 = \underline{p(1-p)/n}$

b) $p^* = 3/40 = \underline{,075}$, $std^*(p^*) = \sqrt{p^*(1-p^*)/n} = \sqrt{(3/40)\cdot(1-3/40)/40} = \underline{,0416}$

c) H_0: $p = ,05$, H_1: $p > ,05$, $T = X$, R: $X > k$, $P(R \mid p = ,05) = P(X > k \mid p = ,05) \le ,05$.

$P^* = P(X > 2 \mid p = ,05) = 1 - P(X \le 2 \mid p = ,05) = \sum_{i=0}^{2}\binom{40}{i} \cdot ,05^i \cdot ,95^{40-i} =$

$1-(,129+,271+,278) = ,322$, dvs. ikke grunnlag på nivå 5 % til å hevde at $p > ,05$

Oppgave 7.19 (E) Ved en bensinstasjon er det ukentlige salget X av bensin, normalfordelt med forventning μ og standardavvik $\sigma = 500$ liter.

Anta at $\mu = 8000$ liter er kjent.

a) Finn sannsynligheten for at bensinstasjonen selger over 8500 liter en uke.

b) Hvor stort lager må stasjonen ha for at man skal være 95 % sikker på at lageret ikke skal tømmes i løpet av en uke?

c) Anta at salget hver uke er uavhengig av salget andre uker.

Finn sannsynligheten for at det totale salget i løpet av 4 uker er mindre enn 30 000 liter.

Innehaveren av bensinstasjonen har fått mistanke om at det ukentlige salget har sunket (dvs. at $\mu < 8000$ liter), og bestemmer seg for å undersøke dette nærmere. Han noterer bensinsalget i 9 uker og finner gjennomsnittet $\bar{x} = 7800$ liter.

d) Er det på grunnlag av dette resultatet grunn til å tro at $\mu < 8000$ liter? Formuler problemet som et hypotesetestingsproblem og gjennomfør testen med 5 % signifikansnivå (σ antas kjent lik 500 liter).

e) Beregn testens styrke dersom $\mu = 7500$ liter.

Løsningsforslag: ([x] = liter)

a) $P(X > 8500) = \Phi\left(-\left(\frac{8500-8000}{500}\right)\right) = \Phi(-1) = \underline{,1587}$

b) La k betegne nødvendig lager.

$P(X < k) = ,95 \Rightarrow \Phi\left(\frac{k-8000}{500}\right) = ,95 \Rightarrow \frac{k-8000}{500} = z_{,05} = 1{,}645 \Rightarrow \underline{k = 8825 \text{ liter}}$

c) $Y = X_1 + \cdots + X_4$ er $N(4\mu, \sqrt{4}\sigma) = N(32000{,}1000)$.
$P(Y < 30000) = \Phi\left(\frac{30000-32000}{1000}\right) = \Phi(-2) = \underline{{,}0228}$

d) H_0: $\mu = \mu_0 = 8000$, H_1: $\mu < 8000$, $T = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{9}}$, R: $T < -z_{,05} = -1{,}645$.

Gjennomfører testen: $T = \dfrac{7800 - 8000}{500/3} = -1{,}2$.

Konklusjon: Ikke grunnlag på nivå 5 % å konkludere at $\mu < 8000$

e) Styrkefunksjon: $\gamma(\mu) = P(R \mid \mu) = P(\frac{\overline{X}-\mu+\mu-9000}{1500} < -z_\alpha \mid \mu) = \Phi\left(-z_\alpha - \frac{\mu-9000}{1500}\right)$
Setter inn $\mu = 7500$: $\gamma(7500) = \Phi\left(-1{,}645 - \frac{7500-8000}{500/3}\right) = \Phi(1{,}36) = \underline{{,}91}$

Oppgave 7.20 (E) La X være vekten til en tilfeldig student ved tidligere BIH. X antas normalfordelt $N(\mu,\sigma)$ der $\mu = 73$ kg og $\sigma = 7$ kg.

a) Finn sannsynligheten for at en tilfeldig student veier mellom 60 og 75 kg.

Anta at vi har et tilfeldig utvalg på 50 studenter. La Y være antallet av disse som veier mellom 60 og 75 kg.

b) Hvilken sannsynlighetsfordeling får Y? Finn forventning og varians til Y. Finn tilnærmet $P(Y > 35)$.

c) En av heisene er beregnet å tåle vekten 800 kg. Hvor mange studenter kan ta heisen samtidig dersom vi krever at sannsynligheten for overbelastning skal bli mindre enn eller lik 0,05?

Heisselskapet syntes de fikk for mange reparasjoner på heisen på grunn av overbelastning, selv om studentene fulgte reglene for maksimalt antall personer i heisen. De mente at dette måtte skyldes at studentene veide mer enn forutsatt, dvs. at $\mu > 73$ kg. For å kontrollere dette ble 10 tilfeldige personer kontrollveid. Tilsammen veide de 785 kg.

d) Lag en test med 5 % signifikansnivå for å avgjøre om heisselskapet har rett.

e) Beregn testens styrke dersom $\mu = 80$ kg.

Løsningsforslag: ($[x]$ = kg).

a) $p = P(60 < X < 75) = \Phi\left(\frac{75-73}{7}\right) - \Phi\left(\frac{60-73}{7}\right) = \Phi({,}29) - \Phi(-1{,}86) = {,}614 - {,}031 = \underline{{,}583}$

b) Y er Bino(n, p) med $n = 50$ og $p = {,}583$ fra a).
$\mathrm{E}Y = np = 50 \cdot {,}583 = \underline{29{,}15}$, $\mathrm{Var}Y = np(1-p) = 29{,}15 \cdot (1 - {,}583) = \underline{12{,}16}$
$P(Y > 35) = 1 - P(Y \le 34) \approx 1 - \Phi\left(\frac{34-29{,}15+{,}5}{\sqrt{12{,}16}}\right) = \Phi(-1{,}53) = \underline{{,}063}$

c) $W = X_1 + \cdots + X_k \sim N(k\mu, \sqrt{k}\sigma)$.
$P(W \ge 800) \le {,}05 \Rightarrow \Phi\left(\frac{800-k\mu}{\sqrt{k}\cdot\sigma}\right) \ge {,}95 \Rightarrow \frac{800-k\mu}{\sqrt{k}\cdot\sigma} \ge z_{,05} = 1{,}645$

Erstatter vi ulikhetstegnet over med likhetstegn får vi følgende andregradsligning (i $k^{1/2}$) til bestemmelse av k:

$\mu k + z_{,05} \cdot \sigma \cdot \sqrt{k} - 800 = 0$

Ligningen over gir reell løsning k = 10,44, følgelig kan heisen ta 10 studenter samtidig.

d) $H_0 : \mu = 73,\ H_1 : \mu > 73, \mathrm{T} = \dfrac{\overline{X} - 73}{\sigma / \sqrt{10}},\ R\!: T > z_{,05} = 1{,}645$

Gjennomfører testen: $T = \dfrac{785/10 - 73}{7/\sqrt{10}} = 2{,}48 > 1{,}645$

Konklusjonen: Grunnlag på nivå 5 % å konkludere at $\mu > 73$.

e) Styrkefunksjon: $\gamma(\mu) = P(R \mid \mu) = P\left(\frac{\overline{X} - \mu + \mu - 73}{\sigma/\sqrt{n}} > z_\alpha \mid \mu\right) = \Phi\left(-\left(z_\alpha - \frac{\mu - 73}{\sigma/\sqrt{n}}\right)\right)$

$\gamma(80) = \Phi\left(-\left(1{,}645 - \frac{80-73}{7/\sqrt{10}}\right)\right) = \Phi(1{,}52) = \underline{0{,}935}$

Oppgave 7.21 (E) Resistansen i en motstandstråd er normalfordelt $N(\mu,\sigma)$ med μ = 23,4 Ω og σ = 0,25 Ω.

a) Hva er sannsynligheten for at en motstand har en resistans større enn 24,0 Ω?

b) Hva er sannsynligheten for at forskjellen i absoluttverdi mellom resistansen til to motstander er større enn 0,5 Ω?

Motstandene kommer i pakker på 100 stk. De skal brukes i enheter hvor resistansen må ligge mellom 23,0 Ω og 24,0 Ω. Bedriften har avtale med leverandøren om å returnere pakken dersom mer enn 5 stk. ikke holder mål.

c) Beregn tilnærmet sannsynligheten for at pakken må returneres.

I praksis tar bedriften en stikkprøve på 5 stk, dersom det er mer enn 1 motstand som ikke holder sendes pakken tilbake. Anta at det i en pakke er 7 motstander som ikke holder.

d) Bestem sannsynligheten for at bedriften sender pakken tilbake.

I en periode har bedriften måttet sende mange pakker tilbake. De mener at μ > 23,4 Ω. De tok derfor en stikkprøve på 10 motstander og fant følgende verdier for resistansen (målt i Ω):

23,2 24,7 23,8 24,2 24,8 23,2 25,1 23,9 23,4 23,6

e) Lag en test med 5 % signifikansnivå for å avgjøre om bedriften har rett.

Løsningsforslag: X = resistansen i tilfeldig motstandstråd, $[X]$ = Ω.

a) $P(X > 24{,}0) = \Phi\left(-\frac{24{,}0-23{,}4}{,25}\right) = \Phi(-2{,}4) = \underline{,0082}$

b) $P(|X_2 - X_1| > ,5) = 1 - P(-,5 < X_2 - X_1 < ,5)$. $X_2 - X_1$ er $N(0, \sqrt{2} \cdot ,25) \Rightarrow$

$1-P(-,5 < X_2 - X_1 <,5) = 1-\Phi\left(\frac{,5}{\sqrt{2}\cdot,25}\right)+\Phi\left(-\frac{,5}{\sqrt{2}\cdot,25}\right) \approx 2\cdot\Phi(-1{,}41) = \underline{,159}$

c) $p = 1-P(23{,}0 < X < 24{,}0) = 1-(\Phi\left(\frac{24-23{,}4}{,25}\right)-\Phi\left(\frac{23-23{,}4}{,25}\right)) =,063$

La Y betegne antall ubrukelig motstandståder, $Y \sim$ Bino(n, p) med n = 100 og p = ,063. Antar Y tn. $N(np, \sqrt{np(1-p)})$ og får:

$P(Y > 5) = 1-P(Y \le 5) = 1-\Phi\left(\frac{5-np+,5}{\sqrt{np(1-p)}}\right) = 1-\Phi\left(\frac{5-100\cdot,063+,5}{\sqrt{100\cdot,063\cdot,937}}\right) = \Phi(,329) = \underline{,629}$

d) La W betegne antall defekte av n = 5 som tas tilfeldig ut blant N = 100 med D = 7 defekte. W er da hyp(n, D, N)-fordelt og vi får:

$P(W > 1) = 1-P(W \le 1) = 1-\frac{\binom{7}{0}\cdot\binom{100-7}{5-0}}{\binom{100}{5}}-\frac{\binom{7}{1}\cdot\binom{100-7}{5-1}}{\binom{100}{5}} = 1-,690-,272 = \underline{,038}$

e) $H_0\!: \mu = 23{,}4,\ H_1\!: \mu > 23{,}4,\ T = \dfrac{\overline{X}-23{,}4}{S/\sqrt{n}},\ R\!: T > t_{,05}(n-1) = t_{,05}(9) = 1{,}83$

$\bar{x} = 23{,}99, s = 2{,}72 \Rightarrow T = \frac{23{,}99-23{,}4}{2{,}72/\sqrt{10}} = 2{,}72 > 1{,}83$. Grunnlag i data på nivå 5 % å konkludere at $\mu > 23{,}4$.

Oppgave 7.22 (E) Ved bruk av et bestemt medisinsk preparat A får gjennomgående halvparten av pasientene uheldige ettervirkninger, dvs. p = P(ettervirkninger) = 0,5.

a) Anta at 10 pasienter bruker medisinen. La X være antall pasienter som får uheldige ettervirkninger. Hvilken sannsynlighetsfordeling får X? Finn E(X) og Var(X). Finn sannsynligheten for at 3 eller færre får uheldige ettervirkninger.

b) En modifisert versjon B av preparatet skal utprøves. Det påstås at de uheldige ettervirkningene nå er redusert, dvs. p < 0,5. For å sjekke denne påstanden prøves B på n = 10 pasienter. Det viser seg at 3 av pasientene får ettervirkninger.

Kan vi på grunnlag av dette resultatet påstå at sannsynligheten for ettervirkninger er redusert?

Formuler spørsmålet som et hypotesetestingsproblem, og besvar ved hjelp av beregninger gjort i a) (5 % signifikansnivå).

Løsningsforslag:

a) Antar uavhengighet mellom pasientene. Da er X tn. Bino(*10*, 0,5).
$\mathrm{E}X = np = 10\cdot 0{,}5 = \underline{5},\ \mathrm{Var}X = np(1-p) = 5\cdot(1-0{,}5) = \underline{2{,}5}$

$P(X \le 3) = \underline{,172}$ (fra tabell).

b) Lar Y betegne antall av $n = 10$ tilfeldige pasienter som får ringvirkninger med preparat B. Y er da tn. Bino$(10, p)$ med ukjent p.

$H_0\!: p = 0{,}5,\ H_1\!: p < 0{,}5, T = Y, R\!: T \le k, P(T \le k \mid p =,5) \le,05$

Fordi vi fra a) har at $P(T \le 3 \mid p = ,5) = ,172 > ,05$ må k være mindre enn 3, følgelig: Ikke grunnlag på nivå 5 % å konkludere at preparat B gir mindre sannsynlighet for ettervirkning enn preparat A.

Oppgave 7.23 (E) Produsenten av en bestemt type tabletter merker tablettene med et tall μ, som angir den mengde aktivt stoff hver tablett skal inneholde. På grunn av ukontrollerbare variasjoner i produksjonsprosessen, vil imidlertid det eksakte innhold aktivt stoff ikke være μ, men må oppfattes som en stokastisk variabel X med (kjent) forventning μ og varians σ^2. Vi antar at X er normalfordelt.

Anta $\mu = 100$ mg og $\sigma = 2$ mg.

a) En tablett anses som defekt dersom mengden aktivt stoff avviker mer enn 4 mg fra μ. Finn sannsynligheten for at en tilfeldig tablett er defekt.

b) Tablettene pakkes på brett med 10 tabletter pr. brett. La S være total mengde aktivt stoff i de 10 tablettene. Hvilken sannsynlighetsfordeling får S? Finn sannsynligheten for at total mengde aktivt stoff i de 10 tablettene overstiger 1010 mg.

c) Brettene selges i esker med 5 brett i hver eske. La Y være antall brett i en eske der totalmengden av aktivt stoff overstiger 1010 mg. Hvilken sannsynlighetsfordeling får Y? Finn sannsynligheten for at minst 2 av brettene i en eske har mer enn 1010 mg aktivt stoff.

d) Produsenten av tablettene tar daglig stikkprøver fra produksjonen for å kontrollere verdien av μ. 10 tabletter tas tilfeldig fra dagens produksjon, og blir analysert i bedriftens laboratorium. Dersom det er grunn til å tro at $\mu \ne 100$ mg må produksjonsprosessen stoppes og justeres. En dag ble følgende verdier (mg) målt for mengden av aktivt stoff:

   103 98 99 104 101 100 104 99 102 103
   Er det ut fra disse verdiene grunn til å tro at $\mu \ne 100$ mg?

   Formuler spørsmålet som et hypoteseprøvingsproblem og test med 5 % signifikansnivå.

e) Bestem testens styrkefunksjon. Hva er sannsynligheten for at prosessen blir stanset og justert dersom μ i virkeligheten er 102 mg?

f) Hvordan ville du utført testen i d) dersom σ var ukjent? Utfør testingen også i dette tilfellet.

Løsningsforslag: $[X] = \text{mg}$.

a) $P(|X-\mu|<4) = P(-4 < X-\mu > 4) = 1 - P(-4 < X-\mu < 4) =$
$1 - P(\frac{-4}{\sigma} < Z = \frac{X-\mu}{\sigma} < \frac{4}{\sigma}) = 1 - (\Phi(2) - \Phi(-2)) = 2\Phi(-2) = \underline{,0455}$

b) $S \sim N(n\mu, \sqrt{n}\sigma) = \underline{N(1000, \sqrt{10}\cdot 2)}$, $p = P(S > 1010) = \Phi\left(-\frac{1010-1000}{\sqrt{10}\cdot 2}\right) = \underline{,0569}$

c) $Y \sim \text{Bino}(5, p)$ med $p = ,0569$ fra b). $P(Y \geq 2) = 1 - P(Y \leq 1) =$
$1 - \binom{5}{0} \cdot ,0569^0 \cdot ,9431^5 - \binom{5}{1} \cdot ,0569^1 \cdot ,9431^4 = \underline{,0288}$

d) $H_0\!: \mu = 100,\ H_1\!: \mu \neq 100,\ T = \dfrac{\overline{X} - 100}{2/\sqrt{10}},\ R\!: |T| > z_{\alpha/2} = z_{,025} = 1{,}96$
Data gir $\bar{x} = 101{,}3,\ s = 2{,}21,\ T = (101{,}3 - 100)/(2/\sqrt{10}) = 2{,}06,\ |T| = 2{,}06 > 1{,}96$
Konklusjon: <u>Grunnlag</u> i data på nivå 5 % å konkludere at $\mu \neq 100$.

e) Styrkefunksjon: $\gamma(\mu) = P(R\,|\,\mu) = P(|T| > z_{\alpha/2}\,|\,\mu) = 1 - P(-z_{\alpha/2} < T < z_{\alpha/2}) =$
$1 - P(-z_{\alpha/2} < \frac{S-\mu+\mu-100}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1 - P(-z_{\alpha/2} - \frac{\mu-100}{\sigma/\sqrt{n}} < Z = \frac{S-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2} - \frac{\mu-100}{\sigma/\sqrt{n}}) =$
$1 - \Phi\left(z_{\alpha/2} - \frac{\mu-100}{\sigma/\sqrt{n}}\right) + \Phi\left(-z_{\alpha/2} - \frac{\mu-100}{\sigma/\sqrt{n}}\right)$
$\gamma(102) = 1 - \Phi\left(1{,}96 - \frac{102-100}{2/\sqrt{10}}\right) + \Phi\left(-1{,}96 - \frac{102-100}{2/\sqrt{10}}\right) = 1 - \Phi(-1{,}20) + \Phi(-5{,}12) = \underline{,88}$

f) t-test: $T = \dfrac{\overline{X} - \mu}{S/\sqrt{n}},\ R\!: |T| > t_{,025}(9) = 2{,}26$ der $S/\sqrt{n}$ nå er empirisk standardavvik til $\overline{X}$ og $t_{,025}(9)$ er øvre 2,5 % fraktil i stud(n–1) = stud(9)-fordelingen. Data gir:
$T = \dfrac{101{,}30 - 100}{2{,}214/\sqrt{10}} = 1{,}86,\ |T| = 1{,}86 < 2{,}26$. Konklusjon: <u>Ikke</u> forkast H_0.

Oppgave 7.24 (E) Noen fysikere var uenige om størrelsen på lyshastigheten. En gruppe, «de ortodokse», påstod at den var 300000 km/s, mens en annen, «skeptikerne», hevdet at det var helt usannsynlig at den var «et så rundt tall», og trodde altså ikke på «de ortodokses» påstand. Man ble enige om å utføre 9 målinger, og at måleresultatene kunne antas å være uavhengige, forventningsrette og normalfordelte $N(\mu, 300)$.

a) Sett opp en nullhypotese og en alternativhypotese, og forklar hva det vil si å gjennomføre en test med signifikansnivå 5 %.

b) Formuler testen og gjennomfør den med følgende måleresultater:

300508 299150 299350 300137 299891 298592 300015 299799 298937

c) Bestem testens styrkefunksjon og beregn styrken for μ = 299700, 299900 og 300100. Forklar hva styrken uttrykker.

d) Det ble uenighet om hvor egnet testen var. Man besluttet derfor å foreta nye målinger. Man krevde at testen fortsatt skulle ha signifikansnivå $\alpha = 0{,}05$, men nå skulle styrken være minst 0,95 i $\mu = 299700$. I mellomtiden hadde «de ortodokse» med et annet eksperiment slått fast at lyshastigheten i hvert fall ikke var høyere enn 300000 km/s, og det godtok skeptikerne.

- Hva er en rimelig mothypotese nå?
- Hvor stor må n, antall målinger, være i den nye testen?

<u>Løsningsforslag</u>: X = tilfeldig måling av lyshastighet, $[X]$ = km/s.

a) $H_0: \mu = EX = 300000,\ H_1: \mu \neq 300000$

Å gjennomføre en test med signifikansnivå 5 % vil si å undersøke om estimert avvik fra 300000 er så stor at det er maksimalt 5 % sannsynlighet å observere et så stort avvik eller større dersom $\mu = 300000$.

b) $T = \dfrac{\overline{X} - 300000}{300/\sqrt{n}}$, $R\!: |T| > z_{\alpha/2} = z_{,025} = 1{,}96$

Data gir $\bar{x} = 299597{,}7$ og $T = \dfrac{299597{,}7 - 300000}{300/\sqrt{9}} = -4{,}00,\ |T| = 4{,}00 > 1{,}96$

Konklusjon: <u>Grunnlag</u> i data på nivå 5 % til å konkludere at $\mu \neq 300000$

c) $\mu_0 = 300000,\ \gamma(\mu) = P(R\,|\,\mu) = P(\,|T| > z_{\alpha/2}\,|\,\mu) = 1 - P(-z_{\alpha/2} < T < z_{\alpha/2}) =$

$1 - P(-z_{\alpha/2} < \frac{\overline{X} - \mu + \mu - 300000}{300/\sqrt{n}} < z_{\alpha/2}) = 1 - P(-z_{\alpha/2} - \frac{\mu - \mu_0}{300/\sqrt{n}} < Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2} - \frac{\mu - \mu_0}{300/\sqrt{n}}) =$

$1 - \Phi\left(z_{\alpha/2} - \frac{\mu - \mu_0}{300/\sqrt{n}}\right) + \Phi\left(-z_{\alpha/2} - \frac{\mu - \mu_0}{300/\sqrt{n}}\right) = \underline{1 - \Phi\left(1{,}96 - \frac{\mu - 300000}{100}\right) + \Phi\left(-1{,}96 - \frac{\mu - 300000}{100}\right)}$

Innsatt $\mu = 299700$, 299900 og 300100 får vi:

$\underline{\gamma(299700) = {,}851,\ \ \gamma(299900) = {,}170,\ \ \gamma(300100) = {,}170}$

Styrken uttrykker sannsynligheten for å forkaste H_0 (konkludere at $\mu \neq 300000$) som funksjon av den sanne verdien til μ.

d) $\underline{H_1\!: T < -z_\alpha = -z_{,95} = -1{,}645}$

Styrkefunksjonen bli nå modifisert til uttrykket $\gamma(\mu) = \Phi\left(-z_\alpha - \frac{\mu - \mu_0}{300/\sqrt{n}}\right)$.

$\gamma(\mu) \geq {,}95 \Rightarrow \Phi\left(-z_{,05} - \frac{\mu - \mu_0}{300/\sqrt{n}}\right) \geq {,}95 \Rightarrow -z_{,05} - \frac{\mu - \mu_0}{300/\sqrt{n}} \geq z_{,05} = 1{,}645 \Rightarrow$

$n \geq \left(\frac{2 \cdot z_{,05} \cdot 300}{\mu - \mu_0}\right)^2 = \left(\frac{2 \cdot 1{,}645 \cdot 300}{299700 - 300000}\right)^2 = 10{,}82$, dvs. <u>$n = 11$</u> målinger må foretas.

Oppgave 7.25 (E) Forbrukerkontoret i en bestemt by har mottatt klager på en bestemt pizzaprodusent. Denne produsenten hevder at deres store pepperonipizzaer i gjennomsnitt inneholder 60 gram pepperoni. Publikum mener at vekten av pepperoni må være betydelig lavere. En konsulent ved forbrukerkontoret får i oppdrag å utføre en hypotesetest. Vekten av pepperoni (pr. pizza) antas normalfordelt med forventningsverdi μ og standardavvik på 15 gram. Hypotesene som skal testes er:

H_0: $\mu = 60$ mot H_1: $\mu < 60$

Konsulenten velger ut én pizza for inspeksjon.

a) Foreta hypotesetest med signifikansnivå $\alpha = 0{,}05$. Hva blir konklusjonen dersom den observerte pepperonivekten er 39 gram?

b) Finn det minste signifikansnivået som gjør at vi forkaster nullhypotesen med en pepperonivekt på 39 gram.

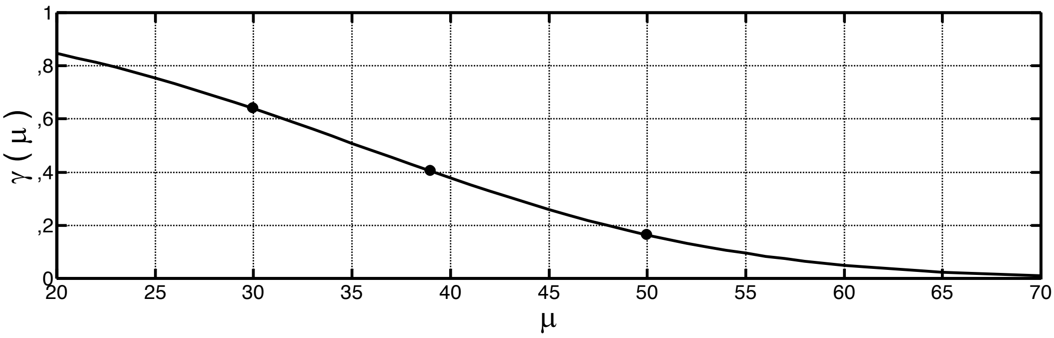
c) La $\alpha = 0{,}05$. Beregn styrken til testen i de tilfellene der $\mu = 30$, $\mu = 39$ og $\mu = 50$ (gram). Skissér styrkefunksjonen. Hvilken informasjon gir styrkefunksjonen?

Løsningsforslag: X = vekten av tilfeldig pizza, $[X]$ = gram.

a) $T = (X - 60)/15,\ R\!: T < -z_{,05} = -1{,}645$. $T(x = 39) = (39–60)/15 = -1{,}4$.
Konklusjon: Ikke grunnlag på nivå 5 % til å hevde at $\mu < 60$.

b) $P(T < -1{,}4 \mid \mu = 60) = \Phi(-1{,}4) = \underline{8{,}1\,\%}$

c) $\gamma(\mu) = P(R \mid \mu) = P((X - \mu + \mu - 60)/15 < -z_{,05}) = \underline{\Phi\left(-z_{,05} - (\mu - 60)/15\right)}$
Setter inn $z_{,05}$ = 1,656 og får: $\underline{\gamma(30) = ,639,\ \gamma(39) = ,403,\ \gamma(50) = ,164}$



Styrkefunksjonen er vist i figuren ovenfor og angir sannsynligheten for å forkaste H_0 (konkludere at $\mu < 60$) som funksjon av μ.

Oppgave 7.26 (E) I en befolkningsgruppe antar vi at fødselsvekten X (i kg) til et vilkårlig nyfødt barn er å betrakte som en normalfordelt variabel. Gjennomsnittet μ i fordelingen antas å være 3,5 kg og standardavviket σ = 0,35 kg. Man har en stund hatt mistanke om at den delen av befolkningsgruppen som bor i Tåkedalen, et område med mye forurensning, får barn med lavere fødselsvekt enn den øvrige delen av befolkningsgruppen. For å undersøke dettte nærmere, registrerte man fødselsvekten til 1000 Tåkedal-barn over en tidsperiode. Dette resulterte i en gjennomsnittsvekt på $\bar{x} = 3{,}48$ kg. Anta at vektene til Tåkedal-barna er uavhengige og $N(\mu, 0{,}35)$-fordelte.

a) Foreta hypotesetest av H_0: $\mu = 3{,}50$ mot H_1: $\mu < 3{,}50$ med signifikansnivå på 1 %. Hva blir konklusjonen med $\bar{x} = 3{,}48$ kg?

b) Hvor mange barn må undersøkes hvis en ønsker at μ med 95 % sikkerhet skal ligge innenfor et konfidensintervall med bredde 0,02?

c) La oss se bort fra signifikansnivået i a). Hva er det minste signifikansnivået som gjør at en kan forkaste H_0 til fordel for H_1 når $\bar{x}$=3,48 kg?

Løsningsforslag:

a) $\overline{X} \sim N(\mu, ,35/\sqrt{n}),\ T = \dfrac{\overline{X} - 3{,}50}{,35/\sqrt{n}},\ R\!: T < -z_\alpha = -z_{,01} = -2{,}33$

$T(\bar{x} = 3{,}48) = (3{,}48 - 3{,}50) / (,35 / \sqrt{1000}) = -1{,}81$

Konklusjon: Ikke grunnlag på nivå 1 % til å konkludere at $\mu < 3{,}50$.

b) 95 % KI for μ gitt ved $\bar{X} \mp z_{\alpha/2} \cdot \sigma / \sqrt{n}$ med bredde $2 \cdot z_{\alpha/2} \cdot \sigma / \sqrt{n}$.

$2 \cdot z_{\alpha/2} \cdot \sigma / \sqrt{n} \leq ,02 \Rightarrow n \geq \left(\frac{2 \cdot 1{,}96 \cdot ,35}{,02}\right)^2 = 4705{,}96$, dvs. $\underline{n = 4706}$ må undersøkes.

c) $P(T < \frac{3{,}48-3{,}50}{,35/\sqrt{1000}}) = \Phi\left(\frac{3{,}48-3{,}50}{,35/\sqrt{1000}}\right) = \Phi(-1.81) = ,035$, $\underline{\alpha_{\min} = 3{,}5\,\%}$.

Oppgave 7.27 (E) En brusautomat er konstruert slik at den - hvis den fungerer riktig - gir porsjoner på 218 g. På grunn av tilfeldigheter varierer imidlertid porsjonene fra gang til gang, slik at de følger en normalfordeling med standardavvik på 6,2 g. I det siste har mange kunder klaget og hevdet at automaten gir for lite brus. Ledelsen bestemmer seg for å undersøke om klagene er berettiget, og foretar derfor kontrollveiing av 16 tilfeldig valgte porsjoner.

a) Hvilken nullhypotese og mothypotese bør ledelsen teste i denne situasjonen?

Ledelsen velger å gjennomføre en test med signifikansnivå $\alpha = 0{,}05$. Forklar hva dette betyr.

b) Formulér testen.

Gjennomfør testen gitt at gjennomsnittsvekten til de 16 porsjonene er 214,6 g.

Gjennomfør også testen for det tilfellet der gjennomsnittsvekten er 216,5 g.

c) Beregn styrken til testen i punktet 216. Hva forteller svaret deg?

Beregn dessuten styrken til testen i punktene 212 og 214, og skissér styrkefunksjonen. Hva innebærer det at styrkefunksjonen er strengt avtagende i dette tilfellet?

d) Noen i ledelsen synes at testen er for usikker og for lite følsom. De foreslår i stedet å velge signifikansnivået $\alpha = 0{,}01$ og dessuten kreve at styrken av testen i punktet 216 skal være minst 0,90. Formulér den nye testen, og bestem det minste antall porsjoner vi i dette tilfellet må kontrollveie. Kan vi nå forkaste nullhypotesen hvis gjennomsnittsvekten av de kontrollveiede porsjonene er 216,5 g?

Løsningsforslag: X = vekt i gram til tilfeldig person. Antar $\bar{X}$ tn. $N(\mu, 6{,}2/\sqrt{n})$.

a) $\underline{H_0: \mu = 218,\ H_1: \mu < 218}$. 5 % nivå betyr at det maksimalt skal være 5 % sannsynlighet for feilaktig å konkludere med for lite (forventet) brusinnhold.

b) $T = \dfrac{\bar{X} - 218}{6{,}2 / \sqrt{n}}$, $R: T < -z_\alpha = -z_{,05} = -1{,}645$

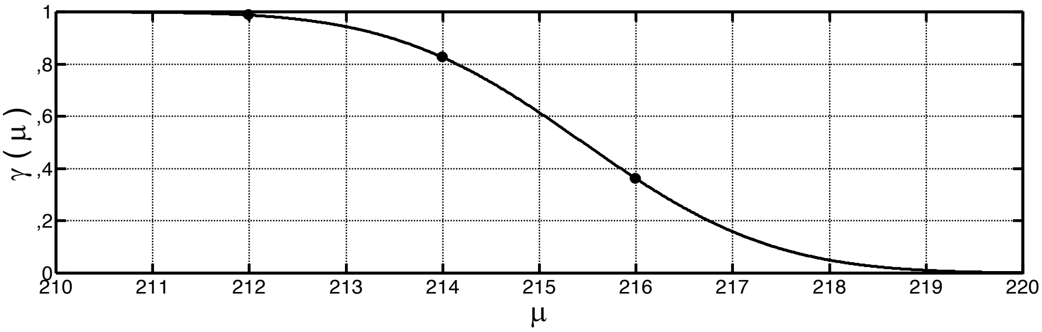
$T(214{,}6) = \frac{214{,}6-218}{6{,}2/\sqrt{16}} = -2{,}19$ Konklusjon: Forkast H_0.

$T(216{,}5) = \frac{216{,}5-218}{6{,}2/\sqrt{16}} = -{,}97$ Konklusjon: Ikke forkast H_0.

c) $\gamma(\mu) = P(R \mid \mu) = P(T < -z_\alpha \mid \mu) = P(\frac{\overline{X}-\mu}{6{,}2/\sqrt{n}} + \frac{\mu-218}{6{,}2/\sqrt{n}} \mid \mu) = \Phi(-z_\alpha - \frac{\mu-218}{6{,}2/\sqrt{n}} \mid \mu)$

Med $z_{,05}$ og n = 16 får vi γ(216) = Φ(–,35) = <u>,36</u>, dvs. at det er 36 % sannsynlighet for å forkaste testen dersom forventet brusinnhold er 216 g.

Tilsvarende fås $\gamma(212) = \Phi(2{,}23) = \underline{,99}$, $\gamma(214) = \Phi(,94) = \underline{,83}$.



Styrkefunksjonen er skissert i figuren ovenfor. At den er avtagende innebærer at sannsynligheten for å konkludere med for lavt (forventet) brusinnhold øker jo lavere (forventet) brusinnhold det er.

d) $\gamma(216) = \Phi\left(-z_{,01} - \frac{216-218}{6{,}2/\sqrt{n}}\right) \geq ,90 \Rightarrow -z_{,01} + \frac{2}{6{,}2}\sqrt{n} \geq z_{,10} \Rightarrow n \geq \left(\frac{(1{,}28+2{,}33)\cdot 6{,}2}{2}\right)^2 = 125{,}2$

dvs. vi må undersøke <u>n = 126</u> porsjoner. T og R blir da:

$$T = \frac{\overline{X}-218}{6{,}2/\sqrt{126}},\ R{:}\ T < -z_{,01} = -2{,}33,\ T(\bar{x} = 216{,}5) = -2{,}72,\ \ \underline{\text{Forkast } H_0}$$

Oppgave 7.28 (E) En bilfabrikk hevder at motoren i deres nyeste modell yter 135 hk i gjennomsnitt ved turtallet 4000 omdr./min., når anbefalt drivstoff benyttes. Noen bilkjøpere tror at ytelsen er lavere, og ber et frittstående biltestefirma måle motorstyrken i henhold til fabrikkens spesifikasjoner. 16 tilfeldig utvalgte biler av denne modellen ble undersøkt. Firmaet fant at motorstyrken i gjennomsnitt var 128 hk. Anta at motorstyrken er normalfordelt med forventning μ og standardavvik σ = 14 hk.

a) Formulér en passende null-hypotese og en mothypotese. Kan bilkjøperne påstå at de har rett dersom de er villige til å ta en feilrisiko på 1 %?

b) Beregn testens styrke for μ = 128 hk.

c) Hvor mange biler må tas med i undersøkelsen for at styrken for μ = 128 hk skal bli minst 90 %, når signifikansnivået fortsatt skal være 1 %?

<u>Løsningsforslag</u>: X = motorytelse i hestekrefter til tilfeldig valgt bil.

a) $H_0{:}\,\mu = 135,\ H_1{:}\,\mu < 135.\ \overline{X} \sim N(\mu, 14/\sqrt{n}),\ T = \frac{\overline{X}-135}{14/\sqrt{n}},\ R{:}\ T < -z_{,01} = 2{,}33$

$T(\bar{x} = 128) = \frac{128-135}{14/\sqrt{16}} = -2$. Konklusjon: <u>Ikke</u> grunnlag på nivå 1 % å konkludere at bilene har for lav ytelse.

b) $\gamma(\mu) = P(R \mid \mu) = P(T < -z_{,01} \mid \mu) = P(\frac{\overline{X}-\mu+\mu-135}{14/\sqrt{16}} < -z_{,01} \mid \mu) = \Phi(-z_{,01} - \frac{\mu-135}{14/\sqrt{16}})$

$\gamma(128) = \Phi(-2{,}33 - \frac{128-135}{14/\sqrt{16}}) = \Phi(-{,}33) = \underline{{,}37}$

c) $\gamma(128) = \Phi\left(-z_{,01} - \frac{128-135}{14/\sqrt{n}}\right) \geq {,}90 \Rightarrow -z_{,01} + \frac{7}{14}\sqrt{n} \geq z_{,10} \Rightarrow$

$n \geq \left(2 \cdot (1{,}28 + 2{,}33) \cdot 6{,}2\right)^2 = 52{,}1.$ Konklusjon: $\underline{n = 53}$ biler må undersøkes.

Kapittel 8

To populasjoner

Oppgave 8.1 For å sammenligne to metoder for opplæring av industriarbeidere, blir 20 arbeidere valgt ut til et forsøk. Av disse blir 10 valgt tilfeldig for å prøve metode 1, og de 10 andre prøver metode 2. Etter opplæringen utfører arbeiderne en individuell test der forventet tid (i minutter) brukt på testen er målet på hvor god metoden er. Følgende data oppnås:

	Tid [minutter]									
Met. 1	15	20	11	23	16	21	18	16	27	24
Met. 2	23	31	13	19	23	17	28	26	25	28

a) Kan du konkludere fra dataene at forventet tid er signifikant mindre etter trening med metode 1 enn med 2?

(test med $\alpha = 0{,}05$)

b) Formuler de forutsetningene du gjør om populasjonsfordelingene.

c) Konstruer et 95 % konfidensintervall for forskjellen i forventet tid for de to metodene.

<u>Løsningsforslag</u>: La X_i og Y_i betegne tid (i minutter) til forsøksperson nr. i som prøver henholdsvis metode 1 og 2.

a) Antar $\overline{X}$ er $N(\mu_1, \sigma/\sqrt{n})$ og at $\overline{Y}$ er $N(\mu_2, \sigma/\sqrt{n})$.

$$H_0\colon \mu_1 = \mu_2,\ H_1\colon \mu_1 \neq \mu_2,\ T = \frac{\overline{X}-\overline{Y}}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}},\ R\colon |T| > z_{\alpha/2} = z_{,025} = 1{,}96.$$

Innsatt data får vi: $\bar{x} = 19{,}1,\ \bar{y} = 23{,}3,\ s_x = 4{,}82,\ s_y = 5{,}56$

$$s_p = \sqrt{\tfrac{9\cdot 4{,}82^2 + 9\cdot 5{,}56^2}{10+10-2}} = 5{,}20,\ T = \tfrac{19{,}1-23{,}3}{5{,}20\sqrt{\frac{1}{10}+\frac{1}{10}}} = -1{,}81,\ |T| = 1{,}81 < 1{,}96.$$

Konklusjon: <u>Ikke</u> grunnlag i data på nivå 5 % å konkludere at $\mu_1 \neq \mu_2$.

b) Har antatt at $X_1, \ldots, X_n$ er uif $N(\mu_1, \sigma)$ og at $Y_1, \ldots, Y_n$ er uif $N(\mu_2, \sigma)$.

c) 95 % KI for $(\mu_1 - \mu_2)$: $\bar{x} - \bar{y} \mp t_{,025}(18)\cdot s_p\sqrt{\frac{1}{10}+\frac{1}{10}} = \underline{(-9{,}09\ ,\ 0{,}69)}$, $t_{,025}(18) = 2{,}10$.

Oppgave 8.2 En sosiolog ønsker å sammenligne fødselsraten til kvinner i to stammer A og B i Øst-Afrika. Fra hver stamme blir det plukket ut et tilfeldig utvalg på 100 kvinner i aldersgruppen 50-60 år, og antall barn hver kvinne har født blir registrert. Følgende frekvensfordeling oppnås:

	0	1	2	3	4	5	6	7	8	Tot
frekv. A	6	14	18	25	19	11	5	2	0	100
frekv. B	0	3	8	18	30	19	15	5	2	100

a) Beregn middelverdi og standardavvik til hver frekvensfordeling

b) Indikerer dataene en signifikant forskjell i forventet antall barn født av kvinner i stamme B?

c) Konstruer et 98 % konfidensintervall for differansen mellom de to populasjons-forventningene.

Løsningsforslag:

a) $\bar{x}_A = \frac{1}{100}\cdot(0\cdot 6+1\cdot 14+\cdots+8\cdot 0)=\underline{3},\ \overline{x_A^2}=\frac{1}{100}\cdot(0^2\cdot 6+1^2\cdot 14+\cdots+8^2\cdot 0)=11{,}68$

$s_A=\sqrt{\overline{x_A^2}-\bar{x}_A^2}=\sqrt{11{,}68-3^2}=\underline{1{,}64}$. Tilsvarende: $\bar{x}_B=\underline{4{,}29},\ s_B=\underline{1{,}50}$

b) Antar $\bar{X}_A \sim N(\mu_A,\sigma_A),\ \bar{X}_B \sim N(\mu_B,\sigma_B)$.

$$H_0:\mu_A-\mu_B=0,\ H_1:\mu_A-\mu_B\neq 0,\ T=\frac{\bar{X}_A-\bar{X}_B}{\sqrt{S_A^2/n_A+S_B^2/n_B}},\ R:|T|>z_{\alpha/2}=z_{,025}=1{,}96$$

Innsatt data fra a): $T=\frac{3-4{,}29}{\sqrt{1{,}54^2/100+1{,}50^2/100}}=-6{,}00$. Konklusjon: Grunnlag i data på nivå 5 % til å konkludere at $\mu_B\neq\mu_A$.

c) Tn. 98 % KI for $\mu_A-\mu_B$: $\bar{x}_A-\bar{x}_B\mp z_{,01}\cdot\sqrt{s_A^2/n_A+s_B^2/n_B}=$

$3-4{,}29\mp 2{,}33\cdot\sqrt{1{,}54^2/100+1{,}50^2/100}=\underline{(-1{,}81\,,-{,}77)}$

Oppgave 8.3 Målinger av gripestyrken til venstre- og høyre-handa til 10 keivhendte skribenter registreres:

	Person									
	1	2	3	4	5	6	7	8	9	10
Venstre	140	90	125	130	95	121	85	97	131	110
Høyre	138	87	110	132	96	120	86	90	129	100

a) Underbygger dataene påstanden at keivhendte har en større gripestyrke i venstrehånda enn i høyrehånda ?

b) Konstruer et 90 % konfidensintervall for forventet forskjell.

Løsningsforslag: La $D_i=V_i-H_i$ betegne differansen til gripestyrke mellom venstre og høyre hånd til forsøksperson nr. *i*.

a) $\bar{D}\sim N(\delta,\sigma_D/\sqrt{n}),\ H_0:\delta=0, H_1:\delta>0, T=\dfrac{\bar{D}}{S_D/\sqrt{n}}, R:T>t_{,05}(9)=1{,}83$

Data gir $\bar{x}=112{,}4,\ \bar{y}=108{,}8,\ \bar{d}=\bar{x}-\bar{y}=3{,}6,\ s_D=5{,}46, T=\frac{3{,}6}{5{,}46/\sqrt{10}}=2{,}09>1{,}83$

Konklusjon: Grunnlag i data på nivå 5 % å konkludere at forventet gripestyrke med venstre hånd er større enn med høyre for kjeivhendte.

b) Tn. 90 % KI for δ: $\bar{d} \mp t_{,05}(9) \cdot s_D / \sqrt{n} = 3{,}6 \mp 1{,}83 \cdot 5{,}46 / \sqrt{10} = \underline{(,44\ ,\ 6{,}76)}$

Oppgave 8.4 En ønsker å sammenlikne det surstoffforbruk (mm^3 pr. time) en ørret har når den svømmer i henholdsvis hurtig (A) og i langsomt (B) rennende vann, og gjør et forsøk med i alt 8 nummererte ørreter som svømmer en viss tid i A og samme tid i B. For hver ørret avgjør en ved loddtrekning om den først skal svømme i A eller B. I hvert enkelt forsøk blir surstoffforbruket bestemt, og resultatet blir:

	Ørret nr:							
	1	2	3	4	5	6	7	8
A	107	98	87	118	96	125	131	106
B	94	69	91	87	97	112	107	80

Se på differansene X = A – B:

X: 13 29 –4 31 –1 13
24 26

og anta at X er normalfordelt. Test om ørretens surstofforbruk er større når den svømmer i hurtig rennende vann enn når den svømmer i langsomt rennende vann. Bruk 5 % signifikansnivå.

Løsningsforslag:

Antar $\bar{X}$ er $N(\delta, \sigma/\sqrt{n}), [X] = \text{mm}^3 / \text{time}$

$$H_0{:}\delta = 0, H_1{:}\delta > 0, T = \frac{\bar{D}}{S_D/\sqrt{n}}, R{:}T > t_{,05}(7) = 1{,}89. \text{ Data gir } T = \frac{16{,}38}{13{,}44/\sqrt{8}} = 3{,}45$$

Konklusjon: Grunnlag i data på nivå 5 % å konkludere at forventet surstofforbruk er større i hurtig enn i langsomt rennende vann.

Oppgave 8.5 To metoder A og B for måling av PH-verdier i en oppløsning skal sammenlignes. En gruppe studenter måler 10 ulike oppløsninger med begge målemetodene. En har mistanke om at de to metodene ikke gir samme gjennomsnittlige måleresultat.

Måleresultater:

Nr. (i)	1	2	3	4	5	6	7	8	9	10
A	8,14	7,19	6,75	7,00	7,64	7,01	6,98	6,30	7,52	7,40
B	7,91	7,48	7,31	7,36	7,62	7,38	7,40	6,60	7,49	7,35

Antagelser:
Metode A gir måleresultater X_i som er normalfordelt $N(\mu_i, \sigma_1)$.

Metode B gir måleresultater Y_i som er normalfordelt $N(\mu_i+\Delta,\sigma_2)$.

Differansene $Y_i - X_i$ antas uavhengige og normalfordelte med standardavvik σ.

a) Finn et punktestimat for Δ.

b) Bestem et 95 % konfidensintervall for Δ.

c) Finn et 99 % konfidensintervall på formen $[0,b]$ for σ.

Løsningsforslag:

a) $D_i = Y_i - X_i, \Delta^* = \overline{D} = \overline{Y} - \overline{X} = 7{,}39 - 7{,}193 = \underline{,197}$

b) 95 % KI for Δ: $\Delta^* \mp t_{,025}(9) \cdot s_D / \sqrt{10} = -,197 \mp 2{,}26 \cdot ,258 / \sqrt{10} = \underline{(,0126\ ,\ ,381)}$

c) 99 % KI for σ: $[0\ ,\ \sqrt{\frac{(n-1)\cdot s_D^2}{\chi^2_{,01}(9)}}\,] = [0\ ,\ \sqrt{\frac{9\cdot ,258^2}{2{,}088}}\,] = \underline{[0\ ,\ ,54]}$

Oppgave 8.6 To typer voltmetre A og B skal sammenlignes. En har mistanke om at voltmetre av typen B systematisk viser høyere verdi for spenningen enn voltmetre av type A. Seks studenter fikk hver utdelt et voltmeter av hver type og målte hver sin fritt valgte spenning med begge voltmetrene.

La X_i være målt spenning med voltmeter av type A, og Y_i målt spenning med voltmeter av type B.

Måleresultatene ble:

Stud. nr.	1	2	3	4	5	6
X_i [V]	5,0	2,1	6,9	8,3	9,4	11,8
Y_i [V]	5,3	2,1	7,1	8,2	9,7	12,0

Anta $Z_i = Y_i - X_i$ er uif $N(\Delta,\sigma)$.

a) Konstruer på grunnlag av observasjonene et 90 % konfidensintervall for Δ. Vil du på grunnlag av observasjonene påstå at B gir høyere verdier enn A?

b) Konstruer et 99 % konfidensintervall på formen $[0,b]$ for σ.

Løsningsforslag:

a) 90 % KI for Δ: $\bar{z} \mp t_{,05}(5) \cdot s_Z / \sqrt{6} = -,15 \mp 2{,}015 \cdot ,164 / \sqrt{6} = \underline{(-,285\ ,\ -,015)}$

Siden konfidensintervallet ikke dekker 0 kan vi på nivå 10 % påstå at B forventes å gi høyere verdi enn A.

b) 90 % KI $[0\ ,\ b]$ for Δ: $[0\ ,\ \sqrt{\frac{(n-1)\cdot s_z^2}{\chi^2_{,01}(5)}}\,] = [0\ ,\ \sqrt{\frac{5\cdot ,164^2}{,554}}\,] = \underline{[0\ ,\ ,49]}$

Kommentar: Siden datagrunnlaget er oppgitt med kun 1 desimal, som er i samme størrelsesorden som s_z, vil vi ha en diskretiseringseffekt som gjør vår normaltilnærmelse noe tvilsom. En måte å undersøke denne diskretiserings-effekten er ved hjelp av Monte Carlo simulering.

Oppgave 8.7 En PC-leverandør er interessert i å undersøke om bruk av fargeskjermer vil virke gunstig ved tekstbehandling. En gruppe på 10 personer ble bedt om å skrive inn en bestemt tekst på maskiner med svart-hvitt-skjermer. En annen gruppe på 10 personer ble bedt om å skrive inn den samme teksten mot fargeskjermer.

Tiden (i minutter) som hver av deltakerne brukte på å skrive inn teksten, ble målt.

Måleverdiene ($x_1,\ldots,x_{10}$) med svart-hvitt-skjerm, og ($y_1,\ldots,y_1$) med fargeskjerm er vist nedenfor:

x: 10,2 13,1 13,9 14,1 12,5 10,5 13,4 14,2 14,4 13,5
y: 11,5 12,3 13,1 10,5 9,9 11,3 13,4 11,0 13,2 12,8

Du kan i et av punktene nedenfor få bruk for følgende generelle konfidensintervall:

$$\bar{x}-\bar{y} \mp t_{\alpha/2} \cdot s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

der $t_{\alpha/2}$ er $\alpha/2$-fraktilen i t-fordelingen med $n_1 + n_2 - 2$ frihetsgrader, og

$$s_p^2 = \frac{(n_1-1)s_x^2+(n_2-1)s_y^2}{n_1+n_2-2}$$

Regnehjelp:

$\Sigma\, x_i = 129{,}8$, $\Sigma\, x_i^2 = 1705{,}0$, $\Sigma\, y_i = 119$, $\Sigma\, y_i^2 = 1429{,}7$

a) Beregn $\bar{x}$, $\bar{y}$ og de observerte standardavvikene s_x og s_y for dataene over.

b) Bestem et 95 % konfidensintervall for forskjellen mellom forventet tidsforbruk for de to gruppene. Vi antar normalfordeling i begge grupper, og uavhengighet både innen hver gruppe og gruppene imellom. Hvilken konklusjon kan du trekke på grunnlag av resultatet?

Gruppen som først skrev teksten inn mot fargeskjerm, ble dernest bedt om å skrive inn teksten på maskiner med svart-hvitt-skjerm. (Vi forutsetter at teksten er av en slik karakter at det ikke er noen fordel å ha skrevet den en gang tidligere). Måleverdiene ($z_1,\ldots,z_{10}$) er vist nedenfor i samme rekkefølge som observasjonene ($y_1,\ldots,y_{10}$):

z: 12,0 12,4 13,5 11,4 12,0 11,4
12,8 12,9 13,6 12,9

c) Bestem et 95 % KI for forskjellen mellom forventet tidsforbruk mot svart-hvitt-skjerm (z_i) og fargeskjerm (y_i). Anta normalfordeling. Tolk resultat. Oppgitt: $\bar{z}=12{,}44$, $s_{z\text{-}y}=\,{,}84$.

Løsningsforslag: Antar $\overline{X} \sim N(\mu_X, \sigma / \sqrt{n_X})$, $\overline{Y} \sim N(\mu_Y, \sigma / \sqrt{n_Y})$.

a) $\bar{x} = 129{,}8 / 10 = \underline{12{,}98}$, $s_x = \sqrt{\frac{1}{9} \cdot (1705{,}0 - 129{,}8^2 / 10)} = \underline{1{,}498}$

$\bar{y} = 119 / 10 = \underline{11{,}9}$, $s_y = \sqrt{\frac{1}{9} \cdot (1429{,}7 - 119^2 / 10)} = \underline{1{,}229}$

b) $s_p = \sqrt{\frac{9 \cdot 1{,}498^2 + 9 \cdot 1{,}229^2}{10+10-2}} = 1{,}37$

95 % KI for $\mu_X - \mu_Y$:

$\bar{x} - \bar{y} \mp t_{,025}(18) \cdot s_p \sqrt{\frac{1}{10} + \frac{1}{10}} = 12{,}98 - 11{,}9 \mp 2{,}10 \cdot 1{,}37 \sqrt{\frac{1}{10} + \frac{1}{10}} = \underline{(-{,}21\,,\, 2{,}37)}$

Siden konfidensintervallet ikke inneholder 0, kan vi på nivå 5 % ikke konkludere at det er noen signifikant forskjell på de to metodene.

c) $D_i = Z_i - Y_i \sim N(\delta, \sigma_{z-y})$. 95 % KI for δ:

$\bar{d} \mp t_{,025}(9) \cdot s_{z-y} / \sqrt{10} = 12{,}44 - 11{,}9 \mp 2{,}26 \cdot {,}84 / \sqrt{10} = \underline{(-{,}06\,,\, 1{,}14)}$

Siden konfidensintervallet ikke inneholder 0 kan vi på nivå 5 % ikke konkludere at det er signifikant forskjell i tidsforbruk mellom farge- og svart/hvit skjerm.

Oppgave 8.8 Man ønsker å undersøke om vitamin B_1 er et effektivt middel til å stimulere veksten av sopp. Til dette formålet blir 10 sopper valgt tilfeldig og 5 av disse får behandling av vitamin B_1. Vektene på soppene etter avsluttet forsøk var følgende:

Beh. med B_1:	27	34	20	28	20
Ikke beh.:	18	14.5	13.5	12.5	23

a) Beregn empirisk middelverdi, varians og standardavvik for de to gruppene.

Det antas heretter at vektfordelingen på behandlet sopp er $N(\mu_1, \sigma)$, og at vektfordelingen for ubehandlet sopp er $N(\mu_2, \sigma)$.

b) Finn forventningsrette estimater for $\mu_1 - \mu_2$ og σ^2.

c) Man ønsker å finne ut om det er grunnlag for å påstå at vitamin B_1 har en positiv effekt. Formulér dette som et hypotesetestingsproblem og test med 1 % signifikansnivå. Hva er din konklusjon ?

Løsningsforslag: La X og Y betegne vekt av sopp henholdsvis med og uten behandling.

a) $\Sigma x = 27 + \cdots + 20 = 129$, $\Sigma x^2 = 27^2 + \cdots + 20^2 = 3469$, $\Sigma y = 81{,}5$, $\Sigma y^2 = 1401{,}8$

$\bar{x} = 129 / 5 = \underline{25{,}8}$, $s_x^2 = \frac{1}{4} \cdot (3469 - 129^2 / 5) = \underline{35{,}2}$, $s_x = \sqrt{35{,}2} = \underline{5{,}93}$

$\bar{y} = 81{,}5 / 5 = \underline{16{,}3}$, $s_y^2 = \frac{1}{4} \cdot (1401{,}8 - 81{,}5^2 / 5) = \underline{18{,}34}$, $s_y = \sqrt{18{,}34} = \underline{4{,}28}$

b) $(\mu_1 - \mu_2)^* = \bar{x} - \bar{y} = 25{,}8 - 16{,}3 = \underline{9{,}5}$

$$(\sigma^2)^* = \frac{(5-1)\cdot s_x^2 + (5-1)\cdot s_y^2}{5+5-2} = \frac{4\cdot 35{,}2 + 4\cdot 18{,}34}{8} = \underline{26{,}77}$$

c) $H_0\!: \mu_1 - \mu_2 = 0,\ H_1\!: \mu_1 - \mu_2 > 0,\ T = \dfrac{\bar{X} - \bar{Y}}{S\sqrt{\frac{1}{5}+\frac{1}{5}}},\ R\!: T > t_{,01}(8) = 2{,}90$

Data gir: $T = \dfrac{25{,}8 - 16{,}3}{\sqrt{26{,}77}\cdot\sqrt{\frac{1}{5}+\frac{1}{5}}} = 2{,}903$. Konklusjon: Det er <u>såvidt grunnlag</u> i data på nivå 5 % til å konkludere at behandling B gir signifikant vekstvirkning. Antall sifre i mellomregningene er her avgjørende for konklusjonen.

Oppgave 8.9 To galluper utført av Norges Markedsdata for september 1981 og januar 1982 viste blant annet følgende resultater for preferanse i %:

	Sep.81	Jan.82
Vil stemme AP	36,4	38,7
Vil stemme H	23,6	30,5
Totalt antall som oppga preferanse.	1145	1134

a) Er endringen i andelen som vil stemme Høyre signifikant?

b) Finn et 95 % konfidensintervall for økningen til AP. Angi signifikanssannsynligheten.

<u>Løsningsforslag</u>:

Benytter subskript 1 for å betegne Sep.81 og subskript 2 for å betegne Jan. 82. Lar videre p betegne den ukjente andelen vi ville funnet om alle N stemmeberettigede var blitt spurt, og lar $n << N$ betegne antallet i utvalget. $p^* = X/n$ betegner estimator for p, der X er antallet i utvalget som svarer de vil stemme på det aktuelle partiet. Vi antar uavhengighet mellom alle besvarelsene.

a) Sannsynlighetsmodell: $p_1^* = X_1/n_1$ tn. $N(p_1, \sqrt{p_1(1-p_1)/n_1})$ og $p_2^* = X_2/n_2$ tn. $N(p_2, \sqrt{p_2(1-p_2)/n_2})$. På grunn av antagelsen om uavhengighet så vil differansen mellom års-estimatorene, $\Delta p^* = p_2^* - p_1^*$ være tn. normalfordelt:

$$\Delta p^* \text{ tn. } N(p_2 - p_1, \sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2})$$

Tester: $H_0\!:\ \Delta p = p_2 - p_1 = 0,\quad H_1\!:\ \Delta p \neq 0$

Testobservator: $T = \dfrac{p_2^* - p_1^*}{\sqrt{p_1^*(1-p_1^*)/n_1 + p_2^*(1-p_2^*)/n_2}}$ tn. $N(0,1)$ under H_0.

Forkastingsområde: $R\!:\ |T| > z_{\alpha/2} = z_{,025} = 1{,}96$

Gjennomføring av testen: $T = \dfrac{,305 - ,236}{\sqrt{,236 \cdot ,764 / 1145 + ,305 \cdot ,695 / 1134}} = 3{,}72$

$|\,T\,| = 3{,}72 > 1{,}96$, dvs. dataene gir grunnlag på nivå 5 % til å konkludere at det har vært en signifikant endring i andelen som vil stemme Høyre. Verdien 3,72 tilsvarer øvre 10^{-4} fraktil i $N(0,1)$-fordelingen, følgelig er signifikans-sannynligheten (*p*-verdien) $\underline{P^* = ,0002}$ (tosidig test).

b) Utifra samme sannsynlighetsmodell som i a) får vi følgende 95 % KI for økningen i AP's tilslutnging:

$\Delta p^* \mp z_{,025}\sqrt{p_1^*(1-p_1^*)/n_1 + p_2^*(1-p_2^*)/n_2} =$

$,387 - ,364 \mp 1{,}96\sqrt{,364 \cdot ,636 / 1145 + ,387 \cdot ,613 / 1134} = \underline{(-1{,}7\,\%\ ,\ 6{,}3\,\%)}$

Kapittel 9

Lineær regresjon

Oppgave 9.1 Gitt følgende (x, y)-data:

x	29,0	18,6	3,5	30,4	13,5	19,6	12,6
y	2,48	1,66	0,44	2,77	1,94	1,82	1,09

a) Tegn verdiene inn i et sprednings- diagram.

b) Beregn $\bar{x}$ og $\bar{y}$, og trekk på øyemål den rette linja gjennom ($\bar{x}, \bar{y}$) som ser ut til å passe best mulig til data.

Anta i det følgende at x-ene er konstante, kjente tall uten statistisk usikkerhet, mens y-ene er uavhengige realiseringer av en stokastisk variabel Y som er $N(a+bx, \sigma)$-fordelt.

c) Beregn regresjonslinja $y_r^* = a^* + b^* \cdot x$ og tegn denne inn i diagrammet. Får du bra overensstemmelse med linja du trakk på øyemål?

d) Estimer σ.

e) Estimer std(a^*), std(b^*) og std($y_r^* \mid x = 10{,}0$).

f) Anta at du har 4 uavhengige y-observasjoner i stedet for bare en for hver x-verdi. Med hvilken faktor vil std(b^*) og std(a^*) avta? Hva med Var(S^2)?

<u>Løsningsforslag</u>:

Hjelpestørrelser: $\Sigma x = 172{,}2$, $\Sigma x^2 = 2848{,}5$, $\Sigma y = 12{,}2$, $\Sigma y^2 = 25{,}037$, $\Sigma x \cdot y = 264{,}14$,
$S_x^2 = \Sigma x^2 - (\Sigma x)^2 / n = 537{,}1$, $S_y^2 = \Sigma y^2 - (\Sigma y)^2 / n = 3{,}774$, $S_{xy} = \Sigma xy - \Sigma x \Sigma y / n = 42{,}45$

a) Spredningsdiagram, se neste figur, '+' angir middelverdi

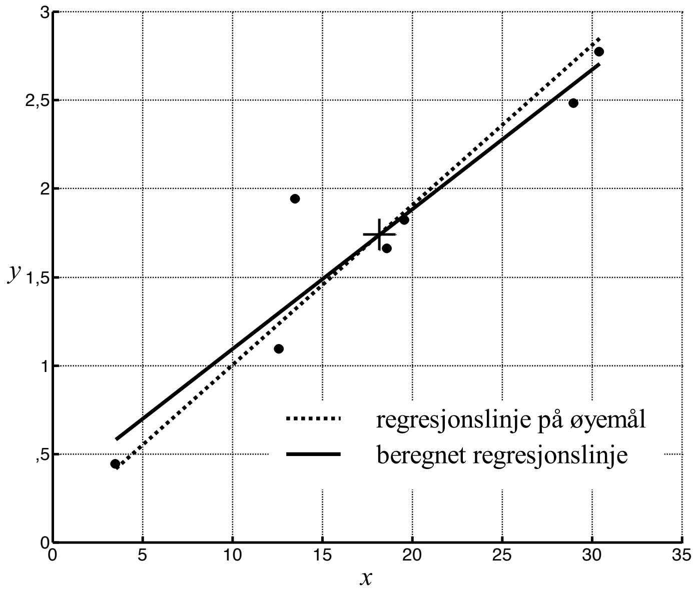
b) $\bar{x} = \Sigma x / n = 172{,}2 / 7 = \underline{18{,}17}$, $\bar{y} = \Sigma y / n = 12{,}2 / 7 = \underline{1{,}74}$. Regresjonslinje på øyemål i neste figur (prikket kurve).

c) $b^* = S_{xy} / S_x^2 = 42{,}45 / 537{,}1 = \underline{{,}079}$, $a^* = \bar{y} - b^* \cdot \bar{x} = 1{,}74 – {,}079 \cdot 18{,}17 = \underline{{,}305}$
Regresjonslinja $y_r^* = {,}305 + {,}079 \cdot x$ i neste figur viser brukbart samsvar med b).

d) $\mathrm{SSE} = \Sigma(y_i - a^* - b^* \cdot x_i) = S_y^2 - (b^*)^2 \cdot S_x^2 = 3{,}774 – {,}079^2 \cdot 537{,}1 = {,}422$
$S = \sigma^* = \sqrt{\mathrm{SSE} / (n-2)} = \sqrt{{,}422 / 5} = \underline{{,}291}$

e) $\mathrm{std}^*(a^*) = s \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_x^2}} = \underline{{,}253}$, $\mathrm{std}^*(b^*) = s / S_x = {,}291 / \sqrt{537{,}1} = \underline{{,}0126}$

$$\text{std}^*(y_r^* \mid x=10) = s\sqrt{\tfrac{1}{n} + \tfrac{(10-\bar{x})^2}{S_x^2}} = \underline{,150}$$



f) std(b^*) og std(a^*) er proporsjonale med $1/\sqrt{n}$. Når n firedobles avtar derfor disse standardavvikene med en faktor $\underline{2}$. La S_n^2 betegne S^2 med n observasjonspar.

$$\text{Var}(S_n^2) = 2\sigma^4/(n-2),\ \text{Var}(S_{4n}^2) = 2\sigma^4/(4n-2) \Rightarrow \tfrac{\text{Var}(S_{4n}^2)}{\text{Var}(S_n^2)} = \tfrac{n-2}{4n-2} = \underline{5/26}$$

Oppgave 9.2 Sammenhengen mellom en plutselig forandring (sprang) i input til et 1. ordens reguleringssystem og output y (sprangrespons) fra systemet er

(1) $\quad y = a(1 - e^{-t/T})$

der t er tiden fra forandringen trer i kraft, T er en tidskonstant som er mindre jo raskere systemet er og a er output når t går mot uendelig.

a) Anta at a er kjent og T ukjent. Transformer den ikke-lineære sammenhengen i lign. (1) over til en lineær sammenheng

(2) $\quad y' = a' + b' \cdot t$

Hva blir sammenhengen mellom de transformerte størrelsene y', a' og b' og de opprinnelige: y, a og T?

b) Anta fremdeles at a er kjent og bestem et uttrykk for minste kvadraters-estimatoren b'^* for b'. Hva blir den tilsvarende estimatoren T^* for T?

c) Gitt følgende sammenhengende verdier for t [sek] og y [V]:

t	1	2	3	4	5	6	7	8	9	100
y	1,85	2,98	5,96	6,09	7,32	7,44	7,65	8,20	8,29	10,00

Undersøk om det er grunn til å anta at systemet er et 1. ordens system, og estimer i såfall tidskonstanten T for systemet. Nevn en fallgrube du kan gå i her.

Løsningsforslag:

a) $y = a(1-e^{-t/T}) \Rightarrow 1 - y/a = e^{-t/T} \Rightarrow \ln(1-y/a) = (-\frac{1}{T})\cdot t$

Sammenlignes siste ligning ovenfor med ligninga $y' = a' + b'\cdot t$ ser vi at $\underline{y' = \ln(1-y/a), a' = 0, b' = -1/T}$.

b) $Q = \Sigma(y'_i - b'\cdot t_i)^2$, $\frac{\partial Q}{\partial b'} = 0 \Rightarrow \Sigma(y'_i - b'\cdot t_i)\cdot t_i = 0 \Rightarrow (b')^* = \underline{\frac{\Sigma y'_i\cdot t_i}{\Sigma t_i^2}}$

$T^* = -1/(b')^* = \underline{-\Sigma(t_i^2)/\Sigma(y'_i\cdot t_i)}$

c) Vi transformerer y-verdiene i tabellen til y'-verdier og får for stigende y-verder:

−0,205 −0,354 −0,906 −0,939 −1,317 −1,363 −1,448 −1,715 −1,766 −∞

Fra tabellen antas at asympotisk verdi kan settes til a = 10, fordi denne y-verdien er målt ved tidspunkt t = 100 som er langt større enn de øvrige t-veridene. En enkel indikasjon fås ved å undersøke korrelasjonskoeffisienten r mellom t og y'. Punktet (t,y) = (100, 10) bør imidlertid utelukkes, da denne må betraktes som en utligger som kunstig vil påvirke r-verdien. Vi transformerer y-verdiene i tabellen til y'-verdier og får for stigende y-verder:

y': −0,205 −0,354 −0,906 −0,939 −1,317 −1,363 −1,448 −1,715 −1,766 −∞

Utelater siste verdi og finner at $r = \text{corr}^*(y',t) = -0{,}97$. Dette er en sterk indikasjon på at det er rimelig å anta at vi har et tilnærmet 1. ordens reguleringssystem der tidskonstanten T^* kan estimeres som følger:

$$\sum_{i=1}^{9} t_i^2 = 285,\ \sum_{i=1}^{9} y'_i\cdot t_i = -61{,}9,\ T^* = -\Sigma(t^2)/\Sigma(y'\cdot t) = -285/61{,}9 = \underline{4{,}60}\ \text{(sek)}.$$

Oppgave 9.3 En temperaturmåler viser en bortimot perfekt lineær sammenheng mellom temperatur, X, og utslag på instrumentet, Y (alle størrelser i °C):

(3) $Y = X + W$

der W er en $N(a,\sigma)$-variabel. W ved et tidspunkt t_1 er uavhengig av W ved et annet tidspunkt t_2. Imidlertid viser det seg at temperaturmåleren er noe ukalibrert. For å kalibrere instrumentet, har fabrikanten i løpet av en uke målt temperaturen ved 0 °C og funnet følgende med basis i n = 1460 uavhengige observasjoner:

$$\bar{y} = 0{,}34\ ^{\circ}\mathrm{C}, \quad S_y = 1{,}23\ ^{\circ}\mathrm{C}$$

a) Hvilken av de angitte størrelser beskriver hvor ukalibrert instrumentet er?

b) Bestem et uttrykk for minste kvadraters estimatoren a^* for a og beregn a^*-verdien med basis i fabrikantens måleresultater.

c) Estimer standardavviket til a^* med basis i fabrikantens målinger.

Løsningsforslag: Enhet: °C

a) $\underline{a}$ angir hvor ukalibrert instrumentet er.

b) Med $x = 0$ fås $Y = W \Leftrightarrow Y = a + V, V \sim N(0,\sigma)$.
$Q = \Sigma(Y_i - \mathrm{E}Y_i)^2 = \Sigma(Y_i - a)^2$, $\frac{\partial Q}{\partial a} = 0 \Rightarrow \Sigma(Y_i - a) \Rightarrow \underline{a^* = \bar{Y}}$.
Data gir $a^* = \underline{0{,}34\ ^{\circ}\mathrm{C}}$

c) $\mathrm{std}^*(a^*) = s_y / \sqrt{n} = 1{,}23 / \sqrt{1460} = \underline{{,}032\ ^{\circ}\mathrm{C}}$

Oppgave 9.4 I situasjonene beskrevet nedenfor skal du, om mulig, identifisere hva som er uavhengig variabel (x) og hva som er avhengig responsvariabel (y):

a) I et medisinsk forsøk undersøkes sammenhengen mellom styrke av en sovepille og antall søvntimer.

b) I et havbruksanlegg måles fiskevekst som funksjon av fórmengde.

c) På en oljeplattform måles hver enkelt bølgehøyde samtidig som den resulterende variasjon i strekkspenningen i et konstruksjonselement på plattformen måles.

d) I et tilfeldig utvalg på 1000 elever fra videregående skole sammelignes matematikk- og engelsk-karakterene på individbasis.

Løsningsforslag:

a) x = styrke av sovepille, y = responsvariabel.

b) x = fórmengde, y = fiskevekst.

c) Naturlig å definere x som bølgehøyde og y som strekkspenning, fordi det er bølgelasten som forårsaker strekkspenningen og ikke omvendt. I dette tilfellet er imidlertid ikke x en «kontrollerbar» variabel.

d) Her er det ikke naturlig å definere en uavhengig variabel.

Oppgave 9.5 Gitt følgende (x,y)-data:

x	−1	0	1
y	−1	0	1

a) Beregn korrelasjonskoeffisienten r.

b) Legg til to nye datapar $(x,y) = (0,a)$ og $(x,y) = (0,-a)$, og vis at r nå blir $r = (1+a^2)^{-1/2}$. Hva går r mot når a går mot uendelig?

Løsningsforslag:

Hjelpestørrelser: $\Sigma x = 0$, $\Sigma y = 0$, $\Sigma x^2 = 2$, $\Sigma y^2 = 2$, $\Sigma xy = 2$

a) $r = \dfrac{\Sigma xy - \Sigma x \Sigma y / 3}{\sqrt{(\Sigma x^2 - (\Sigma x)^2/3)\cdot(\Sigma y^2 - (\Sigma y)^2/3)}} = \dfrac{2-0}{\sqrt{2\cdot 2}} = \underline{1}$

b) Forandringer: $\Sigma y^2 = 2 + 2a^2$, $r = \dfrac{2}{\sqrt{2\cdot(2+2a^2)}} = \dfrac{1}{\sqrt{1+a^2}}$, $\underline{\text{q.e.d.}}$

Når a går mot uendelig ser vi at r går mot $\underline{0}$.

Oppgave 9.6 Gitt følgende $(x$,y)-data:

x	-1	1	1	-1
y	-1	-1	1	1

a) Beregn korrelasjonskoeffisienten r.

b) Legg til to nye datapar $(x$,y$) = (-a,-a)$ og $(x,y) = (a,a)$, og vis at r nå blir $r = (1 + 2/a^2)^{-1}$. Hva går r mot når a går mot uendelig?

Løsningsforslag:

Hjelpestørrelser: $\Sigma x = 0$, $\Sigma y = 0$, $\Sigma x^2 = 4$, $\Sigma y^2 = 4$, $\Sigma xy = 0$.

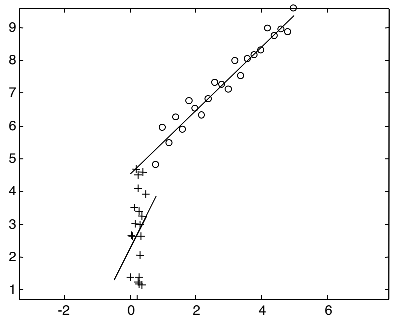
a) $r = 0/\sqrt{4\cdot 4} = \underline{0}$.

b) Forandringer: $\Sigma x^2 = 4 + 2a^2$, $\Sigma y^2 = 4 + 2a^2$, $\Sigma xy = 2a^2$

$$r = \frac{2a^2 - 0}{\sqrt{(4+2a^2)\cdot(4+2a^2)}} = \frac{2a^2}{4+2a^2} = \frac{a^2}{2+a^2} = \frac{1}{1+2/a^2}, \ \underline{\text{q.e.d.}}$$

Når a går mot uendlig så går r mot $\underline{1}$.

Oppgave 9.7 I figuren nedenfor ser dataene ut til å følge to forskjellige rette linjer. Det er tilpasset en rett linje til dataene merket «+» og en rett linje tilpasset dataene merket med sirkler.



Kan du gi en forklaring på hvorfor linja tilpasset «+»-dataene tilsynelatende gir en så dårlig tilpasning?

Løsningsforslag:

Ser vi nøyere på '+' dataene, ser vi at utstrekningen i horisontalretning er svært liten i forhold til utstrekning i vertikalretningen. Ved en slik skeiv skalering vil alltid observasjonspar se ut til å ligge på en rett linje (enten horisontal eller vertikal). Dette er nok et poeng som man av og til kan lure på utnyttes bevisst for å gi inntrykk av en sterkere korrelasjon enn hva som er reelt. Hadde vi «strukket» x-aksen ville vi sett at '+' dataene er svært ukorrelerte, og enhver rett linje ville gitt like god (eller dårlig) tilpasning.

Oppgave 9.8 Utled uttrykkene $b^* = S_{xy}/S_x^2$ og $a^* = \bar{y} - b^* \cdot \bar{x}$ for minste kvadraters-estimatorene a^* og b^*, ved å løse de to ligningene du får ved å derivere minste kvadrat-summen

$$Q = \Sigma(y_i - a - bx_i)^2$$

med hensyn på a og b, og sette de deriverte lik null.

Løsningsforlslag: Dropper subskript i unntatt i starten.

$$\tfrac{\partial Q}{\partial a} = 0 \Rightarrow \Sigma(y_i - a - bx_i)\cdot x_i = 0 \Rightarrow \Sigma x_i y_i - a\Sigma x_i - b\Sigma x_i^2 = 0 \Rightarrow a\Sigma x + b\Sigma x^2 = \Sigma xy \qquad (1)$$

$$\tfrac{\partial Q}{\partial b} = 0 \Rightarrow \Sigma(y_i - a - bx_i) = 0 \Rightarrow na + b\Sigma x = \Sigma y \Rightarrow a + b\bar{x} = \bar{y} \qquad (2)$$

(1) – (2)·Σx gir: $b(\Sigma x^2 - \bar{x}\Sigma x) = \Sigma xy - (\Sigma x)\cdot\bar{y} \Rightarrow b^* = \dfrac{\Sigma xy - \Sigma x\Sigma y/n}{\Sigma x^2 - (\Sigma x)^2/n} = \dfrac{S_{xy}}{S_x^2}$, q.e.d.

(2) $\Rightarrow a^* = \bar{y} - b^* \cdot \bar{x}$, q.e.d.

Oppgave 9.9 Vis at $\Sigma(x_i - \bar{x})(y_i - \bar{y}) = \Sigma(x_i - \bar{x})y_i = \Sigma(y_i - \bar{y})x_i$ (sum fra i = 1 til n).

Løsningsforslag:

Til hjelp: $\Sigma(x_i - \bar{x}) = \Sigma x_i - \Sigma\bar{x} = \Sigma x_i - n\bar{x} = \Sigma x_i - n\cdot\frac{1}{n}\Sigma x_i = \Sigma x_i - \Sigma x_i = 0$.

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = \Sigma x_i(y_i - \bar{y}) - \bar{x}\underbrace{\Sigma(y_i - \bar{y})}_{=0} = \Sigma x_i(y_i - \bar{y})$, q.e.d.

$$\Sigma(x_i-\bar{x})(y_i-\bar{y})=\Sigma(x_i-\bar{x})y_i-\bar{y}\underbrace{\Sigma(x_i-\bar{x})}_{=0}=\Sigma(x_i-\bar{x})y_i,\ \underline{\text{q.e.d.}}$$

Oppgave 9.10 Vis at $\Sigma\left(\frac{1}{n}-\frac{\bar{x}}{S_x^2}(x_i-\bar{x})\right)^2=\frac{1}{n}+\frac{(\bar{x})^2}{S_x^2}$ (sum fra i = 1 n).

Løsningsforslag:

$$\Sigma\left(\tfrac{1}{n}-\tfrac{\bar{x}}{S_x^2}(x_i-\bar{x})\right)^2=\Sigma\left(\tfrac{1}{n^2}-\tfrac{2\bar{x}}{S_x^2}(x_i-\bar{x})+\tfrac{\bar{x}^2}{S_x^4}(x_i-\bar{x})^2\right)=\Sigma\tfrac{1}{n^2}-0+\tfrac{\bar{x}^2}{S_x^4}\underbrace{\Sigma(x_i-\bar{x})^2}_{=S_x^2}$$

$$=\tfrac{1}{n}+\tfrac{\bar{x}^2}{S_x^2},\ \underline{\text{q.e.d.}}$$

Oppgave 9.11 Vis at $\text{Cov}(a^*,b^*)=-\frac{\bar{x}}{S_x^2}\sigma^2$

Anta at vi hadde omdefinert vår regresjonsmodell som følger:

$y_r = a' + b'\cdot(x - \bar{x})$

Hva ville nå minste kvadraters-estimatorene a'^* og b'^* blitt, og hva ville Cov(a'^*,b'^*) blitt i dette tilfellet?

Løsningsforslag:

$$\text{Cov}(a^*,b^*)=\text{Cov}(\bar{Y}-b^*\bar{x},b^*)=\text{Cov}(\bar{Y},b^*)-\bar{x}\cdot\text{Var}b^*=\text{Cov}(\bar{Y},b^*)-\tfrac{\bar{x}^2}{S_x^2}\sigma^2$$

$$\text{Cov}(\bar{Y},b^*)=\text{Cov}(\bar{Y},\tfrac{\Sigma Y_i(x_i-\bar{x})}{S_x^2})=\tfrac{1}{S_x^2}\cdot\text{Cov}(\bar{Y},\Sigma Y_i(x_i-\bar{x}))$$

Siden Y_i og Y_j er ukorrelerte for $i \neq j$ får vi:

$$\text{Cov}(\bar{Y},\Sigma Y_i(x_i-\bar{x}))=\tfrac{1}{n}\cdot\Sigma\text{Cov}(Y_i,Y_i(x_i-\bar{x}))=\tfrac{\sigma^2}{n}\Sigma(x_i-\bar{x})=0$$

Følgelig: $\text{Cov}(\bar{Y},b^*)=0$ og $\text{Cov}(a^*,b^*)=-\frac{\bar{x}}{S_x^2}\sigma^2$, $\underline{\text{q.e.d.}}$

$$y_r=a+bx=a'+b'(x-\bar{x})=(a'-b'\bar{x})+b'x\Rightarrow b'=b,\ a'-b'\bar{x}=a,\ a'=a+b'\bar{x}=a+b\bar{x}$$

Følgelig: $(b')^*=b^*=\underline{S_{xy}/S_x^2}$, $(a')^*=a^*+b^*\bar{x}=\bar{y}-b^*\bar{x}+b^*\bar{x}=\underline{\bar{y}}$

$$\text{Cov}((a')^*,(b')^*)=\text{Cov}(\bar{y},b^*)=\underline{0}$$

Oppgave 9.12 Gitt residualene $E_i^* = Y_i^* - a^* - b^*\cdot x_i$, i = 1, 2,…,n

der $b^* = S_{xy}/S_x^2$ og $a^* = \bar{Y} - b^*\cdot\bar{x}$ i den lineære regresjonsmodellen $Y_i = a + b\cdot x_i + E_i$, i = 1, 2,…,n. $E_1,E_2,\ldots,E_n$ er antatt å være uif $N(0, \sigma)$-variable.

Vis at

$$\mathrm{Var}\left(E_i^*\right) = \sigma^2\left(1 - \frac{1}{n}\left(1 + \frac{\left(x_i - \bar{x}\right)^2}{S_x^2 / n}\right)\right), \quad \mathrm{Cov}\left(E_i^*, E_j^*\right) = -\frac{2}{n}\left(1 + \frac{\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{S_x^2 / n}\right)$$

Løsningsforslag:

$$E_i^* = Y_i - a^* - b^* x_i = (Y_i - \bar{Y}) - b^*(x_i - \bar{x})$$

$$\mathrm{Var}E_i^* = \mathrm{Var}(Y_i - \bar{Y}) + (x_i - \bar{x})^2 \,\mathrm{Var}b^* - 2\cdot\mathrm{Cov}(Y_i - \bar{Y}, b^*(x_i - \bar{x}))$$

$$\mathrm{Var}(Y_i - \bar{Y}) = \mathrm{Var}(Y_i - \tfrac{1}{n}Y_i - \tfrac{1}{n}\sum_{j\neq i} Y_j) = \mathrm{Var}(\tfrac{n-1}{n}Y_i - \tfrac{1}{n}\sum_{j\neq i} Y_j) = \left(\tfrac{n-1}{n}\right)^2\sigma^2 + \tfrac{n-1}{n^2}\sigma^2 = \tfrac{n-1}{n}\sigma^2$$

$$(x_i - \bar{x})^2\,\mathrm{Var}b^* = \tfrac{(x_i - \bar{x})^2}{S_x^2}\sigma^2$$

$$\mathrm{Cov}(Y_i - \bar{Y}, b^*(x_i - \bar{x})) = (x_i - \bar{x})\cdot(\mathrm{Cov}(Y_i, b^*) - \underbrace{\mathrm{Cov}(\bar{Y}, b^*)}_{=0,\text{ se } 9.11}) = (x_i - \bar{x})\cdot\mathrm{Cov}(Y_i, b^*)$$

$$\mathrm{Cov}(Y_i, b^*) = \mathrm{Cov}(Y_i, \tfrac{\Sigma Y_j (x_j - \bar{x})}{S_x^2}) = \tfrac{(x_i - \bar{x})}{S_x^2}\sigma^2$$

$$\Rightarrow \mathrm{Var}E_i^* = \tfrac{n-1}{n}\sigma^2 + \tfrac{(x_i - \bar{x})^2}{S_x^2}\sigma^2 - 2\tfrac{(x_i - \bar{x})^2}{S_x^2}\sigma^2 = \sigma^2\cdot\left(1 - \tfrac{1}{n}\left(1 + \tfrac{(x_i - \bar{x})^2}{S_x^2/n}\right)\right), \ \underline{\text{q.e.d.}}$$

$$\mathrm{Cov}(E_i^*, E_j^*) = \mathrm{Cov}((Y_i - \bar{Y}) - b^*(x_i - \bar{x}), (Y_j - \bar{Y}) - b^*(x_j - \bar{x})) = \mathrm{Cov}(Y_i - \bar{Y}, Y_j - \bar{Y})$$

$$-\mathrm{Cov}(b^*(x_i - \bar{x}), Y_j - \bar{Y}) - \mathrm{Cov}(Y_i - \bar{Y}, b^*(x_j - \bar{x})) + \mathrm{Cov}(b^*(x_i - \bar{x}), b^*(x_j - \bar{x}))$$

$$\mathrm{Cov}(Y_i - \bar{Y}, Y_j - \bar{Y}) = \mathrm{Cov}(Y_i, Y_j) - \mathrm{Cov}(Y_i, \bar{Y}) - \mathrm{Cov}(\bar{Y}, Y_j) + \mathrm{Cov}(\bar{Y}, \bar{Y}) =$$

$$0 - \tfrac{\sigma^2}{n} - \tfrac{\sigma^2}{n} + \tfrac{\sigma^2}{n} = \sigma^2 / n$$

$$\mathrm{Cov}(b^*(x_i - \bar{x}), Y_j - \bar{Y}) = (x_i - \bar{x})\cdot\mathrm{Cov}(\tfrac{\Sigma Y_i (x_i - \bar{x})}{S_x^2}, Y_j - \bar{Y})$$

$$= \tfrac{(x_i - \bar{x})(x_j - \bar{x})}{S_x^2}\sigma^2 - \tfrac{\sigma^2}{S_x^2}\cdot\tfrac{1}{n}\Sigma(x_i - \bar{x}) = \tfrac{(x_i - \bar{x})(x_j - \bar{x})}{S_x^2}\sigma^2$$

Tilsvarende: $\mathrm{Cov}(Y_i - \bar{Y}, b^*(x_j - \bar{x})) = \tfrac{(x_i - \bar{x})(x_j - \bar{x})}{S_x^2}\sigma^2$. Følgelig:

$$\mathrm{Cov}(E_i^*, E_j^*) = -\tfrac{\sigma^2}{n} - \tfrac{(x_i - \bar{x})(x_j - \bar{x})\sigma^2}{S_x^2} - \tfrac{(x_i - \bar{x})(x_j - \bar{x})\sigma^2}{S_x^2} + \tfrac{(x_i - \bar{x})(x_j - \bar{x})\sigma^2}{S_x^2}$$

$$= -\frac{\sigma^2}{n}\left(1 + \tfrac{(x_i - \bar{x})(x_j - \bar{x})}{S_x^2/n}\right), \ \underline{\text{q.e.d.}}$$

Oppgave 9.13 (E) Tabellen viser middeltemperaturen i Bergen de siste 30 år. x_i er år og Y_i er temperatur i °C.

x_i	Y_i	x_i	Y_i
1961	7,7	1976	7,7
1964	7,3	1979	7,2
1967	7,3	1982	7,7
1970	7,4	1985	7,4
1973	7,3	1988	7,9

Verdiene i tabellen er middeltemperaturen over 3-årsperioder. Temperaturen 7,4 °C for 1970 betyr middeltemperaturen for årene 1969-1970-1971. (Dette har ingen betydning for bestemmelse av regresjonslinjen). De 8 første temperaturene er målt på Frediksberg, de to siste på Florida. Vi velger å ignorere dette i denne sammenheng.

a) Bestem regresjonslinja

$y_r^*(x) = a^* + b^* \cdot x$

ved en lineær regresjon. Tegn spredningsdiagram og tegn linja inn i diagrammet. Det oppgis at

$$S_x^2 = \sum_{i=1}^{10} (x_i - \bar{x})^2 = 742{,}5 \qquad \sum_{i=1}^{10} (x_i - \bar{x}) y_i = 6{,}45$$

b) Er det grunnlag for, på bakgrunn av den gitte tabellen, å hevde at temperaturen i Bergen viser en stigende tendens?

Test H_0: $b = 0$ mot H_1: $b > 0$

Benytt et signifikansnivå på 5 %

Det oppgis at $$\text{SSE} = \sum_{i=1}^{10} \left(Y_i - a^* - b^* x_i\right)^2 = 0{,}4530$$

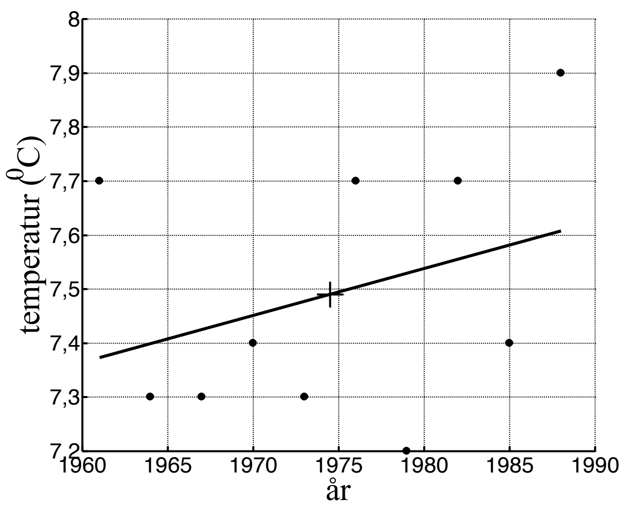
<u>Løsningsforslag</u>:

a) $\bar{x} = 1974{,}5, \bar{y} = 74{,}9, b^* = 6{,}45/742{,}5 = {,}00869, a^* = \bar{y} - b^* \cdot \bar{x} = -9{,}668$

Regresjonslinja blir følgelig $y_r^* = \underline{-9{,}668 + {,}00869x}$, se neste figur.

b) $T = \dfrac{b^* - b_0}{S / S_x}$, $R: T > t_{,05}(n-2) = t_{,05}(8) = 1{,}860$,

$$s = \sqrt{\text{SSE}/(n-2)} = \sqrt{{,}4530/8} = {,}24 \Rightarrow T = \frac{{,}00869 - 0}{{,}24/\sqrt{742{,}5}} = {,}99$$



Konklusjon: Ikke grunnlag på nivå 5 % å hevde signifikant temperaturøkning.

Oppgave 9.14 (E) I en fremstillingsprosess for syntetiske fibre til bruk i tekstiler, inngår krymping av fibrene ved koking i trykkoker. For å finne ut hvordan temperaturen innvirker på krympningsgraden, utføres krympning av fiberprøver ved 11 forskjellige temperaturer x_i (°C), $i = 1,2,\ldots,11$. Krympningsgraden Y_i (%) avleses hver gang. Resultatet ble:

x_i	120	122	124	126	128	130	132	134	136	138	140
Y_i	3,3	3,7	3,9	4,1	4,5	4,4	4,8	5,0	5,6	5,5	5,9

Vi antar en lineær regresjonsmodell er brukbar, dvs. at

$EY_i = a + bx_i$ (a, b ukjente parametre)

Vi antar dessuten at $Y_1,\ldots,Y_{11}$ er uavhengige og normalfordelte med varians σ^2.

a) Bestem den estimerte regresjonslinja $y_r^* = a^* + b^*x$. Tegn spredningsdiagram og tegn den estimerte regresjonslinja inn i diagrammet.

b) Test hypotesen H_0: $b = 0,15$ mot alternativet H_1: $b \neq 0,15$. Bruk 5 % signifikansnivå.

Regnehjelp:

$$\sum_{i=1}^{11}(x_i-\bar{x})^2 = 440,\quad \sum_{i=1}^{11}(x_i-\bar{x})Y_i = 54,8,\quad \text{SSE} = \sum_{i=1}^{11}\left(Y_i - a^* - b^*x_i\right)^2 = 0,165$$

Løsningsforslag:

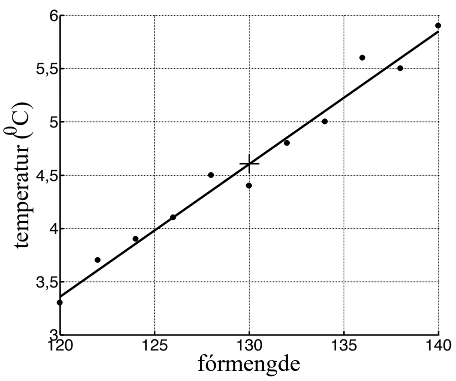
a) $\bar{x} = 130, \bar{y} = 4,609, b^* = S_{xy}/S_x^2 = 54,8/440 = ,1245, a^* = \bar{y} - b^*\cdot\bar{x} = -11,58$

Regresjonslinja blir: $y_r^* = -11,58 + ,1245x$

b) $s = \sqrt{\text{SSE}/(n-2)} = \sqrt{,165/9} = ,135$

$$|T| = \left|\frac{,1245 - ,15}{,135/\sqrt{440}}\right| = 3,88 > t_{,025}(9) = 1,83$$

Konklusjon: Grunnlag på nivå 5 % å hevde at $b \neq ,15$.



Oppgave 9.15 (E) En forskergruppe vil undersøke hvordan proteininnholdet i foret påvirker veksten av regnbueørret. Det ble utført forsøk med 10 grupper ørret, hvor hver gruppe fikk forskjellig andel protein, x_i (%), i foret. Gjennomsnittlig tilvekst Y_i i løpet av en gitt periode ble målt i hver gruppe. Resultatet ble:

x_i	18	22	26	30	34	38	42	46	50	54
Y_i	178	215	223	244	232	255	248	261	264	259

a) Bestem empirisk korr.koeffisient r

$Y_1, Y_2, \ldots, Y_{10}$ antas uavhengige og normalfordelte med samme ukjente varians σ^2 og forventningsverdier $E(Y_i) = a + bx_i$, i = 1,2,…,10.

b) Bestem den estimerte regresjonslinja $y_r^* = a^* + b^*x$ ved lineær regresjon.

c) Tegn spredningsdiagram og tegn regresjonslinja inn i diagrammet.

d) Finn et 95 % konfidensintervall for den sanne regresjonskoeffisienten b.

Regnehjelp:

$$S_x^2 = \sum_{i=1}^{10}(x_i - \bar{x})^2 = 1320, \qquad \sum_{i=1}^{10}(y_i - \bar{y})^2 = 6507{,}6$$

$$\sum_{i=1}^{11}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{11}(x_i - \bar{x})Y_i = 2608, \ \text{SSE} = \sum_{i=1}^{11}\left(Y_i - a^* - b^* x_i\right)^2 = 1354{,}8$$

Løsningsforslag:

a) $r = \dfrac{S_{xy}}{S_x S_y} = \dfrac{2608}{\sqrt{1320}\cdot\sqrt{6507{,}6}} = \underline{,89}$

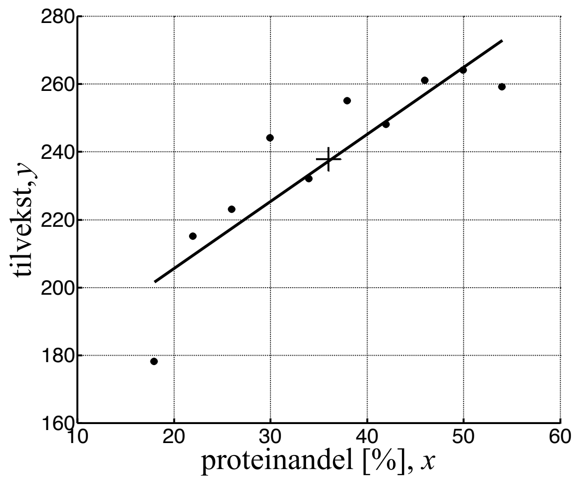
$\bar{x} = 36, \bar{y} = 237{,}9,$

b) $b^* = S_{xy} / S_x^2 = 2608/1320 = 1{,}98,$

$a^* = \bar{y} - b^* \cdot \bar{x} = 166$

Regresjonslinje: $y_r^* = \underline{166 + 1{,}98x}$.

c) Se figur.



d) 95 % KI for b: $b^* \mp t_{,025}(n-2)\cdot S / S_x$. Innsatt $b^* = 1{,}98$, $s = \sqrt{1354{,}8/8} = 13{,}0$, $t_{,025}(8) = 2{,}306$, $S_x = \sqrt{1320} = 36{,}3$ fås intervallet $\underline{(1{,}15\,,2{,}81)}$.

Oppgave 9.16 (E) Tabellen viser flygods (x_i) og tilsvarende fortjeneste (Y_i) for 10 flyselskap (data fra 1973).

Flyselskap	flygods (10^9 kg)	fortj. (mill kr)	Flyselskap	flygods (10^9 kg)	fortj. (mill kr)
Pan American	860	188	Seaboard	359	53
Flying Tiger	681	120	North West	246	52
United	645	135	Eastern	207	56
American	529	114	Delta	176	56
TWA	475	98	Continental	144	29

a) Tegn spredningsdiagram.

Anta at en lineær regresjonsmodell er brukbar, dvs. at $E(Y_i) = a + bx_i$. Finn den estimerte regresjonslinja $y_r^* = a^* + b^* x$.

b) Det blir hevdet at $b = 0,250$, men vi har mistanke om at b er mindre.

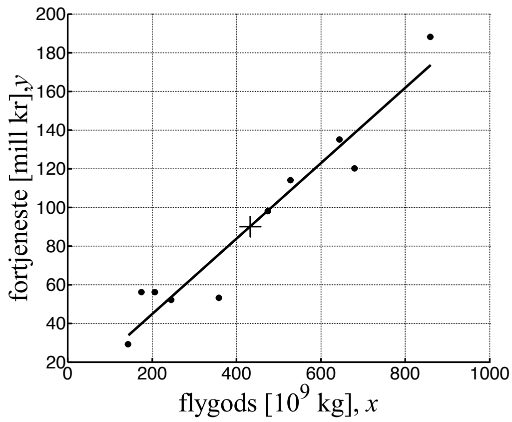
Test H_0: $b = 0,250$ mot H_1: $b < 0,250$ med 5 % signifikansnivå.

Regnehjelp:

$$\sum_{i=1}^{10}(x_i - \bar{x})y_i = 105362 \text{ , } \sum_{i=1}^{10}(x_i - \bar{x})^2 = 540842, \sum_{i=1}^{10}\left(y_i - a' - b^* x_i\right)^2 = 1489$$

Løsningsforslag:

a)



$$\bar{x} = 432,2, \bar{y} = 90,1, b^* = S_{xy} / S_x^2 = 105362 / 540842 = ,1948, a^* = \bar{y} - b^* \cdot \bar{x} = 5,907$$

Regresjonslinje: $\underline{y_r^* = 5,90 + ,195x}$.

$$s = \sqrt{\text{SSE} / (n-2)} = \sqrt{1489 / 8} = 13,64,$$

b) $T = -\dfrac{,195 - ,25}{13,64 / \sqrt{540842}} = -2,97$ Siden $T < -t_{,05}(8) = -1,86$ kan vi på nivå 5 % konkludere at $\underline{b < ,25}$

Oppgave 9.17 (E) Oljeinntektene Y_i fra en bestemt del av Nordsjøen har over en periode på 12 mndr. $(x_1, x_2, \ldots, x_{12})$ vært som følger (i mrd kr):

x_i	1	2	3	4	5	6	7	8	9	10	11	12
Y_i	2,0	1,9	1,8	2,1	2,0	1,9	2,2	2,3	2,1	2,2	2,3	2,4

a) Vis at den estimerte regresjonslinja ved en lineær regresjon er gitt ved

$y_r^* = 1{,}827 + 0{,}042 \cdot x$, når det opplyses at $\sum_{i=1}^{12}(x_i - \bar{x})^2 = 143 \quad \sum_{i=1}^{12}(x_i - \bar{x})Y_i = 6$

b) Lag spredningsdiagram og tegn inn den estimerte regresjonslinja.

c) La b være stigningstallet for den sanne regresjonslinja, slik at $EY_i = a + bx_i$. Lag et 95 % konfidensintervall for b. Oppgitt: $\sum_{i=1}^{12}(y_i - 1{,}827 - 0{,}042x_i)^2 = 0{,}128$

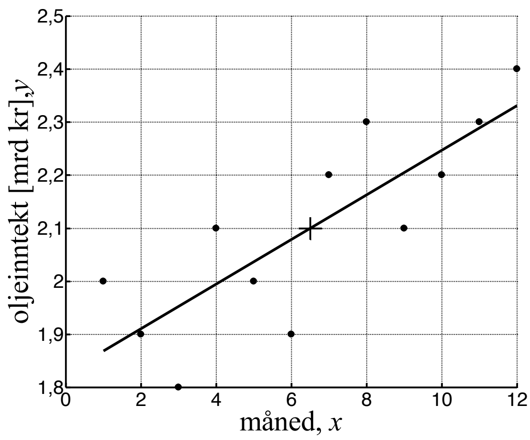
Løsningsforslag:

$\bar{x} = 6{,}50, \bar{y} = 2{,}10,$

a) $b^* = S_{xy} / S_x = 6/143 = {,}042,$

$a^* = \bar{y} - b^* \cdot \bar{x} = 1{,}827,$ q.e.d.

b) Se figur.



c) 95 % KI for b: $b^* \mp t_{,025}(n-2) \cdot S / S_x$.

Innsatt

$b^* = {,}042,\ s = \sqrt{{,}128/10} = {,}113,$

$t_{,025}(10) = 2{,}228,\ S_x = \sqrt{143} = 11{,}96$ fås intervallet $({,}021\ ,\ {,}063)$

Kapittel 10

Variansanalyse

Oppgave 10.1 En fabrikant av stålplater vil analysere stålstyrken til 3 forskjellige stålkvaliteter A, B og C, ett parti av hver type. Styrken måles i form av stålets flytegrense. Henholdsvis 100, 70 og 30 tilfeldige prøver tas fra A-, B- og C-partiet.

a) Hva er forsøksenhetene her?
b) Hva er forsøksfaktorene i forsøket?
c) Hva er faktornivåene?
d) Hvilke behandlinger har vi?
e) Hvilke populasjoner har vi?
f) Hvor mange replikater er det?
g) Hva er responsvariabel

Løsningsforslag:

a) Forsøksenheter: stålprøvene.

b) Forsøksfaktorer: stålkvalitet A, B og C.

c) Faktornivåer: ett nivå pr. forsøksfaktor.

d) 3 behandlinger, hver stålkvalitet gir én behandling.

e) Vi har 3 populasjoner, en pr. stålkvalitet. For en gitt stålkvalitet består populasjonen av alle de flytegrenser vi ville målt dersom alle prøvene i hele partiet ble undersøkt.

f) 100 replikater av A, 70 av B og 30 av C.

g) Responsvariabel: flytegrense.

Oppgave 10.2 Gitt y_{ij}-dataene

A	B	C
10	21	11
14	22	16
14	15	12
14	18	20
15		13

a) Beregn $\bar{y}_1, \bar{y}_2, \bar{y}_3$ og $\bar{y}$

b) Beregn residualene og sjekk at de summerer til null.

c) Beregn de nødvendige størrelser og sett opp ANOVA-tabell.

d) Test om det er signifikante forskjeller mellom noen av de tre

behandlings-middel-
verdiene på nivå 5 %.

e) Dersom testen i d) gir forkasting, bestem hvilke behandlinger som er signifikant forskjellige basert på 95 % konfidensintervall for forskjell på to og to behandlinger.

Løsningsforslag:

a) $\bar{y}_1 = (10+\cdots+5)/5 = \underline{13{,}4}, \bar{y}_2 = (21+\cdots+18)/4 = \underline{19{,}0}, \bar{y}_3 = (11+\cdots+13)/5 = \underline{14{,}4}$

$\bar{y} = (n_1\bar{y}_1 + n_2\bar{y}_2 + n_3\bar{y}_3)/n = (5\cdot 13{,}4 + 4\cdot 19{,}0 + 5\cdot 14{,}4)/(5+4+5) = \underline{15{,}36}$

b) $y_{ij} - \bar{y}_j$: $10-13{,}4,\ldots,15-13{,}4, 21-19{,}0,\ldots,18-19{,}0, 11-14{,}4,\ldots,13-14{,}4$

Summeres alle residualene får vi null, generelt får vi et resultat i nærheten av null grunnet avrundingsfeil.

c) Behandlings-kvadratsum: $SS_T = \Sigma n_j(\bar{y}_j - \bar{y})^2 =$

$5\cdot(13{,}4-15{,}36)^2 + 4\cdot(19{,}0-15{,}36)^2 + 5\cdot(14{,}4-15{,}36)^2 = \underline{76{,}81}$

Total kvadratsum: $S_y^2 = \sum_{i,j}(y_{ij} - \bar{y})^2 = (10-15{,}36)^2 + \cdots (13-15{,}36)^2 = \underline{173{,}99}$

Residualsum: $SSE = S_y^2 - SS_T = 173{,}99 - 76{,}81 = \underline{97{,}18}$

ANOVA-tabell:

Kilde	Kvadratsum	d.f.	Kvadratmiddel	F
Behandling	$SS_T = 76{,}8$	$k-1=2$	$MS_T = 76{,}8/2 = 38{,}4$	$\frac{SS_T/(k-1)}{SSE/(n-k)}$
Feil	$SSE = 97{,}2$	$n-k=11$	$MSE = 97{,}2/11 = 8{,}84$	
Total	$S_y^2 = 174{,}0$			$= 38{,}4/8{,}84$ $= 4{,}34$

d) $F = 4{,}34 > F_{,05}(2,11) = 3{,}982$. Konklusjon: Grunnlag på nivå 5 % å hevde at det er behandlingsforskjell.

e) Konstruerer 95 % KI for A–B, A–C og B–C.

Hjelpestørrelser: $s = \sqrt{SSE/11} = 2{,}97$, $t = t_{,025}(11) = 2{,}201$.

A–B: $\bar{y}_1 - \bar{y}_2 \mp t\cdot s\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = 13{,}4 - 19{,}0 \mp 2{,}201\cdot 2{,}97\cdot\sqrt{\frac{1}{5}+\frac{1}{4}} = \underline{(-9{,}99\,,\,-1{,}21)}$

A–C: $\bar{y}_1 - \bar{y}_3 \mp t\cdot s\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_3}} = 13{,}4 - 14{,}4 \mp 2{,}201\cdot 2{,}97\cdot\sqrt{\frac{1}{5}+\frac{1}{5}} = \underline{(-5{,}13\,,\,3{,}13)}$

B–C: $\bar{y}_2 - \bar{y}_3 \mp t\cdot s\cdot\sqrt{\frac{1}{n_2}+\frac{1}{n_3}} = 19{,}0 - 14{,}4 \mp 2{,}201\cdot 2{,}97\cdot\sqrt{\frac{1}{4}+\frac{1}{5}} = \underline{(0{,}21\,,\,8{,}99)}$

Konklusjon: B signifikant best.

Oppgave 10.3 Gitt følgende svært forenklede ANOVA-tabell:

$SS_T = 20 \quad k\text{-}1 = 4$
$SSE = 100 \quad n\text{-}k = 40$

a) Hvor mange behandlinger k og observasjoner n har vi?

b) Hvilken F-fordeling er aktuell her?

c) Vis at det på signifikansnivå 5 % ikke er grunnlag for å hevde at det er forskjellig effekt av behandlingene.

d) Betyr konklusjonen i c) at det ikke er forskjellig effekt av behandlingene?

Løsningsforslag:

a) $k = 4 + 1 = \underline{5}$ behandlinger, $n = 40 + k = \underline{45}$ observasjoner.

b) $F(k-1, n-k) = \underline{F(4{,}40)\text{ - fordelingen}}$ er aktuell her.

c) $F = \dfrac{SS_T/(k-1)}{SSE/(n-k)} = \dfrac{20/4}{100/40} = 2 < F_{,05}(4{,}60) = 2{,}525 < F_{,05}(4{,}40)$

Konklusjon: Ikke grunnlag for å hevde forskjellig behandlingseffekt.

d) Nei, selv om det er forskjellig behandlingseffekt behøver ikke nullhypotesen om ingen effekt bli forkastet.

Oppgave 10.4 Ytelsen til 3 forskjellige typer PC'er med samme prosessortype (Pentium 200 Pro), skal sammenlignes ved å måle CPU-tiden på en bestemt regne-operasjon. Tilfeldige PC'er fra hver av de 3 leverandørene plukkes ut og CPU-tiden måles. Resultatet i målte sekunder er gitt nedenfor:

$\bar{y}_1 = 12{,}11,\ \bar{y}_2 = 11{,}87,\ \bar{y}_3 = 12{,}02$
$n_1 = 5,\ n_2 = 4,\ n_3 = 9,\ SSE = 0{,}0058$

a) Beregn de nødvendige størrelser og lag ANOVA-tabell.

b) Test på signifikansnivå 5 % om det er grunnlag for å konkludere at det er forskjell på noen av de tre PC- typene. Finn eventuelt hvilken PC som du kan konkludere er «signifikant best».

c) Ville ytelsen være avgjørende for hvilken av de 3 PC-typene du ville velge?

Løsningsforslag:

a) $\bar{y} = (n_1\bar{y}_1 + n_2\bar{y}_2 + n_3\bar{y}_3)/n = (5\cdot 12{,}11 + 4\cdot 11{,}87 + 9\cdot 12{,}02)/(5+4+9) = \underline{12{,}01}$

Behandlings-kvadratsum: $SS_T = \Sigma_j\, n_j(\bar{y}_j - \bar{y})^2 =$

$5\cdot(12{,}11-12{,}01)^2 + 4\cdot(11{,}87-12{,}01)^2 + 9\cdot(12{,}02-12{,}01)^2 = \underline{,129}$

ANOVA-tabell:

Kilde	Kvadratsum	d.f.	Kvadratmiddel	F
Behandling	$SS_T = ,129$	$k - 1 = 2$	$MS_T = ,129/2 = ,0645$	$\dfrac{SS_T / (k-1)}{SSE / (n-k)}$
Feil	$SSE = ,0058$	$n - k = 15$	$MSE = ,0058/15 = 3{,}9 \cdot 10^{-4}$	$= ,0645/3{,}9 \cdot 10^{-4}$
Total	$S_y^2 = ,1348$			$= 165{,}4$

b) $F = 165{,}4 >> F_{,05}(2{,}15) = 3{,}682$. Konklusjon: Signifikant PC-forskjell.

Hvis én PC er signifikant best må det være nr. 2. Beregner 95 % KI for PC2−PC1 og PC2−PC3 for å undersøke dette.

Hjelpestørrelser: $s = \sqrt{SSE / (n-k)} = \sqrt{,0058 / 15} = ,0197$, $t = t_{,025}(15) = 2{,}131$.

PC2−PC1: $\bar{y}_2 - \bar{y}_1 \mp t \cdot s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (-,268\,,\, -,212)$

PC2−PC3: $\bar{y}_2 - \bar{y}_3 \mp t \cdot s \cdot \sqrt{\frac{1}{n_2} + \frac{1}{n_3}} = (-,173\,,\, -,127)$

Konklusjon: PC nr. 2 har signifikant større ytelse enn de andre.

d) Ytelsen er neppe avgjørende for valg av PC i dette tilfellet. Selv om PC nr. 2 har bedre ytelse, er forskjellen så liten fra de andre at det trolig er andre egenskaper ved PC'en som ville være mer avgjørende for valget.

Oppgave 10.5 4 biltyper skal sammenlignes for å se om det er signifikante forskjeller i bensinforbruk. Tilfeldige biler plukkes ut, og de kjører nøyaktig den samme distansen i samme hastighet. Resultatene i form av liter pr. mil er:

A	B	C
0,436	0,591	0,480
0,421	0,585	0,529
0,364	0,645	0,528
0,301	0,590	0,530
0,407		0,514
0,359		

a) Beregn de nødvendige størrelser og sett opp ANOVA-tabell.

b) Test om det er signifikant forskjell på midlere bensinforbruk for de forskjellige biltypene på nivå 5 %.

c) Tidligere tester har gitt sterke indikasjoner på at A bruker minst bensin av de fire. Gir dataene grunnlag for å konkludere med dette på nivå 5 %?

Løsningsforslag:

a) $n_1 = 6, \bar{y}_1 = ,381, n_2 = 4, \bar{y}_2 = ,603, n_3 = 5, \bar{y}_3 = ,516, n_4 = 6, \bar{y}_4 = ,507$
$\bar{y} = \Sigma n_j \bar{y}_j / n = 10{,}32/21 = ,4914$
Behandlingskvadratsum: $SS_T = \Sigma n_j (\bar{y}_j - \bar{y})^2 = ,127$
Total kvadratsum: $S_y^2 = \Sigma_{i,j}(y_{ij} - \bar{y})^2 = (,436 - ,4914)^2 + \cdots + (,502 - ,4914)^2 = ,157$
Residualsum: $SSE = S_y^2 - SS_T = ,157 - ,127 = ,030$

Kilde	Kvadratsum	d.f.	Kvadratmiddel	F
Behandling	SS_T = ,127	$k-1=3$	MS_T = ,127/3 = ,0423	$\dfrac{SS_T/(k-1)}{SSE/(n-k)}$ =,0423/,0018 = 23,5
Feil	SSE = ,030	$n-k=17$	MSE =,030/17 = ,0018	
Total	S_y^2 =,157			

b) $F = 23{,}5 > F_{,05}(3{,}17) = 3{,}197$. Konklusjon: Signifikant forskjell på bensinbruk.

c) Konstruerer 95 % KI for A–B, A–C og A–D.
Hjelpestørrelser: $s = \sqrt{SSE/(n-k)} = \sqrt{,030/17} = ,042,\ t = t_{,025}(17) = 2{,}110$

A–B: $\bar{y}_1 - \bar{y}_2 \mp t \cdot s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = ,381 - ,603 \mp 2{,}110 \cdot ,042 \cdot \sqrt{\frac{1}{6} + \frac{1}{4}} = (-,279\ ,\ -,165)$
A–C: $\bar{y}_1 - \bar{y}_3 \mp t \cdot s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_3}} = ,381 - ,516 \mp 2{,}110 \cdot ,042 \cdot \sqrt{\frac{1}{6} + \frac{1}{5}} = (-,189\ ,\ -,081)$
A–D: $\bar{y}_1 - \bar{y}_4 \mp t \cdot s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_4}} = ,381 - ,507 \mp 2{,}110 \cdot ,042 \cdot \sqrt{\frac{1}{6} + \frac{1}{6}} = (-,177\ ,\ -,075)$

Konklusjon: Grunnlag på nivå 5 % å konkluderre at A bruker minst bensin av de fire. Siden det på forhånd var sterke indikasjoner på at A brukte minst, kunne vi her nøyd oss med 90 % konfidensintervaller, så konklusjonen her har faktisk nivå 2,5 %.

Oppgave 10.6 En lege (Ola) påstår at det ikke er noen påviselig forskjell mellom 3 ulike sovepiller A, B og C. En annen lege (Kari) påstår imidlertid hårdnakket at type A er best. Et tilfeldig utvalg av villige pasienter med søvnproblemer får i løpet av tre ulike perioder tildelt de 3 forskjellige sovepillene. De vet ikke hvilket middel de får fra gang til gang, og rekkefølgen trekkes vilkårlig. Gjennomsnittlig antall søvntimer pr. natt blir registrert, og gir følgende resultat (subskript 1,2 og 3 angir henholdsvis A, B og C):

$\bar{y}_1 = 7{,}34,\ \ \bar{y}_2 = 6{,}92,\ \ \bar{y}_3 = 7{,}21$
$n_1 = 41,\ \ n_2 = 28,\ \ n_3 = 34,\ \ SSE = 10{,}1$

a) Hvordan ville du analysert dataene for å verifisere ditt syn dersom du var Ola? Og dersom du var Kari?

b) Gir data sterkt grunnlag for å stole mer på den ene av de to legene? I såfall hvem og hvorfor?

Løsningsforslag:

a) Hvis jeg var Ola ville jeg nøyet meg med en simultan test for å verifisere at denne ikke ville påvise noen signifikant forskjell mellom A, B og C. Dersom jeg var Kari ville jeg konstruere konfidensintervaller for A-B og A-C for å påvise at begge kun ville inneholde positive intervallgrenser.

b) $\bar{y} = \Sigma n_j \bar{y}_j / n = 7{,}18$, Behandlingskvadratsum: $\mathrm{SS_T} = \Sigma_j n_j (\bar{y}_j - \bar{y})^2 = 2{,}97$

$$F = \frac{\mathrm{SS_T}/(k-2)}{\mathrm{SSE}/(n-k)} = \frac{2{,}97/2}{10{,}1/100} = 14{,}7 > F_{,05}(2{,}101) = 3{,}087$$

Følgelig: På nivå 5 % kan det konkluderes med forskjell, og jeg ville ikke stolt på Ola.

Vi konstruerer så 95 % KI for A-B og A-C.

Hjelpestørrelser: $s = \sqrt{\mathrm{SSE}/(n-k)} = \sqrt{10{,}1/100} = {,}318$, $t = t_{,025}(100) = 1{,}984$

A–B: $\bar{y}_1 - \bar{y}_2 \mp t \cdot s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 7{,}34 - 6{,}92 \mp 1{,}984 \cdot {,}318 \cdot \sqrt{\frac{1}{41} + \frac{1}{28}} = ({,}265\ ,\ {,}575)$

A–C: $\bar{y}_1 - \bar{y}_3 \mp t \cdot s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_3}} = 7{,}34 - 7{,}21 \mp 1{,}984 \cdot {,}318 \cdot \sqrt{\frac{1}{6} + \frac{1}{5}} = (-{,}016\ ,\ {,}276)$

Konklusjon: Vi ser at A er signifikant bedre enn B, mens A ikke er signifikant bedre enn C, selv om det er nære på. Jeg ville derfor heller ikke stolt helt på Kari, selv om jeg synes dataene i større grader støtter hennes oppfatninger enn Ola sine.

Kapittel 11

Monte Carlo simulering

Mange av oppgavene henviser til følgende slumptall generert med $U[0,1]$-generator:

R_1 = 0,8617
R_2 = 0,3841
R_3 = 0,5595
R_4 = 0,1789

Oppgave 11.1 La R være en U[0,1]-variabel. Hva blir oppsettet for å generere en tilfeldig $U[-3,10]$-variabel X? Hva blir X- verdien på basis av R_1?

Løsningsforslag:

$X = -3 + (10 - (-3))R = \underline{-3 + 13R}$. $\quad R_1 = ,8617 \Rightarrow X = -3 + 13 \cdot ,8617 = \underline{8,202}$

Oppgave 11.2 En eksponensialfordelt variabel X har forventning E(X) = 0,1. Bestem på enkleste måte 2 tilfeldige X-verdier på basis av R_1 og R_2.

Løsningsforslag:

$X = -,1 \cdot \ln(R),\ X_1 = -,1 \cdot \ln(,8617) = \underline{,01488},\ X_2 = -,1 \cdot \ln(,3841) = \underline{,09569}$

Oppgave 11.3 Bestem på grunnlag av R_1 til R_4 fire tilfeldige binomiske variabel-verdier X_1 - X_4 fra binomisk fordeling med parametre n = 12 og p = 0,3.

Løsningsforslag: Tabellen nedenfor viser kumulativ verdi $F(x)$ for Bino(12, ,3).

x	0	1	2	3	4	5	6	7	8	9	10	11	12
F	,014	,085	,253	,493	,724	,882	,951	,991	,998	1,00	1,00	1,00	1,00

Finner den minste F-verdi som er større enn R, og leser av tilsvarende x-verdi.

$X_1 = \underline{5},\ X_2 = \underline{3},\ X_3 = \underline{4},\ X_4 = \underline{2}$

Oppgave 11.4 Bestem på grunnlag av R_1 - R_4 fire tilfeldige Poissonvariable fra Poisson- fordelingen med parameter λ = 3.

Løsningsforslag: Tabellen nedenfor viser kumulativ verdi $F(x)$ for Po(3).

x	0	1	2	3	4	5	6	7	8	9	10	11	12
F	,050	,199	,423	,647	,815	,916	,967	,988	,996	,999	1,00	1,00	1,00

Finner den minste F-verdi som er større enn R, og leser av tilsvarende x-verdi.
$X_1 = \underline{5},\ X_2 = \underline{2},\ X_3 = \underline{3},\ X_4 = \underline{1}$

Oppgave 11.5 Bestem k i tetthetsfunksjonen $f(x) = k \cdot e^{-5x}$, $x \geq 0$. Generer deretter 2 tilfeldige variabelverdier fra fordelinga på enkleste måte basert på R_1 og R_2.

Løsningsforslag:

$$\int_0^\infty ke^{-5x}dx = \left[-\frac{k}{5}e^{-5x}\right]_0^\infty = -0 + k/5 = k/5 = 1 \Rightarrow k = \underline{5}$$

Kumulativ fordeling: $F(x) = \int_0^x 5e^{-5t}dt = \left[-e^{-5t}\right]_0^x = 1 - e^{-5x}$

Invers kumulativ fordeling: $F = 1 - e^{-5x} \Rightarrow x = -\frac{1}{5} \cdot \ln(1-F)$

Siden F er U(0,1)-fordelt, så vil også $1-F$ være U(0,1)-fordelt. Følgelig vil $X_{rand} = -\frac{1}{5} \cdot \ln(R)$ være en tilfeldig variabel fra $f(x)$ når R er U(0,1)-fordelt.
$\Rightarrow X_1 = -\frac{1}{5} \cdot \ln(,8617) = \underline{,0298},\ X_2 = -\frac{1}{5} \cdot \ln(,3841) = \underline{,1914}$

Oppgave 11.6 Bestem syklus og sykluslengde til generatoren $x_{n+1} = 3 \cdot x_n \pmod{10}$ med $x_0 = 4$.

$$x_1 = 3 \cdot x_0 \ (\mathrm{mod}\,10) = 3 \cdot 4 \ (\mathrm{mod}\,10) = 12 \ (\mathrm{mod}\,10) = 2$$
$$x_2 = 3 \cdot x_1 \ (\mathrm{mod}\,10) = 3 \cdot 2 \ (\mathrm{mod}\,10) = 6 \ (\mathrm{mod}\,10) = 6$$
$$x_3 = 3 \cdot x_2 \ (\mathrm{mod}\,10) = 3 \cdot 6 \ (\mathrm{mod}\,10) = 18 \ (\mathrm{mod}\,10) = 8$$
$$x_4 = 3 \cdot x_3 \ (\mathrm{mod}\,10) = 3 \cdot 8 \ (\mathrm{mod}\,10) = 24 \ (\mathrm{mod}\,10) = 4$$

Siden vi startet med verdien 4 og har fått verdien 4 igjen vil syklusen gjenta seg. Syklusen er følgelig: $\underline{4,2,6,8}$ med sykluslengde på $\underline{4}$.

Oppgave 11.7 En frekvenstest for en $U[0,1]$ generator gav følgende resultat (O_i = antall tall mellom $(i-1) \cdot 0{,}2$ og $i \cdot 0{,}2$, $i = 1, 2, 3, 4$ og 5): $O_1 = 14$, $O_2 = 27$, $O_3 = 19$, $O_4 = 30$, $O_5 = 10$. Bestem verdi til testobservatoren og test på nivå 5% ved hjelp av Pearson's kjikvadrat tilpasningstest.

Løsningsforslag:

Vi har totalt $n = 100$ observasjoner og 5 like brede intervaller, hvilket for en idéell $U[0,1]$ generator gir $E_i = 20, i = 1, \ldots, 5$. Vi får da:

$$T = \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} = \frac{(14-20)^2}{20} + \cdots + \frac{(10-20)^2}{20} = \frac{286}{20} = \underline{14{,}3} > \chi^2_\alpha(5-1) = 9{,}49$$

Konklusjon: Grunnlag på nivå 5 % å konkludere at generatoren ikke er perfekt.

Oppgave 11.8 Generer to tilfeldige $N(-1{,}2)$- variable på basis av R_1 og R_2 og på basis av tilnærmelse til Φ^{-1}.

Løsningsforslag: Henter konstantene c_0, c_1, c_2, d_1, d_2 og d_3 fra læreboka.

$$R = R_1 = ,8617 > ,5 \Rightarrow t = \sqrt{-2\ln(1-R_1)} = 1{,}9891, Z = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} = 1{,}0880$$

$$X_1 = -1 + 2 \cdot Z = -1 + 2 \cdot 1{,}0880 = \underline{1{,}1760}$$

$R = R_2 = ,3841 < ,5$. Vi erstatter da R med $1-R$ i uttrykket for t, beregner Z som over, men skifter fortegn til Z før vi beregner X. Vi får:

$$t = \sqrt{-2 \cdot \ln(1-(1-R))} = \sqrt{-2 \cdot \ln(R)} = \sqrt{-2 \cdot \ln(,3841)} = 1{,}3834$$

$$Z = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} = ,2943, \quad X_2 = -1 + 2 \cdot (-Z) = -1 - 2 \cdot ,2943 = \underline{-1{,}5886}$$

Oppgave 11.9 Generer to tilfeldige $N(3,\ 0{,}1)$- variable på basis av R_1 og R_2 og effektiv metode.

Løsningsforslag:

$$Z_1 = \sqrt{-2 \cdot \ln(R_1)} \cdot \cos(2\pi R_2) = \sqrt{-2 \cdot \ln(,8617)} \cdot \cos(2\pi \cdot ,3841) = -,4072$$

$$Z_2 = \sqrt{-2 \cdot \ln(R_1)} \cdot \sin(2\pi R_2) = \sqrt{-2 \cdot \ln(,8617)} \cdot \sin(2\pi \cdot ,8341) = ,3631$$

$$X_1 = 3 + ,1 \cdot Z_1 = 3 + ,1 \cdot (-,4072) = \underline{2{,}959},\ X_2 = 3 + ,1 \cdot Z_2 = 3 + ,1 \cdot (,3631) = \underline{3{,}036}$$

Oppgave 11.10 Generer et tilfeldig variabelpar fra binormal fordeling $N_2(0{,}1;0{,}1;-0{,}5)$ med basis i R_1 og R_2.

Løsningsforslag:

$$X = \sqrt{-2 \cdot \ln(R_1)} \cdot \sin(2\pi R_2) = \sqrt{-2 \cdot \ln(,8617)} \cdot \sin(2\pi \cdot ,3841) = \underline{,3631}$$

$$Y = \sqrt{-2 \cdot \ln(R_1)} \cdot \left(\sqrt{1-\rho^2} \cdot \cos(2\pi R_2) + \rho \cdot \sin(2\pi R_2)\right) =$$

$$\sqrt{-2 \cdot \ln(,8617)} \cdot \left(\sqrt{1-(-,5)^2} \cdot \cos(2\pi \cdot ,3841) - ,5 \cdot \sin(2\pi \cdot ,3841)\right) = \underline{-,5342}$$

Kapittel 12

Shewart-diagrammer

Oppgave 12.1 Temperaturen X i 3 forskjellige lagertanker med væske som er regulert til å holde en temperatur på 9,5 °C, måles ved 38 forskjellige tidspunkt. Ett og samme reguleringssystem regulerer temperaturen i de tre tankene. Gjennomsnittet av alle målingene viste en temperatur på 9,47 °C. Midlere variasjonsbredde var 0,60 °C.

a) Bestem styregrensene for et $\overline{X}$-diagram.

b) Bestem styregrensene for et R-diagram.

c) Ved et senere tidspunkt ble gjennomsnittstemperaturen til 3 forskjellige tanker funnet å være 10,2 °C mens variasjonsbredden var 0,9 °C. Gir dette grunn til alarm?

<u>Løsningsforslag</u>:

a) $n = 3$, $\overline{\overline{X}} = 9{,}47$ °C, $\overline{R} = 0{,}60$ °C, og fra tabell (med $n = 3$) finner vi at $A_2 = 1{,}023$.

$$LCL_{\overline{X}} = \overline{\overline{X}} - A_2 \cdot \overline{R} = 9{,}47 - 1{,}023 \cdot 0{,}60 = 8{,}86, \text{ dvs. } \underline{8{,}86\ ^\circ\text{C}}$$

$$UCL_{\overline{X}} = \overline{\overline{X}} - A_2 \cdot \overline{R} = 9{,}47 + 1{,}023 \cdot 0{,}60 = 10{,}04, \text{ dvs. } \underline{10{,}04\ ^\circ\text{C}}$$

b) Fra tabell finner vi at $D_4 = 2{,}574$ for $n = 3$.

$$LCL_R = \underline{0},\ UCL_R = D_4 \cdot \overline{R} = 2{,}574 \cdot {,}60 = 1{,}54, \text{ dvs. } \underline{1{,}54\ ^\circ\text{C}}$$

c) Siden 10,2 °C er større enn øvre kontrollgrense (10,04 °C) i $\overline{X}$-diagrammet er det <u>grunn</u> til alarm.

Oppgave 12.2 En fabrikk produserer blant annet elektriske motstander på 10 Ω. For å undersøke om produksjonsprosessen er stabil, måles 5 og 5 motstander med jevne tidsmellomrom. Resultatene av målinger (i Ω) ved 20 forskjellige tidspunkt er gjengitt nedenfor.

tid	X [Ω]	$\overline{X}$	R

1	9	9	5	9	4	7,2	5
2	9	9	9	6	6	7,8	3
3	10	9	16	14	17	11,6	7
4	10	9	12	5	15	10,2	10
5	9	9	6	9	6	7,8	3
6	9	9	10	4	4	7,2	6
7	10	9	8	20	11	11,6	12
8	9	9	14	6	7	9,0	8
9	9	9	8	9	10	9,0	2
10	10	10	12	7	24	12,6	17
11	10	10	19	5	18	12,4	14
12	9	9	10	7	7	8,4	3
13	10	9	10	5	6	8,0	5
14	10	9	8	14	10	10,2	6
15	10	9	5	4	7	7,0	6
16	10	9	11	8	4	8,4	7
17	9	9	16	17	9	12,0	8
18	9	9	10	6	7	8,2	4
19	9	9	15	7	11	10,2	8
20	10	9	16	10	8	10,6	8

a) En av X-verdiene ved tidspunkt 3 er feil. Finn denne og angi rett verdi.

b) Angi styregrensene og konstruer et $\overline{X}$-R diagram.

c) Gir diagrammet i b) grunnlag for å undersøke prosessen nærmere?

Løsningsforslag:

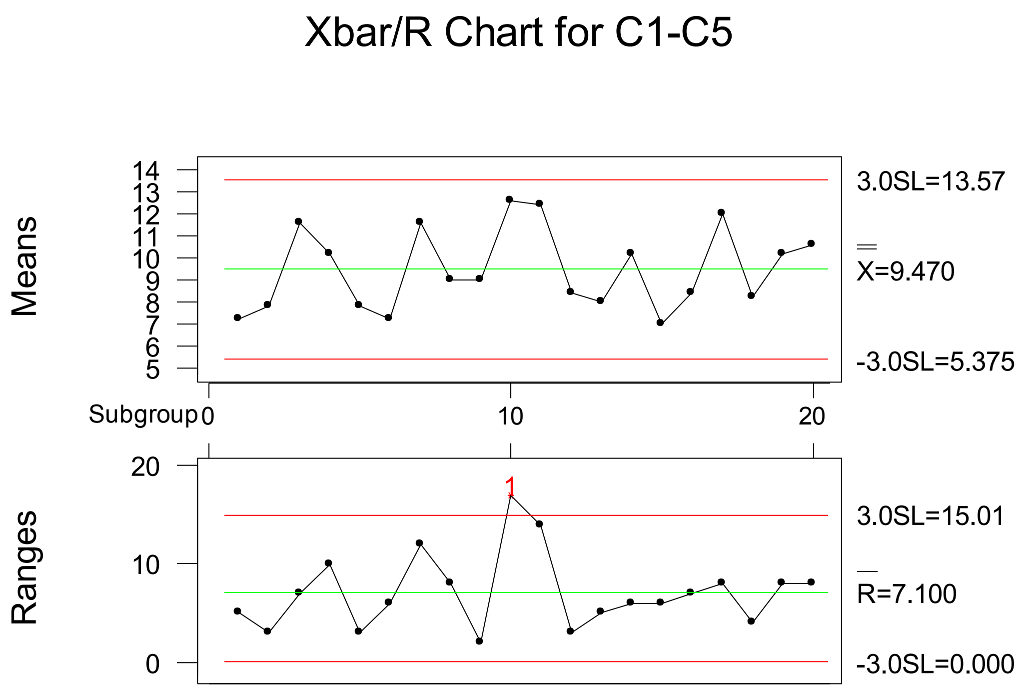
a) Ved tidspunkt 3 skal siste verdi være 9, ikke 17.

b) Fra tabellen ovenfor finner vi at $n = 5$, $\overline{\overline{x}} = 9{,}47\ \Omega$, $\overline{R} = 7{,}1\ \Omega$ som gir

$$LCL_{\overline{X}} = \overline{\overline{X}} - A_2 \cdot \overline{R} = 9{,}47\,\Omega - {,}577 \cdot 7{,}1\,\Omega = \underline{5{,}37\ \Omega}$$
$$UCL_{\overline{X}} = 9{,}47\,\Omega + {,}577 \cdot 7{,}1\,\Omega = \underline{13{,}57\,\Omega}$$

Minitabutskriften av $\overline{X}$-R diagrammet fremkommer ved først å legge $20 \cdot 5$ matrisa med data i kolonne c1-c5. Deretter benyttes kommandoene:

```
MTB> %rxbarr;
SUBC> rsub c1-c5;
SUBC> rbar;
SUBC> test 1.
```



c) Med basis i *R*-diagrammet (én alarm merket med 1) ser vi at det er grunn til å studere prosessen nærmere.

Oppgave 12.3 En bedrift produserer en vare der hver enhet skal veie 1 kg. Det har vist seg at den lokale variasjonen ved et gitt tidspunkt er neglisjerbar, men at nivået kan variere fra dag til dag. For å overvåke prosessen, måler man vekten av en vareenhet pr. dag over en periode på 20 dager, og får følgende avvik fra et kg (angitt i gram, suksessive tidspunkt fra venstre mot høyre):

13 0 −17 3 18 −1 −1 −12 23 16
−16 7 −3 1 9 3 7 16 −12 4

a) Konstruer et *XmR*-diagram.

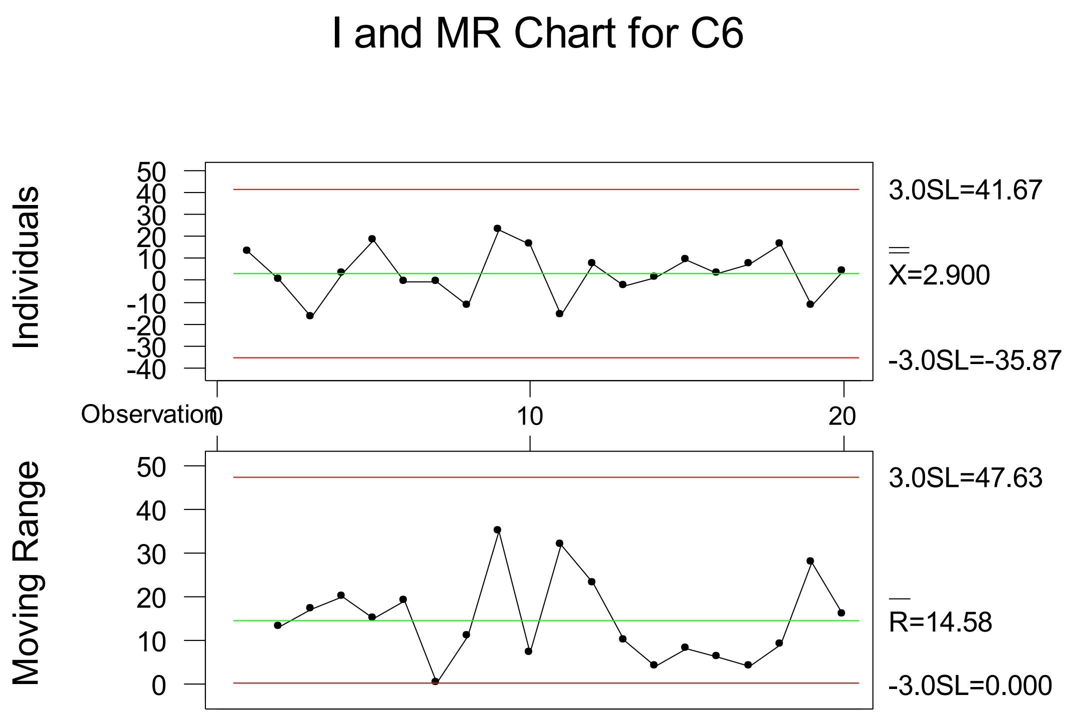
b) På dag 21 måles en ny verdi på 1,040 kg. Gir denne verdien grunnlag for alarm?

Løsningsforslag:

a) Vi legger dataene inn i kolonne c6 i Minitab og benytter kommandoene

```
MTB> %imrchart c6;
SUBC> rspan 2;
SUBC> test 1.
```

Dette gir følgende utskrift av *XmR*-diagrammet:



La oss se om grensene stemmer. Fra dataene finner vi at $\bar{x} = 2{,}9$ g og $\overline{mR} = 14{,}58$ g. Vi får da:

$LNPL_X = \overline{X} - 2{,}660 \cdot \overline{mR} = 2{,}9\,g - 2{,}660 \cdot 14{,}58\,g = -35{,}9\,g$, OK!

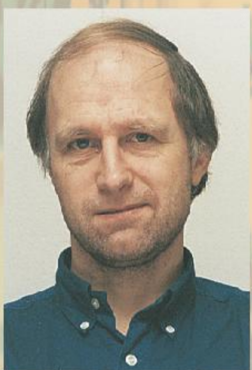
$UNPL_X = \overline{X} + 2{,}660 \cdot \overline{mR} = 2{,}9\,g + 2{,}660 \cdot 14{,}58\,g = 41{,}68\,g$, trolig avrundingsavvik.

$UCL_R = 3{,}268 \cdot \overline{mR} = 3{,}268 \cdot 14{,}58\,g = 47{,}65\,g$, trolig avrundingsavvik.

c) måleverdi på 1,040 kg på dag 21 tilsvarer en x-verdi på 40 g (avvik fra 1 kg) og en R-verdi på 40 g – 4 g = 36 g. Ingen av verdiene faller utenfor styregrensene, det er derfor ikke grunn til alarm.

Boka inneholder oppgavetekst og komplette løsningsforslag til alle oppgavene i læreboka *Statistikk og sannsynlighetsregning* av samme forfatter. Læreboka er spesielt tilpasset Ingeniørutdanningsrådets rammeplan for grunnlagsfaget statistikk (2 vekttall) ved høgskolenes ingeniørutdanninger. Av over 200 oppgaver i boka er det ca. 80 eksamensoppgaver fra tidligere statistikkeksamener for ingeniørstudenter ved forskjellige høgskoler i Norge.

Fordi oppgavene er av generell og anvendt karakter på nivå med vanlige innføringskurs i statistikk, forventes oppgaveboka å ha interesse langt utover ingeniørstudentene. Oppgavene dekker blant annet emnene beskrivende statistikk, sannsynlighetsregning, diskrete og kontinuerlige fordelinger, estimering, hypotesetesting, envegs variansanalyse, lineær regresjon, Monte Carlo-simulering og Shewart-diagrammer.



Alf Harbitz er forsker og statistiker ved Havforskningsinstituttets avdeling i Tromsø. Han er dr.scient. i fysikk og har arbeidet med matematisk statistikk og fysikk i næringsliv og innen universitet og høgskole fra slutten av 1970-åra.

ISBN 978-82-7674-536-8