MAT110

Statistikk 1

Løsninger til eksamensoppgaver 2012 - 2019

Per Kristian Rekdal
Bård-Inge Pettersen

Høgskolen i Molde
Vitenskapelig høgskole i logistikk

Forord

Løsninger til eksamensoppgaver:

Dette er en samling av løsninger til eksamenssoppgaver i emnet "*MAT110 Statistikk 1*" ved Høgskolen i Molde fra 2015 − 2019.

Det finnes også en tilhørende samling med komplette eksamensoppgaver til disse løsningene. Samlingen med oppgaver finnes i en egen PDF-fil, separert fra dette løsningsheftet.

Gratis:

Både samlingen med løsninger og tilhørende samling med eksamensoppgaver kan lastes ned gratis via Høgskolen i Molde sin åpne kursportal www.himoldeX.no.

Bård-Inge Pettersen
Per Kristian Rekdal

Innhold

Kapittel 1

Hovedeksamen 2012

LØSNING: Eksamen 1. juni 2012

MAT110 Statistikk I

Oppgave 1: (sannsynlighetsregning)

a) $\underline{\underline{P(A \cup B) \;=\; P(A) + P(B)}}$, $\underline{\underline{P(\,A \cap B) \;=\; P(A) \cdot P(B)}}$

b) Den *spesielle* addisjonssetningen , Den *spesielle* multiplikasjonssetningen

c) Den *spesielle* addisjonssetningen forutsetter:
Begivenhetene A og B er disjunkte, dvs. $A \cap B = \varnothing$, dvs. A og B inntreffer aldri samtidig.

Den *spesielle* multiplikasjonssetningen forutsetter:
Begivenhetene A og B er uavhengige, dvs. $P(A|B) = P(A)$ eller $P(B|A) = P(B)$.

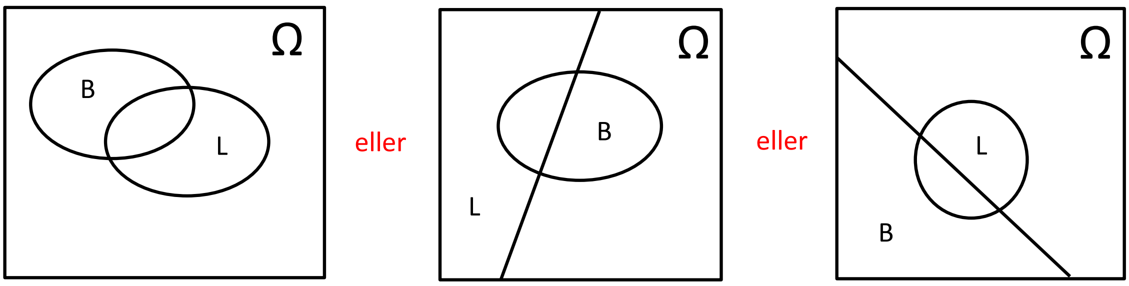
■

Oppgave 2: (betingede sannsynligheter, økonomi)

a) i) Ut fra opplysningene i oppgaveteksten ser vi umiddelbart at: $\underline{\underline{P(B) \;=\; 0.20}}$

ii) Komplementsetningen gir: $\underline{\underline{P(\overline{B})}} \;=\; 1 - 0.20 \;=\; \underline{\underline{0.80}}$

b) Venn-diagram kan tegnes på flere måter. En måte er nok:



Figur 1: Venn-diagram.

c) Multiplikasjonsetningen gir:

$$\underline{\underline{P(B \cap L)}} \;=\; P(L|B) \cdot P(B) \;=\; 0.75 \cdot 0.20 \;=\; \underline{\underline{0.15}} \tag{1}$$

d) Oppsplitting av utfallsrom Ω:

$$\underline{\underline{P(L)}} \;=\; P(L|B) \cdot P(B) \;+\; P(L|\overline{B}) \cdot P(\overline{B}) \tag{2}$$

$$\;=\; 0.75 \cdot 0.20 \;+\; 0.30 \cdot 0.80 = \underline{\underline{0.39}} \tag{3}$$

e) Multiplikasjonsetningen gir:

$$\underline{\underline{P(B|L)}} \quad = \quad \frac{P(B\cap L)}{P(L)} \quad = \quad \frac{0.15}{0.39} \quad = \quad \underline{\underline{0.38}} \tag{4}$$

Alternativt så kan Bayes lov brukes:

$$\underline{\underline{P(B|L)}} \quad = \quad P(L|B)\cdot\frac{P(B)}{P(L)} \quad = 0.75\cdot\frac{0.20}{0.39} \quad = \quad \underline{\underline{0.38}} \tag{5}$$

f) Komplementsetningen gir:

$$\underline{\underline{P(\overline{B}|L)}} \; = \; 1 - P(B|L) \quad = \quad 1 - 0.38 \quad = \quad \underline{\underline{0.62}} \tag{6}$$

g) For å bestemme lønnsomheter så regner vi ut forventet fortjeneste:

$$\underline{E[X]} \quad = \quad \sum_{n=1}^{2} x_i\, P(X = x_i) \quad = \quad x_1 \cdot P(X = x_1) \; + \; x_2 \cdot P(X = x_2) \tag{7}$$

$$= \quad (\,-10\,000 \cdot 0.38 \; + \; 8\,000 \cdot 0.62\,)\,\text{NOK} \quad = \quad \underline{1\,160\ \text{NOK}} \tag{8}$$

<ins>Ja</ins>, siden forventet <ins>fortjeneste er positiv</ins> så er det lønnsomt å gi lån til de nye kundene i lavinntektsgruppen.

■

Oppgave 3: (binomisk fordeling, logistikk)

a) "Forsøksserien" med oppmøte til en flyavgang har følgende egenskaper:

1. For hver passasjèr er det kun 2 mulige utfall, oppmøte (suksess) eller ikke oppmøte (fiasko).
2. I oppgaven antas samme sannsynlighet p for oppmøte for alle n billettkjøperne.
3. I oppgaven antas det at alle billettkjøperne møter opp uavhengig av hverandre.
4. Det gjennomføres et bestemt antall forsøk, dvs. et bestemt antall passasjèrer i dette tilfellet.

Forsøksserien oppfyller dermed kravene til en **binomisk** forsøksserie. Den stokastiske variablene X som beskriver antallet suksesser i denne forsøksserien er derfor **binomisk** fordelt.

b) i) Forventning av $X \sim \text{Bin}[n, p]$:

$$\underline{\underline{E[X]}} \;=\; n \cdot p \;= 120 \cdot 0.95 \;=\; \underline{\underline{114}} \tag{9}$$

ii) Tolkning:

$$\underline{\underline{E[X] \;=\; \text{forventet antall billettkjøpere som faktisk møter opp til sin flyavgang}}}$$

c) i) Variansen til $X \sim \text{Bin}[n, p]$:

$$\underline{\underline{Var[X]}} \;=\; n \cdot p\,(1-p) \;= 120 \cdot 0.95 \cdot (1 - 0.95) \;=\; \underline{\underline{5.7}} \tag{10}$$

ii) Tolkning:

$$\underline{\underline{Var[X]}} = \underline{\underline{\text{forventet varians/usikkerhet i antall billettkjøpere som faktisk møter opp til sin flyavgang}}}$$

d) Forventet inntekt for SAS når $n = 120$:

$$\underline{\underline{E[I]}} = E[a \cdot X] = a \cdot \underbrace{E[X]}_{=n \cdot p} = (\,800 \cdot n \cdot p\,)\,\text{NOK} \tag{11}$$

$$= (\,800 \cdot 120 \cdot 0.95\,)\,\text{NOK} = \underline{\underline{91\,200\text{ NOK}}} \tag{12}$$

e) Forventet inntekt for SAS når $n = 123$:

$$\underline{\underline{E[I]}} = E[a \cdot X] = a \cdot \underbrace{E[X]}_{=n \cdot p} = (\,800 \cdot n \cdot p\,)\,\text{NOK} \tag{13}$$

$$= (\,800 \cdot 123 \cdot 0.95\,)\,\text{NOK} = \underline{\underline{93\,480\text{ NOK}}} \tag{14}$$

f) Sannsynligheten for at det møter opp flere passasjèrer enn flyet har kapasitèt til:

$$\underline{\underline{P(X \geq 121)}} = P(X = 121) + P(X = 122) + P(X = 123) \tag{15}$$

$$= \binom{n}{121} p^{121} (1-p)^{n-121} + \binom{n}{122} p^{122} (1-p)^{n-122} + \binom{n}{123} p^{123} (1-p)^{n-123}$$

$$= \binom{123}{121} 0.95^{121} (1-0.95)^{n-121} + \binom{123}{122} p^{122} (1-0.95)^{n-122} \tag{16}$$

$$+ \binom{123}{123} 0.95^{123} (1-0.95)^{123-123} \tag{17}$$

$$= 0.0378 + 0.0118 + 0.0018 \tag{18}$$

$$= \underline{\underline{0.0514}} \quad \text{(eksakt svar med 4 desimales nøyaktighet)} \tag{19}$$

__Kommentàr__:

Det er meningen at denne oppgaven skal løses på måten som vist i lign.(15). Det er en eksakt løsning. Man trenger nemlig mellomregningene, dvs. lign.(18), i denne oppgaven for å løse oppgave g.

Isolert sett kan deloppgave f også løses ved hjelp av en *tilnærmet* metode. Den tilnærmede metoden gir ikke et så godt svar som det eksakte, selvsagt. Men man får et svar som er i nærheten: Siden (se lign.(7.66) i formelsamling)

$$n \cdot p(1-p) = 5.8 \gtrsim 5 \tag{20}$$

så er X *tilnærmet* en normalfordeling, $X \sim \mathrm{N}[\,E[X], \sigma[X]\,]$, hvor $E[X] = n \cdot p = 116.85$ og $Var[X] = n \cdot p(1-p) = 5.8425$ ($n = 123$ og $p = 0.95$). Uten heltallskorreksjon får man da:

$$\underline{P(x \geq 121)} = 1 - P(Z \leq 1.72) = 1 - 0.9573 = \underline{0.0427} \quad \text{(tilnærmet svar)} \tag{21}$$

Heltallskorreksjon gir ikke alltid et bedre svar, jfr. kommentarèr i læreboken. Derfor er ikke heltallskorreksjon tatt med her.

g) Forventet utgift til SAS ved en slik "overbooking", dvs. når $n = 123$:

$$\underline{\underline{E[U]}} = E[b \cdot Y] = b \cdot E[Y] \tag{22}$$

$$= b \sum_{y=1}^{3} y \, P(Y = y) \tag{23}$$

$$= b \cdot \Big(1 \cdot P(Y = 1) + 2 \cdot P(Y = 2) + 3 \cdot P(Y = 3) \Big) \tag{24}$$

$$= b \cdot \Big(\underbrace{1 \cdot \underbrace{P(X = 121)}_{=0.0378} + 2 \cdot \underbrace{P(X = 122)}_{=0.0118} + 3 \cdot \underbrace{P(X = 123)}_{=0.0018}}_{\text{numeriske resultat hentet fra oppgave f, lign.(18)}} \Big) \tag{25}$$

$$= 5000 \cdot \Big(1 \cdot 0.0378 + 2 \cdot 0.0118 + 3 \cdot 0.0018 \Big) \text{ NOK} \tag{26}$$

$$= \underline{\underline{334 \text{ NOK}}} \tag{27}$$

h) Siden forventet billettinntekter ved overbooking er

$$\underline{E[I] - E[U]} = \Big(93\,480 - 334 \Big) \text{NOK} = \underline{93\,146 \text{ NOK}} \tag{28}$$

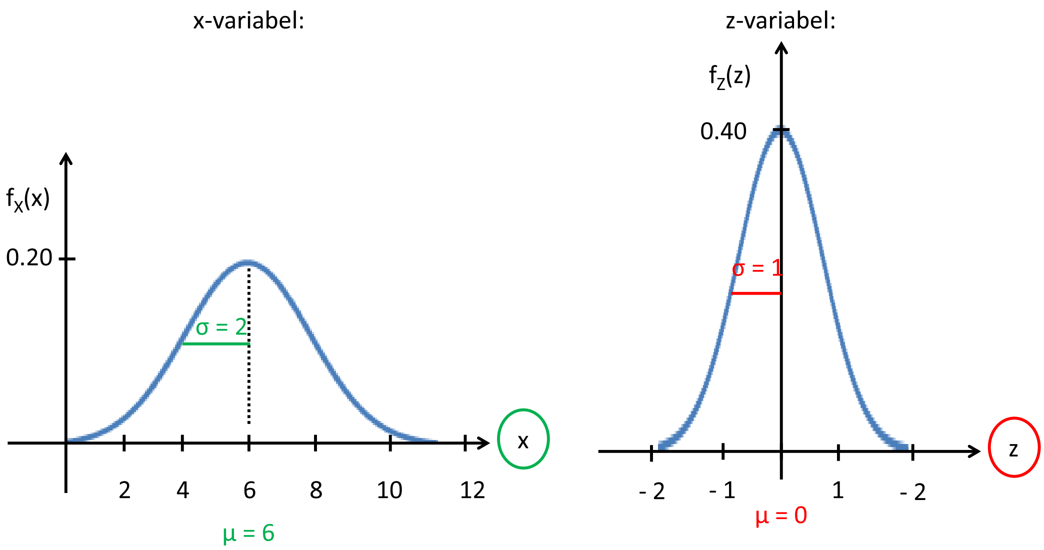
er større enn forventet inntekt ved fullt fly, 91 200 NOK (se oppgave d), så <u>lønner det seg</u> for SAS å overbooke.

■

Oppgave 4: (normalfordeling)

a) Normalfordelingen er en kontinuerlig sannsynlighetsfordeling.

b) Tetthetsfunksjone $f_X(x)$ og $f_Z(z)$:



Figur 2: Tetthetsfunksjone $f_X(x)$ og $f_Z(z)$.

c) Arealet under enhver gyldig tetthetsfunksjon er normert til $\underline{\underline{\ 1\ }}$. ■

Oppgave 5: (sentralgrensesetningen, økonomi og logistikk)

a) Siden

$$\underline{\underline{\sum_{i=0}^{n} p_i}} \quad = \quad p_0 + p_1 + p_2 + p_3 \quad = \quad 0.55 + 0.25 + 0.15 + 0.05 \quad = \quad \underline{\underline{1}} \tag{29}$$

så er den oppgitte sannsynlighetsfordelingen en gyldig fordeling.

b) i) Forventet antall feilleveringer per dag:

$$\underline{\underline{E[X]}} \overset{\text{def.}}{=} \sum_{i=0}^{3} x_i \cdot P(X = x_i) \tag{30}$$

$$= \quad 0 \cdot \underbrace{P(X=0)}_{=0.55} \; + \; 1 \cdot \underbrace{P(X=1)}_{=0.25} \; + \; 2 \cdot \underbrace{P(X=2)}_{=0.15} \; + \; 3 \cdot \underbrace{P(X=3)}_{=0.05} \tag{31}$$

$$= \quad 0 \cdot 0.55 \; + \; 1 \cdot 0.25 \; + \; 2 \cdot 0.15 \; + \; 3 \cdot 0.05 = \underline{\underline{0.70}} \tag{32}$$

ii) For å finne variansen $Var[X]$ må vi først ha $E[X^2]$:

$$\underline{E[X^2]} \overset{\text{def.}}{=} \sum_{i=0}^{3} x_i^2 \cdot P(X = x_i) \tag{33}$$

$$= \quad 0^2 \cdot \underbrace{P(X=0)}_{=0.55} \; + \; 1^2 \cdot \underbrace{P(X=1)}_{=0.25} \; + \; 2^2 \cdot \underbrace{P(X=2)}_{=0.15} \; + \; 3^2 \cdot \underbrace{P(X=3)}_{=0.05} \tag{34}$$

$$= \quad 0^2 \cdot 0.55 \; + \; 1^2 \cdot 0.25 \; + \; 2^2 \cdot 0.15 \; + \; 3^2 \cdot 0.05 = \underline{1.30} \tag{35}$$

Dette innsatt i setningen for varians: (se formelsamling)

$$\underline{\underline{Var[X]}} = E[X^2] - E[X]^2 = 1.30 - 0.70^2 = \underline{\underline{0.81}} \tag{36}$$

c) i) Forventet antall feillevèringer per dag i *gjennomsnitt* over ett år:

$$\underline{\underline{E[\overline{X}]}} = E\left[\frac{X_1 + X_2 + \ ... \ + X_n}{n}\right] = \frac{1}{n}\left(\overbrace{E[X_1] + E[X_2] + \ ... \ + E[X_n]}^{\text{n stk.}}\right) \tag{37}$$

$$= \frac{\cancel{n} E[X]}{\cancel{n}} = E[X] = \underline{\underline{0.70}} \tag{38}$$

NB: Overgangen i lign.(37) gjelder alltid. Uansett om de stokastiske variablene X_i er uavhengige eller ikke.

ii) Variansen til gjennomsnittet av antall solgte biler per dag:

$$\underline{\underline{Var[\overline{X}]}} = Var\left[\frac{X_1 + X_2 + \ ... \ + X_n}{n}\right] \tag{39}$$

$$\overset{\text{uavhengig}}{=} \frac{1}{n^2}\left(\overbrace{Var[X_1] + Var[X_2] + \ ... \ + Var[X_n]}^{\text{n stk.}}\right) \tag{40}$$

$$= \frac{\cancel{n} Var[X]}{n^{\cancel{2}}} \tag{41}$$

$$= \frac{Var[X]}{n} = \frac{0.81}{300} = \underline{\underline{0.0027}} \tag{42}$$

NB: Overgangen i lign.(41) gjelder *kun* dersom de stokastiske variablene X_i er uavhengige.

d) Utfylt versjon av tabellen:

tyngdepunkt	
E[X]	0.70
E[$\overline{X}$]	0.70

spredning	
Var[X]	0.81
Var[$\overline{X}$]	0.0027

Figur 3: Utfylt versjon av tabellen.

Kommentarer:

Forventingene til $\overline{X}$ og X er de samme, dvs.

$$E[\overline{X}] \quad = \quad E[X] \tag{43}$$

Med andre ord: tyngdepunktet til sannsynlighetsfordelingen $P(\overline{X} = \overline{x})$ er sammenfallende med tyngdepunktet til $P(X = x)$.

Men variansen til $\overline{X}$ er mye mindre:

$$Var[\overline{X}] \quad \ll \quad Var[X] \tag{44}$$

Med andre ord: spredningen/usikkerheten til sannsynlighetsfordelingen $P(\overline{X} = \overline{x})$ er mye mindre enn spredningen/usikkerheten til $P(X = x)$.

e) i) Med forutsetningene som formulet i oppgaven så gjelder sentralgrensesetningen.

ii) Ifølge sentralgrensesetningen er da den stokastiske varibalen $\overline{X}$, dvs. gjennomsnittet, **normal**fordelt

$$\underline{\underline{\overline{X} \quad \sim \quad \mathrm{N}[\, E[\overline{X}\,]\, , \, Var[\overline{X}\,]\,] \qquad = \quad \mathrm{N}\left[\, E[\,X\,]\, , \, \frac{Var[\,X\,]}{n}\right]}} \tag{45}$$

iii) En **tommelfingerregel** for at sentralgrensesetningen skal gjelde er:

$$\underline{\underline{n \; \gtrsim \; 30}} \tag{46}$$

dvs. antall forsøk bør være ca. 30 eller mer.

f) Sannsynligheten for at et avisbud har mer enn 180 feilleveringer i året:
(*uten* heltallskorreksjon)

$$\underline{P(X_1 + X_2 + \; ... \; + X_n > 180)} \quad = \quad P\left(\frac{X_1 + X_2 + \; ... \; + X_n}{n} > \frac{180}{n}\right) \tag{47}$$

$$= \quad P(\overline{X} > \frac{180}{n}) \tag{48}$$

$$= \quad 1 \; - \; P(\overline{X} \leq \frac{180}{n}) \tag{49}$$

$$\overset{\text{standardisèr}}{=} \quad 1 \; - \; P\left(\underbrace{\frac{\overline{X} - E[\overline{X}]}{\sigma[\,\overline{X}\,]}}_{= \, \overline{Z}} \leq \frac{\frac{180}{n} - E[\overline{X}]}{\sigma[\overline{X}]}\right) \tag{50}$$

$$= \quad 1 \; - \; P\left(\overline{Z} \; \leq \; \frac{\frac{180}{300} - 0.70}{\sqrt{0.0027}}\right) \tag{51}$$

$$= \quad \underline{1 \; - \; P(\overline{Z} \; \leq \; -1.92)} \tag{52}$$

$$
\begin{aligned}
\underline{\underline{P(X_1+X_2+\ \ldots\ +X_n>180)}} &= 1 - P(\overline{Z} \leq -1.92) && (53)\\
&= \cancel{1} - \left[\cancel{1} - P(\overline{Z} \leq 1.92)\right] && (54)\\
&= P(\overline{Z} \leq 1.92) && (55)\\
&= G(1.92) \overset{\text{tabell}}{=} \underline{\underline{0.9726}} && (56)
\end{aligned}
$$

$\blacksquare$

Kapittel 2

Kontinuasjonseksamen 2012

LØSNING: Eksamen 10. januar 2013

MAT110 Statistikk I

Oppgave 1: (sannsynlighetsregning)

a) Regner ut $P(E) \cdot P(F)$:

$$\underline{\underline{P(E) \cdot P(F)}} \quad = \quad 0.52 \cdot 0.46 \quad = \quad \underline{\underline{0.2392}} \tag{1}$$

b) Ifølge den *spesielle* multiplikajonssetningen vet vi at begivenhetene E og F er uavhengige dersom $P(E) \cdot P(F) = P(E \cap F)$. Siden

$$P(E) \cdot P(F) \quad = \quad 0.2392 \tag{2}$$

og

$$P(E \cap F) \quad = \quad 0.42 \tag{3}$$

så innser vi umiddelbart at E og F er $\underline{\underline{\text{ avhengige }}}$.

c) Ifølge den *spesielle* multiplikajonssetningen vet vi at begivenhetene E og G er uavhengige dersom $P(E) \cdot P(G) = P(E \cap G)$. Siden

$$P(E) \cdot P(G) \quad = \quad 0.52 \cdot 0.38 \quad = \quad 0.1976 \tag{4}$$

og

$$P(E \cap G) \quad = \quad 0.1976 \tag{5}$$

så innser vi umiddelbart at E og G er $\underline{\underline{\text{ uavhengige }}}$.

d) Bruker komplementsetningen:

$$\underline{\underline{P(\overline{B})}} \quad = \quad 1 - P(B) \;=\; 1 - 0.45 \;=\; \underline{\underline{0.55}} \tag{6}$$

e) Sannsynlighet for oppsplitting av utfallsrom Ω:

$$\underline{\underline{P(A)}} \quad = \quad P(A|B) \cdot P(B) \;+\; P(A|\overline{B}) \cdot P(\overline{B}) \tag{7}$$

$$= \quad 0.33 \cdot 0.45 \;+\; 0.82 \cdot 0.55 \;=\; \underline{\underline{0.5995}} \tag{8}$$

■

Oppgave 2: (**binomisk** og **normal** fordeling, transport, logistikk)

a) 4 krav må være oppfylt for at X skal være binomisk fordelt:

1. Hvert forsøk skal ha 2 mulige utfall, s (suksess) eller f (fiasko).
2. Det skal være samme sannsynlighet ($p = 0.90$) for suksess i alle n forsøkene.
3. Alle forsøk er uavhengige.
4. Vi gjennomfører et bestemt antall forsøk, i dette tilfellet $n = 150$.

Alle disse 4 kravene er oppfylt i vårt tilfelle. Derfor er det rimelig å anta at X er binomisk fordelt, dvs. $X \sim \text{Bin}[\,n = 150, p = 0.9\,]$.

b) Forventet antall personer som kommer på utflukten:

$$\underline{\underline{E[X]}} \overset{\text{bin.}}{=} n \cdot p = 150 \cdot 0.90 = \underline{\underline{135}} \tag{9}$$

c) i) Variansen til antall person som kommer på utflukten:

$$\underline{\underline{Var[X]}} \overset{\text{bin.}}{=} n \cdot p \cdot (1-p) = 150 \cdot 0.90 \cdot (1 - 0.90) = \underline{\underline{13.5}} \tag{10}$$

ii) Tilhørende standardavviket $\sigma[X]$:

$$\underline{\underline{\sigma[X]}} = \sqrt{Var[X]} = \sqrt{13.5} \approx \underline{\underline{3.67}} \tag{11}$$

d) i) Betingelse som må være oppfylt så for at en binomisk fordeling kan tilnærmes med en **normal**fordeling: [1]

$$\underline{\underline{n \cdot p\,(1-p)\ \gtrsim\ 5}} \tag{12}$$

ii) Før vårt tilfelle:

$$150 \cdot 0.90\,(1 - 0.90)\ =\ 13.5\ \gtrsim 5 \tag{13}$$

<u>Ja, betingelsen er godt oppfylt for vårt tilfelle.</u>

e) Siden betingelsen i lign.(12) er oppfylt så er biominal fordelingen tilnærmet en **normal**fordeling. Dette gjør det enklere å regne ut sannsynligheten for at alle oppmøtte får plass er:

$$\underline{\underline{P(\text{en tur})}} \quad = \quad P(X \leq 140) \tag{14}$$

$$\overset{\text{standardisèr}}{=} \quad P\Bigg(\underbrace{\frac{X - E[X]}{\sigma[X]}}_{\equiv Z} \leq \underbrace{\frac{140 + 0.5 - E[X]}{\sigma[X]}}_{\equiv Z_0}\Bigg) \tag{15}$$

$$= \quad P\Bigg(Z \leq \underbrace{\frac{140 + 0.5 - 135}{3.67}}_{\equiv Z_0}\Bigg) \tag{16}$$

$$= \quad P(\,Z \leq 1.50\,) \tag{17}$$

$$\overset{\text{tabell}}{=} \quad \underline{\underline{0.9332}} \tag{18}$$

[1] Se setningen på side 26 i formelsamlingen.

f) Det er helt sikkert at alle studenter blir transporttert til øya dersom det kjøres 2 turer. Dermed

$$P(\text{en tur}) \;+\; P(\text{to turer}) \;=\; 1 \tag{19}$$

Men fra oppgave **e** har vi: $P(\text{en tur}) = 0.9332$. Dermed:

$$\underline{\underline{P(\text{to turer})}} \;=\; 1 \;-\; P(\text{to turer}) \tag{20}$$

$$= \;1 \;-\; 0.9332 \;=\; \underline{\underline{0.0668}} \tag{21}$$

g) Forventet utgifter forbundet med å frakte studenter til Hjertøya:

$$\underline{\underline{E[U]}} \;=\; E[\,c\cdot Y\,] \;=\; c\cdot E[Y] \tag{22}$$

$$= \;c\sum_{i=1}^{2} y_i\; P(Y = y_i) \tag{23}$$

$$= \;c\cdot\Big(1\cdot P(Y=1) \;+\; 2\cdot P(Y=2)\Big) \tag{24}$$

$$= \;c\cdot\Big(\underbrace{1\cdot\underbrace{P(X\le 140)}_{=0.9332} + 2\cdot\underbrace{P(X>140)}_{=0.0668}}_{\text{resultat hentet fra oppg. } \mathbf{e} \text{ og } \mathbf{f}\text{, lign.(18) og (21).}}\Big) \tag{25}$$

$$= \;950\cdot\Big(1\cdot 0.9332 \;+\; 2\cdot 0.0668\Big)\;\text{NOK} \tag{26}$$

$$= \;\underline{\underline{1013.46\;\;\text{NOK}}} \tag{27}$$

h) Forventet fortjeneste forbundet med å frakte studenter til Hjertøya:

$$\underline{\underline{E[F]}} = E[\,a \cdot X - c \cdot Y\,] \tag{28}$$

$$= a \cdot E[X] \; - \; \underbrace{c \cdot E[Y]}_{=E[U]} \tag{29}$$

$$= 35 \cdot 135 \; - \; 1013.46 \tag{30}$$

$$= \underline{\underline{\;3711.54 \;\; \text{NOK}\;}} \tag{31}$$

■

Oppgave 3: (sannsynlighetsfordelinger, normalfordelingen)

a) Se vedlegg A.

b) Se vedlegg B.

c) Arealet under begge begge tetthetsfunksjonene $f_X(x)$ og $f_Z(z)$ er begge normert til $\underline{\underline{\;\;1\;\;}}$.

Oppgave 4: (sentralgrensesetningen, økonomi og logistikk)

a) Siden

$$\underline{\underline{\sum_{i=0}^{n} p_i}} = p_0 + p_1 + p_2 + p_3 = 0.70 + 0.15 + 0.10 + 0.05 = \underline{\underline{1}} \tag{32}$$

så er den oppgitte sannsynlighetsfordelingen en **gyldig** fordeling.

b) i) Forventning $E[X]$:

$$\underline{\underline{E[X]}} \overset{\text{def.}}{=} \sum_{i=0}^{3} x_i \cdot P(X = x_i) \tag{33}$$

$$= 0 \cdot \underbrace{P(X=0)}_{=0.70} + 1 \cdot \underbrace{P(X=1)}_{=0.15} + 2 \cdot \underbrace{P(X=2)}_{=0.10} + 3 \cdot \underbrace{P(X=3)}_{=0.05} \tag{34}$$

$$= 0 \cdot 0.70 + 1 \cdot 0.15 + 2 \cdot 0.10 + 3 \cdot 0.05 = \underline{\underline{0.50}} \tag{35}$$

ii) Tolking:

$$\underline{\underline{\begin{aligned} E[X] = \ &\text{forventet antall feilleveringer for et tilfeldig valgt bud} \\ &\text{i Stockholm en tilfeldig valgt dag} \end{aligned}}} \tag{36}$$

c) i) For å finne variansen $Var[X]$ må vi først ha $E[X^2]$:

$$\underline{E[X^2]} \stackrel{\text{def.}}{=} \sum_{i=0}^{3} x_i^2 \cdot P(X = x_i) \tag{37}$$

$$= 0^2 \cdot \underbrace{P(X=0)}_{=0.70} + 1^2 \cdot \underbrace{P(X=1)}_{=0.15} + 2^2 \cdot \underbrace{P(X=2)}_{=0.10} + 3^2 \cdot \underbrace{P(X=3)}_{=0.05} \tag{38}$$

$$= 0^2 \cdot 0.70 + 1^2 \cdot 0.15 + 2^2 \cdot 0.10 + 3^2 \cdot 0.05 = \underline{1.0} \tag{39}$$

Dette innsatt i setningen for varians: (se formelsamling)

$$\underline{\underline{Var[X]}} = E[X^2] - E[X]^2 = 1.0 - 0.50^2 = \underline{\underline{0.75}} \tag{40}$$

ii) Tolking:

$$\underline{\underline{\begin{aligned} Var[X] = \;& \text{forventet variasjon/spredning i antall feilleveringer} \\ & \text{for et tilfeldig valgt bud i Stockholm en tilfeldig valgt dag} \end{aligned}}} \tag{41}$$

d) i) Forventet antall feilleveringer per dag i *gjennomsnitt* over ett år for et bud hos Bring:

$$\underline{\underline{E[\,\overline{X}\,]}} \;=\; E\left[\frac{X_1+X_2+\;...\;+X_n}{n}\right] \;=\; \frac{1}{n}\left(\overbrace{E[X_1]+E[X_2]+\;...\;+E[X_n]}^{\text{n stk.}}\right) \tag{42}$$

$$= \;\frac{\cancel{n}\,E[X]}{\cancel{n}} \;=\; E[X] \;=\; \underline{\underline{0.50}} \tag{43}$$

NB: Overgangen i lign.(42) gjelder alltid. Uansett om de stokastiske variablene X_i er uavhengige eller ikke.

ii) Variansen til *gjennomsnittet* over ett år av antall feilleveringer per dag for et bud hos Bring:

$$\underline{\underline{Var[\,\overline{X}\,]}} \;=\; Var\left[\frac{X_1+X_2+\;...\;+X_n}{n}\right] \tag{44}$$

$$\overset{\text{uavhengig}}{=} \;\frac{1}{n^2}\left(\overbrace{Var[X_1]+Var[X_2]+\;...\;+Var[X_n]}^{\text{n stk.}}\right) \tag{45}$$

$$= \;\frac{\cancel{n}\,Var[X]}{n^{\cancel{2}}} \tag{46}$$

$$= \;\frac{Var[X]}{n} \;=\; \frac{0.75}{312} \;\approx\; \underline{\underline{0.0024}} \tag{47}$$

NB: Overgangen i lign.(45) gjelder *kun* dersom de stokastiske variablene X_i er uavhengige.

e) Forventingene til $\overline{X}$ og X er de samme, dvs.:

$$\underbrace{E[\,\overline{X}\,]}_{=\,0.50} \quad = \quad \underbrace{E[X]}_{=\,0.50} \tag{48}$$

Med andre ord: tyngdepunktet til sannsynlighetsfordelingen $P(\overline{X} = \overline{x})$ er sammenfallende med tyngdepunktet til $P(X = x)$.

Men variansen til $\overline{X}$ er mye mindre:

$$\underbrace{Var[\,\overline{X}\,]}_{=\,0.0024} \quad \ll \quad \underbrace{Var[\,X\,]}_{=\,0.75} \tag{49}$$

Med andre ord: spredningen/usikkerheten til sannsynlighetsfordelingen $P(\overline{X} = \overline{x})$ er mye mindre enn spredningen/usikkerheten til $P(X = x)$.

f) i) Med forutsetningene som formulet i oppgaven så gjelder sentralgrensesetningen.

ii) Ifølge sentralgrensesetningen er da den stokastiske varibalen $\overline{X}$, dvs. gjennomsnittet, **normal**fordelt:

$$\overline{X} \;\sim\; \mathrm{N}[\,E[\,\overline{X}\,]\,,\,Var[\,\overline{X}\,]\,] \quad = \quad \mathrm{N}\left[\,E[\,X\,]\,,\,\frac{Var[\,X\,]}{n}\,\right] \tag{50}$$

iii) En **tommelfingerregel** for at sentralgrensesetningen skal gjelde er:

$$n \;\gtrsim\; 30 \tag{51}$$

dvs. antall forsøk bør være ca. 30 eller mer.

g) Sannsynligheten for at det gjøres mer enn 200 feilleveringer i året per bud:
(*uten* heltallskorreksjon)

$$\underline{\underline{P(X_1 + X_2 + \ \dots \ + X_n > 200)}} = P\left(\frac{X_1 + X_2 + \ \dots \ + X_n}{n} > \frac{200}{n}\right) \tag{52}$$

$$= P(\overline{X} > \frac{200}{n}) \tag{53}$$

$$= 1 - P(\overline{X} \leq \frac{200}{n}) \tag{54}$$

$$\overset{\text{standardiser}}{=} 1 - P\Bigg(\underbrace{\frac{\overline{X} - E[\overline{X}]}{\sigma[\overline{X}]}}_{= \overline{Z}} \leq \frac{\frac{200}{n} - E[\overline{X}]}{\sigma[\overline{X}]}\Bigg) \tag{55}$$

$$= 1 - P\left(\overline{Z} \ \leq \ \frac{\frac{200}{312} - 0.50}{\sqrt{0.0024}}\right) \tag{56}$$

$$= 1 - P(\overline{Z} \ \leq \ 2.88) \tag{57}$$

$$= 1 - G(2.88) \tag{58}$$

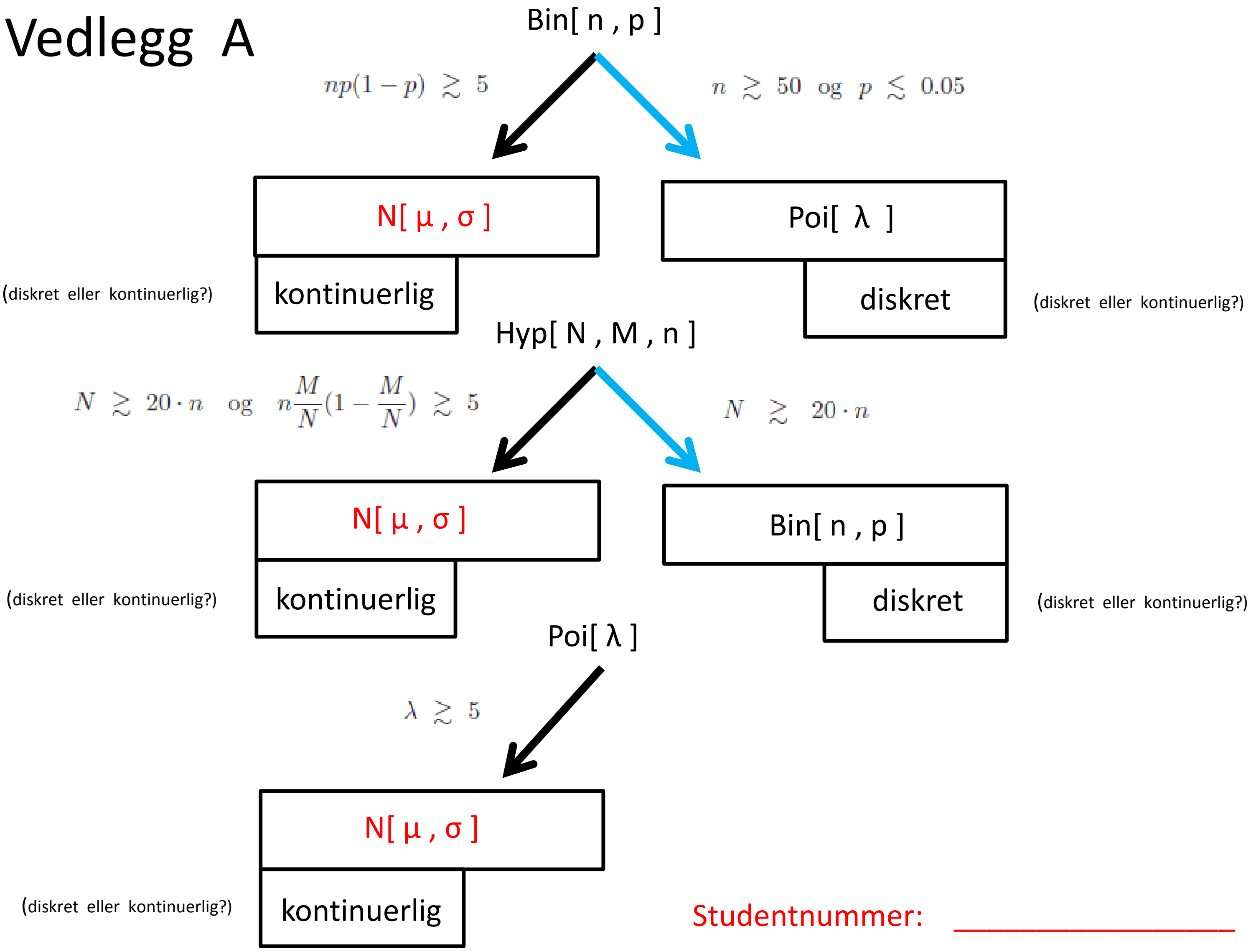
$$\overset{\text{tabell}}{=} 1 - 0.9980 \tag{59}$$

$$= \underline{\underline{0.0020}} \tag{60}$$

■

Vedlegg A

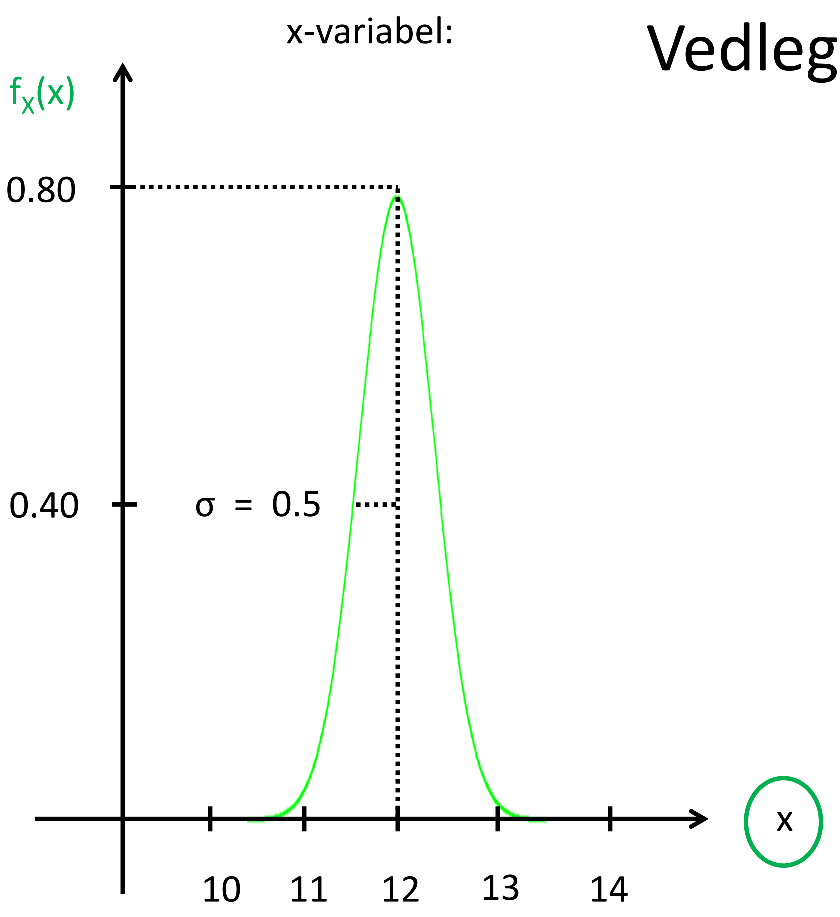
Vedlegg A

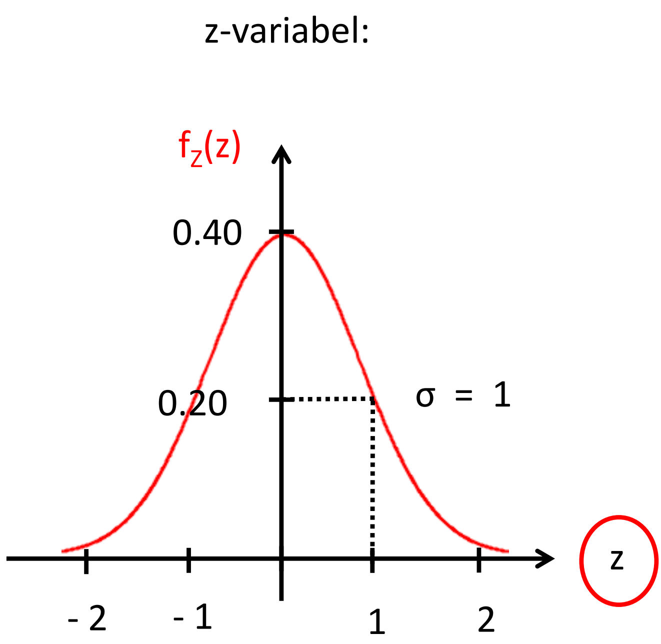


Studentnummer: ____________________

Vedlegg B

Vedlegg B





Studentnummer: ____________________

Kapittel 3

Hovedeksamen 2013

LØSNING: Eksamen 30. mai 2013

"*MAT110 Statistikk 1*", 2013

Oppgave 1: (sentrale formler, oversikt)

Se vedlegg A.

■

Oppgave 2: (sannsynlighetsregning, logistikk)

a) Tolkning:

$$\underline{\underline{P(B_2|B_1)}} = \text{sannsynligheten for at bølgehøyden er for stor dag 2,} \quad \underline{\underline{\text{gitt at den er for stor dag 1}}} \tag{1}$$

b) Bruker definisjonen av uavhengighet, $P(B_2|B_1) = P(B_2)$:
Siden

$$\underbrace{\underbrace{P(B_2|B_1)}_{=\ 0.70} \neq \underbrace{P(B_2)}_{=\ 0.05}}_{\text{venstre side IKKE like høyre side}} \tag{2}$$

så følger det at begivenhetene B_1 og B_2 **ikke** er uavhengige.

c) Bruker definisjonen av betinget sannsynlighet:

$$\underline{\underline{P(B_1 \cap B_2)}} \stackrel{\text{bet.}}{=} \overbrace{P(B_2|B_1)}^{=0.70} \cdot \overbrace{P(B_1)}^{=0.05} \tag{3}$$

$$= 0.70 \cdot 0.05 = \underline{\underline{0.035}} \tag{4}$$

d) Bruk den generelle addisjonssetningen:

$$\underline{\underline{P(B_1 \cup B_2)}} \stackrel{\text{add.}}{=} \overbrace{P(B_1)}^{=0.05} + \overbrace{P(B_2)}^{=0.05} - \overbrace{P(B_1 \cap B_2)}^{=0.035} \tag{5}$$

$$= 0.05 + 0.05 - 0.035 = \underline{\underline{0.065}} \tag{6}$$

e) Bruk F.EKS. total sannsynliget:

$$\underline{\underline{P(B_1 \cap \overline{B}_2)}} \overset{\text{tot.}}{=} \overbrace{P(B_1)}^{=\ 0.05} - \overbrace{P(B_1 \cap B_2)}^{=\ 0.035} \tag{7}$$

$$= 0.05 - 0.035 = \underline{\underline{0.015}} \tag{8}$$

ELLER definisjonen på <u>betinget</u> sannsynlighet:

$$\underline{\underline{P(B_1 \cap \overline{B}_2)}} \overset{\text{bet.}}{=} \overbrace{P(\overline{B}_2|B_1)}^{=\ 1-P(B_2|B_1)} \cdot P(B_1) \tag{9}$$

$$= [1 - \overbrace{P(B_2|B_1)}^{=\ 0.70}] \cdot \overbrace{P(B_1)}^{=\ 0.05} \tag{10}$$

$$= [1 - 0.70] \cdot 0.05 = \underline{\underline{0.015}} \tag{11}$$

f) Tolkning:

$$\underline{\underline{P(B_1 \cap \overline{B}_2)}} = \underline{\underline{\text{sannsynligheten for at bølgehøyden er for stor dag 1, men ikke for høy dag 2}}} \tag{12}$$

g) Bruk definisjonen av betinget sannsynlighet:

$$P(\overline{B}_2|\overline{B}_1) = \frac{\overbrace{P(\overline{B}_2 \cap \overline{B}_1)}^{=\ 1-P(B_1 \cup B_2)}}{\underbrace{P(\overline{B}_1)}_{=\ 1-P(B_1)}} \tag{13}$$

I telleren er den ene "*tvillingsetningen*" benyttet. I nevneren er komplementsetningen benyttet.

$$\underline{\underline{P(\overline{B}_2|\overline{B}_1)}} = \frac{1 - \overbrace{P(B_2 \cup B_1)}^{=\ 0.065}}{1 - \underbrace{P(B_1)}_{=0.05}} = \frac{1 - 0.065}{1 - 0.05} = \underline{\underline{0.98}} \tag{14}$$

Kommentar:

Man kan også løse denne oppgaven ved å bruke komplementsetningen og total sannsynlighet. Men det er selvfølgelig mest naturlig å løse oppgaven på den måten som fotnoten i oppgaveteksten legger opp til.

■

Oppgave 3: (normalfordeling, sentralgrensesetningen, økonomi)

a) Sannsynligheten for at en tilfeldig valgt konto har minst ett overtrekk per måned:

$$\underline{\underline{P(X \geq 1)}} = 1 - P(X \leq 0) \tag{15}$$

$$= 1 - \underbrace{P(X = 0)}_{\overset{\text{tabell}}{=} 0.57} = 1 - 0.57 = \underline{\underline{0.43}} \tag{16}$$

b) i) Forventet antall overtrekk for en tilfeldig valgt konto:

$$\underline{\underline{E[X]}} \overset{\text{def.}}{=} \sum_{i=0}^{4} x_i \cdot P(X = x_i) \tag{17}$$

$$= 0 \cdot \underbrace{P(X = 0)}_{=0.57} + 1 \cdot \underbrace{P(X = 1)}_{=0.13} + 2 \cdot \underbrace{P(X = 2)}_{=0.18} + 3 \cdot \underbrace{P(X = 3)}_{=0.10} + 4 \cdot \underbrace{P(X = 3)}_{=0.02}$$

$$= 0 \cdot 0.57 + 1 \cdot 0.13 + 2 \cdot 0.18 + 3 \cdot 0.10 + 4 \cdot 0.02 = \underline{\underline{0.87}} \tag{18}$$

ii) For å finne variansen $Var[X]$ så regner vi først ut $E[X^2]$:

$$\underline{E[X^2]} \overset{\text{def.}}{=} \sum_{i=0}^{3} x_i^2 \cdot P(X = x_i) \tag{19}$$

$$= 0^2 \cdot \underbrace{P(X = 0)}_{=0.57} + 1^2 \cdot \underbrace{P(X = 1)}_{=0.13} + 2^2 \cdot \underbrace{P(X = 2)}_{=0.18} + 3^2 \cdot \underbrace{P(X = 3)}_{=0.10} + 4^2 \cdot \underbrace{P(X = 3)}_{=0.02}$$

$$= 0^2 \cdot 0.57 + 1^2 \cdot 0.13 + 2^2 \cdot 0.18 + 3^2 \cdot 0.10 + 4^2 \cdot 0.02 = \underline{2.07} \tag{20}$$

Dette innsatt i setningen for '*varianssetningen*': (se formelsamlingen, lign.(5.8))

$$\underline{\underline{Var[X]}} = E[X^2] - E[X]^2 = 2.07 - 0.87^2 = \underline{\underline{1.3131}} \tag{21}$$

c) i) Tolkning:

$$\underline{\underline{E[\,\overline{X}\,]}} = \underline{\underline{\text{forventet antall overtrekk per måned i gjennomsnitt for kundene i Sparebanken Møre}}} \tag{22}$$

ii) Forventet antall overtrekk i gjennomsnitt: $(\ n = 500\ )$

$$\underline{\underline{E[\,\overline{X}\,]}} = E\left[\ \frac{1}{n}\Big(X_1 + X_2 + \ ...\ + X_n\Big)\ \right] \tag{23}$$

$$\overset{\text{alltid}}{=} \frac{1}{n}\Big(E[X_1] + E[X_2] + \ ...\ + \ E[X_n]\Big) \tag{24}$$

$$= \frac{1}{n}\Big(\ \underbrace{E[X] + E[X] + \ ...\ + E[X]}_{=\ n}\ \Big) = \frac{1}{\cancel{n}}\cancel{n}\cdot\underbrace{E[X]}_{=0.87} = \underline{\underline{0.87}} \tag{25}$$

NB: Overgangen i lign.(23) til (24) gjelder alltid. Uansett om de stokastiske variablene X_i er uavhengige eller ikke.

d) i) Tolkning:

$$\underline{\underline{Var[\,\overline{X}\,]}} \;=\; \underline{\underline{\text{forventet variasjon/spredning i antall overtrekk per måned i gjennomsnitt for kundene i Sparebanken Møre}}} \tag{26}$$

ii) Variansen til gjennomsnittet av antall overtrekk per måned: $\quad (\; n = 500 \;)$

$$\underline{\underline{Var[\,\overline{X}\,]}} = Var\left[\frac{1}{n} \left(X_1 + X_2 + \;...\; + X_n \right) \right] \tag{27}$$

$$\overset{\text{uavh.}}{=} \frac{1}{n^2} \left(\underbrace{Var[X_1] + Var[X_2] + \;...\; + Var[X_n]}_{n \cdot Var[X]} \right) \tag{28}$$

$$= \frac{1}{n^{\cancel{2}}} \cancel{n} \cdot \underbrace{Var[X]}_{=1.3131} = \frac{1.3131}{500} = \underline{\underline{0.002626}} \tag{29}$$

NB: Overgangen i lign.(27) til (28) gjelder kun dersom de stokastiske variablene X_i er uavhengige.

e) Siden

1. antall overtrekk for de forskjellige kontoene er uavhengige: (oppgitt i oppgaveteksten)

   $X_i \;\sim\;$ er uavhengige for alle $i = 1, 2, 3, ..., n$

2. alle X_i har samme sannsynlighetsfordeling: (oppgitt i oppgaveteksten)

   $X_i \;\sim\;$ samme sannsynlighetsfordeling for alle $i = 1, 2, 3, ..., n$

3. antall "forsøk", dvs. antall overtrekk, $n = 500$ er tilstrekkelig stort [1]

så gjelder **sentralgrensesetningen**. Dermed er $\overline{X}$ $\underline{\underline{\text{ normalfordelt }}}$.

[1] Husk: Antall forsøk n for at sentralgrensesetningen skal gjelde er avhengig av situasjonen. Men en **tommelfingerregel** er at vi bør ha $n \gtrsim 30$.

f) Fra oppgavene foran ser vi at:

$$\underbrace{E[\,\overline{X}\,]}_{=\ 0.87} \quad = \quad \underbrace{E[X]}_{=\ 0.87} \tag{30}$$

og at

$$\underbrace{Var[\,\overline{X}\,]}_{=\ 0.002626} \quad \ll \quad \underbrace{Var[X]}_{=\ 1.3131} \tag{31}$$

Det betyr at sannsynlighetfordelingen til X, dvs. $P(X = x)$ gitt ved tabell i oppgaveteksten, og sannsynlighetfordelingen til $\overline{X}$, dvs. $\overline{X} \sim N[\,E[X], \frac{VarX}{n}\,]$ har samme tyngdepunkt, men mye mindre varians / usikkerhet.

g) Sannsynligheten for at samlet antall overtrekk per måned er større enn 400:

$$\underline{\underline{P(\,X_1 + X_2 + \ \dots\ + X_n > 400\,)}} \quad = \quad P\left(\frac{X_1 + X_2 + \ \dots\ + X_n}{n} > \frac{400}{n}\right) \tag{32}$$

$$= \quad P(\,\overline{X} > \frac{400}{n}\,) \tag{33}$$

$$= \quad 1 \;-\; P(\,\overline{X} \leq \frac{400}{n}\,) \tag{34}$$

$$\stackrel{\text{standardisèr}}{=} \quad 1 \;-\; P\left(\underbrace{\frac{\overline{X} - E[\,\overline{X}\,]}{\sigma[\,\overline{X}\,]}}_{=\overline{Z}} \;\leq\; \frac{\frac{400}{n} - E[\,\overline{X}\,]}{\sigma[\,\overline{X}\,]}\right) \tag{35}$$

$$= \quad 1 \;-\; P(\overline{Z} \;\leq\; \frac{\frac{400}{500} - 0.87}{\sqrt{0.002626}}) \tag{36}$$

$$= \quad 1 \;-\; P(\overline{Z} \;\leq\; -1.37) \tag{37}$$

$$= \quad 1 \;-\; \left(1 - \underbrace{P(\overline{Z} \;\leq\; 1.37)}_{=\;G(1.37)}\right) \tag{38}$$

$$= \quad \cancel{1} \;-\; \cancel{1} \;+\; \underbrace{G(1.37)}_{\stackrel{\text{tabell}}{=}\;0.9147} \tag{39}$$

$$= \quad \underline{\underline{0.9147}} \tag{40}$$

Kommentar:

Legg merke til at det er $\overline{X}$ som skal standardiseres. Ikke X. Det betyr at vi må bruke $E[\,\overline{X}\,] = 0.87$ og $\sigma[\,\overline{X}\,] = \sqrt{0.002626}$, (ikke $E[\,X\,] = 0.87$ og $\sigma[\,X\,] = \sqrt{1.3131}$).

h) Det skal være 95 % sannsynlighet for at det samlede antall overtrekk per måned er mindre eller lik en øvre grense X_{grense}.
Denne grensen er dermed bestemt av ligningen:

$$P(X_1 + X_2 + \ \dots \ + X_n \leq X_{\text{grense}}) \quad = \quad 0.95 \tag{41}$$

$$P\left(\frac{X_1 + X_2 + \ \dots \ + X_n}{n} \leq \underbrace{\frac{X_{\text{grense}}}{n}}_{= \overline{X}_{\text{grense}}} \right) \quad = \quad 0.95 \tag{42}$$

$$P(\overline{X} \leq \overline{X}_{\text{grense}}) \quad = \quad 0.95 \tag{43}$$

Vi standardisèrer lign.(43):

$$P\left(\underbrace{\frac{\overline{X} - E[\overline{X}\,]}{\sigma[\overline{X}\,]}}_{= \overline{Z}} \leq \underbrace{\frac{\overline{X}_{\text{grense}} - E[\overline{X}\,]}{\sigma[\overline{X}\,]}}_{= \overline{Z}_{\text{grense}}} \right) \quad = \quad 0.95 \tag{44}$$

$$P(\overline{Z} \ \leq \ \overline{Z}_{\text{grense}}) \quad = \quad 0.95 \tag{45}$$

Ved "omvendt tabelloppslag" ser vi at 0.9495 og 0.9505 ligger midt mellom 0.95. Dette tilsvarer at *argumentet* er 1.645:

$$\overline{Z}_{\text{grense}} \quad = \quad 1.645 \tag{46}$$

Dermed:

$$\overline{Z}_{\text{grense}} \quad = \quad \frac{\overline{X}_{\text{grense}} - E[\overline{X}]}{\sigma[\overline{X}]} \tag{47}$$

$$\underline{\overline{X}_{\text{grense}}} \quad = \quad \overline{Z}_{\text{grense}} \cdot \sigma[\overline{X}] \ + \ E[\overline{X}] \tag{48}$$

$$= \quad 1.645 \cdot \sqrt{0.002626} \ + \ 0.87 \quad \approx \quad \underline{0.9543} \tag{49}$$

Siden $\overline{X}_{\text{grense}} = \frac{X_{\text{grense}}}{n}$ får vi:

$$\underline{\underline{X_{\text{grense}}}} = \overline{X}_{\text{grense}} \cdot n \tag{50}$$

$$= 0.9543 \cdot 500 \approx \underline{\underline{478}} \tag{51}$$

Det er 95 % sannsynlighet for at det samlede antall overtrekk per måned er mindre enn 478.

■

Oppgave 4: (binomisk fordeling, logistikk og økonomi)

a) "Forsøksseriene" med produksjon og transport av lysarmatur og lysrør har følgende egenskaper:

1. Kun 2 mulige utfall, defekt/ødelagt ("suksess") og ikke defekt/ikke ødelagt ("fiasko").
2. Det er samme sannsynlighet p_d og p_t for alle lysrørene.
3. Lysrørene er, per antagelse, uavhengige, både hva produksjon og transport angår.
4. Det gjennomføres et bestemt antall "forsøk", dvs. et bestemt antall lysrør n produseres og transporteres.

Forsøksseriene oppfyller dermed kravene til en **binomisk** forsøksserie. De stokastiske variablene D og T er derfor **binomisk** fordelt.

b) i) Forventning av $D \sim \text{Bin}[n, p_d]$:

$$\underline{\underline{E[D]}} \;=\; n \cdot p_d \;= 25 \cdot 0.05 \;=\; \underline{\underline{1.25}} \tag{52}$$

ii) Tolkning:

$$\underline{\underline{E[D] \;=\; \text{forventet antall defekte lysrør i en produksjonsserie på } n = 25}}$$

c) i) Variansen til $D \sim \text{Bin}[n, p_d]$:

$$\underline{\underline{Var[D]}} \;=\; n \cdot p_d\,(1 - p_d) \;= 25 \cdot 0.05 \cdot (1 - 0.05) \;=\; \underline{\underline{1.1875}} \tag{53}$$

ii) Tolkning:

$$\underline{\underline{Var[D]}} = \underline{\underline{\text{forventet varians/usikkerhet i antall defekte lysrør i en produksjonsserie på } n = 25}}$$

d) Sannsynligheten for at mer enn 2 lysrør er defekte i en forsendelse:

$$\underline{\underline{P(D > 2)}} = 1 - P(D \leq 2) \tag{54}$$

$$= 1 - \Big(P(D=0) + P(D=1) + P(D=2) \Big) \tag{55}$$

$$= 1 - P(D=0) - P(D=1) - P(D=2) \tag{56}$$

$$= 1 - \binom{n}{0} p_d^0 (1-p_d)^{n-0} + \binom{n}{1} p_d^1 (1-p_d)^{n-1} + \binom{n}{2} p_d^2 (1-p_d)^{n-2} \tag{57}$$

$$= 1 - \binom{n}{0} 0.05^0 (1 - 0.05^{n-0}) - \binom{25}{1} 0.05^1 (1 - 0.05^{n-1}) - \binom{25}{2} 0.05^2 (1 - 0.05^{n-2})$$

$$= 1 - 0.2774 - 0.3650 - 0.2305 \tag{58}$$

$$= \underline{\underline{0.1271}} \qquad \text{(svar med 4 desimalers nøyaktighet)} \tag{59}$$

Kommentar: (denne kommentaren er ikke nødvendig å ha med på eksamensbesvarelsen)

Siden

$$n \cdot p_d \,(1 - p_d) \;=\; 25 \cdot 0.05\,(1 - 0.05) \;=\; 1.1875 \;\ll\; 5 \tag{60}$$

så er ikke D tilnærmet en normalfordeling. I dette tilfellet er derfor det ikke noe alternativ å løse denne oppgaven tilnærmet via en normalfordeling og tilhørende tabelloppslag. Her må man faktisk gjøre utregningen som vist ovenfor.

e) Ta forventningen av uttrykket for fortjenesten F som er oppgitt i oppgaven:

$$\underline{\underline{E[\,F\,]}} \;=\; E[\;(n - D - T)\cdot i \;-\; n\cdot(k + k_t)\;] \tag{61}$$

$$= E[\;n\cdot i - D\cdot i - T\cdot i \;-\; n\cdot(k+k_t)\;] \tag{62}$$

$$= \underbrace{E[\,n\cdot i\,]}_{=\,n\cdot i} \;-\; \underbrace{E[\,D\cdot i\,]}_{=\,E[\,D\,]\cdot i} \;-\; \underbrace{E[\,T\cdot i\,]}_{=\,E[\,T\,]\cdot i} \;-\; \underbrace{E[\,n\cdot(k+k_t)\,]}_{=\,n\cdot(k+k_t)} \tag{63}$$

$$= n\cdot i \;-\; \underbrace{E[\,D\,]}_{=\,n\cdot p_d}\cdot i \;-\; \underbrace{E[\,T\,]}_{=\,n\cdot p_t}\cdot i \;-\; n\cdot(k+k_t) \tag{64}$$

$$= n\cdot p_d\cdot i \;-\; n\cdot p_t\cdot i \;-\; n\cdot(k+k_t) \tag{65}$$

$$= \underline{\underline{n\Big[\,(1 - p_d - p_t)\cdot i \;-\; (k+k_t)\Big]}} \quad , \quad \text{q.e.d.} \tag{66}$$

f) Størst forventning oppnås i det tilfellet når utgiften er minst. Bruker tipset i fotnoten:

Bring:

$$\underline{E[F]} = n \cdot \Bigg[(1 - p_d) \cdot i - k \; - \; \underbrace{(\, p_t \cdot i + k_t \,)}_{\text{NB !}} \Bigg] \tag{67}$$

$$= n \cdot \Bigg[(1 - p_d) \cdot i - k \; - \; (\, 0.15 \cdot 1\,700 + 275 \,) \text{ NOK} \Bigg] \tag{68}$$

$$= \underline{n \cdot \Bigg[(1 - p_d) \cdot i - k \; - \; 530 \text{ NOK} \Bigg]} \tag{69}$$

DHL:

$$\underline{E[F]} = n \cdot \Bigg[(1 - p_d) \cdot i - k \; - \; \underbrace{(\, p_t \cdot i + k_t \,)}_{\text{NB !}} \Bigg] \tag{70}$$

$$= n \cdot \Bigg[(1 - p_d) \cdot i - k \; - \; (\, 0.04 \cdot 1\,700 + 750 \,) \text{ NOK} \Bigg] \tag{71}$$

$$= \underline{n \cdot \Bigg[(1 - p_d) \cdot i - k \; - \; 818 \text{ NOK} \Bigg]} \tag{72}$$

Konklusjon:

Bring har minst forventet utgift.
For å få størst forventet inntekt bør derfor Glamox velge $\underline{\underline{\text{ Bring }}}$.

Kommentar:

Den siste fotnoten i eksamensoppgaven inneholdt en feil. Den gamle, feilaktige ligningen som var oppgitt i fotnoten var: $E[\,F\,] = n \cdot (1 - p_d) \cdot i - k \;-\; n \cdot p_t \cdot i + k_t$. Denne feilen skal ikke på noen måte belastes studentene: ved sensur blir dette tatt full høyde for. Det det betyr, blant annet, at de som har brukt fotnoten slik den opprinnelig stod, får full poengsum på denne deloppgaven dersom alt ellers er rett.
For kompletthets skyld vises her også hvordan løsningen blir dersom man bruker den feilaktige ligningen:

f) Størst forventning oppnås i det tilfellet når utgiften er minst.

Bring:

$$\underline{E[\,F\,]} = n \cdot (1 - p_d) \cdot i - k \;-\; n \cdot p_t \cdot i + k_t \tag{73}$$

$$= n \cdot (1 - p_d) \cdot i - k \;-\; (\;25 \cdot 0.15 \cdot 1\,700 \;+\; 275\;)\ \text{NOK} \tag{74}$$

$$= \underline{n \cdot (1 - p_d) \cdot i - k \;-\; 6\,650\ \ \text{NOK}} \tag{75}$$

DHL:

$$\underline{E[\,F\,]} = n \cdot (1 - p_d) \cdot i - k \;-\; n \cdot p_t \cdot i + k_t \tag{76}$$

$$= n \cdot (1 - p_d) \cdot i - k \;-\; (\;25 \cdot 0.04 \cdot 1\,700 \;+\; 750\;)\ \text{NOK} \tag{77}$$

$$= \underline{n \cdot (1 - p_d) \cdot i - k \;-\; 2\,450\ \ \text{NOK}} \tag{78}$$

Med feil formel

Konklusjon:

DHL har minst forventet utgift.
For å få størst forventet inntekt bør derfor Glamox velge DHL.

■

Vedlegg A

Beskrivende statistikk
(utvalg av observasjoner)

Stokastiske variabler
(sanns.-fordeling av stok. var. X)

Side 1 **(av 2)**

Empirisk gjennomsnitt: (formel)

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Forventning: (diskret) (formel)

$$E[X] = \sum_{i=1}^{m} x_i \cdot P(X = x_i)$$

Kommentar:

Lokaliseringsmål: tyngdepunkt

Empirisk varians: (formel)

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \overline{x})^2$$

Varians: (diskret) (formel)

$$Var[X] = \sum_{i=1}^{m} (x_i - E[X])^2 \cdot P(X = x_i)$$

Kommentar:

Spredningsmål: varians

Empirisk kovarians: (formel)

$$S_{xy} = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

Kovarians: (formel)

$$Cov[X,Y] = E[(X - E[X])(Y - E[Y])]$$

Kommentar:

Et mål på lineær samvariasjon. Ikke normalisert.

Beskrivende statistikk **(utvalg av observasjoner)**	**Stokastiske variabler** **(sanns.-fordeling av stok. var. X)**

Side 2 **(av 2)**

Empirisk korrelasjonskoeffisient:	Korrelasjonskoeffisient: (formel)
$R_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$	$\rho[X,Y] = \frac{Cov[X,Y]}{\sqrt{Var[X]} \cdot \sqrt{Var[Y]}}$

Kommentar:

Et mål på lineær samvariasjon, korrelasjon.
Normalisert, ligger i intervallet: $-1 \leq$ korr. koeff. ≤ 1

Empirisk korrelasjonskoeffisient:	Korrelasjonskoeffisient: (formel)
$R_{xy} = -1$	$\rho[X,Y] = -1$

Kommentar:

Sterk negativ korrelasjon.

Empirisk korrelasjonskoeffisient:	Korrelasjonskoeffisient: (formel)
$R_{xy} = 1$	$\rho[X,Y] = 1$

Kommentar:

Sterk positiv korrelasjon.

Kapittel 4

Kontinuasjonseksamen 2013

LØSNING: Eksamen 6. jan. 2014

"MAT110 Statistikk 1"

Oppgave 1: (revisjon)

a) Dette er en tellesituasjon med uniformt utfallsrom. Da kan vi bruke urnemodellen.

$$\underline{\underline{P_{\text{A11}}}} = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{12}{2000} = \underline{\underline{0.006}} \tag{1}$$

$$\underline{\underline{P_{\text{A12}}}} = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{24}{8000} = \underline{\underline{0.003}} \tag{2}$$

b) Samme metode som **a**:

$$\underline{\underline{P_{\text{B11}}}} = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{20}{4000} = \underline{\underline{0.005}} \tag{3}$$

$$\underline{\underline{P_{\text{B12}}}} = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{2}{1000} = \underline{\underline{0.002}} \tag{4}$$

c) i) $P_{\text{A11}} > P_{\text{B11}} \Rightarrow$ $\underline{\underline{\text{strategi A}}}$ er best for 2011.
ii) $P_{\text{A12}} > P_{\text{B12}} \Rightarrow$ $\underline{\underline{\text{strategi A}}}$ er best for 2012.

Altså strategi A er best for begge årene hver for seg.

d) Strategi A og B når man ser begge årene under ett:

$$\underline{\underline{P_A}} = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{12+24}{2000+8000} = \underline{\underline{0.0036}} \tag{5}$$

$$\underline{\underline{P_B}} = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{20+2}{4000+1000} = \underline{\underline{0.0044}} \tag{6}$$

Dermed:

$P_A < P_B \quad \Rightarrow \quad \underline{\underline{\text{strategi B}}}$ er best når man ser begge årene under ett.

e) Dersom vi ser på 2011 og 2012 hver for seg så er strategi A best begge årene, jfr. oppgave **c**. Dersom vi ser på begge årene under ett så er strategi B best, jfr. oppgave **d**.

Kommentar:
Altså, selv om strategi A er best både for 2011 og 2012 hver for seg så er strategi B best begge årene sett under ett. [1] [2]

f) I oppgaven står det at bedriften er sikker på at det er flest bilag med feil i 2011. Derfor er det best (oppnår størst sannsynlighet) å gjøre flest stikkprøver i 2011, slik som i strategi B.

■

---

[1] En alternativ og noe mer kompakt og matematisk formulering av dette er:
. Selv om

$$P_{A11} > P_{B11} \tag{7}$$
$$P_{A12} > P_{B12} \tag{8}$$

så er:

$$P_A < P_B \tag{9}$$

(På eksamen kan man velge om man vil formulere seg med ord eller matematisk).

[2] Dette fenomenet er velkjent i statistikk og kalles Yule-Simpsons paradoks.

Oppgave 2: (logistikk) [3]

a) Forventet etterspørsel av aviser en gitt dag, $E[D]$:

$$\begin{aligned}
\underline{\underline{E[D]}} &= \sum_{i=0}^{5} d_i P(D = d_i) && (10)\\
&= 0 \cdot 0.1 + 1 \cdot 0.05 + 2 \cdot 0.15 + 3 \cdot 0.3 + 4 \cdot 0.25 + 5 \cdot 0.15 && (11)\\
&= \underline{\underline{\ 3\ }} && (12)
\end{aligned}$$

b) En funksjon av en tilfeldig variabel er bare en ny tilfeldig variabel. Derfor:
$S = \min(D, q)$ er en stokastisk variabel fordi den er en funksjon av D, hvor D er en stokastisk variabel.

c) Sannsynlighetsfordelingen er gyldig dersom $\sum_{i=0}^{5} P(S = s_i) = 1$.
La oss derfor se om dette er tilfelle:

$$\begin{aligned}
\underline{\underline{\sum_{i=0}^{5} P(S = s_i)}} &= P(S = 0) + P(S = 1) + \ ...\ + P(S = 5) && (13)\\
&= 0.1 + 0.05 + 0.15 + 0.7 + 0 + 0 = \underline{\underline{\ 1\ }} && (14)
\end{aligned}$$

d) i) Forventet antall solgte aviser en gitt dag når avisgutten bestiller $q = 3$ aviser: [4]

$$\begin{aligned}
\underline{\underline{E[S]}} &= \sum_{i=0}^{5} s_i P(S = s_i) && (15)\\
&= 0 \cdot 0.1 + 1 \cdot 0.05 + 2 \cdot 0.15 + 3 \cdot 0.7 + 4 \cdot 0 + 5 \cdot 0 && (16)\\
&= \underline{\underline{\ 2.45\ }} && (17)
\end{aligned}$$

[3] Problemet i denne oppgaven er kjent som "*avisguttens dilemma*" eller "*avisguttens problem*". Dette grunnleggende problemet beskriver tilbud og etterspørsel i ubalanse.

[4] Bruker sannsynlighetsfordelingen til $P(S = s_i)$ som oppgitt i oppgaven.

ii) Tolkning: $E[S]$ er forventet antall solgte aviser en gitt dag dersom avisgutten bestiller $q = 3$ aviser.

e) Teknisk forklaring:
Når avisgutten bestiller $q = 3$ aviser så er denne nye øvre grensen mindre enn den opprinnelige øvre grensen på 5 aviser. Siden sannsynlighetfordelingen $P(S = s_i)$ er den samme som $P(D = d_i)$ frem til $s = q - 1 = 2$ så må $E[D] > E[S]$.

En mer ikke-teknisk forklaring aksepteres også:
Man kan ikke selge flere aviser enn markedet etterspør.
Det er derfor rimelig at forventet etterspørsel $E[D]$ er større enn forventet salg $E[S]$.

f) Forventet fortjeneste $E[\pi(q)]$: [5]

$$\underline{E[\pi(q)]} = E[\, r\, S - w\, q\,] = \underline{rE[S] - w\, q} \tag{20}$$

For tilfellet $q = 3$ er $E[S] = 2.45$, jfr. oppgave **d**. Med $w = 5$ NOK og $r = 20$ NOK får vi:

$$\underline{\underline{E[\pi(q)]}} \quad = \quad (20 \cdot 2.45 - 5 \cdot 3) \text{ NOK} \quad = \quad \underline{\underline{34 \text{ NOK}}} \tag{21}$$

[5] Her bruker vi regneregelene: (a og b er konstanter)

$$\begin{aligned} E[aX + bY] &= aE[X] + bE[Y] \,, & (18) \\ E[a] &= a \,, & (19) \end{aligned}$$

som man kan finne i kap. 5 i formelsamlingen.

g) Med innkjøpspris $w = 5$ og utslagspris $r = 20$ fås:

$$P(D \le q^*) = 1 - \frac{w}{r} \tag{22}$$

$$P(D \le q^*) = 1 - \frac{5}{20} \tag{23}$$

$$P(D \le q^*) = 0.75 \tag{24}$$

$$P\left(\underbrace{\frac{D-\mu}{\sigma}}_{\equiv Z} \le \underbrace{\frac{q^*-\mu}{\sigma}}_{\equiv Z_0}\right) \overset{\text{standardiser}}{=} 0.75 \tag{25}$$

$$P(Z \le Z_0) = 0.75 \tag{26}$$

Ved "omvendt tabelloppslag" ser vi at $Z_0' = 0.67$ tilsvarer $P(Z' \le Z_0') = 0.7486$. Videre ser vi at $Z_0'' = 0.68$ tilsvarer $P(Z'' \le Z_0'') = 0.7517$. Vi skal ha 0.75, som er ca. midt i mellom. Dermed:

$$Z_0 = 0.675 \tag{27}$$

Vi løser:

$$Z_0 = \frac{q^*-\mu}{\sigma} \tag{28}$$

med hensyn på q^*: ($\mu = 3$ og $\sigma = 1.5$)

$$\underline{q^*} = \mu + Z_0 \cdot \sigma \tag{29}$$

$$= 3 + 0.675 \cdot 1.5 = \underline{4.0125} \tag{30}$$

Avisgutten må bestille $\underline{\underline{q^* \approx 4}}$ aviser for å få størst mulig fortjeneste.

h) Forventet etterspørsel av aviser er $\mu = 3$ per dag.
Fortjenesten blir størst når avisgutten bestiller $q^* \approx 4$ aviser per dag. Altså

$$q^* > \mu \ , \tag{31}$$

dvs. det lønner seg å bestille flere aviser enn det man forventer å selge.
Dette fordi man **taper mye mer** [6] **på tapt salg** enn på aviser han ikke får solgt.

■

[6] Avisgutten taper $\frac{r}{w} = \frac{20 \text{ NOK}}{5 \text{ NOK}} = 4$ ganger mer på tapt salg enn å brenne inne med aviser han ikke får solgt.

Oppgave 3: (økonomi)

a) Siden S_i er uavhengige så er "og"-sannsynligheten kun produktet av hver enkelt ubetinget sannsynlighet: [7]

$$\underline{\underline{P(S_1 \cap S_2)}} \;=\; P(S_1) \cdot P(S_2) \;=\; \frac{1}{6} \cdot \frac{1}{6} \;=\; \frac{1}{36} \;\approx\; \underline{\underline{0.028}} \tag{32}$$

b) Den generelle **addisjonssetningen** gir oss sammenhengen mellom "og"-sannsynligheter og "eller"-sannsynligheter. Vi kjenner "og"-sannsynligheten fra oppgave **a**.
Dermed:

$$\underline{\underline{P(S_1 \cup S_2)}} \;=\; P(S_1) + P(S_2) - \overbrace{P(S_1 \cap S_2)}^{=\frac{1}{36}} \tag{33}$$

$$= \frac{1}{6} + \frac{1}{6} - \frac{1}{36} \;=\; \frac{11}{36} \;\approx\; \underline{\underline{0.306}} \tag{34}$$

c) Sannsynligheten for at det snør dag nr. 2 gitt at det snødde dag nr. 1 finnes ved å bruke **multiplikasjonssetningen**:

$$\underline{\underline{P(S_2|S_1)}} \;=\; \frac{P(S_1 \cap S_2)}{P(S_1)} \;=\; \frac{\frac{2}{\cancel{30}}}{\frac{5}{\cancel{30}}} \;=\; \frac{2}{5} \;=\; \underline{\underline{0.4}} \tag{35}$$

d) Begivenheten $S_1 \cap S_2 \cap S_3$ betyr at det snør $\underline{\underline{\text{både dag 1, dag 2 og dag 3}}}$.

[7] Se den "**spesielle multiplikasjonssetning**" i formelsamlingen.

e) **Multiplikasjonssetningen** anvendt på $P(S_1 \cap S_2 \cap S_3)$ gir: [8]

$$\begin{aligned} P(S_1 \cap S_2 \cap S_3) &= P(S_3 \cap S_2 \cap S_1) && (37) \\ &= P(S_3|S_2 \cap S_1)P(S_2 \cap S_1) && (38) \end{aligned}$$

Multiplikasjonssetningen anvendt på $P(S_2 \cap S_1)$ gir:

$$\begin{aligned} \underline{\underline{P(S_1 \cap S_2 \cap S_3)}} &= P(S_3|S_2 \cap S_1)P(S_2 \cap S_1) && (39) \\ &= \underbrace{P(S_3|S_2 \cap S_1)}_{=\,0.6}\underbrace{P(S_2|S_1)}_{=\,0.4}\underbrace{P(S_1)}_{=\,\frac{1}{6}} && (40) \\ &= 0.6 \cdot 0.4 \cdot \frac{1}{6} = \underline{\underline{0.04}} && (41) \end{aligned}$$

hvor $P(S_3|S_2 \cap S_1) = 0.6$ var oppgitt i oppgaven, $P(S_2|S_1) = 0.4$ fant vi i oppgave **d** og $P(S_1) = \frac{1}{6}$ var også oppgitt i oppgaven.

■

[8] I formelsamlingen er **multiplikasjonssetningen** formulert på følgende måte:

$$P(A \cap B) = P(A|B)P(B) \quad (36)$$

I lign.(38) anvender vi denne ligningen med $A = S_3$ og $B = S_2 \cap S_1$.

Oppgave 4: (økonomi)

a) Sannsynligheten for at en tilfeldig valgt førsteårsstudent består 40 sp eller mer:

$$\underline{\underline{P(X \geq 40)}} = P(X = 40) + P(X = 45) + P(X = 50) + P(X = 55) + P(X = 60) \quad (42)$$

$$= 0.06 + 0.06 + 0.10 + 0.10 + 0.23 = \underline{\underline{0.55}} \quad (43)$$

b) Antall førsteårstudenter som vil bestå 40 sp eller mer når det er 250 studenter:

$$250 \cdot P(X \geq 40) = 250 \cdot 0.55 = \underline{\underline{137.5}} \quad (44)$$

c) Sannsynligheten $P(25 \leq X \leq 35)$:

$$\underline{\underline{P(25 \leq X \leq 35)}} = P(X = 25) + P(X = 30) + P(X = 35) \quad (45)$$

$$= 0.03 + 0.04 + 0.04 = \underline{\underline{0.11}} \quad (46)$$

d) Tolkning:
$P(25 \leq X \leq 35)$ er sannsynligheten for at en tilfeldig valgt førsteårsstudent består mellom 25 og 35 studiepoeng, dvs. sannsynligheten for at en tilfeldig valgt førsteårs-student består 25, 30 eller 35 sp.

e) Forventningen $E[\,\overline{X}\,]$:

$$\underline{\underline{E[\,\overline{X}\,]}} \;=\; E\left[\,\frac{1}{n}\left(X_1 + X_2 + \;...\; + X_n\right)\,\right] \tag{47}$$

$$\overset{\text{alltid}}{=} \;\frac{1}{n}\Big(E[X_1] + E[X_2] + \;...\; + \;E[X_n]\Big) \tag{48}$$

$$= \;\frac{1}{n}\Big(\;\underbrace{E[X] + E[X] + \;...\; + E[X]}_{=\; n\cdot E[X]}\;\Big) \tag{49}$$

$$= \;\frac{1}{\cancel{n}}\,\cancel{n}\cdot E[X] \;=\; E[X] \;=\; \underline{\underline{35}} \tag{50}$$

siden $E[X] = 35$, som oppgitt i oppgaven. [9]

f) Tolkning:

$E[\,\overline{X}\,]$ er **forventningen av gjennomsnittlig** antall beståtte studiepoeng i løpet av første studieår for alle n studentene.

■

[9]Legg merke til at vi i lign.(48) bruker regnereglen: $(a, b$ er konstanter$)$

$$E[\,aX + bY\,] \;=\; aE[X] + bE[Y]\;, \tag{51}$$

som man finner i formelsamlingen.

Kapittel 5

Hovedeksamen 2014

LØSNING: Eksamen 9. mai 2014

"*MAT110 Statistikk 1*", 2014

Oppgave 1: (revisjon)

a) Komplementsetningen for $P(\overline{\text{objekt}}|\overline{K})$:

$$\underline{\underline{P(\overline{\text{objekt}}|\overline{K})}} = 1 - \overbrace{P(\text{objekt}|\overline{K})}^{=0.05} = 1 - 0.05 = \underline{\underline{0.95}} \tag{1}$$

b) Formelen for total oppsplitting: (se formelsamlingen side 28 eller 31)

$$\underline{\underline{P(\text{objekt})}} = \overbrace{P(\text{objekt}|\overline{K})}^{=0.05} \overbrace{P(\bar{K})}^{=1-P(K)=0.999} + \overbrace{P(\text{objekt}|K)}^{=0.80} \overbrace{P(K)}^{=0.001} \tag{2}$$

$$= 0.05 \cdot 0.999 + 0.80 \cdot 0.001 = \underline{\underline{0.05075}} \tag{3}$$

c) Bayes formel:

$$\underline{\underline{P(\overline{K}|\text{objekt})}} = \overbrace{P(\text{objekt}|\overline{K})}^{=0.05} \cdot \frac{\overbrace{P(\overline{K})}^{=0.999}}{\underbrace{P(\text{objekt})}_{=0.05075}} = 0.05 \cdot \frac{0.999}{0.05075} = \underline{\underline{0.9842}} \tag{4}$$

d) <u>Kommentar</u>:

Svaret i oppgave **c** sier at det store flertallet av bedrifter, hele 98 %, som blir klassifisert som konkursobjekter faktisk *ikke* går konkurs.
Revisjonsselskapet KPMG bør derfor ikke trekke konklusjoner som baserer seg kun på dette verktøyet. [1]

■

[1] Merknad: (behøver ikke være med i eksamensbesvarelsen)
Når det gjelder hva som gir riktige og hva som gir falske indiksjoner så bør man være presis:
$P(\text{objekt}|\overline{K}) = 0.05$ betyr at verktøyet gir 5 % falske prediksjoner for bedrifter som *ikke går konkurs*.
$P(\overline{K}|\text{objekt}) = 0.9842$ betyr at verktøyet gir 98.42 % falske prediksjoner for bedrifter som er *klassifisert som konkursobjekt*.

Oppgave 2: (logistikk)

a) Siden denne oppgaven dreier seg om antall begivenheter innenfor et gitt tidsintervall (30 minutter), altså en rate, så vil den stokastiske variabelen X beskrives av en [2]

$$\underline{\underline{\text{Poissonfordeling}}} \tag{5}$$

b) I oppgaveteksten opplyses det at det i gjennomsnitt kommer 920 personbiler per 4 timer til fergekaien. Antall biler som kommer hver halvtime λ er derfor:

$$\underline{\underline{\lambda}} \;=\; 920 \cdot \frac{0.5\ \cancel{\text{time}}}{4\ \cancel{\text{timer}}} \;=\; \underline{\underline{115}} \quad , \quad \text{q.e.d.} \tag{6}$$

c) i) Fra oppgave 2**a** og 2**b** vet vi at $X \sim \text{Poi}[\lambda]$, hvor $\lambda = 115$. Forventet antall biler $E[X]$ som ankommer fergekaien mellom to ferger blir derfor: [3]

$$\underline{\underline{E[X]}} \;=\; \lambda \;=\; \underline{\underline{115}} \tag{7}$$

ii) Standardavviket $\sigma[X]$ av antall biler som ankommer fergekaien mellom to ferger er da:

$$\underline{\underline{\sigma[X]}} \;=\; \sqrt{Var[X]} \;=\; \sqrt{\lambda} \;=\; \sqrt{115} \;=\; \underline{\underline{10.72}} \tag{8}$$

[2] Her er ikke "*suksess*"-sannsynligheten p og antall forsøk n oppgitt. Så en binomisk sannsynlighetsfordeling er ikke hensiktsmessig for situasjonen som beskrevet i oppgaven.

[3] Se f.eks. formelsamlingen side 49 eller side 61.

d) i) En diskret Poisson sannsynlighetsfordelingen Poi[λ] kan, under bestemte betingelser, med god tilnærmelse beskrives av en normalfordeling:

$$\text{Poi}[\lambda] \quad \longrightarrow \quad \text{N}[\mu = \lambda, \sigma = \sqrt{\lambda}] \tag{9}$$

ii) Fra side 60 (eller side 58) i formelsamlingen ser vi at dersom

$$\lambda \gtrsim 5 \ , \tag{10}$$

så gjelder påstanden fra oppgave 2**d** i.

iii) Siden $\lambda = 115 \gg 5$ så er betingelsen i lign.(10) oppfylt og tilnærmelsen i lign.(9) gjelder i vårt tilfelle.

e) Ved å bruke resultatet fra oppgave 2**d** i kan vi finne sannsynligheten for at ikke alle bilene får plass i fergen dersom det i utgangspunktet står 20 biler i fergekø: (Oppgitt i oppgaven: Fergene har kapasitet på $X_0 = 125$ biler.)

$$\underline{P(X > X_0 - 20)} = P(X > 125 - 20) \tag{11}$$

$$= 1 - P(X \leq 105) \tag{12}$$

$$\overset{\text{standardiser}}{=} 1 - P\left(\underbrace{\frac{X - E[X]}{\sigma[X]}}_{= Z} \leq \frac{105 - E[X]}{\sigma[X]} \right) \tag{13}$$

$$= 1 - P\left(Z \leq \frac{105 - \lambda}{\sqrt{\lambda}} \right) \tag{14}$$

$$= 1 - P\left(Z \leq \frac{105 - 115}{\sqrt{115}} \right) \tag{15}$$

$$= 1 - P(Z \leq -0.93) \tag{16}$$

$$= \cancel{1} - \left(\cancel{1} - P(Z \leq 0.93) \right) = \underline{P(Z \leq 0.93)} \tag{17}$$

Dermed:

$$\underline{\underline{P(X > X_0 - 20)}} = P(Z \leq 0.93) \tag{18}$$

$$= \underbrace{G(0.93)}_{\text{se tabell}} = \underline{\underline{0.8238}} \tag{19}$$

Dersom det i utgangspunktet er 20 biler i fergekø i tillegg de til som ankommer fergekaien med konstant rate λ så er det $\underline{\underline{\ 82.38\ \%\ }}$ sannsynlighet for at ikke alle bilene får plass i fergen.

f) *Fjord 1* ønsker at det skal være 95 % sannsynlighet for at alle bilene kommer med. La X_{kap} være den ukjente kapasiteten til fergen som vi ønsker å finne. Denne er bestemt av:

$$P(X \leq X_{\text{kap}}) = 0.95 \tag{20}$$

$$P\left(\underbrace{\frac{X - E[X]}{\sigma[X]}}_{\equiv Z} \leq \underbrace{\frac{X_{\text{kap}} - E[X]}{\sigma[X]}}_{\equiv Z_{\text{kap}}}\right) \overset{\text{standardiser}}{=} 0.95 \tag{21}$$

$$P(Z \leq Z_{\text{kap}}) = 0.95 \tag{22}$$

Ved "omvendt tabelloppslag" ser vi at svaret 0.95 ligger midt mellom argumentene $Z_{\text{kap}} = 1.64$ og $Z_{\text{kap}} = 1.65$. Dermed:

$$Z_{\text{kap}} = 1.645 \tag{23}$$

Løser med hensyn på X_{kap} alene:

$$Z_{\text{kap}} = \frac{X_{\text{kap}} - E[X]}{\sigma[X]} \tag{24}$$

$$
\begin{aligned}
\underline{\underline{X_{\text{kap}}}} \quad &= \quad Z_{\text{kap}} \cdot \overbrace{\sigma[X]}^{=\sqrt{\lambda}} \;+\; \overbrace{E[X]}^{=\lambda} && (25)\\
&= \quad Z_{\text{kap}} \cdot \sqrt{\lambda} \;+\; \lambda && (26)\\
&= \quad 1.645 \cdot \sqrt{115} \;+\; 115 \quad = \quad 132.64 \quad \approx \quad \underline{\underline{133}} && (27)
\end{aligned}
$$

For at det skal være 95 % sikkert at alle i fergekøen skal komme med fergen så må fergen ha en kapasitet på $\underline{\underline{\text{133 personbiler}}}$.

■

Oppgave 3: (økonomi)

a) 4 forutsetninger må være oppfylt for at Y skal være binomisk fordelt:

1. Hvert forsøk skal ha 2 mulige utfall, s (suksess) eller f (fiasko):
   Enten så kjøper en passasjer en brus, eller så gjør han/hun det ikke.
2. Det skal være samme sannsynlighet $p = 0.11$ for suksess i alle n forsøkene.
3. Alle forsøk er uavhengige:
   Fergepassasjerene kjøper brus uavhengige av hverandre.
4. Vi gjennomfører et bestemt antall forsøk, n:
   Fra oppgave 2**d** vet vi at det kommer $\lambda = 115$ biler hver halvtime, dvs. mellom to ferger. Selv om antall ankomne biler egentlig er en stokastisk størrelse så antas det i denne oppgaven at det "*faktisk kommer λ antall biler til fergekaien mellom to fergeravganger*". I oppgaven får vi også opplyst at det i utgangspunktet ikke er noen biler i fergekø. I tillegg får vi opplyst at det i gjennomsnitt er 3 passasjerer i hver bil. Derfor er: $n = 3\lambda = 3 \cdot 115 = 345$ er et *bestemt* antall "forsøk".

Alle de 4 forutsetningene for en binomisk fordeling er oppfylt. Derfor er det rimelig å anta at Y er binomisk fordelt, dvs. $\underline{\underline{Y \sim \text{Bin}[\,n = 3\lambda = 345, p = 0.11\,].}}$

b) Siden $Y \sim \text{Bin}[\,n = 3\lambda = 345, p = 0.11\,]$ så er forventet antall solgte brus på en gitt overfart ifølge modell 1:

$$\underline{\underline{E[Y]}} \overset{\text{Bin.}}{=} n \cdot p = 345 \cdot 0.11 = \underline{\underline{37.95}} \tag{28}$$

c) i) Variansen til antall solgte brus på en gitt overfart ifølge modell 1:

$$\underline{\underline{Var[Y]}} \overset{\text{Bin.}}{=} n \cdot p \cdot (1-p) = 345 \cdot 0.11 \cdot (1 - 0.11) \approx \underline{\underline{33.78}} \tag{29}$$

ii) Tilhørende standardavviket $\sigma[X]$:

$$\underline{\underline{\sigma[X]}} = \sqrt{Var[X]} = \sqrt{33.78} = \underline{\underline{5.81}} \tag{30}$$

d) i) Betingelse som må være oppfylt så for at en binomisk fordeling kan tilnærmes med en normalfordeling er: [4]

$$\underline{\underline{n \cdot p\,(1-p)\ \gtrsim\ 5}} \tag{31}$$

ii) For vårt tilfelle:

$$345 \cdot 0.11\,(1-0.11) \;=\; 33.78\ \gg 5 \tag{32}$$

<u>Ja, betingelsen er godt oppfylt for vårt tilfelle.</u>

e) For å regne ut sannsynligheten for at det selges mer enn 45 brus på en overfart så benytter vi at $Y \sim \mathrm{Bin}[\,n, p\,]$ kan tilnærmes med en normalfordeling:

$$\underline{\underline{P(Y > 45)}} \;=\; 1 \;-\; P(Y \leq 45) \tag{33}$$

$$=\; 1 \;-\; P\Big(\underbrace{\frac{Y - E[Y]}{\sigma[Y]}}_{=\,Z} \;\leq\; \frac{45 - E[Y]}{\sigma[Y]}\Big) \tag{34}$$

$$=\; 1 \;-\; P\left(Z \;\leq\; \frac{45-38}{5.81}\right) \tag{35}$$

$$=\; 1 \;-\; P(Z \;\leq\; 1.21) \tag{36}$$

$$=\; 1 \;-\; \underbrace{G(1.21)}_{\overset{\text{tabell}}{=}\ 0.8869} \tag{37}$$

$$=\; 1 \;-\; 0.8869 \;=\; \underline{\underline{0.1131}} \tag{38}$$

[4] Se side 60 i formelsamlingen.

f) Sannsynligheten at en tilfeldig valgt passasjer kjøper mer enn en brus:

$$\underline{\underline{P(B > 1)}} = P(B = 2) + P(B = 3) \tag{39}$$

$$= 0.02 + 0.01 = \underline{\underline{0.03}} \tag{40}$$

Alternativt kan denne løses på følgende måte:

$$\underline{\underline{P(B > 1)}} = 1 - P(B \leq 1) \tag{41}$$

$$= 1 - \Big(P(B = 0) + P(B = 1) \Big) \tag{42}$$

$$= 1 - 0.93 - 0.04 = \underline{\underline{0.03}} \tag{43}$$

Det er nok at oppgaven løses på en måte.

g) Forventet antall brus solgt totalt i løpet av en overfart: $\quad (\; n = 3\lambda = 345 \;)$

$$\underline{\underline{E[B_{\text{tot}}]}} = E[\, B_1 + B_2 + \; ... \; + B_n \,] \tag{44}$$

$$\overset{\text{alltid}}{=} \underbrace{E[B_1] + E[B_2] + \; ... \; + E[N_n]}_{n \; = \; 345} = n \cdot \underbrace{E[B]}_{=0.11} = 345 \cdot 0.11 = \underline{\underline{37.95}} \tag{45}$$

NB: Overgangen i lign.(44) til (45) gjelder alltid. Uansett om de stokastiske variablene B_i er uavhengige eller ikke.

h) Variansen til antall brus solgt totalt i løpet av en overfart: $\quad (\ n = 3\lambda = 345 \ )$

$$\underline{\underline{Var[B_{\text{tot}}]}} = Var[\,B_1 + B_2 + \ \dots \ + B_n\,] \tag{46}$$

$$\overset{\text{uavhengig}}{=} Var[B_1] + Var[B_2] + \ \dots \ + Var[B_n] \tag{47}$$

$$= n \cdot \underbrace{Var[B]}_{=\,\sigma^2[B]} = n \cdot \sigma^2[B] = 345 \cdot 0.4449^2 \approx \underline{\underline{68.28}} \tag{48}$$

NB: Overgangen i lign.(46) til (47) gjelder kun dersom de stokastiske variablene B_i er uavhengige.

i) Siden

1. fergepassasjerene kjøper brus uavhengig av hverandre:
   $B_i \sim$ er uavhengige for alle $i = 1, 2, 3, ..., n$
2. alle bruskjøpere har samme sannsynlighetsfordeling for B_i:
   $B_i \sim$ samme sannsynlighetsfordeling for alle $i = 1, 2, 3, ..., n$
3. antall "forsøk", dvs. potensielle bruskjøpere, $n = 3\lambda = 345$, er tilstrekkelig stort [5]

så gjelder **sentralgrensesetningen**. Dermed er B_{tot} normalfordelt med god tilnærmelse:

$$\underline{\underline{B_{\text{tot}} \sim \mathrm{N}\big[\,E[B_{\text{tot}}], Var[B_{\text{tot}}]\,\big]}} \tag{49}$$

[5] Husk: Antall forsøk n for at sentralgrensesetningen skal gjelde er avhengig av situasjonen. Men en **tommelfingerregel** er at vi bør ha $n \gtrsim 30$.

j) For å regne ut sannsynligheten for at det selges mer enn 45 brus på en overfart så benytter vi at B_{tot} kan tilnærmes med en normalfordeling:

$$\underline{\underline{P(B_{\text{tot}} > 45)}} = 1 - P(B_{\text{tot}} \leq 45) \tag{50}$$

$$= 1 - P\Bigg(\underbrace{\frac{B_{\text{tot}} - E[B_{\text{tot}}]}{\sigma[B_{\text{tot}}]}}_{=Z_{\text{tot}}} \leq \frac{45 - E[B_{\text{tot}}]}{\sigma[B_{\text{tot}}]}\Bigg) \tag{51}$$

$$= 1 - P(Z_{\text{tot}} \leq \frac{45 - 37.95}{\sqrt{68.28}}) \tag{52}$$

$$= 1 - P(Z_{\text{tot}} \leq 0.85) \tag{53}$$

$$= 1 - \underbrace{G(0.85)}_{\text{se tabell}} \tag{54}$$

$$= 1 - 0.8023 = \underline{\underline{0.1977}} \tag{55}$$

k) I modell 1 kjøper fergepassasjerene enten ingen brus eller en brus.
I modell 2, derimot, tas det høyde for at fergepassasjerene kan kjøpe mer enn en brus.
Derfor er det rimelig at variansen til modell 2 er større enn variansen til modell 1:

$$\underbrace{Var[B_{\text{tot}}] = 68.28}_{\text{modell 2}} > \underbrace{Var[Y] = 33.78}_{\text{modell 1}} \tag{56}$$

l) Forventet fortjeneste $E[F]$ på brussalget til *Fjord1* for en overfart: [6]

$$\underline{E[F]} \;=\; E[\,(p-k)B_{\text{tot}} - k_{\text{fast}}\,] \;=\; \underline{(p-k)E[B_{\text{tot}}] - k_{\text{fast}}} \tag{57}$$

I oppgaven er det opplyst at $p = 20$ NOK, $k = 6$ NOK og $k_{\text{fast}} = 175$ NOK.
I tillegg fant vi i oppgave 3**i** at $E[B_{\text{tot}}] = 37.95$.
Dermed:

$$\underline{\underline{E[F]}} \;=\; \big(\,(20-6)\cdot 37.95 - 175\,\big)\text{ NOK} \;=\; \underline{\underline{356.30\;\;\text{NOK}}} \tag{58}$$

■

[6] Bruk gjerne regnereglene på side 37 i formelsamlingen.

Oppgave 4: (økonomi)

a) Vi bruker læresetningen på side 66 i formelsamlingen for å finne minste kvadraters **regresjonslinje** for x og y. Parametrene $\hat{\beta}$ og $\hat{\alpha}$ er da: (dropper benevning her)

$$\underline{\hat{\beta}} = \frac{S_{xy}}{S_x^2} = \frac{18}{2.5} = \underline{7.2} \tag{59a}$$

$$\underline{\hat{\alpha}} = \overline{y} - \beta\overline{x} = 284 - 7.2 \cdot 3 = \underline{262.4} \tag{59b}$$

Minste kvadraters lineære regresjonslinje $\hat{y} = \hat{\alpha} + \hat{\beta}\, x$ blir dermed:
(se lign.(11.8) i formelsamlingen)

$$\underline{\underline{\hat{y} = 262.4 + 7.2 \cdot x}} \tag{60}$$

b) i) En regresjonslinje mellom variablene x og y sier at for en gitt verdi av x så kan y estimeres/predikeres via regresjonslinjen.

ii) Parameteren $\hat{\beta}$ er stigningstallet for en lineær regresjonslinje.
Parameteren angir estimert/predikert endring $\hat{y}$ når x endres med èn enhet.

iii) Parameteren $\hat{\beta} = 7.2$ i oppgave **a**.
Det betyr at aksjeprisen til Norwegian øker med 7.2 NOK per dag.

c) Regresjonslinjen i lign.(60) predikerer at aksjekursen til Norwegian er:

$$\underline{\underline{\hat{y}(12)}} = (262.4 + 7.2 \cdot 12)\ \text{NOK} = \underline{\underline{348.8\ \ \text{NOK}}} \tag{61}$$

etter 12 dager.

d) i) Forklaringskraften R^2 er normallisert til: $0 \leq R^2 \leq 1$.

ii) Forklaringskraften R^2 er enhets<u>u</u>avhengig, den er benevningsløs.

iii) Forklaringskraften R^2 er et mål på:

- i hvor stor grad x_i kan brukes til å **forutsi** y_i
- hvor godt de faktiske observasjonene y_i **passer** med lineær regresjonen $\hat{y}_i$
- styrken i **samvariasjon** mellom prediksjonene $\hat{y}_i$ og de faktiske obervasjonene y_i

Det er nok at man nevner èn eller lignende av disse kulepunkt-kommentarene.

e) Forklaringskraften R^2 kan leses direkte fra Excel-utskriften:
(Se cellen som heter "*R Square*" i Excel-utskriften): [7]

$$\underline{\underline{R^2 = 0.9744}} \tag{62}$$

f) Kommentar til svaret i oppgave **e**:

At $R^2 = 0.9744$ betyr at for en gitt dag innenfor perioden hvor den lineære trenden gjelder så kan vi "i stor grad", tilsvarende 97.44 %, forutsi aksjekursen.
Vi sier at modellen har <u>stor forklaringskraft</u>.

[7] Man kan også regne ut forklaringskraften R^2 "for hånd" via definisjonen $R^2 = 1 - SSE/SST$. Men det er mye mer arbeidskrevende. Fint at dataprogrammer (som f.eks. Excel) kan hjelpe oss med slikt.

Kapittel 6

Kontinuasjonseksamen 2014

Høgskolen i Molde
Vitenskapelig høgskole i logistikk

LØSNING: Eksamen 8. januar 2015

"*MAT110 Statistikk 1*", høst 2014

Oppgave 1: (økonomi)

a) Matematisk formulering ihht. notasjon:

$$P(G) = 0.75 \tag{1}$$
$$P(D) = 0.25 \tag{2}$$

b) Matematisk formulering ihht. notasjon:

$$P(L|G) = 0.10 \tag{3}$$
$$P(L|D) = 0.50 \tag{4}$$

c) Vi skal finne $P(L)$ når vi kjenner sannsynlighetene i oppgave **a** og **b**.
Da kan vi bruke oppsplitting av utfallsrom:

$$\underline{P(L)} \overset{\text{oppspl.}}{=} P(L|G)P(G) + P(L|D)P(D) \tag{5}$$

$$= 0.10 \cdot 0.75 + 0.50 \cdot 0.25 = \underline{0.20} \tag{6}$$

Det er $\underline{\underline{20\ \%}}$ av kundemassen som har lav inntekt.

d) Vi skal finne $P(D|L)$. Siden vi vet $P(L|D)$ og de tilhørende ubetingede sannsynlighetene så kan vi bruke **Bayes lov**:

$$\underline{P(D|L)} \overset{\text{Baye}}{=} P(L|D)\,\frac{P(D)}{P(L)} \tag{7}$$

$$= 0.50 \cdot \frac{0.25}{0.20} = \underline{\;0.625\;} \tag{8}$$

Sannsynligheten for at en kunde med lav inntekt har **dårlig** betalingsevne er $\underline{\underline{62.5\ \%}}$.

e) Vi skal finne $P(G|L)$. Vi kjenner $P(D|L) = 0.625$ fra oppgave **d**.
Dermed kan vi bruke **komplement**setningen:

$$\underline{P(G|L)} \overset{\text{komp.}}{=} 1 - P(D|L) \tag{9}$$

$$= 1 - 0.625 = \underline{\;0.375\;} \tag{10}$$

Sannsynligheten for at en kunde med lav inntekt har **god** betalingsevne er $\underline{\underline{37.5\ \%}}$.

f) Forventet fortjeneste $E[F]$ for en kunde med lav inntekt er:
($f_G = 6000$ NOK og $f_D = -4000$ NOK)

$$\underline{\underline{E[F]}} = \sum_{X=G,D} f_X \cdot P(X|L) \tag{11}$$

$$= f_G \cdot P(G|L) + f_D \cdot P(D|L) \tag{12}$$

$$= (\,6000 \cdot 0.375 - 4000 \cdot 0.625\,)\,\text{NOK} = \underline{\underline{-250\ \text{NOK}}} \tag{13}$$

Man behøver ikke gjøre det så formelt som ovenfor. Dette er også OK:

$$\underline{\underline{E[F]}} = 6000 \cdot P(G|L) - 4000 \cdot P(D|L) \tag{14}$$

$$= (\,6000 \cdot 0.375 - 4000 \cdot 0.625\,)\,\text{NOK} = \underline{\underline{-\,250\ \text{NOK}}} \tag{15}$$

g) Siden forventet fortjeneste $E[F]$ for en kunde med lav inntekt er negativ så **taper** banken på denne kundegruppen.

I det lange løp er det derfor $\underline{\underline{\text{ikke lønnsomt}}}$ å gi lån til kunder med av inntekt.

■

Oppgave 2: (petroleumslogistikk)

a) Siden $X \sim \text{Poi}[X]$ og 1 "hendelse" per uke:

$$\underline{P(X=1)} \quad = \quad \frac{\lambda^1}{1!}\, e^{-\lambda} \tag{16}$$

$$\stackrel{\lambda=0.6}{=} \quad \frac{0.6^1}{1!}\, e^{-0.6} \quad \stackrel{\text{kalkis}}{=} \quad \underline{0.3293} \tag{17}$$

Sannsynligheten for at det skjer 1 "hendelse" per uke er $\underline{\underline{32.93\,\%}}$.

b) Siden $X \sim \text{Poi}[X]$ og mer enn 1 "hendelse" per uke:

$$\underline{P(X>1)} \quad = \quad 1 \; - \; P(X \leq 1) \tag{18}$$

$$= \quad 1 \; - \; \Big(P(X=0) \; + \; P(X=1) \Big) \tag{19}$$

$$= \quad 1 \; - \; \frac{\lambda^0}{0!}\, e^{-\lambda} \; - \; \frac{\lambda^1}{1!}\, e^{-\lambda} \tag{20}$$

$$\stackrel{\lambda=0.6}{=} \quad 1 \; - \; \frac{0.6^0}{0!}\, e^{-0.6} \; - \; \frac{0.6^1}{1!}\, e^{-0.6} \tag{21}$$

$$\stackrel{\text{kalkis}}{=} \quad 1 \; - \; 0.5488 \; - \; 0.3293 \tag{22}$$

$$= \quad \underline{0.1219} \tag{23}$$

Sannsynligheten for at det skjer mer enn 1 "hendelse" per uke er $\underline{\underline{12.19\,\%}}$.

c) Forventet antall “hendelser” per måned:

$$\underline{\underline{E[Y]}} = E[X_1 + X_2 + X_3 + X_4] \tag{24}$$

$$\overset{\text{alltid}}{=} E[X_1] + E[X_2] + E[X_3] + E[X_3] = \overbrace{\lambda + \lambda + \lambda + \lambda}^{\text{4 stk.}} = 4\,\lambda \tag{25}$$

$$= 4 \cdot 0.6 = \underline{\underline{2.4}} \tag{26}$$

NB: Overgangen i lign.(24) til (25) gjelder alltid. Uansett om de stokastiske variablene X_i er uavhengige eller ikke.
(Se lign.(5.12) på side 37 i formelsamlingen fra 2014).

d) En sum av Poisson fordelinger $X_1 + X_2 + X_3 + X_4$ er også Poisson fordelt, dvs. Y er Poisson fordelt.
Den stokastiske variabelen Y er da Poisson fordelt med forventning $\lambda_Y = E[Y] = 2.4$.
Sannsynligheten for at det skjer *mer enn* 1 “hendelse” per måned er da:

$$\underline{\underline{P(Y > 1)}} = 1 - P(Y \leq 1) \tag{27}$$

$$= 1 - \Big(P(Y = 0) + P(Y = 1) \Big) \tag{28}$$

$$= 1 - \frac{\lambda_Y^0}{0!}\, e^{-\lambda_Y} - \frac{\lambda_Y^1}{1!}\, e^{-\lambda_Y} \tag{29}$$

$$\overset{\lambda_Y = 2.4}{=} 1 - \frac{2.4^0}{0!}\, e^{-2.4} - \frac{2.4^1}{1!}\, e^{-2.4} \tag{30}$$

$$\overset{\text{kalkis}}{=} 1 - 0.091 - 0.2177 \tag{31}$$

$$= \underline{\underline{0.6916}} \tag{32}$$

e) Siden

$$P(Y > 1) = 0.6916 \quad \text{og} \quad P(X > 1) = 0.1219 \tag{33}$$

så ser vi at $P(Y > 1)$ er mye større enn $P(X > 1)$. Siden Y beskriver en periode på 1 måned mens X beskriver kun 1 uke så er virker det rimelig fornuftig at:

$$P(Y > 1) \gg P(X > 1) \tag{34}$$

NB:
Selv om Y beskriver en periode på 1 måned og X kun for 1 uke, så kan vi IKKE si at $\underbrace{P(Y > 1) = 4 \cdot P(X > 1)}_{\text{galt}}$.

∎

Oppgave 3: (økonomi og logistikk)

a) Siden

$$\underline{\underline{\sum_{i=0}^{n} p_i}} \quad = \quad p_0 + p_1 + p_2 \quad = \quad 0.95 + 0.03 + 0.02 \quad = \quad \underline{\underline{1}} \tag{35}$$

så er den oppgitte sannsynlighetsfordelingen en **gyldig** fordeling.

b) Forventning $E[X]$:

$$\underline{\underline{E[X]}} \quad \stackrel{\text{def.}}{=} \quad \sum_{i=0}^{3} x_i \cdot P(X = x_i) \tag{36}$$

$$= \quad 0 \cdot \underbrace{P(X=0)}_{=0.95} \quad + \quad 1 \cdot \underbrace{P(X=1)}_{=0.03} \quad + \quad 2 \cdot \underbrace{P(X=2)}_{=0.02} \tag{37}$$

$$= \quad 0 \cdot 0.95 \quad + \quad 1 \cdot 0.03 \quad + \quad 2 \cdot 0.02 \quad = \quad \underline{\underline{0.07}} \tag{38}$$

Tolking:

$\underline{\underline{E[X] = \text{ forventet antall produksjonsfeil for en tilfeldig valgt ananasbrus}}}$

c) For å finne variansen $Var[X]$ så regner vi først ut $E[X^2]$:

$$\underline{E[X^2]} \quad \stackrel{\text{def.}}{=} \quad \sum_{i=0}^{3} x_i^2 \cdot P(X = x_i) \tag{39}$$

$$= \quad 0^2 \cdot \underbrace{P(X=0)}_{=0.95} \quad + \quad 1^2 \cdot \underbrace{P(X=1)}_{=0.03} \quad + \quad 2^2 \cdot \underbrace{P(X=2)}_{=0.02} \tag{40}$$

$$= \quad 0^2 \cdot 0.95 \quad + \quad 1^2 \cdot 0.03 \quad + \quad 2^2 \cdot 0.02 \quad = \quad \underline{0.11} \tag{41}$$

Dette innsatt i setningen for varians: (se formelsamling) [1]

$$\underline{\underline{Var[X]}} \quad = \quad E[X^2] - E[X]^2 \quad = \; 0.11 - 0.07^2 \; = \; \underline{\underline{0.1051}} \tag{42}$$

Tolking:

$$\underline{\underline{\begin{aligned} Var[X] = \; & \text{forventet variasjon/spredning i antall produksjonsfeil} \\ & \text{for en tilfeldig valgt ananasbrus} \end{aligned}}} \tag{43}$$

d) Forventet inntekt per dag på ananasbrus:

$$\underline{E[\,I\,]} \quad = \quad E[\,p\,n\,0.95 \; - \; b\,n\,X\,] \tag{44}$$

$$= \quad p\,n\,0.95 \; - \; b\,n \, \underbrace{E[\,X\,]}_{=\;0.07} \tag{45}$$

$$= \quad (\,15 \cdot 4500 \cdot 0.95 \; - \; 23 \cdot 4500 \cdot 0.07\,)\ \text{NOK} \quad = \quad \underline{56\,880\ \text{NOK}} \tag{46}$$

Forventet inntekt per dag på ananasbrus er $\underline{\underline{56\,880\ \text{NOK}}}$.

e) i) Forventet antall produksjonsfeil av ananasbrus per dag:

$$\underline{\underline{E[\,Y\,]}} \quad = \quad E\,[\,X_1 + X_2 + \; ... \; + X_n\,] \quad = \quad \overbrace{E[X_1] + E[X_2] + \; ... \; + E[X_n]}^{\text{n stk.}} \tag{47}$$

$$= \quad n\,E[X] \quad = \quad 4500 \cdot 0.07 \quad = \quad \underline{\underline{315}} \tag{48}$$

NB: Overgangen i lign.(47) gjelder alltid. Uansett om de stokastiske variablene X_i er uavhengige eller ikke.

[1] Siden sannsynlighetensfordelingen $P(X)$ er så liten, se tabellen i oppgaveteksten, så kunne vi også lett regnet ut $Var[X]$ via definisjonen.

ii) Variansen til antall produksjonsfeil av ananasbrus:

$$\underline{\underline{Var[Y]}} = Var[X_1 + X_2 + \ ... \ + X_n] \tag{49}$$

$$\overset{\text{uavh.}}{=} \overbrace{Var[X_1] + Var[X_2] + \ ... \ + Var[X_n]}^{\text{n stk.}} \tag{50}$$

$$= \ n\,Var[X] \ = \ 4500 \cdot 0.1051 \ = \ \underline{\underline{472.95}} \tag{51}$$

NB: Overgangen i lign.(49) gjelder *kun* dersom de stokastiske variablene X_i er uavhengige.

f) i) Med forutsetningene som formulet i oppgaven så gjelder sentralgrensesetningen.

ii) Ifølge sentralgrensesetningen er da den stokastiske varibalen Y **normal**fordelt:

$$\underline{\underline{Y \ \sim \ \mathrm{N}[\,E[Y]\,, Var[Y]\,]}} \tag{52}$$

iii) En **tommelfingerregel** for at sentralgrensesetningen skal gjelde er:

$$\underline{\underline{n \ \gtrsim \ 30}} \tag{53}$$

dvs. antall forsøk bør være ca. 30 eller mer.

g) Siden Y er (tilnærmet) normalfordelt så kan vi bruke dette når vi skal regne ut sannsynligheten $P(Y > 350)$: [2]

$$\underline{P(Y > 350)} \overset{\text{standardiser}}{=} P\left(\underbrace{\frac{Y - E[Y]}{\sigma[\,Y\,]}}_{=\,Z} > \frac{350 - E[X]}{\sigma[X]}\right) \tag{54}$$

$$= \quad P\left(Z \ > \ \frac{350 - 315}{\sqrt{472.95}}\right) \tag{55}$$

$$= \quad P(\,Z \ > \ 1.61\,) \tag{56}$$

$$= \quad 1 \ - \ P(\,Z \ \leq \ 1.61\,) \tag{57}$$

$$= \quad 1 \ - \ \underbrace{G(\,1.61\,)}_{=\,0.9463} \tag{58}$$

$$\overset{\text{tabell}}{=} \quad 1 \ - \ 0.9463 \tag{59}$$

$$= \quad \underline{0.0537} \tag{60}$$

Sannsynligheten for at antall produksjonsfeil når et uakseptabelt nivå, dvs. mer enn 350, er $\underline{\underline{5.37\ \%}}$.

■

[2] I oppgaven stod det at vi **ikke** behøver heltallskorreksjon. Derfor utelater vi det.

Oppgave 4: (logistikk)

a) Regresjonsanalyse er teori og metoder for å analysere og utnytte samvariasjon mellom variable.

b) Setningen for minste kvadraters lineære regresjonslinje: [3]

La $(x_1, y_2), (x_2, y_2), ..., (x_n, y_n)$ være observasjonspar/datasett.
Minste kvadraters lineære regresjonslinje er da gitt ved:

$$\boxed{\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta}\,x \;, \qquad (61)}$$

hvor

$$\boxed{\begin{aligned} \hat{\beta} \;&=\; \frac{S_{xy}}{S_x^2} && (62) \\ \hat{\alpha} \;&=\; \overline{y} \;-\; \hat{\beta}\,\overline{x} && (63) \end{aligned}}$$

og

$$S_x^2 \;=\; \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \quad , \quad S_{xy} \;=\; \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \tag{64}$$

samt $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ og $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

c) i) Forklaringskraften R^2 er normallisert til: $0 \leq R^2 \leq 1$.

ii) Forklaringskraften R^2 er enhetsuavhengig, den er benevningsløs.

iii) Forklaringskraften R^2 er et mål på:

- i hvor stor grad x_i kan brukes til å **forutsi** y_i
- hvor godt de faktiske observasjonene y_i **passer** med lineær regresjonen $\hat{y}_i$
- styrken i **samvariasjon** mellom prediksjonene $\hat{y}_i$ og de faktiske obervasjonene y_i

Det er nok at man nevner èn eller lignende av disse kulepunkt-kommentarene.

[3] Se side 66 i formelsamlingen fra 2014.

d) Gjenneomsnittene x og y er:

$$\underline{\underline{\overline{x}}} \quad = \quad \frac{1}{n}\sum_{i=1}^{n} x_i \tag{65}$$

$$= \quad \frac{1}{5}\left(2009 + 2010 + 2011 + 2012 + 2013\right) \;=\; \underline{\underline{2011}} \tag{66}$$

$$\underline{\underline{\overline{y}}} \quad = \quad \frac{1}{n}\sum_{i=1}^{n} y_i \tag{67}$$

$$= \quad \frac{1}{5}\left(228 + 240 + 245 + 248 + 258\right) \text{ tusen tonn} \;=\; \underline{\underline{243.8 \text{ tusen tonn}}} \tag{68}$$

Nå bruker vi læresetningen på side 66 i formelsamlingen fra 2014 for å finne minste kvadraters **regresjonslinje** for x og y.

Parametrene $\hat{\beta}$ og $\hat{\alpha}$ er da: (dropper benevningen)

$$\underline{\hat{\beta}} \quad = \quad \frac{S_{xy}}{S_x^2} \;=\; \frac{17}{2.5} \;=\; \underline{6.8} \tag{69a}$$

$$\underline{\hat{\alpha}} \quad = \quad \overline{y} \;-\; \beta\overline{x} \;=\; 243.8 - 6.8 \cdot 2011 \;=\; \underline{-13\,431} \tag{69b}$$

Minste kvadraters lineære regresjonslinje $\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta}\,x$ blir dermed:
(se lign.(11.8) i formelsamlingen fra 2014)

$$\underline{\underline{\hat{y} \;=\; -13\,431 \;+\; 6.8 \cdot x}} \tag{70}$$

e) Bruker regresjonslinjen $\hat{y} = \hat{y}(x)$ fra oppgave **d**, dvs. lign.(70):

$$\underline{\hat{y}(2017)} \quad = \quad -13\,431 \;+\; 6.8 \cdot 2017 \;=\; \underline{284.6} \tag{71}$$

Regresjonslinjen predikerer at etterspørselen av slurry vil være $\underline{\underline{\text{284.6 tusen tonn}}}$ i 2017.

f) Parameteren $\hat{\beta}$ er stigningstallet for en lineær regresjonslinje.
Parameteren angir estimert/predikert endring $\hat{y}$ når x endres med èn enhet.
Siden (se lign.(69a))

$$\hat{\beta} = 6.8 \tag{72}$$

så innser vi at den gjennomsnittlige etterspørselen av slurry øker med $\underline{\underline{\text{6.8 tusen tonn}}}$ i året.

g) Gjennomsnittelig årlig etterspørsel på 305 tusen tonn inntreffer, ifølgede regresjonslinjen, når $\hat{y}(x) = 305$:

$$305 \quad = \quad -13\,431 \;+\; 6.8 \cdot x \tag{73}$$

og løser med hensyn på x:

$$\underline{x} \;=\; \frac{305 \;+\; 13\,431}{6.8} \;=\; \underline{2020} \tag{74}$$

Firmaet må ha større tank i $\underline{\underline{2020}}$ dersom kapasiteten på lagertanken ikke skal være en begrensende faktor.

h) Forklaringskraften R^2 kan leses direkte fra Excel-utskriften:
(Se cellen som heter "*R Square*" i Excel-utskriften): [4]

$$\underline{\underline{R^2 \;=\; 0.9538}} \tag{75}$$

i) Kommentar til svaret i oppgave **h**:

At $R^2 = 0.9538$ betyr at for en gitt dag innenfor perioden hvor den lineære trenden gjelder så kan vi "i stor grad", tilsvarende 95.38 %, forutsi etterspørselen av slurry.
Modellen har $\underline{\underline{\text{stor forklaringskraft}}}$.

■

[4] Man kan også regne ut forklaringskraften R^2 "for hånd" via definisjonen $R^2 = 1 - SSE/SST$. Men det er mye mer arbeidskrevende. Fint at dataprogrammer (som f.eks. Excel) kan hjelpe oss med slikt.

Kapittel 7

Hovedeksamen 2015

LØSNING: Eksamen 28. mai 2015

"*MAT110 Statistikk 1*", vår 2015

Oppgave 1: (revisjon)

a) Situasjonen som beskrives i oppgaven kan modelleres med en urne.
I denne urnen er fordelingen kjent, M antall bilag med feil og $N = 100\,000$ bilag totalt.
Den stokastiske variabelen X har derfor en hypergeometrisk fordeling.

b) Siden X er hypergeometrisk fordelt så er forventingen:

$$E[X] \;=\; n\,\frac{M}{N} \tag{1}$$

Løser med hensynpå M alane og får:

$$\underline{\underline{M}} \;=\; \frac{E[X]N}{n} \;=\; \frac{10 \cdot 100\,000}{1000} \;=\; \underline{\underline{1000}} \quad , \qquad \text{q.e.d.} \tag{2}$$

c) Siden X er hypergeometrisk fordelt så er variansen:

$$\underline{Var[X]} \;=\; \frac{N-n}{N-1}\,n\,\frac{M}{N}\left(1-\frac{M}{N}\right) \tag{3}$$

$$=\; \frac{100\,000-1000}{100\,000-1}\,1000\,\frac{1000}{100\,000}\left(1-\frac{1000}{100\,000}\right) \;=\; \underline{9.8} \tag{4}$$

og tilhørende standardavvik:

$$\underline{\underline{\sigma[X]}} \ = \ \sqrt{Var[X]} \ = \ \sqrt{9.8} \ = \ \underline{\underline{3.13}} \quad , \quad \text{q.e.d.} \tag{5}$$

d) i) En hypergeometrisk fordeling kan med god tilnæremelse beskrives av en normalfordeling dersom:

$$N \ \gtrsim \ 20 \cdot n \tag{6}$$

$$n \, \frac{M}{N} \left(1 - \frac{M}{N} \right) \ \gtrsim \ 5 \tag{7}$$

ii) For vårt tilfelle er:

$$100\,000 \ \gtrsim \ 20 \cdot 1000 \tag{8}$$

$$1000 \, \frac{1000}{100\,000} \left(1 - \frac{1000}{100\,000} \right) \ \gtrsim \ 5 \tag{9}$$

som gir

$$100\,000 \ \gtrsim \ 20\,000 \tag{10}$$

$$9.9 \ \gtrsim \ 5 \tag{11}$$

dvs. kritereiene er oppfylte i vårt tilfelle.

e) og **g**) Se vedlegg.

f) I oppgaven er det oppgitt at den stokastiske variabelen X er normalfordelt med $X \sim \mathrm{N}\big[\, E[X] = 10, \sigma[X]\,\big]$. Dermed løses denne oppgaven med (omvendt) tabelloppslag [1]. Standardiserer:

$$P(X \geq g) \quad = \quad 0.10 \tag{12}$$

$$P\Bigg(\underbrace{\frac{X - E[X]}{\sigma[x]}}_{\equiv Z} \geq \underbrace{\frac{g - E[X]}{\sigma[X]}}_{\equiv Z_0}\Bigg) \quad \overset{\text{standardiser}}{=} \quad 0.10 \tag{13}$$

$$P(Z \geq Z_0) \quad = \quad 0.10 \tag{14}$$

Ulikheten skal være "rett vei", altså mindre enn eller lik:

$$P(Z \geq Z_0) \;=\; 0.10 \tag{15}$$

$$1 - P(Z < Z_0) \;=\; 0.10 \tag{16}$$

$$P(Z < Z_0) \;=\; 0.90 \tag{17}$$

Ved "omvendt tabelloppslag" ser vi at Z_0 må være 1.28: [2]

$$Z_0 \;=\; 1.28 \tag{19}$$

Vi løser

$$Z_0 \;=\; \frac{g - E[X]}{\sigma[X]} \tag{20}$$

med hensyn på g: ($E[X] = 10$ og $\sigma[X] = 3.13$ fra oppgave **c**)

$$\underline{g} \;=\; E[X] \;+\; Z_0 \cdot \sigma[X] \tag{21}$$

$$=\; 10 \;+\; 1.28 \cdot 3.13 \;=\; \underline{14} \tag{22}$$

[1]Husk at tabellen på side 65 i 2015-formelsamlingen dreier seg KUN om normalfordeling.

[2]For en kontinuerlig sannsynlighetsfordeling er sannsynligheten i et punkt lik null. Dermed kan vi skrive:

$$P(Z \leq Z_0) = P(Z < Z_0) + \underbrace{P(Z = Z_0)}_{=0} = P(Z < Z_0) \tag{18}$$

Dermed kan vi finne fortsatt bruke tabellen i formelsamlingen selv om tabellen sier noe om $G(Z_0) \equiv P(Z \leq Z_0)$ mens vi skal finne $P(Z < Z_0)$.

Dette betyr at revisoren underkjenner regnskapet dersom han finner 14 feil eller mer blant de $n = 1000$ bilagene han trekker.

$\blacksquare$

Oppgave 2: (økonomi / logistikk)

a) Siden den stokastiske variabelen D_1 er oppgitt i oppgaven til å være normalfordelt så finner vi sannsynligheten $P(D_1 > 21\,000)$ ved standardisering og tilhørende tabelloppslag:

$$\underline{P(D_1 > 21\,000)} \overset{\text{standardiser}}{=} P\Bigg(\underbrace{\frac{D_1 - \mu_1}{\sigma_1}}_{=\,Z} > \frac{21\,000 - \mu_1}{\sigma_1}\Bigg) \tag{23}$$

$$= P\left(Z > \frac{21\,000 - 15\,000}{3000}\right) \tag{24}$$

$$= P(Z > 2) \tag{25}$$

$$= 1 - P(Z \leq 2) \tag{26}$$

$$= 1 - \underbrace{G(2)}_{=\,0.9772} \tag{27}$$

$$\overset{\text{tabell}}{=} 1 - 0.9772 \tag{28}$$

$$= \underline{0.0228} \tag{29}$$

Sannsynligheten for at mer enn 21 000 trofaste aviskjøpere kjøper Dagbladet en gitt dag er $\underline{\underline{2.28\ \%}}$.

b) Det er oppgitt i oppgaven at den totale etterspørselen D_{tot} av aviser dersom det har skjedd en spesiell nyhet er:

$$D_{tot} = D_1 + D_2 \tag{30}$$

Ved å bruke formel (5.13) i formelsamlingen fra 2015 finner vi da:

$$\underline{\underline{E[D_{tot}]}} = E[D_1 + D_2] \tag{31}$$

$$\overset{\text{alltid}}{=} \underbrace{E[D_1]}_{=\mu_1} + \underbrace{E[D_2]}_{=\mu_2} = \mu_1 + \mu_2 = 15\,000 + 18\,000 = \underline{\underline{33\,000}} \tag{32}$$

NB: Overgangen i lign.(31) til (32) gjelder alltid. Uansett om de stokastiske variablene X_i er uavhengige eller ikke.

c) Variansen til den totale etterspørselen D_{tot} av aviser dersom det har skjedd en spesiell nyhet er:

$$\underline{\underline{Var[\,D_{tot}\,]}} = Var\,[\,D_1 + D_2\,] \tag{33}$$

$$\overset{\text{uavh.}}{=} Var[D_1] + Var[D_2] \tag{34}$$

$$= \sigma_1^2 + \sigma_2^2 = 3000^2 + 4000^2 = \underline{\underline{25\,000\,000}} \tag{35}$$

NB: Overgangen fra lign.(33) til lign.(34) gjelder *kun* dersom de stokastiske variablene D_1 og D_2 er <u>u</u>avhengige.

d) Forventede fortjenesten til Dagbladet $E[F]$ per dag dersom en spesiell nyhet har skjedd:

$$\underline{\underline{E[F]}} = E\Big[\,(p-k)D_{tot} \;-\; c\,\Big] \tag{36}$$

$$= (p-k)E[D_{tot}] \;-\; c \;=\; \Big(\,(25-12)\cdot 33\,000 - 400\,000\,\Big)\text{ NOK} \tag{37}$$

$$= \underline{\underline{29\,000\text{ NOK}}} \tag{38}$$

Her har vi brukt at $E[D_{tot}] = 33\,000$ fra oppgave **2b**, se lign.(32).

e) Siden D_{tot} er oppgitt i oppgaven til å være normalfordelt så finner vi sannsynligheten $P(D_{tot} > 21\,000)$ ved standardisering og tilhørende tabelloppslag: (bruker at $\sigma[D_{tot}] = \sqrt{Var[D_{tot}]} = \sqrt{25\,000\,000} = 5000$ fra oppgave **2c**)

$$\underline{P(D_{tot} > 21\,000|S)} \overset{\text{standardiser}}{=} P\left(\underbrace{\frac{D_{tot}-E[D_{tot}]}{\sigma[D_{tot}]}}_{=\,Z_{tot}}\Bigg|S > \frac{21\,000-33\,000}{5000}\right) \tag{39}$$

$$= P\left(Z_{tot} \;>\; \frac{21\,000-33\,000}{5000}\right) \tag{40}$$

$$= P(\,Z_{tot} > -\,2.4\,) \tag{41}$$

$$= 1 - P(\,Z_{tot} \leq -\,2.4\,) \tag{42}$$

$$= \cancel{1} - \Big(\cancel{1} - P(\,Z_{tot} \leq 2.4\,)\Big) \tag{43}$$

$$= P(\,Z_{tot} \;\leq\; 2.4\,) \tag{44}$$

$$\overset{\text{tabell}}{=} G(2.4) \tag{45}$$

$$= \underline{\;0.9918\;} \tag{46}$$

Sannsynligheten for at mer enn 21 000 personer kjøper Dagbladet en gitt dag, dersom vi vet at det har skjedd en spesielle nyhet, er $\underline{\underline{99.18\ \%}}$.

f) Oppgaven spør etter sannsynligheten $P(D_{tot} > 21\,000)$.
Oppsplitting av utfallsrom Ω: [3]

$$\underline{P(D_{tot} > 21\,000)} \overset{\text{oppspl.}}{=} \underbrace{P(D_{tot} > 21\,000|S)}_{=0.9918} \cdot \overbrace{P(S)}^{=0.20} + \underbrace{P(D_{tot} > 21\,000|\overline{S})}_{=P(D_1>21\,000)=0.028} \cdot \overbrace{P(\overline{S})}^{=1-0.20} \tag{47}$$

$$= 0.9918 \cdot 0.20 + 0.0228 \cdot (1 - 0.20) = \underline{0.2166} \tag{48}$$

Sannsynligheten for at det en gitt dag etterspørres mer enn 21 000 aviser dersom vi ikke vet om det skjer en spesiell nyhet eller ikke den aktuelle dagen, er $\underline{\underline{21.66\,\%}}$.

■

[3] Se side 32 i formelsamlingen 2015.

Oppgave 3: (petroleumslogistikk)

a) Siden

$$\underline{\underline{\sum_{i=0}^{n} P(X = x_i)}} = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) \quad (49)$$

$$= 0.09 + 0.28 + 0.41 + 0.17 + 0.05 = \underline{\underline{1}} \quad (50)$$

b) Sannsynligheten for at det leveres inn 3 rapporter eller mer per dag:

$$\underline{\underline{P(X \geq 3)}} = P(X=3) + P(X=4) = 0.17 + 0.05 = \underline{\underline{0.22}} \quad (51)$$

c) i) Forventet antall innleverte rapporter per dag:

$$\underline{\underline{E[X]}} \overset{\text{def.}}{=} \sum_{i=0}^{4} x_i \cdot P(X = x_i) \quad (52)$$

$$= 0 \cdot \underbrace{P(X=0)}_{=0.09} + 1 \cdot \underbrace{P(X=1)}_{=0.28} + 2 \cdot \underbrace{P(X=2)}_{=0.41} + 3 \cdot \underbrace{P(X=3)}_{=0.17} + 4 \cdot \underbrace{P(X=4)}_{=0.05}$$

$$= 0 \cdot 0.09 + 1 \cdot 0.28 + 2 \cdot 0.41 + 3 \cdot 0.17 + 4 \cdot 0.05 = \underline{\underline{1.81}} \quad (53)$$

ii) For å finne variansen $Var[X]$ så regner vi først ut $E[X^2]$:

$$\underline{E[X^2]} \overset{\text{def.}}{=} \sum_{i=0}^{3} x_i^2 \cdot P(X = x_i) \quad (54)$$

$$= 0^2 \cdot \underbrace{P(X=0)}_{=0.09} + 1^2 \cdot \underbrace{P(X=1)}_{=0.28} + 2^2 \cdot \underbrace{P(X=2)}_{=0.41} + 3^2 \cdot \underbrace{P(X=3)}_{=0.17} + 4^2 \cdot \underbrace{P(X=4)}_{=0.05}$$

$$= 0^2 \cdot 0.09 + 1^2 \cdot 0.28 + 2^2 \cdot 0.41 + 3^2 \cdot 0.17 + 4^2 \cdot 0.05 = \underline{4.25} \quad (55)$$

Dette innsatt i setningen for '*varianssetningen*': [4]

$$\underline{\underline{Var[X]}} \quad = \quad E[X^2] - E[X]^2 \quad = 4.25 - 1.81^2 \quad = \underline{\underline{0.9739}} \tag{56}$$

d) i) Tolkning:

$$\underline{\underline{E[\,\overline{X}\,] \;=\; \text{forventet antall innleverte rapporter per dag i gjennomsnitt over et helt år}}}$$

ii) Forventet antall innleverte rapporter i gjennomsnitt per dag: $\quad (\; n = 365 \;)$

$$\underline{\underline{E[\,\overline{X}\,]}} \quad = \quad E\left[\, \frac{1}{n}\left(X_1 + X_2 + \;\ldots\; + X_n \right) \,\right] \tag{57}$$

$$\overset{\text{alltid}}{=} \quad \frac{1}{n}\left(E[X_1] + E[X_2] + \;\ldots\; + \; E[X_n] \right) \tag{58}$$

$$= \quad \frac{1}{n}\left(\; \underbrace{E[X] + E[X] + \;\ldots\; + E[X]}_{=\; n\cdot E[X]} \; \right) \quad = \quad \frac{1}{\cancel{n}} \, \cancel{n} \cdot \underbrace{E[X]}_{=1.81} \quad = \quad \underline{\underline{1.81}} \tag{59}$$

NB: Overgangen i lign.(57) til (58) gjelder alltid. Uansett om de stokastiske variablene X_i er uavhengige eller ikke.

[4]Se side 39 i formelsamlingen fra 2015.

e) i) Tolkning:

$$\underline{\underline{Var[\,\overline{X}\,]}} \;=\; \underline{\underline{\text{forventet variasjon/spredning i antall innleverte rapporter per dag i gjennomsnitt over et helt år}}} \tag{60}$$

ii) Variansen til antall innleverte rapporter per dag i gjennomsnitt over et år: $(\; n = 365 \;)$

$$\underline{\underline{Var[\,\overline{X}\,]}} \;=\; Var\left[\,\frac{1}{n}\Big(X_1 + X_2 + \;...\; + X_n\Big)\,\right] \tag{61}$$

$$\overset{\text{uavh.}}{=} \;\frac{1}{n^2}\Big(\underbrace{Var[X_1] + Var[X_2] + \;...\; + Var[X_n]}_{n\cdot Var[X]}\Big) \tag{62}$$

$$= \;\frac{1}{n^{\cancel{2}}}\,\cancel{n}\cdot\underbrace{Var[X]}_{=0.9739} \;=\; \frac{0.9739}{365} \;=\; \underline{\underline{0.002668}} \tag{63}$$

NB: Overgangen i lign.(61) til (62) gjelder fordi de stokastiske variablene X_i er uavhengige.

f) Siden

1. antall innleverte rapporter per dag er uavhengige: (oppgitt i oppgaveteksten)

   $X_i \;\sim\;$ er uavhengige for alle $i = 1, 2, 3, ..., n$

2. alle X_i har samme sannsynlighetsfordeling: (oppgitt i oppgaveteksten)

   $X_i \;\sim\;$ samme sannsynlighetsfordeling for alle $i = 1, 2, 3, ..., n$

3. antall "forsøk", dvs. antall dager, $n = 365$ er tilstrekkelig stort [5]

så gjelder **sentralgrensesetningen**. Dermed er $\overline{X}$ normalfordelt.

[5]Husk: Antall forsøk n for at sentralgrensesetningen skal gjelde er avhengig av situasjonen. Men en **tommelfingerregel** er at vi bør ha $n \gtrsim 30$.

g) Fra oppgavene **2c**, **2d** og **2e** foran ser vi at:

$$\underbrace{E[\,\overline{X}\,]}_{=\,1.81} \quad = \quad \underbrace{E[X]}_{=\,1.81} \tag{64}$$

og at

$$\underbrace{Var[\,\overline{X}\,]}_{=\,0.002668} \quad \ll \quad \underbrace{Var[X]}_{=\,0.9739} \tag{65}$$

Det betyr at sannsynlighetfordelingen til X, dvs. $P(X = x)$ gitt ved tabell i oppgaveteksten, og sannsynlighetfordelingen til $\overline{X}$, dvs. $\overline{X} \sim N[\,E[X], \sqrt{\frac{VarX}{n}}\,]$ har samme tyngdepunkt, men mye mindre varians / usikkerhet.

h) Sannsynligheten for at antall innleverte rapporter i løpet av et år overstiger 700: ($n = 365$)

$$\underline{\underline{P(X_1 + X_2 + \ \dots \ + X_n > 700)}} \quad = \quad P\left(\frac{X_1 + X_2 + \ \dots \ + X_n}{n} > \frac{700}{n}\right) \tag{66}$$

$$= \quad P\left(\overline{X} > \frac{700}{n}\right) \tag{67}$$

$$= \quad 1 \ - \ P\left(\overline{X} \leq \frac{700}{n}\right) \tag{68}$$

$$\overset{\text{standardiser}}{=} \quad 1 \ - \ P\Bigg(\underbrace{\frac{\overline{X} - E[\overline{X}]}{\sigma[\overline{X}]}}_{=\overline{Z}} \ \leq \ \frac{\frac{700}{n} - E[\overline{X}]}{\sigma[\overline{X}]}\Bigg) \tag{69}$$

$$= \quad 1 \ - \ P\left(\overline{Z} \ \leq \ \frac{\frac{700}{365} - 1.81}{\sqrt{0.002668}}\right) \tag{70}$$

$$= \quad 1 \ - \ P(\overline{Z} \ \leq \ 2.09) \tag{71}$$

$$= \quad 1 \ - \ G(2.09) \tag{72}$$

$$= \quad 1 \ - \ 0.9817 \tag{73}$$

$$= \quad \underline{\underline{0.0183}} \tag{74}$$

Kommentar:

Legg merke til at det er $\overline{X}$ som skal standardiseres. Ikke X. Det betyr at vi må bruke $E[\overline{X}] = 1.81$ og $\sigma[\overline{X}] = \sqrt{0.002668}$ (ikke $E[X] = 1.81$ og $\sigma[X] = \sqrt{0.9739}$).

■

Oppgave 4: (økonomi)

a) **i**) R_{xy} er normalisert og ligger mellom -1 og 1, dvs.: $\underline{\underline{-1 \leq R_{xy} \leq 1}}$.

ii) R_{xy} er et mål på lineær korrelasjon mellom obervasjonene x og y.
Det er et mål på i hvor stor grad det er en lineær sammmenheng mellom obervasjonene x og y.

iii) R_{xy} er enhetsuavhengig, dvs. ingen enhet. [6]

b) Uttrykket for korrelasjonskoeffisienten finner vi i formelsamlingen. Det gir:

$$\underline{\underline{R_{xy}}} = \frac{S_{xy}}{S_x \cdot S_y} \tag{75}$$

$$= \frac{-110.83}{\sqrt{116.67} \cdot \sqrt{109.54}} = \underline{\underline{-0.98}} \tag{76}$$

c) At $R_{xy} = -0.98$ betyr at det er en sterk negativ korrelasjon mellom x og y, dvs. store x hører sammen med små y og omvendt.
Det er altså en sterk lineær sammenheng mellom x og y med negativt stigningstall.

[6] Dette betyr at R_{xy} har samme verdi uansett hva slags enhet man bruker for å regne ut $R_{xy} = S_{xy}/(S_x \cdot S_y)$.

d) Vi skal finne regresjonslinjen. Formelen for minste kvadraters lineære regresjonslinje står i formelsamlingen:

$$\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta}x \tag{77}$$

hvor parametrene $\hat{\beta}$ og $\hat{\alpha}$ er: (dropper benevningen)

$$\underline{\hat{\beta}} \;=\; \frac{S_{xy}}{S_x^2} \;=\; \frac{-110.83}{116.67} \;=\; \underline{-0.95} \tag{78a}$$

$$\underline{\hat{\alpha}} \;=\; \overline{y} \;-\; \beta\overline{x} \;=\; 40.43-(-0.95)\cdot 15 \;=\; \underline{54.68} \tag{78b}$$

Minste kvadraters lineære regresjonslinje $\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta}\,x$ blir dermed:

$$\underline{\underline{\hat{y} \;=\; 54.68 \;-\; 0.95\cdot x}} \tag{79}$$

e) Bruker regresjonslinjen $\hat{y} = \hat{y}(x)$ fra oppgave **4d**, dvs. lign.(79):

$$\underline{\hat{y}(17)} \;=\; 54.68 \;-\; 0.95\cdot 17 \;=\; \underline{38.53} \tag{80}$$

Regresjonslinjen predikerer at kvadratmeterprisen for en leilighet som er 17 km utenfor sentrum er $\underline{\underline{38\,530\ \text{NOK}/m^2}}$.

f) Gjennomsnittlig kvadratmeterpris i Norge er 35 000 NOK/m^2.
Ifølge regresjonslinjen, når $\hat{y}(x) = 35$, er da:

$$35 \quad = \quad 54.68 \; - \; 0.95 \cdot x \tag{81}$$

Løser med hensyn på x:

$$\underline{\;x\;} \quad = \quad \frac{54.68 \; - \; 35}{0.95} \quad = \quad \underline{20.72} \tag{82}$$

Ifølge regresjonslinjen oppnår man landsgjennomsnittet i kvadratmeterpris
20.72 km utenfor sentrum.

g) Forklaringskraften R^2 kan leses direkte fra Excel-utskriften:
(Se cellen som heter "*R Square*" i Excel-utskriften): [7]

$$\underline{\underline{R^2 \; = \; 0.9613}} \tag{83}$$

h) Kommentar til svaret i oppgave **4g**:

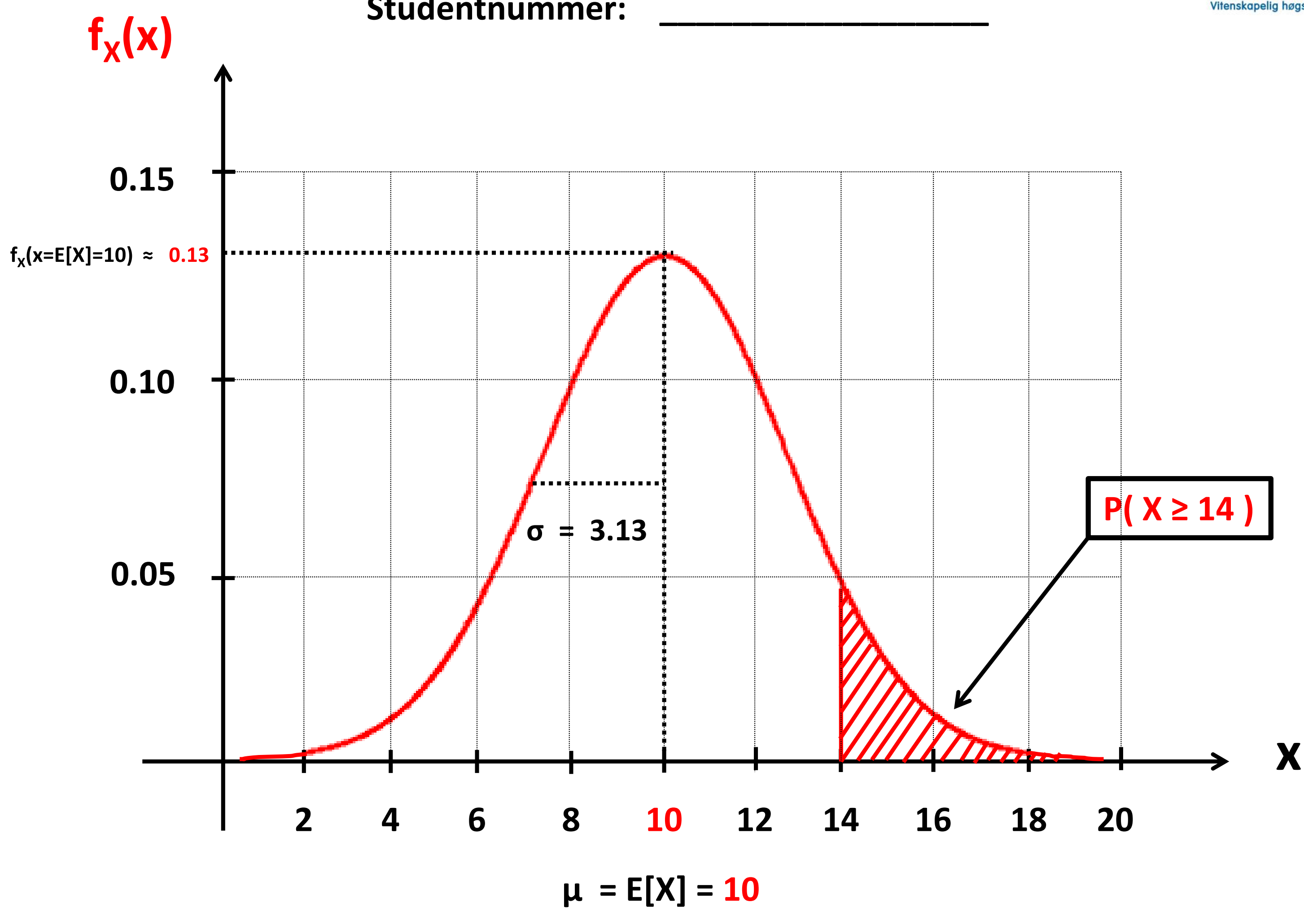
At $R^2 = 0.9613$ betyr at regresjonslinjen vil i "i stor grad" forutsi kvadrameterprisen
på en leilighet for en gitt avstand fra sentrum.
Modellen har stor forklaringskraft.

■

[7] Man kan også regne ut forklaringskraften R^2 "for hånd" via definisjonen $R^2 = 1 - SSE/SST$. Men det er mye mer arbeidskrevende. Fint at dataprogrammer (som f.eks. Excel) kan hjelpe oss med slikt.

Vedlegg

Studentnummer: ___________________



Kapittel 8

Kontinuasjonseksamen 2015

LØSNING: Eksamen 8. januar 2016

"*MAT110 Statistikk 1*", vår 2015

Oppgave 1: (logistikk)

a) Siden lastebildene ankommer varemottaket uavhengige av hverandre så er dette et telleproblem. Dermed er sannsynlighetene P_1, P_2 og P_3 gitt ved:

$$\underline{\underline{P_1}} = \frac{n_1}{n} = \frac{5}{10} = \underline{\underline{0.5}} \tag{1}$$

$$\underline{\underline{P_2}} = \frac{n_2}{n} = \frac{3}{10} = \underline{\underline{0.3}} \tag{2}$$

$$\underline{\underline{P_3}} = \frac{n_3}{n} = \frac{2}{10} = \underline{\underline{0.2}} \tag{3}$$

b) Sannsynlighetene P_1, P_2 og P_3 utgjør en **gyldig** sannsynlighetsfordeling fordi de summeres opp til èn:

$$\underline{\underline{P_1 + P_2 + P_3}} = 0.5 + 0.3 + 0.2 = \underline{\underline{1}} \tag{4}$$

c) Den spesifikke sannsynlighetfordelingen $P(X = x_i)$ er oppgitt.
Dermed bruker vi definisjonen av forventning:

$$\underline{E[X]} \stackrel{\text{def.}}{=} \sum_{i=1}^{3} x_i \cdot P(X = x_i) \tag{5}$$

$$= 10 \cdot \underbrace{P(X = 10)}_{=0.5} + 20 \cdot \underbrace{P(X = 20)}_{=0.3} + 30 \cdot \underbrace{P(X = 30)}_{=0.2}$$

$$= 10 \cdot 0.5 + 20 \cdot 0.3 + 30 \cdot 0.2 = \underline{17} \tag{6}$$

Forventet behandlingstid $E[X]$ for å laste av lasten til en tilfeldig valgt lastebil er <u>17 minutter</u>.

d) Den spesifikke sannsynlighetfordelingen $P(X = x_i)$ er oppgitt.
Dermed bruker vi definisjonen av varians:

$$\underline{\underline{Var[X]}} \stackrel{\text{def.}}{=} \sum_{i=1}^{3} \Big(x_i - E[X] \Big)^2 \cdot P(X = x_i) \tag{7}$$

$$= (10 - 17)^2 \cdot \underbrace{P(X = 10)}_{=0.5} + (20 - 17)^2 \cdot \underbrace{P(X = 20)}_{=0.3} + (30 - 17)^2 \cdot \underbrace{P(X = 30)}_{=0.2}$$

$$= (10 - 17)^2 \cdot 0.5 + (20 - 17)^2 \cdot 0.3 + (30 - 17)^2 \cdot 0.2 = \underline{\underline{61}} \tag{8}$$

med tilhørende standardavvik $\sigma[X]$:

$$\underline{\underline{\sigma[X]}} \stackrel{\text{def.}}{=} \sqrt{Var[X]} = \sqrt{61} = \underline{\underline{7.81}} \tag{9}$$

e) Fortventet ventetid $E[V]$:

$$\underline{E[V]} = E[X_1 + X_2 + X_3] \tag{10}$$

$$= E[X_1] \ + \ E[X_2] \ + \ E[X_3] \tag{11}$$

$$= E[X] \ + \ E[X] \ + \ E[X] \tag{12}$$

$$= 3\,E[X] \ = \ 3 \cdot 17 \ = \ \underline{51} \tag{13}$$

Forventet behandlingstid $E[V]$ dersom det står 3 tilfeldige lastebiler foran deg i kø, er <u>51 minutter</u>.

f) Fortventet varians $Var[V]$:

$$\underline{\underline{Var[V]}} = Var[X_1 + X_2 + X_3] \tag{14}$$

$$\overset{\text{uavh.}}{=} Var[X_1] \ + \ Var[X_2] \ + \ Var[X_3] \tag{15}$$

$$= Var[X] \ + \ Var[X] \ + \ Var[X] \tag{16}$$

$$= 3\,Var[X] \ = \ 3 \cdot 61 \ = \ \underline{\underline{183}} \tag{17}$$

med tilhørende standardavvik $\sigma[V]$:

$$\underline{\underline{\sigma[V]}} \overset{\text{def.}}{=} \sqrt{Var[V]} \ = \ \sqrt{183} \ = \ \underline{\underline{13.5}} \tag{18}$$

g) Siden lastebildene ankommer varemottaket uavhengige av hverandre så er den simultane sannsynligheten bare produktet av sannsynlighetene:

$$\underline{\underline{p(30,30,30)}} \equiv P\big(\, X_1 = 30 \;\text{ og }\; X_2 = 30 \;\text{ og }\; X_3 = 30\,\big) \tag{19}$$

$$\stackrel{\text{uavh.}}{=} P(X=30)\cdot P(X=30)\cdot P(X=30) \tag{20}$$

$$= \Big(\underbrace{P(X=30)}_{=0.2}\Big)^3 = 0.2^3 = \underline{\underline{0.008}} \tag{21}$$

h) Tolkning:

$p(30,30,30)$ = sannsynligheten for at alle 3 lastebilene som står foran deg i kø, er store lastebilder som tar 30 minutter hver å laste av

i) Den eneste måten at ventetiden kan bli 90 minutter på er at alle 3 lastebildene som er foran deg i kø, er store: $P(V=90) = p(30,30,30)$. Dermed:

$$\underline{P(V \le 80)} = 1 - \overbrace{P(V=90)}^{=\;p(30,30,30)} \tag{22}$$

$$= 1 - 0.008 = \underline{0.992} \tag{23}$$

Sannsynligheten for at du må vente 80 minutter eller mindre er $\underline{\underline{P(V \le 80) = 0.992}}$.

j) Igjen bruker vi det faktum at lastebilene ankommer uavhengige av hverandre.
Da er de simultane sannsynlighetene bare produktet av sannsynlighetene. Dermed:

$$
\begin{aligned}
\underline{P(V=50)} \;&=\; \overbrace{p(10,10,30) \;+\; p(10,30,10) \;+\; p(30,10,10)}^{\text{disse tre sanns. er like}} \\
&\;+\; \underbrace{p(20,20,10) \;+\; p(20,10,20) \;+\; p(10,20,20)}_{\text{disse tre sanns. er like}} \qquad (24)
\end{aligned}
$$

$$
= \; 3\cdot p(10,10,30) \;+\; 3\cdot p(20,20,10) \qquad (25)
$$

$$
= \; 3\Big(p(10,10,30) \;+\; p(20,20,10)\Big) \qquad (26)
$$

$$
= \; 3\Big(0.5\cdot 0.5\cdot 0.2 \;+\; 0.3\cdot 0.3\cdot 0.5\Big) \;=\; \underline{0.285} \qquad (27)
$$

Sannsynligheten for at du må vente 50 minutter er $\underline{\underline{P(V=50)=0.285}}$.

■

Oppgave 2: (økonomi / aksjer)

a) Bruker setningen for total sannsynlighet for å finne sannsynligheten for at en tilfeldig valgt medlem av *Econa* aldri leser næringslivsaviser:

$$\underline{\underline{P(A)}} \overset{\text{total}}{=} P(A \cap H) + P(A \cap \overline{H}) \tag{28}$$

$$= 0.04 + 0.21 = \underline{\underline{0.25}} \tag{29}$$

b) Begivenhetene i tabellen er disjunkte fordi begivenhetene aldri inntreffer samtidig. Man kan f.eks. ikke både leser aviser regelmessig og aldri.

c) Siden begivenhetene er disjunkte så er situasjonen helt analog med situasjonen for oppsplitting av utfallsrom. Sannsynligheten $P(H)$ blir dermed:

$$\underline{\underline{P(H)}} \overset{\text{oppspl.}}{=} P(H \cap R) + P(H \cap I) + P(H \cap A) \tag{30}$$

$$= 0.18 + 0.10 + 0.04 = \underline{\underline{0.32}} \tag{31}$$

d) Den generelle multiplikasjonssetningen gir:

$$\underline{\underline{P(H|A)}} = \frac{P(H \cap A)}{P(A)} = \frac{0.04}{0.25} = \underline{\underline{0.16}} \tag{32}$$

e) Man kan igjen bruke den generelle multiplikasjonssetningen: (husk at: $H \cap A = A \cap H$)

$$\underline{\underline{P(A|H)}} \overset{\text{mult.}}{=} \frac{P(A \cap H)}{P(H)} = \frac{0.04}{0.25} = \underline{\underline{0.125}} \tag{33}$$

eller Bayes lov

$$\underline{\underline{P(A|H)}} \overset{\text{Baye}}{=} P(H|A)\,\frac{P(A)}{P(H)} = 0.16\,\frac{0.25}{0.32} = \underline{\underline{0.125}} \tag{34}$$

På eksamen er det nok at dere løser oppgaven på èn av måtene.

f) Tolkning:

$$\underline{\underline{P(A|H)}} = \underline{\underline{\text{sannsynligheten for at et medlem av } \textit{Econa} \text{ som har handlet aksjer på aksjemarkedet det siste året, aldri leser finansaviser}}} \tag{35}$$

g) Sannsynligheten $P\big(H|\overline{R}\big) = P\big(H|(A \cup I)\big)$ kan vi finne via den generelle multiplikasjonssetningen:

$$P\big(H|\overline{R}\big) = \frac{P\big(H \cap \overline{R}\big)}{P(\overline{R})} \tag{36}$$

eller, siden $\overline{R} = A \cup I$,

$$P\big(H|(A \cup I)\big) = \frac{P\big(H \cap (A \cup I)\big)}{P(A \cup I)} \tag{37}$$

Siden begivenhetene A og I disjunkte, så kan vi bruke den spesielle addisjonssetningen:

$$\underline{P(A\cup I)} \overset{\text{add.}}{=} P(A) + P(I) = \underbrace{(0.04+0.21)}_{\text{se tabell eller oppg. }\mathbf{1a}} + \underbrace{(0.10+0.31)}_{\text{se tabell i oppgaven}} = \underline{0.66} \tag{38}$$

Tabellen i oppgaven gir videre at: [1]

$$\underline{P\big(H\cap(A\cup I)\big)} = P\Big((H\cap A)\cup(H\cap I)\Big) \overset{\text{add.}}{=} P(H\cap A) + P(H\cap I) \tag{39}$$

$$= 0.04 + 0.10 = \underline{0.14} \tag{40}$$

Innsatt i lign.(37):

$$\underline{\underline{P\big(H|(A\cup I)\big)}} = \frac{P\big(H\cap(A\cup I)\big)}{P(A\cup I)} = \frac{0.14}{0.66} \approx \underline{\underline{0.21}} \tag{41}$$

■

[1] Man får også full poengscore selv om man ikke har med mellomregningene i lign.(39). Jeg er kanskje litt overtydelig her.

Oppgave 3: (logistikk)

a) Det er oppgitt i oppgaven at $X_i \sim \text{Poi}[\lambda]$.
Siden $X_G \sim \text{Poi}[\lambda]$ så finner man sannsynligheten for at det skjer 2 utrykninger i løpet av en uke i Gjemnes på følgende måte:

$$\underline{\underline{P(X_G = 2)}} = \frac{\lambda_G^2}{2!}\, e^{-\lambda_G} \tag{42}$$

$$\overset{\lambda_G = 1.9}{=} \frac{1.9^2}{2!}\, e^{-1.9} \overset{\text{kalkis}}{=} \underline{\underline{0.2700}} \tag{43}$$

b) Det er oppgitt i oppgaven at $X_G \sim \text{Poi}[\lambda]$.
Sannsynligheten for at det skjer mer enn 2 utrykninger i Gjemnes i løpet av en uke:

$$\underline{\underline{P(X_G > 2)}} = 1 - P(X_G \leq 2) \tag{44}$$

$$= 1 - \left(\overbrace{P(X_G = 0) + P(X_G = 1)}^{\text{Oppgitt i oppgaven.}} + \overbrace{P(X_G = 2)}^{\text{Fra oppg. } \mathbf{3a}} \right) \tag{45}$$

$$= 1 - 0.1496 - 0.2842 - 0.2700 \tag{46}$$

$$= \underline{\underline{0.2962}} \tag{47}$$

c) Forventet antall akutte utrykninger i Eide, Fræna og Gjemnes til sammen per uke:

$$\underline{\underline{E[Y]}} = E[X_E + X_F + X_G] \tag{48}$$

$$\stackrel{\text{alltid}}{=} E[X_E] + E[X_F] + E[X_G] = \lambda_E + \lambda_F + \lambda_G \tag{49}$$

$$= 1.5 + 2.3 + 1.9 = \underline{\underline{5.7}} \tag{50}$$

NB: Overgangen i lign.(48) til (49) gjelder alltid. Uansett om de stokastiske variablene X_i er uavhengige eller ikke. (Se lign.(5.13) på side 40 i formelsamlingen).

d) En sum av Poisson fordelinger $X_E + X_F + X_G$ er også Poisson fordelt, dvs. Y er Poisson fordelt med forventing $\lambda_Y = E[Y] = 5.7$.

Sannsynligheten for at det skjer *mer enn* 2 akutte utrykningen i Eide, Fræna og Gjemnes tilsammen oer uke er da:

$$\underline{\underline{P(Y > 2)}} = 1 - P(Y \leq 2) \tag{51}$$

$$= 1 - \Big(P(Y=0) + P(Y=1) + P(Y=2) \Big) \tag{52}$$

$$= 1 - \frac{\lambda_Y^0}{0!} e^{-\lambda_Y} - \frac{\lambda_Y^1}{1!} e^{-\lambda_Y} - \frac{\lambda_Y^2}{2!} e^{-\lambda_Y} \tag{53}$$

$$\stackrel{\lambda_Y = 5.7}{=} 1 - \frac{5.7^0}{0!} e^{-5.7} - \frac{5.7^1}{1!} e^{-5.7} - \frac{5.7^2}{2!} e^{-5.7} \tag{54}$$

$$\stackrel{\text{kalkis}}{=} 1 - 0.0033 - 0.0191 - 0.0544 \tag{55}$$

$$= \underline{\underline{0.9232}} \tag{56}$$

e) Siden

$$P(Y > 2) = 0.9232 \quad \text{og} \quad P(X > 2) = 0.2962 \tag{57}$$

så ser vi at $P(Y > 2)$ er mye større enn $P(X_G > 2)$. Siden Y beskriver summen av antall akutte utrykninger Eide, Fræna og Gjemnes tilsammen mens X_G kun beskriver antall utrykninger i Gjemnes alene, så er virker det rimelig fornuftig at:

$$P(Y > 2) \gg P(X > 2) \tag{58}$$

<u>NB</u>:

Selv om Y beskriver summen av akutte utrykninger i Eide, Fræna og Gjemnes tilsammen så kan vi IKKE slik at:

$$\underbrace{P(Y > 2) \ = \ P(X_E > 2) + P(X_F > 2) + P(X_G > 2)}_{\text{galt}} \tag{59}$$

f) Forventet antall akutte utrykninger i året i Gjemnes kommune:

$$\underline{\underline{E[\,X_{\text{år}}\,]}} \ = \ E\,[\,X_1 + X_2 + \ \ldots \ + X_n\,] \ = \ \overbrace{E[X_1] + E[X_2] + \ \ldots \ + E[X_n]}^{\text{n stk.}} \tag{60}$$

$$= \ n \underbrace{E[X]}_{=\lambda_G = 1.9} \ = \ 52 \cdot 1.9 \ = \ \underline{\underline{98.8}} \tag{61}$$

NB: Overgangen i lign.(60) gjelder alltid. Uansett om de stokastiske variablene X_i er uavhengige eller ikke.

g) Variansen til antall akutte utrykninger i året for Gjemnes kommune?

$$\underline{\underline{Var[\,X_{\text{år}}\,]}} \quad = \quad Var\,[\,X_1 + X_2 + \;...\; + X_n\,] \tag{62}$$

$$\overset{\text{uavh.}}{=} \quad \overbrace{Var[X_1] + Var[X_2] + \;...\; + Var[X_n]}^{\text{n stk.}} \tag{63}$$

$$= \quad n\,\underbrace{Var[X]}_{=\lambda_G=1.9} \quad = \quad 52 \cdot 1.9 \quad = \quad \underline{\underline{98.8}} \tag{64}$$

NB: Overgangen i lign.(62) gjelder *kun* dersom de stokastiske variablene X_i er uavhengige.

h) i) Med forutsetningene som formulert i oppgaven så gjelder sentralgrensesetningen.

ii) Ifølge sentralgrensesetningen er da den stokastiske variabelen $X_{\text{år}}$ **normal**fordelt:

$$\underline{\underline{X_{\text{år}} \quad \sim \quad \mathrm{N}[\,E[\,X_{\text{år}}\,]\,,\, Var[\,X_{\text{år}}\,]\,]}} \tag{65}$$

iii) En **tommelfingerregel** for at sentralgrensesetningen skal gjelde er:

$$\underline{\underline{n \;\gtrsim\; 30}} \tag{66}$$

dvs. antall forsøk bør være ca. 30 eller mer.

i) Siden $X_{\text{år}}$ er (tilnærmet) normalfordelt så kan vi bruke dette når vi skal regne ut sannsynligheten $P(X_{\text{år}} > 110)$: [2]

$$\underline{P(X_{\text{år}} > 110)} \overset{\text{standardiser}}{=} P\Bigg(\underbrace{\frac{X_{\text{år}} - E[X_{\text{år}}]}{\sigma[X_{\text{år}}]}}_{=Z} > \frac{110 - E[X_{\text{år}}]}{\sigma[X_{\text{år}}]}\Bigg) \tag{67}$$

$$= P\Bigg(Z > \frac{110 - 98.8}{\sqrt{98.8}}\Bigg) \tag{68}$$

$$= P(Z > 2.13) \tag{69}$$

$$= 1 - P(Z \leq 1.13) \tag{70}$$

$$= 1 - \underbrace{G(1.13)}_{=0.8708} \tag{71}$$

$$\overset{\text{tabell}}{=} 1 - 0.8708 \tag{72}$$

$$= \underline{0.1292} \tag{73}$$

Sannsynligheten for at det blir mer enn 110 akutte utrykninger i året i Gjemnes kommune er $\underline{\underline{12.92\ \%}}$.

■

[2]I oppgaven stod det at vi **ikke** behøver heltallskorreksjon. Derfor utelater vi det.

Oppgave 4: (økonomi)

a) **i**) R_{xy} er normalisert og ligger mellom -1 og 1, dvs.: $\underline{\underline{-1 \leq R_{xy} \leq 1}}$.

ii) R_{xy} er et mål på lineær korrelasjon mellom obervasjonene x og y.
Det er et mål på i hvor stor grad det er en lineær sammmenheng mellom obervasjonene x og y.

iii) R_{xy} er enhetsuavhengig, dvs. ingen enhet. [3]

b) Uttrykket for korrelasjonskoeffisienten finner vi i formelsamlingen. Det gir:

$$\underline{\underline{R_{xy}}} = \frac{S_{xy}}{S_x \cdot S_y} \tag{74}$$

$$= \frac{10.36}{\sqrt{6} \cdot \sqrt{19.7}} = \underline{\underline{0.95}} \tag{75}$$

c) At $R_{xy} = 0.95$ betyr at det er en sterk positiv korrelasjon mellom x og y, dvs. store x hører sammen med små y og omvendt.

Det er altså en sterk lineær sammenheng mellom x og y med positivt stigningstall.

[3]Dette betyr at R_{xy} har samme verdi uansett hva slags enhet man bruker for å regne ut $R_{xy} = S_{xy}/(S_x \cdot S_y)$.

d) Vi skal finne regresjonslinjen. Formelen for minste kvadraters lineære regresjonslinje står i formelsamlingen:

$$\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta}x \tag{76}$$

hvor parametrene $\hat{\beta}$ og $\hat{\alpha}$ er: (dropper benevningen)

$$\underline{\hat{\beta}} \;=\; \frac{S_{xy}}{S_x^2} \;=\; \frac{10.36}{6} \;=\; \underline{1.73} \tag{77}$$

$$\underline{\hat{\alpha}} \;=\; \overline{y} \;-\; \beta\overline{x} \;=\; 33.625 - 1.73 \cdot 4.5 \;=\; \underline{25.84} \tag{78}$$

Minste kvadraters lineære regresjonslinje $\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta}\,x$ blir dermed:

$$\underline{\underline{\hat{y} \;=\; 25.84 \;+\; 1.73 \cdot x}} \tag{79}$$

e) Bruker regresjonslinjen $\hat{y} = \hat{y}(x)$ fra oppgave **4d**, dvs. lign.(79):

$$\underline{\hat{y}(12)} \;=\; 25.84 \;+\; 1.73 \cdot 12 \;=\; \underline{46.56} \tag{80}$$

Dersom den lineære treden holder seg også i 2016 så predikerer regresjonslinjen at Brunvoll vil få omtrent $\underline{\underline{\text{47 serviceordrer}}}$.

f) Med 50 serviceordrer i kvartalet er $\hat{y}(x) = 50$. Ifølge regresjonslinjen er da:

$$\hat{y}(x) \overset{\text{Eq.(80)}}{=} 25.84 \; + \; 1.73 \cdot x \tag{81}$$

$$50 = 25.84 \; + \; 1.73 \cdot x \tag{82}$$

Løser med hensyn på x:

$$\underline{x} = \frac{50 \; - \; 25.84}{1.74} = 13.97 \; \approx \; \underline{14} \tag{83}$$

Ifølge regresjonslinjen så vil antall serviceordrer bli 50 i kvartalet,
2. kvartal i 2017.

g) Forklaringsstyrken R^2 er: (se formelsamling)

$$\underline{\underline{R^2}} \overset{\text{def.}}{=} 1 - \frac{SSE}{SST} \tag{84}$$

$$= 1 - \frac{12.73}{137.88} = \underline{\underline{0.9077}} \tag{85}$$

h) Kommentar til svaret i oppgave **4g**:

At $R^2 = 0.9077$ betyr at regresjonslinjen vil "i stor grad" forutsi antall serviceordrer per kvartal innenfor regresjonslinjens gyldighetsområde hvor vi har observasjoner.

■

Kapittel 9

Hovedeksamen 2016

Høgskolen i Molde
Vitenskapelig høgskole i logistikk

LØSNING: Eksamen 27. mai 2016

"*MAT110 Statistikk 1*", vår 2016

Oppgave 1: (logistikk , økonomi)

a) Sannsynligheten for at "*det ikke er ledig plass på ferge 2 er 0.25* ":

$$P(\overline{L}_2) \;=\; 0.25 \tag{1}$$

Sannsynligheten for at "*det er ledig plass på ferge 1 og ferge 2 er 0.70* "

$$P(L_1 \overbrace{\cap}^{\text{og}} L_2) \;=\; 0.70 \tag{2}$$

b) Vi skal finne "eller"-sannsynligheten $P(L_1 \cup L_2)$.

Vi kjenner "og"-sannsynligheten $P(L_1 \overbrace{\cap}^{\text{og}} L_2)$, og vi skal finne "eller"-sannsynligheten. Setningen som forbinder disse sannsynlighetene er den generelle addisjonssetningen:

$$\overbrace{P(L_1 \underbrace{\cup}_{\text{eller}} L_2)}^{\text{skal finne}} \;=\; P(L_1) + P(L_2) \;-\; \overbrace{P(L_1 \underbrace{\cap}_{\text{og}} L_2)}^{=\,0.70} \tag{3}$$

hvor $P(L_1 \cap L_2) = 0.70$ var oppgitt i oppgaven, og
hvor $P(L_1)$ og $P(L_2)$ finnes via kompementsetningen.

Dermed:

$$\underline{P(L_1 \cup L_2)} = \overbrace{P(L_1)}^{=\ 1-0.15} + \overbrace{P(L_2)}^{=\ 1-0.25} - \overbrace{P(L_1 \cap L_2)}^{=\ 0.70} \tag{4}$$

$$= 0.85 + 0.75 - 0.70 = \underline{0.90} \tag{5}$$

Altså, det er 90 % sannsynlighet for at det er ledig plass på ferge 1 eller ferge 2 (eller begge).

c) Vi skal finne sannsynligheten $P(L_1 \cap \overline{L}_2)$.

Sannsynligheten $P(L_1 \cap L_2) = 0.70$ var oppgitt i oppgaven, og vi skal finne $P(L_1 \cap \overline{L}_2)$.
Setningen som forbinder disse sannsynlighetene er den total sannsynlighet,
se formelsamlingen:

$$\underbrace{P(L_1)}_{=\ 1-0.15} = \underbrace{P(L_1 \cap L_2)}_{=\ 0.70} + \overbrace{P(L_1 \underbrace{\cap}_{\text{og}} \overline{L}_2)}^{\text{skal finne}} \tag{6}$$

som gir

$$\underline{P(L_1 \cap \overline{L}_2)} = P(L_1) - P(L_1 \cap L_2) \tag{7}$$

$$= 0.85 - 0.70 = \underline{0.15} \tag{8}$$

Altså, det er 15 % sannsynlighet for at det er ledig plass på ferge 1,
men ikke ledig på ferge 2.

d) Sannsynligheten $P(L_1 \cup L_2) = 0.90$ fant vi i oppgave **1b**, og vi skal finne $P(\overline{L}_1 \cap \overline{L}_2)$. Setningen som forbinder disse sannsynlighetene er en av tvillingsetningene, se formelsamlingen:

$$P(\overline{L}_1 \overbrace{\cap}^{\text{og}} \overline{L}_2) \;=\; 1 - P(L_1 \overbrace{\cup}^{\text{eller}} L_2) \tag{9}$$

som gir

$$\underline{P(\overline{L}_1 \overbrace{\cap}^{\text{og}} \overline{L}_2)} \;=\; 1 \;-\; \underbrace{P(L_1 \overbrace{\cup}^{\text{eller}} L_2)}_{=\,0.90} \tag{10}$$

$$=\; 1 - 0.90 \;=\; \underline{0.10} \tag{11}$$

Altså, det er $\underline{\underline{\text{10 \% sannsynlighet}}}$ for at begge fergene er fulle.

e) Bring sine forventede utgifter pga. fulle ferger for en tur mellom Molde og Volda er:

$$\underline{\underline{E[U]}} \;=\; \sum_{u=1}^{4} u_i \cdot P(U = u_i) \tag{12}$$

$$=\; \Big(\, 0 \cdot P(L_1 \cap L_2) \;+\; 1200 \cdot \overbrace{P(\overline{L}_1 \cap L_2)}^{=\,0.05} \tag{13}$$

$$+\; 800 \cdot P(L_1 \cap \overline{L}_2) \;+\; 2000 \cdot P(\overline{L}_1 \cap \overline{L}_2) \,\Big) \text{ NOK} \tag{14}$$

$$=\; \Big(0 \cdot 0.70 \;+\; 1200 \cdot 0.05 \;+\; 800 \cdot 0.15 \;+\; 2000 \cdot 0.10 \Big) \text{ NOK} \tag{15}$$

$$=\; \underline{\underline{380 \text{ NOK}}} \tag{16}$$

■

Oppgave 2: (logistikk , økonomi , "*avisguttens problem*")

a) Siden D er Poissonfordelt så er $E[D] = \lambda$:

$$\underline{\underline{E[D]}} \;=\; \lambda \;=\; \underline{\underline{81}} \tag{17}$$

b) Siden D er Poissonfordelt så er $\sigma[D] = \sqrt{\lambda}$:

$$\underline{\underline{\sigma[D]}} \;=\; \sqrt{\lambda} \;=\; \sqrt{81} \;=\; \underline{\underline{9}} \tag{18}$$

c) Se side 73 og 75 i 2016-versjonen av formelsamlingen ser vi læresetningen som sier at dersom λ er stor nok, typisk $\underline{\underline{\lambda \gg 5}}$, så er en Poissonfordeling med god tilnærmelse normalfordelt.

Siden $\lambda = 81 \gg 5$ så er vi her langt innenfor denne grensen.

d) Siden D er tilnærmet normalfordelt og finn sannsynligheten ved standadisering ($X \to Z$) og deretter tabelloppslag: [1]

$$\underline{P(D > 80)} \overset{\text{standardiser}}{=} P\left(\overbrace{\frac{D - E[D]}{\sigma[D]}}^{=Z} > \frac{80 - E[D]}{\sigma[D]}\right) \tag{19}$$

$$= P\left(Z > \frac{80 - 81}{\sqrt{81}}\right) \tag{20}$$

$$= \underline{P(Z > -0.11)} \tag{21}$$

$$= 1 - P(Z > 0.11) \tag{22}$$

$$= \cancel{1} - \left(\cancel{1} - P(Z \leq 0.11)\right) \tag{23}$$

$$= \underbrace{P(Z \leq 0.11)}_{=0.5438}$$

$$\overset{\text{tabell}}{=} \underline{0.5438} \tag{24}$$

Sannsynligheten for at det etterspørres mer enn 80 aviser en tilfeldig dag er $\underline{\underline{54.38\ \%}}$.

e) Forventningen $E[S]$ er gitt ved

$$\underline{\underline{E[S]}} = \underbrace{\sum_{s=1}^{84} sP(S = s)}_{=49.89} + 85 \cdot \underbrace{P(S = 85)}_{=0.3429} \tag{25}$$

$$= 49.89 + 85 \cdot 0.3429 \tag{26}$$

$$= \underline{\underline{79}} \tag{27}$$

[1] I oppgaven stod det ikke noe om at vi behøver heltallskorreksjon. Derfor tar vi det ikke med her.

f) Tolkning: $E[S]$ er forventet antall *solgte* aviser en gitt dag dersom avisgutten bestiller $q = 85$ aviser.

g) Fra oppgavene **2a** og **2e** har vi:

$$E[D] \;=\; 81 \quad , \quad E[S] \;=\; 79 \tag{28}$$

altså

$$\text{forventet etterspørsel} \;=\; 81 \quad > \quad \text{forventet salg} \;=\; 79 \tag{29}$$

Man kan ikke selge flere aviser enn markedet etterspør.
Det er derfor rimelig at forventet etterspørsel $E[D]$ er større enn forventet salg $E[S]$, dvs. det er rimelig at $E[D] > E[S]$. [2]

h) Forventet fortjeneste $E[\pi(q)]$: [3]

$$\underline{E[\pi(q)]} \;=\; E[\,r\,S - w\,q\,] \;=\; \underline{rE[S] - w\,q} \tag{32}$$

For tilfellet $q = 85$ er $E[S] = 79$, jfr. oppgave **2e**. Med $w = 5$ NOK og $r = 20$ NOK får vi:

$$\underline{\underline{E[\pi(q)]}} \;=\; \Big(20 \cdot 79 - 5 \cdot 85\Big)\ \text{NOK} \;=\; \underline{\underline{1\,155\ \text{NOK}}} \tag{33}$$

[2]En teknisk forklaring som man ikke behøver å ha med i besvarelsen på eksamen: Når avisgutten bestiller $q = 85$ aviser så er dette den øvre grensen for hvor mange aviser han han selge. Siden sannsynlighetfordelingen $P(S = s_i)$ er den samme som $P(D = d_i)$ frem til $s = q - 1 = 84$ så må $E[D] > E[S]$.

[3]Her bruker vi regneregelene: (a og b er konstanter)

$$E[aX + bY] \;=\; aE[X] + bE[Y]\ , \tag{30}$$
$$E[a] \;=\; a\ , \tag{31}$$

som man finner i formelsamlingen.

i) I oppgaven er det oppgitt at den stokastiske variabelen D er normalfordelt, med $D \sim \mathrm{N}\big[\,\mu = 81, \sigma = 9\,\big]$. Dermed løses denne oppgaven med (omvendt) tabelloppslag. [4]

Med innkjøpspris $w = 5$ og utslagspris $r = 20$ fås:

$$P(D \leq q^*) \quad = \quad 1 - \frac{w}{r} \tag{34}$$

$$P(D \leq q^*) \quad = \quad 1 - \frac{5}{20} \tag{35}$$

$$P(D \leq q^*) \quad = \quad 0.75 \tag{36}$$

$$P\Bigg(\underbrace{\frac{D-\mu}{\sigma}}_{\equiv Z} \leq \underbrace{\frac{q^*-\mu}{\sigma}}_{\equiv Z_0}\Bigg) \quad \overset{\text{standardiser}}{=} \quad 0.75 \tag{37}$$

$$P(Z \leq Z_0) \quad = \quad 0.75 \tag{38}$$

Ved "omvendt tabelloppslag" ser vi at $Z_0' = 0.67$ tilsvarer $P(Z' \leq Z_0') = 0.7486$. Videre ser vi at $Z_0'' = 0.68$ tilsvarer $P(Z'' \leq Z_0'') = 0.7517$. Vi skal ha 0.75, som er ca. *midt i mellom*. Dermed:

$$Z_0 \quad = \quad 0.675 \tag{39}$$

Vi løser:

$$Z_0 \quad = \quad \frac{q^*-\mu}{\sigma} \tag{40}$$

med hensyn på q^*: $\quad (\ \mu = 81 \text{ og } \sigma = 9\ )$

$$\underline{q^*} \quad = \quad \mu \ + \ Z_0 \cdot \sigma \tag{41}$$

$$= \quad 81 \ + \ 0.675 \cdot 9 \quad = \quad \underline{87} \tag{42}$$

Avisgutten må bestille $\underline{\underline{q^* = 87}}$ aviser for å få størst mulig fortjeneste.

[4] Husk at tabellen på side 63 i 2016-formelsamlingen dreier seg KUN om normalfordeling.

j) Forventet etterspørsel av aviser er $E[D] = \mu = 81$ per dag.
Fortjenesten blir størst når avisgutten bestiller $q^* = 87$ aviser per dag. Altså

$$\begin{aligned} q^* &> E[D] \\ 87 &> 81 \ , \end{aligned} \tag{43}$$

dvs. det lønner seg å bestille flere aviser enn det man forventer å selge.
Dette fordi:

Man **taper mye mer på tapt salg** enn på aviser han ikke får solgt.

Utdypende kommentar:

Dersom avisgutten brenner inne med for mange aviser så taper han "bare" 5 NOK per avis.

Men dersom han har for lite aviser så mister han salg,
dvs. han mister inntekten $(20 - 5)$ NOK $= 15$ NOK per avis.

Tapt salg er altså mye dyrere for avisgutten enn å brenne inne med aviser.
Derfor er det "mindre ille" å brenne inne med aviser enn å gå tom for aviser.
Det er forklaringen på at han bestiller flere aviser enn han forventer å selge.

PS: Man behøver ikke ha med den utdypende kommentaren i eksamensbesvarelsen.
Det er nok med forklaringen som er innrammet.

■

Oppgave 3: (logistikk)

a) Siden den stokastiske variabelen X har et tellbart antall mulige verdier så er X $\underline{\underline{\text{diskret}}}$.

b) Enhver gyldig sannsynlighetsfordeling er $\underline{\underline{\text{normalisert til 1}}}$, dvs. $\sum_i P(X = x_i) = 1$.
For at denne normaliseringsbetingelsen skal være oppfylt så må:
$P(X = 2) = 0.20$.

c) Forventning $E[X]$:

$$\underline{\underline{E[X]}} \stackrel{\text{def.}}{=} \sum_{i=0}^{5} x_i \cdot P(X = x_i) \tag{44}$$

$$\begin{aligned} &= 0 \cdot \overbrace{P(X = 0)}^{=0.25} + 1 \cdot \overbrace{P(X = 1)}^{=0.30} + 2 \cdot \overbrace{P(X = 2)}^{=0.20} \\ &\quad + 3 \cdot \underbrace{P(X = 3)}_{=0.15} + 4 \cdot \underbrace{P(X = 4)}_{=0.10} + 5 \cdot \underbrace{P(X = 5)}_{=0} \end{aligned} \tag{45}$$

$$= 0 \cdot 0.25 + 1 \cdot 0.30 + 2 \cdot 0.20 + 3 \cdot 0.15 + 4 \cdot 0.10 + 5 \cdot 0 \tag{46}$$

$$= \underline{\underline{1.55}} \tag{47}$$

d) Sannsynligheten for at det skjer høyst èn trafikkulykke på den aktuelle vegstrekningen over en periode på to år er:

$$\begin{aligned} P(Y \leq 1) \;&=\; P(X_1 = 0) \cdot P(X_2 = 0) \\ &+\; P(X_1 = 1) \cdot P(X_2 = 0) \;+\; P(X_1 = 0) \cdot P(X_2 = 1) \end{aligned} \tag{48}$$

Men $P(X_1 = 0) = P(X_2 = 0) = 0.25$ osv., så:

$$\underline{\underline{P(Y \leq 1)}} \;=\; P(X = 0)^2 \;+\; 2P(X = 0)P(X = 1) \tag{49}$$

$$=\; 0.25^2 \;+\; 2 \cdot 0.25 \cdot 0.30 \;=\; \underline{\underline{0.2125}} \tag{50}$$

e) Forventet antall trafikkulykker på den aktuelle strekningen over en periode på $n = 30$ år:

$$\underline{\underline{E[Y]}} \;=\; E[X_1 + X_2 + \; ... \; + X_n] \;=\; \overbrace{E[X_1] + E[X_2] + \; ... \; + E[X_n]}^{\text{n stk.}} \tag{51}$$

$$=\; n\,E[X] \;=\; 30 \cdot 1.55 \;=\; \underline{\underline{46.5}} \tag{52}$$

NB: Overgangen i lign.(51) gjelder alltid. Uansett om de stokastiske variablene X_i er uavhengige eller ikke.

f) Variansen til antall trafikkylykker på den aktuelle vegstrekningen over en periode på på $n = 30$ år:

$$\underline{\underline{Var[Y]}} = Var[X_1 + X_2 + \ ... \ + X_n] \tag{53}$$

$$\stackrel{\text{uavh.}}{=} \overbrace{Var[X_1] + Var[X_2] + \ ... \ + Var[X_n]}^{\text{n stk.}} \tag{54}$$

$$= n \cdot \underbrace{Var[X]}_{=1.6475} = 30 \cdot 1.6475 = \underline{\underline{49.425}} \tag{55}$$

NB: Overgangen i lign.(53) gjelder *kun* dersom de stokastiske variablene X_i er uavhengige.

g) i) Med forutsetningene som formulet i oppgaven så gjelder sentralgrensesetningen.

ii) Ifølge sentralgrensesetningen er da den stokastiske varibalen Y **normal**fordelt:

$$\underline{\underline{Y \sim \mathrm{N}\big[\, E[Y], Var[Y] \,\big]}} \tag{56}$$

iii) En **tommelfingerregel** for at sentralgrensesetningen skal gjelde er:

$$\underline{\underline{n \gtrsim 30}} \tag{57}$$

dvs. antall forsøk bør være ca. 30 eller mer.

h) Siden Y er (tilnærmet) normalfordelt så kan vi bruke dette når vi skal regne ut sannsynligheten $P(Y > 50)$: [5]

$$\underline{P(Y > 50)} \overset{\text{standardiser}}{=} P\left(\underbrace{\frac{Y - E[Y]}{\sigma[Y]}}_{=\,Z} > \frac{50 - E[Y]}{\sigma[Y]} \right) \tag{58}$$

$$= P\left(Z > \frac{50 - 46.5}{\sqrt{49.425}} \right) \tag{59}$$

$$= P(Z > 0.50) \tag{60}$$

$$= 1 - P(Z \leq 0.50) \tag{61}$$

$$= 1 - \underbrace{G(0.50)}_{=\,0.6915} \tag{62}$$

$$\overset{\text{tabell}}{=} 1 - 0.6915 \tag{63}$$

$$= \underline{0.3085} \tag{64}$$

Sannsynligheten for at det skjer mer enn 50 trafikkulykker på den aktuelle vegstrekningen over en periode på $n = 30$ år er $\underline{\underline{30.85\ \%}}$.

■

[5] I oppgaven stod det ikke noe om at vi behøver heltallskorreksjon. Derfor tar vi det ikke med her.

Oppgave 4: (økonomi)

a) **i**) S_{xy} kan ha alle reelle verdier, dvs. $S_{xy} \in \langle -\infty, \infty \rangle$.

ii) Tallverdien til S_{xy} er vanskelig å tolke fordi:

- størrelsen S_{xy} kan gi "store" eller "små" verdier som vi må sammenligne med andre relevante tall for å kunne forstå bedre

- S_{xy} er enhetsavhengig og gir dermed ulikt resultat avhengig av hva slags enhet man velger

b) **i**) R_{xy} er normalisert og ligger mellom -1 og 1, dvs.: $\underline{\underline{-1 \leq R_{xy} \leq 1}}$.

ii) R_{xy} er et mål på lineær korrelasjon mellom obervasjonene x og y. Det er et mål på i hvor stor grad det er en lineær sammmenheng mellom obervasjonene x og y.

iii) R_{xy} er enhetsuavhengig, dvs. ingen enhet. [6]

c) Uttrykket for korrelasjonskoeffisienten finner man i formelsamlingen. Det gir:

$$\underline{\underline{R_{xy}}} = \frac{S_{xy}}{S_x \cdot S_y} \tag{65}$$

$$= \frac{-0.1531}{\sqrt{0.15625} \cdot \sqrt{0.15325}} = \underline{\underline{-0.99}} \tag{66}$$

[6]Dette betyr at R_{xy} har samme verdi uansett hva slags enhet man bruker for å regne ut $R_{xy} = S_{xy}/(S_x \cdot S_y)$.

d) At $R_{xy} = -0.99$ betyr at det er en veldig $\underline{\underline{\text{sterk negativ korrelasjon}}}$ mellom x og y.

Det er altså en svært $\underline{\underline{\text{sterk lineær sammenheng}}}$ mellom x og y med negativt stigningstall.

e) Vi skal finne regresjonslinjen. Formelen for minste kvadraters lineære regresjonslinje står i formelsamlingen:

$$\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta} x \tag{67}$$

hvor parametrene $\hat{\beta}$ og $\hat{\alpha}$ er: (dropper enheten/benevningen)

$$\underline{\hat{\beta}} \;=\; \frac{S_{xy}}{S_x^2} \;=\; \frac{-0.1531}{0.15625} \;=\; \underline{-0.98} \tag{68}$$

$$\underline{\hat{\alpha}} \;=\; \overline{y} \;-\; \beta\overline{x} \;=\; 0.73 - (-0.98) \cdot 3.5 \;=\; \underline{4.16} \tag{69}$$

Minste kvadraters lineære regresjonslinje $\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta}\, x$ blir dermed:

$$\underline{\underline{\hat{y} \;=\; 4.16 \;-\; 0.98 \cdot x}} \tag{70}$$

f) Bruker regresjonslinjen $\hat{y} = \hat{y}(x)$ fra oppgave **4e**, dvs. lign.(70):

$$\underline{\hat{y}(3.4)} \;=\; 4.16 \;-\; 0.98 \cdot 3.4 \;=\; \underline{0.828} \tag{71}$$

Dersom prisen på en Volvo V70 er 340 000 NOK, dvs. $x = 3.4$, så predikerer regresjonslinjen at det vil selges $\underline{\underline{\text{828 biler per dag}}}$.

g) Dersom ledelsen i Volvo ønsker å selge 1 000 biler per dag, $\hat{y}(x) = 1$, så får man ifølge regresjonslinjen:

$$\hat{y}(x) \overset{\text{Eq.(71)}}{=} 4.16 \; - \; 0.98 \cdot x \tag{72}$$

$$1 = 4.16 \; - \; 0.98 \cdot x \tag{73}$$

Løser med hensyn på x:

$$\underline{x} = \frac{4.16 - 1}{0.98} = \underline{3.22} \tag{74}$$

Ifølge regresjonslinjen så vil Volvo selge 1 000 biler per dag dersom de setter prisen til $\underline{\underline{322\,000 \text{ NOK.}}}$

h) Forklaringsstyrken R^2 er: (se formelsamling)

$$\underline{\underline{R^2}} \overset{\text{def.}}{=} 1 - \frac{SSE}{SST} \tag{75}$$

$$= 1 - \frac{0.01275}{0.613} = \underline{\underline{0.9792}} \tag{76}$$

i) Kommentar til svaret i oppgave **4h**:

At $R^2 = 0.9792$ betyr at regresjonslinjen vil "i stor grad" forutsi antall solgte biler per per dag innenfor regresjonslinjens gyldighetsområde hvor vi har observasjoner.

■

Kapittel 10

Kontinuasjonseksamen 2016

LØSNING: Eksamen 5. jan 2017

"*MAT110 Statistikk 1*", vår 2016

Oppgave 1: (produksjonsplanlegging , økonomi)

a) Sannsynligheten for at malingen ikke er defekt etter 1 time er:

$$\underline{\underline{P(\overline{D}_1)}} \;=\; 1 - P(D_1) \;=\; 1 - 0.30 \;=\; \underline{\underline{0.70}} \tag{1}$$

hvor vi har brukt komplementsetningen.

b) Sannsynligheten for at malingen ikke er defekt etter 2 time
dersom vi vet at den var defekt etter 1 timer, er:

$$\underline{\underline{P(\overline{D}_2|D_1)}} \;=\; 1 - P(\overline{D}_2|D_1) \;=\; 1 - 0.10 \;=\; \underline{\underline{0.90}} \tag{2}$$

hvor vi har brukt komplementsetningen igjen.

c) Uttrykket $P(\overline{D}_2 \cap D_1)$ er sannsynligheten for at
malingen er "*defekt*" etter 1 time og "*ikke defekt*" etter 2 timer,
sannsynligheten for at malingen er OK etter etter 2 timer.

d) Den generelle multiplikasjonssetningen gir:

$$\underline{\underline{P(\overline{D}_2 \cap D_1)}} = P(\overline{D}_2|D_1)P(D_1) \tag{3}$$

$$= 0.90 \cdot 0.30 = \underline{\underline{0.27}} \quad , \quad \text{q.e.d.} \tag{4}$$

e) Vi sannsynlighetstreet som oppgitt i oppgaven finner vi at:

$$\underline{\underline{P(\overline{D}_4 \cap D_3 \cap D_2 \cap D_1)}} = P(\overline{D}_4|D_3 \cap D_2 \cap D_1)P(D_3|D_2 \cap D_1)P(D_2|D_1)P(D_1) \tag{5}$$

$$= 1 \cdot 0.02 \cdot 0.10 \cdot 0.30 = \underline{\underline{0.0006}} \quad , \quad \text{q.e.d.} \tag{6}$$

f) Bruker resulatene fra tidligere deloppgaver samt den oppgitte sannsynligheten $P(\overline{D}_3 \cap D_2 \cap D_1) = 0.0294$ til å finne forventet antall timer som Jotun bruker på blanding og justeringer av malingen:

$$\underline{\underline{E[X]}} = \sum_{i=1}^{4} x_i \cdot P(X = x_i) \tag{7}$$

$$\begin{aligned} &= 1 \cdot \overbrace{P(\overline{D}_1)}^{=\,0.70} + 2 \cdot \overbrace{P(\overline{D}_2 \cap D_1)}^{=\,0.27} \\ &+ 3 \cdot \underbrace{P(\overline{D}_3 \cap D_2 \cap D_1)}_{=\,0.0294} + 4 \cdot \underbrace{P(\overline{D}_4 \cap D_3 \cap D_2 \cap D_1)}_{=\,0.0006} \end{aligned} \tag{8}$$

$$= \big(1 \cdot 0.70 + 2 \cdot 0.27 + 3 \cdot 0.0294 + 4 \cdot 0.0006\big) \text{ timer}$$

$$= \underline{\underline{1.3306 \ \text{ timer}}} \tag{9}$$

g) Siden

$$\underline{\underline{\sum_{i=1}^{4} P(X = x_i)}} \quad = \quad P(X = x_1) \ + \ P(X = x_2) \ + \ P(X = x_3) \ + \ P(X = x_4) \tag{10}$$

$$\begin{aligned} &= \quad P(\overline{D}_1) \ + \ P(\overline{D}_2 \cap D_1) \\ &+ \quad P(\overline{D}_3 \cap D_2 \cap D_1) \ + \ P(\overline{D}_4 \cap D_3 \cap D_2 \cap D_1) \end{aligned} \tag{11}$$

$$= \quad 0.70 \ + \ 0.27 \ + \ 0.0294 \ + \ 0.0006 \tag{12}$$

$$= \quad \underline{\underline{1}} \tag{13}$$

så er sannsynlighetsfordelingen $P(X = x_i)$ en gyldig fordeling.

■

Oppgave 2: (økonomi , aksjer)

a) Fortjenesten F er:

$$\boxed{F \;=\; \text{antall aksjer} \cdot \text{pris per aksje om ett år} \;\;-\;\; \text{antall aksjer} \cdot \text{pris per aksje i dag} \qquad (14)}$$

for de tre selskapene så får vi:

$$
\begin{aligned}
\underline{\underline{F}} \;&=\; \underbrace{\text{antall aksjer i } A}_{=\,N\cdot a} \cdot \underbrace{\text{pris per aksje for selskap } A \text{ om ett år}}_{=\,X} \\
&+\; \underbrace{\text{antall aksjer i } B}_{=\,N\cdot b} \cdot \underbrace{\text{pris per aksje for selskap } B \text{ om ett år}}_{=\,Y} \\
&+\; \underbrace{\text{antall aksjer i } C}_{=\,N\cdot c} \cdot \underbrace{\text{pris per aksje for selskap } C \text{ om ett år}}_{=\,Z} \qquad (15)
\end{aligned}
$$

$$
\begin{aligned}
&-\; \underbrace{\text{antall aksjer i } A}_{=\,N\cdot a} \cdot \underbrace{\text{pris per aksje for selskap } A \text{ i dag}}_{=\,100} \qquad (16) \\
&-\; \underbrace{\text{antall aksjer i } B}_{=\,N\cdot b} \cdot \underbrace{\text{pris per aksje for selskap } B \text{ i dag}}_{=\,105} \\
&-\; \underbrace{\text{antall aksjer i } C}_{=\,N\cdot c} \cdot \underbrace{\text{pris per aksje for selskap } C \text{ i dag}}_{=\,130} \qquad (17)
\end{aligned}
$$

$$= \; NaX \;+\; NbY \;+\; NcZ \quad - \quad Na100 \;+\; Nb105 \;+\; Nc130 \qquad (18)$$

$$= \; \underline{\underline{Na\big(X-100\big) \;+\; Nb\big(Y-105\big) \;+\; Nc\big(Z-130\big)}} \qquad , \qquad \text{q.e.d.} \qquad (19)$$

b) Den forventede fortjenesten $E[F]$ ved et eventuelt salg av aksjene om ett år er:

$$\underline{E[F]} \quad = \quad E\Big[\ Na(X-100)\ +\ Nb(Y-105)\ +\ Nc(Z-130)\Big] \tag{20}$$

$$= \quad Na\Big(\underbrace{E[X]}_{=90}-100\Big)\ +\ Nb\Big(\underbrace{E[Y]}_{=125}-105\Big)\ +\ Nc\Big(\underbrace{E[Z]}_{=180}-130\Big) \tag{21}$$

$$= \quad Na\big(90-100\big)\ +\ Nb\big(125-105\big)\ +\ Nc\big(180-130\big) \tag{22}$$

$$= \quad Na(-10)\ +\ Nb\,20\ +\ Nc\,50 \tag{23}$$

$$= \quad \underline{\underline{10N\Big(-a\ +\ 2b\ +\ 5c\Big)}} \qquad , \qquad \text{q.e.d.} \tag{24}$$

c) Fra formelsamlingen vet vi formelen for variansen av en lineær kombinasjon av to stokastiske variabler X og Y, nemlig:

$$\overbrace{Var[aX+bY]}^{\text{variasjon/(spredning)}} \quad = \quad a^2\,Var[X]\ +\ b^2\,Var[Y]\ +\ 2ab\cdot\overbrace{Cov[X,Y]}^{\underline{\text{sam}}\text{variasjon}} \tag{25}$$

Dersom X og Y er uavhengige så er $Cov[X,Y]=0$.

For vårt tilfelle er de stokastiske variablene uavhengige. Da er også $Cov[X,Y,Z]=0$. Variansen $Var[F]$ til fortjenesten ved et eventuelt salg om ett år blir da:

$$\underline{\underline{Var[F]}} = Var\Big[\, Na\big(X-100\big) \;+\; Nb\big(Y-105\big) \;+\; Nc\big(Z-130\big)\,\Big] \tag{26}$$

$$\begin{aligned} &= \big(Na\big)^2 \underbrace{Var[X]}_{=\,100} + \big(Nb\big)^2 \underbrace{Var[Y]}_{=\,200} + \big(Nc\big)^2 \underbrace{Var[Z]}_{=\,600} \\ &- \big(Na\big)^2 \underbrace{Var[100]}_{=\,0} + \big(Nb\big)^2 \underbrace{Var[105]}_{=\,0} + \big(Nc\big)^2 \underbrace{Var[130]}_{=\,0} \end{aligned} \tag{27}$$

$$= N^2\Big(100a^2 \;+\; 200b^2 \;+\; 600c^2\Big) \tag{28}$$

$$= \underline{\underline{100N^2\left(a^2 \;+\; 2b^2 \;+\; 6c^2\right)}} \qquad , \qquad \text{q.e.d.} \tag{29}$$

d) Dersom man ønsker minst mulig risiko i investeringen så tilsvarer det at variansen $Var[F]$ er minst mulig. I oppgaven er det oppgitt at denne variansen er minst når $a = 0.6$, $b = 0.3$ og $c = 0.1$.

Antall aksjer som må kjøpes i de forskjellige selskapene for å oppnå dette er da:

$$\underline{\underline{Na}} = 100\,000 \cdot 0.6 = \underline{\underline{60\,000}} \tag{30}$$

$$\underline{\underline{Nb}} = 100\,000 \cdot 0.3 = \underline{\underline{30\,000}} \tag{31}$$

$$\underline{\underline{Nc}} = 100\,000 \cdot 0.1 = \underline{\underline{10\,000}} \tag{32}$$

e) Fra oppgave **3b** vet vi at den forventede fortjenetsen $E[F]$ er

$$E[F] \;=\; 10N\Big(-a \;+\; 2b \;+5c\Big) \tag{33}$$

Fra denne ligningen ser vi umiddelbart at $E[F]$ blir størst når vi kun investerer kun investerer i selskap C, dvs.

$$a = 0 \qquad , \qquad b = 0 \qquad , \qquad c = 1 \tag{34}$$

Verdien på den forventede fortjenesten $E[F]$ blir da:

$$\underline{\underline{E[F]}} \quad = \quad 10N \cdot 5c \tag{35}$$

$$= \quad 10 \cdot 100\,000 \cdot 5 \cdot 1 \;\; \text{NOK} \;=\; \underline{\underline{5\,000\,000 \;\; \text{NOK}}} \tag{36}$$

altså 5 mill. NOK.

f) Forventningene er: [1]

$$\underline{\underline{E[X]}} \;=\; E[\,90+10\epsilon\,] \;=\; 90 \;+\; 10\,\underbrace{E[\epsilon]}_{=0} \;=\; \underline{\underline{90\;\;\text{NOK}}} \tag{37}$$

$$\underline{\underline{E[Y]}} \;=\; E[\,125+10\sqrt{2}\epsilon\,] \;=\; 125 \;+\; 10\sqrt{2}\,\underbrace{E[\epsilon]}_{=0} \;=\; \underline{\underline{125\;\;\text{NOK}}} \tag{38}$$

$$\underline{\underline{E[X]}} \;=\; E[\,180+10\sqrt{6}\epsilon\,] \;=\; 180 \;+\; 10\sqrt{6}\,\underbrace{E[\epsilon]}_{=0} \;=\; \underline{\underline{180\;\;\text{NOK}}} \tag{39}$$

g) Variansene er: (se lign.(25))

$$\underline{\underline{Var[X]}} \;=\; Var[\,90+10\epsilon\,] \;=\; 0 \;+\; 10^2\,\underbrace{Var[\epsilon]}_{=1} \;=\; \underline{\underline{100\;\;\text{NOK}^2}} \tag{40}$$

$$\underline{\underline{Var[Y]}} \;=\; Var[\,125+10\sqrt{2}\epsilon\,] \;=\; 0 \;+\; (10\sqrt{2})^2\,\underbrace{Var[\epsilon]}_{=1} \;=\; \underline{\underline{200\;\;\text{NOK}^2}} \tag{41}$$

$$\underline{\underline{Var[X]}} \;=\; Var[\,180+10\sqrt{6}\epsilon\,] \;=\; 0 \;+\; (10\sqrt{6})^2\,\underbrace{Var[\epsilon]}_{=1} \;=\; \underline{\underline{600\;\;\text{NOK}^2}} \tag{42}$$

h) Ved å sammenligne svarene i oppgavene **3f** og **3g** med forventningene og variansene som er oppgitt i oppgaven så innser vi at forventningene og variansene til de individuelle stokastiske variablene <u>er de samme</u> i disse to tilfellene. [2]

[1]Kommentar: ϵ kan oppfattes som en **markedsindikator**. "Ting går bra" når denne idikatoren er positiv, og tilsvarende dårlig når ϵ er negativ. (En slik kommentar behøver man ikke ha med på eksamensbesvarelsen.)

[2]Dette til tross for at de stokastiske variablene er uavhengige i det ene tilfellet og avhengige i det andre.

i) Siden X er normalfordelt så kan man finne sannsynligheten

$$P\Big(E[X] - \sigma[X] \leq X \leq E[X] + \sigma[X]\Big) \tag{43}$$

ved oppslag i formelsamlingen. Vi finner: [3]

$$\underline{P\Big(E[X] - \sigma[X] \leq X \leq E[X] + \sigma[X]\Big) \;=\; 0.682} \tag{44}$$

Altså:
Sannsynligheten for at aksjekursen for selskap A ligger mellom 80 NOK og 100 NOK om ett år er $\underline{\underline{68.2\ \%}}$. [4]

j) Fortjenesten F om ett år er:

$$F \;=\; 10N\big(a + \sqrt{2}b + \sqrt{6}c\big)\,\epsilon \tag{45}$$

Variansen til fortjenesten F er da: (se lign.(25))

$$\underline{\underline{Var[F]}} \;=\; Var\Big[\,10N\Big(a + \sqrt{2}b + \sqrt{6}c\Big)\,\epsilon\,\Big] \tag{46}$$

$$= \;\Big(10N\big(a + \sqrt{2}b + \sqrt{6}c\big)\Big)^2\,\underbrace{Var[\epsilon]}_{=\,1} \tag{47}$$

$$= \;\underline{\underline{(10N)^2\Big(a + \sqrt{2}b + \sqrt{6}c\Big)^2}} \qquad , \qquad \text{q.e.d.} \tag{48}$$

[3]Se side 65 i 2016-versjonen av formelsamlingen.
[4]Legg merke til at siden $Var[X] = 100\ \text{NOK}^2$ så er $\sigma[X] = 10$ NOK. Dessuten er $E[X] = 90$ NOK.

k) Avhengige aksjer:

Når aksjene er avhengige slik som beskrevet i oppgaven så er $a = 1$, $b = 0$ og $c = 0$.
Det betyr at det er minst risiko å plassere alle $N = 100\,000$ aksjene i selskap A.

Uavhengige aksjer:

Når aksjene er uavhengige, derimot, så er $a = 0.6$, $b = 0.3$ og $c = 0.1$.
Det betyr at det minst risiko å spre aksjene over alle tre selskapene med en fordeling som i oppgave **3d**.

■

Oppgave 3: (logistikk)

a) Forventet antall akutte utrykninger i fem kommunene til sammen per uke:

$$\underline{\underline{E[Y]}} = E[X_1 + X_2 + X_3 + X_4 + E_5] \tag{49}$$

$$\stackrel{\text{alltid}}{=} E[X_1] + E[X_2] + E[X_3] + E[X_4] + E[X_5] \tag{50}$$

$$= \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 \tag{51}$$

$$= 0.97 + 2.43 + 4.29 + 5.02 + 7.21 = \underline{\underline{19.92}} \tag{52}$$

NB: Overgangen i lign.(49) til (50) gjelder alltid. Uansett om de stokastiske variablene X_i er uavhengige eller ikke.
(Se lign.(5.11) på side 39 i formelsamlingen fra 2016).

b) For en Poissonfordeling er forventing og varians er like, dvs. $E[Y] = Var[Y] = 19.92.$ [5]
Standardavviket til antall akutte utrykninger i fem kommunene til sammen per uke er dermed: [6]

$$\underline{\underline{\sigma[Y]}} = \sqrt{Var[Y]} = \sqrt{19.92} = \underline{\underline{4.46}} \tag{53}$$

c) En Poissonfordelt stokastisk variabel er med god tilnærmelse normalfordelt dersom $\lambda \gtrsim 5$.
Siden $\underline{\underline{\lambda_Y = E[Y] = 19.92 \gg 5}}$ så er da Y normalfordelt:

$$Y \sim \mathrm{N}\big[E[Y], \sigma[Y]\big] \tag{54}$$

[5] Se f.eks. side 58 i formelsamlngen fra 2016.

[6] I denne oppgaven behøves det ingen utregninger.

d) Siden Y er (tilnærmet) normalfordelt så kan vi bruke dette når vi skal regne ut sannsynligheten $P(Y > 25)$: [7]

$$\underline{P(Y > 25)} \overset{\text{standardiser}}{=} P\left(\underbrace{\frac{Y - E[Y]}{\sigma[\,Y\,]}}_{=\,Z} > \frac{25 - E[Y]}{\sigma[Y]} \right) \tag{55}$$

$$= \quad P\left(Y \;>\; \frac{25 - 19.92}{\sqrt{19.92}} \right) \tag{56}$$

$$= \quad P(\,Z \;>\; 1.14\,) \tag{57}$$

$$= \quad 1 \;-\; P(\,Z \;\leq\; 1.14\,) \tag{58}$$

$$= \quad 1 \;-\; \underbrace{G(\,1.14\,)}_{=\;0.8729} \tag{59}$$

$$\overset{\text{tabell}}{=} \quad 1 \;-\; 0.8729 \tag{60}$$

$$= \quad \underline{0.1271} \tag{61}$$

Sannsynligheten for at det blir mer enn 25 akutte utrykninger per uke i de 5 aktuelle kommunene er $\underline{\underline{12.71\ \%}}$.

[7] I oppgaven stod det at vi **ikke** behøver heltallskorreksjon. Derfor utelater vi det.

e) Siden $X_2 \sim \text{Poi}[\lambda_2]$ (oppgitt i oppgaven) så finner man sannsynligheten for at det skjer 2 utrykninger i løpet av en uke i Eide på følgende måte: [8]

$$\underline{\underline{P(X_2 = 2)}} \quad = \quad \frac{\lambda_2^2}{2!}\, e^{-\lambda_2} \tag{62}$$

$$\overset{\lambda_2=2.43}{=} \quad \frac{2.43^2}{2!}\, e^{-2.43} \quad \overset{\text{kalkis}}{=} \quad \underline{\underline{0.2599}} \tag{63}$$

f) Forventet antall akutte utrykninger i året i Eide kommune:

$$\underline{\underline{E[\, Z_{\text{år}} \,]}} \quad = \quad E\,[\, Z_1 + Z_2 + \;\ldots\; + Z_n \,] \quad = \quad \overbrace{E[Z_1] + E[Z_2] + \;\ldots\; + E[Z_n]}^{\text{n stk.}} \tag{64}$$

$$= \quad n\, \underbrace{E[Z_j]}_{\lambda_j = 2.43} \quad = \quad 52 \cdot 2.43 \quad = \quad \underline{\underline{126.36}} \tag{65}$$

NB: Overgangen i lign.(64) gjelder alltid. Uansett om de stokastiske variablene Z_i er uavhengige eller ikke.

g) Variansen til antall akutte utrykninger i året for Eide kommune:

$$\underline{\underline{Var[\, Z_{\text{år}} \,]}} \quad = \quad Var\,[\, Z_1 + Z_2 + \;\ldots\; + Z_n \,] \tag{66}$$

$$\overset{\text{uavh.}}{=} \quad \overbrace{Var[Z_1] + Var[Z_2] + \;\ldots\; + Var[Z_n]}^{\text{n stk.}} \tag{67}$$

$$= \quad n\, \underbrace{Var[Z_j]}_{\lambda_j = 2.43} \quad = \quad 52 \cdot 2.43 \quad = \quad \underline{\underline{126.36}} \tag{68}$$

NB: Overgangen i lign.(66) gjelder *kun* dersom de stokastiske variablene Z_j er <u>u</u>avhengige.

[8] Se side 56 i formelsamlingen fra 2016.

h) i) Med forutsetningene som formulert i oppgaven så gjelder sentralgrensesetningen.

ii) Ifølge sentralgrensesetningen er da den stokastiske variabelen $Z_{\text{år}}$ **normal**fordelt:

$$\underline{\underline{Z_{\text{år}} \quad \sim \quad \mathrm{N}[\, E[\, Z_{\text{år}}\,]\,,\, Var[\, Z_{\text{år}}\,]\,]}} \tag{69}$$

iii) En **tommelfingerregel** for at sentralgrensesetningen skal gjelde er:

$$\underline{\underline{n \;\gtrsim\; 30}} \tag{70}$$

dvs. antall forsøk bør være ca. 30 eller mer.

i) Siden $Z_{\text{år}}$ er (tilnærmet) normalfordelt så kan vi bruke dette når vi skal regne ut sannsynligheten $P(Z_{\text{år}} < 115)$: [9]

$$
\begin{aligned}
\underline{P(Z_{\text{år}} < 115)} &\overset{\text{standardiser}}{=} P\left(\underbrace{\frac{Z_{\text{år}} - E[Z_{\text{år}}]}{\sigma[\,Z_{\text{år}}\,]}}_{=\,Z} < \frac{115 - E[Z_{\text{år}}]}{\sigma[Z_{\text{år}}]}\right) && (71)\\
&= P\left(Z \; < \; \frac{115 - 126.36}{\sqrt{126.36}}\right) && (72)\\
&= P(\,Z \; < -1.01\,) && (73)\\
&= 1 \; - \; P(\,Z \; < \; 1.01\,) && (74)\\
&= 1 \; - \; \underbrace{G(\,1.01\,)}_{=\,0.8438} && (75)\\
&\overset{\text{tabell}}{=} 1 \; - \; 0.8738 && (76)\\
&= \underline{0.1562} && (77)
\end{aligned}
$$

Sannsynligheten for at det blir mindre enn 115 akutte utrykninger i året i Eide kommune er $\underline{\underline{15.62\ \%}}$.

■

[9] I oppgaven stod det at vi **ikke** behøver heltallskorreksjon. Derfor utelater vi det.

Oppgave 4: (logistikk)

a) Uttrykket for korrelasjonskoeffisienten finner man i formelsamlingen. Det gir:

$$\underline{\underline{R_{xy}}} = \frac{S_{xy}}{S_x \cdot S_y} \tag{78}$$

$$= \frac{6863.26}{\sqrt{8\,284\,549} \cdot \sqrt{5.76828}} = \underline{\underline{0.99}} \tag{79}$$

b) At $R_{xy} = 0.99$ betyr at det er en veldig <u>sterk positiv korrelasjon</u> mellom x og y.

Det er altså en svært <u>sterk lineær sammenheng</u> mellom x og y med positivt stigningstall.

c) Vi skal finne regresjonslinjen. Formelen for minste kvadraters lineære regresjonslinje står i formelsamlingen:

$$\hat{y} = \hat{\alpha} + \hat{\beta}x \tag{80}$$

hvor parametrene $\hat{\beta}$ og $\hat{\alpha}$ er: (dropper enheten/benevningen)

$$\underline{\hat{\beta}} = \frac{S_{xy}}{S_x^2} = \frac{6863.26}{8\,284\,549} = \underline{0.000828} \tag{81}$$

$$\underline{\hat{\alpha}} = \overline{y} - \beta\overline{x} = 3.984 - 0.000828 \cdot 5745.4 = -\underline{0.77} \tag{82}$$

Minste kvadraters lineære regresjonslinje $\hat{y} = \hat{\alpha} + \hat{\beta}\,x$ blir dermed:

$$\underline{\underline{\hat{y} = -0.77 + 0.000828 \cdot x}} \tag{83}$$

d) Bruker regresjonslinjen $\hat{y} = \hat{y}(x)$ fra oppgave **4c**, dvs. lign.(83):

$$\underline{\hat{y}(7492)} \quad = \quad -0.77 \; + \; 0.000828 \cdot 7492 \quad = \quad \underline{5.43} \tag{84}$$

Rauma kommune med 7492 innyggere vil, ifølge regresjonslinjen i lign.(84), ha $\underline{\underline{\text{5.43 utrykninger per dag}}}$.

e) Løser mhp. x alene:

$$\hat{y} \quad \overset{\text{Eq.(84)}}{=} \quad \hat{\alpha} \; + \; \hat{\beta} \cdot x \tag{85}$$

$$\hat{y} \; - \; \hat{\alpha} \quad = \quad \hat{\beta} \cdot x \tag{86}$$

$$\underline{x \; = \; \frac{\hat{y} - \alpha}{\beta}} \tag{87}$$

Med $\hat{y} = 6$ utrykninger per uke, $\hat{\alpha} = -0.77$ og $\hat{\beta} = 0.000838$ så får vi:

$$\underline{x} \; = \; \frac{6 - (-0.77)}{0.000828} \; = \; \underline{8176} \tag{88}$$

Ifølge regresjonslinjen så må det være minst $\underline{\underline{\text{8176 innbyggere}}}$ for at Helse M&R må ha èn ekstra person på telefonberedskap. [10]

■

[10] Dette svaret er litt følsomt for avrundinger av koeffisientene $\hat{\alpha}$ og $\hat{\beta}$. Man kan få full score på denne oppgaven selv om man ikke får akkurat svaret 8176.

Kapittel 11

Hovedeksamen 2017

LØSNING: Eksamen 24. mai 2017

"*MAT110 Statistikk 1*", vår 2017

Oppgave 1: (logistikk)

a) Betingelsene

- uniformt utfallsrom
- utfallene ikke påvirker hverandre, dvs. de er uavhengige

må være oppfylte for at urnmodellen skal gjelde.

I oppgaven antar man at pakkene er uavhengige av hverandre.
I tillegg antas det i oppgaven at det er lik sannsynlighet for å trekke de forskjellige pakkene fra urnen, altså uniformt utfallsrom.

Ja, urnemodellen gjelder for vårt tilfelle med pakker.

b) Med kun ett trekk ($s = 1$) så finner man sannsynlighetene ved svært korte utregninger:

$$\underline{\underline{P(\overline{S} \cap \overline{F})}} = \frac{\#\text{ g}}{\#\text{ m}} = \frac{70}{100} = \underline{\underline{0.70}} \tag{1}$$

$$\underline{\underline{P(\overline{S} \cap F)}} = \frac{\#\text{ g}}{\#\text{ m}} = \frac{15}{100} = \underline{\underline{0.15}} \tag{2}$$

$$\underline{\underline{P(S \cap \overline{F})}} = \frac{\#\text{ g}}{\#\text{ m}} = \frac{10}{100} = \underline{\underline{0.10}} \tag{3}$$

$$\underline{\underline{P(S \cap F)}} = \frac{\#\text{ g}}{\#\text{ m}} = \frac{5}{100} = \underline{\underline{0.05}} \tag{4}$$

c) Bruker resultatene fra oppgave **1b** og regner ut
JetPak sine forventede utgifter per pakke:

$$\begin{aligned}\underline{\underline{E[U]}} &= \sum_{i=1}^{4} u_i \cdot P(U = u_i) \\ &= \left(0 \cdot \overbrace{P(\overline{S} \cap \overline{F})}^{=\,0.70} + 500 \cdot \overbrace{P(\overline{S} \cap F)}^{=\,0.15} + 900 \cdot \overbrace{P(S \cap \overline{F})}^{=\,0.10} + 1\,500 \cdot \overbrace{P(S \cap F)}^{=\,0.05} \right) \text{NOK} \qquad (5) \\ &= \underline{\underline{240 \text{ NOK}}} \qquad (6)\end{aligned}$$

d) Fra figuren i oppgaver ser vi at de 4 aktuelle kombinasjonene av begivenheter
ikke overlapper. Derfor er de disjunkte. [1]

[1] Ekstrakommentar som man ikke behøver å ha med på eksamen:
I dette tilfellet overlapper de akkurat ikke - i analogi med et puslespill.

e) Bruker addisjonssetningen og resultater fra oppgave **1a**:

$$\underline{\underline{P\big((\overline{S}\cap F)\cup(S\cap\overline{F})\big)}} = \overbrace{P\big((\overline{S}\cap F)\big)}^{=\ 0.10} + \overbrace{P\big((S\cap\overline{F})\big)}^{=\ 0.15} - \overbrace{P\big((\overline{S}\cap F)\cap(S\cap\overline{F})\big)}^{=\ 0\quad(\text{disjunkt})} \tag{7}$$

$$= 0.10 + 0.15 = \underline{\underline{0.25}} \quad , \quad \text{q.e.d.} \tag{8}$$

Merk: Fra oppgave **1d** vet vi at $\overline{S}\cap F$ og $S\cap\overline{F}$ er disjunkte.

f) Med kun ett trekk ($s = 1$) så finner man den aktuelle sannsynligheten ved en svært korte utregninger:

$$\underline{\underline{P\big((\overline{S}\cap F)\cup(S\cap\overline{F})\big)}} = \frac{\#\ \mathrm{g}}{\#\ \mathrm{m}} = \frac{15 + 10}{100} = \underline{\underline{0.25}} \tag{9}$$

altså sammen sannsynlighet som i oppgave **1e** – slik som det skal være.

g) Oppsplitting av utfallsrom:

$$\underline{\underline{P(S)}} = P(S\cap F) + P(S\cap\overline{F}) = 0.05 + 0.10 = \underline{\underline{0.15}} \tag{10}$$

h) Bruker den ene tvillingsetningen samt den oppgitte sannsynligheten:

$$\underline{\underline{P\big(S \cap F\big)}} = 1 - P(\overline{S} \cup \overline{F}) \tag{11}$$

$$= 1 - 0.95 = \underline{\underline{0.05}} \tag{12}$$

som stemmer med resultatet fra oppgave **1b**.

Dette er slik det må være siden vi i oppgavene **1b** og **1h** regner ut samme sannsynlighet med to forskjellige metoder.

i) I oppgaven er det oppgitt at:

- Rekkefølgen pakkene trekkes i betyr ikke noe
- Når en pakke er trykket så legger man den ikke tilbake i urnen, dvs. uten tilbakelegging

Dermed tilsvarer dette situasjon 3.

j) Sannsynligheten P_{mix} for at Jetpak leverer 3 pakker fra hver av de 4 aktuelle begivenhetene:

$$\underline{\underline{P_{\text{mix}}}} = \frac{\#\ \text{g}}{\#\ \text{m}} \tag{13}$$

$$= \frac{\#\ \text{komb. } S \cap \overline{F} \cdot \#\ \text{komb. } \overline{S} \cap F \cdot \#\ \text{komb. } S \cap F \cdot \#\ \text{komb. } \overline{S} \cap \overline{F}}{\text{totalt}} \tag{14}$$

$$= \frac{\binom{10}{3} \cdot \binom{15}{3} \cdot \binom{5}{3} \cdot \binom{70}{3}}{\binom{100}{12}} = \underline{\underline{2.85 \cdot 10^{-5}}} \tag{15}$$

altså svært liten sannsynlighet. [2]

■

[2] Legg merke til at "summeregelen" er opfyllt i lign.(15).

Oppgave 2: (økonomi)

a) "Forsøksserien" med $n = 120$ bedrifter som kan gå konkurs har følgende **egenskaper**:

1. Hver bedrift har kun 2 mulige utfall, konkurs eller ikke konkurs.
2. I oppgaven antas det at alle bedriftene har samme sannsynlighet p $(= 0.05)$ for å gå konkurs.
3. Eventuell konkurs i en bedrift er uavhengig om en anne bedrift går konkurs.
4. Vi gjennomfører et bestemt antall forsøk, $n = 120$ i dette tilfellet.

Alle de 4 forutsetningene for en binomisk fordeling er oppfylt.
Derfor er det rimelig å anta at X er binomisk fordelt, dvs.

$$X \sim \text{Bin}[\, n = 120, p = 0.05\,] \tag{16}$$

b) Sannsynligheten for at nøyaktig 2 av de 120 bedriftene går konkurs i løpet av et gitt år:

$$\underline{\underline{P(X = 2)}} = \binom{n}{2} p^2 \, (1 - p)^{n-2} \tag{17}$$

$$= \binom{120}{2} 0.05^2 \, (1 - 0.05)^{120-2} = \underline{\underline{0.0420}} \tag{18}$$

c) Sannsynligheten for at høyst 2 av de 120 bedrifter går konkurs i løpet av et gitt år:

$$\underline{\underline{P(X \leq 2)}} = P(X=0) + P(X=1) + P(X=2) \tag{19}$$

$$= \binom{n}{0} p^0 (1-p)^{n-0} + \binom{n}{1} p^1 (1-p)^{n-1} + \binom{n}{2} p^2 (1-p)^{n-2} \tag{20}$$

$$= \binom{120}{0} 0.05^0 (1-0.05)^{120-0} \tag{21}$$

$$+ \binom{120}{1} 0.05^1 (1-0.05)^{120-1} \tag{22}$$

$$+ \binom{120}{2} 0.05^2 (1-0.05)^{120-2} \tag{23}$$

$$= 0.0021 + 0.0134 + 0.0420 = \underline{\underline{0.0575}} \tag{24}$$

d) 10 observasjoner:
Den generelle trenden er – over en periode på 10 år – at konkurs-sannsynligheten er 5 % i løpet av et år.
Konkurs-sannsynligheten på 5 % er altså basert på 10 observasjoner.

En observasjon:
Journalisten, derimot, konkluderer med at konkurs-sannsynligheten for 2017 er 1.7 % som kun er basert på èn observasjon fra 2016, altså kun basert på ett år.
Journalisten tar altså ikke med dataene fra de 9 foregående årene, og vil derfor ha et dårligere grunnlag/estimat for konkurs-sannsynligheten enn om alle 10 observasjonene hadde blitt inkludert.

Konklusjon:
Journalisten baserer sin konklusjon på et for snevert grunnlag – kun en observasjon istedet for 10.

e) *i*) Betingelse som må være oppfylt så for at en binomisk fordeling kan tilnærmes med en normalfordeling er: [3]

$$\underline{\underline{n \cdot p\,(1-p) \;\gtrsim\; 5}} \tag{25}$$

ii) For vårt tilfelle:

$$120 \cdot 0.05\,(1-0.05) \;=\; 5.7 \quad\gtrsim\quad 5 \tag{26}$$

Ja, betingelsen er såvidt oppfylt for vårt tilfelle.

[3]Se formelsamlingen.

f) Sannsynligheten for at høyst 2 av de 120 bedrifter går konkurs når vi bruker tilnærmelsen om at X er normalfordelt:

$$\underline{\underline{P(X \leq 2)}} = P\Bigg(\underbrace{\frac{X - E[X]}{\sigma[X]}}_{= Z} \leq \frac{2 - n \cdot p}{\sqrt{n \cdot p(1-p)}}\Bigg) \tag{27}$$

$$= P\left(Z \leq \frac{2 - 120 \cdot 0.05}{\sqrt{120 \cdot 0.05\,(1 - 0.05)}}\right) \tag{28}$$

$$= P(Z \leq -1.68) \tag{29}$$

$$= 1 - \underbrace{P(Z \leq 1.68)}_{\overset{\text{tabell}}{=} 0.9535} \tag{30}$$

$$= 1 - 0.9535 \tag{31}$$

$$= \underline{\underline{0.0465}} \tag{32}$$

g) Fra oppgave **2c**: $P(X \leq 2) = 0.0575$
Fra oppgave **2f**: $P(X \leq 2) = 0.0465$

Ved å bruke normalfordelingen som tilnærmelse for binomialfordelingen så bommer vi med drøy prosent. Men svarene er "ganske" like.

Siden vi kun såvidt oppfyller kriteriet - jfr. oppgave **2e** *ii* - og siden dette er en tilnærmelse så må man regne med litt unøyaktighet.

h) Siden konkurs-sannsynligeheten avhenger av ytre faktorer så er det ikke lenger rimelig å anta at konkurs eller ikke i de disse IKT-bedriftene er uavhengige.

Fra oppgave **2a** vet vi da at èn av betingelsene ikke er oppfylt for at vi skal ha en binomisk forsøksserie - betingelsen om uavhengighet. Derfor:

Nei, det er $\underline{\underline{\text{ikke rimelig}}}$ at X er binomisk fordelt lenger.

■

Oppgave 3: (logistikk)

a) Siden den stokastiske variabelen X har et tellbart antall mulige verdier så er X $\underline{\underline{\text{diskret}}}$.

b) Enhver gyldig sannsynlighetsfordeling er $\underline{\underline{\text{normalisert til 1}}}$, dvs. $\sum_{i=0}^{5} P(X = x_i) = 1$.
For at denne normaliseringsbetingelsen skal være oppfylt så må:
$P(X = 3) = 0.29$.

c) Forventning $E[X]$:

$$\underline{\underline{E[X]}} \overset{\text{def.}}{=} \sum_{i=0}^{5} x_i \cdot P(X = x_i) \tag{33}$$

$$= \quad 0 \cdot \overbrace{P(X=0)}^{=0.03} + 1 \cdot \overbrace{P(X=1)}^{=0.08} + 2 \cdot \overbrace{P(X=2)}^{=0.15}$$

$$+ 3 \cdot \underbrace{P(X=3)}_{=0.29} + 4 \cdot \underbrace{P(X=4)}_{=0.25} + 5 \cdot \underbrace{P(X=5)}_{=0.20} \tag{34}$$

$$= 0 \cdot 0.03 + 1 \cdot 0.08 + 2 \cdot 0.15 + 3 \cdot 0.29 + 4 \cdot 0.25 + 5 \cdot 0.20 \tag{35}$$

$$= \quad \underline{\underline{3.25}} \tag{36}$$

d) I oppgaven var det oppgitt at:

$$E[X^2] \;=\; 12.29 \tag{37}$$

Dette innsatt i "*varianssetningen*": (se formelsamling) [4]

$$\underline{Var[X]} \;=\; E[X^2] - E[X]^2 \quad = \; 12.29 - 3.25^2 \;=\; \underline{1.7275} \tag{38}$$

som gir standardavviket

$$\underline{\underline{\sigma[X]}} \;=\; \sqrt{Var[X]} \tag{39}$$

$$\;=\; \sqrt{1.7275} \;=\; \underline{\underline{1.3143}} \tag{40}$$

Tolking:

$$\underline{\underline{\sigma[X] \;=\; \text{avvik/spredning i antall utleide biler per dag for AVIS}}}$$

[4]Istedet for å bruke "*varianssetningen*" kan man bruke definisjonen av varians, $Var[X] \stackrel{\text{def.}}{=} E\big[\,(X - E[X])^2\,\big]$, for å finne svaret. Begge tilnærmelser er like riktige. Og gir selvsagt samme svar. Men siden $E[X^2]$ var oppgitt så gir tilnærmelsen med varianssetningen minst arbeid.

e) Sannsynligheten for overskudd en gitt dag er $P(F > 0)$.
Overskudd $F > 0$ inntreffer når antall utleide biler per dag er $X > b/a = 3.03$.
Dermed:

$$\begin{aligned} \underline{\underline{P(F > 0)}} &= P(X \geq 4) \\ &= P(X = 4) \;+\; P(X = 5) \;=\; 0.25 \;+\; 0.20 \;=\; \underline{\underline{0.45}} \end{aligned} \tag{41}$$

f) Forventet fortjeneste $E[F]$ en gitt dag:

$$\underline{\underline{E[\,F\,]}} \;=\; E\,[\,aX - b\,] \;=\; a\,\overbrace{E[X]}^{=3.25} - b \tag{42}$$

$$= \Big(620 \cdot 3.25 \;-\; 1880\Big)\ \text{NOK} \;=\; \underline{\underline{135\ \ \text{NOK}}} \tag{43}$$

hvor vi har brukt svaret fra oppgave **3c**, dvs. $E[X] = 3.25$.

g) Variansen til fortjeneste $Var[F]$ en gitt dag:

$$\underline{\underline{Var[\,F\,]}} \;=\; Var\,[\,aX - b\,] \;=\; a^2\,\overbrace{Var[X]}^{=1.7275} - 0 \tag{44}$$

$$= 620^2 \cdot 1.7275\ \ \text{NOK}^2/\text{dag}^2 \;=\; \underline{\underline{664\,051\ \ \text{NOK}^2/\text{dag}^2}} \tag{45}$$

hvor vi har brukt svaret fra oppgave **3d**, dvs. $Var[X] = 1.7275$.

h) Som fotnoten i oppgaven sier:
Det er totalt 6 måter å få leie ut to biler på over en periode på tre dager:

$$\underline{\underline{P(2 \text{ biler})}} = p_0 \cdot p_1 \cdot p_1 + p_1 \cdot p_0 \cdot p_1 + p_1 \cdot p_1 \cdot p_0 + p_2 \cdot p_0 \cdot p_0 + p_0 \cdot p_2 \cdot p_0 + p_0 \cdot p_0 \cdot p_2 \tag{46}$$

$$= p_0 \cdot p_1^2 + p_1^2 \cdot p_0 + p_1^2 \cdot p_0 + p_2 \cdot p_0^2 + p_0^2 \cdot p_2 + p_0^2 \cdot p_2 \tag{47}$$

$$= \underline{\underline{3\, p_0\, p_1^2 + 3\, p_0^2\, p_2}} \quad , \quad \text{q.e.d.} \tag{48}$$

i) Sannsynligheten $P(2 \text{ biler})$ for at AVIS leier ut to biler i løpet av en periode på tre dager: (tallene finner vi i tabellen i oppgaven)

$$\underline{\underline{P(2 \text{ biler})}} \overset{\text{Eq.(48)}}{=} 3\, p_0\, p_1^2 + 3\, p_0^2\, p_2 \tag{49}$$

$$= 3 \cdot 0.03 \cdot 0.08^2 + 3 \cdot 0.03^2 \cdot 0.15 = \underline{\underline{0.000981}} \tag{50}$$

altså i underkant av 0.1 %.

j) i) Med forutsetningene som formulet i oppgaven så gjelder sentralgrensesetningen.

ii) Ifølge sentralgrensesetningen er da den stokastiske variabelen Y **normal**fordelt:

$$\underline{\underline{Y \;\sim\; \mathrm{N}\big[\, E[Y\,]\,, Var[Y\,]\,\big]}} \tag{51}$$

iii) En **tommelfingerregel** for at sentralgrensesetningen skal gjelde er:

$$\underline{\underline{n \;\gtrsim\; 30}} \tag{52}$$

dvs. antall forsøk bør være ca. 30 eller mer.

k) Forventet fortjeneste i løpet av et helt år $E[\,Y\,]$: [5]

$$\underline{\underline{E[\,Y\,]}} \;=\; E\,[\,F_1 + F_2 + \;...\; + F_n\,] \;\stackrel{\text{alltid}}{=}\; \overbrace{E[F_1] + E[F_2] + \;...\; + E[F_n]}^{\text{n stk.}} \tag{53}$$

$$=\; n\,E[F] \;=\; 250 \cdot 135\ \text{ NOK} \;=\; \underline{\underline{33\,750\ \text{ NOK}}} \tag{54}$$

hvor vi har brukt svaret fra oppgave **3f**, dvs. $E[F] = 135$ NOK.

[5]NB: Overgangen i lign.(53) gjelder alltid - uansett om de stokastiske variablene F_i er uavhengige eller ikke.

l) Variansen til fortjenesten over et helt år er:

$$\underline{\underline{Var[Y]}} \quad = \quad Var\left[\, F_1 + F_2 + \ \dots\ + F_n \,\right] \tag{55}$$

$$\stackrel{\text{uavh.}}{=} \quad \overbrace{Var[F_1] + Var[F_2] + \ \dots\ + Var[F_n]}^{\text{n stk.}} \tag{56}$$

$$= \quad n \cdot \underbrace{Var[F]}_{=664\,051} \quad = \quad 250 \cdot 664\,051 \quad \text{NOK}^2/\text{dag}^2 \quad = \quad \underline{\underline{166\,012\,750 \quad \text{NOK}^2/\text{dag}^2}}$$

hvor vi har brukt svaret fra oppgave **3g**, dvs. $Var[F] = 664\,051 \quad \text{NOK}^2/\text{dag}^2$.

NB: Overgangen i lign.(55) gjelder *kun* dersom de stokastiske variablene F_i er <u>u</u>avhengige.

■

Oppgave 4: (økonomi)

a) **i**) En regresjonslinje mellom variablene x og y sier at for en gitt verdi av x så kan y estimeres/predikeres via regresjonslinjen.

ii) Parameteren $\hat{\beta}$ er stigningstallet for en lineær regresjonslinje. Parameteren angir estimert/predikert endring $\hat{y}$ når x endres med èn enhet.

iii) Parameteren $\hat{\beta}$ er oppgitt i oppgaven: $\hat{\beta} = 1168.5$
Det betyr at antall utrykninger øker med 1168.5 per år.

b) Regresjonslinjen predikerer at antall utrykninger i år 2020, dvs. $x = 10$, er:

$$\underline{\underline{\hat{y}(10)}} \;=\; 5226.5 \;+\; 1168.5 \cdot 10 \;=\; \underline{\underline{16\,911.5}} \tag{57}$$

c) Korrelasjonskoeffisienten R_{xy} er definert ved: (se formelsamling)

$$R_{xy} \;=\; \frac{S_{xy}}{S_x S_y} \tag{58}$$

Definisjonen av koeffisienten $\hat{\beta}$ i regresjonslinjen er: (se formelsamling)

$$\hat{\beta} \;=\; \frac{S_{xy}}{S_x^2} \tag{59}$$

og løser denne med hensyn på S_{xy} alene:

$$S_{xy} \;=\; \hat{\beta} S_x^2 \tag{60}$$

Setter lign.(60) inn i lign.(58):

$$\underline{\underline{R_{xy}}} = \frac{S_{xy}}{S_x S_y} \tag{61}$$

$$= \frac{\hat{\beta} S_x^{\cancel{2}}}{\cancel{S_x} S_y} = \underline{\underline{\hat{\beta}\,\frac{S_x}{S_y}}} \quad , \quad \text{q.e.d.} \tag{62}$$

d) Korrelasjonskoeffisienten R_{xy} for observasjonene blir da:

$$\underline{\underline{R_{xy}}} = \hat{\beta}\,\frac{S_x}{S_y} \tag{63}$$

$$= 1168.5\,\frac{1.87}{2213.9} = \underline{\underline{0.99}} \tag{64}$$

e) Siden $R_{xy} = 0.99$ så er det et <u>sterk positiv korrelasjon</u> mellom x og y.
Det er en <u>nesten perfekt lineær sammenheng</u> mellom x og y,
med positivt stigningstall.

f) Definisjonen av koeffisienten $\hat{\alpha}$: (se formelsamling)

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x} \tag{65}$$

og denne med hensyn på $\overline{y}$ alene

$$\underline{\underline{\overline{y}}} = \hat{\alpha} + \hat{\beta}\overline{x} \tag{66}$$

$$= 5226.5 + 1168.5 \cdot 3.5 = \underline{\underline{9316.3}} \tag{67}$$

som er gjennomsnittlig antall utrykninger per år når man tar gjennomsnittet over perioden $2011 - 2016$.

■

Kapittel 12

Kontinuasjonseksamen 2017

LØSNING: Eksamen 3. jan. 2018

"MAT110 Statistikk 1"

Oppgave 1: (økonomi , aksjer)

a) Siden den stokastiske variabelen X har et tellbart antall mulige verdier
så er X diskret.

b) Enhver gyldig sannsynlighetsfordeling er normalisert til 1, dvs. $\sum_i P(X = x_i) = 1$.
For at denne normaliseringsbetingelsen skal være oppfylt så må:
$P(X = 220) = 0.60$.

c) Sannsynligheten $P(X \leq 200)$ finner vi ved avlesning fra tabellen i opppgaven:

$$\underline{\underline{P(X \leq 200)}} = \sum_{X \leq 200} P(X = x_i) \tag{1}$$

$$= P(X < 180) + P(X = 180) + P(X = 200) \tag{2}$$

$$= 0 + 0.05 + 0.05 = \underline{\underline{0.10}} \tag{3}$$

Sannsynligheten for at prisen på aksjene er mindre enn eller lik 200 NOK om ett år er:
$P(X \leq 200) = 0.10$

d) Forventning $E[X]$:

$$\underline{E[X]} \stackrel{\text{def.}}{=} \sum_i x_i P(X = x_i) \tag{4}$$

$$\begin{aligned} &= 180\overbrace{P(X=180)}^{=0.05} + 200\overbrace{P(X=200)}^{=0.05} + 220\overbrace{P(X=220)}^{=0.60} \\ &+ 240\underbrace{P(X=240)}_{=0.20} + 260\underbrace{P(X=260)}_{=0.10} \end{aligned} \tag{5}$$

$$\begin{aligned} &= 180 \cdot 0.05 \quad + \quad 200 \cdot 0.05 \quad + \quad 220 \cdot 0.60 \\ &+ 240 \cdot 0.20 \quad + \quad 260 \cdot 0.10 \end{aligned} \tag{6}$$

$$= \underline{225} \tag{7}$$

Forventet aksjekurs om ett år er $\underline{\underline{E[X] \, = \, 225 \text{ NOK}}}$

e) Varians $Var[X]$:

$$\underline{Var[X]} \overset{\text{def.}}{=} \sum_i \big(x_i - E[X]\big)^2 \cdot P(X = x_i) \tag{8}$$

$$\begin{aligned}
&= (180-225)^2 \overbrace{P(X=180)}^{=0.05} + (200-225)^2 \overbrace{P(X=200)}^{=0.05} \\
&+ (220-225)^2 \overbrace{P(X=220)}^{=0.60} + (240-225)^2 \underbrace{P(X=240)}_{=0.20} \\
&+ (240-225)^2 \underbrace{P(X=240)}_{=0.10}
\end{aligned} \tag{9}$$

$$\begin{aligned}
&= (180-225)^2 \cdot 0.05 + (200-225)^2 \cdot 0.05 + (220-225)^2 \cdot 0.60 \\
&+ (240-225)^2 \cdot 0.20 + (260-225)^2 \cdot 0.10
\end{aligned} \tag{10}$$

$$= \underline{315} \tag{11}$$

Forventet varians om ett år er $\underline{\underline{Var[X] = 315 \text{ NOK}^2}}$

$\blacksquare$

Oppgave 2: (økonomi , aksjer)

a) Fortjenesten F er:

$$\boxed{F \;=\; \text{antall aksjer} \cdot \text{pris per aksje om ett år} \;\; - \;\; \text{antall aksjer} \cdot \text{pris per aksje i dag} \qquad (12)}$$

for de tre selskapene så får vi:

$$\begin{aligned}
\underline{\underline{F}} \;&=\; \underbrace{\text{antall aksjer i } A}_{=\,N\cdot a} \cdot \underbrace{\text{pris per aksje for selskap } A \text{ om ett år}}_{=\,X} \\
&+\; \underbrace{\text{antall aksjer i } B}_{=\,N\cdot b} \cdot \underbrace{\text{pris per aksje for selskap } B \text{ om ett år}}_{=\,Y} \\
&+\; \underbrace{\text{antall aksjer i } C}_{=\,N\cdot c} \cdot \underbrace{\text{pris per aksje for selskap } C \text{ om ett år}}_{=\,Z} \qquad (13)
\end{aligned}$$

$$\begin{aligned}
&-\; \underbrace{\text{antall aksjer i } A}_{=\,N\cdot a} \cdot \underbrace{\text{pris per aksje for selskap } A \text{ i dag}}_{=\,230} \qquad (14) \\
&-\; \underbrace{\text{antall aksjer i } B}_{=\,N\cdot b} \cdot \underbrace{\text{pris per aksje for selskap } B \text{ i dag}}_{=\,205} \\
&-\; \underbrace{\text{antall aksjer i } C}_{=\,N\cdot c} \cdot \underbrace{\text{pris per aksje for selskap } C \text{ i dag}}_{=\,235} \qquad (15)
\end{aligned}$$

$$= \; NaX \;+\; NbY \;+\; NcZ \;\;-\;\; Na230 \;-\; Nb205 \;-\; Nc235 \qquad (16)$$

$$= \; \underline{\underline{Na\big(X - 230\big) \;+\; Nb\big(Y - 205\big) \;+\; Nc\big(Z - 235\big)}} \qquad , \qquad \text{q.e.d.} \qquad (17)$$

b) Den forventede fortjenesten $E[F]$ ved et eventuelt salg av aksjene om ett år er: [1]

$$\underline{\underline{E[F]}} = E\Big[\; Na(X-230) \;+\; Nb(Y-205) \;+\; Nc(Z-235)\Big] \tag{18}$$

$$= Na\Big(\underbrace{E[X]}_{=225}-230\Big) \;+\; Nb\Big(\underbrace{E[Y]}_{=215}-205\Big) \;+\; Nc\Big(\underbrace{E[Z]}_{=250}-235\Big) \tag{19}$$

$$= Na\big(\underbrace{225-230}_{=-5}\big) \;+\; Nb\big(\underbrace{215-205}_{=\,10}\big) \;+\; Nc\big(\underbrace{250-235}_{=\,15}\big) \tag{20}$$

$$= Na(-5) \;+\; Nb\,10 \;+\; Nc\,15 \tag{21}$$

$$= \underline{\underline{N\Big(-5a \;+\; 10b \;+\; 15c\Big)}} \qquad , \qquad \text{q.e.d.} \tag{22}$$

c) At leddet $-5a$ er negativt i $E[F]$ betyr at man forventer
å <u>tape</u> på å investere i selskap A dersom man selger aksjene etter ett år.

d) Fra lign.(22) i oppgave **2b** ser vi umiddelbart at $E[F]$ blir størst når vi
<u>kun investerer i selskap C</u>, dvs.:

$$a = 0 \qquad , \qquad b = 0 \qquad , \qquad c = 1 \tag{23}$$

Verdien på den forventede fortjenesten $E[F]$ blir da:

$$\underline{\underline{E[F]}} = N\,15c \tag{24}$$

$$= 50\,000 \cdot 15 \cdot 1 \;\; \text{NOK} \;=\; \underline{\underline{750\,000 \;\; \text{NOK}}} \tag{25}$$

[1] I lign.(19) har vi benyttet oss av at forventingen til en kostant, er konstant: $E[\,Na\,230\,] = Na\,230$.

e) Fra formelsamlingen vet vi formelen for variansen av en lineær kombinasjon av to stokastiske variabler X og Y, nemlig:

$$\overbrace{Var[aX+bY]}^{\text{variasjon/(spredning)}} \;=\; a^2\,Var[X] \;+\; b^2\,Var[Y] \;+\; 2ab\cdot\overbrace{Cov[X,Y]}^{\underline{\text{sam}}\text{variasjon}} \tag{26}$$

Dersom X og Y er <u>u</u>avhengige så er $Cov[X,Y]=0$.

For vårt tilfelle er de stokastiske variablene <u>u</u>avhengige. Da er også $Cov[X,Y,Z]=0$. Variansen $Var[F]$ til fortjenesten ved et eventuelt salg om ett år blir da: [2]

$$\underline{\underline{Var[F]}} \;=\; Var\Big[\,Na\big(X-230\big) \;+\; Nb\big(Y-205\big) \;+\; Nc\big(Z-235\big)\,\Big] \tag{27}$$

$$= \big(Na\big)^2 Var[X] \;+\; \big(Nb\big)^2 \underbrace{Var[Y]}_{=\;1.5\,Var[X]} \;+\; \big(Nc\big)^2 \underbrace{Var[Z]}_{=\;3\,Var[X]}$$

$$- \;\underbrace{Var[Na\,230]}_{=\;0} \;+\; \underbrace{Var[Nb\,205]}_{=\;0} \;+\; \underbrace{Var[Nc\,235]}_{=\;0} \tag{28}$$

$$= \underbrace{Var[X]}_{=\;\sigma^2}\,N^2\Big(a^2 \;+\; 1.5b^2 \;+\; 3c^2\Big) \tag{29}$$

$$= \underline{\underline{\sigma^2\,N^2\Big(a^2 \;+\; 1.5b^2 \;+\; 3c^2\Big)}} \qquad , \qquad \text{q.e.d.} \tag{30}$$

[2] I lign.(28) har vi benyttet oss av at variansen av en konstant er null: $Var[\,Na\,230\,]=0$.

f) Dersom man ønsker minst mulig risiko i investeringen så tilsvarer det at variansen $Var[F]$ er minst mulig. I oppgaven er det oppgitt at denne variansen er minst når $a = 0.5$, $b = 0.33$ og $c = 0.17$.

Antall aksjer som må kjøpes i de forskjellige selskapene for å oppnå dette er da:

$$\underline{\underline{Na}} = 50\,000 \cdot 0.5 = \underline{\underline{25\,000}} \tag{31}$$

$$\underline{\underline{Nb}} = 50\,000 \cdot 0.33 = \underline{\underline{16\,500}} \tag{32}$$

$$\underline{\underline{Nc}} = 50\,000 \cdot 0.17 = \underline{\underline{8\,500}} \tag{33}$$

g) Fortjenesten F om ett år er:

$$F = R\,N\left(a + \sqrt{1.5}b + \sqrt{3}c\right) + k \tag{34}$$

hvor k er en konstant.

Variansen til fortjenesten F er da: [3] (se lign.(26))

$$\underline{\underline{Var[F]}} = Var\Big[R\,N\Big(a + \sqrt{1.5}b + \sqrt{3}c\Big) + k\Big] \tag{35}$$

$$= \underbrace{Var[R]}_{=\sigma^2}\left(N\left(a + \sqrt{1.5}b + \sqrt{3}c\right)\right)^2 + \underbrace{Var[k]}_{=0} \tag{36}$$

$$= \underline{\underline{\sigma^2 N^2\Big(a + \sqrt{1.5}b + \sqrt{3}c\Big)^2}} \quad , \qquad \text{q.e.d.} \tag{37}$$

[3]Husk at variansen av en konstant er null: $Var[k] = 0$.

h) I modell 2 varierer aksjeprisene X, Y og Z "*i takt*" som oppgitt i fotnoten i oppgaveteksten. Derfor er det ingenting å hente på å spre aksjene på de tre selskapene.

Det som bestemmer hvilket selskap vi bør investere i da er det selskapet som har minst følsomhet for rentenivået via den usikre faktoren R, dvs. "*renteindikatoren*".

Fra lign.(37) her i løsningforslaget eller en av de oppgitte ligningene i oppgaveteksten ser vi at selskap A er minst følsom for endringer i rentenivået.

I modell 2 er derfor $Var[F]$ blir minst når vi investerer alt i selskap A. [4]

■

[4]Kommentar:
R kan oppfattes som en **renteindikator** som sier noe om når marketdet går bra eller ikke. Markedet går bra når denne indikatoren er positiv, og tilsvarende dårlig når R er negativ. (Kommentaren i denne fotnoten behøver man ikke ha med på eksamensbesvarelsen.)

Oppgave 3: (logistikk)

a) **i**) Tolkning:

$$\underline{\underline{E[Y]}} \quad = \quad \text{forventet antall dager forsinkelser per år ved leveranse av} \qquad$$
$$\underline{\underline{n \text{ antall båter med alumina i året}}} \tag{38}$$

ii) Forventet antall dager forsinkelser i året:

$$\underline{\underline{E[Y]}} = E\big[X_1 + X_2 + \ \dots \ + X_n\big] \tag{39}$$

$$\overset{\text{alltid}}{=} E[X_1] + E[X_2] + \ \dots \ + \ E[X_n] \tag{40}$$

$$= \underbrace{E[X] + E[X] + \ \dots \ + E[X]}_{=\ n} \quad = \quad n \underbrace{E[X_i]}_{=0} \quad = \quad \underline{\underline{0}} \tag{41}$$

NB: 1) Overgangen i lign.(39) til (40) gjelder alltid.
Uansett om de stokastiske variablene X_i er uavhengige eller ikke.

2) Legg merke til at vi ikke trenger å benytte oss av den numeriske verdien $n = 40$.
Utregningene i denne deloppgaven gjelder for alle n.

b) i) Tolkning:

$$\underline{\underline{Var[Y]}} = \text{forventet variasjon/spredning i antall dager med forsinkelser}$$
$$\underline{\underline{\text{etter } n \text{ turer i året mellom Brasil og Norge}}} \qquad (42)$$

ii) Variansen til antall dager med forsinkelser: ($n = 40$)

$$\underline{\underline{Var[Y]}} = Var\left[\, X_1 + X_2 + \ \dots\ + X_n \,\right] \qquad (43)$$

$$\overset{\text{uavh.}}{=} \underbrace{Var[X_1] + Var[X_2] + \ \dots\ + Var[X_n]}_{n \cdot Var[X_i]} \qquad (44)$$

$$= n \underbrace{Var[X_i]}_{= \left(\frac{aL}{M}\right)^2} = \underline{\underline{n\left(\frac{aL}{M}\right)^2}} \quad , \quad \text{q.e.d.} \qquad (45)$$

NB: Overgangen i lign.(43) til (44) gjelder kun dersom de stokastiske variablene X_i er uavhengige.

c) Fra deloppgavene foran og opplysingene i oppgaveteksten ser vi at:

$$\underbrace{Var[Y]}_{= n\left(\frac{aL}{M}\right)^2} > \underbrace{Var[X_i]}_{= \left(\frac{aL}{M}\right)^2} \qquad (46)$$

Hver leveranse har en usikkerhet/varians $Var[X_i]$.
Når man summerer n slike varianser så får man n ganger større varians siden alle leveransene X_i'ene er uavhengige. [5]

[5]Sagt på en mer folkelig måte: Dersom man summerer mange usikkerheter så får man en større usikkerhet.

d) Siden

1. de forskjellige turene er $\underline{\text{u}}$avhengige: (oppgitt i oppgaveteksten)

   $X_i \;\sim$ er $\underline{\text{u}}$avhengige for alle $i = 1, 2, 3, ..., n$

2. alle X_i har samme sannsynlighetsfordeling: (oppgitt i oppgaveteksten)

   $X_i \;\sim$ samme sannsynlighetsfordeling for alle $i = 1, 2, 3, ..., n$

3. antall "forsøk", dvs. antall overtrekk, $n = 40$ er tilstrekkelig stort [6]

så gjelder **sentralgrensesetningen**. Dermed er Y tilnærmet normalfordelt.

e) Sannsynlighet for at det er mindre enn 10 dagers forsinkelse i løpet av et år med $n = 40$ turer: [7]

$$\underline{P(Y < 10)} \stackrel{\text{standardiser}}{=} P\Bigg(\underbrace{\frac{Y - E[Y]}{\sigma[T]}}_{=Z} \le \frac{10 - E[Y]}{\sigma[Y]}\Bigg) \tag{47}$$

$$= P\left(Z \le \frac{10 - 0}{\sqrt{n\left(\frac{aL}{M}\right)^2}}\right) \tag{48}$$

$$= P\Bigg(Z \le \underbrace{\frac{10}{\sqrt{n}\,\frac{aL}{M}}}_{=\,Z_0}\Bigg) = 0.95 \tag{49}$$

som gir, med $Z_0 = \frac{10}{\sqrt{n}\,\frac{aL}{M}}$,

$$P\big(Z \le Z_0\big) = 0.95 \tag{50}$$

[6] Husk: Antall forsøk n for at sentralgrensesetningen skal gjelde er avhengig av situasjonen. Men en **tommelfingerregel** er at vi bør ha $n \gtrsim 30$.

[7] Siden normalfordelingen er kontinuerlig så spiller det ingen rolle om vi har $<$ eller $\le$ i lign.(47).

Ved “omvendt tabelloppslag” ser vi at 0.9495 og 0.9505 ligger midt mellom 0.95. Dette tilsvarer at *argumentet* er 1.645:

$$Z_0 = 1.645 \tag{51}$$

Dermed:

$$Z_0 = \frac{10\,M}{\sqrt{n}\,\frac{aL}{\cancel{M}}\,\cancel{M}} \tag{52}$$

$$Z_0 = \frac{10M}{\sqrt{n}\,aL} \tag{53}$$

og løser med hensyn på L alene:

$$\underline{M} = \frac{Z_0\,\sqrt{n}\,aL}{10} \tag{54}$$

$$= \frac{1.645 \cdot \sqrt{40} \cdot 0.16 \cdot 18}{10} = \underline{3} \tag{55}$$

Det må settes inn $\underline{\underline{M = 3 \text{ båter}}}$ mellom Brasil og Norge for å oppnå den ønskede sikkerheten i leveringspresisjonen.

■

Oppgave 4: (logistikk)

a) Gjennomsnitt av x_i:

$$\underline{\underline{\overline{x}}} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{56}$$

$$= \frac{1}{5}\Big(2013 + 2014 + 2015 + 2016 + 2017\Big) = \underline{\underline{2015}} \tag{57}$$

Gjennomsnitt av y_i:

$$\underline{\underline{\overline{y}}} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{58}$$

$$= \frac{1}{5}\Big(762 + 788 + 794 + 825 + 834\Big) = \underline{\underline{800.6}} \tag{59}$$

b) Uttrykket for korrelasjonskoeffisienten finner man i formelsamlingen. Det gir:

$$\underline{\underline{R_{xy}}} = \frac{S_{xy}}{S_x \cdot S_y} \tag{60}$$

$$= \frac{45.25}{\sqrt{2.5}\cdot\sqrt{850.8}} = \underline{\underline{0.98}} \tag{61}$$

c) At $R_{xy} = 0.98$ betyr at det er en veldig <u>sterk positiv korrelasjon</u> mellom x og y.

Det er altså en svært <u>sterk lineær sammenheng</u> mellom x og y med positivt stigningstall.

d) Vi skal finne regresjonslinjen. Formelen for minste kvadraters lineære regresjonslinje står i formelsamlingen:

$$\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta}x \tag{62}$$

hvor parametrene $\hat{\beta}$ og $\hat{\alpha}$ er: (dropper enheten/benevningen)

$$\underline{\hat{\beta}} \;=\; \frac{S_{xy}}{S_x^2} \;=\; \frac{45.25}{2.5} \;=\; \underline{18.1} \tag{63}$$

$$\underline{\hat{\alpha}} \;=\; \overline{y} \;-\; \beta\overline{x} \;=\; 800.6 - 18.1 \cdot 2015 \;=\; \underline{-35\,670.9} \tag{64}$$

Minste kvadraters lineære regresjonslinje $\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta}\,x$ blir dermed:

$$\underline{\underline{\hat{y} \;=\; -\,35\,670.9 \;+\; 18.1x}} \qquad , \qquad \text{q.e.d.} \tag{65}$$

e) Løser mhp. x alene:

$$\hat{y} \;\overset{\text{Eq.(62)}}{=}\; \hat{\alpha} \;+\; \hat{\beta}x \tag{66}$$

$$\hat{y} \;-\; \hat{\alpha} \;=\; \hat{\beta}x \tag{67}$$

$$\underline{x \;=\; \frac{\hat{y} - \alpha}{\beta}} \tag{68}$$

Med $\hat{y} = 900$, $\hat{\alpha} = -35\,670.9$ og $\hat{\beta} = 18.1$ så får vi:

$$\underline{x} \;=\; \frac{900 - (-35\,670.9)}{18.1} \;=\; \underline{2020.5} \tag{69}$$

Ifølge regresjonslinjen vil kapasiteten bli sprengt $\underline{\underline{\text{midten av år}\;\; 2020}}$.

■

Kapittel 13

Hovedeksamen 2018

LØSNING: Eksamen 22. mai 2018

"*MAT110 Statistikk 1*", vår 2018

Oppgave 1: (logistikk)

a) Sannsynlighetene p_i, med $i = 1, 2, 3, ..., 8$ utgjør en gyldig sannsynlighetsfordeling fordi:

$$\underline{\underline{\sum_{i=1}^{8} p_i}} = \frac{5}{350} + \frac{40}{350} + \frac{25}{350} + \frac{80}{350} \tag{1}$$

$$+ \frac{8}{350} + \frac{68}{350} + \frac{22}{350} + \frac{102}{350} \tag{2}$$

$$= \frac{150 \ + \ 200}{350} = \frac{350}{350} = \underline{\underline{1}} \tag{3}$$

b) Fra figuren i oppgaven ser vi at de to begivenhetene S_1 og S_2 ikke overlapper.[1]
Derfor er de disjunkte.

[1] I likhet med brikkene i et puslespill så overlapper S_1 og S_2 akkurat ikke.

c) Siden sannsynlighetene $p_1 = \frac{5}{350}$, $p_2 = \frac{40}{350}$ osv. er alle $\underline{\underline{\text{ulike}}}$,
altså $\underline{\underline{\text{ikke uniformt utfallsrom}}}$, så gjelder ikke urnemodellen for vår sannsynlighetsmodell. [2]

d) Fra første ligningen i oppgaveteksten finner vi:

$$P(S_2 \cap H \cap A) \;=\; p_5 \;=\; \frac{8}{350} \tag{4}$$

og

$$P(S_2 \cap H) \;=\; p_5 + p_7 \;=\; \frac{8}{350} + \frac{22}{350} \;=\; \frac{30}{350} \tag{5}$$

Dermed har vi det vi trenger for å regne ut den betingede sannsynligheten:

$$\underline{\underline{P(A|S_2 \cap H)}} \;=\; \frac{P(S_2 \cap H \cap A)}{P(S_2 \cap H)} \;=\; \frac{\frac{8}{\cancel{350}}}{\frac{30}{\cancel{350}}} \;=\; \underline{\underline{\frac{8}{30}}} \qquad (= 0.2667) \tag{6}$$

e) Tolkning: [3]

$\underline{\underline{P(A|S_2 \cap H)}}$ = sannsynligheten for at et tilbud blir aksptert dersom det er laget av selger 2 og verdien av tilbudet er høy

■

[2]Fra figuren i oppgaven ser vi også at kulene har forskjellige størrelse. Det visualiserer at sannsynlighetene p_i er forskjellige.

[3]Det er flere måter å formulere seg på her. Løsningsforslaget er bare én mulig formulering.

Oppgave 2: (logistikk - prognostisering)

a) Y beskriver en forsøksserie som har følgende egenskaper:

1. Hvert tilbud har kun 2 mulige utfall, akseptert (suksess) eller ikke akseptert (fiasko).
2. Alle tilbudene har samme sannsynlighet p for å bli akseptert, hvor $p = 0.2667$ i vårt tilfelle.
3. I oppgaven antas det at tilbudene er uavhengig av hverandre.
4. Det gjennomføres et bestemt antall forsøk, $n = 4$ i vårt tilfelle.

Forsøksserien oppfyller dermed kravene til en **binomisk** forsøksserie.
Den stokastiske variablene Y er da binomisk fordelt: $Y \sim \text{Bin}\big[\, n = 4, p = 0.2667 \,\big]$.

b) Siden Y binomsik fordelt, $Y \sim \text{Bin}\big[\, n = 4, p = 0.2667 \,\big]$,
så finner vi formelen for sannsynlighet i formelsamlingen:

$$\underline{\underline{P(Y \geq 2)}} = 1 - P(Y = 0) - P(Y = 1) \tag{7}$$

$$= 1 - \binom{n}{0} p^0 \, (1-p)^{n-0} - \binom{n}{1} p^1 \, (1-p)^{n-1} \tag{8}$$

$$= 1 - \binom{4}{0} 0.2667^0 \, (1 - 0.2667)^{4-0} - \binom{4}{1} 0.2667^1 \, (1 - 0.2667)^{4-1} \tag{9}$$

$$= 1 - 0.2892 - 0.4207 \tag{10}$$

$$= \underline{\underline{0.2901}} \tag{11}$$

c) Siden Y er binomisk fordelt, $Y \sim \text{Bin}\left[\, n = 4, p = 0.2667 \,\right]$,
så finner vi formelen for forventing

$$\underline{\underline{E[Y]}} \;=\; n \cdot p \;=\; 4 \cdot 0.2667 \;=\; \underline{\underline{1.0668}} \tag{12}$$

Tolkning:

$$\underline{\underline{E[Y] \;=\; \text{forventet antall tilbud som blir aksepert av de 4 tilbudene}}}$$

d) Bruker regnereglene for forventingen i formelsamlingen og
finner forventningen av V:

$$E[V] \;=\; E\left[\, v_1 Y_1 \;+\; v_2 Y_2 \;+\; v_3 Y_3 \;+\; v_4 Y_4 \,\right] \tag{13}$$

$$\;=\; v_1 E[Y_1] \;+\; v_2 E[Y_2] \;+\; v_3 E[Y_3] \;+\; v_4 E[Y_4] \tag{14}$$

hvor $E[Y_1] = E[Y_2] = E[Y_3] = E[Y_4] = p$ (oppgitt i en fotnote i oppgaven). Dermed:

$$\underline{\underline{E[V]}} \;=\; v_1 p + v_2 p + v_3 p + v_4 p \tag{15}$$

$$\;=\; \underline{\underline{\big(v_1 + v_2 + v_3 + v_3 \big)\, p}} \quad , \qquad \text{q.e.d.} \tag{16}$$

hvor p er en felles faktor som vi faktoriserer utenfor.

e) Bruker lign.(16) og setter inn tallene for v_i og p som oppgitt i oppgaven:

$$\underline{\underline{E[V]}} = \big(v_1 + v_2 + v_3 + v_3\big)\,p \tag{17}$$

$$= \Big(40\,000 + 33\,000 + 105\,000 + 75\,000\Big)\,0.2667 \quad \text{NOK} \;=\; \underline{\underline{67\,475 \quad \text{NOK}}} \tag{18}$$

f) Bruker regnereglene for varians i formelsamlingen og finner variansen av V:

$$Var[V] \quad = \quad Var\big[\,v_1Y_1 + v_2Y_2 + v_3Y_3 + v_4Y_4\,\big] \tag{19}$$

$$\overset{\text{uavh.}}{=} \quad v_1^2 Var[Y_1] \;+\; v_2^2 Var[Y_2] \;+\; v_3^2 Var[Y_3] \;+\; v_4^2 Var[Y_4] \tag{20}$$

siden covariansen er null da alle Y_i'ene er uavhengige.
I en fotnote i oppgaven er det oppgitt at: $(\;Y_i \;\sim\; \text{Bin}[\,n_i = 1, p\,]\;)$

$$Var[Y_1] \;=\; Var[Y_2] \;=\; Var[Y_3] \;=\; Var[Y_4] \;=\; p(1-p) \tag{21}$$

Variansen blir da:

$$\underline{\underline{Var[V]}} \;=\; v_1^2 p(1-p) \;+\; v_2^2 p(1-p) \;+\; v_3^2 p(1-p) \;+\; v_4^2 p(1-p) \tag{22}$$

$$= \;\underline{\underline{\big(v_1^2 \;+\; v_2^2 \;+\; v_3^2 \;+\; v_4^2\big)\,p(1-p)}} \qquad , \qquad \text{q.e.d.} \tag{23}$$

hvor $p(1-p)$ er en felles faktor som vi faktoriserer utenfor.

g) Siden alle verdiene v_i er forskjellige så er også forventingsverdiene til leddene i lineærkombinasjonen forskjellige.

Dermed er en av forutsetningene til sentralgrenseteoremet brutt.

For at sentralgrenseteoremet skal gjelde må forventingensverdiene til leddene i lineær kombinasjonen være like.

Dermed er ikke summen $V = v_1Y_1 + v_2Y_2 + ... + v_{200}Y_{200}$ normalfordelt selv om tilbudene, dvs. Y_i'ene, er uavhengige og aksept-sannsynlighetene p er like.

■

Oppgave 3: (logistikk og økonomi)

a) Forventet etterspørsel av antall kurver med jordbær per dag på dagtid:

$$\underline{\underline{E[D_{\text{tot}}]}} = E\big[\, D_1 + D_2 + D_3 + D_4 \,\big] \tag{24}$$

$$= E[D_1] + E[D_2] + E[D_3] + E[D_4] \tag{25}$$

$$= 20 + 50 + 30 + 10 \tag{26}$$

$$= \underline{\underline{110}} \qquad , \qquad \text{q.e.d.} \tag{27}$$

Merk:
Overgangen i lign.(25) gjelder alltid, uansett om D_i'ene er uavhengige eller ikke.

b) Siden de stokastiske variablene D_i er uavhengige, så er tilhørende kovarians lik null.
Variansen blir dermed: (se formelsamlingen)

$$\underline{\underline{Var[D_{\text{tot}}]}} = Var\big[\, D_1 + D_2 + D_3 + D_4 \,\big] \tag{28}$$

$$\overset{\text{uavh.}}{=} Var[D_1] + Var[D_2] + Var[D_3] + Var[D_4] \tag{29}$$

$$= 16 + 36 + 88 + 4 \tag{30}$$

$$= \underline{\underline{144}} \tag{31}$$

c) Det er oppgitt i oppgaven at alle D_i er normalfordelte:

$$D_i \;\sim\; N\big[\; E[D_i]\,,\, Var[D_i] \;\big] \tag{32}$$

I tillegg er det oppgitt i oppgaven at alle D_i'ene er uavhengige. [4]
Fra læresetningen i formelsamlingen [5] vet vi da at summen
$D_{\text{tot}} = D_1 + D_2 + D_3 + D_4$ også er normalfordelt, dvs.:

$$\underline{\underline{D_{\text{tot}} \;\sim\; N\big[\; E[D_{\text{tot}}]\,,\, Var[D_{\text{tot}}] \;\big]7}} \tag{33}$$

hvor $E[D_{\text{tot}}] = 110$ og $Var[D_{\text{tot}}] = 144$ fra oppgave **3a** og **3b**, henholdsvis.

[4] At D_i'ene er uavhengige er det samme som å si at etterspørselen av jordbær i de 4 butikkene er uavhengige.
[5] Se side 108 i formelsamlingen fra 2018.

d) Med innkjøpspris (kostnad for Bunnpris) $c = 30$ NOK og utslagspris $p = 40$ NOK fås:

$$P(D_{\text{tot}} \leq q^*) = 1 - \frac{c}{p} \tag{34}$$

$$P(D_{\text{tot}} \leq q^*) = 1 - \frac{30}{40} \tag{35}$$

$$P(D_{\text{tot}} \leq q^*) = 0.25 \tag{36}$$

$$P\left(\underbrace{\frac{D_{\text{tot}} - E[D_{\text{tot}}]}{\sigma[D_{\text{tot}}]}}_{\equiv Z} \leq \underbrace{\frac{q^* - E[D_{\text{tot}}]}{\sigma[D_{\text{tot}}]}}_{\equiv z_0}\right) \overset{\text{standardiser}}{=} 0.25 \tag{37}$$

$$P(Z \leq z_0) = 0.25 \tag{38}$$

z_0 finnes ved "omvendt tabelloppslag".
Men sannsynligheten $P(Z \leq z_0) = 0.25$ finner ikke i tabellen i formelsamlingen.
Sannsynlighetene i formelsamlingen "stopper" ved 0.5000.
Men av symmetrigrunner så finner man z_0 er negativ og gitt ved:

$$z_0 = -0.675 \tag{39}$$

Vi løser:

$$z_0 = \frac{q^* - E[D_{\text{total}}]}{\sigma[D_{\text{tot}}]} \tag{40}$$

med hensyn på q^*:

$$\underline{q^*} = E[D_{\text{tot}}] + z_0 \cdot \sigma[D_{\text{tot}}] \tag{41}$$

$$= 110 - 0.675 \cdot \sqrt{144} = \underline{101.9} \tag{42}$$

Bunnpris må bestille $\underline{\underline{q^* \approx 102}}$ kurver med jordbær per gang for å få størst mulig forventet fortjeneste.

e) For å finne p_{red} må vi først regne ut sannsynligheten.
Deretter kan vi regne ut p_{red} via den oppgitte ligningen i oppgaveteksten.

Fra oppgave **3c** vet vi at D_{tot} er normalfordelt.
Fra oppgave **3a** og **3b** vet vi at $E[D_{\text{tot}}] = 110$ og $\sigma[D_{\text{tot}}] = \sqrt{Var[D_{\text{tot}}]} = \sqrt{144} = 12$.
Dermed kan vi regne ut sannsynligheten $P(D_{\text{tot}} \leq q^*)$ via standardisering og tabelloppslag:

$$\underline{P(D_{\text{tot}} \leq q^*)} = P\Bigg(\underbrace{\frac{D_{\text{tot}} - E[D_{\text{tot}}]}{\sigma[D_{\text{tot}}]}}_{=Z} \leq \frac{q^* - E[D_{\text{tot}}]}{\sigma[D_{\text{tot}}]}\Bigg) \tag{43}$$

$$= P\left(Z \leq \frac{116 - 110}{12}\right) \tag{44}$$

$$= P(Z \leq 0.5) \tag{45}$$

$$\stackrel{\text{tabell}}{=} \underline{0.6915} \tag{46}$$

Ligningen oppgitt i oppgaveteksten er:

$$P(D_{\text{tot}} \leq q^*) = \frac{p - c}{p - p_{\text{red}}} \tag{47}$$

$$p - p_{\text{red}} = \frac{p - c}{P(D_{\text{tot}} \leq q^*)} \tag{48}$$

som gir:

$$\underline{p_{\text{red}}} = p - \frac{p - c}{P(D_{\text{tot}} \leq q^*)} \tag{49}$$

$$= \left(40 - \frac{40 - 30}{0.6915}\right) \text{NOK} = \underline{25.5 \text{ NOK}} \tag{50}$$

Bunnpris må sette den reduserte prisen til $\underline{\underline{p_{\text{red}} = 25.5 \text{ NOK}}}$
for å maksimere den forventede fortjetensten.

■

Oppgave 4: (økonomi)

a) **i**) R_{xy} er normalisert og ligger mellom -1 og 1, dvs.: $\underline{\underline{-1 \leq R_{xy} \leq 1}}$.

ii) R_{xy} er et mål på lineær korrelasjon mellom obervasjonene x og y.
Det er et mål på i hvor stor grad det er en lineær sammmenheng mellom obervasjonene x og y.

iii) R_{xy} er enhetsuavhengig, dvs. ingen enhet. [6]

b) Uttrykket for korrelasjonskoeffisienten finner vi i formelsamlingen. Det gir:

$$\underline{\underline{R_{xy}}} = \frac{S_{xy}}{S_x \cdot S_y} \tag{51}$$

$$= \frac{2\,084\,821}{\sqrt{37\,364} \cdot \sqrt{133\,388\,393}} = \underline{\underline{0.93}} \tag{52}$$

c) At $R_{xy} = 0.93$ betyr at det er en sterk positiv korrelasjon mellom x og y, dvs. store x hører sammen med store y og omvendt.
Det er altså en sterk lineær sammenheng mellom x og y med positivt stigningstall.

[6] Dette betyr at R_{xy} har samme verdi uansett hva slags enhet man bruker for å regne ut $R_{xy} = S_{xy}/(S_x \cdot S_y)$.

d) Vi skal finne regresjonslinjen. Formelen for minste kvadraters lineære regresjonslinje står i formelsamlingen:

$$\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta}x \tag{53}$$

hvor parametrene $\hat{\beta}$ og $\hat{\alpha}$ er: (dropper benevningen)

$$\underline{\hat{\beta}} \;=\; \frac{S_{xy}}{S_x^2} \;=\; \frac{2\,084\,821}{37\,364} \;=\; \underline{55.8} \tag{54a}$$

$$\underline{\hat{\alpha}} \;=\; \overline{y} \;-\; \beta\overline{x} \;=\; 43\,437.5 - 55.8 \cdot 607.5 \;=\; \underline{9\,539} \tag{54b}$$

Minste kvadraters lineære regresjonslinje $\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta}\,x$ blir dermed:

$$\underline{\underline{\hat{y} \;=\; 9\,539 \;+\; 55.8\,x}} \tag{55}$$

e) At en person tjener 1 million NOK tilsvarer $x = 1000$ i antall 1000 NOK:
Bruker regresjonslinjen $\hat{y} = \hat{y}(x)$ fra oppgave **4d**, dvs. lign.(55):

$$\underline{\hat{y}(1000)} \;=\; \Big(9\,539 \;+\; 55.8 \cdot 1000\Big) \;\text{NOK/m}^2 = \;\underline{65\,338 \;\;\text{NOK/m}^2} \tag{56}$$

Ifølge regresjonslinjen så kommer kjøperent til å en leilighet med en kvadratmeterpris på utenfor sentrum er $\underline{\underline{65\,338 \text{ NOK/m}^2}}$.

f) Gjennomsnittlig kvadratmeterpris i Norge er 30 000 NOK/m^2.
Ifølge regresjonslinjen, når $\hat{y}(x) = 30\,000$, er da:

$$30\,000 \quad = \quad 9\,539 \; + \; 55.8\, x \tag{57}$$

Løser med hensyn på x:

$$\underline{x} \; = \; \frac{30\,000 \; - \; 9\,539}{55.8} \; = \; \underline{367} \tag{58}$$

Ifølge regresjonslinjen har kjøperen av leiligheten en årsinntekt på $\underline{\underline{367\,000 \text{ NOK}}}$.

g) Forklaringskraften R^2 kan leses direkte fra Excel-utskriften:
(Se cellen som heter "*R Square*" i Excel-utskriften): [7]

$$\underline{\underline{R^2 \; = \; 0.8721}} \tag{59}$$

h) Kommentar til svaret i oppgave **4g**:

At $R^2 = 0.8721$ betyr at regresjonslinjen vil i "i stor grad" forutsi kvadrameterprisen på en leilighet for en person med en gitt årslønn.
Modellen har $\underline{\underline{\text{stor forklaringskraft}}}$, tilsvarende $R^2 = 0.8721$.

■

[7] Man kan også regne ut forklaringskraften R^2 "for hånd" via definisjonen $R^2 = 1 - SSE/SST$. Men det er mye mer arbeidskrevende. Fint at dataprogrammer (som f.eks. Excel) kan hjelpe oss med slikt.

Kapittel 14

Kontinuasjonseksamen 2018

LØSNING: Eksamen 10. sept. 2018

"*MAT110 Statistikk 1*", vår 2018

Oppgave 1: (logistikk)

a) Sannsynlighetene p_i, med $i = 1, 2, 3, ..., 8$ utgjør en gyldig sannsynlighetsfordeling dersom:

$$\underline{\underline{\sum_{i=1}^{8} p_i}} = \frac{5}{350} + \frac{40}{350} + \frac{25}{350} + \frac{80}{350} + \frac{8}{350} + \frac{68}{350} + \frac{22}{350} + p_8 = 1 \tag{1}$$

Løser med hensyn på p_8:

$$\underline{\underline{p_8}} = 1 - \left(\frac{5}{350} + \frac{40}{350} + \frac{25}{350} + \frac{80}{350} + \frac{8}{350} + \frac{68}{350} + \frac{22}{350} \right) \tag{2}$$

$$= 1 - \frac{248}{350} \tag{3}$$

$$= \frac{350}{350} - \frac{248}{350} \tag{4}$$

$$= \frac{350 - 248}{350} \tag{5}$$

$$= \underline{\underline{\frac{102}{350}}} \quad , \qquad \text{q.e.d.} \tag{6}$$

b) Fra figuren tilhørende oppgaven (utfallsrom) ser vi at utfallene i utfallsrommet ikke overlapper. [1]
Derfor er utfallene disjunkte.

c) Siden sannsynlighetene $p_1 = \frac{5}{350}$, $p_2 = \frac{40}{350}$ osv. er alle ulike, altså ikke uniformt utfallsrom, så gjelder ikke urnemodellen for vår sannsynlighetsmodell. [2]

d) Definisjonen av begivenheten for aksepterte tilbud A:

$$\underline{\underline{A \;=\; \Big\{ \; (\color{red}{1}, h, a) \; , \; (\color{red}{1}, \overline{h}, a) \;\;,\;\; (\color{blue}{2}, h, a) \; , \; (\color{blue}{2}, \overline{h}, a) \; \Big\}}}$$

Definisjonen av begivenheten for høye tilbud H:

$$\underline{\underline{H \;=\; \Big\{ \; (\color{red}{1}, h, a) \; , \; (\color{red}{1}, h, \overline{a}) \;\;,\;\; (\color{blue}{2}, h, a) \; , \; (\color{blue}{2}, h, \overline{a}) \; \Big\}}} \tag{7}$$

[1] I likhet med brikkene i et puslespill så overlapper utfallene i utfallsromet akkurat ikke.
[2] Fra figuren tilhørende oppgave 3 i oppgaveteksten (urne) ser vi også at kulene har forskjellige størrelse. Det visualiserer at sannsynlighetene p_i er forskjellige.

e) Ved å bruke sannsynlighetene oppgitt i lign.(1) i oppgaveteksten og samme fremgangsmåten som i lign.(5) i oppgaveteksten, så innser man at begivenheten H for høye tilbud er gitt ved:

$$\begin{aligned} \underline{\underline{P(H)}} &= p_1 + p_3 + p_5 + p_7 \\ &= \frac{5}{350} + \frac{25}{350} + \frac{8}{350} + \frac{22}{350} = \frac{60}{350} = \underline{\underline{0.1714}} \end{aligned} \tag{8}$$

f) Fra første ligningen i oppgaveteksten finner vi:

$$P(S_1 \cap H \cap A) = p_1 = \frac{5}{350} \tag{9}$$

og

$$P(S_1 \cap H) = p_1 + p_3 = \frac{5}{350} + \frac{25}{350} = \frac{30}{350} \tag{10}$$

Dermed har vi det vi trenger for å regne ut den betingede sannsynligheten:

$$\underline{\underline{P(A|S_1 \cap H)}} = \frac{P(S_1 \cap H \cap A)}{P(S_1 \cap H)} = \frac{\frac{5}{\cancel{350}}}{\frac{30}{\cancel{350}}} = \underline{\underline{\frac{5}{30}}} \quad (= 0.1667) \tag{11}$$

e) Tolkning: [3]

$\underline{\underline{P(A|S_1 \cap H)}}$ = sannsynligheten for at et tilbud blir aksptert gitt at tilbudet er <u>laget av selger 1 og verdien av tilbudet er høyt</u>

■

[3] Det er flere måter å formulere seg på her. Løsningsforslaget er bare én mulig formulering.

Oppgave 2: (logistikk - prognostisering)

a) Y beskriver en forsøksserie som har følgende egenskaper:

1. Hvert tilbud har kun 2 mulige utfall, akseptert (suksess) eller ikke akseptert (fiasko).
2. Alle tilbudene har samme sannsynlighet p for å bli akseptert, hvor $p = 0.1667$ i vårt tilfelle.
3. I oppgaven antas det at tilbudene er uavhengig av hverandre.
4. Det gjennomføres et bestemt antall forsøk, $n = 4$ i vårt tilfelle.

Forsøksserien oppfyller dermed kravene til en **binomisk** forsøksserie.
Den stokastiske variablene Y er da binomisk fordelt: $Y \sim \text{Bin}\big[\, n = 4, p = 0.1667 \,\big]$.

b) Siden Y binomisk fordelt, $Y \sim \text{Bin}\big[\, n = 4, p = 0.1667 \,\big]$,
så finner vi formelen for sannsynlighet i formelsamlingen:

$$\underline{\underline{P(Y = 2)}} = \binom{n}{2} p^1 \, (1 - p)^{n-2} \tag{12}$$

$$= \binom{4}{2} 0.1667^1 \, (1 - 0.1667)^{4-2} = \underline{\underline{0.1158}} \tag{13}$$

c) Siden Y er binomisk fordelt, $Y \sim \text{Bin}\big[\, n = 4, p = 0.1667 \,\big]$,
så finner vi formelen for forventing

$$\underline{\underline{E[Y]}} \;=\; n \cdot p \;=\; 4 \cdot 0.1667 \;=\; \underline{\underline{0.6668}} \tag{14}$$

Tolkning:

$$\begin{aligned} \underline{\underline{E[Y]}} \quad &= \quad \text{forventet antall tilbud som blir aksepert} \\ & \quad\;\; \underline{\underline{\text{av de 4 høyre tilbudene laget av selger 1}}} \end{aligned} \tag{15}$$

d) Bruker regnereglene for forventingen i formelsamlingen og
finner forventningen av V:

$$E[V] \;=\; E\big[\, v_1 Y_1 \;+\; v_2 Y_2 \;+\; v_3 Y_3 \;+\; v_4 Y_4 \,\big] \tag{16}$$

$$\;=\; v_1 E[Y_1] \;+\; v_2 E[Y_2] \;+\; v_3 E[Y_3] \;+\; v_4 E[Y_4] \tag{17}$$

hvor $E[Y_1] = E[Y_2] = E[Y_3] = E[Y_4] = p$ (oppgitt i en fotnote i oppgaven). Dermed:

$$\underline{\underline{E[V]}} \;=\; v_1 p + v_2 p + v_3 p + v_4 p \tag{18}$$

$$\;=\; \underline{\underline{\big(v_1 + v_2 + v_3 + v_3\big)\, p}} \qquad , \qquad \text{q.e.d.} \tag{19}$$

hvor p er en felles faktor som vi faktoriserer utenfor.

e) Bruker lign.(19) og setter inn tallene for v_i og p som oppgitt i oppgaven:

$$\underline{\underline{E[V]}} = \big(v_1 + v_2 + v_3 + v_3\big)\, p \tag{20}$$

$$= \Big(40\,000 + 33\,000 + 105\,000 + 75\,000\Big) \cdot 0.1667 \ \text{ NOK} = \underline{\underline{42\,175 \ \text{ NOK}}} \tag{21}$$

f) Bruker regnereglene for varians i formelsamlingen og finner variansen av V:

$$Var[V] = Var\big[\, v_1Y_1 + v_2Y_2 + v_3Y_3 + v_4Y_4 \,\big] \tag{22}$$

$$\overset{\text{uavh.}}{=} v_1^2 Var[Y_1] \ + \ v_2^2 Var[Y_2] \ + \ v_3^2 Var[Y_3] \ + \ v_4^2 Var[Y_4] \tag{23}$$

siden alle Y_i'ene er uavhengige.
I en fotnote i oppgaven er det oppgitt at: $\left(Y_i \sim \text{Bin}[\, n_i = 1, p \,] \right)$ [4]

$$Var[Y_1] = Var[Y_2] = Var[Y_3] = Var[Y_4] = p(1-p) \tag{24}$$

Variansen blir da:

$$\underline{\underline{Var[V]}} = v_1^2 p(1-p) \ + \ v_2^2 p(1-p) \ + \ v_3^2 p(1-p) \ + \ v_4^2 p(1-p) \tag{25}$$

$$= \underline{\underline{\big(v_1^2 \ + \ v_2^2 \ + \ v_3^2 \ + \ v_4^2\big)\, p(1-p)}} \quad , \quad \text{q.e.d.} \tag{26}$$

hvor $p(1-p)$ er en felles faktor som vi faktoriserer utenfor.

[4] En *bernullifordeling* er bare et spesialtilfelle av en binomialfordelingen: en bernullifordeling er en binomialfordelingen med $n_i = 1$. Siden bernullifordeling ikke er nevnt eksplisitt i forelesningene eller kompendiet så brukes navnet "binomialfordelingen" både i oppgaven og løsningen.

g) Siden alle verdiene v_i er forskjellige så er også forventingsverdiene til leddene V_i i lineærkombinasjonen forskjellige.

Dermed er en av forutsetningene til sentralgrenseteoremet brutt [5] og sentralgrenseteoremet kan ikke garanatere at V blir normalfordelt.

Dermed er ikke summen $V = v_1 Y_1 + v_2 Y_2 + ... + v_{200} Y_{200}$ normalfordelt selv om tilbudene, dvs. Y_i'ene, er uavhengige og aksept-sannsynlighetene p er like.

■

[5] For at sentralgrenseteoremet skal gjelde så må sannsynlighetsfordelingene til de stokastiske variablene V_i i en lineærkombinasjon være uavhengige og identiske slik at forventingene og variansene i lineær kombinasjonen er like. De stokastiske variablene Y_i *er* faktisk identiske og uavhengige, men forventingsverdiene og variansene til de ulike leddene, dvs. til V_i, i lineær kombinasjonen er forskjellige siden v_i er forskjellige.

Oppgave 3: (logistikk og økonomi)

a) Forventet etterspørsel av antall kg med bananer per dag på dagtid:

$$\underline{\underline{E[D_{\text{tot}}]}} = E\left[\, D_1 + D_2 + D_3 + D_4 \,\right] \tag{27}$$

$$= E[D_1] + E[D_2] + E[D_3] + E[D_4] \tag{28}$$

$$= 40 + 100 + 60 + 20 \tag{29}$$

$$= \underline{\underline{220}} \qquad , \qquad \text{q.e.d.} \tag{30}$$

Merk:
Overgangen i lign.(28) gjelder alltid, uansett om D_i'ene er uavhengige eller ikke.

b) Siden de stokastiske variablene D_i er uavhengige, så er tilhørende kovarians lik null.
Variansen blir dermed: (se formelsamlingen)

$$\underline{\underline{Var[D_{\text{tot}}]}} = Var\left[\, D_1 + D_2 + D_3 + D_4 \,\right] \tag{31}$$

$$\overset{\text{uavh.}}{=} Var[D_1] + Var[D_2] + Var[D_3] + Var[D_4] \tag{32}$$

$$= 64 + 144 + 352 + 16 \tag{33}$$

$$= \underline{\underline{576}} \tag{34}$$

c) Det er oppgitt i oppgaven at alle D_i er normalfordelte:

$$D_i \;\sim\; N\big[\; E[D_i]\,,\, Var[D_i] \;\big] \tag{35}$$

I tillegg er det oppgitt i oppgaven at alle D_i'ene er uavhengige. [6]
Fra læresetningen i formelsamlingen [7] vet vi da at summen
$D_{\text{tot}} = D_1 + D_2 + D_3 + D_4$ også er normalfordelt, dvs.:

$$\underline{\underline{D_{\text{tot}} \;\sim\; N\big[\; E[D_{\text{tot}}]\,,\, Var[D_{\text{tot}}] \;\big]}} \tag{36}$$

hvor $E[D_{\text{tot}}] = 220$ og $Var[D_{\text{tot}}] = 576$ fra oppgave **3a** og **3b**, henholdsvis.

[6] At D_i'ene er uavhengige er det samme som å si at etterspørselen av bananer i de 4 butikkene er uavhengige.
[7] Se side 108 i formelsamlingen fra 2018.

d) Med innkjøpspris (kostnad for Kiwi) $c = 8$ NOK og utslagspris $p = 12.9$ NOK fås:

$$P(D_3 \leq q_3^*) = 1 - \frac{c}{p} \tag{37}$$

$$P(D_3 \leq q_3^*) = 1 - \frac{8}{12.9} \tag{38}$$

$$P(D_3 \leq q^*) = 0.38 \tag{39}$$

$$P\Bigg(\underbrace{\frac{D_3 - E[D_3]}{\sigma[D_3]}}_{\equiv Z} \leq \underbrace{\frac{q_3^* - E[D_3]}{\sigma[D_3]}}_{\equiv z_0} \Bigg) \overset{\text{standardiser}}{=} 0.38 \tag{40}$$

$$P(Z \leq z_0) = 0.38 \tag{41}$$

z_0 finnes ved "omvendt tabelloppslag".Men z_0 var oppgitt i oppgaven. Dermed slipper man å gjøre dette omvendte tabelloppslaget. [8]

$$z_0 = -0.30 \tag{42}$$

Vi løser:

$$z_0 = \frac{q_3^* - E[D_3]}{\sigma[D_3]} \tag{43}$$

med hensyn på q^*:

$$\underline{q_3^*} = E[D_3] + z_0 \cdot \sigma[D_3] \tag{44}$$

$$= 60 - 0.30 \cdot \sqrt{352} = \underline{54.4} \tag{45}$$

Bunnpris må bestille $\underline{\underline{q_3^* \approx 54.4}}$ kg med banan per gang for å få størst mulig forventet fortjeneste.

[8] Følgende kommentar behøver man ikke ha med i eksamensbesvarelsen fordi $z_0 = -0.30$ er oppgitt: Sannsynligheten $P(Z \leq z_0) = 0.38$ finnes ikke direkte i tabellen i formelsamlingen fordi sannsynlighetene i formelsamlingen ikke viser sannsynligheter under 0.5000. Siden $P(Z = z_0)$ er symmetrisk om origo så er $P(Z \leq -z_0) = P(Z \geq z_0)$. (Helt generelt er en funksjon $f(x)$ symmetrisk om y-aksen dersom $f(x) = f(-x)$.) Arealet under en sannsynlighetsfunksjon $P(Z = z_0)$ er en sannsynlighet som er normalisert, dvs. $P(Z \leq z_0) + P(Z \geq z_0) = 1$. Ved å kombinere disse sammenhengene gir: $P(Z \leq -z_0) = 1 - 0.38 = 0.62$.

e) For å finne p_{red} må vi først regne ut sannsynligheten.
Deretter kan vi regne ut p_{red} via den oppgitte ligningen i oppgaveteksten.

Fra oppgaveteksten vet vi at D_i er normalfordelt.
Fra tabellen i oppgave 3 vet vi at $E[D_3] = 60$ og $\sigma[D_3] = \sqrt{Var[D_3]} = \sqrt{352}$.
Dermed kan vi regne ut sannsynligheten $P(D_3 \leq q_3^*)$ via standardisering og tabelloppslag:

$$\underline{P(D_3 \leq q^*)} = P\left(\underbrace{\frac{D_3 - E[D_3]}{\sigma[D_3]}}_{=Z} \leq \frac{q_3^* - E[D_3]}{\sigma[D_3]}\right) \tag{46}$$

$$= P\left(Z \leq \frac{70 - 60}{\sqrt{352}}\right) \tag{47}$$

$$= P(Z \leq 0.53) \tag{48}$$

$$\overset{\text{tabell}}{=} \underline{0.7019} \tag{49}$$

Ligningen oppgitt i oppgaveteksten er:

$$P(D_3 \leq q_3^*) = \frac{p - c}{p - p_{\text{red}}} \tag{50}$$

$$p - p_{\text{red}} = \frac{p - c}{P(D_3 \leq q_3^*)} \tag{51}$$

som gir:

$$\underline{p_{\text{red}}} = p - \frac{p - c}{P(D_3 \leq q_3^*)} \tag{52}$$

$$= \left(12.9 - \frac{12.9 - 8}{0.7019}\right) \text{NOK} = \underline{5.92 \text{ NOK}} \tag{53}$$

Bunnpris må sette den reduserte prisen til $\underline{\underline{p_{\text{red}} = 5.92 \text{ NOK}}}$
for å maksimere den forventede fortjetensten.

■

Oppgave 4: (økonomi)

a) Vi bruker læresetningen på side 66 i formelsamlingen for å finne minste kvadraters **regresjonslinje** for x og y. Parametrene $\hat{\beta}$ og $\hat{\alpha}$ er da: (dropper benevning her)

$$\underline{\hat{\beta}} \;=\; \frac{S_{xy}}{S_x^2} \;=\; \frac{18}{2.5} \;=\; \underline{7.2} \tag{54}$$

$$\underline{\hat{\alpha}} \;=\; \overline{y} \;-\; \beta\overline{x} \;=\; 284 - 7.2 \cdot 3 \;=\; \underline{\underline{262.4}} \tag{55}$$

Minste kvadraters lineære regresjonslinje $\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta}\,x$ blir dermed:
(se lign.(11.7) i formelsamlingen fra 2018)

$$\underline{\underline{\hat{y} \;=\; 262.4 \;+\; 7.2\,x}} \tag{56}$$

b) i) En regresjonslinje mellom variablene x og y sier at for en gitt verdi av x så kan y estimeres/predikeres via regresjonslinjen.

ii) Parameteren $\hat{\beta}$ er stigningstallet for en lineær regresjonslinje. Parameteren angir estimert/predikert endring $\hat{y}$ når x endres med èn enhet.

iii) Parameteren $\hat{\beta} = 7.2$ i oppgave **4a**. Det betyr at aksjeprisen til Telenor øker med 7.2 NOK per dag.

c) Regresjonslinjen i lign.(56) predikerer at aksjekursen til Telenor er:

$$\underline{\underline{\hat{y}(12)}} \;=\; \big(262.4 \;+\; 7.2 \cdot 12\big)\;\text{NOK} \;=\; \underline{\underline{348.8\;\;\text{NOK}}} \tag{57}$$

etter 12 dager.

d) i) Forklaringsstyrken R^2 er normallisert til: $0 \le R^2 \le 1$.

ii) Forklaringsstyrken R^2 er enhetsuavhengig, den er benevningsløs.

iii) Forklaringsstyrken R^2 er et mål på:

- i hvor stor grad x_i kan brukes til å **predikere** y_i
- hvor godt de faktiske observasjonene y_i **passer** med lineær regresjonen $\hat{y}_i$
- styrken i **samvariasjon** mellom prediksjonene $\hat{y}_i$ og de faktiske obervasjonene y_i

Det er nok at man nevner èn eller lignende av disse kulepunkt-kommentarene.

e) Forklaringsstyrken R^2 kan leses direkte fra Excel-utskriften:
(Se cellen som heter "*R Square*" i Excel-utskriften): [9]

$$\underline{\underline{R^2 \;=\; 0.9744}} \tag{58}$$

f) Kommentar til svaret i oppgave **4e**:

At $R^2 = 0.9744$ betyr at for en gitt dag innenfor den aktuelle perioden hvor den lineære trenden gjelder så ligger observasjonene svært tett på regresjonslinjen. [10]

[9] Man kan også regne ut forklaringstyrken R^2 "for hånd" via definisjonen $R^2 = 1 - SSE/SST$. Men det er mye mer arbeidskrevende. Fint at dataprogrammer (som f.eks. Excel) kan hjelpe oss med slikt.

[10] Litt mer presist og litt mer teknisk: At $R^2 = 0.9744$ betyr at 97.44 % av variansen til observasjonene ligger innenfor regresjonslinjen. (Denne fotnoten er bare en ekstrainformasjon som ikke kreves at studentene har med i sin eksamensbesvarelse.

At $R^2 = 0.9744$ betyr at for en gitt dag innenfor perioden hvor den lineære trenden gjelder så kan vi "i stor grad", tilsvarende 97.44 %, forutsi aksjekursen.
Vi sier at modellen har stor forklaringsstyrke.

■

Kapittel 15

Testeksamen 2019

LØSNING: TESTeksamen mai 2019

"*MAT110 Statistikk 1*", vår 2019

Versjon 04

Oppgave 1: (telling av lakselus)

a) Utfallsrommet Ω_3 er gitt ved

$$\underline{\underline{\Omega_3}} = \underline{\underline{\begin{aligned} \Big\{ & (\overline{l_1}, \overline{l_2}, \overline{l_3}) \;,\; (l_1, l_2, l_3) \;, \\ & (l_1, \overline{l_2}, \overline{l_3}) \;,\; (\overline{l_1}, l_2, l_3) \;, \\ & (\overline{l_1}, l_2, \overline{l_3}) \;,\; (l_1, \overline{l_2}, l_3) \;, \\ & (\overline{l_1}, \overline{l_2}, l_3) \;,\; (l_1, l_2, \overline{l_3}) \quad \Big\} \end{aligned}}}$$

b) Utfallsrommet Ω_3 har et endelig antall utfall.
Derfor er utfallsrommet <u>diskret</u>.

c) De blå utfallene i utfallsrommet Ω_3 er utfall hvor minst to av de tilfeldig trekte fiskene har lakselus:

$$
\begin{aligned}
\Omega_3 \quad = \quad \Big\{ \; & (\overline{l_1}, \overline{l_2}, \overline{l_3}) \;\; , \;\; (l_1, l_2, l_3) \;\; , \\
& (l_1, \overline{l_2}, \overline{l_3}) \;\; , \;\; (\overline{l_1}, l_2, l_3) \;\; , \\
& (\overline{l_1}, l_2, \overline{l_3}) \;\; , \;\; (l_1, \overline{l_2}, l_3) \;\; , \\
& (\overline{l_1}, \overline{l_2}, l_3) \;\; , \;\; (l_1, l_2, \overline{l_3}) \quad \Big\}
\end{aligned}
$$

dvs. begivenheten S er:

$$
\begin{aligned}
\underline{\underline{S}} \quad = \quad \Big\{ \; & \quad\quad\quad , \;\; (l_1, l_2, l_3) \;\; , \\
& \quad\quad\quad , \;\; (\overline{l_1}, l_2, l_3) \;\; , \\
& \quad\quad\quad , \;\; (l_1, \overline{l_2}, l_3) \;\; , \\
& \quad\quad\quad , \;\; (l_1, l_2, \overline{l_3}) \quad \Big\}
\end{aligned}
$$

som er en delmengde av utfallsrommet Ω_3.

d) Sannsynligheten for begivenheten S er:

$$
\underline{\underline{p(S)}} = P(l_1, l_2, l_3) + P(\overline{l_1}, l_2, l_3) + P(l_1, \overline{l_2}, l_3) + P(l_1, l_2, \overline{l_3}) \tag{1}
$$

$$
= p^3 + (1-p)p^2 + p(1-p)p + p^2(1-p) \tag{2}
$$

$$
= \underline{\underline{p^3 + 3p^2(1-p)}} \tag{3}
$$

Med $p = 0.5$ blir den numeriske verdien:

$$\underline{\underline{p(S)}} = p^3 + 3p^2(1-p) \tag{4}$$

$$= 0.5^3 + 3 \cdot 0.5^2(1-0.5) \tag{5}$$

$$= 4 \cdot 0.5^3 \tag{6}$$

$$= \underline{\underline{0.5}} \tag{7}$$

e) Utfallsrommet Ω_{10} har 2^{10} forskjellige utfall.
Det nytter dermed ikke å skrive opp alle utfallene eksplisitt som vi gjorde i oppgave **1a**.

Vi beskriver derfor utfallsrommet ved hjelp av symbolikk:

$$\underline{\underline{\Omega_{10} = \left\{ (a_1, a_2, \ldots, a_{10}) \;\middle|\; a_i \in \{l_i, \overline{l_i}\},\ i = 1, 2, \ldots, 10 \right\}}} \tag{8}$$

f) Sannsynligheten for et gitt vilkårlig utfall er:

$$\underline{P(u)} = P(a_1, a_2, \ldots, a_{10}) = \underline{P(a_1)P(a_2)\cdots P(a_{10})} \tag{9}$$

I oppgaven er det oppgitt at $p = 0.5$, hvor $p =$ sannsynligheten for at en vilkårlig valgt fisk har lakselus. Det betyr, for $a_i = l_i$

$$\underline{P(l_i)} = p = \underline{0.5} \tag{10}$$

Tilsvarende for $a_i = \overline{l_i}$

$$\underline{P(\overline{l_i})} = 1 - p = 1 - 0.5 = \underline{0.5} \tag{11}$$

Konklusjon:
$P(a_i) = 0.5$ uansett om $a_i = l_i$ eller $a_i = \overline{l_i}$.
Dermed har alle mulige utfall samme sannsynlighet, dvs. uniformt utfallsrom. [1]

Dersom p *ikke* var akkurat lik 0.5, ville modellen *ikke* blitt uniform, siden utfallene da hadde fått forskjellige sannsynligheter for å inntreffe, avhengig av hvor mange lakselus man fikk.

[1] Det viktige i denne oppgaven er at alle mulige utfall samme sannsynlighet, ikke at den numeriske verdien er $P(u) = 0.5^{10}$.

g) Siden utfallsrommet er uniformt (se oppgave **1f**) så kan vi bruke urnemodellen.
La:

$$A = \text{begivenheten at 7 av de 10 trekte fiskene har lakselus} \tag{12}$$

Dermed: [2]

$$\underline{\underline{P(A)}} = \frac{\text{gunstige}}{\text{mulige}} = \frac{\binom{10}{7}}{2^{10}} = \frac{120}{1024} = \underline{\underline{0.1172}} \tag{14}$$

siden

$$\text{gunstige} = \binom{10}{7} \quad (\text{ situasjon 3 }) \tag{15}$$

$$\text{mulige} = 2^{10} \tag{16}$$

hvor antall gunstige er funnet ved å telle hvor mange måter vi kan trekke 10 kuler uten tilbakelegging fra en urne hvor det er 7 røde kuler og 3 grønne kuler, dvs. totalt 10 kuler i urnen. [3]

---

[2] Siden rekkefølgen ikke spiller noen rolle og siden vi ikke har tilbakelegging så har vi situasjon 3. Dermed:

$$\text{gunstige} = \binom{10}{7} \quad (\text{binomialkoeffisienten}) \tag{13}$$

[3] Siden vi i oppgave **1f** viste at utfallsrommet er uniformt og siden hintet i oppgaven sier at man kan bruke urnemodellen så legges det opp til å løse oppgaven på den måten. Det er ikke meningen at dere skal løse oppgaven på to måter. men vi nevner likevel at man kan alternativt løse oppgaven via stokastiske variabler:

$$X = \text{antall trekte fisker med lakselus i en forsøksserie på totalt } n = 10 \text{ forsøk} \tag{17}$$

Siden det er kun er to mulige utfall for et forsøk og siden fiskene antas å være uavhengige av hverandre når gjelder lakselus så innser vi at X er binomisk fordelt:

$$X \sim \text{Bin}[n = 10, p = 0.5] \tag{18}$$

Dermed:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \tag{19}$$

Med $n = 10$, $x = 7$ og $p = 0.5$ får vi:

$$\underline{\underline{P(X = 7)}} = \binom{10}{7} 0.5^7 (1-0.5)^{10-7} = 120 \cdot 0.5^{10} = \underline{\underline{0.1172}} \tag{20}$$

altså det samme som lign.(14).

h) La:

$$\begin{aligned} B &= \text{være begivenheten hvor fisk nr.1 og fisk nr.10 har lakselus} \\ C &= \text{være begivenheten hvor minst 8 av de 10 trekte fiskene har lakselus} \end{aligned} \tag{21}$$

Oppgaven spør om sannsynligheten for at C inntreffer gitt B, dvs:
(betinget sannsynlighet)

$$\boxed{P(C|B) \;=\; \frac{P(C \cap B)}{P(B)} \qquad (22)}$$

Urnemodellen: [4]

$$\underline{P(C \cap B)} = \frac{\text{gunstige}}{\text{mulige}} = \frac{\overbrace{\binom{8}{6} + \binom{8}{7} + \binom{8}{8}}^{\text{situasjon 3}}}{2^8} = \frac{28 + 8 + 1}{256} = \underline{0.1445} \tag{23}$$

$$P(B) = \frac{\text{gunstige}}{\text{mulige}} = \frac{2^8}{2^{10}} = \frac{1}{2^2} = 0.25 \tag{24}$$

hvor antall gunstige finnes ved å trekke 8 kuler uten tilbakelegging fra en urne hvor det er 6, 7 og 8 røde kuler hhv.

Dermed:

$$\underline{\underline{P(C|B)}} = \frac{P(C \cap B)}{P(B)} = \frac{0.1445}{0.25} = \underline{\underline{0.5781}} \tag{25}$$

[4] Det legges opp til at man skal bruke urnemodellen her. Vi nevner likevel at også denne oppgaven kan løses ved å bruke stokastiske variabler analogt til fotnoten i oppgave **1g**.

Oppgave 2: (stokastiske variabler og forventningverdi - telling av lakselus)

a) Tolkning:

$$\underline{\underline{P(X \leq j)}} \quad = \quad \text{sannsynligheten for at en tilfeldig valgt fisk} \tag{26}$$

$$\underline{\underline{\text{fra et oppdrettsanlegg med lusetall mellom 0.1 og 0.3 har } j \text{ eller mindre antall lakselus}}}$$

b) Sannsynlighetene $P(X = j)$ med $j = 0, 1, 2, 3, 4, 5$ utgjør en gyldig sannsynlighetsfordeling fordi den er normert til 1:

$$\underline{\underline{\sum_{j=0}^{5} P(X = j)}} \quad = \quad P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \tag{27}$$

$$+ \; P(X = 4) + P(X = 5)$$

$$= \quad 0.85 + 0.12 + 0.0185 + 0.0070 + 0.0035 + 0.0010 \tag{28}$$

$$= \quad \underline{\underline{1}} \tag{29}$$

c) Forventningsverdien $E[X]$:

$$\underline{\underline{E[X]}} \quad = \quad \sum_{j=0}^{5} x_j \, P(X = j) \tag{30}$$

$$= \quad \sum_{j=0}^{5} j \, P(X = j) \tag{31}$$

$$= \quad 0 \cdot 0.85 \; + \; 1 \cdot 0.12 \; + \; 2 \cdot 0.0185 \; + \; 3 \cdot 0.0070 \; + \; 4 \cdot 0.0035 \; + \; 5 \cdot 0.0010 \tag{32}$$

$$= \quad \underline{\underline{0.1970}} \tag{33}$$

d) Variansen $Var[X]$:

$$\underline{\underline{Var[X]}} = \sum_{j=0}^{5} \Big(x_j - E[X]\Big)^2 P(X = j) \tag{34}$$

$$= \sum_{j=0}^{3} \Big(j - E[X]\Big)^2 P(X = j) \tag{35}$$

$$= \Big(0 - 0.1970\Big)^2 \cdot 0.85 + \Big(1 - 0.1970\Big)^2 \cdot 0.12$$

$$+ \Big(2 - 0.1970\Big)^2 \cdot 0.0185 + \Big(3 - 0.1970\Big)^2 \cdot 0.0070 \tag{36}$$

$$+ \Big(4 - 0.1970\Big)^2 \cdot 0.0035 + \Big(5 - 0.1970\Big)^2 \cdot 0.0010 \tag{37}$$

$$= \underline{\underline{0.2992}} \tag{38}$$

med tilhørende standardavvik $\sigma[X]$:

$$\underline{\underline{\sigma[X]}} = \sqrt{Var[X]} \tag{39}$$

$$= \sqrt{0.2992} = \underline{\underline{0.5470}} \tag{40}$$

e) Forventet mengde medisin $E[X]$: [5]

$$\underline{E[M]} = E[aX^2 + bX + c] \tag{41}$$

$$= \underline{aE[X^2] + bE[X] + c} \tag{42}$$

hvor (fra oppgave **2c**)

$$E[X] = 0.1970 \tag{43}$$

Varianssetningen

$$Var[X] = E[X^2] - E[X]^2 \tag{44}$$

gir

$$\boxed{E[X^2] = Var[X] + E[X]^2 \qquad (45)}$$

hvor (fra oppgave **2d**)

$$Var[X] = 0.2992 \tag{46}$$

Dette innsatt i lign.(42) gir

$$\underline{E[M]} = aE[X^2] + bE[X] + c \tag{47}$$

$$= a\Big(Var[X] + E[X]^2\Big) + bE[X] + c \tag{48}$$

$$= 1.2\Big(0.2992 + 0.1970^2\Big) + 2.3 \cdot 0.1970 + 1.4 \tag{49}$$

$$= \underline{2.2587} \tag{50}$$

Forklaring $E[M]$:
Den forventede mengden med medisin som må til for å fjerne lakselusen til fisken er
2.2587 gram medisin per 100 gram fôr.

[5]Her brukes regnereglene for forventning, se formelsamling.

Kommentar:

Man kan også regne ut $E[X^2]$ i lign.(44) via definisjonen av forventning: [6]

$$\underline{E[X^2]} = \sum_{j=0}^{5} x_j^2 \, P(X = j) \tag{51}$$

$$= \sum_{j=0}^{5} j^2 \, P(X = j) \tag{52}$$

$$= 0^2 \cdot 0.85 \; + \; 1^2 \cdot 0.12 \; + \; 2^2 \cdot 0.0185 \; + \; 3^2 \cdot 0.0070 \tag{53}$$

$$= 4^2 \cdot 0.0035 \; + \; 5^2 \cdot 00.10 \tag{54}$$

$$= \underline{0.3380} \tag{55}$$

[6]Det er nok at man gjør det på èn av måtene.

f) Sannsynligheten for at gjennomsnittet $\overline{X}_t$ ligger under lusegrensen:

$$P(\overline{X}_t \leq a) \tag{56}$$

hvor $\quad a = 0.2.$

Siden $X_{kt} \overset{\text{i.i.d.}}{\sim} X \ \sim \ P(X=j)$ med 50 variabler, dvs. $k = 1, 2, \ldots, 50$, så er, ifølge sentralgrensesetningen, gjennomsnittet

$$\overline{X}_t \ = \ \frac{1}{n}\sum_{k=1}^{n} X_{kt} \tag{57}$$

normalfordelt:

$$\overline{X}_t \ \sim \ N\Big[\ E[\overline{X}_t] \ , \ \sigma[\overline{X}_t] \ \Big] \tag{58}$$

hvor

$$E[\,\overline{X}_t\,] \ = \ \overbrace{E[X] \ = \ 0.1970}^{\text{oppg. } \mathbf{2c}} \tag{59}$$

$$\sigma[\,\overline{X}_t\,] \ = \ \overbrace{\frac{\sigma[X]}{\sqrt{n}} \ = \ \frac{0.5470}{\sqrt{50}}}^{\text{oppg. } \mathbf{2d}} \tag{60}$$

Standardiserer:

$$\underline{P(\overline{X}_t \leq a)} \ \overset{\text{standardiser}}{=} \ P\left(\frac{\overline{X}_t - E[\overline{X}_t]}{\sigma[\overline{X}_t]} \ \leq \ \frac{a - E[\overline{X}_t]}{\sigma[\overline{X}_t]} \right) \tag{61}$$

$$= \ P\left(\underbrace{\frac{\overline{X}_t - E[X]}{\sigma[X]/\sqrt{n}}}_{\equiv Z} \ \leq \ \frac{a - E[X]}{\sigma[X]/\sqrt{n}} \right) \tag{62}$$

$$= \ P\left(Z \ \leq \ \frac{0.2 - 0.1970}{0.5470/\sqrt{50}} \right) \tag{63}$$

$$= \ \underline{P\Big(Z \ \leq \ 0.04 \Big)} \tag{64}$$

Ulikhetstegnet er "rett" vi sammenlignet med tabellen i formelsamlingen.
Tabelloppslag:

$$\underline{\underline{P(\overline{X}_t \leq a)}} = \overbrace{P\big(Z \leq 0.04\big)}^{=\, 0.5160} = \underline{\underline{0.5160}} \tag{65}$$

$\blacksquare$

Oppgave 3: (statistisk modell og estimatorer)

a) Siden $X \sim \text{NBin}[\lambda, \theta]$ så er:

$$E[X_i] \;=\; \lambda \tag{66}$$

for alle $i = 1, 2, ..., N$ fiskene. Forventningen av estimatoren $\hat{\lambda}$ blir da:

$$\begin{aligned}
\underline{\underline{E[\,\hat{\lambda}\,]}} \quad &= \quad E[\,\overline{X}\,] &(67)\\
&= \quad E\Big[\,\frac{1}{N}\sum_{i=1}^{N} X_i\,\Big] &(68)\\
&= \quad \frac{1}{N}\Big(\,\underbrace{E[X_1]}_{\lambda} \;+\; \underbrace{E[X_2]}_{\lambda} \;+\; \underbrace{E[X_3]}_{\lambda} \;+\; ... \;+\; \underbrace{E[X_N]}_{\lambda}\,\Big) &(69)\\
&= \quad \frac{\cancel{N}\,\lambda}{\cancel{N}} \;=\; \underline{\underline{\lambda}} &(70)
\end{aligned}$$

dvs. estimatoren $\hat{\lambda}$ er forventingsrett.

Realiseringen av $\hat{\lambda}$ er gitt ved: [7]

$$\underline{\underline{\hat{\lambda}(\,x_1, \ldots, x_N\,)}} \;=\; \overline{x} \;=\; \underline{\underline{0.1970}} \tag{71}$$

[7] Tallet $\overline{x} = 0.1970$ var oppgitt i tabell 1 i oppgavene.

b) Vi skal finne en estimator $\hat{\theta}$ for θ basert på ligningen

$$\sigma^2 \; = \; \lambda \; + \; \frac{\lambda^2}{\theta} \tag{72}$$

hvor $\hat{\sigma}^2 = S_X^2$ er en estimator for σ.

Løser nå lign. (73) mhp. θ:

$$\theta \; = \; \frac{\lambda^2}{\sigma^2 - \lambda} \tag{73}$$

Estimatorer: (dette er oppgitt i oppgaven)

$$\hat{\lambda} \; = \; \overline{X} \tag{74}$$

$$\hat{\sigma^2} \; = \; S_X^2 \tag{75}$$

Dette innsatt i lign.(73):

$$\underline{\underline{\hat{\theta}}} \; = \; \frac{\hat{\lambda}^2}{\hat{\sigma}^2 - \hat{\lambda}} \; = \; \underline{\underline{\frac{\overline{X}^2}{S_X^2 - \overline{X}}}} \tag{76}$$

Realiseringen av $\hat{\theta}$ er gitt ved: [8]

$$\underline{\underline{\hat{\theta}(\, x_1, \ldots, x_N \,)}} \; = \; \frac{\overline{x}^2}{s_X^2 - \overline{x}} \; = \; \frac{0.1970^2}{0.2970 - 0.1970} \; = \; \underline{\underline{0.3881}} \tag{77}$$

[8] Tallene $\overline{x} = 0.1970$ og $s_X^2 = 0.2970$ var oppgitt i tabell 1 i oppgavene.

c) Siden vi skal regne ut et *asymptotisk* konfidensintervall, dvs. for tilfellet $N \to \infty$, for estimatoren $\hat{\lambda}$ gitt ved:

$$\hat{\lambda} \;=\; \overline{X} \;=\; \frac{1}{N}\sum_{i=1}^{N} X_i \tag{78}$$

må vi bruke sentralgrensesetningen til å konstruere konfidensintervallet.

Oppgitt i oppgaven:

$$X_1,\, X_2,\, X_3,\, \ldots\, , X_N \;\overset{\text{i.i.d.}}{\sim}\; X \;\sim\; \text{NBin}[\,\lambda,\, \theta] \tag{79}$$

Vi har at:

$$E[\,\hat{\lambda}\,] \;=\; E[\,\overline{X}\,] \;=\; \lambda \qquad \text{(oppgitt i oppgaven)} \tag{80}$$

$$\sigma[\,\hat{\lambda}\,] \;=\; \sigma[\,\overline{X}\,] \;=\; \sqrt{\frac{\lambda + \frac{\lambda^2}{\theta}}{N}} \qquad (Var[X] \text{ oppgitt i oppgaven}) \tag{81}$$

Sentralgrensesetningen sier da at den skalerte variabelen Z er standard normalfordelt:

$$\underline{Z} \;=\; \frac{\overline{X} - E[\overline{X}]}{\sigma[\overline{X}]} \;=\; \frac{\overline{X} - \lambda}{\sqrt{\frac{\lambda + \frac{\lambda^2}{\theta}}{N}}} \;\overset{n\to\infty}{\sim}\; N[0,1] \tag{82}$$

Sentralgrensesetningen gir at

$$\lim_{N\to\infty} P\big(\, z_{\alpha/2} \,\leq\, Z \leq z_{1-\alpha/2} \,\big) \;=\; 1-\alpha \tag{83}$$

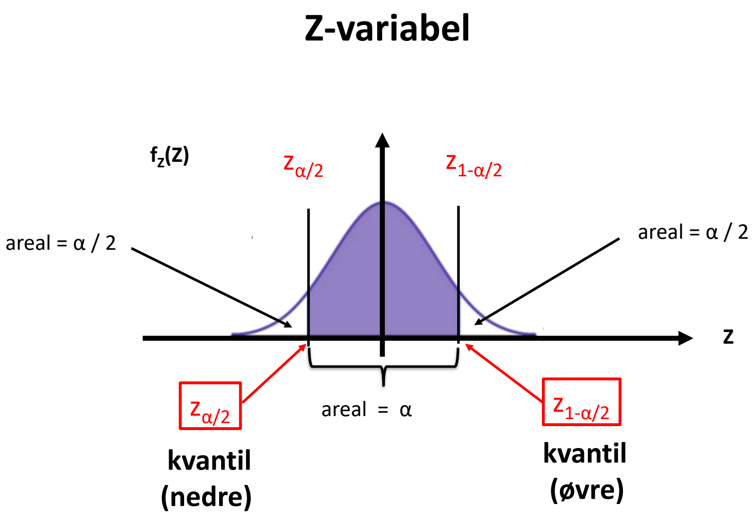
hvor (se figur 1)

$$z_{\alpha/2} \;=\; \text{nedre kvantil til normalfordelingen} \tag{84}$$
$$z_{1-\alpha/2} \;=\; \text{øvre kvantil til normalfordelingen} \tag{85}$$

Vi må transformere lign.(83) over på formen

$$\boxed{\lim_{N\to\infty} P_\lambda\big(\, LB_N^\mu \,\leq\, \lambda \leq UB_N^\mu \,\big) \;=\; 1-\alpha \qquad (86)}$$

som er definisjonen av et $(1-\alpha)\,100\,\%$-konfidensintervall for λ.



Figur 1: Kvantil.

Setter inn for Z fra lign.(82) inn i lign.(83):

$$\lim_{N\to\infty} P\left(z_{\alpha/2} \leq \frac{\overline{X} - \lambda}{\sqrt{\frac{\lambda + \frac{\lambda^2}{\theta}}{N}}} \leq z_{1-\alpha/2} \right) \quad = \quad 1 - \alpha \tag{87}$$

Deretter multipliserer vi uttrykket i $P(\cdots)$ med $\sqrt{\frac{\lambda+\frac{\lambda^2}{\theta}}{N}}$:

$$\lim_{N\to\infty} P\left(z_{\alpha/2}\sqrt{\frac{\lambda + \frac{\lambda^2}{\theta}}{N}} \leq \overline{X} - \lambda \leq z_{1-\alpha/2}\sqrt{\frac{\lambda + \frac{\lambda^2}{\theta}}{N}} \right) \quad = \quad 1 - \alpha \tag{88}$$

Deretter multipliserer vi uttrykket i $P(\cdots)$ med -1 hvor vi må huske at retningen på ulikhettegnene snus:

$$\lim_{N\to\infty} P\left(- z_{\alpha/2}\sqrt{\frac{\lambda + \frac{\lambda^2}{\theta}}{N}} \geq -\overline{X} + \lambda \geq -z_{1-\alpha/2}\sqrt{\frac{\lambda + \frac{\lambda^2}{\theta}}{N}} \right) \quad = \quad 1 - \alpha \tag{89}$$

Deretter summerer vi uttrykket i $P(\cdots)$ med $\overline{X}$ for å sette λ alene i midten:

$$\lim_{N\to\infty} P\left(\overline{X} - z_{\alpha/2}\sqrt{\frac{\lambda + \frac{\lambda^2}{\theta}}{N}} \geq \lambda \geq \overline{X} - z_{1-\alpha/2}\sqrt{\frac{\lambda + \frac{\lambda^2}{\theta}}{N}} \right) \quad = \quad 1 - \alpha \tag{90}$$

Deretter snur vi høyre og venstre side i $P(\cdots)$ hvor vi må huske at retningen på ulikhettegnene snus:

$$\lim_{N\to\infty} P\left(\overline{X} - z_{1-\alpha/2}\sqrt{\frac{\lambda + \frac{\lambda^2}{\theta}}{N}} \leq \lambda \leq \overline{X} - z_{\alpha/2}\sqrt{\frac{\lambda + \frac{\lambda^2}{\theta}}{N}} \right) \quad = \quad 1 - \alpha \tag{91}$$

Deretter erstatter vi λ med $\overline{X}$ og θ med $\hat{\theta}$ på høyre og venstre side i $P(\cdots)$:

$$\lim_{N\to\infty} P\left(\overline{X} - z_{1-\alpha/2}\sqrt{\frac{\overline{X} + \frac{\overline{X}^2}{\hat{\theta}}}{N}} \leq \lambda \leq \overline{X} - z_{\alpha/2}\sqrt{\frac{\overline{X} + \frac{\overline{X}^2}{\hat{\theta}}}{N}} \right) \quad = \quad 1-\alpha \tag{92}$$

Lign.(92) er et uttrykk på formen gitt i lign.(86) med

$$LB_N^\lambda \quad = \quad \overline{X} \; - \; z_{1-\alpha/2}\sqrt{\frac{\overline{X} + \frac{\overline{X}^2}{\hat{\theta}}}{N}} \tag{93}$$

$$UB_N^\lambda \quad = \quad \overline{X} \; - \; z_{\alpha/2}\sqrt{\frac{\overline{X} + \frac{\overline{X}^2}{\hat{\theta}}}{N}} \tag{94}$$

Vi har dermed vist at

$$\underline{\underline{[\; LB_N^\lambda \; , \; UB_N^\lambda \;] \; = \; \left[\overline{X} - z_{1-\alpha/2}\sqrt{\frac{\overline{X} + \frac{\overline{X}^2}{\hat{\theta}}}{N}} \; , \; \overline{X} - z_{\alpha/2}\sqrt{\frac{\overline{X} + \frac{\overline{X}^2}{\hat{\theta}}}{N}} \right]}} \quad , \quad \text{q.e.d.} \tag{95}$$

er et $(1-\alpha)\,100\,\%$ asymptotisk konfidensintervall for λ.

d) Se på uttrykket

$$\overline{X} + \frac{\overline{X}^2}{\hat{\theta}} \tag{96}$$

i konfidensintervallet i lign.(95). Fra lign.(76) har vi:

$$\hat{\theta} \;=\; \frac{\overline{X}^2}{S_X^2 - \overline{X}} \tag{97}$$

Innsatt:

$$\underline{\overline{X} + \frac{\overline{X}^2}{\hat{\theta}}} \;=\; \overline{X} + \frac{\overline{X}^2}{\frac{\overline{X}^2}{S_X^2 - \overline{X}}} \tag{98}$$

$$=\; \overline{X} + \frac{\overline{X}^2}{\overline{X}^2}(S_X^2 - \overline{X}) \;=\; \overline{X} + S_X^2 - \overline{X} \;=\; \underline{S_X^2} \tag{99}$$

Innsatt i lign. (95) gir

$$\underline{\underline{\left[\; LB_N^\lambda \;,\; UB_N^\lambda \;\right] \;=\; \left[\overline{X} - z_{1-\alpha/2}\frac{S_X}{\sqrt{N}} \;,\; \overline{X} - z_{\alpha/2}\frac{S_X}{\sqrt{N}}\right]}} \quad , \quad \text{q.e.d.} \tag{100}$$

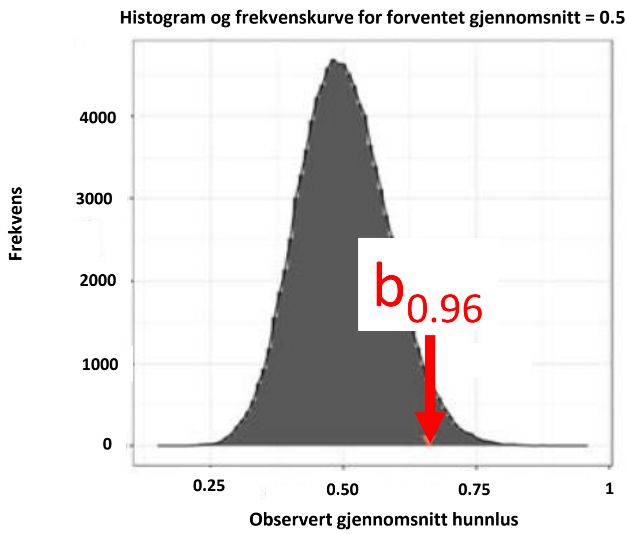
■

Oppgave 4: (krav som indikerer overskridelse av lusegrensa)

a) Krav 1 kan skrives som: [9]

$$\underline{\underline{P(\ \overline{X}_t\ >\ b_{0.96}\ )\ =\ 0.04}} \tag{102}$$

gitt at det sanne lusetallet er under grensen, dvs. gitt at $E[X] < 0.2$, hvor $b_{0.96}$ er 96 %-kvantilen for $\overline{X}_t$. (se se figur 2)



Figur 2: Kvantilen $b_{0.96}$ til $\overline{X}_t$.

[9] Man kan også inkludere den røde setningen inn i P med en mer kompakt notasjon:

$$\underline{\underline{P_{E[X]<a}(\ \overline{X}_t\ >\ b_{0.96}\ )\ =\ 0.04}} \tag{101}$$

b) Sannsynligheten for at krav 2 inntreffer er gitt ved: [10]

$$\underline{\underline{P(B) \;=\; P\big(\, \overline{X}_t > a \;\; \textbf{og} \;\; \overline{X}_{t+1} > a \;\; \textbf{og} \;\; \overline{X}_{t+2} > a \;\; \textbf{og} \; \overline{X}_{t+3} > a \,\big)}} \tag{104}$$

gitt at det sanne lusetallet er under grensen, dvs. gitt at $E[X] < 0.2$.

■

[10]Man kan også inkludere den røde setningen inn i P med en mer kompakt notasjon:

$$\underline{\underline{P(B) \;=\; P_{E[X<0.2]}\big(\, \overline{X}_t > a \;\; \textbf{og} \;\; \overline{X}_{t+1} > a \;\; \textbf{og} \;\; \overline{X}_{t+2} \geq a \;\; \textbf{og} \; \overline{X}_{t+3} > a \,\big)}} \tag{103}$$

Kapittel 16

Hovedeksamen 2019

LØSNING: Eksamen 21. mai 2019

"*MAT110 Statistikk 1*", vår 2019

Versjon 02

Oppgave 1: (telling av lakselus)

a) Utfallsrommet Ω har et endelig antall utfall.
Derfor er utfallsrommet diskret.

b) Hvert forsøk har 2 mulige utfall. Et eksperiment består av en sekvens med n forsøk.
Utfallsrommet Ω har dermed 2^n mulige utfall. [1]

c) Eksperimentet:

$$l_1 \ , \ l_2 \ , \ l_3 \ , \ \bar{l}_4 \tag{1}$$

betyr at:

- den tilfeldig valgte fisken fra merd 1 har lakselus
- den tilfeldig valgte fisken fra merd 2 har lakselus
- den tilfeldig valgte fisken fra merd 3 har lakselus
- den tilfeldig valgte fisken fra merd 4 har ikke lakselus

[1] Med $n = 4$ får vi $2^4 = 16$, som stemmer med Ω oppgitt i oppgaven.

Med $p = 0.5$:

$$\underline{\underline{P(l_1, l_2, l_3, \overline{l_4})}} = p^3(1-p)^1 \tag{2}$$

$$= 0.5^3 \cdot (1-0.5)^1 = 0.5^4 = \underline{\underline{0.0625}} \tag{3}$$

d) Ved å se på ligningen for begivenheten S som oppgitt i ligningen så innser vi at:

$$\begin{aligned}
\underline{\underline{p(S)}} = \; & P(l_1, l_2, l_3, l_4) \\
+ \; & P(\overline{l_1}, l_2, l_3, l_4) + P(l_1, \overline{l_2}, l_3, l_4) \\
+ \; & P(l_1, l_2, \overline{l_3}, l_4) + P(l_1, l_2, l_3, \overline{l_4}) \\
+ \; & P l_1, l_2, (\overline{l_3}, \overline{l_4}) + P(l_1, \overline{l_2}, l_3, \overline{l_4}) + P(l_1, \overline{l_2}, \overline{l_3}, l_4) \\
+ \; & P(\overline{l_1}, \overline{l_2}, l_3, l_4) + P(\overline{l_1}, l_2, \overline{l_3}, l_4) + P(\overline{l_1}, l_2, l_3, \overline{l_4})
\end{aligned} \tag{4}$$

$$= \underline{\underline{p^4 + 4p^3(1-p)^1 + 6p^2(1-p)^2}} \tag{5}$$

e) Oppgaven ber oss da finne den betingede sannsynligheten: [2]

$$\boxed{P(A|S) \;=\; \frac{P(A\cap S)}{P(S)} \qquad (6)}$$

hvor

$$P(A\cap S) \;=\; 6\cdot 0.5^4 \qquad \text{(oppgitt i oppgaven)} \tag{7}$$

Fra oppgave **1d** kjenner vi $P(S)$, se lign.(5).
Med $p = 0.5$:

$$P(S) \;=\; 11\cdot 0.5^4 \tag{8}$$

Den betingede sannsynligheten i lign.(6) blir da:

$$\underline{\underline{P(A|S)}} \;=\; \frac{P(A\cap S)}{P(S)} \tag{9}$$

$$\;=\; \frac{6\cdot \cancel{0.5^4}}{11\cdot \cancel{0.5^4}} \;=\; \underline{\underline{\frac{6}{11}}} \qquad (= 0.5454) \tag{10}$$

■

[2]Det er dette som er litt vanskelig med dette oppgaven, å innse at det er den betingede sannsynligheten som oppgaven spør om.

Oppgave 2: (stokastiske variabler og forventningverdi - telling av lakselus)

a) Tolkning:

$$\underline{\underline{P(X = j)}} \quad = \quad \text{sannsynligheten for at en tilfeldig valgt fisk} \tag{11}$$

$$\underline{\underline{\text{fra et oppdrettsanlegg med lusetall mellom 0.4 og 0.6 har } j \text{ antall lakselus}}}$$

b) Sannsynlighetene $P(X = j)$ med $j = 0, 1, 2, 3$ utgjør en gyldig sannsynlighetsfordeling fordi den er normert til 1:

$$\underline{\underline{\sum_{j=0}^{3} P(X = j)}} \quad = \quad P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \tag{12}$$

$$= \quad 0.72 + 0.13 + 0.08 + 0.07 \quad = \quad \underline{\underline{1}} \tag{13}$$

c) Forventningsverdien $E[X]$:

$$\underline{\underline{E[X]}} \quad = \quad \sum_{j=0}^{3} x_j \, P(X = j) \tag{14}$$

$$= \quad \sum_{j=0}^{3} j \, P(X = j) \tag{15}$$

$$= \quad 0 \cdot 0.72 \; + \; 1 \cdot 0.13 \; + \; 2 \cdot 0.08 \; + \; 3 \cdot 0.07 \tag{16}$$

$$= \quad \underline{\underline{0.5}} \tag{17}$$

d) Variansen $Var[X]$:

$$\underline{\underline{Var[X]}} = \sum_{j=0}^{3} \Big(x_j - E[X]\Big)^2 P(X = j) \tag{18}$$

$$= \sum_{j=0}^{3} \Big(j - E[X]\Big)^2 P(X = j) \tag{19}$$

$$= \Big(0 - 0.5\Big)^2 \cdot 0.72 \; + \; \Big(1 - 0.5\Big)^2 \cdot 0.12 + \Big(2 - 0.5\Big)^2 \cdot 0.08 \; + \; \Big(3 - 0.5\Big)^2 \cdot 0.07 \tag{20}$$

$$= \underline{\underline{0.83}} \tag{21}$$

med tilhørende standardavvik $\sigma[X_1]$:

$$\underline{\underline{\sigma[X]}} = \sqrt{Var[X]} \tag{22}$$

$$= \sqrt{0.83} = \underline{\underline{0.911}} \tag{23}$$

e) Sannsynligheten for at gjennomsnittet $\overline{X}_t$ ligger over lusegrensen:

$$P(\overline{X}_t > 0.5) \tag{24}$$

Siden $X_{kt} \overset{\text{i.i.d.}}{\sim} X \sim P(X=j)$ med 50 variabler, dvs. $k = 1, 2, \ldots, 50$, så er, ifølge sentralgrensesetningen, gjennomsnittet

$$\overline{X}_t \;=\; \frac{1}{n}\sum_{k=1}^{n} X_{kt} \tag{25}$$

normalfordelt:

$$\overline{X}_t \;\sim\; N\Big[\; E[\overline{X}_t] \;,\; \sigma[\overline{X}_t] \;\Big] \tag{26}$$

hvor

$$E[\,\overline{X}_t\,] \;=\; \overbrace{E[X] \;=\; 0.5}^{\text{oppg. } \mathbf{2c}} \tag{27}$$

$$\sigma[\,\overline{X}_t\,] \;=\; \overbrace{\frac{\sigma[X]}{\sqrt{n}} \;=\; \frac{0.911}{\sqrt{50}}}^{\text{oppg. } \mathbf{2d}} \tag{28}$$

Standardiserer:

$$\underline{P(\overline{X}_t > 0.5)} \;\overset{\text{standardiser}}{=}\; P\left(\frac{\overline{X}_t - E[\overline{X}_t]}{\sigma[\overline{X}_t]} \;>\; \frac{0.5 - E[\overline{X}_t]}{\sigma[\overline{X}_t]}\right) \tag{29}$$

$$=\; P\Bigg(\underbrace{\frac{\overline{X}_t - E[X]}{\sigma[X]/\sqrt{n}}}_{\equiv Z} \;>\; \frac{0.5 - E[X]}{\sigma[X]/\sqrt{n}}\Bigg) \tag{30}$$

$$=\; P\left(Z \;>\; \frac{0.5 - 0.5}{0.911/\sqrt{50}}\right) \tag{31}$$

$$=\; \underline{P\Big(Z \;>\; 0\Big)} \tag{32}$$

Siden ulikhetstegnet er "feil" vi sammenlignet med tabellen i formelsamlingen så må vi:

$$\underline{P(\overline{X}_t > 0.5)} = P\big(Z > 0\big) \tag{33}$$

$$= \underline{1 - P\big(Z < 0\big)} \tag{34}$$

Via tabelloppslag i formelsamlingen finner vi nå:

$$\underline{\underline{P(\overline{X}_t > 0.5)}} = 1 - \overbrace{P\big(Z < 0\big)}^{=\,0.5} \tag{35}$$

$$= 1 - 0.5 = \underline{\underline{0.5}} \tag{36}$$

f) Siden sannsynligheten for at $\overline{X}_t > 0.5$ er 50 %, så er det helt tilfeldig om gjennomsnittet av de 50 tilfeldig valgte fiskene en gitt uke t er innenfor lusegrensen eller ikke.

Kriteriet $\overline{X}_t > 0.5$ er derfor ikke et godt kriterium for å avgjøre om lusegresen er brutt eller ikke.

$\underline{\underline{\text{Nei}}}$, Mattilsynet bør derfor ikke konkludere om lusegrensen er brutt eller ikke basert på gjennomsnittet $\overline{X}_t > 0.5$.

■

Oppgave 3: (statistisk modell og estimatorer)

a) Siden $X \sim \text{Poi}[\lambda]$ så er:

$$E[X_i] \;=\; \lambda \tag{37}$$

for alle $i = 1, 2, ..., N$ fiskene. Forventningen av estimatoren $\hat{\lambda}$ blir da:

$$\begin{aligned}
\underline{\underline{E[\,\hat{\lambda}\,]}} \quad &= \quad E[\,\overline{X}\,] && (38)\\
&= \quad E\Big[\,\frac{1}{N}\sum_{i=1}^{N} X_i\,\Big] && (39)\\
&= \quad \frac{1}{N}\Big(\,\underbrace{E[X_1]}_{\lambda} \;+\; \underbrace{E[X_2]}_{\lambda} \;+\; \underbrace{E[X_3]}_{\lambda} \;+\; ... \;+\; \underbrace{E[X_N]}_{\lambda}\,\Big) && (40)\\
&= \quad \frac{\cancel{N}\,\lambda}{\cancel{N}} \quad = \quad \underline{\underline{\lambda}} && (41)
\end{aligned}$$

dvs. estimatoren $\hat{\lambda}$ er forventingsrett.

b) Siden vi skal regne ut et *asymptotisk* konfidensintervall, dvs. for tilfellet $N \to \infty$, for estimatoren λ gitt ved:

$$\hat{\lambda} \;=\; \overline{X} \;=\; \frac{1}{N}\sum_{i=1}^{N} X_i \tag{42}$$

må vi bruke sentralgrensesetningen til å konstruere konfidensintervallet.

Oppgitt i oppgaven:

$$X_1,\, X_2,\, X_3,\, \dots\, , X_N \;\overset{\text{i.i.d.}}{\sim}\; X \;\sim\; \text{Poi}[\,\lambda\,] \tag{43}$$

Også oppgitt i oppgaven:

$$E[\,\hat{\lambda}\,] \;=\; E[\,\overline{X}\,] \;=\; \lambda \quad \text{(se oppgave } \mathbf{3b}\text{)} \tag{44}$$

$$\sigma[\,\hat{\lambda}\,] \;=\; \sigma[\,\overline{X}\,] \;=\; \sqrt{\frac{\lambda}{N}} \tag{45}$$

Sentralgrensesetningen sier da at den skalerte variabelen Z er standard normalfordelt:

$$\underline{Z} \;=\; \frac{\overline{X} - E[\overline{X}]}{\sigma[\overline{X}]} \;=\; \underline{\frac{\overline{X} - \lambda}{\sqrt{\frac{\lambda}{N}}}} \;\overset{n\to\infty}{\sim}\; N[0,1] \tag{46}$$

Sentralgrensesetningen gir at

$$\lim_{N\to\infty} P\big(\, z_{\alpha/2} \,\le\, Z \le z_{1-\alpha/2}\,\big) \quad = \quad 1-\alpha \tag{47}$$

hvor (se figur 1)

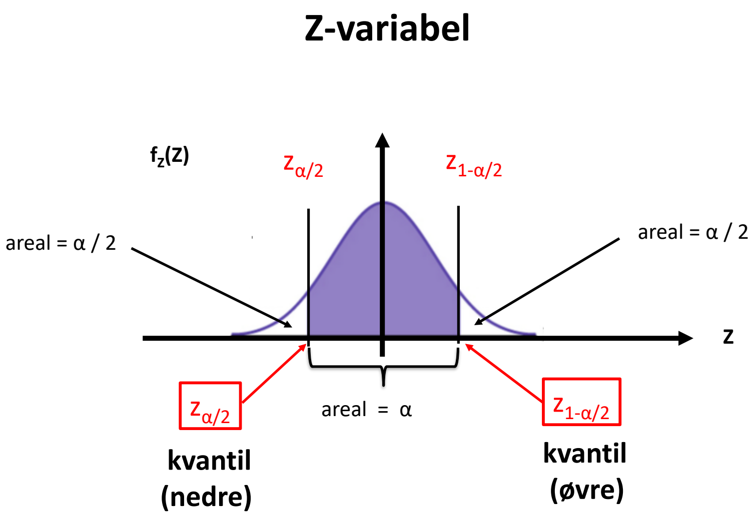
$$z_{\alpha/2} \quad = \quad \text{nedre kvantil til normalfordelingen} \tag{48}$$

$$z_{1-\alpha/2} \quad = \quad \text{øvre kvantil til normalfordelingen} \tag{49}$$

Vi må transformere lign.(47) over på formen

$$\boxed{\lim_{N\to\infty} P_\lambda\big(\, LB_N^\mu \,\le\, \lambda \le UB_N^\mu\,\big) \quad = \quad 1-\alpha \qquad (50)}$$

som er definisjonen av et $(1-\alpha)\,100\,\%$-konfidensintervall for λ.



Figur 1: Kvantil.

Setter inn for Z fra lign.(46) inn i lign.(47):

$$\lim_{N\to\infty} P\left(z_{\alpha/2} \leq \frac{\overline{X} - \lambda}{\sqrt{\frac{\lambda}{N}}} \leq z_{1-\alpha/2} \right) \;\; = \;\; 1 - \alpha \tag{51}$$

Deretter multipliserer vi uttrykket i $P(\cdots)$ med $\sqrt{\frac{\lambda}{N}}$:

$$\lim_{N\to\infty} P\left(z_{\alpha/2}\sqrt{\frac{\lambda}{N}} \leq \overline{X} - \lambda \leq z_{1-\alpha/2}\sqrt{\frac{\lambda}{N}} \right) \;\; = \;\; 1 - \alpha \tag{52}$$

Deretter multipliserer vi uttrykket i $P(\cdots)$ med -1 hvor vi må huske at retningen på ulikhettegnene snus:

$$\lim_{N\to\infty} P\left(- z_{\alpha/2}\sqrt{\frac{\lambda}{N}} \geq -\overline{X} + \lambda \geq -z_{1-\alpha/2}\sqrt{\frac{\lambda}{N}} \right) \;\; = \;\; 1 - \alpha \tag{53}$$

Deretter summerer vi uttrykket i $P(\cdots)$ med $\overline{X}$ for å sette λ alene i midten:

$$\lim_{N\to\infty} P\left(\overline{X} - z_{\alpha/2}\sqrt{\frac{\lambda}{N}} \geq \lambda \geq \overline{X} - z_{1-\alpha/2}\sqrt{\frac{\lambda}{N}} \right) \;\; = \;\; 1 - \alpha \tag{54}$$

Deretter snur vi høyre og venstre side i $P(\cdots)$ hvor vi må huske at retningen på ulikhettegnene snus:

$$\lim_{N\to\infty} P\left(\overline{X} - z_{1-\alpha/2}\sqrt{\frac{\lambda}{N}} \leq \lambda \leq \overline{X} - z_{\alpha/2}\sqrt{\frac{\lambda}{N}} \right) \;\; = \;\; 1 - \alpha \tag{55}$$

Deretter erstatter vi λ med $\overline{X}$ på høyre og venstre side i $P(\cdots)$, dvs. setter $\lambda = \overline{X}$:

$$\lim_{N\to\infty} P\left(\overline{X} - z_{1-\alpha/2}\sqrt{\frac{\overline{X}}{N}} \leq \lambda \leq \overline{X} - z_{\alpha/2}\sqrt{\frac{\overline{X}}{N}} \right) \;=\; 1-\alpha \tag{56}$$

Lign.(56) er et uttrykk på formen gitt i lign.(50) med

$$LB_N^{\lambda} \;=\; \overline{X} \;-\; z_{1-\alpha/2}\sqrt{\frac{\overline{X}}{N}} \tag{57}$$

$$UB_N^{\lambda} \;=\; \overline{X} \;-\; z_{\alpha/2}\sqrt{\frac{\overline{X}}{N}} \tag{58}$$

Vi har dermed vist at

$$\underline{\underline{\left[\, LB_N^{\lambda} \,,\, UB_N^{\lambda} \,\right] \;=\; \left[\overline{X} - z_{1-\alpha/2}\sqrt{\frac{\overline{X}}{N}} \,,\, \overline{X} - z_{\alpha/2}\sqrt{\frac{\overline{X}}{N}} \right]}} \quad , \quad \text{q.e.d.} \tag{59}$$

er et $(1-\alpha)\,100\,\%$ asymptotisk konfidensintervall for λ.

c) Fra tabell 1 i oppgaven har vi at:

$$\overline{x} \;=\; 0.499 \quad , \quad s_x^2 \;=\; 0.894 \tag{60}$$

altså

$$\overline{x} \;<\; s_x^2 \tag{61}$$

dvs. det er en tendens til <u>overspredning</u> i datanene.

d) I oppgaven er det oppgitt at:

$$E[X] \;=\; \lambda \tag{62}$$

$$Var[X] \;=\; \lambda + \frac{\lambda^2}{\theta} \tag{63}$$

Da innser vi at:

$$E[X] \;<\; Var[X] \qquad \text{(overspredning)} \tag{64}$$

siden $\lambda > 0$ og $\theta > 0$, dvs. positive. Derfor er negativ binomial fordelingen en <u>passende modell</u> dersom dataene har overspredning.

■

Oppgave 4: (krav som indikerer overskridelse av lusegrensa)

a) Kvantilen $b_{0.96}$ til $\overline{X}_t$ er definert ved:

$$P(\ \overline{X}_t > b_{0.96}\ ) \ = \ 0.04 \tag{65}$$

Siden (lign.(66) er oppgitt i oppgaven)

$$X_{1t}, X_{2t}, \ldots, X_{nt} \overset{i.i.d.}{\sim} X \ \sim \ \text{Poi}[\lambda] \tag{66}$$

hvor $n = 50$, dvs. variablene er Poisson-fordelt, så er:

$$E[\,\overline{X}_t\,] \ = \ \lambda \tag{67}$$

$$\sigma[\,\overline{X}_t\,] \ = \ \sqrt{\frac{\lambda}{n}} \tag{68}$$

Sentralgrensesetningen sier da at den skalerte variabelen Z er tilnærmet standard normalfordelt fordi $n = 50$: [3]

$$\underline{Z} \ = \ \frac{\overline{X} - E[\overline{X}_t]}{\sigma[\overline{X}_t]} \ = \ \underline{\frac{\overline{X}_t - \lambda}{\sqrt{\frac{\lambda}{n}}}} \ \overset{n\to\infty}{\sim} \ N[0,1] \tag{69}$$

[3]En *tommelfingerregel* er at det må være $\gtrsim 30$ forsøk for at sentralgrensesetningen skal gjelde.

Standardierer lign.(65):

$$P\left(\underbrace{\frac{\overline{X}_t - \lambda}{\sqrt{\frac{\lambda}{n}}}}_{=\,Z} > \underbrace{\frac{b_{0.96} - \lambda}{\sqrt{\frac{\lambda}{n}}}}_{=\,z_{0.96}} \right) = 0.04 \tag{70}$$

hvor

$$z_{0.96} = \frac{b_{0.96} - \lambda}{\sqrt{\frac{\lambda}{n}}} \tag{71}$$

er 96 %-kvantilen til den standardiserte normalfordelingen.
Løser lign.(71) mhp. $b_{0.96}$:

$$\underline{b_{0.96} = \lambda + z_{0.96}\sqrt{\frac{\lambda}{n}}} \tag{72}$$

Tabelloppslag:

$$z_{0.96} = 1.75 \tag{73}$$

Numerisk verdi:

$$\underline{b_{0.96}} = \lambda + z_{0.96}\sqrt{\frac{\lambda}{n}} = 0.49 + 1.75 \cdot \sqrt{\frac{0.49}{50}} = \underline{0.663} \tag{74}$$

Siden

$$b_{0.96} = 0.663 \quad < \quad \overline{x}_t = 0.68 \tag{75}$$

konkluderer vi med at <u>kravet er overskridet</u>,
dvs. laksegrensa er overskredet ihht. krav 1.

b) Krav 2:

$$P(B) \;=\; P(\,\overline{X}_t > a \;\; \textbf{og} \;\; \overline{X}_{t+1} > a \;\; \textbf{og} \;\; \overline{X}_{t+2} > a \;\; \textbf{og} \; \overline{X}_{t+3} > a) \tag{76}$$

Siden lusetellingene er uavhengige av hverandre fra uke til uke, har vi at:

$$\begin{aligned}
\underline{\underline{P(B)}} \quad &= \quad P\Big(\,\overline{X}_t > a \;\; \textbf{og} \;\; \overline{X}_{t+1} > a \;\; \textbf{og} \;\; \overline{X}_{t+2} > a \;\; \textbf{og} \;\; \overline{X}_{t+3} > a\,\Big) && (77) \\
&= \quad P(\,\overline{X}_t > a\,)\,P(\,\overline{X}_{t+1} > a\,)P(\,\overline{X}_{t+2} > a\,)\,P(\,\overline{X}_{t+3} > a\,) && (78) \\
&= \quad 0.50^4 && (79) \\
&= \quad \underline{\underline{0.0625}} && (80)
\end{aligned}$$

■

Kapittel 17

Kontinuasjonseksamen 2019

LØSNING: Eksamen 9. sep. 2019

"*MAT110 Statistikk 1*", vår 2019

Versjon 01

Oppgave 1: (telling av lakselus)

a) Utfallsrommet Ω har et endelig antall utfall.
Derfor er utfallsrommet diskret.

b) Hvert forsøk har 2 mulige utfall. Et eksperiment består av en sekvens med n forsøk.
Utfallsrommet Ω har dermed 2^n mulige utfall. [1]

c) Eksperimentet:

$$l_1 \ , \ \overline{l_2} \ , \ l_3 \tag{1}$$

betyr at:

- den tilfeldig valgte fisken fra merd 1 har lakselus
- den tilfeldig valgte fisken fra merd 2 har ikke lakselus
- den tilfeldig valgte fisken fra merd 3 har lakselus

[1]Med $n = 3$ får vi $2^3 = 8$, som stemmer med Ω oppgitt i oppgaven.

Siden eksperimentene er uavhengige, så er:

$$\underline{\underline{P(l_1, \overline{l_2}, l_3)}} = P(l_1)P(\overline{l_2})P(l_3) \tag{2}$$

$$= p \cdot (1-p) \cdot p = 0.3^2 \cdot (1-0.3) \cdot 0.3 = \underline{\underline{0.063}} \tag{3}$$

siden $p = 0.3$ (oppgitt i oppgaven).

d) Ved å se på ligningen for begivenheten S som oppgitt i ligningen så innser vi at:

$$\underline{\underline{P(S)}} = P(l_1, l_2, l_3) + P(\overline{l_1}, l_2, l_3) + P(l_1, \overline{l_2}, l_3) + P(l_1, l_2, \overline{l_3}) \tag{4}$$

$$= P(l_1)P(l_2)P(l_3) + P(\overline{l_1})P(l_2)P(l_3) + P(l_1)P(\overline{l_2})P(l_3) + P(l_1)P(l_2)P(\overline{l_3})$$

$$= p^3 + (1-p)^1 p^2 + p(1-p)^1 p + p^2(1-p)^1 \tag{5}$$

$$= \underline{\underline{p^3 + 3p^2(1-p)^1}} \tag{6}$$

e) Begivenheten $A \cap S$: [2]

$$A \cap S = \left\{ (\overline{l_1}, l_2, l_3) \ , \ (l_1, \overline{l_2}, l_3) \ , \ (l_1, l_2, \overline{l_3}) \right\}$$

[2] $A \cap S$ består av utfallene i S, se lign.(9) i oppgaveteksten, hvor kun én av de tilfeldig valgte fiskene ikke har lakselus.

f) Oppgaven ber oss da finne den betingede sannsynligheten: [3]

$$\boxed{P(A|S) \;=\; \frac{P(A\cap S)}{P(S)} \qquad (7)}$$

hvor

$$P(A\cap S) \;=\; 3p^2(1-p) \;=\; 0.189 \qquad \text{(oppgitt i oppgaven)} \tag{8}$$

Fra oppgave **1d** kjenner vi $P(S)$, se lign.(6).
Med $p = 0.3$:

$$P(S) \;=\; p^3 \;+\; 3p^2(1-p)^1 \;=\; 0.216 \tag{9}$$

Den betingede sannsynligheten i lign.(7) blir da:

$$\underline{\underline{P(A|S)}} \;=\; \frac{P(A\cap S)}{P(S)} \tag{10}$$

$$\;=\; \frac{0.189}{0.216} \;=\; \underline{\underline{0.875}} \tag{11}$$

■

[3]Det er dette som er litt vanskelig med dette oppgaven, å innse at det er den betingede sannsynligheten som oppgaven spør om.

Oppgave 2: (stokastiske variabler og forventningverdi - telling av lakselus)

a) Tolkning:

$$\underline{\underline{P(X \geq j)}} \quad = \quad \text{sannsynligheten for at en tilfeldig valgt fisk} \tag{12}$$

$$\underline{\underline{\text{fra et oppdrettsanlegg med lusetall mellom 0.01 og 0.4 har } j \text{ antall lakselus eller mer}}}$$

b) Sannsynlighetene $P(X = j)$ med $j = 0, 1, 2, 3$ utgjør en gyldig sannsynlighetsfordeling fordi den er normert til 1:

$$\underline{\underline{\sum_{j=0}^{3} P(X = j)}} \quad = \quad P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \tag{13}$$

$$= \quad 0.88 + 0.07 + 0.04 + 0.01 \quad = \quad \underline{\underline{1}} \tag{14}$$

c) Forventningsverdien $E[X]$: [4]

$$\underline{\underline{E[X]}} \quad = \quad \sum_{j=0}^{3} x_j \, P(X = j) \tag{15}$$

$$= \quad \sum_{j=0}^{3} j \, P(X = j) \tag{16}$$

$$= \quad 0 \cdot 0.88 \; + \; 1 \cdot 0.07 \; + \; 2 \cdot 0.04 \; + \; 3 \cdot 0.01 \tag{17}$$

$$= \quad \underline{\underline{0.18}} \tag{18}$$

[4]Med X, dvs. antall hunnlus (se definisjon i oppgaven), så er: $x_j = j$.

d) Variansen $Var[X]$:

$$\underline{\underline{Var[X]}} = \sum_{j=0}^{3} \Big(x_j - E[X]\Big)^2 P(X = j) \tag{19}$$

$$= \sum_{j=0}^{3} \Big(j - E[X]\Big)^2 P(X = j) \tag{20}$$

$$= \Big(0 - 0.18\Big)^2 \cdot 0.88 \ + \ \Big(1 - 0.18\Big)^2 \cdot 0.07$$

$$+ \Big(2 - 0.18\Big)^2 \cdot 0.04 \ + \ \Big(3 - 0.18\Big)^2 \cdot 0.01 \tag{21}$$

$$= \underline{\underline{0.2876}} \tag{22}$$

med tilhørende standardavvik $\sigma[X_1]$:

$$\underline{\underline{\sigma[X]}} = \sqrt{Var[X]} \tag{23}$$

$$= \sqrt{0.2876} = \underline{\underline{0.5363}} \tag{24}$$

e) Sannsynligheten for at gjennomsnittet $\overline{X}_t$ ligger under lusegrensen:

$$P(\overline{X}_t \leq a) \tag{25}$$

Siden $X_{kt} \overset{\text{i.i.d.}}{\sim} X \sim P(X = j)$ med 50 variabler, dvs. $k = 1, 2, \ldots, 50$, så er, ifølge sentralgrensesetningen, gjennomsnittet

$$\overline{X}_t \;=\; \frac{1}{n}\sum_{k=1}^{n} X_{kt} \tag{26}$$

normalfordelt:

$$\overline{X}_t \;\sim\; N\Big[\; E[\overline{X}_t] \;,\; \sigma[\overline{X}_t] \;\Big] \tag{27}$$

hvor

$$\mu \;=\; E[\,\overline{X}_t\,] \;=\; \overbrace{E[X] \;=\; 0.18}^{\text{oppg. } \mathbf{2c}} \tag{28}$$

$$\sigma \;=\; \sigma[\,\overline{X}_t\,] \;=\; \overbrace{\frac{\sigma[X]}{\sqrt{n}} \;=\; \frac{0.5363}{\sqrt{50}}}^{\text{oppg. } \mathbf{2d}} \tag{29}$$

Standardiserer:

$$\underline{P(\overline{X}_t \leq 0.2)} \;\overset{\text{standardiser}}{=}\; P\Bigg(\underbrace{\frac{\overline{X}_t - E[\overline{X}_t]}{\sigma[\overline{X}_t]}}_{\equiv Z} \;\leq\; \frac{0.2 - \mu}{\sigma} \Bigg) \tag{30}$$

$$=\; P\Bigg(Z \;\leq\; \frac{0.2 - 0.18}{0.5363/\sqrt{50}} \Bigg) \tag{31}$$

$$=\; \underline{P\Big(Z \;\leq\; 0.26 \Big)} \tag{32}$$

Via tabelloppslag i formelsamlingen finner vi nå:

$$\underline{\underline{P(\overline{X}_t \leq 0.2)}} = P\Big(Z \leq 0.26\Big) \tag{33}$$

$$= \underline{\underline{0.6026}} \tag{34}$$

f) Siden det "bare" er 60.26 % sannsynlighet for at det målte lusetallet $\overline{X}_t$ til oppdrettsanlegget i uke t er under lusegrensa så er et relativt tilfeldig om et oppdrettsanlegg er under lusegrensa eller ikke. [5]

Nei, basert på det målte gjennomsnittet $\overline{X}_t$ bør ikke Mattilsynet konkludere med at lusetallet til oppdrettsanlegget ligger under lusegrensa.

■

[5] Helt tilfeldig er det dersom det hadde vært 50 %.

Oppgave 3: (statistisk modell og estimatorer)

a) Siden $X \sim \text{Poi}[\lambda]$ så er:

$$E[X_i] \;=\; \lambda \tag{35}$$

for alle $i = 1, 2, ..., N$ fiskene. Forventningen av estimatoren $\hat{\lambda}$ blir da:

$$\begin{aligned}
\underline{\underline{E[\,\hat{\lambda}\,]}} \quad &= \quad E[\,\overline{X}\,] && (36)\\
&= \quad E\Big[\,\frac{1}{N}\sum_{i=1}^{N} X_i\,\Big] && (37)\\
&= \quad \frac{1}{N}\Big(\,\underbrace{E[X_1]}_{\lambda} + \underbrace{E[X_2]}_{\lambda} + \underbrace{E[X_3]}_{\lambda} + ... + \underbrace{E[X_N]}_{\lambda}\,\Big) && (38)\\
&= \quad \frac{\cancel{N}\lambda}{\cancel{N}} \;=\; \underline{\underline{\lambda}} && (39)
\end{aligned}$$

dvs. estimatoren $\hat{\lambda}$ er forventingsrett.

b) Siden vi skal regne ut et *asymptotisk* konfidensintervall, dvs. for tilfellet $N \to \infty$, for estimatoren λ gitt ved:

$$\hat{\lambda} \;=\; \overline{X} \;=\; \frac{1}{N}\sum_{i=1}^{N} X_i \tag{40}$$

må vi bruke sentralgrensesetningen til å konstruere konfidensintervallet.

Oppgitt i oppgaven:

$$X_1,\, X_2,\, X_3,\, \ldots\, , X_N \;\stackrel{\text{i.i.d.}}{\sim}\; X \;\sim\; \text{Poi}[\,\lambda\,] \tag{41}$$

Også oppgitt i oppgaven:

$$E[\,\hat{\lambda}\,] \;=\; E[\,\overline{X}\,] \;=\; \lambda \quad (\text{se oppgave } \mathbf{3b}) \tag{42}$$

$$\sigma[\,\hat{\lambda}\,] \;=\; \sigma[\,\overline{X}\,] \;=\; \sqrt{\frac{\lambda}{N}} \tag{43}$$

Sentralgrensesetningen sier da at den skalerte variabelen Z er standard normalfordelt:

$$\underline{Z} \;=\; \frac{\overline{X} - E[\overline{X}]}{\sigma[\overline{X}]} \;=\; \underline{\frac{\overline{X} - \lambda}{\sqrt{\frac{\lambda}{N}}}} \;\stackrel{n\to\infty}{\sim}\; N[0,1] \tag{44}$$

Sentralgrensesetningen gir at

$$\lim_{N\to\infty} P\big(z_{\alpha/2} \le Z \le z_{1-\alpha/2} \big) \;=\; 1-\alpha \tag{45}$$

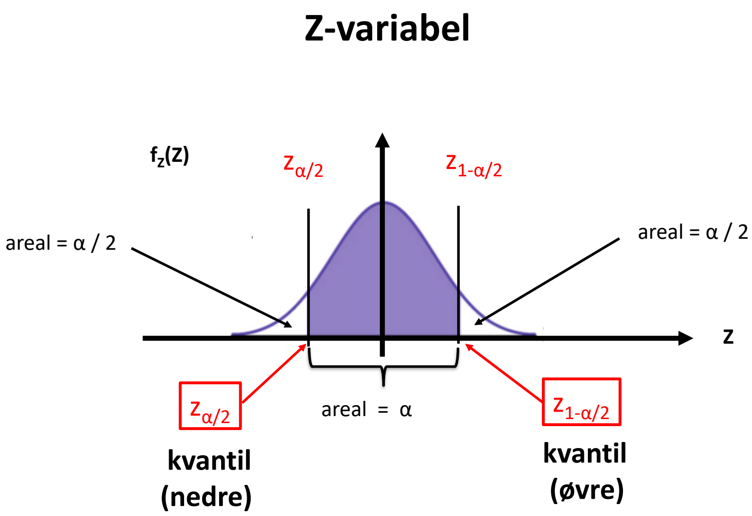
hvor (se figur 1)

$$z_{\alpha/2} \;=\; \text{nedre kvantil til normalfordelingen} \tag{46}$$
$$z_{1-\alpha/2} \;=\; \text{øvre kvantil til normalfordelingen} \tag{47}$$

Vi må transformere lign.(45) over på formen

$$\boxed{\lim_{N\to\infty} P_\lambda\big(LB_N^\mu \le \lambda \le UB_N^\mu \big) \;=\; 1-\alpha \qquad (48)}$$

som er definisjonen av et $(1-\alpha)\,100\,\%$-konfidensintervall for λ.



Figur 1: Kvantil.

Setter inn for Z fra lign.(44) inn i lign.(45):

$$\lim_{N\to\infty} P\left(z_{\alpha/2} \leq \frac{\overline{X} - \lambda}{\sqrt{\frac{\lambda}{N}}} \leq z_{1-\alpha/2}\right) \quad = \quad 1 - \alpha \tag{49}$$

Deretter multipliserer vi uttrykket i $P(\cdots)$ med $\sqrt{\frac{\lambda}{N}}$:

$$\lim_{N\to\infty} P\left(z_{\alpha/2}\sqrt{\frac{\lambda}{N}} \leq \overline{X} - \lambda \leq z_{1-\alpha/2}\sqrt{\frac{\lambda}{N}}\right) \quad = \quad 1 - \alpha \tag{50}$$

Deretter multipliserer vi uttrykket i $P(\cdots)$ med -1 hvor vi må huske at retningen på ulikhetstegnene snus:

$$\lim_{N\to\infty} P\left(- z_{\alpha/2}\sqrt{\frac{\lambda}{N}} \geq -\overline{X} + \lambda \geq -z_{1-\alpha/2}\sqrt{\frac{\lambda}{N}}\right) \quad = \quad 1 - \alpha \tag{51}$$

Deretter summerer vi uttrykket i $P(\cdots)$ med $\overline{X}$ for å sette λ alene i midten:

$$\lim_{N\to\infty} P\left(\overline{X} - z_{\alpha/2}\sqrt{\frac{\lambda}{N}} \geq \lambda \geq \overline{X} - z_{1-\alpha/2}\sqrt{\frac{\lambda}{N}}\right) \quad = \quad 1 - \alpha \tag{52}$$

Deretter snur vi høyre og venstre side i $P(\cdots)$ hvor vi må huske at retningen på ulikhettegnene snus:

$$\lim_{N\to\infty} P\left(\overline{X} - z_{1-\alpha/2}\sqrt{\frac{\lambda}{N}} \leq \lambda \leq \overline{X} - z_{\alpha/2}\sqrt{\frac{\lambda}{N}}\right) \quad = \quad 1 - \alpha \tag{53}$$

Deretter erstatter vi λ med $\overline{X}$ på høyre og venstre side i $P(\cdots)$, dvs. setter $\lambda = \overline{X}$:

$$\lim_{N\to\infty} P\left(\overline{X} - z_{1-\alpha/2}\sqrt{\frac{\overline{X}}{N}} \leq \lambda \leq \overline{X} - z_{\alpha/2}\sqrt{\frac{\overline{X}}{N}} \right) \;=\; 1-\alpha \tag{54}$$

Lign.(54) er et uttrykk på formen gitt i lign.(48) med

$$LB_N^{\lambda} \;=\; \overline{X} \;-\; z_{1-\alpha/2}\sqrt{\frac{\overline{X}}{N}} \tag{55}$$

$$UB_N^{\lambda} \;=\; \overline{X} \;-\; z_{\alpha/2}\sqrt{\frac{\overline{X}}{N}} \tag{56}$$

Vi har dermed vist at

$$\underline{\underline{[\; LB_N^{\lambda} \;,\; UB_N^{\lambda} \;] \;=\; \left[\overline{X} - z_{1-\alpha/2}\sqrt{\frac{\overline{X}}{N}} \;,\; \overline{X} - z_{\alpha/2}\sqrt{\frac{\overline{X}}{N}} \right]}} \quad , \quad \text{q.e.d.} \tag{57}$$

er et $(1-\alpha)\,100\,\%$ asymptotisk konfidensintervall for λ.

c) Fra tabell 1 i oppgaven har vi at:

$$\overline{x} \ = \ 0.097 \quad , \quad s_x^2 \ = \ 0.126 \tag{58}$$

altså

$$\overline{x} \ < \ s_x^2 \tag{59}$$

dvs. det er en tendens til overspredning i datanene.

d) I oppgaven er det oppgitt at:

$$E[X] \ = \ \lambda \tag{60}$$

$$Var[X] \ = \ \lambda + \frac{\lambda^2}{\theta} \tag{61}$$

Da innser vi at:

$$E[X] \ < \ Var[X] \qquad \text{(overspredning)} \tag{62}$$

siden $\lambda > 0$ og $\theta > 0$, dvs. positive. Derfor er negativ binomial fordelingen en passende modell dersom dataene har overspredning.

■

Oppgave 4: (krav som indikerer overskridelse av lusegrensa)

a) Kvantilen $b_{0.95}$ til $\overline{X}_t$ er definert ved:

$$P(\ \overline{X}_t > b_{0.95}\ ) \ = \ 0.05 \tag{63}$$

Siden (lign.(64) er oppgitt i oppgaven)

$$X_{1t}, X_{2t}, \ldots, X_{nt} \ \overset{i.i.d.}{\sim} \ X \ \sim \ \text{Poi}[\lambda] \tag{64}$$

hvor $n = 50$, dvs. variablene er Poisson-fordelt, så er:

$$E[\,\overline{X}_t\,] \ = \ \lambda \tag{65}$$

$$\sigma[\,\overline{X}_t\,] \ = \ \sqrt{\frac{\lambda}{n}} \tag{66}$$

Sentralgrensesetningen sier da at den skalerte variabelen Z
er tilnærmet standard normalfordelt fordi $n = 50$: [6]

$$\underline{Z} \ = \ \frac{\overline{X} - E[\overline{X}_t]}{\sigma[\overline{X}_t]} \ = \ \underline{\frac{\overline{X}_t - \lambda}{\sqrt{\frac{\lambda}{n}}}} \ \overset{n\to\infty}{\sim} \ N[0,1] \tag{67}$$

[6] En *tommelfingerregel* er at det må være $\gtrsim 30$ forsøk for at sentralgrensesetningen skal gjelde.

Standardierer lign.(63):

$$P\Bigg(\underbrace{\frac{\overline{X}_t - \lambda}{\sqrt{\frac{\lambda}{n}}}}_{= Z} > \underbrace{\frac{b_{0.95} - \lambda}{\sqrt{\frac{\lambda}{n}}}}_{= z_{0.95}} \Bigg) = 0.05 \tag{68}$$

hvor

$$z_{0.95} = \frac{b_{0.95} - \lambda}{\sqrt{\frac{\lambda}{n}}} \tag{69}$$

er 95 %-kvantilen til den standardiserte normalfordelingen.
Løser lign.(69) mhp. $b_{0.95}$:

$$\underline{b_{0.95} = \lambda + z_{0.95}\sqrt{\frac{\lambda}{n}}} \tag{70}$$

Omvendt tabelloppslag:

$$z_{0.95} = 1.65 \tag{71}$$

Numerisk verdi:

$$\underline{b_{0.95}} = \lambda + z_{0.95}\sqrt{\frac{\lambda}{n}} = 0.17 + 1.65 \cdot \sqrt{\frac{0.17}{50}} = \underline{0.2662} \tag{72}$$

Siden

$$b_{0.95} = 0.2662 \quad > \quad \overline{x}_t = 0.25 \tag{73}$$

konkluderer vi med at kravet er $\underline{\underline{\text{ikke overskridet}}}$,
dvs. laksegrensa er ikke overskredet ihht. krav 1.

b) Krav 2:

$$P(B) = P(\,\overline{X}_t > a \;\; \textbf{og} \;\; \overline{X}_{t+1} > a \;\; \textbf{og} \;\; \overline{X}_{t+2} > a \;\; \textbf{og}\; \overline{X}_{t+3} > a) \tag{74}$$

Siden lusetellingene er uavhengige av hverandre fra uke til uke, har vi at:

$$\begin{aligned}
\underline{\underline{P(B)}} &= P\Big(\,\overline{X}_t > a \;\; \textbf{og} \;\; \overline{X}_{t+1} > a \;\; \textbf{og} \;\; \overline{X}_{t+2} > a \;\; \textbf{og} \;\; \overline{X}_{t+3} > a\,\Big) && (75)\\
&= P(\,\overline{X}_t > a\,)\,P(\,\overline{X}_{t+1} > a\,)P(\,\overline{X}_{t+2} > a\,)\,P(\,\overline{X}_{t+3} > a\,) && (76)\\
&= 0.50^4 && (77)\\
&= \underline{\underline{0.0625}} && (78)
\end{aligned}$$

■