



V-2017

MAT110

Statistikk 1

Løsningsforslag til
eksamensoppgaver

2012 - 2016

Per Kristian Rekdal



Høgskolen i Molde
Vitenskapelig høgskole i logistikk

Innhold

1	LØSNING: Eksamen 1. juni 2012	7
2	LØSNING: Eksamen 10. januar 2013	23
3	LØSNING: Eksamen 30. mai 2013	41
4	LØSNING: Eksamen 6. januar 2014	61
5	LØSNING: Eksamen 9. mai 2014	73
6	LØSNING: Eksamen 8. januar 2015	89
7	LØSNING: Eksamen 28. mai 2015	105
8	LØSNING: Eksamen 8. januar 2016	125
9	LØSNING: Eksamen 27. mai 2016	143
10	LØSNING: Eksamen 5. januar 2017	159

Forord

Løsningsforslag:

Dette er en [samling av løsningsforslag](#) til gamle eksamensoppgaver i emnet “*MAT110 Statistikk I*” ved Høgskolen i Molde. Samlingen inneholder løsningsforslag tilhørende totalt 10 eksamensoppgaver, i perioden fra og med 2012 til og med 2016.

Det finnes også en tilhørende samling med selve eksamensoppgavene til disse løsningsforslagene. Samlingen med eksamensoppgaver finnes i et eget hefte, separert fra dette løsningsheftet.

Gratis:

Både samlingen med oppgaver og tilhørende samling med komplette løsningsforslag kan lastes ned [gratis](#) via Høgskolen i Molde sin åpne kursportal www.himoldeX.no.

Hvordan bruke denne samlingen av tidligere eksamensoppgaver med løsningsforslag?:

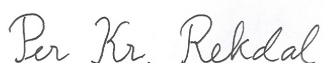
Det anbefales å [regne gjennom](#) gamle eksamensoppgaver før eksamen. Dersom man gjør det så får man en god pekepinn på hva som kreves på eksamensdagen. [Sett av 4 timer](#), prøv så godt du kan uten løsningsforslag. Etter at de 4 timene er over, rett din egen eksamensbesvarelse. Og sett gjerne karakter på deg selv.

Ikke bare i eksamensperioden, men også ellers i semesteret kan det være lurt å regne gjennom gamle eksamensoppgaver. Men gå gjennom teorien før man gjør oppgaver. Da får man bedre utbytte av oppgaveløsningen.

Videoer:

Komplette sett med forelesningsvideoer fra 2013, 2014, 2015 og 2016 finnes på www.himoldeX.no.

Per Kristian Rekdal



Copyright © Høgskolen i Molde, januar 2017.

Kapittel 1

LØSNING: Eksamen 1. juni 2012

“MAT110 Statistikk 1”

Oppgave 1: (sannsynlighetsregning)

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

b) Den spesielle addisjonssetningen , Den spesielle multiplikasjonssetningen

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

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

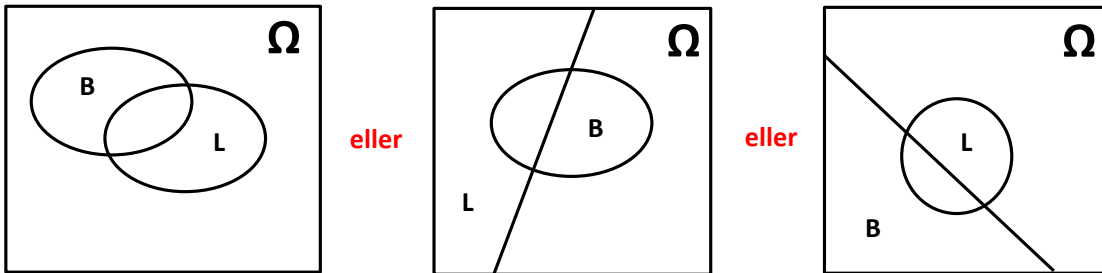
■

Oppgave 2: (økonomi)

a) i) Ut fra opplysningene i oppgaveteksten ser vi umiddelbart at: $P(B) = 0.20$

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

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



Figur 1.1: Venn-diagram.

c) Vi skal finne $P(B \cap L)$. Da kan vi bruke multiplikasjonsetningen:

$$\underline{\underline{P(B \cap L)}} = P(L|B) \cdot P(B) = 0.75 \cdot 0.20 = \underline{\underline{0.15}} \quad (1.1)$$

d) Vi skal finne $P(L)$. Oppsplitting av utfallsrom Ω :

$$\underline{\underline{P(L)}} = P(L|B) \cdot P(B) + P(L|\bar{B}) \cdot P(\bar{B}) \quad (1.2)$$

$$= 0.75 \cdot 0.20 + 0.30 \cdot 0.80 = \underline{\underline{0.39}} \quad (1.3)$$

e) Vi skal vise $P(B|L) = 0.38$. Det kan vises via multiplikasjonsetningen:

$$\underline{\underline{P(B|L)}} = \frac{P(B \cap L)}{P(L)} = \frac{0.15}{0.39} = \underline{\underline{0.38}} \quad , \quad \text{q.e.d.} \quad (1.4)$$

Alternativt så kan Bayes lov brukes:

$$\underline{\underline{P(B|L)}} = P(L|B) \cdot \frac{P(B)}{P(L)} = 0.75 \cdot \frac{0.20}{0.39} = \underline{\underline{0.38}} \quad , \quad \text{q.e.d.} \quad (1.5)$$

f) Vi skal finne $P(\bar{B}|L)$. Komplementsetningen gir:

$$\underline{\underline{P(\bar{B}|L)}} = 1 - P(B|L) = 1 - 0.38 = \underline{\underline{0.62}} \quad (1.6)$$

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

$$\underline{\underline{E[X]}} = \sum_{n=1}^2 x_i P(X = x_i) \quad (1.7)$$

$$= x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) \quad (1.8)$$

$$= \left(-10\,000 \cdot 0.38 + 8\,000 \cdot 0.62 \right) \text{NOK} = \underline{\underline{1\,160 \text{ NOK}}} \quad (1.9)$$

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

■

Oppgave 3: (logistikk)

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

1. For hver passasjer 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 passasjerer 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 $X \sim \text{Bin}[n, p]$:

$$\underline{\underline{E[X]}} = n \cdot p = 120 \cdot 0.95 = \underline{\underline{114}} \quad (1.10)$$

ii) Tolkning:

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

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

$$\underline{\underline{Var[X]}} = n \cdot p(1 - p) = 120 \cdot 0.95 \cdot (1 - 0.95) = \underline{\underline{5.7}} \quad (1.11)$$

ii) Tolkning:

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

d) Forventet inntekt $E[I]$ for SAS når $n = 120$:

$$\underline{\underline{E[I]}} = E[a \cdot X] = a \cdot \underbrace{E[X]}_{=n \cdot p} = 800 \cdot n \cdot p \text{ NOK} \quad (1.12)$$

$$= 800 \cdot 120 \cdot 0.95 \text{ NOK} = \underline{\underline{91\,200 \text{ NOK}}} \quad (1.13)$$

e) Forventet inntekt $E[I]$ for SAS når $n = 123$:

$$\underline{\underline{E[I]}} = E[a \cdot X] = a \cdot \underbrace{E[X]}_{=n \cdot p} = 800 \cdot n \cdot p \text{ NOK} \quad (1.14)$$

$$= 800 \cdot 123 \cdot 0.95 \text{ NOK} = \underline{\underline{93\,480 \text{ NOK}}} \quad (1.15)$$

f) Sannsynligheten for at det møter opp flere passasjerer enn flyet har kapasitet til:

$$\underline{P(X \geq 121)} = P(X = 121) + P(X = 122) + P(X = 123) \quad (1.16)$$

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

$$= \binom{123}{121} 0.95^{121} (1-0.95)^{n-121} + \binom{123}{122} p^{122} (1-0.95)^{n-122} \quad (1.17)$$

$$+ \binom{123}{123} 0.95^{123} (1-0.95)^{123-123} \quad (1.18)$$

$$= 0.0378 + 0.0118 + 0.0018 \quad (1.19)$$

$$= \underline{0.0514} \quad (\text{eksakt svar med 4 desimales nøyaktighet}) \quad (1.20)$$

Kommentar:

Det er meningen at denne oppgaven skal løses på måten som vist i lign.(1.16). Det er en eksakt løsning. Man trenger nemlig mellomregningene, dvs. lign.(1.19), 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)

$$n \cdot p(1-p) = 5.8 \gtrsim 5 \quad (1.21)$$

så er X *tilnærmet* en normalfordeling, $X \sim N[E[X], \sigma[X]]$, 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:

$$\underline{P(x \geq 121)} = 1 - P(Z \leq 1.72) = 1 - 0.9573 = \underline{0.0427} \quad (\text{tilnærmet svar}) \quad (1.22)$$

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

g) Forventet utgift $E[U]$ til SAS ved en slik “overbooking”, dvs. når $n = 123$:

$$\underline{\underline{E[U]}} = E[b \cdot Y] = b \cdot E[Y] \quad (1.23)$$

$$= b \sum_{y=1}^3 y P(Y = y) \quad (1.24)$$

$$= b \cdot \left(1 \cdot P(Y = 1) + 2 \cdot P(Y = 2) + 3 \cdot P(Y = 3) \right) \quad (1.25)$$

$$= b \cdot \left(\underbrace{1 \cdot P(X = 121)}_{=0.0378} + 2 \cdot \underbrace{P(X = 122)}_{=0.0118} + 3 \cdot \underbrace{P(X = 123)}_{=0.0018} \right) \quad (1.26)$$

numeriske resultat hentet fra oppgave 3f, lign.(1.19)

$$= 5000 \cdot \left(1 \cdot 0.0378 + 2 \cdot 0.0118 + 3 \cdot 0.0018 \right) \text{ NOK} \quad (1.27)$$

$$= \underline{\underline{334 \text{ NOK}}} \quad (1.28)$$

h) Siden forventet billettinntekter ved overbooking er

$$\underline{\underline{E[I] - E[U]}} = \left(93\,480 - 334 \right) \text{ NOK} = \underline{\underline{93\,146 \text{ NOK}}} \quad (1.29)$$

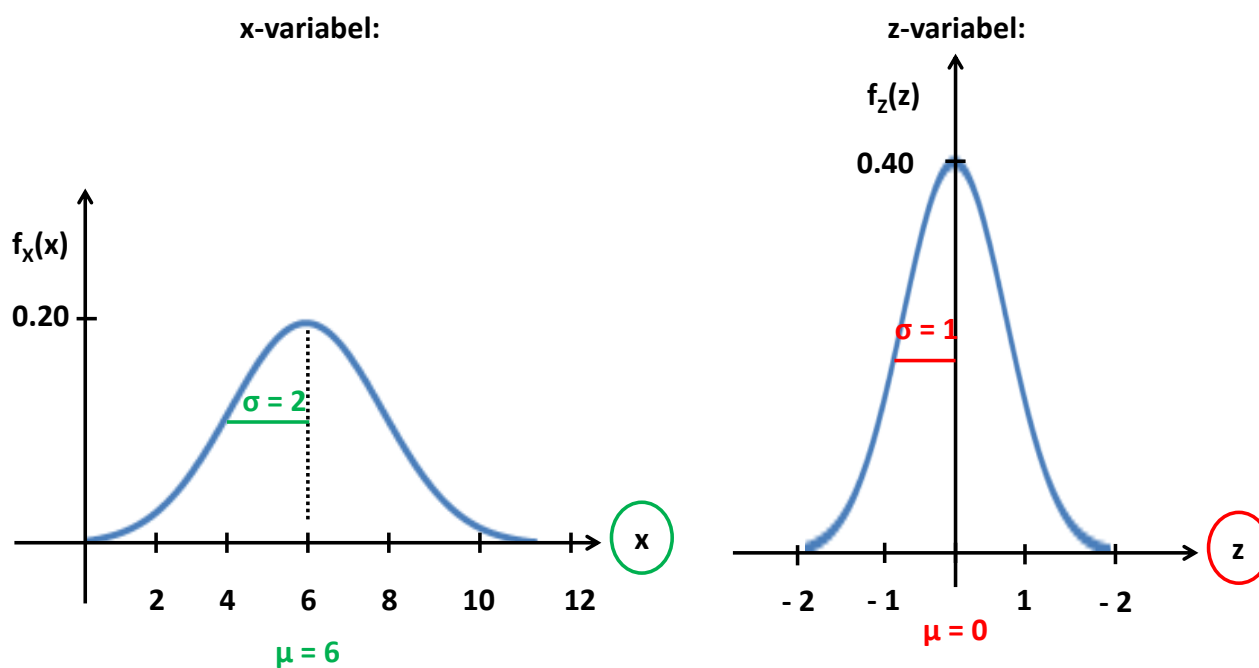
er større enn forventet inntekt ved fullt fly, 91 200 NOK (se oppgave 3d), så lønner det seg for SAS å overbooke.

■

Oppgave 4: (normalfordeling)

a) Normalfordelingen er en kontinuerlig sannsynlighetsfordeling.

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



Figur 1.2: Tetthetsfunksjone $f_X(x)$ og $f_Z(z)$.

c) Arealet under enhver gyldig tetthetsfunksjon er normert til 1.

■

Oppgave 5: (økonomi og logistikk)

a) Siden

$$\sum_{i=0}^n p_i = p_0 + p_1 + p_2 + p_3 = 0.55 + 0.25 + 0.15 + 0.05 = \underline{\underline{1}} \quad (1.30)$$

så er den oppgitte sannsynlighetsfordelingen en gyldig fordeling.

b) i) Forventet antall feilleveringer per dag:

$$\underline{\underline{E[X]}} \stackrel{\text{def.}}{=} \sum_{i=0}^3 x_i \cdot P(X = x_i) \quad (1.31)$$

$$\begin{aligned} &= 0 \cdot \underbrace{P(X=0)}_{=0.55} + 1 \cdot \underbrace{P(X=1)}_{=0.25} + 2 \cdot \underbrace{P(X=2)}_{=0.15} + 3 \cdot \underbrace{P(X=3)}_{=0.05} \\ &= 0 \cdot 0.55 + 1 \cdot 0.25 + 2 \cdot 0.15 + 3 \cdot 0.05 = \underline{\underline{0.70}} \end{aligned} \quad (1.32)$$

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

$$\underline{\underline{E[X^2]}} \stackrel{\text{def.}}{=} \sum_{i=0}^3 x_i^2 \cdot P(X = x_i) \quad (1.33)$$

$$\begin{aligned} &= 0^2 \cdot \underbrace{P(X=0)}_{=0.55} + 1^2 \cdot \underbrace{P(X=1)}_{=0.25} + 2^2 \cdot \underbrace{P(X=2)}_{=0.15} + 3^2 \cdot \underbrace{P(X=3)}_{=0.05} \\ &= 0^2 \cdot 0.55 + 1^2 \cdot 0.25 + 2^2 \cdot 0.15 + 3^2 \cdot 0.05 = \underline{\underline{1.30}} \end{aligned} \quad (1.34)$$

Dette innsatt i setningen for varians: (se formelsamling)

$$\underline{\underline{Var[X]}} = E[X^2] - E[X]^2 = 1.30 - 0.70^2 = \underline{\underline{0.81}} \quad (1.35)$$

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

$$\underline{\underline{E[\bar{X}]}} = E\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] = \frac{1}{n} \left(\overbrace{E[X_1] + E[X_2] + \dots + E[X_n]}^{n \text{ stk.}} \right) \quad (1.36)$$

$$= \frac{\varkappa E[X]}{\varkappa} = E[X] = \underline{\underline{0.70}} \quad (1.37)$$

NB: Overgangen i lign.(1.36) gjelder alltid. Uansett om de stokastiske variablene X_i er uavhengige eller ikke.

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

$$\underline{\underline{Var[\bar{X}]}} = Var\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] \quad (1.38)$$

$$\stackrel{\text{uavhengig}}{=} \frac{1}{n^2} \left(\overbrace{Var[X_1] + Var[X_2] + \dots + Var[X_n]}^{n \text{ stk.}} \right) \quad (1.39)$$

$$= \frac{\varkappa Var[X]}{n^2} \quad (1.40)$$

$$= \frac{Var[X]}{n} = \frac{0.81}{300} = \underline{\underline{0.0027}} \quad (1.41)$$

NB: Overgangen i lign.(1.40) gjelder *kun* dersom de stokastiske variablene X_i er uavhengige.

d) Utfylt versjon av tabellen:

tyngdepunkt		spredning	
$E[X]$	0.70	$\text{Var}[X]$	0.81
$E[\bar{X}]$	0.70	$\text{Var}[\bar{X}]$	0.0027

Figur 1.3: Utfylt versjon av tabellen.

Kommentarer:

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

$$E[\bar{X}] = E[X] \quad (1.42)$$

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

Men variansen til \bar{X} er mye mindre:

$$\text{Var}[\bar{X}] \ll \text{Var}[X] \quad (1.43)$$

Med andre ord: spredningen/usikkerheten til sannsynlighetsfordelingen $P(\bar{X} = \bar{x})$ 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 variabelen \bar{X} , dvs. gjennomsnittet, **normalfordelt**

$$\underline{\underline{\bar{X} \sim N[E[\bar{X}], Var[\bar{X}]]}} = N\left[E[X], \frac{Var[X]}{n}\right] \quad (1.44)$$

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

$$\underline{\underline{n \gtrsim 30}} \quad (1.45)$$

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

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

$$\underline{P(X_1 + X_2 + \dots + X_n > 180)} = P\left(\frac{X_1 + X_2 + \dots + X_n}{n} > \frac{180}{n}\right) \quad (1.46)$$

$$= P\left(\bar{X} > \frac{180}{n}\right) \quad (1.47)$$

$$= 1 - P\left(\bar{X} \leq \frac{180}{n}\right) \quad (1.48)$$

$$\stackrel{\text{standardiser}}{=} 1 - P\left(\underbrace{\frac{\bar{X} - E[\bar{X}]}{\sigma[\bar{X}]}}_{= \bar{Z}} \leq \frac{\frac{180}{n} - E[\bar{X}]}{\sigma[\bar{X}]}\right) \quad (1.49)$$

$$= 1 - P\left(\bar{Z} \leq \frac{\frac{180}{300} - 0.70}{\sqrt{0.0027}}\right) \quad (1.50)$$

$$= \underline{1 - P(\bar{Z} \leq -1.92)} \quad (1.51)$$

$$\underline{\underline{P(X_1 + X_2 + \dots + X_n > 180)}} = 1 - P(\bar{Z} \leq -1.92) \quad (1.52)$$

$$= 1 - \left[1 - P(\bar{Z} \leq 1.92) \right] \quad (1.53)$$

$$= P(\bar{Z} \leq 1.92) \quad (1.54)$$

$$= G(1.92) \stackrel{\text{tabell}}{=} \underline{\underline{0.9726}} \quad (1.55)$$

■

Kapittel 2

LØSNING: Eksamen 10. januar 2013

“MAT110 Statistikk 1”

Oppgave 1: (sannsynlighetsregning)

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

$$\underline{\underline{P(E) \cdot P(F)}} = 0.52 \cdot 0.46 = \underline{\underline{0.2392}} \quad (2.1)$$

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

$$P(E) \cdot P(F) = 0.2392 \quad (2.2)$$

og

$$P(E \cap F) = 0.42 \quad (2.3)$$

så innser vi umiddelbart at E og F er avhengige.

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

$$P(E) \cdot P(G) = 0.52 \cdot 0.38 = 0.1976 \quad (2.4)$$

og

$$P(E \cap G) = 0.1976 \quad (2.5)$$

så innser vi umiddelbart at E og G er uavhengige.

- d) Man kan finne sannsynligheten for at aksje B ikke stiger med komplementsetningen:

$$\underline{\underline{P(\bar{B})}} = 1 - P(B) = 1 - 0.45 = \underline{\underline{0.55}} \quad (2.6)$$

- e) Man kan finne sannsynligheten for at aksje A stiger ved å bruke oppsplitting av utfallsrom Ω :

$$\underline{\underline{P(A)}} = P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B}) \quad (2.7)$$

$$= 0.33 \cdot 0.45 + 0.82 \cdot 0.55 = \underline{\underline{0.5995}} \quad (2.8)$$

■

Oppgave 2: (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 $E[X]$ som kommer på utflukten:

$$\underline{E[X]} \stackrel{\text{bin.}}{=} n \cdot p = 150 \cdot 0.90 = \underline{135} \quad (2.9)$$

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

$$\underline{Var[X]} \stackrel{\text{bin.}}{=} n \cdot p \cdot (1 - p) = 150 \cdot 0.90 \cdot (1 - 0.90) = \underline{13.5} \quad (2.10)$$

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

$$\underline{\sigma[X]} = \sqrt{Var[X]} = \sqrt{13.5} \approx \underline{3.67} \quad (2.11)$$

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

$$\underline{\underline{n \cdot p(1 - p) \gtrsim 5}} \quad (2.12)$$

- ii) Før vårt tilfelle:

$$150 \cdot 0.90(1 - 0.90) = 13.5 \gtrsim 5 \quad (2.13)$$

Ja, betingelsen er godt oppfylt for vårt tilfelle.

- e) Siden betingelsen i lign.(2.12) er oppfylt så er binominal fordelingen tilnærmet en **normalfordeling**. Dette gjør det enklere å regne ut sannsynligheten for at alle oppmøtte får plass er:

$$\underline{\underline{P(\text{en tur})}} = P(X \leq 140) \quad (2.14)$$

$$\stackrel{\text{standardiser}}{=} P\left(\underbrace{\frac{X - E[X]}{\sigma[X]}}_{\equiv Z} \leq \underbrace{\frac{140 + 0.5 - E[X]}{\sigma[X]}}_{\equiv Z_0}\right) \quad (2.15)$$

$$= P\left(Z \leq \underbrace{\frac{140 + 0.5 - 135}{3.67}}_{\equiv Z_0}\right) \quad (2.16)$$

$$= P(Z \leq 1.50) \quad (2.17)$$

$$\stackrel{\text{tabell}}{=} \underline{\underline{0.9332}} \quad (2.18)$$

¹Se setningen på side 26 i formelsamlingen.

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

$$P(\text{en tur}) + P(\text{to turer}) = 1 \quad (2.19)$$

Men fra oppgave **2e** har vi: $P(\text{en tur}) = 0.9332$. Dermed:

$$\underline{\underline{P(\text{to turer})}} = 1 - P(\text{to turer}) \quad (2.20)$$

$$= 1 - 0.9332 = \underline{\underline{0.0668}} \quad (2.21)$$

- g) Forventet utgifter $E[U]$ forbundet med å frakte studenter til Hjertøya:

$$\underline{\underline{E[U]}} = E[c \cdot Y] = c \cdot E[Y] \quad (2.22)$$

$$= c \sum_{i=1}^2 y_i P(Y = y_i) \quad (2.23)$$

$$= c \cdot \left(1 \cdot P(Y = 1) + 2 \cdot P(Y = 2) \right) \quad (2.24)$$

$$= c \cdot \left(\underbrace{1 \cdot P(X \leq 140)}_{=0.9332} + 2 \cdot \underbrace{P(X > 140)}_{=0.0668} \right) \quad (2.25)$$

resultat hentet fra oppg. **2e** og **2f**, lign.(2.18) og (2.21).

$$= 950 \cdot \left(1 \cdot 0.9332 + 2 \cdot 0.0668 \right) \text{ NOK} \quad (2.26)$$

$$= \underline{\underline{1013.46 \text{ NOK}}} \quad (2.27)$$

h) Forventet fortjeneste $E[F]$ forbundet med å frakte studenter til Hjertøya:

$$\underline{\underline{E[F]}} = E[a \cdot X - c \cdot Y] \quad (2.28)$$

$$= a \cdot E[X] - \underbrace{c \cdot E[Y]}_{= E[U]} \quad (2.29)$$

$$= 35 \cdot 135 - 1013.46 \quad (2.30)$$

$$= \underline{\underline{3711.54 \text{ NOK}}} \quad (2.31)$$

■

Oppgave 3: (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 1.

Oppgave 4: (økonomi og logistikk)

a) Siden

$$\underline{\underline{\sum_{i=0}^n p_i}} = p_0 + p_1 + p_2 + p_3 = 0.70 + 0.15 + 0.10 + 0.05 = \underline{\underline{1}} \quad (2.32)$$

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

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

$$\underline{\underline{E[X]}} \stackrel{\text{def.}}{=} \sum_{i=0}^3 x_i \cdot P(X = x_i) \quad (2.33)$$

$$= 0 \cdot \underbrace{P(X = 0)}_{=0.70} + 1 \cdot \underbrace{P(X = 1)}_{=0.15} + 2 \cdot \underbrace{P(X = 2)}_{=0.10} + 3 \cdot \underbrace{P(X = 3)}_{=0.05} \quad (2.34)$$

$$= 0 \cdot 0.70 + 1 \cdot 0.15 + 2 \cdot 0.10 + 3 \cdot 0.05 = \underline{\underline{0.50}} \quad (2.35)$$

ii) Tolking:

$$\underline{\underline{E[X] = \text{forventet antall feilleveringer for et tilfeldig valgt bud} \\ \text{i Stockholm en tilfeldig valgt dag}}} \quad (2.36)$$

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

$$\underline{E[X^2]} \stackrel{\text{def.}}{=} \sum_{i=0}^3 x_i^2 \cdot P(X = x_i) \quad (2.37)$$

$$\begin{aligned} &= 0^2 \cdot \underbrace{P(X = 0)}_{=0.70} + 1^2 \cdot \underbrace{P(X = 1)}_{=0.15} + 2^2 \cdot \underbrace{P(X = 2)}_{=0.10} + 3^2 \cdot \underbrace{P(X = 3)}_{=0.05} \\ &= 0^2 \cdot 0.70 + 1^2 \cdot 0.15 + 2^2 \cdot 0.10 + 3^2 \cdot 0.05 = \underline{1.0} \end{aligned} \quad (2.38)$$

Dette innsatt i “varianssetningen”: (se formelsamling)

$$\underline{\underline{Var[X]}} = E[X^2] - E[X]^2 = 1.0 - 0.50^2 = \underline{\underline{0.75}} \quad (2.39)$$

ii) Tolking:

$Var[X]$ = forventet variasjon/spredning i antall feilleveringer
for et tilfeldig valgt bud i Stockholm en tilfeldig valgt dag

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

$$\underline{\underline{E[\bar{X}]}} = E \left[\frac{X_1 + X_2 + \dots + X_n}{n} \right] = \frac{1}{n} \left(\overbrace{E[X_1] + E[X_2] + \dots + E[X_n]}^{n \text{ stk.}} \right) \quad (2.40)$$

$$= \frac{\cancel{n} E[X]}{\cancel{n}} = E[X] = \underline{\underline{0.50}} \quad (2.41)$$

NB: Overgangen i lign.(2.40) 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:

$$\underline{\underline{Var[\bar{X}]}} = Var \left[\frac{X_1 + X_2 + \dots + X_n}{n} \right] \quad (2.42)$$

$$\stackrel{\text{uavhengig}}{=} \frac{1}{n^2} \left(\overbrace{Var[X_1] + Var[X_2] + \dots + Var[X_n]}^{n \text{ stk.}} \right) \quad (2.43)$$

$$= \frac{\cancel{n} Var[X]}{\cancel{n}^2} \quad (2.44)$$

$$= \frac{Var[X]}{n} = \frac{0.75}{312} \approx \underline{\underline{0.0024}} \quad (2.45)$$

NB: Overgangen i lign.(2.43) gjelder *kun* dersom de stokastiske variablene X_i er uavhengige.

- e) Forventningene til \bar{X} og X er de samme, dvs.:

$$\boxed{\underbrace{E[\bar{X}]}_{= 0.50} = \underbrace{E[X]}_{= 0.50} \quad (2.46)}$$

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

Men variansen til \bar{X} er mye mindre:

$$\boxed{\underbrace{Var[\bar{X}]}_{= 0.0024} \ll \underbrace{Var[X]}_{= 0.75} \quad (2.47)}$$

Med andre ord: spredningen/usikkerheten til sannsynlighetsfordelingen $P(\bar{X} = \bar{x})$ 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 variablen \bar{X} , dvs. gjennomsnittet, **normalfordelt**:

$$\underline{\underline{\bar{X} \sim N[E[\bar{X}], Var[\bar{X}]] = N\left[E[X], \frac{Var[X]}{n}\right]}} \quad (2.48)$$

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

$$\underline{\underline{n \gtrsim 30}} \quad (2.49)$$

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)

$$\underline{\underline{P(X_1 + X_2 + \dots + X_n > 200)}} = P\left(\frac{X_1 + X_2 + \dots + X_n}{n} > \frac{200}{n}\right) \quad (2.50)$$

$$= P\left(\bar{X} > \frac{200}{n}\right) \quad (2.51)$$

$$= 1 - P\left(\bar{X} \leq \frac{200}{n}\right) \quad (2.52)$$

$$\stackrel{\text{standardiser}}{=} 1 - P\left(\underbrace{\frac{\bar{X} - E[\bar{X}]}{\sigma[\bar{X}]}}_{= \bar{Z}} \leq \frac{\frac{200}{n} - E[\bar{X}]}{\sigma[\bar{X}]}\right) \quad (2.53)$$

$$= 1 - P\left(\bar{Z} \leq \frac{\frac{200}{312} - 0.50}{\sqrt{0.0024}}\right) \quad (2.54)$$

$$= 1 - P(\bar{Z} \leq 2.88) \quad (2.55)$$

$$= 1 - G(2.88) \quad (2.56)$$

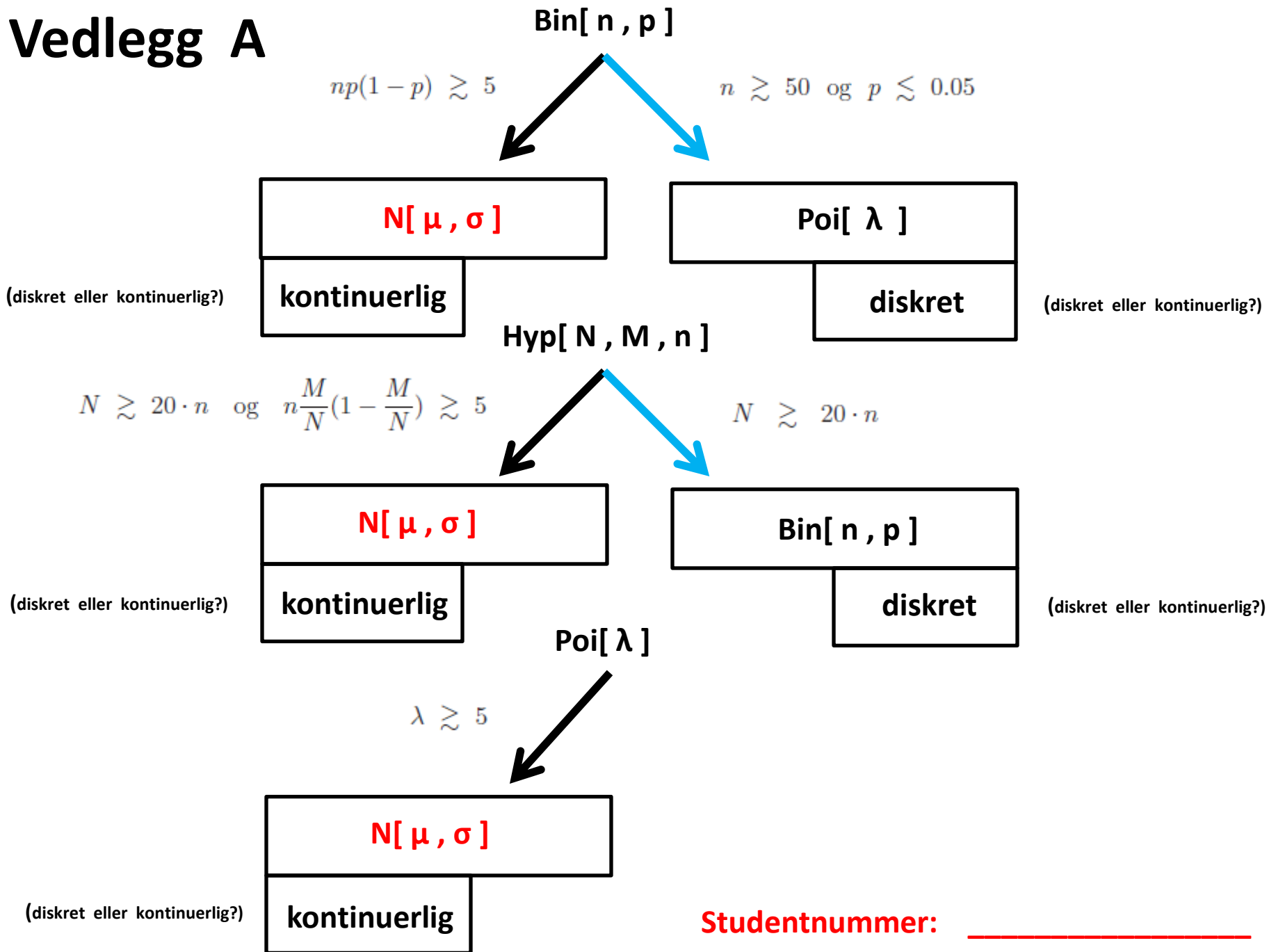
$$\stackrel{\text{tabell}}{=} 1 - 0.9980 \quad (2.57)$$

$$= \underline{\underline{0.0020}} \quad (2.58)$$

■

Vedlegg A

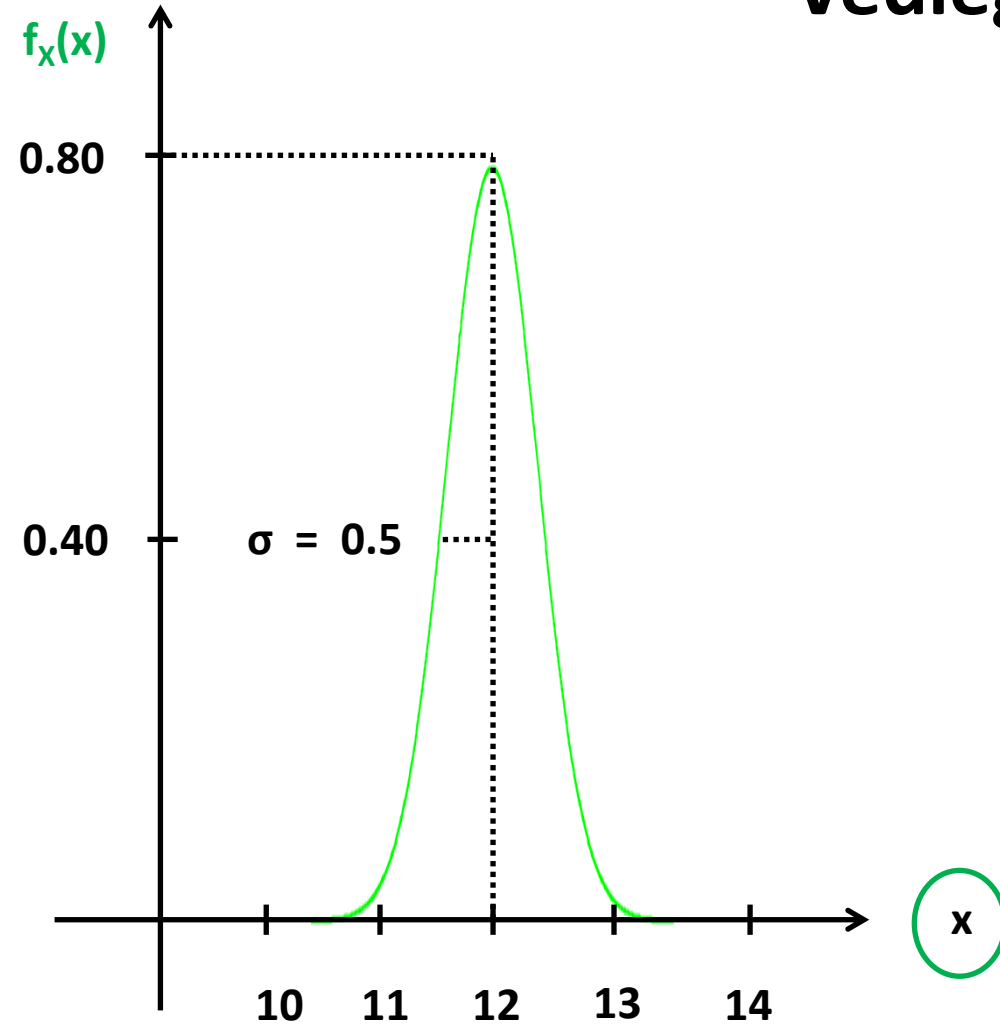
Vedlegg A



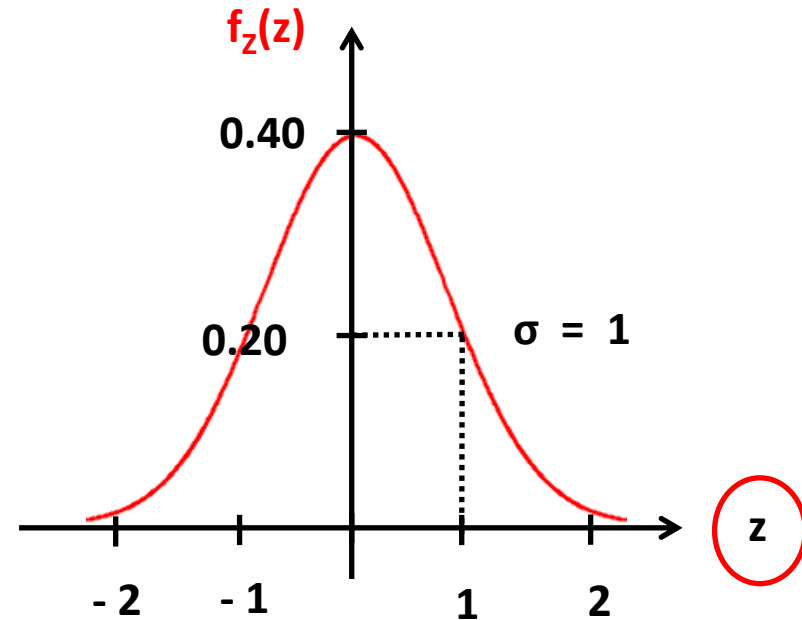
Vedlegg B

Vedlegg B

x-variabel:



z-variabel:



Studentnummer: _____

Kapittel 3

LØSNING: Eksamen 30. mai 2013

“MAT110 Statistikk 1”

Oppgave 1: (sentrale **formler**, oversikt)

Se vedlegg A.



Oppgave 2: (logistikk)

a) Tolkning:

$$\underline{\underline{P(B_2|B_1)}} = \text{sannsynligheten for at bølgehøyden er for stor dag 2,} \\ \underline{\underline{\text{gitt at den er for stor dag 1}}} \quad (3.1)$$

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

$$\underbrace{P(B_2|B_1)}_{= 0.70} \neq \underbrace{P(B_2)}_{= 0.05} \quad (3.2)$$

venstre side IKKE like høyre side

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

c) For å finne sannsynligheten for at bølgehøyden er for stor både dag 1 og dag 2, dvs. $P(B_1 \cap B_2)$, så kan man bruke **definisjonen** av betinget sannsynlighet:

$$\underline{\underline{P(B_1 \cap B_2)}} \stackrel{\text{bet.}}{=} \overbrace{P(B_2|B_1)}^{=0.70} \cdot \overbrace{P(B_1)}^{=0.05} \quad (3.3)$$

$$= 0.70 \cdot 0.05 = \underline{\underline{0.035}} \quad (3.4)$$

d) For å finne sannsynligheten for at bølgehøyden er for stor både dag 1 eller dag 2, dvs. $P(B_1 \cup B_2)$, så kan man bruke den generelle **addisjonssetningen**:

$$\underline{\underline{P(B_1 \cup B_2)}} \stackrel{\text{add.}}{=} \overbrace{P(B_1)}^{=0.05} + \overbrace{P(B_2)}^{=0.05} - \overbrace{P(B_1 \cap B_2)}^{=0.035} \quad (3.5)$$

$$= 0.05 + 0.05 - 0.035 = \underline{\underline{0.065}} \quad (3.6)$$

e) Bruk F.EKS. **total** sannsynliget:

$$\underline{\underline{P(B_1 \cap \bar{B}_2)}} \stackrel{\text{tot.}}{=} \overbrace{P(B_1)}^{= 0.05} - \overbrace{P(B_1 \cap B_2)}^{= 0.035} \quad (3.7)$$

$$= 0.05 - 0.035 = \underline{\underline{0.015}} \quad (3.8)$$

ELLER definisjonen på betinget sannsynlighet:

$$\underline{\underline{P(B_1 \cap \bar{B}_2)}} \stackrel{\text{bet.}}{=} \overbrace{P(\bar{B}_2|B_1)}^{= 1-P(B_2|B_1)} \cdot P(B_1) \quad (3.9)$$

$$= \left(1 - \overbrace{P(B_2|B_1)}^{= 0.70}\right) \cdot \overbrace{P(B_1)}^{= 0.05} \quad (3.10)$$

$$= (1 - 0.70) \cdot 0.05 = \underline{\underline{0.015}} \quad (3.11)$$

f) Tolkning:

$$\underline{\underline{P(B_1 \cap \bar{B}_2)}} = \text{sannsynligheten for at bølgehøyden er for stor dag 1,} \\ \underline{\underline{\text{men ikke for høy dag 2}}} \quad (3.12)$$

g) Bruk definisjonen av **betinget** sannsynlighet:

$$P(\bar{B}_2|\bar{B}_1) = \frac{\overbrace{P(\bar{B}_2 \cap \bar{B}_1)}^{= 1-P(B_1 \cup B_2)}}{\underbrace{P(\bar{B}_1)}_{= 1-P(B_1)}} \quad (3.13)$$

I telleren er den ene "*tvillingsetningen*" benyttet. I nevneren er **komplement**setningen benyttet.

$$\underline{\underline{P(\bar{B}_2|\bar{B}_1)}} = \frac{1 - \overbrace{P(B_2 \cup B_1)}^{= 0.065}}{1 - \underbrace{P(B_1)}_{= 0.05}} = \frac{1 - 0.065}{1 - 0.05} = \underline{\underline{0.98}} \quad (3.14)$$

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: (økonomi)

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

$$\underline{\underline{P(X \geq 1)}} = 1 - P(X \leq 0) \quad (3.15)$$

$$= 1 - \underbrace{P(X = 0)}_{\substack{\text{tabell} \\ 0.57}} = 1 - 0.57 = \underline{\underline{0.43}} \quad (3.16)$$

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

$$\underline{\underline{E[X]}} \stackrel{\text{def.}}{=} \sum_{i=0}^4 x_i \cdot P(X = x_i) \quad (3.17)$$

$$\begin{aligned} &= 0 \cdot \underbrace{P(X = 0)}_{=0.57} + 1 \cdot \underbrace{P(X = 1)}_{=0.13} + 2 \cdot \underbrace{P(X = 2)}_{=0.18} + 3 \cdot \underbrace{P(X = 3)}_{=0.10} + 4 \cdot \underbrace{P(X = 3)}_{=0.02} \\ &= 0 \cdot 0.57 + 1 \cdot 0.13 + 2 \cdot 0.18 + 3 \cdot 0.10 + 4 \cdot 0.02 = \underline{\underline{0.87}} \quad (3.18) \end{aligned}$$

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

$$\underline{\underline{E[X^2]}} \stackrel{\text{def.}}{=} \sum_{i=0}^3 x_i^2 \cdot P(X = x_i) \quad (3.19)$$

$$\begin{aligned} &= 0^2 \cdot \underbrace{P(X = 0)}_{=0.57} + 1^2 \cdot \underbrace{P(X = 1)}_{=0.13} + 2^2 \cdot \underbrace{P(X = 2)}_{=0.18} + 3^2 \cdot \underbrace{P(X = 3)}_{=0.10} + 4^2 \cdot \underbrace{P(X = 3)}_{=0.02} \\ &= 0^2 \cdot 0.57 + 1^2 \cdot 0.13 + 2^2 \cdot 0.18 + 3^2 \cdot 0.10 + 4^2 \cdot 0.02 = \underline{\underline{2.07}} \quad (3.20) \end{aligned}$$

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

$$\underline{\underline{Var[X]}} = E[X^2] - E[X]^2 = 2.07 - 0.87^2 = \underline{\underline{1.3131}} \quad (3.21)$$

c) i) Tolkning:

$$\underline{\underline{E[\bar{X}]}} = \text{forventet antall overtrekk per måned i gjennomsnitt} \\ \underline{\underline{\text{for kundene i Sparebanken Møre}}} \quad (3.22)$$

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

$$\underline{\underline{E[\bar{X}]}} = E \left[\frac{1}{n} \left(X_1 + X_2 + \dots + X_n \right) \right] \quad (3.23)$$

$$\stackrel{\text{alltid}}{=} \frac{1}{n} \left(E[X_1] + E[X_2] + \dots + E[X_n] \right) \quad (3.24)$$

$$= \frac{1}{n} \left(\underbrace{E[X] + E[X] + \dots + E[X]}_{=n} \right) = \frac{1}{n} n \cdot \underbrace{E[X]}_{=0.87} = \underline{\underline{0.87}} \quad (3.25)$$

NB: Overgangen i lign.(3.23) til (3.24) gjelder **alltid**. Uansett om de stokastiske variablene X_i er uavhengige eller ikke.

d) i) Tolkning:

$$\underline{\underline{Var[\bar{X}]}} = \text{forventet variasjon/spredning i antall overtrekk per måned} \\ \text{i gjennomsnitt for kundene i Sparebanken Møre} \quad (3.26)$$

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

$$\underline{\underline{Var[\bar{X}]}} = Var\left[\frac{1}{n}\left(X_1 + X_2 + \dots + X_n\right)\right] \quad (3.27)$$

$$\stackrel{\text{uavh.}}{=} \frac{1}{n^2} \underbrace{\left(Var[X_1] + Var[X_2] + \dots + Var[X_n]\right)}_{n \cdot Var[X]} \quad (3.28)$$

$$= \frac{1}{n^2} \underbrace{\cancel{\mu} \cdot Var[X]}_{=1.3131} = \frac{1.3131}{500} = \underline{\underline{0.002626}} \quad (3.29)$$

NB: Overgangen i lign.(3.27) til (3.28) 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 $i = 1, 2, 3, \dots, n$
2. alle X_i har samme sannsynlighetsfordeling: (oppgitt i oppgaveteksten)
 $X_i \sim$ samme sannsynlighetsfordeling for alle $i = 1, 2, 3, \dots, n$
3. antall “forsøk”, dvs. antall overtrekk, $n = 500$ er tilstrekkelig stort¹

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

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

f) Fra oppgavene foran ser vi at:

$$\underbrace{E[\bar{X}]}_{= 0.87} = \underbrace{E[X]}_{= 0.87} \quad (3.30)$$

og at

$$\underbrace{Var[\bar{X}]}_{= 0.002626} \ll \underbrace{Var[X]}_{= 1.3131} \quad (3.31)$$

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

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

$$\underline{\underline{P\left(X_1 + X_2 + \dots + X_n > 400\right)}} = P\left(\frac{X_1 + X_2 + \dots + X_n}{n} > \frac{400}{n}\right) \quad (3.32)$$

$$= P\left(\bar{X} > \frac{400}{n}\right) \quad (3.33)$$

$$= 1 - P\left(\bar{X} \leq \frac{400}{n}\right) \quad (3.34)$$

$$\stackrel{\text{standardiser}}{=} 1 - P\left(\underbrace{\frac{\bar{X} - E[\bar{X}]}{\sigma[\bar{X}]}}_{=\bar{Z}} \leq \frac{\frac{400}{n} - E[\bar{X}]}{\sigma[\bar{X}]}\right) \quad (3.35)$$

$$= 1 - P\left(\bar{Z} \leq \frac{\frac{400}{500} - 0.87}{\sqrt{0.002626}}\right) \quad (3.36)$$

$$= 1 - P(\bar{Z} \leq -1.37) \quad (3.37)$$

$$= 1 - \left(1 - \underbrace{P(\bar{Z} \leq 1.37)}_{= G(1.37)}\right) \quad (3.38)$$

$$= 1 - 1 + \underbrace{G(1.37)}_{\stackrel{\text{tabell}}{=} 0.9147} \quad (3.39)$$

$$= \underline{\underline{0.9147}} \quad (3.40)$$

Kommentar:

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

$$P(X_1 + X_2 + \dots + X_n \leq X_{\text{grense}}) = 0.95 \quad (3.41)$$

$$P\left(\frac{X_1 + X_2 + \dots + X_n}{n} \leq \underbrace{\frac{X_{\text{grense}}}{n}}_{= \bar{X}_{\text{grense}}}\right) = 0.95 \quad (3.42)$$

$$P(\bar{X} \leq \bar{X}_{\text{grense}}) = 0.95 \quad (3.43)$$

Vi standardiserer lign.(3.43):

$$P\left(\underbrace{\frac{\bar{X} - E[\bar{X}]}{\sigma[\bar{X}]}}_{= \bar{Z}} \leq \underbrace{\frac{\bar{X}_{\text{grense}} - E[\bar{X}]}{\sigma[\bar{X}]}}_{= \bar{Z}_{\text{grense}}}\right) = 0.95 \quad (3.44)$$

$$P(\bar{Z} \leq \bar{Z}_{\text{grense}}) = 0.95 \quad (3.45)$$

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

$$\bar{Z}_{\text{grense}} = 1.645 \quad (3.46)$$

Dermed:

$$\bar{Z}_{\text{grense}} = \frac{\bar{X}_{\text{grense}} - E[\bar{X}]}{\sigma[\bar{X}]} \quad (3.47)$$

$$\bar{X}_{\text{grense}} = \bar{Z}_{\text{grense}} \cdot \sigma[\bar{X}] + E[\bar{X}] \quad (3.48)$$

$$= 1.645 \cdot \sqrt{0.002626} + 0.87 \approx \underline{0.9543} \quad (3.49)$$

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

$$\underline{\underline{X_{\text{grense}}}} = \bar{X}_{\text{grense}} \cdot n \quad (3.50)$$

$$= 0.9543 \cdot 500 \approx \underline{\underline{478}} \quad (3.51)$$

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

■

Oppgave 4: (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 $D \sim \text{Bin}[n, p_d]$:

$$\underline{\underline{E[D]}} = n \cdot p_d = 25 \cdot 0.05 = \underline{\underline{1.25}} \quad (3.52)$$

ii) Tolkning:

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

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

$$\underline{\underline{Var[D]}} = n \cdot p_d (1 - p_d) = 25 \cdot 0.05 \cdot (1 - 0.05) = \underline{\underline{1.1875}} \quad (3.53)$$

ii) Tolkning:

$Var[D]$ = 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:

$$\underline{\underline{P(D > 2)}} = 1 - P(D \leq 2) \quad (3.54)$$

$$= 1 - \left(P(D = 0) + P(D = 1) + P(D = 2) \right) \quad (3.55)$$

$$= 1 - P(D = 0) - P(D = 1) - P(D = 2) \quad (3.56)$$

$$= 1 - \binom{n}{0} p_d^0 (1 - p_d)^{n-0} + \binom{n}{1} p_d^1 (1 - p_d)^{n-1} + \binom{n}{2} p_d^2 (1 - p_d)^{n-2} \quad (3.57)$$

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

$$= 1 - 0.2774 - 0.3650 - 0.2305 \quad (3.58)$$

$$= \underline{\underline{0.1271}} \quad (\text{svar med 4 desimalers nøyaktighet}) \quad (3.59)$$

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

Siden

$$n \cdot p_d (1 - p_d) = 25 \cdot 0.05 (1 - 0.05) = 1.1875 \ll 5 \quad (3.60)$$

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:

$$\underline{\underline{E[F]}} = E[(n - D - T) \cdot i - n \cdot (k + k_t)] \quad (3.61)$$

$$= E[n \cdot i - D \cdot i - T \cdot i - n \cdot (k + k_t)] \quad (3.62)$$

$$= \underbrace{E[n \cdot i]}_{= n \cdot i} - \underbrace{E[D \cdot i]}_{= E[D] \cdot i} - \underbrace{E[T \cdot i]}_{= E[T] \cdot i} - \underbrace{E[n \cdot (k + k_t)]}_{= n \cdot (k + k_t)} \quad (3.63)$$

$$= n \cdot i - \underbrace{E[D] \cdot i}_{= n \cdot p_d} - \underbrace{E[T] \cdot i}_{= n \cdot p_t} - n \cdot (k + k_t) \quad (3.64)$$

$$= n \cdot p_d \cdot i - n \cdot p_t \cdot i - n \cdot (k + k_t) \quad (3.65)$$

$$= \underline{\underline{n \left[(1 - p_d - p_t) \cdot i - (k + k_t) \right]}} \quad , \quad \text{q.e.d.} \quad (3.66)$$

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

Bring:

$$\underline{E[F]} = n \cdot \left[(1 - p_d) \cdot i - k - \underbrace{(p_t \cdot i + k_t)}_{\text{NB!}} \right] \quad (3.67)$$

$$= n \cdot \left[(1 - p_d) \cdot i - k - (0.15 \cdot 1\,700 + 275) \text{ NOK} \right] \quad (3.68)$$

$$= \underline{n \cdot \left[(1 - p_d) \cdot i - k - 530 \text{ NOK} \right]} \quad (3.69)$$

DHL:

$$\underline{E[F]} = n \cdot \left[(1 - p_d) \cdot i - k - \underbrace{(p_t \cdot i + k_t)}_{\text{NB!}} \right] \quad (3.70)$$

$$= n \cdot \left[(1 - p_d) \cdot i - k - (0.04 \cdot 1\,700 + 750) \text{ NOK} \right] \quad (3.71)$$

$$= \underline{n \cdot \left[(1 - p_d) \cdot i - k - 818 \text{ NOK} \right]} \quad (3.72)$$

Konklusjon:

Bring har minst forventet utgift.

For å få størst forventet inntekt bør derfor Glamox velge Bring.



Vedlegg A

**(Husk å skrive studentnummer
på vedlegget, begge sider.)**

Beskrivende statistikk
(utvalg av observasjoner)

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

Side 1 (av 2)

Empirisk gjennomsnitt: (formel)

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

Forventning: (diskret) (formel)

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

Kommentar:

Lokaliseringsmål: tyngdepunkt

Empirisk varians: (formel)

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

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)

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

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:

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

Korrelasjonskoeffisient: (formel)

$$\rho[X, Y] = \frac{Cov[X, Y]}{\sqrt{Var[X]} \cdot \sqrt{Var[Y]}}$$

Kommentar:

Et mål på lineær samvariasjon, korrelasjon.

Normalisert, ligger i intervallet: $-1 \leq \text{korr. koeff.} \leq 1$

Empirisk korrelasjonskoeffisient:

$$R_{xy} = -1$$

Korrelasjonskoeffisient: (formel)

$$\rho[X, Y] = -1$$

Kommentar:

Sterk negativ korrelasjon.

Empirisk korrelasjonskoeffisient:

$$R_{xy} = 1$$

Korrelasjonskoeffisient: (formel)

$$\rho[X, Y] = 1$$

Kommentar:

Sterk positiv korrelasjon.

Kapittel 4

LØSNING: Eksamen 6. januar 2014

“MAT110 Statistikk 1”

Oppgave 1: (revisjon)

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

$$\underline{\underline{P_{A11}}} = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{12}{2000} = \underline{\underline{0.006}} \quad (4.1)$$

$$\underline{\underline{P_{A12}}} = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{24}{8000} = \underline{\underline{0.003}} \quad (4.2)$$

b) Samme metode som **1a**:

$$\underline{\underline{P_{B11}}} = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{20}{4000} = \underline{\underline{0.005}} \quad (4.3)$$

$$\underline{\underline{P_{B12}}} = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{2}{1000} = \underline{\underline{0.002}} \quad (4.4)$$

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

$$\underline{\underline{P_A}} = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{12 + 24}{2000 + 8000} = \underline{\underline{0.0036}} \quad (4.5)$$

$$\underline{\underline{P_B}} = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{20 + 2}{4000 + 1000} = \underline{\underline{0.0044}} \quad (4.6)$$

Dermed:

$P_A < P_B \Rightarrow$ 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 1c.
Dersom vi ser på begge årene under ett så er strategi B best, jfr. oppgave 1d.

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.

■

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

$$P_{A11} > P_{B11} \quad (4.7)$$

$$P_{A12} > P_{B12} \quad (4.8)$$

så er:

$$P_A < P_B \quad (4.9)$$

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

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

Oppgave 2: (logistikk)³

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

$$\underline{\underline{E[D]}} = \sum_{i=0}^5 d_i P(D = d_i) \quad (4.10)$$

$$= 0 \cdot 0.1 + 1 \cdot 0.05 + 2 \cdot 0.15 + 3 \cdot 0.3 + 4 \cdot 0.25 + 5 \cdot 0.15 = \underline{\underline{3}} \quad (4.11)$$

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:

$$\underline{\underline{\sum_{i=0}^5 P(S = s_i)}} = P(S = 0) + P(S = 1) + \dots + P(S = 5) \quad (4.12)$$

$$= 0.1 + 0.05 + 0.15 + 0.7 + 0 + 0 = \underline{\underline{1}} \quad (4.13)$$

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

$$\underline{\underline{E[S]}} = \sum_{i=0}^5 s_i P(S = s_i) \quad (4.14)$$

$$= 0 \cdot 0.1 + 1 \cdot 0.05 + 2 \cdot 0.15 + 3 \cdot 0.7 + 4 \cdot 0 + 5 \cdot 0 = \underline{\underline{2.45}} \quad (4.15)$$

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

⁴Bruker 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

$$\underline{E[\pi(q)]} = E[rS - wq] = \underline{rE[S] - wq} \quad (4.18)$$

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

$$\underline{\underline{E[\pi(q)]}} = (20 \cdot 2.45 - 5 \cdot 3) \text{ NOK} = \underline{\underline{34 \text{ NOK}}} \quad (4.19)$$

⁵Her bruker vi regneregelen: (a og b er konstanter)

$$E[aX + bY] = aE[X] + bE[Y], \quad (4.16)$$

$$E[a] = a, \quad (4.17)$$

som man kan finne i kap. 5 i formelsamlingen.

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

$$P(D \leq q^*) = 1 - \frac{w}{r} \quad (4.20)$$

$$P(D \leq q^*) = 1 - \frac{5}{20} \quad (4.21)$$

$$P(D \leq q^*) = 0.75 \quad (4.22)$$

$$P\left(\underbrace{\frac{D - \mu}{\sigma}}_{\equiv Z} \leq \underbrace{\frac{q^* - \mu}{\sigma}}_{\equiv Z_0}\right) \stackrel{\text{standardiser}}{=} 0.75 \quad (4.23)$$

$$P(Z \leq Z_0) = 0.75 \quad (4.24)$$

Ved “[omvendt tabeloppslag](#)” 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:

$$Z_0 = 0.675 \quad (4.25)$$

Vi løser:

$$Z_0 = \frac{q^* - \mu}{\sigma} \quad (4.26)$$

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

$$\underline{q^*} = \mu + Z_0 \cdot \sigma \quad (4.27)$$

$$= 3 + 0.675 \cdot 1.5 = \underline{4.0125} \quad (4.28)$$

Avisgutten må bestille $q^* \approx 4$ 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

$$q^* > \mu , \quad (4.29)$$

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.



⁶Avisgutten 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

$$\underline{\underline{P(S_1 \cap S_2)}} = P(S_1) \cdot P(S_2) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} \approx \underline{\underline{0.028}} \quad (4.30)$$

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

$$\underline{\underline{P(S_1 \cup S_2)}} = P(S_1) + P(S_2) - \overbrace{P(S_1 \cap S_2)}^{= \frac{1}{36}} \quad (4.31)$$

$$= \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36} \approx \underline{\underline{0.306}} \quad (4.32)$$

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

$$\underline{\underline{P(S_2|S_1)}} = \frac{P(S_1 \cap S_2)}{P(S_1)} = \frac{\frac{2}{30}}{\frac{5}{30}} = \frac{2}{5} = \underline{\underline{0.4}} \quad (4.33)$$

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

⁷Se den “spesielle multiplikasjonssetning” i formelsamlingen.

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

$$P(S_1 \cap S_2 \cap S_3) = P(S_3 \cap S_2 \cap S_1) \quad (4.35)$$

$$= P(S_3|S_2 \cap S_1)P(S_2 \cap S_1) \quad (4.36)$$

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

$$\underline{\underline{P(S_1 \cap S_2 \cap S_3)}} = P(S_3|S_2 \cap S_1)P(S_2 \cap S_1) \quad (4.37)$$

$$= \underbrace{P(S_3|S_2 \cap S_1)}_{= 0.6} \underbrace{P(S_2|S_1)}_{= 0.4} \underbrace{P(S_1)}_{= \frac{1}{6}} \quad (4.38)$$

$$= 0.6 \cdot 0.4 \cdot \frac{1}{6} = \underline{\underline{0.04}} \quad (4.39)$$

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 **3d** og $P(S_1) = \frac{1}{6}$ var også oppgitt i oppgaven.

■

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

$$P(A \cap B) = P(A|B)P(B) \quad (4.34)$$

I lign.(4.36) 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:

$$\underline{\underline{P(X \geq 40)}} = P(X = 40) + P(X = 45) + P(X = 50) + P(X = 55) + P(X = 60) \quad (4.40)$$

$$= 0.06 + 0.06 + 0.10 + 0.10 + 0.23 = \underline{\underline{0.55}} \quad (4.41)$$

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

$$250 \cdot P(X \geq 40) = 250 \cdot 0.55 = \underline{\underline{137.5}} \quad (4.42)$$

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

$$\underline{\underline{P(25 \leq X \leq 35)}} = P(X = 25) + P(X = 30) + P(X = 35) \quad (4.43)$$

$$= 0.03 + 0.04 + 0.04 = \underline{\underline{0.11}} \quad (4.44)$$

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 $E[\bar{X}]$:

$$\underline{\underline{E[\bar{X}]}} = E\left[\frac{1}{n}\left(X_1 + X_2 + \dots + X_n\right)\right] \quad (4.45)$$

$$\stackrel{\text{alltid}}{=} \frac{1}{n}\left(E[X_1] + E[X_2] + \dots + E[X_n]\right) \quad (4.46)$$

$$= \frac{1}{n}\left(\underbrace{E[X] + E[X] + \dots + E[X]}_{= n \cdot E[X]}\right) \quad (4.47)$$

$$= \frac{1}{n} \cdot n \cdot E[X] = E[X] = \underline{\underline{35}} \quad (4.48)$$

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

f) Tolkning:

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

■

⁹Legg merke til at vi i lign.(4.46) bruker regnereglen: $(a, b$ er konstanter)

$$E[aX + bY] = aE[X] + bE[Y], \quad (4.49)$$

som man finner i formelsamlingen.

Kapittel 5

LØSNING: Eksamen 9. mai 2014

“MAT110 Statistikk 1”

Oppgave 1: (revisjon)

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

$$\underline{\underline{P(\overline{\text{objekt}}|\overline{K})}} = 1 - \overbrace{P(\text{objekt}|\overline{K})}^{=0.05} = 1 - 0.05 = \underline{\underline{0.95}} \quad (5.1)$$

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

$$\underline{\underline{P(\text{objekt})}} = \overbrace{P(\text{objekt}|\overline{K})}^{=0.05} \overbrace{P(\overline{K})}^{=1-P(K)=0.999} + \overbrace{P(\text{objekt}|K)}^{=0.80} \overbrace{P(K)}^{=0.001} \quad (5.2)$$

$$= 0.05 \cdot 0.999 + 0.80 \cdot 0.001 = \underline{\underline{0.05075}} \quad (5.3)$$

c) Bayes formel:

$$\underline{\underline{P(\bar{K}|\text{objekt})}} = \overbrace{P(\text{objekt}|\bar{K})}^{=0.05} \cdot \frac{\overbrace{P(\bar{K})}^{=0.999}}{\underbrace{P(\text{objekt})}_{=0.05075}} = 0.05 \cdot \frac{0.999}{0.05075} = \underline{\underline{0.9842}} \quad (5.4)$$

d) Kommentar:

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



¹Merknad: (behøver ikke være med i eksamensbesvarelsen)

Når det gjelder hva som gir riktige og hva som gir falske indiksjoner så bør man være presis:

$P(\text{objekt}|\bar{K}) = 0.05$ betyr at verktøyet gir 5 % falske prediksjoner for bedrifter som *ikke* går konkurs.

$P(\bar{K}|\text{objekt}) = 0.9842$ 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

$$\underline{\text{Poissonfordeling}} \quad (5.5)$$

- 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:

$$\underline{\lambda} = 920 \cdot \frac{0.5 \text{ time}}{4 \text{ timer}} = \underline{115}, \quad \text{q.e.d.} \quad (5.6)$$

- c) i) Fra oppgave **2a** og **2b** 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

$$\underline{E[X]} = \lambda = \underline{115} \quad (5.7)$$

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

$$\underline{\sigma[X]} = \sqrt{\text{Var}[X]} = \sqrt{\lambda} = \sqrt{115} = \underline{10.72} \quad (5.8)$$

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

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

- d) i) En diskret Poisson sannsynlighetsfordelingen $\text{Poi}[\lambda]$ kan, under bestemte betingelser, med god tilnærming beskrives av en normalfordeling:

$$\text{Poi}[\lambda] \longrightarrow N[\mu = \lambda, \sigma = \sqrt{\lambda}] \quad (5.9)$$

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

$$\lambda \gtrsim 5, \quad (5.10)$$

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

- iii) Siden $\lambda = 115 \gg 5$ så er betingelsen i lign.(5.10) oppfylt og tilnærmelsen i lign.(5.9) gjelder i vårt tilfelle.

- e) Ved å bruke resultatet fra oppgave **2d 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.)

$$\underline{P(X > X_0 - 20)} = P(X > 125 - 20) \quad (5.11)$$

$$= 1 - P(X \leq 105) \quad (5.12)$$

$$\stackrel{\text{standardiser}}{=} 1 - P\left(\underbrace{\frac{X - E[X]}{\sigma[X]}}_{= Z} \leq \frac{105 - E[X]}{\sigma[X]}\right) \quad (5.13)$$

$$= 1 - P\left(Z \leq \frac{105 - \lambda}{\sqrt{\lambda}}\right) \quad (5.14)$$

$$= 1 - P\left(Z \leq \frac{105 - 115}{\sqrt{115}}\right) \quad (5.15)$$

$$= 1 - P(Z \leq -0.93) \quad (5.16)$$

$$= 1 - \left(1 - P(Z \leq 0.93)\right) = \underline{P(Z \leq 0.93)} \quad (5.17)$$

Dermed:

$$\underline{\underline{P(X > X_0 - 20)}} = P(Z \leq 0.93) \quad (5.18)$$

$$= \underbrace{G(0.93)}_{\text{se tabell}} = \underline{\underline{0.8238}} \quad (5.19)$$

Dersom det i utgangspunktet er 20 biler i fergekø i tillegg de til som ankommer fergekaien med konstant rate λ så er det 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:

$$P(X \leq X_{\text{kap}}) = 0.95 \quad (5.20)$$

$$P\left(\underbrace{\frac{X - E[X]}{\sigma[X]}}_{\equiv Z} \leq \underbrace{\frac{X_{\text{kap}} - E[X]}{\sigma[X]}}_{\equiv Z_{\text{kap}}}\right) \stackrel{\text{standardiser}}{=} 0.95 \quad (5.21)$$

$$P(Z \leq Z_{\text{kap}}) = 0.95 \quad (5.22)$$

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:

$$Z_{\text{kap}} = 1.645 \quad (5.23)$$

Løser med hensyn på X_{kap} alene:

$$Z_{\text{kap}} = \frac{X_{\text{kap}} - E[X]}{\sigma[X]} \quad (5.24)$$

$$\underline{\underline{X_{\text{kap}}}} = Z_{\text{kap}} \cdot \overbrace{\sigma[X]}^{=\sqrt{\lambda}} + \overbrace{E[X]}{=\lambda} \quad (5.25)$$

$$= Z_{\text{kap}} \cdot \sqrt{\lambda} + \lambda \quad (5.26)$$

$$= 1.645 \cdot \sqrt{115} + 115 = 132.64 \approx \underline{\underline{133}} \quad (5.27)$$

For at det skal være 95 % sikkert at alle i fergekøen skal komme med fergen så må fergen ha en kapasitet på 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 2d 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. $Y \sim \text{Bin}[n = 3\lambda = 345, p = 0.11]$.

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:

$$\underline{\underline{E[Y]}} \stackrel{\text{Bin.}}{=} n \cdot p = 345 \cdot 0.11 = \underline{\underline{37.95}} \quad (5.28)$$

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

$$\underline{\underline{\text{Var}[Y]}} \stackrel{\text{Bin.}}{=} n \cdot p \cdot (1 - p) = 345 \cdot 0.11 \cdot (1 - 0.11) \approx \underline{\underline{33.78}} \quad (5.29)$$

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

$$\underline{\underline{\sigma[X]}} = \sqrt{\text{Var}[X]} = \sqrt{33.78} = \underline{\underline{5.81}} \quad (5.30)$$

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

$$\underline{\underline{n \cdot p(1-p) \gtrsim 5}} \quad (5.31)$$

- ii) For vårt tilfelle:

$$345 \cdot 0.11(1 - 0.11) = 33.78 \gg 5 \quad (5.32)$$

Ja, betingelsen er godt oppfylt for vårt tilfelle.

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

$$\underline{\underline{P(Y > 45)}} = 1 - P(Y \leq 45) \quad (5.33)$$

$$= 1 - P\left(\underbrace{\frac{Y - E[Y]}{\sigma[Y]}}_{= Z} \leq \frac{45 - E[Y]}{\sigma[Y]}\right) \quad (5.34)$$

$$= 1 - P\left(Z \leq \frac{45 - 38}{5.81}\right) \quad (5.35)$$

$$= 1 - P(Z \leq 1.21) \quad (5.36)$$

$$= 1 - \underbrace{G(1.21)}_{\substack{\text{tabell} \\ 0.8869}} \quad (5.37)$$

$$= 1 - 0.8869 = \underline{\underline{0.1131}} \quad (5.38)$$

⁴Se side 60 i formelsamlingen.

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

$$\underline{\underline{P(B > 1)}} = P(B = 2) + P(B = 3) \quad (5.39)$$

$$= 0.02 + 0.01 = \underline{\underline{0.03}} \quad (5.40)$$

Alternativt kan denne løses på følgende måte:

$$\underline{\underline{P(B > 1)}} = 1 - P(B \leq 1) \quad (5.41)$$

$$= 1 - \left(P(B = 0) + P(B = 1) \right) \quad (5.42)$$

$$= 1 - 0.93 - 0.04 = \underline{\underline{0.03}} \quad (5.43)$$

Det er nok at oppgaven løses på en måte.

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

$$\underline{\underline{E[B_{tot}]}} = E[B_1 + B_2 + \dots + B_n] \quad (5.44)$$

$$\stackrel{\text{alltid}}{=} \underbrace{E[B_1] + E[B_2] + \dots + E[B_n]}_{n = 345} = n \cdot \underbrace{E[B]}_{=0.11} = 345 \cdot 0.11 = \underline{\underline{37.95}} \quad (5.45)$$

NB: Overgangen i lign.(5.44) til (5.45) 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: ($n = 3\lambda = 345$)

$$\underline{\underline{Var[B_{tot}]}} = Var[B_1 + B_2 + \dots + B_n] \quad (5.46)$$

$$\stackrel{\text{uavhengig}}{=} Var[B_1] + Var[B_2] + \dots + Var[B_n] \quad (5.47)$$

$$= \underbrace{n \cdot Var[B]}_{= \sigma^2[B]} = n \cdot \sigma^2[B] = 345 \cdot 0.4449^2 \approx \underline{\underline{68.28}} \quad (5.48)$$

NB: Overgangen i lign.(5.46) til (5.47) 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 $i = 1, 2, 3, \dots, n$

2. alle bruskjøpere har samme sannsynlighetsfordeling for B_i :

$B_i \sim$ samme sannsynlighetsfordeling for alle $i = 1, 2, 3, \dots, n$

3. antall “forsøk”, dvs. potensielle bruskjøpere, $n = 3\lambda = 345$, er **tilstrekkelig stort** ⁵

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

$$\underline{\underline{B_{tot} \sim N[E[B_{tot}], Var[B_{tot}]]}} \quad (5.49)$$

⁵Husk: Antall forsøk n for at sentralgrensesetningen skal gjelde er avhengig av situasjonen. Men en **tommel-fingerregel** 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 **normal**fordeling:

$$\underline{\underline{P(B_{\text{tot}} > 45)}} = 1 - P(B_{\text{tot}} \leq 45) \quad (5.50)$$

$$= 1 - P\left(\underbrace{\frac{B_{\text{tot}} - E[B_{\text{tot}}]}{\sigma[B_{\text{tot}}]}}_{= Z_{\text{tot}}} \leq \frac{45 - E[B_{\text{tot}}]}{\sigma[B_{\text{tot}}]}\right) \quad (5.51)$$

$$= 1 - P\left(Z_{\text{tot}} \leq \frac{45 - 37.95}{\sqrt{68.28}}\right) \quad (5.52)$$

$$= 1 - P(Z_{\text{tot}} \leq 0.85) \quad (5.53)$$

$$= 1 - \underbrace{G(0.85)}_{\text{se tabell}} \quad (5.54)$$

$$= 1 - 0.8023 = \underline{\underline{0.1977}} \quad (5.55)$$

- 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:

$$\underbrace{Var[B_{\text{tot}}] = 68.28}_{\text{modell 2}} > \underbrace{Var[Y] = 33.78}_{\text{modell 1}} \quad (5.56)$$

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

$$\underline{E[F]} = E[(p - k)B_{\text{tot}} - k_{\text{fast}}] = \underline{(p - k)E[B_{\text{tot}}] - k_{\text{fast}}} \quad (5.57)$$

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

Dermed:

$$\underline{\underline{E[F]}} = ((20 - 6) \cdot 37.95 - 175) \text{ NOK} = \underline{\underline{356.30 \text{ NOK}}} \quad (5.58)$$

■

⁶Bruk 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)

$$\hat{\beta} = \frac{S_{xy}}{S_x^2} = \frac{18}{2.5} = \underline{7.2} \quad (5.59)$$

$$\hat{\alpha} = \bar{y} - \beta\bar{x} = 284 - 7.2 \cdot 3 = \underline{262.4} \quad (5.60)$$

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

$$\underline{\underline{\hat{y} = 262.4 + 7.2 \cdot x}} \quad (5.61)$$

- 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 Norwegian øker med 7.2 NOK per dag.

- c) Regresjonslinjen i lign.(5.61) predikerer at aksjekursen til Norwegian er:

$$\underline{\underline{\hat{y}(12) = (262.4 + 7.2 \cdot 12) \text{ NOK} = 348.8 \text{ NOK}}} \quad (5.62)$$

etter 12 dager.

d) 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 observasjonene y_i

Det er nok at man nevner én 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

$$\underline{\underline{R^2 = 0.9744}} \quad (5.63)$$

f) Kommentar til svaret i oppgave 4e:

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 forklaringskraft.

■

⁷Man 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

LØSNING: Eksamen 8. januar 2015

“MAT110 Statistikk 1”

Oppgave 1: (økonomi)

a) Matematisk formulering ihht. notasjon:

$$P(G) = 0.75 \quad (6.1)$$

$$P(D) = 0.25 \quad (6.2)$$

b) Matematisk formulering ihht. notasjon:

$$P(L|G) = 0.10 \quad (6.3)$$

$$P(L|D) = 0.50 \quad (6.4)$$

c) Vi skal finne $P(L)$ når vi kjenner sannsynlighetene i oppgave **1a** og **1b**.
Da kan vi bruke oppsplitting av utfallsrom Ω :

$$\underline{P(L)} \stackrel{\text{oppspl.}}{=} P(L|G)P(G) + P(L|D)P(D) \quad (6.5)$$

$$= 0.10 \cdot 0.75 + 0.50 \cdot 0.25 = \underline{0.20} \quad (6.6)$$

Det er 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**:

$$\underline{P(D|L)} \stackrel{\text{Baye}}{=} P(L|D) \frac{P(D)}{P(L)} \quad (6.7)$$

$$= 0.50 \cdot \frac{0.25}{0.20} = \underline{0.625} \quad (6.8)$$

Sannsynligheten for at en kunde med lav inntekt har **dårlig** betalingsevne er 62.5 %.

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

$$\underline{P(G|L)} \stackrel{\text{komp.}}{=} 1 - P(D|L) \quad (6.9)$$

$$= 1 - 0.625 = \underline{0.375} \quad (6.10)$$

Sannsynligheten for at en kunde med lav inntekt har **god** betalingsevne er 37.5 %.

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

$$\underline{E[F]} = \sum_{X=G,D} f_X \cdot P(X|L) \quad (6.11)$$

$$= f_G \cdot P(G|L) + f_D \cdot P(D|L) \quad (6.12)$$

$$= (6000 \cdot 0.375 - 4000 \cdot 0.625) \text{ NOK} = \underline{-250 \text{ NOK}} \quad (6.13)$$

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

$$\underline{E[F]} = 6000 \cdot P(G|L) - 4000 \cdot P(D|L) \quad (6.14)$$

$$= (6000 \cdot 0.375 - 4000 \cdot 0.625) \text{ NOK} = \underline{-250 \text{ NOK}} \quad (6.15)$$

- 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 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:

$$\underline{P(X = 1)} = \frac{\lambda^1}{1!} e^{-\lambda} \quad (6.16)$$

$$\stackrel{\lambda=0.6}{=} \frac{0.6^1}{1!} e^{-0.6} \stackrel{\text{kalkis}}{=} \underline{0.3293} \quad (6.17)$$

Sannsynligheten for at det skjer 1 “hendelse” per uke er 32.93%.

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

$$\underline{P(X > 1)} = 1 - P(X \leq 1) \quad (6.18)$$

$$= 1 - \left(P(X = 0) + P(X = 1) \right) \quad (6.19)$$

$$= 1 - \frac{\lambda^0}{0!} e^{-\lambda} - \frac{\lambda^1}{1!} e^{-\lambda} \quad (6.20)$$

$$\stackrel{\lambda=0.6}{=} 1 - \frac{0.6^0}{0!} e^{-0.6} - \frac{0.6^1}{1!} e^{-0.6} \quad (6.21)$$

$$\stackrel{\text{kalkis}}{=} 1 - 0.5488 - 0.3293 \quad (6.22)$$

$$= \underline{0.1219} \quad (6.23)$$

Sannsynligheten for at det skjer mer enn 1 “hendelse” per uke er 12.19%.

c) Forventet antall “hendelser” per måned:

$$\underline{E[Y]} = E[X_1 + X_2 + X_3 + X_4] \quad (6.24)$$

$$\stackrel{\text{alltid}}{=} E[X_1] + E[X_2] + E[X_3] + E[X_3] = \overbrace{\lambda + \lambda + \lambda + \lambda}^{4 \text{ stk.}} = 4\lambda \quad (6.25)$$

$$= 4 \cdot 0.6 = \underline{2.4} \quad (6.26)$$

NB: Overgangen i lign.(6.24) til (6.25) 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:

$$\underline{P(Y > 1)} = 1 - P(Y \leq 1) \quad (6.27)$$

$$= 1 - \left(P(Y = 0) + P(Y = 1) \right) \quad (6.28)$$

$$= 1 - \frac{\lambda_Y^0}{0!} e^{-\lambda_Y} - \frac{\lambda_Y^1}{1!} e^{-\lambda_Y} \quad (6.29)$$

$$\stackrel{\lambda_Y=2.4}{=} 1 - \frac{2.4^0}{0!} e^{-2.4} - \frac{2.4^1}{1!} e^{-2.4} \quad (6.30)$$

$$\stackrel{\text{kalkis}}{=} 1 - 0.091 - 0.2177 \quad (6.31)$$

$$= \underline{0.6916} \quad (6.32)$$

e) Siden

$$P(Y > 1) = 0.6916 \quad \text{og} \quad P(X > 1) = 0.1219 \quad (6.33)$$

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:

$$P(Y > 1) \gg P(X > 1) \quad (6.34)$$

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

$$\sum_{i=0}^n p_i = p_0 + p_1 + p_2 = 0.95 + 0.03 + 0.02 = \underline{\underline{1}} \quad (6.35)$$

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

b) Forventning $E[X]$:

$$\underline{\underline{E[X]}} \stackrel{\text{def.}}{=} \sum_{i=0}^3 x_i \cdot P(X = x_i) \quad (6.36)$$

$$= 0 \cdot \underbrace{P(X=0)}_{=0.95} + 1 \cdot \underbrace{P(X=1)}_{=0.03} + 2 \cdot \underbrace{P(X=2)}_{=0.02} \quad (6.37)$$

$$= 0 \cdot 0.95 + 1 \cdot 0.03 + 2 \cdot 0.02 = \underline{\underline{0.07}} \quad (6.38)$$

Tolking:

$E[X] =$ forventet antall produksjonsfeil for en tilfeldig valgt ananasbrus

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

$$\underline{\underline{E[X^2]}} \stackrel{\text{def.}}{=} \sum_{i=0}^3 x_i^2 \cdot P(X = x_i) \quad (6.39)$$

$$= 0^2 \cdot \underbrace{P(X=0)}_{=0.95} + 1^2 \cdot \underbrace{P(X=1)}_{=0.03} + 2^2 \cdot \underbrace{P(X=2)}_{=0.02} \quad (6.40)$$

$$= 0^2 \cdot 0.95 + 1^2 \cdot 0.03 + 2^2 \cdot 0.02 = \underline{\underline{0.11}} \quad (6.41)$$

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

$$\underline{\underline{Var[X]}} = E[X^2] - E[X]^2 = 0.11 - 0.07^2 = \underline{\underline{0.1051}} \quad (6.42)$$

Tolking:

$$\underline{\underline{Var[X] = \text{forventet variasjon/spredning i antall produksjonsfeil} \\ \text{for en tilfeldig valgt ananasbrus}}} \quad (6.43)$$

d) Forventet inntekt per dag på ananasbrus:

$$\underline{E[I]} = E[pn \cdot 0.95 - bnX] \quad (6.44)$$

$$= pn \cdot 0.95 - bn \underbrace{E[X]}_{= 0.07} \quad (6.45)$$

$$= (15 \cdot 4500 \cdot 0.95 - 23 \cdot 4500 \cdot 0.07) \text{ NOK} = \underline{\underline{56\,880 \text{ NOK}}} \quad (6.46)$$

Forventet inntekt per dag på ananasbrus er 56 880 NOK.

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

$$\underline{\underline{E[Y]}} = E[X_1 + X_2 + \dots + X_n] = \overbrace{E[X_1] + E[X_2] + \dots + E[X_n]}^{n \text{ stk.}} \quad (6.47)$$

$$= nE[X] = 4500 \cdot 0.07 = \underline{\underline{315}} \quad (6.48)$$

NB: Overgangen i lign.(6.47) gjelder alltid. Uansett om de stokastiske variablene X_i er uavhengige eller ikke.

¹Siden 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:

$$\underline{\underline{Var[Y]}} = Var[X_1 + X_2 + \dots + X_n] \quad (6.49)$$

$$\stackrel{\text{uavh.}}{=} \overbrace{Var[X_1] + Var[X_2] + \dots + Var[X_n]}^{n \text{ stk.}} \quad (6.50)$$

$$= n Var[X] = 4500 \cdot 0.1051 = \underline{\underline{472.95}} \quad (6.51)$$

NB: Overgangen i lign.(6.49) 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 variabelen Y **normalfordelt**:

$$\underline{\underline{Y}} \sim N[E[Y], Var[Y]] \quad (6.52)$$

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

$$\underline{\underline{n}} \gtrsim 30 \quad (6.53)$$

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

$$\underline{P(Y > 350)} \stackrel{\text{standardiser}}{=} P\left(\underbrace{\frac{Y - E[Y]}{\sigma[Y]}}_{= Z} > \frac{350 - E[X]}{\sigma[X]}\right) \quad (6.54)$$

$$= P\left(Z > \frac{350 - 315}{\sqrt{472.95}}\right) \quad (6.55)$$

$$= P(Z > 1.61) \quad (6.56)$$

$$= 1 - P(Z \leq 1.61) \quad (6.57)$$

$$= 1 - \underbrace{G(1.61)}_{= 0.9463} \quad (6.58)$$

$$\stackrel{\text{tabell}}{=} 1 - 0.9463 \quad (6.59)$$

$$= \underline{0.0537} \quad (6.60)$$

Sannsynligheten for at antall produksjonsfeil når et uakseptabelt nivå, dvs. mer enn 350, er 5.37 %.

■

²I 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 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ være observasjonspar/datasett. Minste kvadraters lineære **regresjonslinje** er da gitt ved:

$$\hat{y} = \hat{\alpha} + \hat{\beta}x, \quad (6.61)$$

hvor

$$\hat{\beta} = \frac{S_{xy}}{S_x^2} \quad (6.62)$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad (6.63)$$

og

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad S_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (6.64)$$

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

c) i) Forklaringskraften R^2 er normalisert 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 observasjonene y_i

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

³Se side 66 i formelsamlingen fra 2014.

d) Gjennomsnittene x og y er:

$$\underline{\bar{x}} = \frac{1}{n} \sum_{i=1}^n x_i \quad (6.65)$$

$$= \frac{1}{5} (2009 + 2010 + 2011 + 2012 + 2013) = \underline{\underline{2011}} \quad (6.66)$$

$$\underline{\bar{y}} = \frac{1}{n} \sum_{i=1}^n y_i \quad (6.67)$$

$$= \frac{1}{5} (228 + 240 + 245 + 248 + 258) \text{ tusen tonn} = \underline{\underline{243.8 \text{ tusen tonn}}} \quad (6.68)$$

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)

$$\underline{\hat{\beta}} = \frac{S_{xy}}{S_x^2} = \frac{17}{2.5} = \underline{\underline{6.8}} \quad (6.69a)$$

$$\underline{\hat{\alpha}} = \bar{y} - \beta\bar{x} = 243.8 - 6.8 \cdot 2011 = \underline{\underline{-13\,431}} \quad (6.69b)$$

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

$$\underline{\underline{\hat{y} = -13\,431 + 6.8 \cdot x}} \quad (6.70)$$

e) Bruker regresjonslinjen $\hat{y} = \hat{y}(x)$ fra oppgave **d**, dvs. lign.(6.70):

$$\underline{\hat{y}(2017)} = -13\,431 + 6.8 \cdot 2017 = \underline{\underline{284.6}} \quad (6.71)$$

Regresjonslinjen predikerer at etterspørselen av slurry vil være 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 en enhet. Siden (se lign.(6.69a))

$$\hat{\beta} = 6.8 \quad (6.72)$$

så innser vi at den gjennomsnittlige etterspørselen av slurry øker med 6.8 tusen tonn i året.

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

$$305 = -13\,431 + 6.8 \cdot x \quad (6.73)$$

og løser med hensyn på x :

$$\underline{x} = \frac{305 + 13\,431}{6.8} = \underline{2020} \quad (6.74)$$

Firmaet må ha større tank i 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

$$\underline{\underline{R^2 = 0.9538}} \quad (6.75)$$

- i) Kommentar til svaret i oppgave 4h:

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 stor forklaringskraft.

■

⁴Man 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

LØSNING: Eksamen 28. mai 2015

“MAT110 Statistikk 1”

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 forventningen:

$$E[X] = n \frac{M}{N} \quad (7.1)$$

Løser med hensyn på M alene og får:

$$\underline{\underline{M}} = \frac{E[X]N}{n} = \frac{10 \cdot 100\,000}{1000} = \underline{\underline{1000}} \quad , \quad \text{q.e.d.} \quad (7.2)$$

c) Siden X er hypergeometrisk fordelt så er variansen:

$$\underline{Var[X]} = \frac{N-n}{N-1} n \frac{M}{N} \left(1 - \frac{M}{N}\right) \quad (7.3)$$

$$= \frac{100\,000 - 1000}{100\,000 - 1} 1000 \frac{1000}{100\,000} \left(1 - \frac{1000}{100\,000}\right) = \underline{9.8} \quad (7.4)$$

og tilhørende standardavvik:

$$\underline{\underline{\sigma[X]}} = \sqrt{Var[X]} = \sqrt{9.8} = \underline{\underline{3.13}} \quad , \quad \text{q.e.d.} \quad (7.5)$$

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

$$N \gtrsim 20 \cdot n \quad (7.6)$$

$$n \frac{M}{N} \left(1 - \frac{M}{N}\right) \gtrsim 5 \quad (7.7)$$

ii) For vårt tilfelle er:

$$100\,000 \gtrsim 20 \cdot 1000 \quad (7.8)$$

$$1000 \frac{1000}{100\,000} \left(1 - \frac{1000}{100\,000}\right) \gtrsim 5 \quad (7.9)$$

som gir

$$100\,000 \gtrsim 20\,000 \quad (7.10)$$

$$9.9 \gtrsim 5 \quad (7.11)$$

dvs. kriteriene er oppfylte i vårt tilfelle.

e) og g) Se vedlegg.

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

$$P(X \geq g) = 0.10 \quad (7.12)$$

$$P\left(\underbrace{\frac{X - E[X]}{\sigma[X]}}_{\equiv Z} \geq \underbrace{\frac{g - E[X]}{\sigma[X]}}_{\equiv Z_0}\right) \stackrel{\text{standardiser}}{=} 0.10 \quad (7.13)$$

$$P(Z \geq Z_0) = 0.10 \quad (7.14)$$

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

$$P(Z \geq Z_0) = 0.10 \quad (7.15)$$

$$1 - P(Z < Z_0) = 0.10 \quad (7.16)$$

$$P(Z < Z_0) = 0.90 \quad (7.17)$$

Ved “[omvendt tabelloppslag](#)” ser vi at Z_0 må være 1.28: ²

$$Z_0 = 1.28 \quad (7.19)$$

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

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

$$P(Z \leq Z_0) = P(Z < Z_0) + \underbrace{P(Z = Z_0)}_{=0} = P(Z < Z_0) \quad (7.18)$$

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 $P(Z < Z_0)$.

Vi løser

$$Z_0 = \frac{g - E[X]}{\sigma[X]} \quad (7.20)$$

med hensyn på g : ($E[X] = 10$ og $\sigma[X] = 3.13$ fra oppgave **1c**)

$$\underline{g} = E[X] + Z_0 \cdot \sigma[X] \quad (7.21)$$

$$= 10 + 1.28 \cdot 3.13 = \underline{\underline{14}} \quad (7.22)$$

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



Oppgave 2: (økonomi og 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:

$$\underline{P(D_1 > 21\,000)} \stackrel{\text{standardiser}}{=} P\left(\underbrace{\frac{D_1 - \mu_1}{\sigma_1}}_{= Z} > \frac{21\,000 - \mu_1}{\sigma_1}\right) \quad (7.23)$$

$$= P\left(Z > \frac{21\,000 - 15\,000}{3000}\right) \quad (7.24)$$

$$= P(Z > 2) \quad (7.25)$$

$$= 1 - P(Z \leq 2) \quad (7.26)$$

$$= 1 - \underbrace{G(2)}_{= 0.9772} \quad (7.27)$$

$$\stackrel{\text{tabell}}{=} 1 - 0.9772 \quad (7.28)$$

$$= \underline{0.0228} \quad (7.29)$$

Sannsynligheten for at mer enn 21 000 trofaste aviskjøpere kjøper Dagbladet en gitt dag er 2.28 %.

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

$$D_{tot} = D_1 + D_2 \quad (7.30)$$

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

$$\underline{\underline{E[D_{tot}]}} = E[D_1 + D_2] \quad (7.31)$$

$$\stackrel{\text{alltid}}{=} \underbrace{E[D_1]}_{=\mu_1} + \underbrace{E[D_2]}_{=\mu_2} = \mu_1 + \mu_2 = 15\,000 + 18\,000 = \underline{\underline{33\,000}} \quad (7.32)$$

NB: Overgangen i lign.(7.31) til (7.32) 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:

$$\underline{\underline{Var[D_{tot}]}} = Var[D_1 + D_2] \quad (7.33)$$

$$\stackrel{\text{uavh.}}{=} Var[D_1] + Var[D_2] \quad (7.34)$$

$$= \sigma_1^2 + \sigma_2^2 = 3000^2 + 4000^2 = \underline{\underline{25\,000\,000}} \quad (7.35)$$

NB: Overgangen fra lign.(7.33) til lign.(7.34) gjelder *kun* dersom de stokastiske variablene D_1 og D_2 er uavhengige.

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

$$\underline{E[F]} = E[(p - k)D_{tot} - c] \quad (7.36)$$

$$= (p - k)E[D_{tot}] - c = \left((25 - 12) \cdot 33\,000 - 400\,000 \right) \text{ NOK} \quad (7.37)$$

$$= \underline{29\,000 \text{ NOK}} \quad (7.38)$$

Her har vi brukt at $E[D_{tot}] = 33\,000$ fra oppgave **2b**, se lign.(7.32).

- 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 $\sigma[D_{tot}] = \sqrt{\text{Var}[D_{tot}]} = \sqrt{25\,000\,000} = 5000$ fra oppgave **2c**)

$$\underline{P(D_{tot} > 21\,000|S)} \stackrel{\text{standardiser}}{=} P\left(\underbrace{\frac{D_{tot} - E[D_{tot}]}{\sigma[D_{tot}]}}_{= Z_{tot}} \mid S > \frac{21\,000 - 33\,000}{5000}\right) \quad (7.39)$$

$$= P\left(Z_{tot} > \frac{21\,000 - 33\,000}{5000}\right) \quad (7.40)$$

$$= P(Z_{tot} > -2.4) \quad (7.41)$$

$$= 1 - P(Z_{tot} \leq -2.4) \quad (7.42)$$

$$= 1 - \left(1 - P(Z_{tot} \leq 2.4)\right) \quad (7.43)$$

$$= P(Z_{tot} \leq 2.4) \quad (7.44)$$

$$\stackrel{\text{tabell}}{=} G(2.4) \quad (7.45)$$

$$= \underline{0.9918} \quad (7.46)$$

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

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

$$\underline{P(D_{tot} > 21\,000)} \stackrel{\text{oppspl.}}{=} \underbrace{P(D_{tot} > 21\,000|S)}_{=0.9918} \cdot \overbrace{P(S)}^{=0.20} + \underbrace{P(D_{tot} > 21\,000|\bar{S})}_{=P(D_1 > 21\,000) = 0.028} \cdot \overbrace{P(\bar{S})}^{=1-0.20} \quad (7.47)$$

$$= 0.9918 \cdot 0.20 + 0.0228 \cdot (1 - 0.20) = \underline{0.2166} \quad (7.48)$$

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 21.66%.

■

³Se side 32 i formelsamlingen 2015.

Oppgave 3: (petroleumslogistikk)

a) Siden

$$\sum_{i=0}^n P(X = x_i) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) \quad (7.49)$$

$$= 0.09 + 0.28 + 0.41 + 0.17 + 0.05 = \underline{1} \quad (7.50)$$

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

$$\underline{P(X \geq 3)} = P(X = 3) + P(X = 4) = 0.17 + 0.05 = \underline{0.22} \quad (7.51)$$

c) i) Forventet antall innleverte rapporter per dag:

$$\underline{E[X]} \stackrel{\text{def.}}{=} \sum_{i=0}^4 x_i \cdot P(X = x_i) \quad (7.52)$$

$$\begin{aligned} &= 0 \cdot \underbrace{P(X = 0)}_{=0.09} + 1 \cdot \underbrace{P(X = 1)}_{=0.28} + 2 \cdot \underbrace{P(X = 2)}_{=0.41} + 3 \cdot \underbrace{P(X = 3)}_{=0.17} + 4 \cdot \underbrace{P(X = 4)}_{=0.05} \\ &= 0 \cdot 0.09 + 1 \cdot 0.28 + 2 \cdot 0.41 + 3 \cdot 0.17 + 4 \cdot 0.05 = \underline{1.81} \quad (7.53) \end{aligned}$$

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

$$\underline{E[X^2]} \stackrel{\text{def.}}{=} \sum_{i=0}^3 x_i^2 \cdot P(X = x_i) \quad (7.54)$$

$$\begin{aligned} &= 0^2 \cdot \underbrace{P(X = 0)}_{=0.09} + 1^2 \cdot \underbrace{P(X = 1)}_{=0.28} + 2^2 \cdot \underbrace{P(X = 2)}_{=0.41} + 3^2 \cdot \underbrace{P(X = 3)}_{=0.17} + 4^2 \cdot \underbrace{P(X = 4)}_{=0.05} \\ &= 0^2 \cdot 0.09 + 1^2 \cdot 0.28 + 2^2 \cdot 0.41 + 3^2 \cdot 0.17 + 4^2 \cdot 0.05 = \underline{4.25} \quad (7.55) \end{aligned}$$

Dette innsatt i setningen for ‘*varianssetningen*’:⁴

$$\underline{Var[X]} = E[X^2] - E[X]^2 = 4.25 - 1.81^2 = \underline{0.9739} \quad (7.56)$$

d) i) Tolkning:

$E[\bar{X}] =$ forventet antall innleverte rapporter per dag i gjennomsnitt over et helt år

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

$$\underline{E[\bar{X}]} = E\left[\frac{1}{n} \left(X_1 + X_2 + \dots + X_n \right)\right] \quad (7.57)$$

$$\stackrel{\text{alltid}}{=} \frac{1}{n} \left(E[X_1] + E[X_2] + \dots + E[X_n] \right) \quad (7.58)$$

$$= \frac{1}{n} \left(\underbrace{E[X] + E[X] + \dots + E[X]}_{=n \cdot E[X]} \right) = \frac{1}{n} \cdot \underbrace{E[X]}_{=1.81} = \underline{1.81} \quad (7.59)$$

NB: Overgangen i lign.(7.57) til (7.58) gjelder **alltid**. Uansett om de stokastiske variablene X_i er uavhengige eller ikke.

⁴Se side 39 i formelsamlingen fra 2015.

e) i) Tolkning:

$$\underline{\underline{Var[\bar{X}]}} = \text{forventet variasjon/spredning i antall innleverte rapporter per dag} \\ \underline{\underline{\text{i gjennomsnitt over et helt \u00e5r}}} \quad (7.60)$$

ii) Variansen til antall innleverte rapporter per dag i gjennomsnitt over et \u00e5r: ($n = 365$)

$$\underline{\underline{Var[\bar{X}]}} = Var \left[\frac{1}{n} \left(X_1 + X_2 + \dots + X_n \right) \right] \quad (7.61)$$

$$\stackrel{\text{uavh.}}{=} \frac{1}{n^2} \left(\underbrace{Var[X_1] + Var[X_2] + \dots + Var[X_n]}_{n \cdot Var[X]} \right) \quad (7.62)$$

$$= \frac{1}{n^2} \cdot \underbrace{Var[X]}_{=0.9739} = \frac{0.9739}{365} = \underline{\underline{0.002668}} \quad (7.63)$$

NB: Overgangen i lign.(7.61) til (7.62) 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 $i = 1, 2, 3, \dots, n$
2. alle X_i har samme sannsynlighetsfordeling: (oppgitt i oppgaveteksten)
 $X_i \sim$ samme sannsynlighetsfordeling for alle $i = 1, 2, 3, \dots, n$
3. antall "fors\u00f8k", dvs. antall dager, $n = 365$ er tilstrekkelig stort ⁵

s\u00e5 gjelder **sentralgrensesetningen**. Dermed er \bar{X} normalfordelt.

⁵Husk: Antall fors\u00f8k n for at sentralgrensesetningen skal gjelde er avhengig av situasjonen. Men en **tommel-fingerregel** er at vi b\u00f8r ha $n \gtrsim 30$.

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

$$\underbrace{E[\bar{X}]}_{= 1.81} = \underbrace{E[X]}_{= 1.81} \quad (7.64)$$

og at

$$\underbrace{Var[\bar{X}]}_{= 0.002668} \ll \underbrace{Var[X]}_{= 0.9739} \quad (7.65)$$

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

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

$$\underline{\underline{P\left(X_1 + X_2 + \dots + X_n > 700\right)}} = P\left(\frac{X_1 + X_2 + \dots + X_n}{n} > \frac{700}{n}\right) \quad (7.66)$$

$$= P\left(\bar{X} > \frac{700}{n}\right) \quad (7.67)$$

$$= 1 - P\left(\bar{X} \leq \frac{700}{n}\right) \quad (7.68)$$

$$\stackrel{\text{standardiser}}{=} 1 - P\left(\underbrace{\frac{\bar{X} - E[\bar{X}]}{\sigma[\bar{X}]}}_{=\bar{Z}} \leq \frac{\frac{700}{n} - E[\bar{X}]}{\sigma[\bar{X}]}\right) \quad (7.69)$$

$$= 1 - P\left(\bar{Z} \leq \frac{\frac{700}{365} - 1.81}{\sqrt{0.002668}}\right) \quad (7.70)$$

$$= 1 - P(\bar{Z} \leq 2.09) \quad (7.71)$$

$$= 1 - G(2.09) \quad (7.72)$$

$$= 1 - 0.9817 \quad (7.73)$$

$$= \underline{\underline{0.0183}} \quad (7.74)$$

Kommentar:

Legg merke til at det er \bar{X} som skal standardiseres. Ikke X . Det betyr at vi må bruke $E[\bar{X}] = 1.81$ og $\sigma[\bar{X}] = \sqrt{0.002668}$ (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.: $\underline{\underline{-1 < R_{xy} < 1}}$.
- ii) R_{xy} er et mål på lineær korrelasjon mellom observasjonene x og y .
Det er et mål på i hvor stor grad det er en lineær sammenheng mellom observasjonene x og y .
- iii) R_{xy} er enhetsuavhengig, dvs. ingen enhet.⁶

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

$$\underline{\underline{R_{xy}}} = \frac{S_{xy}}{S_x \cdot S_y} \quad (7.75)$$

$$= \frac{-110.83}{\sqrt{116.67} \cdot \sqrt{109.54}} = \underline{\underline{-0.98}} \quad (7.76)$$

- 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 $R_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$.

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

$$\hat{y} = \hat{\alpha} + \hat{\beta}x \quad (7.77)$$

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

$$\hat{\beta} = \frac{S_{xy}}{S_x^2} = \frac{-110.83}{116.67} = \underline{-0.95} \quad (7.78a)$$

$$\hat{\alpha} = \bar{y} - \beta\bar{x} = 40.43 - (-0.95) \cdot 15 = \underline{54.68} \quad (7.78b)$$

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

$$\underline{\underline{\hat{y} = 54.68 - 0.95 \cdot x}} \quad (7.79)$$

- e) Bruker regresjonslinjen $\hat{y} = \hat{y}(x)$ fra oppgave 4d, dvs. lign.(7.79):

$$\underline{\hat{y}(17)} = 54.68 - 0.95 \cdot 17 = \underline{38.53} \quad (7.80)$$

Regresjonslinjen predikerer at kvadratmeterprisen for en leilighet som er 17 km utenfor sentrum er 38 530 NOK/m².

- f) Gjennomsnittlig kvadratmeterpris i Norge er 35 000 NOK/ m^2 .
Ifølge regresjonslinjen, når $\hat{y}(x) = 35$, er da:

$$35 = 54.68 - 0.95 \cdot x \quad (7.81)$$

Løser med hensyn på x :

$$\underline{x} = \frac{54.68 - 35}{0.95} = \underline{20.72} \quad (7.82)$$

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

$$\underline{\underline{R^2 = 0.9613}} \quad (7.83)$$

- h) Kommenter til svaret i oppgave 4g:

At $R^2 = 0.9613$ betyr at regresjonslinjen vil i “i stor grad” **forutsi** kvadratmeterprisen på en leilighet for en gitt avstand fra sentrum.
Modellen har stor forklaringskraft.



⁷Man 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.

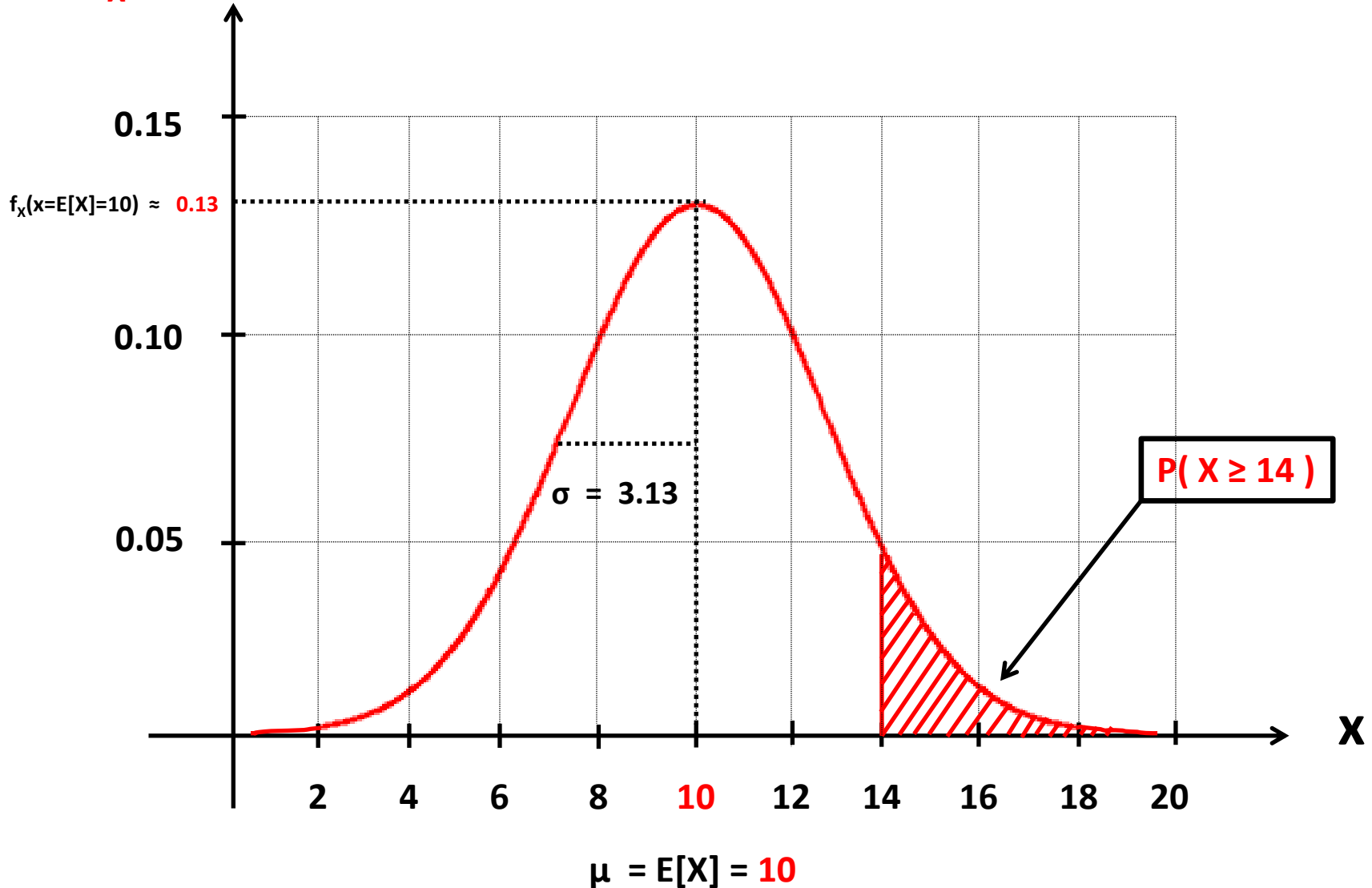
Vedlegg



Høgskolen i Molde
Vitenskapelig høgskole i logistikk

Studentnummer: _____

$f_X(x)$



Kapittel 8

LØSNING: Eksamen 8. januar 2016

“MAT110 Statistikk 1”

Oppgave 1: (logistikk)

- a) Siden lastebilene ankommer varemottaket **uavhengige** av hverandre så er dette et **telleproblem**. Dermed er sannsynlighetene P_1 , P_2 og P_3 gitt ved:

$$\underline{P_1} = \frac{n_1}{n} = \frac{5}{10} = \underline{0.5} \quad (8.1)$$

$$\underline{P_2} = \frac{n_2}{n} = \frac{3}{10} = \underline{0.3} \quad (8.2)$$

$$\underline{P_3} = \frac{n_3}{n} = \frac{2}{10} = \underline{0.2} \quad (8.3)$$

- b) Sannsynlighetene P_1 , P_2 og P_3 utgjør en **gyldig** sannsynlighetsfordeling fordi de summeres opp til èn:

$$\underline{\underline{P_1 + P_2 + P_3}} = 0.5 + 0.3 + 0.2 = \underline{1} \quad (8.4)$$

- c) Den spesifikke sannsynlighetfordelingen $P(X = x_i)$ er oppgitt. Dermed bruker vi [definisjonen](#) av forventning:

$$\underline{E[X]} \stackrel{\text{def.}}{=} \sum_{i=1}^3 x_i \cdot P(X = x_i) \quad (8.5)$$

$$\begin{aligned} &= 10 \cdot \underbrace{P(X = 10)}_{=0.5} + 20 \cdot \underbrace{P(X = 20)}_{=0.3} + 30 \cdot \underbrace{P(X = 30)}_{=0.2} \\ &= 10 \cdot 0.5 + 20 \cdot 0.3 + 30 \cdot 0.2 = \underline{17} \end{aligned} \quad (8.6)$$

Forventet behandlingstid $E[X]$ for å laste av lasten til en tilfeldig valgt lastebil er 17 minutter.

- d) Den spesifikke sannsynlighetfordelingen $P(X = x_i)$ er oppgitt. Dermed bruker vi [definisjonen](#) av varians:

$$\underline{\underline{Var[X]}} \stackrel{\text{def.}}{=} \sum_{i=1}^3 \left(x_i - E[X] \right)^2 \cdot P(X = x_i) \quad (8.7)$$

$$\begin{aligned} &= (10 - 17)^2 \cdot \underbrace{P(X = 10)}_{=0.5} + (20 - 17)^2 \cdot \underbrace{P(X = 20)}_{=0.3} + (30 - 17)^2 \cdot \underbrace{P(X = 30)}_{=0.2} \\ &= (10 - 17)^2 \cdot 0.5 + (20 - 17)^2 \cdot 0.3 + (30 - 17)^2 \cdot 0.2 = \underline{\underline{61}} \end{aligned} \quad (8.8)$$

med tilhørende standardavvik $\sigma[X]$:

$$\underline{\underline{\sigma[X]}} \stackrel{\text{def.}}{=} \sqrt{Var[X]} = \sqrt{61} = \underline{\underline{7.81}} \quad (8.9)$$

e) Fortventet ventetid $E[V]$:

$$\underline{E[V]} = E[X_1 + X_2 + X_3] \quad (8.10)$$

$$= E[X_1] + E[X_2] + E[X_3] \quad (8.11)$$

$$= E[X] + E[X] + E[X] \quad (8.12)$$

$$= 3E[X] = 3 \cdot 17 = \underline{51} \quad (8.13)$$

Forventet behandlingstid $E[V]$ dersom det står 3 tilfeldige lastebiler foran deg i kø, er 51 minutter.

f) Fortventet varians $Var[V]$:

$$\underline{Var[V]} = Var[X_1 + X_2 + X_3] \quad (8.14)$$

$$\stackrel{\text{uavh.}}{=} Var[X_1] + Var[X_2] + Var[X_3] \quad (8.15)$$

$$= Var[X] + Var[X] + Var[X] \quad (8.16)$$

$$= 3Var[X] = 3 \cdot 61 = \underline{183} \quad (8.17)$$

med tilhørende standardavvik $\sigma[V]$:

$$\underline{\sigma[V]} \stackrel{\text{def.}}{=} \sqrt{Var[V]} = \sqrt{183} = \underline{13.5} \quad (8.18)$$

- g) Siden lastebilene ankommer varemottaket uavhengige av hverandre så er den simultane sannsynligheten bare produktet av sannsynlighetene:

$$\underline{p(30, 30, 30)} \equiv P(X_1 = 30 \text{ og } X_2 = 30 \text{ og } X_3 = 30) \quad (8.19)$$

$$\stackrel{\text{uavh.}}{=} P(X = 30) \cdot P(X = 30) \cdot P(X = 30) \quad (8.20)$$

$$= \left(\underbrace{P(X = 30)}_{=0.2} \right)^3 = 0.2^3 = \underline{0.008} \quad (8.21)$$

- h) Tolkning:

$p(30, 30, 30)$ = sannsynligheten for at alle 3 lastebilene som står foran deg i kø, er store lastebiler som tar 30 minutter hver å laste av

- i) Den eneste måten at ventetiden kan bli 90 minutter på er at alle 3 lastebilene som er foran deg i kø, er store: $P(V = 90) = p(30, 30, 30)$. Dermed:

$$\underline{P(V \leq 80)} = 1 - \overbrace{P(V = 90)}^{= p(30,30,30)} \quad (8.22)$$

$$= 1 - 0.008 = \underline{0.992} \quad (8.23)$$

Sannsynligheten for at du må vente 80 minutter eller mindre er $P(V \leq 80) = 0.992$.

- j) Igjen bruker vi det faktum at lastebilene ankommer uavhengige av hverandre. Da er de simultane sannsynlighetene bare produktet av sannsynlighetene. Dermed:

$$\begin{aligned}
 \underline{P(V = 50)} &= \overbrace{p(10, 10, 30) + p(10, 30, 10) + p(30, 10, 10)}^{\text{disse tre sanns. er like}} \\
 &+ \overbrace{p(20, 20, 10) + p(20, 10, 20) + p(10, 20, 20)}^{\text{disse tre sanns. er like}} \quad (8.24)
 \end{aligned}$$

$$= 3 \cdot p(10, 10, 30) + 3 \cdot p(20, 20, 10) \quad (8.25)$$

$$= 3 \left(p(10, 10, 30) + p(20, 20, 10) \right) \quad (8.26)$$

$$= 3 \left(0.5 \cdot 0.5 \cdot 0.2 + 0.3 \cdot 0.3 \cdot 0.5 \right) = \underline{0.285} \quad (8.27)$$

Sannsynligheten for at du må vente 50 minutter er $P(V = 50) = 0.285$.



Oppgave 2: (økonomi og aksjer)

- a) Bruker setningen for total sannsynlighet for å finne sannsynligheten for at en tilfeldig valgt medlem av *Econa* **aldri** leser næringslivsaviser:

$$\underline{\underline{P(A)}} \stackrel{\text{total}}{=} P(A \cap H) + P(A \cap \bar{H}) \quad (8.28)$$

$$= 0.04 + 0.21 = \underline{\underline{0.25}} \quad (8.29)$$

- 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:

$$\underline{\underline{P(H)}} \stackrel{\text{opp spl.}}{=} P(H \cap R) + P(H \cap I) + P(H \cap A) \quad (8.30)$$

$$= 0.18 + 0.10 + 0.04 = \underline{\underline{0.32}} \quad (8.31)$$

- d) Den generelle multiplikasjonssetningen gir:

$$\underline{\underline{P(H|A)}} = \frac{P(H \cap A)}{P(A)} = \frac{0.04}{0.25} = \underline{\underline{0.16}} \quad (8.32)$$

- e) Man kan igjen bruke den generelle multiplikasjonssetningen: (husk at: $H \cap A = A \cap H$)

$$\underline{\underline{P(A|H)}} \stackrel{\text{mult.}}{=} \frac{P(A \cap H)}{P(H)} = \frac{0.04}{0.25} = \underline{\underline{0.125}} \quad (8.33)$$

eller Bayes lov

$$\underline{\underline{P(A|H)}} \stackrel{\text{Baye}}{=} P(H|A) \frac{P(A)}{P(H)} = 0.16 \frac{0.25}{0.32} = \underline{\underline{0.125}} \quad (8.34)$$

På eksamen er det nok at dere løser oppgaven på èn av måtene.

- f) Tolkning:

$$\underline{\underline{P(A|H)}} = \text{sannsynligheten for at et medlem av } Econa \text{ som har handlet aksjer} \\ \underline{\underline{\text{på aksjemarkedet det siste året, aldri leser finansaviser}}} \quad (8.35)$$

- g) Sannsynligheten $P(H|\bar{R}) = P(H|(A \cup I))$ kan vi finne via den generelle multiplikasjonssetningen:

$$P(H|\bar{R}) = \frac{P(H \cap \bar{R})}{P(\bar{R})} \quad (8.36)$$

eller, siden $\bar{R} = A \cup I$,

$$P(H|(A \cup I)) = \frac{P(H \cap (A \cup I))}{P(A \cup I)} \quad (8.37)$$

Siden begivenhetene A og I disjunkte, så kan vi bruke den spesielle addisjonssetningen:

$$\underline{P(A \cup I)} \stackrel{\text{add.}}{=} P(A) + P(I) = \underbrace{(0.04 + 0.21)}_{\text{se tabell eller oppg. 1a}} + \underbrace{(0.10 + 0.31)}_{\text{se tabell i oppgaven}} = \underline{0.66} \quad (8.38)$$

Tabellen i oppgaven gir videre at: ¹

$$\underline{P(H \cap (A \cup I))} = P((H \cap A) \cup (H \cap I)) \stackrel{\text{add.}}{=} P(H \cap A) + P(H \cap I) \quad (8.39)$$

$$= 0.04 + 0.10 = \underline{0.14} \quad (8.40)$$

Innsatt i lign.(8.37):

$$\underline{\underline{P(H|(A \cup I))}} = \frac{P(H \cap (A \cup I))}{P(A \cup I)} = \frac{0.14}{0.66} \approx \underline{\underline{0.21}} \quad (8.41)$$

■

¹Man får også full poengscore selv om man ikke har med mellomregningene i lign.(8.39). 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 Gjennes på følgende måte:

$$\underline{\underline{P(X_G = 2)}} = \frac{\lambda_G^2}{2!} e^{-\lambda_G} \quad (8.42)$$

$$\lambda_G \underline{\underline{= 1.9}} \quad \frac{1.9^2}{2!} e^{-1.9} \stackrel{\text{kalkis}}{=} \underline{\underline{0.2700}} \quad (8.43)$$

- b) Det er oppgitt i oppgaven at $X_G \sim \text{Poi}[\lambda]$.
Sannsynligheten for at det skjer **mer enn 2** utrykninger i Gjennes i løpet av en uke:

$$\underline{\underline{P(X_G > 2)}} = 1 - P(X_G \leq 2) \quad (8.44)$$

$$= 1 - \left(\overbrace{P(X_G = 0) + P(X_G = 1)}^{\text{Oppgitt i oppgaven.}} + \overbrace{P(X_G = 2)}^{\text{Fra oppg. 3a}} \right) \quad (8.45)$$

$$= 1 - 0.1496 - 0.2842 - 0.2700 \quad (8.46)$$

$$= \underline{\underline{0.2962}} \quad (8.47)$$

c) Forventet **antall** akutte **utrykninger** i Eide, Fræna og Gjemnes til sammen per uke:

$$\underline{\underline{E[Y]}} = E[X_E + X_F + X_G] \quad (8.48)$$

$$\stackrel{\text{alltid}}{=} E[X_E] + E[X_F] + E[X_G] = \lambda_E + \lambda_F + \lambda_G \quad (8.49)$$

$$= 1.5 + 2.3 + 1.9 = \underline{\underline{5.7}} \quad (8.50)$$

NB: Overgangen i lign.(8.48) til (8.49) 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 forventning $\lambda_Y = E[Y] = 5.7$.

Sannsynligheten for at det skjer *mer enn* 2 akutte utrykningen i Eide, Fræna og Gjemnes tilsammen oer uke er da:

$$\underline{\underline{P(Y > 2)}} = 1 - P(Y \leq 2) \quad (8.51)$$

$$= 1 - \left(P(Y = 0) + P(Y = 1) + P(Y = 2) \right) \quad (8.52)$$

$$= 1 - \frac{\lambda_Y^0}{0!} e^{-\lambda_Y} - \frac{\lambda_Y^1}{1!} e^{-\lambda_Y} - \frac{\lambda_Y^2}{2!} e^{-\lambda_Y} \quad (8.53)$$

$$\stackrel{\lambda_Y=5.7}{=} 1 - \frac{5.7^0}{0!} e^{-5.7} - \frac{5.7^1}{1!} e^{-5.7} - \frac{5.7^2}{2!} e^{-5.7} \quad (8.54)$$

$$\stackrel{\text{kalkis}}{=} 1 - 0.0033 - 0.0191 - 0.0544 \quad (8.55)$$

$$= \underline{\underline{0.9232}} \quad (8.56)$$

e) Siden

$$P(Y > 2) = 0.9232 \quad \text{og} \quad P(X > 2) = 0.2962 \quad (8.57)$$

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:

$$P(Y > 2) \gg P(X > 2) \quad (8.58)$$

NB:

Selv om Y beskriver summen av akutte utrykninger i Eide, Fræna og Gjemnes tilsammen så kan vi **IKKE** slik at:

$$\underbrace{P(Y > 2) = P(X_E > 2) + P(X_F > 2) + P(X_G > 2)}_{\text{galt}} \quad (8.59)$$

f) Forventet antall akutte utrykninger i året i Gjemnes kommune:

$$\underline{\underline{E[X_{\text{år}}]}} = E[X_1 + X_2 + \dots + X_n] = \overbrace{E[X_1] + E[X_2] + \dots + E[X_n]}^{n \text{ stk.}} \quad (8.60)$$

$$= n \underbrace{E[X]}_{=\lambda_G=1.9} = 52 \cdot 1.9 = \underline{\underline{98.8}} \quad (8.61)$$

NB: Overgangen i lign.(8.60) gjelder alltid. Uansett om de stokastiske variablene X_i er uavhengige eller ikke.

g) Variansen til antall akutte utrykninger i året for Gjemnes kommune?

$$\underline{\underline{Var[X_{\hat{a}r}]}} = Var[X_1 + X_2 + \dots + X_n] \quad (8.62)$$

$$\stackrel{\text{uavh.}}{=} \overbrace{Var[X_1] + Var[X_2] + \dots + Var[X_n]}^{n \text{ stk.}} \quad (8.63)$$

$$= n \underbrace{Var[X]}_{=\lambda_G=1.9} = 52 \cdot 1.9 = \underline{\underline{98.8}} \quad (8.64)$$

NB: Overgangen i lign.(8.62) 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 $X_{\hat{a}r}$ **normalfordelt**:

$$\underline{\underline{X_{\hat{a}r} \sim N[E[X_{\hat{a}r}], Var[X_{\hat{a}r}]]}} \quad (8.65)$$

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

$$\underline{\underline{n \gtrsim 30}} \quad (8.66)$$

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

$$\underline{P(X_{\text{år}} > 110)} \stackrel{\text{standardiser}}{=} P\left(\underbrace{\frac{X_{\text{år}} - E[X_{\text{år}}]}{\sigma[X_{\text{år}}]}}_{= Z} > \frac{110 - E[X_{\text{år}}]}{\sigma[X_{\text{år}}]}\right) \quad (8.67)$$

$$= P\left(Z > \frac{110 - 98.8}{\sqrt{98.8}}\right) \quad (8.68)$$

$$= P(Z > 2.13) \quad (8.69)$$

$$= 1 - P(Z \leq 1.13) \quad (8.70)$$

$$= 1 - \underbrace{G(1.13)}_{= 0.8708} \quad (8.71)$$

$$\stackrel{\text{tabell}}{=} 1 - 0.8708 \quad (8.72)$$

$$= \underline{0.1292} \quad (8.73)$$

Sannsynligheten for at det blir mer enn 110 akutte utrykninger i året i Gjemnes kommune er 12.92 %.

■

²I 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.: $\underline{\underline{-1 < R_{xy} < 1}}$.
- ii) R_{xy} er et mål på lineær korrelasjon mellom observasjonene x og y .
Det er et mål på i hvor stor grad det er en lineær sammenheng mellom observasjonene x og y .
- iii) R_{xy} er enhetsuavhengig, dvs. ingen enhet.³

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

$$\underline{\underline{R_{xy}}} = \frac{S_{xy}}{S_x \cdot S_y} \quad (8.74)$$

$$= \frac{10.36}{\sqrt{6} \cdot \sqrt{19.7}} = \underline{\underline{0.95}} \quad (8.75)$$

- 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.

³Dette 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:

$$\hat{y} = \hat{\alpha} + \hat{\beta}x \quad (8.76)$$

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

$$\hat{\beta} = \frac{S_{xy}}{S_x^2} = \frac{10.36}{6} = \underline{1.73} \quad (8.77)$$

$$\hat{\alpha} = \bar{y} - \beta\bar{x} = 33.625 - 1.73 \cdot 4.5 = \underline{25.84} \quad (8.78)$$

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

$$\underline{\underline{\hat{y} = 25.84 + 1.73 \cdot x}} \quad (8.79)$$

- e) Bruker regresjonslinjen $\hat{y} = \hat{y}(x)$ fra oppgave 4d, dvs. lign.(8.79):

$$\underline{\hat{y}(12)} = 25.84 + 1.73 \cdot 12 = \underline{46.56} \quad (8.80)$$

Dersom den lineære treden holder seg også i 2016 så predikerer regresjonslinjen at Brunvoll vil få omtrent 47 serviceordrer.

f) Med 50 serviceordrer i kvartalet er $\hat{y}(x) = 50$. Ifølge regresjonslinjen er da:

$$\hat{y}(x) \stackrel{\text{Eq.(8.80)}}{=} 25.84 + 1.73 \cdot x \quad (8.81)$$

$$50 = 25.84 + 1.73 \cdot x \quad (8.82)$$

Løser med hensyn på x :

$$\underline{x} = \frac{50 - 25.84}{1.74} = 13.97 \approx \underline{14} \quad (8.83)$$

Ifølge regresjonslinjen så vil antall serviceordrer bli 50 i kvartalet, 2. kvartal i 2017.

g) Forklaringsstyrken R^2 er: (se formelsamling)

$$\underline{\underline{R^2}} \stackrel{\text{def.}}{=} 1 - \frac{SSE}{SST} \quad (8.84)$$

$$= 1 - \frac{12.73}{137.88} = \underline{\underline{0.9077}} \quad (8.85)$$

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

LØSNING: Eksamen 27. mai 2016

“MAT110 Statistikk 1”

Oppgave 1: (logistikk , økonomi)

a) Sannsynligheten for at “det ikke er ledig plass på ferge 2 er 0.25”:

$$P(\bar{L}_2) = 0.25 \quad (9.1)$$

Sannsynligheten for at “det er ledig plass på ferge 1 og ferge 2 er 0.70”

$$\underline{\underline{P(L_1 \overset{\text{og}}{\cap} L_2) = 0.70}} \quad (9.2)$$

b) Vi skal finne “eller”-sannsynligheten $P(L_1 \cup L_2)$.

Vi kjenner “og”-sannsynligheten $P(L_1 \overset{\text{og}}{\cap} L_2)$, og vi skal finne “eller”-sannsynligheten. Setningen som forbinder disse sannsynlighetene er den [generelle addisjonssetningen](#):

$$\underbrace{P(L_1 \cup L_2)}_{\text{skal finne eller}} = P(L_1) + P(L_2) - \underbrace{P(L_1 \overset{\text{og}}{\cap} L_2)}_{= 0.70} \quad (9.3)$$

hvor $P(L_1 \cap L_2) = 0.70$ var oppgitt i oppgaven, og hvor $P(L_1)$ og $P(L_2)$ finnes via komplementsetningen.

Dermed:

$$\underline{P(L_1 \cup L_2)} = \overbrace{P(L_1)}^{= 1-0.15} + \overbrace{P(L_2)}^{= 1-0.25} - \overbrace{P(L_1 \cap L_2)}^{= 0.70} \quad (9.4)$$

$$= 0.85 + 0.75 - 0.70 = \underline{0.90} \quad (9.5)$$

Altså, det er 90 % sannsynlighet for at det er ledig plass på ferge 1 eller ferge 2 (eller begge).

c) Vi skal finne sannsynligheten $P(L_1 \cap \bar{L}_2)$.

Sannsynligheten $P(L_1 \cap L_2) = 0.70$ var oppgitt i oppgaven, og vi skal finne $P(L_1 \cap \bar{L}_2)$. Setningen som forbinder disse sannsynlighetene er den **total sannsynlighet**, se formelsamlingen:

$$\underbrace{P(L_1)}_{= 1-0.15} = \underbrace{P(L_1 \cap L_2)}_{= 0.70} + \overbrace{P(L_1 \cap \bar{L}_2)}^{\text{skal finne}} \quad (9.6)$$

som gir

$$\underline{P(L_1 \cap \bar{L}_2)} = P(L_1) - P(L_1 \cap L_2) \quad (9.7)$$

$$= 0.85 - 0.70 = \underline{0.15} \quad (9.8)$$

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 $P(\bar{L}_1 \cap \bar{L}_2)$. Setningen som forbinder disse sannsynlighetene er en av [tvillingsetningene](#), se formelsamlingen:

$$P(\bar{L}_1 \overset{\text{og}}{\cap} \bar{L}_2) = 1 - P(L_1 \overset{\text{eller}}{\cup} L_2) \quad (9.9)$$

som gir

$$\underline{P(\bar{L}_1 \overset{\text{og}}{\cap} \bar{L}_2)} = 1 - \underbrace{P(L_1 \overset{\text{eller}}{\cup} L_2)}_{= 0.90} \quad (9.10)$$

$$= 1 - 0.90 = \underline{0.10} \quad (9.11)$$

Altså, det er 10 % sannsynlighet for at begge fergene er fulle.

- e) Bring sine [forventede](#) utgifter pga. fulle ferger for en tur mellom Molde og Volda er:

$$\underline{\underline{E[U]}} = \sum_{u=1}^4 u_i \cdot P(U = u_i) \quad (9.12)$$

$$= \left(0 \cdot P(L_1 \cap L_2) + 1200 \cdot \overbrace{P(\bar{L}_1 \cap L_2)}^{= 0.05} \right) \quad (9.13)$$

$$+ \left(800 \cdot P(L_1 \cap \bar{L}_2) + 2000 \cdot P(\bar{L}_1 \cap \bar{L}_2) \right) \text{ NOK} \quad (9.14)$$

$$= \left(0 \cdot 0.70 + 1200 \cdot 0.05 + 800 \cdot 0.15 + 2000 \cdot 0.10 \right) \text{ NOK} \quad (9.15)$$

$$= \underline{\underline{380 \text{ NOK}}} \quad (9.16)$$

■

Oppgave 2: (logistikk , økonomi , “avisguttens problem”)

a) Siden D er Poissonfordelt så er $E[D] = \lambda$:

$$\underline{\underline{E[D]}} = \lambda = \underline{\underline{81}} \quad (9.17)$$

b) Siden D er Poissonfordelt så er $\sigma[D] = \sqrt{\lambda}$:

$$\underline{\underline{\sigma[D]}} = \sqrt{\lambda} = \sqrt{81} = \underline{\underline{9}} \quad (9.18)$$

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 tilnærming normalfordelt.

Siden $\lambda = 81 \gg 5$ så er vi her langt innenfor denne grensen.

- d) Siden D er tilnærmet normalfordelt og finn sannsynligheten ved standardisering ($X \rightarrow Z$) og deretter tabelloppslag: ¹

$$\underline{P(D > 80)} \stackrel{\text{standardiser}}{=} P\left(\overbrace{\frac{D - E[D]}{\sigma[D]}}^{= Z} > \frac{80 - E[D]}{\sigma[D]}\right) \quad (9.19)$$

$$= P\left(Z > \frac{80 - 81}{\sqrt{81}}\right) \quad (9.20)$$

$$= \underline{P(Z > -0.11)} \quad (9.21)$$

$$= 1 - P(Z > 0.11) \quad (9.22)$$

$$= 1 - \left(1 - P(Z \leq 0.11)\right) \quad (9.23)$$

$$= \underbrace{P(Z \leq 0.11)}_{= 0.5438}$$

$$\stackrel{\text{tabell}}{=} \underline{0.5438} \quad (9.24)$$

Sannsynligheten for at det etterspørres mer enn 80 aviser en tilfeldig dag er 54.38 %.

- e) Forventningen $E[S]$ er gitt ved

$$\underline{E[S]} = \underbrace{\sum_{s=1}^{84} sP(S=s)}_{= 49.89} + 85 \cdot \underbrace{P(S=85)}_{= 0.3429} \quad (9.25)$$

$$= 49.89 + 85 \cdot 0.3429 \quad (9.26)$$

$$= \underline{79} \quad (9.27)$$

¹I 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:

$$E[D] = 81 \quad , \quad E[S] = 79 \quad (9.28)$$

altså

$$\text{forventet etterspørsel} = 81 > \text{forventet salg} = 79 \quad (9.29)$$

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

$$E[\pi(q)] = E[rS - wq] = \underline{rE[S] - wq} \quad (9.32)$$

For tilfellet $q = 85$ er $E[S] = 79$, jfr. oppgave **2e**. Med $w = 5$ NOK og $r = 20$ NOK får vi:

$$\underline{E[\pi(q)]} = (20 \cdot 79 - 5 \cdot 85) \text{ NOK} = \underline{1155 \text{ NOK}} \quad (9.33)$$

²En 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 kan selge. Siden sannsynlighetfordelingen $P(S = s_i)$ er den samme som $P(D = d_i)$ frem til $s = q - 1 = 84$ så må $E[D] > E[S]$.

³Her bruker vi regneregelen: (a og b er konstanter)

$$E[aX + bY] = aE[X] + bE[Y] , \quad (9.30)$$

$$E[a] = a , \quad (9.31)$$

som man finner i formelsamlingen.

- i) I oppgaven er det oppgitt at den stokastiske variabelen D er normalfordelt, med $D \sim N[\mu = 81, \sigma = 9]$. Dermed løses denne oppgaven med (omvendt) tabelloppslag.⁴

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

$$P(D \leq q^*) = 1 - \frac{w}{r} \quad (9.34)$$

$$P(D \leq q^*) = 1 - \frac{5}{20} \quad (9.35)$$

$$P(D \leq q^*) = 0.75 \quad (9.36)$$

$$P\left(\underbrace{\frac{D - \mu}{\sigma}}_{\equiv Z} \leq \underbrace{\frac{q^* - \mu}{\sigma}}_{\equiv Z_0}\right) \stackrel{\text{standardiser}}{=} 0.75 \quad (9.37)$$

$$P(Z \leq Z_0) = 0.75 \quad (9.38)$$

Ved “[omvendt tabelloppslag](#)” ser vi at $Z'_0 = 0.67$ tilsvarende $P(Z' \leq Z'_0) = 0.7486$. Videre ser vi at $Z''_0 = 0.68$ tilsvarende $P(Z'' \leq Z''_0) = 0.7517$. Vi skal ha 0.75, som er ca. *midt i mellom*. Dermed:

$$Z_0 = 0.675 \quad (9.39)$$

Vi løser:

$$Z_0 = \frac{q^* - \mu}{\sigma} \quad (9.40)$$

med hensyn på q^* : ($\mu = 81$ og $\sigma = 9$)

$$\underline{q^*} = \mu + Z_0 \cdot \sigma \quad (9.41)$$

$$= 81 + 0.675 \cdot 9 = \underline{87} \quad (9.42)$$

Avisgutten må bestille $q^* = 87$ aviser for å få størst mulig fortjeneste.

⁴Husk 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

$$\begin{aligned} q^* &> E[D] \\ 87 &> 81, \end{aligned} \tag{9.43}$$

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 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:
 $P(X = 2) = 0.20$.

c) Forventning $E[X]$:

$$\underline{E[X]} \stackrel{\text{def.}}{=} \sum_{i=0}^5 x_i \cdot P(X = x_i) \quad (9.44)$$

$$\begin{aligned} &= 0 \cdot \overbrace{P(X = 0)}^{=0.25} + 1 \cdot \overbrace{P(X = 1)}^{=0.30} + 2 \cdot \overbrace{P(X = 2)}^{=0.20} \\ &\quad + 3 \cdot \underbrace{P(X = 3)}_{=0.15} + 4 \cdot \underbrace{P(X = 4)}_{=0.10} + 5 \cdot \underbrace{P(X = 5)}_{=0} \end{aligned} \quad (9.45)$$

$$= 0 \cdot 0.25 + 1 \cdot 0.30 + 2 \cdot 0.20 + 3 \cdot 0.15 + 4 \cdot 0.10 + 5 \cdot 0 \quad (9.46)$$

$$= \underline{1.55} \quad (9.47)$$

- d) Sannsynligheten for at det skjer høyst èn trafikkulykke på den aktuelle vegstrekningen over en periode på to år er:

$$\begin{aligned}
 P(Y \leq 1) &= P(X_1 = 0) \cdot P(X_2 = 0) \\
 &+ P(X_1 = 1) \cdot P(X_2 = 0) + P(X_1 = 0) \cdot P(X_2 = 1)
 \end{aligned}
 \tag{9.48}$$

Men $P(X_1 = 0) = P(X_2 = 0) = 0.25$ osv., så:

$$\underline{\underline{P(Y \leq 1)}} = P(X = 0)^2 + 2P(X = 0)P(X = 1)
 \tag{9.49}$$

$$= 0.25^2 + 2 \cdot 0.25 \cdot 0.30 = \underline{\underline{0.2125}}
 \tag{9.50}$$

- e) Forventet antall trafikkulykker på den aktuelle strekningen over en periode på $n = 30$ år:

$$\underline{\underline{E[Y]}} = E[X_1 + X_2 + \dots + X_n] = \overbrace{E[X_1] + E[X_2] + \dots + E[X_n]}^{n \text{ stk.}}
 \tag{9.51}$$

$$= n E[X] = 30 \cdot 1.55 = \underline{\underline{46.5}}
 \tag{9.52}$$

NB: Overgangen i lign.(9.51) 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 $n = 30$ år:

$$\underline{\underline{Var[Y]}} = Var [X_1 + X_2 + \dots + X_n] \quad (9.53)$$

$$\stackrel{\text{uavh.}}{=} \overbrace{Var[X_1] + Var[X_2] + \dots + Var[X_n]}^{n \text{ stk.}} \quad (9.54)$$

$$= n \cdot \underbrace{Var[X]}_{=1.6475} = 30 \cdot 1.6475 = \underline{\underline{49.425}} \quad (9.55)$$

NB: Overgangen i lign.(9.53) 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 **normalfordelt**:

$$\underline{\underline{Y}} \sim N[E[Y], Var[Y]] \quad (9.56)$$

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

$$\underline{\underline{n}} \gtrsim 30 \quad (9.57)$$

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

$$\underline{P(Y > 50)} \stackrel{\text{standardiser}}{=} P\left(\underbrace{\frac{Y - E[Y]}{\sigma[Y]}}_{= Z} > \frac{50 - E[Y]}{\sigma[Y]}\right) \quad (9.58)$$

$$= P\left(Z > \frac{50 - 46.5}{\sqrt{49.425}}\right) \quad (9.59)$$

$$= P(Z > 0.50) \quad (9.60)$$

$$= 1 - P(Z \leq 0.50) \quad (9.61)$$

$$= 1 - \underbrace{G(0.50)}_{= 0.6915} \quad (9.62)$$

$$\stackrel{\text{tabell}}{=} 1 - 0.6915 \quad (9.63)$$

$$= \underline{0.3085} \quad (9.64)$$

Sannsynligheten for at det skjer mer enn 50 trafikkulykker på den aktuelle vegstrekningen over en periode på $n = 30$ år er 30.85 %.



⁵I 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.: $-1 < R_{xy} < 1$.

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

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

c) Uttrykket for korrelasjonskoeffisienten finner man i formelsamlingen. Det gir:

$$\underline{R_{xy}} = \frac{S_{xy}}{S_x \cdot S_y} \quad (9.65)$$

$$= \frac{-0.1531}{\sqrt{0.15625} \cdot \sqrt{0.15325}} = \underline{\underline{-0.99}} \quad (9.66)$$

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 sterk negativ korrelasjon mellom x og y .

Det er altså en svært sterk lineær sammenheng mellom x og y med negativt stigningstall.

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

$$\hat{y} = \hat{\alpha} + \hat{\beta}x \quad (9.67)$$

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

$$\hat{\beta} = \frac{S_{xy}}{S_x^2} = \frac{-0.1531}{0.15625} = \underline{-0.98} \quad (9.68)$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 0.73 - (-0.98) \cdot 3.5 = \underline{4.16} \quad (9.69)$$

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

$$\underline{\underline{\hat{y} = 4.16 - 0.98 \cdot x}} \quad (9.70)$$

f) Bruker regresjonslinjen $\hat{y} = \hat{y}(x)$ fra oppgave **4e**, dvs. lign.(9.70):

$$\underline{\hat{y}(3.4)} = 4.16 - 0.98 \cdot 3.4 = \underline{0.828} \quad (9.71)$$

Dersom prisen på en Volvo V70 er 340 000 NOK, dvs. $x = 3.4$, så predikerer regresjonslinjen at det vil selges 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:

$$\hat{y}(x) \stackrel{\text{Eq.(9.71)}}{=} 4.16 - 0.98 \cdot x \quad (9.72)$$

$$1 = 4.16 - 0.98 \cdot x \quad (9.73)$$

Løser med hensyn på x :

$$\underline{x} = \frac{4.16 - 1}{0.98} = \underline{3.22} \quad (9.74)$$

Ifølge regresjonslinjen så vil Volvo selge 1 000 biler per dag dersom de setter prisen til 322 000 NOK.

- h) Forklaringsstyrken R^2 er: (se formelsamling)

$$\underline{R^2} \stackrel{\text{def.}}{=} 1 - \frac{SSE}{SST} \quad (9.75)$$

$$= 1 - \frac{0.01275}{0.613} = \underline{0.9792} \quad (9.76)$$

- 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

LØSNING: Eksamen 5. januar 2017

“MAT110 Statistikk 1”

Oppgave 1: (produksjonsplanlegging , økonomi)

- a) Sannsynligheten for at malingen ikke er defekt etter 1 time er:

$$\underline{\underline{P(\bar{D}_1)}} = 1 - P(D_1) = 1 - 0.30 = \underline{\underline{0.70}} \quad (10.1)$$

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:

$$\underline{\underline{P(\bar{D}_2|D_1)}} = 1 - P(\bar{D}_2|D_1) = 1 - 0.10 = \underline{\underline{0.90}} \quad (10.2)$$

hvor vi har brukt komplementsetningen igjen.

- c) Uttrykket $P(\bar{D}_2 \cap D_1)$ er sannsynligheten for at

malingen er “*defekt*” etter 1 time og “*ikke defekt*” etter 2 timer,
sannsynligheten for at malingen er OK etter 2 timer.

d) Den generelle multiplikasjonssetningen gir:

$$\underline{\underline{P(\bar{D}_2 \cap D_1)}} = P(\bar{D}_2|D_1)P(D_1) \quad (10.3)$$

$$= 0.90 \cdot 0.30 = \underline{\underline{0.27}} \quad , \quad \text{q.e.d.} \quad (10.4)$$

e) Vi sannsynlighetstreet som oppgitt i oppgaven finner vi at:

$$\underline{\underline{P(\bar{D}_4 \cap D_3 \cap D_2 \cap D_1)}} = P(\bar{D}_4|D_3 \cap D_2 \cap D_1)P(D_3|D_2 \cap D_1)P(D_2|D_1)P(D_1) \quad (10.5)$$

$$= 1 \cdot 0.02 \cdot 0.10 \cdot 0.30 = \underline{\underline{0.0006}} \quad , \quad \text{q.e.d.} \quad (10.6)$$

f) Bruker resultatene fra tidligere deloppgaver samt den oppgitte sannsynligheten $P(\bar{D}_3 \cap D_2 \cap D_1) = 0.0294$ til å finne **forventet** antall timer som Jotun bruker på blanding og justeringer av malingen:

$$\underline{\underline{E[X]}} = \sum_{i=1}^4 x_i \cdot P(X = x_i) \quad (10.7)$$

$$= 1 \cdot \overbrace{P(\bar{D}_1)}^{= 0.70} + 2 \cdot \overbrace{P(\bar{D}_2 \cap D_1)}^{= 0.27} \quad (10.8)$$

$$+ 3 \cdot \underbrace{P(\bar{D}_3 \cap D_2 \cap D_1)}_{= 0.0294} + 4 \cdot \underbrace{P(\bar{D}_4 \cap D_3 \cap D_2 \cap D_1)}_{= 0.0006}$$

$$= (1 \cdot 0.70 + 2 \cdot 0.27 + 3 \cdot 0.0294 + 4 \cdot 0.0006) \text{ timer}$$

$$= \underline{\underline{1.3306 \text{ timer}}} \quad (10.9)$$

g) Siden

$$\underline{\underline{\sum_{i=1}^4 P(X = x_i)}} = P(X = x_1) + P(X = x_2) + P(X = x_3) + P(X = x_4) \quad (10.10)$$

$$= P(\overline{D_1}) + P(\overline{D_2} \cap D_1) \quad (10.11)$$

$$+ P(\overline{D_3} \cap D_2 \cap D_1) + P(\overline{D_4} \cap D_3 \cap D_2 \cap D_1)$$

$$= 0.70 + 0.27 + 0.0294 + 0.0006 \quad (10.12)$$

$$= \underline{\underline{1}} \quad (10.13)$$

så er sannsynlighetsfordelingen $P(X = x_i)$ en gyldig fordeling.



Oppgave 2: (økonomi , aksjer)

a) Fortjenesten F er:

$$F = \text{antall aksjer} \cdot \text{pris per aksje om ett år} - \text{antall aksjer} \cdot \text{pris per aksje i dag} \quad (10.14)$$

for de tre selskapene så får vi:

$$\begin{aligned} \underline{F} &= \underbrace{\text{antall aksjer i } A}_{= N \cdot a} \cdot \underbrace{\text{pris per aksje for selskap } A \text{ om ett år}}_{= X} \\ &+ \underbrace{\text{antall aksjer i } B}_{= N \cdot b} \cdot \underbrace{\text{pris per aksje for selskap } B \text{ om ett år}}_{= Y} \\ &+ \underbrace{\text{antall aksjer i } C}_{= N \cdot c} \cdot \underbrace{\text{pris per aksje for selskap } C \text{ om ett år}}_{= Z} \end{aligned} \quad (10.15)$$

$$- \underbrace{\text{antall aksjer i } A}_{= N \cdot a} \cdot \underbrace{\text{pris per aksje for selskap } A \text{ i dag}}_{= 100} \quad (10.16)$$

$$- \underbrace{\text{antall aksjer i } B}_{= N \cdot b} \cdot \underbrace{\text{pris per aksje for selskap } B \text{ i dag}}_{= 105}$$

$$- \underbrace{\text{antall aksjer i } C}_{= N \cdot c} \cdot \underbrace{\text{pris per aksje for selskap } C \text{ i dag}}_{= 130} \quad (10.17)$$

$$= NaX + NbY + NcZ - Na100 + Nb105 + Nc130 \quad (10.18)$$

$$= \underline{\underline{Na(X - 100) + Nb(Y - 105) + Nc(Z - 130)}} \quad , \quad \text{q.e.d.} \quad (10.19)$$

b) Den forventede fortjenesten $E[F]$ ved et eventuelt salg av aksjene **om ett år** er:

$$\underline{E[F]} = E\left[Na(X - 100) + Nb(Y - 105) + Nc(Z - 130)\right] \quad (10.20)$$

$$= Na\left(\underbrace{E[X]}_{=90} - 100\right) + Nb\left(\underbrace{E[Y]}_{=125} - 105\right) + Nc\left(\underbrace{E[Z]}_{=180} - 130\right) \quad (10.21)$$

$$= Na(90 - 100) + Nb(125 - 105) + Nc(180 - 130) \quad (10.22)$$

$$= Na(-10) + Nb20 + Nc50 \quad (10.23)$$

$$= \underline{\underline{10N(-a + 2b + 5c)}} \quad , \quad \text{q.e.d.} \quad (10.24)$$

c) Fra formelsamlingen vet vi formelen for variansen av en lineær kombinasjon av to stokastiske variabler X og Y , nemlig:

$$\overbrace{Var[aX + bY]}^{\text{variasjon/(spredning)}} = a^2 Var[X] + b^2 Var[Y] + 2ab \cdot \overbrace{Cov[X, Y]}^{\text{samvariasjon}} \quad (10.25)$$

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:

$$\underline{\underline{Var[F]}} = Var \left[Na(X - 100) + Nb(Y - 105) + Nc(Z - 130) \right] \quad (10.26)$$

$$\begin{aligned} &= (Na)^2 \underbrace{Var[X]}_{= 100} + (Nb)^2 \underbrace{Var[Y]}_{= 200} + (Nc)^2 \underbrace{Var[Z]}_{= 600} \\ &- (Na)^2 \underbrace{Var[100]}_{= 0} + (Nb)^2 \underbrace{Var[105]}_{= 0} + (Nc)^2 \underbrace{Var[130]}_{= 0} \end{aligned} \quad (10.27)$$

$$= N^2 (100a^2 + 200b^2 + 600c^2) \quad (10.28)$$

$$= \underline{\underline{100N^2 (a^2 + 2b^2 + 6c^2)}} \quad , \quad \text{q.e.d.} \quad (10.29)$$

- d) Dersom man ønsker minst mulig risiko i investeringen så tilsvare 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:

$$\underline{\underline{Na}} = 100\,000 \cdot 0.6 = \underline{\underline{60\,000}} \quad (10.30)$$

$$\underline{\underline{Nb}} = 100\,000 \cdot 0.3 = \underline{\underline{30\,000}} \quad (10.31)$$

$$\underline{\underline{Nc}} = 100\,000 \cdot 0.1 = \underline{\underline{10\,000}} \quad (10.32)$$

e) Fra oppgave **3b** vet vi at den forventede fortjenesten $E[F]$ er

$$E[F] = 10N(-a + 2b + 5c) \quad (10.33)$$

Fra denne ligningen ser vi umiddelbart at $E[F]$ blir størst når vi kun investerer kun investerer i selskap C, dvs.

$$a = 0 \quad , \quad b = 0 \quad , \quad c = 1 \quad (10.34)$$

Verdien på den forventede fortjenesten $E[F]$ blir da:

$$\underline{\underline{E[F]}} = 10N \cdot 5c \quad (10.35)$$

$$= 10 \cdot 100\,000 \cdot 5 \cdot 1 \text{ NOK} = \underline{\underline{5\,000\,000 \text{ NOK}}} \quad (10.36)$$

altså 5 mill. NOK.

f) Forventningene er: ¹

$$\underline{\underline{E[X]}} = E[90 + 10\epsilon] = 90 + 10 \underbrace{E[\epsilon]}_{=0} = \underline{\underline{90 \text{ NOK}}} \quad (10.37)$$

$$\underline{\underline{E[Y]}} = E[125 + 10\sqrt{2}\epsilon] = 125 + 10\sqrt{2} \underbrace{E[\epsilon]}_{=0} = \underline{\underline{125 \text{ NOK}}} \quad (10.38)$$

$$\underline{\underline{E[X]}} = E[180 + 10\sqrt{6}\epsilon] = 180 + 10\sqrt{6} \underbrace{E[\epsilon]}_{=0} = \underline{\underline{180 \text{ NOK}}} \quad (10.39)$$

g) Variansene er: (se lign.(10.25))

$$\underline{\underline{Var[X]}} = Var[90 + 10\epsilon] = 0 + 10^2 \underbrace{Var[\epsilon]}_{=1} = \underline{\underline{100 \text{ NOK}^2}} \quad (10.40)$$

$$\underline{\underline{Var[Y]}} = Var[125 + 10\sqrt{2}\epsilon] = 0 + (10\sqrt{2})^2 \underbrace{Var[\epsilon]}_{=1} = \underline{\underline{200 \text{ NOK}^2}} \quad (10.41)$$

$$\underline{\underline{Var[X]}} = Var[180 + 10\sqrt{6}\epsilon] = 0 + (10\sqrt{6})^2 \underbrace{Var[\epsilon]}_{=1} = \underline{\underline{600 \text{ NOK}^2}} \quad (10.42)$$

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** er de samme i disse to tilfellene. ²

¹Kommentar: 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.)

²Dette 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

$$P\left(E[X] - \sigma[X] \leq X \leq E[X] + \sigma[X]\right) \quad (10.43)$$

ved oppslag i formelsamlingen. Vi finner: ³

$$\underline{P\left(E[X] - \sigma[X] \leq X \leq E[X] + \sigma[X]\right) = 0.682} \quad (10.44)$$

Altså:

Sannsynligheten for at aksjekursen for selskap A ligger mellom 80 NOK og 100 NOK om ett år er 68.2 %. ⁴

j) Fortjenesten F om ett år er:

$$F = 10N(a + \sqrt{2}b + \sqrt{6}c) \epsilon \quad (10.45)$$

Variansen til fortjenesten F er da: (se lign.(10.25))

$$\underline{\underline{Var[F]}} = Var\left[10N\left(a + \sqrt{2}b + \sqrt{6}c\right) \epsilon\right] \quad (10.46)$$

$$= \left(10N\left(a + \sqrt{2}b + \sqrt{6}c\right)\right)^2 \underbrace{Var[\epsilon]}_{=1} \quad (10.47)$$

$$= \underline{\underline{(10N)^2\left(a + \sqrt{2}b + \sqrt{6}c\right)^2}}, \quad \text{q.e.d.} \quad (10.48)$$

³Se side 65 i 2016-versjonen av formelsamlingen.

⁴Legg merke til at siden $Var[X] = 100 \text{ NOK}^2$ så er $\sigma[X] = 10 \text{ NOK}$. Dessuten er $E[X] = 90 \text{ NOK}$.

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:

$$\underline{E[Y]} = E[X_1 + X_2 + X_3 + X_4 + E_5] \quad (10.49)$$

$$\stackrel{\text{alltid}}{=} E[X_1] + E[X_2] + E[X_3] + E[X_4] + E[X_5] \quad (10.50)$$

$$= \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 \quad (10.51)$$

$$= 0.97 + 2.43 + 4.29 + 5.02 + 7.21 = \underline{19.92} \quad (10.52)$$

NB: Overgangen i lign.(10.49) til (10.50) 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 forventning 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

$$\underline{\sigma[Y]} = \sqrt{Var[Y]} = \sqrt{19.92} = \underline{4.46} \quad (10.53)$$

c) En Poissonfordelt stokastisk variabel er med god tilnærming normalfordelt dersom $\lambda \gtrsim 5$. Siden $\underline{\lambda_Y = E[Y] = 19.92} \gg 5$ så er da Y normalfordelt:

$$Y \sim N[E[Y], \sigma[Y]] \quad (10.54)$$

⁵Se f.eks. side 58 i formelsamlingen fra 2016.

⁶I 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

$$\underline{P(Y > 25)} \stackrel{\text{standardiser}}{=} P\left(\underbrace{\frac{Y - E[Y]}{\sigma[Y]}}_{= Z} > \frac{25 - E[Y]}{\sigma[Y]}\right) \quad (10.55)$$

$$= P\left(Y > \frac{25 - 19.92}{\sqrt{19.92}}\right) \quad (10.56)$$

$$= P(Z > 1.14) \quad (10.57)$$

$$= 1 - P(Z \leq 1.14) \quad (10.58)$$

$$= 1 - \underbrace{G(1.14)}_{= 0.8729} \quad (10.59)$$

$$\stackrel{\text{tabell}}{=} 1 - 0.8729 \quad (10.60)$$

$$= \underline{0.1271} \quad (10.61)$$

Sannsynligheten for at det blir mer enn 25 akutte utrykninger per uke i de 5 aktuelle kommunene er 12.71 %.

⁷I 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

$$\underline{\underline{P(X_2 = 2)}} = \frac{\lambda_2^2}{2!} e^{-\lambda_2} \quad (10.62)$$

$$\lambda_2 = \underline{\underline{2.43}} \quad \frac{2.43^2}{2!} e^{-2.43} \stackrel{\text{kalkis}}{=} \underline{\underline{0.2599}} \quad (10.63)$$

- f) Forventet antall akutte utrykninger i året i Eide kommune:

$$\underline{\underline{E[Z_{\text{år}}]}} = E[Z_1 + Z_2 + \dots + Z_n] = \overbrace{E[Z_1] + E[Z_2] + \dots + E[Z_n]}^{n \text{ stk.}} \quad (10.64)$$

$$= n \underbrace{E[Z_j]}_{\lambda_j = 2.43} = 52 \cdot 2.43 = \underline{\underline{126.36}} \quad (10.65)$$

NB: Overgangen i lign.(10.64) gjelder **alltid**. Uansett om de stokastiske variablene Z_i er uavhengige eller ikke.

- g) Variansen til antall akutte utrykninger i året for Eide kommune:

$$\underline{\underline{Var[Z_{\text{år}}]}} = Var[Z_1 + Z_2 + \dots + Z_n] \quad (10.66)$$

$$\stackrel{\text{uavh.}}{=} \overbrace{Var[Z_1] + Var[Z_2] + \dots + Var[Z_n]}^{n \text{ stk.}} \quad (10.67)$$

$$= n \underbrace{Var[Z_j]}_{\lambda_j = 2.43} = 52 \cdot 2.43 = \underline{\underline{126.36}} \quad (10.68)$$

NB: Overgangen i lign.(10.66) gjelder *kun* dersom de stokastiske variablene Z_j er uavhengige.

⁸Se 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 $Z_{\hat{a}r}$ **normalfordelt**:

$$\underline{\underline{Z_{\hat{a}r} \sim N[E[Z_{\hat{a}r}], Var[Z_{\hat{a}r}]]}} \quad (10.69)$$

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

$$\underline{\underline{n \gtrsim 30}} \quad (10.70)$$

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

$$\underline{P(Z_{\text{år}} < 115)} \stackrel{\text{standardiser}}{=} P\left(\underbrace{\frac{Z_{\text{år}} - E[Z_{\text{år}}]}{\sigma[Z_{\text{år}}]}}_{= Z} < \frac{115 - E[Z_{\text{år}}]}{\sigma[Z_{\text{år}}]}\right) \quad (10.71)$$

$$= P\left(Z < \frac{115 - 126.36}{\sqrt{126.36}}\right) \quad (10.72)$$

$$= P(Z < -1.01) \quad (10.73)$$

$$= 1 - P(Z < 1.01) \quad (10.74)$$

$$= 1 - \underbrace{G(1.01)}_{= 0.8438} \quad (10.75)$$

$$\stackrel{\text{tabell}}{=} 1 - 0.8738 \quad (10.76)$$

$$= \underline{0.1562} \quad (10.77)$$

Sannsynligheten for at det blir mindre enn 115 akutte utrykninger i året i Eide kommune er 15.62 %.

■

⁹I 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:

$$\underline{\underline{R_{xy}}} = \frac{S_{xy}}{S_x \cdot S_y} \quad (10.78)$$

$$= \frac{6863.26}{\sqrt{8\,284\,549} \cdot \sqrt{5.76828}} = \underline{\underline{0.99}} \quad (10.79)$$

b) At $R_{xy} = 0.99$ betyr at det er en veldig sterk positiv korrelasjon mellom x og y .

Det er altså en svært sterk lineær sammenheng mellom x og y med positivt stigningstall.

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

$$\hat{y} = \hat{\alpha} + \hat{\beta}x \quad (10.80)$$

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

$$\underline{\hat{\beta}} = \frac{S_{xy}}{S_x^2} = \frac{6863.26}{8\,284\,549} = \underline{0.000828} \quad (10.81)$$

$$\underline{\hat{\alpha}} = \bar{y} - \hat{\beta}\bar{x} = 3.984 - 0.000828 \cdot 5745.4 = \underline{-0.77} \quad (10.82)$$

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

$$\underline{\underline{\hat{y} = -0.77 + 0.000828 \cdot x}} \quad (10.83)$$

d) Bruker regresjonslinjen $\hat{y} = \hat{y}(x)$ fra oppgave 4c, dvs. lign.(10.83):

$$\hat{y}(7492) = -0.77 + 0.000828 \cdot 7492 = \underline{5.43} \quad (10.84)$$

Rauma kommune med 7492 innbyggere vil, ifølge regresjonslinjen i lign.(10.84), ha 5.43 utrykninger per dag.

e) Løser mhp. x alene:

$$\hat{y} \stackrel{\text{Eq.(10.84)}}{=} \hat{\alpha} + \hat{\beta} \cdot x \quad (10.85)$$

$$\hat{y} - \hat{\alpha} = \hat{\beta} \cdot x \quad (10.86)$$

$$\underline{x = \frac{\hat{y} - \alpha}{\beta}} \quad (10.87)$$

Med $\hat{y} = 6$ utrykninger per uke, $\hat{\alpha} = -0.77$ og $\hat{\beta} = 0.000838$ så får vi:

$$\underline{x = \frac{6 - (-0.77)}{0.000828} = 8176} \quad (10.88)$$

Ifølge regresjonslinjen så må det være minst 8176 innbyggere for at Helse M&R må ha en ekstra person på telefonberedskap.¹⁰

■

¹⁰Dette 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.