V-2016

MAT110

Statistikk 1

Løsningsforslag til eksamensoppgaver

2012 - 2015

Per Kristian Rekdal

Høgskolen i Molde
Vitenskapelig høgskole i logistikk

Innhold

Forord

Løsningsforslag:

Dette er en samling av løsningsforslag til gamle eksamensoppgaver i emnet "*MAT110 Statistikk 1*" ved Høgskolen i Molde. Samlingen inneholder løsningsforslag tilhørende totalt 8 eksamensoppgaver, i perioden fra og med 2012 til og med 2015.

Det finnes også en tilhørende samling med selve eksamensoppgavene til disse løsningforslagene. 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

Kapittel 1

LØSNING: Eksamen 1. juni 2012

"MAT110 Statistikk 1"

Oppgave 1: (sannsynlighetsregning)

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

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

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

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

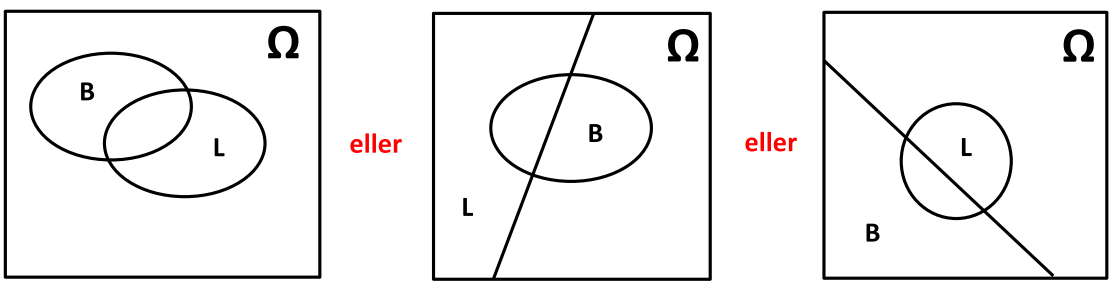
■

Oppgave 2: (økonomi)

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

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

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



Figur 1.1: Venn-diagram.

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

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

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

$$\underline{\underline{P(L)}} \;=\; P(L|B) \cdot P(B) \;+\; P(L|\overline{B}) \cdot P(\overline{B}) \tag{1.2}$$

$$=\; 0.75 \cdot 0.20 \;+\; 0.30 \cdot 0.80 = \underline{\underline{0.39}} \tag{1.3}$$

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

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

Alternativt så kan Bayes lov brukes:

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

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

$$\underline{\underline{P(\overline{B}|L)}} \;=\; 1 - P(B|L) \;=\; 1 - 0.38 \;=\; \underline{\underline{0.62}} \tag{1.6}$$

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

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

$$=\; x_1\cdot P(X=x_1) \;+\; x_2\cdot P(X=x_2) \tag{1.8}$$

$$=\; \Big(-\,10\,000\cdot 0.38 \;+\; 8\,000\cdot 0.62\Big)\,\text{NOK} \;=\; \underline{1\,160\ \text{NOK}} \tag{1.9}$$

<u>Ja</u>, siden forventet <u>fortjeneste er positiv</u> 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 passasjèr er det kun 2 mulige utfall, oppmøte (suksess) eller ikke oppmøte (fiasko).
2. I oppgaven antas samme sannsynlighet p for oppmøte for alle n billettkjøperne.
3. I oppgaven antas det at alle billettkjøperne møter opp uavhengig av hverandre.
4. Det gjennomføres et bestemt antall forsøk, dvs. et bestemt antall passasjèrer i dette tilfellet.

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

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

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

ii) Tolkning:

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

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

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

ii) Tolkning:

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

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

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

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

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

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

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

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

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

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

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

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

$$= 0.0378 \;+\; 0.0118 \;+\; 0.0018 \tag{1.19}$$

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

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)

$$n \cdot p(1-p) \;=\; 5.8 \quad \gtrsim 5 \tag{1.21}$$

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

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

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 $n = 123$:

$$\underline{\underline{E[U]}} = E[b \cdot Y] = b \cdot E[Y] \tag{1.23}$$

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

$$= b \cdot \Big(1 \cdot P(Y=1) + 2 \cdot P(Y=2) + 3 \cdot P(Y=3) \Big) \tag{1.25}$$

$$= b \cdot \Big(\underbrace{1 \cdot \underbrace{P(X=121)}_{=0.0378} + 2 \cdot \underbrace{P(X=122)}_{=0.0118} + 3 \cdot \underbrace{P(X=123)}_{=0.0018}}_{\text{numeriske resultat hentet fra oppgave } \mathbf{3f}\text{, lign.(1.19)}} \Big) \tag{1.26}$$

$$= 5000 \cdot \Big(1 \cdot 0.0378 + 2 \cdot 0.0118 + 3 \cdot 0.0018 \Big) \text{ NOK} \tag{1.27}$$

$$= \underline{\underline{334 \text{ NOK}}} \tag{1.28}$$

h) Siden forventet billettinntekter ved overbooking er

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

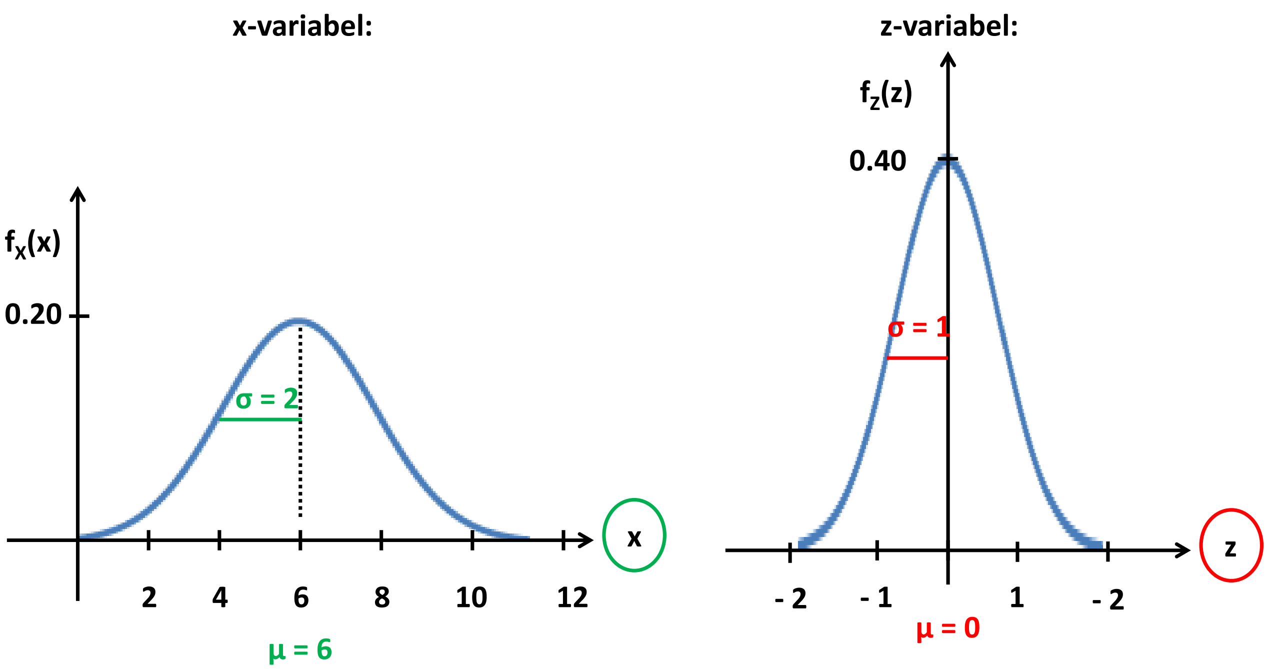
er større enn forventet inntekt ved fullt fly, 91 200 NOK (se oppgave **3d**), så <u>lønner det seg</u> for SAS å overbooke.

■

Oppgave 4: (normalfordeling)

a) Normalfordelingen er en kontinuerlig sannsynlighetsfordeling.

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



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

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

■

Oppgave 5: (økonomi og logistikk)

a) Siden

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

så er den oppgitte sannsynlighetsfordelingen en gyldig fordeling.

b) i) Forventet antall feilleveringer per dag:

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

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

$$= 0 \cdot 0.55 + 1 \cdot 0.25 + 2 \cdot 0.15 + 3 \cdot 0.05 = \underline{\underline{0.70}} \tag{1.32}$$

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

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

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

$$= 0^2 \cdot 0.55 + 1^2 \cdot 0.25 + 2^2 \cdot 0.15 + 3^2 \cdot 0.05 = \underline{1.30} \tag{1.34}$$

Dette innsatt i setningen for varians: (se formelsamling)

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

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

$$\underline{\underline{E[\,\overline{X}\,]}} = E\left[\frac{X_1 + X_2 + \ \ldots\ + X_n}{n}\right] = \frac{1}{n}\left(\overbrace{E[X_1] + E[X_2] + \ \ldots\ + E[X_n]}^{\text{n stk.}}\right) \tag{1.36}$$

$$= \frac{\cancel{n}E[X]}{\cancel{n}} = E[X] = \underline{\underline{0.70}} \tag{1.37}$$

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:

$$\underline{\underline{Var[\,\overline{X}\,]}} = Var\left[\frac{X_1 + X_2 + \ \ldots\ + X_n}{n}\right] \tag{1.38}$$

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

$$= \frac{\cancel{n}Var[X]}{n^{\cancel{2}}} \tag{1.40}$$

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

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

d) Utfylt versjon av tabellen:

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

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

Figur 1.3: Utfylt versjon av tabellen.

Kommentarer:

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

$$\boxed{E[\overline{X}\,] \quad = \quad E[X] \quad (1.42)}$$

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

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

$$\boxed{Var[\,\overline{X}\,] \quad \ll \quad Var[\,X\,] \quad (1.43)}$$

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

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

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

$$\underline{\underline{\overline{X} \;\sim\; \mathrm{N}[\, E[\,\overline{X}\,]\,, Var[\,\overline{X}\,]\,] \qquad = \;\mathrm{N}\left[\, E[\,X\,]\,, \frac{Var[\,X\,]}{n} \right]}} \tag{1.44}$$

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

$$\underline{\underline{n \;\gtrsim\; 30}} \tag{1.45}$$

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

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

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

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

$$= \quad 1 \;-\; P\left(\overline{X} \le \frac{180}{n}\right) \tag{1.48}$$

$$\overset{\text{standardiser}}{=} \quad 1 \;-\; P\left(\underbrace{\frac{\overline{X} - E[\overline{X}]}{\sigma[\,\overline{X}\,]}}_{=\,\overline{Z}} \le \frac{\frac{180}{n} - E[\overline{X}]}{\sigma[\overline{X}]}\right) \tag{1.49}$$

$$= \quad 1 \;-\; P\left(\overline{Z} \;\le\; \frac{\frac{180}{300} - 0.70}{\sqrt{0.0027}}\right) \tag{1.50}$$

$$= \quad \underline{1 \;-\; P\big(\overline{Z} \;\le\; -1.92\big)} \tag{1.51}$$

$$\underline{\underline{P\big(X_1 + X_2 + \ \ldots \ + X_n > 180\big)}} = 1 - P(\overline{Z} \leq -1.92) \tag{1.52}$$

$$= \cancel{1} - \left[\cancel{1} - P(\overline{Z} \leq 1.92) \right] \tag{1.53}$$

$$= P(\overline{Z} \leq 1.92) \tag{1.54}$$

$$= G(1.92) \overset{\text{tabell}}{=} \underline{\underline{0.9726}} \tag{1.55}$$

■

Kapittel 2

LØSNING: Eksamen 10. januar 2013

"MAT110 Statistikk 1"

Oppgave 1: (sannsynlighetsregning)

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

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

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

$$P(E) \cdot P(F) = 0.2392 \tag{2.2}$$

og

$$P(E \cap F) = 0.42 \tag{2.3}$$

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

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

$$P(E) \cdot P(G) \;=\; 0.52 \cdot 0.38 \;=\; 0.1976 \tag{2.4}$$

og

$$P(E \cap G) \;=\; 0.1976 \tag{2.5}$$

så innser vi umiddelbart at E og G er uavhengige.

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

$$\underline{\underline{P(\overline{B})}} \;=\; 1 - P(B) \;=\; 1 - 0.45 \;=\; \underline{\underline{0.55}} \tag{2.6}$$

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

$$\underline{\underline{P(A)}} \;=\; P(A|B) \cdot P(B) \;+\; P(A|\overline{B}) \cdot P(\overline{B}) \tag{2.7}$$

$$=\; 0.33 \cdot 0.45 \;+\; 0.82 \cdot 0.55 \;=\; \underline{\underline{0.5995}} \tag{2.8}$$

■

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:

$$\underline{\underline{E[X]}} \overset{\text{bin.}}{=} n \cdot p = 150 \cdot 0.90 = \underline{\underline{135}} \tag{2.9}$$

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

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

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

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

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

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

ii) Før vårt tilfelle:

$$150 \cdot 0.90\,(1-0.90)\ =\ 13.5\ \gtrsim 5 \tag{2.13}$$

Ja, betingelsen er godt oppfylt for vårt tilfelle.

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

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

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

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

$$=\ P(\,Z \leq 1.50\,) \tag{2.17}$$

$$\stackrel{\text{tabell}}{=}\ \underline{\underline{0.9332}} \tag{2.18}$$

[1] Se setningen på side 26 i formelsamlingen.

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

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

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

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

$$= \;1 \;-\; 0.9332 \;=\; \underline{\underline{0.0668}} \tag{2.21}$$

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

$$\underline{\underline{E[U]}} \;=\; E[\,c \cdot Y\,] \;=\; c \cdot E[Y] \tag{2.22}$$

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

$$= \; c \cdot \Big(1 \cdot P(Y=1) \;+\; 2 \cdot P(Y=2) \Big) \tag{2.24}$$

$$= \; c \cdot \Bigg(\underbrace{1 \cdot \underbrace{P(X \le 140)}_{=\,0.9332} + 2 \cdot \underbrace{P(X > 140)}_{=\,0.0668}}_{\text{resultat hentet fra oppg. 2e og 2f, lign.(2.18) og (2.21).}} \Bigg) \tag{2.25}$$

$$= \; 950 \cdot \Big(1 \cdot 0.9332 \;+\; 2 \cdot 0.0668 \Big) \text{ NOK} \tag{2.26}$$

$$= \; \underline{\underline{1013.46 \;\; \text{NOK}}} \tag{2.27}$$

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

$$\underline{\underline{E[F]}} = E[\, a \cdot X - c \cdot Y \,] \tag{2.28}$$

$$= a \cdot E[X] \; - \; \underbrace{c \cdot E[Y]}_{= E[U]} \tag{2.29}$$

$$= 35 \cdot 135 \; - \; 1013.46 \tag{2.30}$$

$$= \underline{\underline{3711.54 \;\; \text{NOK}}} \tag{2.31}$$

■

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 $\underline{\underline{1}}$.

Oppgave 4: (økonomi og logistikk)

a) Siden

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

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

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

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

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

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

ii) Tolking:

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

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

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

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

$$= 0^2 \cdot 0.70 + 1^2 \cdot 0.15 + 2^2 \cdot 0.10 + 3^2 \cdot 0.05 = \underline{1.0} \tag{2.38}$$

Dette innsatt i "varianssetningen": (se formelsamling)

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

ii) Tolking:

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

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

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

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

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:

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

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

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

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

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

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

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

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

Men <u>variansen</u> til $\overline{X}$ er <u>mye mindre</u>:

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

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

f) i) Med forutsetningene som formulet i oppgaven så gjelder <u>sentralgrensesetningen</u>.

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

$$\underline{\underline{\overline{X} \;\sim\; \mathrm{N}[\,E[\,\overline{X}\,]\,,\,Var[\,\overline{X}\,]\,] \quad = \quad \mathrm{N}\left[\,E[\,X\,]\,,\,\frac{Var[\,X\,]}{n}\,\right]}} \tag{2.48}$$

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

$$\underline{\underline{n \;\gtrsim\; 30}} \tag{2.49}$$

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

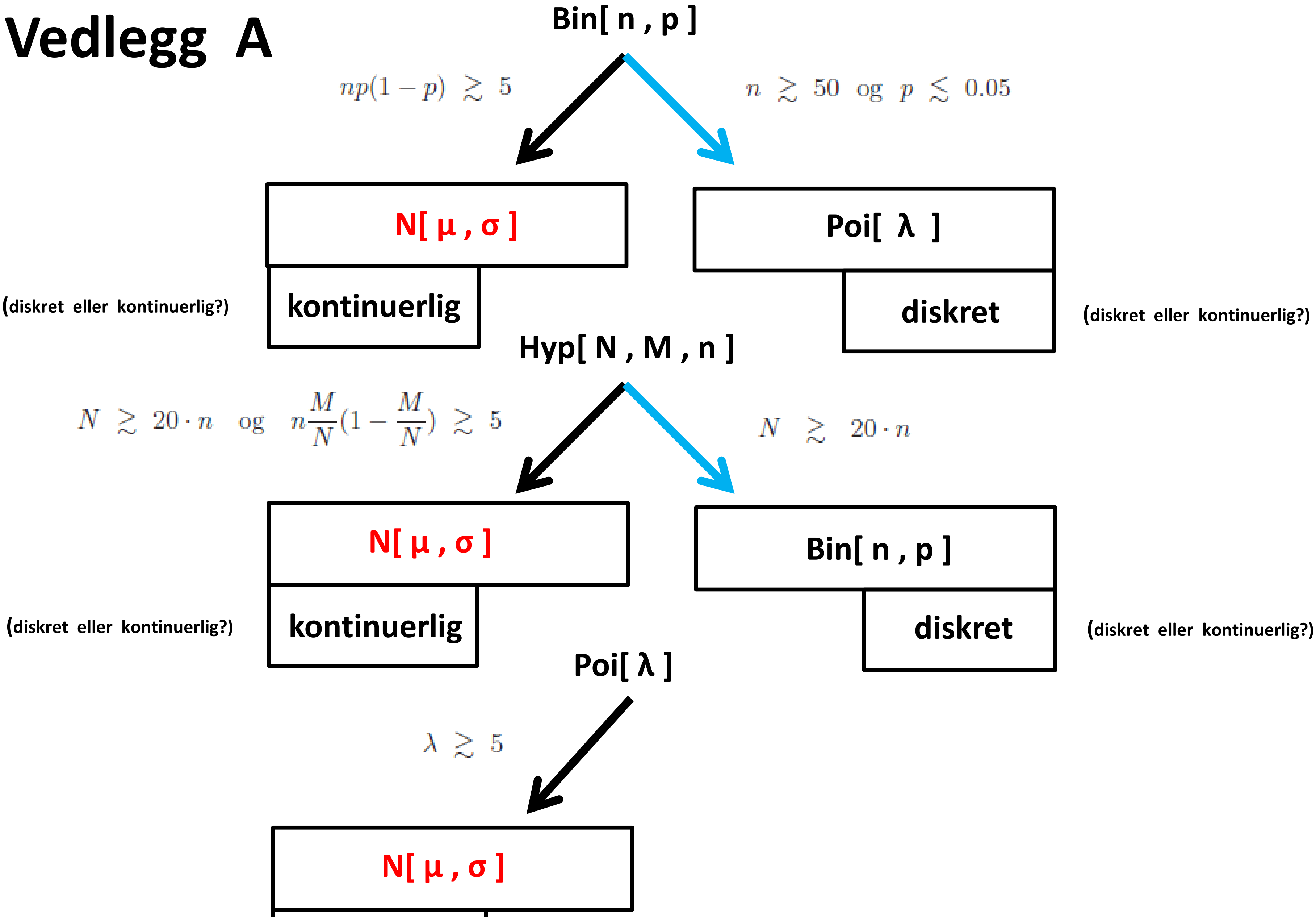
g) Sannsynligheten for at det gjøres mer enn 200 feilleveringer i året per bud: (*uten* heltallskorreksjon)

$$
\begin{aligned}
\underline{\underline{P\Big(X_1 + X_2 + \ \ldots \ + X_n > 200\Big)}} \quad &= \quad P\left(\frac{X_1 + X_2 + \ \ldots \ + X_n}{n} > \frac{200}{n}\right) && (2.50)\\
&= \quad P\left(\overline{X} > \frac{200}{n}\right) && (2.51)\\
&= \quad 1 \ - \ P\left(\overline{X} \leq \frac{200}{n}\right) && (2.52)\\
&\overset{\text{standardiser}}{=} \quad 1 \ - \ P\left(\underbrace{\frac{\overline{X} - E[\overline{X}]}{\sigma[\,\overline{X}\,]}}_{=\,\overline{Z}} \leq \frac{\frac{200}{n} - E[\overline{X}]}{\sigma[\overline{X}]}\right) && (2.53)\\
&= \quad 1 \ - \ P\left(\overline{Z} \ \leq \ \frac{\frac{200}{312} - 0.50}{\sqrt{0.0024}}\right) && (2.54)\\
&= \quad 1 \ - \ P(\overline{Z} \ \leq \ 2.88) && (2.55)\\
&= \quad 1 \ - \ G(2.88) && (2.56)\\
&\overset{\text{tabell}}{=} \quad 1 \ - \ 0.9980 && (2.57)\\
&= \quad \underline{\underline{0.0020}} && (2.58)
\end{aligned}
$$

■

Vedlegg A

Vedlegg A

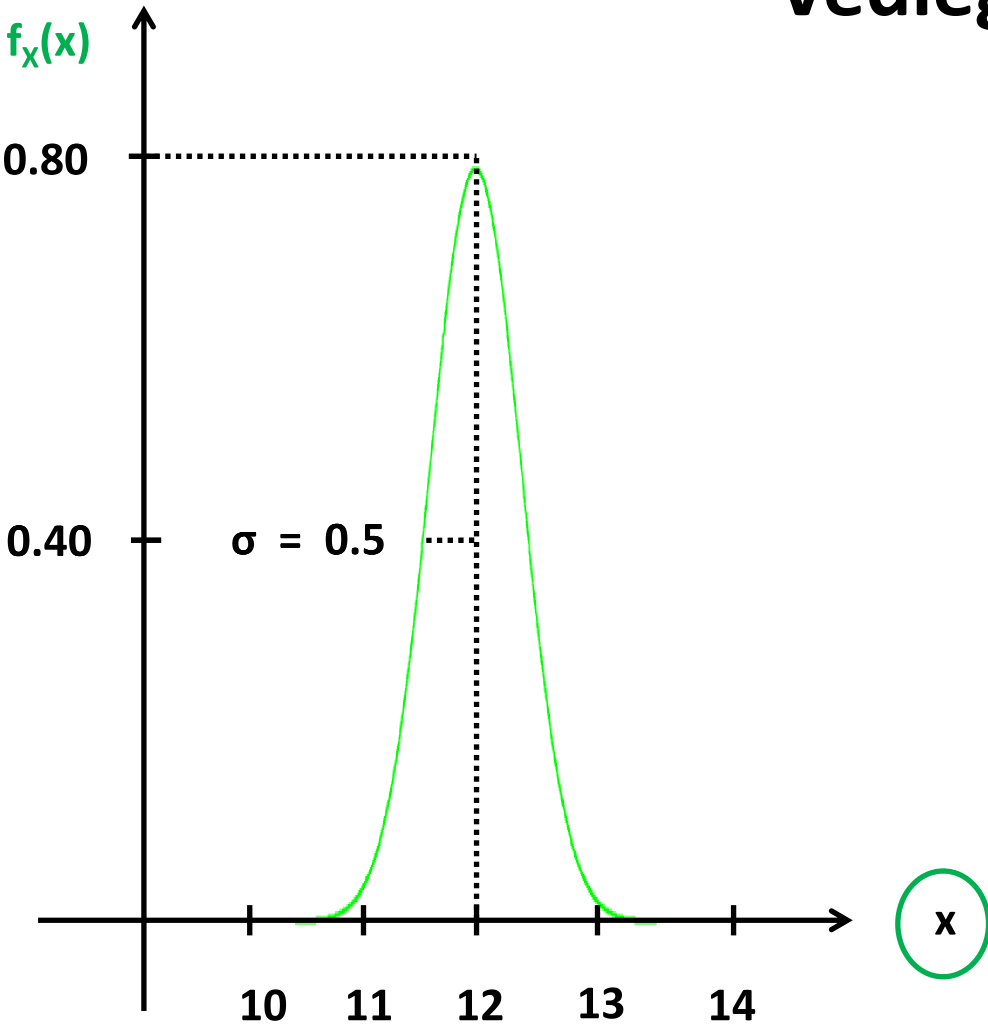


Studentnummer: ____________________

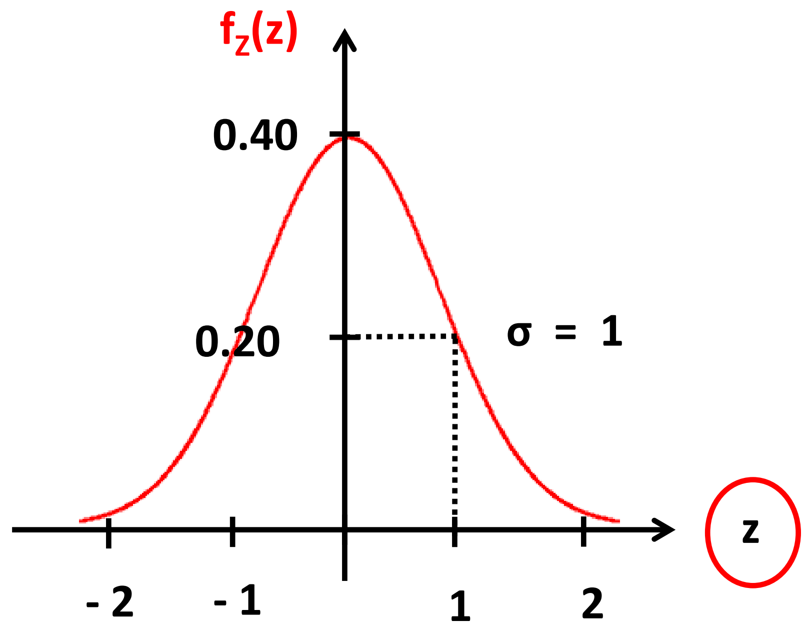
Vedlegg B

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:

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

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

$$\underbrace{\underbrace{P(B_2|B_1)}_{=\ 0.70} \quad \neq \quad \underbrace{P(B_2)}_{=\ 0.05}}_{\text{venstre side IKKE like høyre side}} \tag{3.2}$$

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:

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

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

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:

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

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

e) Bruk F.EKS. total sannsynliget:

$$\underline{\underline{P(B_1 \cap \overline{B}_2)}} \overset{\text{tot.}}{=} \overbrace{P(B_1)}^{=\ 0.05} - \overbrace{P(B_1 \cap B_2)}^{=\ 0.035} \tag{3.7}$$

$$= 0.05 - 0.035 = \underline{\underline{0.015}} \tag{3.8}$$

ELLER definisjonen på betinget sannsynlighet:

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

$$= \Big(1 - \overbrace{P(B_2|B_1)}^{=\ 0.70}\Big) \cdot \overbrace{P(B_1)}^{=\ 0.05} \tag{3.10}$$

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

f) Tolkning:

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

g) Bruk definisjonen av betinget sannsynlighet:

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

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

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

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:

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

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

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

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

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

$$= \quad 0 \cdot 0.57 \; + \; 1 \cdot 0.13 \; + \; 2 \cdot 0.18 \; + \; 3 \cdot 0.10 \; + \; 4 \cdot 0.02 \quad = \; \underline{\underline{0.87}} \tag{3.18}$$

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

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

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

$$= \quad 0^2 \cdot 0.57 \; + \; 1^2 \cdot 0.13 \; + \; 2^2 \cdot 0.18 \; + \; 3^2 \cdot 0.10 \; + \; 4^2 \cdot 0.02 \quad = \; \underline{2.07} \tag{3.20}$$

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

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

c) i) Tolkning:

$$\underline{\underline{E[\,\overline{X}\,]}} = \underline{\underline{\text{forventet antall overtrekk per måned i gjennomsnitt for kundene i Sparebanken Møre}}} \tag{3.22}$$

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

$$\underline{\underline{E[\,\overline{X}\,]}} = E\left[\,\frac{1}{n}\Big(X_1 + X_2 + \ \ldots\ + X_n\Big)\,\right] \tag{3.23}$$

$$\overset{\text{alltid}}{=} \frac{1}{n}\Big(E[X_1] + E[X_2] + \ \ldots\ + \ E[X_n]\Big) \tag{3.24}$$

$$= \frac{1}{n}\Big(\underbrace{E[X] + E[X] + \ \ldots\ + E[X]}_{=\,n}\Big) = \frac{1}{\cancel{n}}\cancel{n}\cdot\underbrace{E[X]}_{=\,0.87} = \underline{\underline{0.87}} \tag{3.25}$$

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:

$$\underline{\underline{Var[\,\overline{X}\,]}} \;=\; \text{forventet variasjon/spredning i antall overtrekk per måned}$$
$$\underline{\underline{\text{ i gjennomsnitt for kundene i Sparebanken Møre}}} \tag{3.26}$$

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

$$\underline{\underline{Var[\,\overline{X}\,]}} \quad = \quad Var\left[\,\frac{1}{n}\left(X_1 + X_2 + \;...\; + X_n\right)\,\right] \tag{3.27}$$

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

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

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 $i = 1, 2, 3, ..., n$

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

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

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

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

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

f) Fra oppgavene foran ser vi at:

$$\underbrace{E[\,\overline{X}\,]}_{=\ 0.87} \quad = \quad \underbrace{E[X]}_{=\ 0.87} \tag{3.30}$$

og at

$$\underbrace{Var[\,\overline{X}\,]}_{=\ 0.002626} \quad \ll \quad \underbrace{Var[X]}_{=\ 1.3131} \tag{3.31}$$

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

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

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

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

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

$$\overset{\text{standardiser}}{=} 1 - P\Bigg(\underbrace{\frac{\overline{X} - E[\,\overline{X}\,]}{\sigma[\,\overline{X}\,]}}_{=\overline{Z}} \leq \frac{\frac{400}{n} - E[\,\overline{X}\,]}{\sigma[\,\overline{X}\,]} \Bigg) \tag{3.35}$$

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

$$= 1 - P(\overline{Z} \ \leq -1.37) \tag{3.37}$$

$$= 1 - \Bigg(1 - \underbrace{P(\overline{Z} \ \leq \ 1.37)}_{=\ G(1.37)} \Bigg) \tag{3.38}$$

$$= \cancel{1} - \cancel{1} + \underbrace{G(1.37)}_{\overset{\text{tabell}}{=}\ 0.9147} \tag{3.39}$$

$$= \underline{\underline{0.9147}} \tag{3.40}$$

Kommentar:

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

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

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

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

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

Vi standardiserer lign.(3.43):

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

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

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

$$\overline{Z}_{\text{grense}} \quad = \quad 1.645 \tag{3.46}$$

Dermed:

$$\overline{Z}_{\text{grense}} \quad = \quad \frac{\overline{X}_{\text{grense}} - E[\overline{X}]}{\sigma[\overline{X}]} \tag{3.47}$$

$$\underline{\overline{X}_{\text{grense}}} \quad = \quad \overline{Z}_{\text{grense}} \cdot \sigma[\overline{X}] \ + \ E[\overline{X}] \tag{3.48}$$

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

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

$$\underline{\underline{X_{\text{grense}}}} = \overline{X}_{\text{grense}} \cdot n \tag{3.50}$$

$$= 0.9543 \cdot 500 \approx \underline{\underline{478}} \tag{3.51}$$

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

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

ii) Tolkning:

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

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

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

ii) Tolkning:

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

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

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

$$= 1 - \Big(P(D = 0) + P(D = 1) + P(D = 2) \Big) \tag{3.55}$$

$$= 1 - P(D = 0) - P(D = 1) - P(D = 2) \tag{3.56}$$

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

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

$$= 1 - 0.2774 - 0.3650 - 0.2305 \tag{3.58}$$

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

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

Siden

$$n \cdot p_d \, (1 - p_d) \quad = \quad 25 \cdot 0.05 \, (1 - 0.05) \quad = \quad 1.1875 \; \ll \; 5 \tag{3.60}$$

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:

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

$$= \quad E\Big[\;\; n \cdot i - D \cdot i - T \cdot i \;\; - \;\; n \cdot (k + k_t) \;\;\Big] \tag{3.62}$$

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

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

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

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

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

Bring:

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

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

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

DHL:

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

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

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

Konklusjon:

Bring har minst forventet utgift.
For å få størst forventet inntekt bør derfor Glamox velge <u>Bring</u>.

■

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)

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

Forventning: (diskret) (formel)

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

Kommentar:

Lokaliseringsmål: tyngdepunkt

Empirisk varians: (formel)

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

Varians: (diskret) (formel)

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

Kommentar:

Spredningsmål: varians

Empirisk kovarians: (formel)

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

Kovarians: (formel)

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

Kommentar:

Et mål på lineær samvariasjon. Ikke normalisert.

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

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

Kommentar:

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

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

Kommentar:

Sterk negativ korrelasjon.

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

Kommentar:

Sterk positiv korrelasjon.

Kapittel 4

LØSNING: Eksamen 6. januar 2014

"MAT110 Statistikk 1"

Oppgave 1: (revisjon)

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

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

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

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

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

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

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

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

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

$$\underline{\underline{P_{\text{A}}}} = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{12+24}{2000+8000} = \underline{\underline{0.0036}} \tag{4.5}$$

$$\underline{\underline{P_{\text{B}}}} = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{20+2}{4000+1000} = \underline{\underline{0.0044}} \tag{4.6}$$

Dermed:

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

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

■

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

$$P_{\text{A11}} > P_{\text{B11}} \tag{4.7}$$
$$P_{\text{A12}} > P_{\text{B12}} \tag{4.8}$$

så er:

$$P_{\text{A}} < P_{\text{B}} \tag{4.9}$$

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

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

Oppgave 2: (logistikk) [3]

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

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

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

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

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

$$= 0.1 + 0.05 + 0.15 + 0.7 + 0 + 0 = \underline{\underline{1}} \tag{4.13}$$

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

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

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

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

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

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

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

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

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

$$\underline{E[\pi(q)]} = E[\, r\, S - w\, q\,] = \underline{r E[S] - w\, q} \tag{4.18}$$

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

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

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

$$\begin{aligned} E[aX + bY] &= aE[X] + bE[Y] \;, & (4.16) \\ E[a] &= a \;, & (4.17) \end{aligned}$$

som man kan finne i kap. 5 i formelsamlingen.

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

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

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

$$P(D \le q^*) = 0.75 \tag{4.22}$$

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

$$P(Z \le Z_0) = 0.75 \tag{4.24}$$

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

$$Z_0 = 0.675 \tag{4.25}$$

Vi løser:

$$Z_0 = \frac{q^* - \mu}{\sigma} \tag{4.26}$$

med hensyn på q^*: $\quad (\; \mu = 3 \text{ og } \sigma = 1.5 \;)$

$$\underline{q^*} = \mu + Z_0 \cdot \sigma \tag{4.27}$$

$$= 3 + 0.675 \cdot 1.5 = \underline{4.0125} \tag{4.28}$$

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

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

$$q^* > \mu \ , \tag{4.29}$$

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.

■

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

Oppgave 3: (økonomi)

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

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

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

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

$$= \; \frac{1}{6} + \frac{1}{6} - \frac{1}{36} \; = \; \frac{11}{36} \; \approx \; \underline{\underline{0.306}} \tag{4.32}$$

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

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

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

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

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

$$P(S_1 \cap S_2 \cap S_3) = P(S_3 \cap S_2 \cap S_1) \tag{4.35}$$

$$= P(S_3|S_2 \cap S_1)P(S_2 \cap S_1) \tag{4.36}$$

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

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

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

$$= 0.6 \cdot 0.4 \cdot \frac{1}{6} = \underline{\underline{0.04}} \tag{4.39}$$

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.

■

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

$$P(A \cap B) = P(A|B)P(B) \tag{4.34}$$

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

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

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 $E[\,\overline{X}\,]$:

$$\underline{\underline{E[\,\overline{X}\,]}} = E\left[\,\frac{1}{n}\left(X_1 + X_2 + \;...\; + X_n\right)\,\right] \tag{4.45}$$

$$\overset{\text{alltid}}{=} \frac{1}{n}\Big(E[X_1] + E[X_2] + \;...\; + \;E[X_n]\Big) \tag{4.46}$$

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

$$= \frac{1}{\cancel{n}}\cancel{n} \cdot E[X] \;=\; E[X] \;=\; \underline{\underline{35}} \tag{4.48}$$

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

f) Tolkning:

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

■

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

$$E[\,aX + bY\,] \;=\; aE[X] + bE[Y]\;, \tag{4.49}$$

som man finner i formelsamlingen.

Kapittel 5

LØSNING: Eksamen 9. mai 2014

"MAT110 Statistikk 1"

Oppgave 1: (revisjon)

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

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

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

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

$$= \;0.05\cdot 0.999 \;+\; 0.80\cdot 0.001 \;=\; \underline{\underline{0.05075}} \tag{5.3}$$

c) Bayes formel:

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

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]

■

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

Oppgave 2: (logistikk)

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

$$\underline{\underline{\text{Poissonfordeling}}} \tag{5.5}$$

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

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

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]

$$\underline{\underline{E[X]}} \;=\; \lambda \;=\; \underline{\underline{115}} \tag{5.7}$$

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

$$\underline{\underline{\sigma[X]}} \;=\; \sqrt{Var[X]} \;=\; \sqrt{\lambda} \;=\; \sqrt{115} \;=\; \underline{\underline{10.72}} \tag{5.8}$$

---

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

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

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

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

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

$$\lambda \;\gtrsim\; 5\;, \tag{5.10}$$

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.)

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

$$= \quad 1 \;-\; P(X \leq 105) \tag{5.12}$$

$$\overset{\text{standardiser}}{=} \quad 1 \;-\; P\Bigg(\underbrace{\frac{X - E[X]}{\sigma[X]}}_{=\,Z} \;\leq\; \frac{105 - E[X]}{\sigma[X]}\Bigg) \tag{5.13}$$

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

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

$$= \quad 1 \;-\; P(Z \;\leq -0.93) \tag{5.16}$$

$$= \quad \cancel{1} \;-\; \Big(\cancel{1} \;-\; P(Z \;\leq\; 0.93)\Big) \quad = \quad \underline{P(Z \;\leq\; 0.93)} \tag{5.17}$$

Dermed:

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

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

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

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

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

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

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

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:

$$Z_{\text{kap}} \quad = \quad 1.645 \tag{5.23}$$

Løser med hensyn på X_{kap} alene:

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

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

$$= Z_{\text{kap}} \cdot \sqrt{\lambda} + \lambda \tag{5.26}$$

$$= 1.645 \cdot \sqrt{115} + 115 = 132.64 \approx \underline{\underline{133}} \tag{5.27}$$

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

■

Oppgave 3: (økonomi)

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

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

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

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

$$\underline{\underline{E[Y]}} \overset{\text{Bin.}}{=} n \cdot p = 345 \cdot 0.11 = \underline{\underline{37.95}} \tag{5.28}$$

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

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

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

$$\underline{\underline{\sigma[X]}} = \sqrt{Var[X]} = \sqrt{33.78} = \underline{\underline{5.81}} \tag{5.30}$$

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

$$\underline{\underline{n \cdot p\,(1-p) \;\gtrsim\; 5}} \tag{5.31}$$

ii) For vårt tilfelle:

$$345 \cdot 0.11\,(1 - 0.11) \;=\; 33.78 \;\gg 5 \tag{5.32}$$

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

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

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

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

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

$$= \;1 \;-\; P(Z \;\leq\; 1.21) \tag{5.36}$$

$$= \;1 \;-\; \underbrace{G(1.21)}_{\overset{\text{tabell}}{=}\;0.8869} \tag{5.37}$$

$$= \;1 \;-\; 0.8869 \;=\; \underline{\underline{0.1131}} \tag{5.38}$$

[4] Se side 60 i formelsamlingen.

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

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

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

Alternativt kan denne løses på følgende måte:

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

$$= \quad 1-\Big(P(B=0)+P(B=1) \Big) \tag{5.42}$$

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

Det er nok at oppgaven løses på en måte.

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

$$\underline{\underline{E[\,B_{\text{tot}}\,]}} \quad = \quad E\big[\,B_1+B_2+\ \ldots\ +B_n\,\big] \tag{5.44}$$

$$\overset{\text{alltid}}{=} \quad \underbrace{E[B_1]+E[B_2]+\ \ldots\ +E[N_n]}_{n\ =\ 345} \quad = \quad n\cdot\underbrace{E[B]}_{=0.11} \quad = \quad 345\cdot 0.11 \quad = \quad \underline{\underline{37.95}} \tag{5.45}$$

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: $\quad (\; n = 3\lambda = 345 \;)$

$$\underline{\underline{Var[B_{\text{tot}}]}} \quad = \quad Var[\, B_1 + B_2 + \; ... \; + B_n \,] \tag{5.46}$$

$$\overset{\text{uavhengig}}{=} \quad Var[B_1] + Var[B_2] + \; ... \; + Var[B_n] \tag{5.47}$$

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

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

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

$$\underline{\underline{B_{\text{tot}} \sim \mathrm{N}\big[\, E[B_{\text{tot}}], Var[B_{\text{tot}}] \,\big]}} \tag{5.49}$$

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

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

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

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

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

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

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

$$= \quad 1 \; - \; 0.8023 \quad = \quad \underline{\underline{0.1977}} \tag{5.55}$$

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:

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

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

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

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:

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

■

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

Oppgave 4: (økonomi)

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

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

$$\underline{\hat{\alpha}} = \overline{y} - \beta\overline{x} = 284 - 7.2 \cdot 3 = \underline{262.4} \tag{5.60}$$

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

$$\underline{\underline{\hat{y} = 262.4 + 7.2 \cdot x}} \tag{5.61}$$

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:

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

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 obervasjonene y_i

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

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

$$\underline{\underline{R^2 \;=\; 0.9744}} \tag{5.63}$$

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.

■

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

Kapittel 6

LØSNING: Eksamen 8. januar 2015

"MAT110 Statistikk 1"

Oppgave 1: (økonomi)

a) Matematisk formulering ihht. notasjon:

$$P(G) = 0.75 \tag{6.1}$$
$$P(D) = 0.25 \tag{6.2}$$

b) Matematisk formulering ihht. notasjon:

$$P(L|G) = 0.10 \tag{6.3}$$
$$P(L|D) = 0.50 \tag{6.4}$$

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

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

$$= 0.10 \cdot 0.75 + 0.50 \cdot 0.25 = \underline{0.20} \tag{6.6}$$

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

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

$$\underline{P(D|L)} \overset{\text{Baye}}{=} P(L|D)\,\frac{P(D)}{P(L)} \tag{6.7}$$

$$= 0.50 \cdot \frac{0.25}{0.20} = \underline{0.625} \tag{6.8}$$

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

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

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

$$= 1 - 0.625 = \underline{0.375} \tag{6.10}$$

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

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

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

$$= f_G \cdot P(G|L) + f_D \cdot P(D|L) \tag{6.12}$$

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

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

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

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

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:

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

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

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

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

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

$$= \quad 1 \; - \; \Big(P(X=0) \; + \; P(X=1) \Big) \tag{6.19}$$

$$= \quad 1 \; - \; \frac{\lambda^0}{0!}\, e^{-\lambda} \; - \; \frac{\lambda^1}{1!}\, e^{-\lambda} \tag{6.20}$$

$$\stackrel{\lambda=0.6}{=} \quad 1 \; - \; \frac{0.6^0}{0!}\, e^{-0.6} \; - \; \frac{0.6^1}{1!}\, e^{-0.6} \tag{6.21}$$

$$\stackrel{\text{kalkis}}{=} \quad 1 \; - \; 0.5488 \; - \; 0.3293 \tag{6.22}$$

$$= \quad \underline{0.1219} \tag{6.23}$$

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

c) Forventet antall "hendelser" per måned:

$$\underline{\underline{E[Y]}} = E\big[X_1 + X_2 + X_3 + X_4\big] \tag{6.24}$$

$$\overset{\text{alltid}}{=} E[X_1] + E[X_2] + E[X_3] + E[X_3] = \overbrace{\lambda + \lambda + \lambda + \lambda}^{4 \text{ stk.}} = 4\,\lambda \tag{6.25}$$

$$= 4 \cdot 0.6 = \underline{\underline{2.4}} \tag{6.26}$$

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:

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

$$= 1 - \Big(P(Y = 0) + P(Y = 1)\Big) \tag{6.28}$$

$$= 1 - \frac{\lambda_Y^0}{0!}\, e^{-\lambda_Y} - \frac{\lambda_Y^1}{1!}\, e^{-\lambda_Y} \tag{6.29}$$

$$\overset{\lambda_Y = 2.4}{=} 1 - \frac{2.4^0}{0!}\, e^{-2.4} - \frac{2.4^1}{1!}\, e^{-2.4} \tag{6.30}$$

$$\overset{\text{kalkis}}{=} 1 - 0.091 - 0.2177 \tag{6.31}$$

$$= \underline{\underline{0.6916}} \tag{6.32}$$

e) Siden

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

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

$$P(Y > 1) \gg P(X > 1) \tag{6.34}$$

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

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

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

b) Forventning $E[X]$:

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

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

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

Tolking:

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

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

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

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

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

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

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

Tolking:

$$\begin{aligned} Var[X] = \; & \text{forventet variasjon/spredning i antall produksjonsfeil} \\ & \text{for en tilfeldig valgt ananasbrus} \end{aligned} \tag{6.43}$$

d) Forventet inntekt per dag på ananasbrus:

$$\underline{E[\,I\,]} \quad = \quad E\Big[p\,n\,0.95 \;-\; b\,n\,X\Big] \tag{6.44}$$

$$= \quad p\,n\,0.95 \;-\; b\,n\,\underbrace{E[\,X\,]}_{=\,0.07} \tag{6.45}$$

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

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

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

$$\underline{\underline{E[\,Y\,]}} \quad = \quad E\,[\,X_1 + X_2 + \;\ldots\; + X_n\,] \quad = \quad \overbrace{E[X_1] + E[X_2] + \;\ldots\; + E[X_n]}^{\text{n stk.}} \tag{6.47}$$

$$= \quad n\,E[X] \quad = \quad 4500 \cdot 0.07 \quad = \quad \underline{\underline{315}} \tag{6.48}$$

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

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

ii) Variansen til antall produksjonsfeil av ananasbrus:

$$\underline{\underline{Var[Y]}} = Var[X_1 + X_2 + \ ... \ + X_n] \tag{6.49}$$

$$\overset{\text{uavh.}}{=} \overbrace{Var[X_1] + Var[X_2] + \ ... \ + Var[X_n]}^{\text{n stk.}} \tag{6.50}$$

$$= n\,Var[X] = 4500 \cdot 0.1051 = \underline{\underline{472.95}} \tag{6.51}$$

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 varibalen Y **normal**fordelt:

$$\underline{\underline{Y \sim \mathrm{N}[\,E[Y]\,, Var[Y]\,]}} \tag{6.52}$$

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

$$\underline{\underline{n \gtrsim 30}} \tag{6.53}$$

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]

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

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

$$= P(Z > 1.61) \tag{6.56}$$

$$= 1 - P(Z \leq 1.61) \tag{6.57}$$

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

$$\overset{\text{tabell}}{=} 1 - 0.9463 \tag{6.59}$$

$$= \underline{0.0537} \tag{6.60}$$

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

■

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

Oppgave 4: (logistikk)

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

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

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

$$\boxed{\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta}\, x \;, \qquad\qquad (6.61)}$$

hvor

$$\boxed{\begin{aligned} \hat{\beta} \;&=\; \frac{S_{xy}}{S_x^2} && (6.62) \\ \hat{\alpha} \;&=\; \overline{y} \;-\; \hat{\beta}\, \overline{x} && (6.63) \end{aligned}}$$

og

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

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

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

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

iii) Forklaringskraften R^2 er et mål på:

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

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

[3] Se side 66 i formelsamlingen fra 2014.

d) Gjenneomsnittene x og y er:

$$\underline{\underline{\overline{x}}} \quad = \quad \frac{1}{n}\sum_{i=1}^{n} x_i \tag{6.65}$$

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

$$\underline{\underline{\overline{y}}} \quad = \quad \frac{1}{n}\sum_{i=1}^{n} y_i \tag{6.67}$$

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

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)

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

$$\underline{\hat{\alpha}} \quad = \quad \overline{y} \; - \; \beta\overline{x} \quad = \quad 243.8 - 6.8 \cdot 2011 \quad = \quad \underline{-13\,431} \tag{6.69b}$$

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

$$\underline{\underline{\hat{y} \;=\; -13\,431 \;+\; 6.8 \cdot x}} \tag{6.70}$$

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

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

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

f) Parameteren $\hat{\beta}$ er stigningstallet for en lineær regresjonslinje.
Parameteren angir estimert/predikert endring $\hat{y}$ når x endres med èn enhet.
Siden (se lign.(6.69a))

$$\hat{\beta} = 6.8 \tag{6.72}$$

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

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

$$305 \quad = \quad -13\,431 \; + \; 6.8 \cdot x \tag{6.73}$$

og løser med hensyn på x:

$$\underline{x} \; = \; \frac{305 \; + \; 13\,431}{6.8} \; = \; \underline{2020} \tag{6.74}$$

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

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

$$\underline{\underline{R^2 \; = \; 0.9538}} \tag{6.75}$$

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 $\underline{\underline{\text{stor forklaringskraft}}}$.

■

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

Høgskolen i Molde
Vitenskapelig høgskole i logistikk

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 forventingen:

$$E[X] = n\frac{M}{N} \tag{7.1}$$

Løser med hensynpå M alane og får:

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

c) Siden X er hypergeometrisk fordelt så er variansen:

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

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

og tilhørende standardavvik:

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

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

$$N \gtrsim 20 \cdot n \tag{7.6}$$

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

ii) For vårt tilfelle er:

$$100\,000 \gtrsim 20 \cdot 1000 \tag{7.8}$$

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

som gir

$$100\,000 \gtrsim 20\,000 \tag{7.10}$$

$$9.9 \gtrsim 5 \tag{7.11}$$

dvs. kritereiene er oppfylte i vårt tilfelle.

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

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

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

$$P\Bigg(\underbrace{\frac{X - E[X]}{\sigma[x]}}_{\equiv Z} \geq \underbrace{\frac{g - E[X]}{\sigma[X]}}_{\equiv Z_0} \Bigg) \quad \overset{\text{standardiser}}{=} \quad 0.10 \tag{7.13}$$

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

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

$$P(Z \geq Z_0) = 0.10 \tag{7.15}$$

$$1 - P(Z < Z_0) = 0.10 \tag{7.16}$$

$$P(Z < Z_0) = 0.90 \tag{7.17}$$

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

$$Z_0 = 1.28 \tag{7.19}$$

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

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

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

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

$$Z_0 = \frac{g - E[X]}{\sigma[X]} \tag{7.20}$$

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

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

$$= 10 + 1.28 \cdot 3.13 = \underline{\underline{14}} \tag{7.22}$$

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:

$$\underline{P(D_1 > 21\,000)} \overset{\text{standardiser}}{=} P\Bigg(\underbrace{\frac{D_1 - \mu_1}{\sigma_1}}_{=\,Z} > \frac{21\,000 - \mu_1}{\sigma_1}\Bigg) \tag{7.23}$$

$$= P\Bigg(Z > \frac{21\,000 - 15\,000}{3000}\Bigg) \tag{7.24}$$

$$= P(Z > 2) \tag{7.25}$$

$$= 1 - P(Z \leq 2) \tag{7.26}$$

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

$$\overset{\text{tabell}}{=} 1 - 0.9772 \tag{7.28}$$

$$= \underline{0.0228} \tag{7.29}$$

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

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

$$D_{tot} = D_1 + D_2 \tag{7.30}$$

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

$$\underline{\underline{E[D_{tot}]}} = E[D_1 + D_2] \tag{7.31}$$

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

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:

$$\underline{\underline{Var[\,D_{tot}\,]}} = Var\,[\,D_1 + D_2\,] \tag{7.33}$$

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

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

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:

$$\underline{\underline{E[F]}} = E\Big[\,(p-k)D_{tot} \; - \; c\,\Big] \tag{7.36}$$

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

$$= \underline{\underline{29\,000 \;\; \text{NOK}}} \tag{7.38}$$

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

$$\underline{P(D_{tot} > 21\,000|S)} \overset{\text{standardiser}}{=} P\Bigg(\underbrace{\frac{D_{tot} - E[D_{tot}]}{\sigma[D_{tot}]}}_{=\,Z_{tot}}\Bigg| S > \frac{21\,000 - 33\,000}{5000}\Bigg) \tag{7.39}$$

$$= P\Bigg(Z_{tot} \; > \; \frac{21\,000 - 33\,000}{5000}\Bigg) \tag{7.40}$$

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

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

$$= \cancel{1} \; - \; \Big(\cancel{1} - P(\,Z_{tot} \leq 2.4\,)\Big) \tag{7.43}$$

$$= P(\,Z_{tot} \;\leq\; 2.4\,) \tag{7.44}$$

$$\overset{\text{tabell}}{=} G(2.4) \tag{7.45}$$

$$= \underline{0.9918} \tag{7.46}$$

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

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

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

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

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 $\underline{\underline{21.66\,\%}}$.

■

[3] Se side 32 i formelsamlingen 2015.

Oppgave 3: (petroleumslogistikk)

a) Siden

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

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

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

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

c) i) Forventet antall innleverte rapporter per dag:

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

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

$$= 0 \cdot 0.09 + 1 \cdot 0.28 + 2 \cdot 0.41 + 3 \cdot 0.17 + 4 \cdot 0.05 = \underline{\underline{1.81}} \quad (7.53)$$

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

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

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

$$= 0^2 \cdot 0.09 + 1^2 \cdot 0.28 + 2^2 \cdot 0.41 + 3^2 \cdot 0.17 + 4^2 \cdot 0.05 = \underline{4.25} \tag{7.55}$$

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

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

d) i) Tolkning:

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

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

$$\underline{\underline{E[\,\overline{X}\,]}} = E\left[\frac{1}{n}\left(X_1 + X_2 + \ ...\ + X_n\right)\right] \tag{7.57}$$

$$\overset{\text{alltid}}{=} \frac{1}{n}\Big(E[X_1] + E[X_2] + \ ...\ + E[X_n]\Big) \tag{7.58}$$

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

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

[4] Se side 39 i formelsamlingen fra 2015.

e) i) Tolkning:

$$\underline{\underline{Var[\,\overline{X}\,]}} \;=\; \text{forventet variasjon/spredning i antall innleverte rapporter per dag}$$
$$\underline{\underline{\text{i gjennomsnitt over et helt år}}} \qquad (7.60)$$

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

$$\underline{\underline{Var[\,\overline{X}\,]}} \;=\; Var\left[\,\frac{1}{n}\left(X_1 + X_2 + \;...\; + X_n\right)\,\right] \qquad (7.61)$$

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

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

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 $i = 1, 2, 3, ..., n$

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

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

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

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

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

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

$$\underbrace{E[\,\overline{X}\,]}_{=\ 1.81} \quad = \quad \underbrace{E[X]}_{=\ 1.81} \tag{7.64}$$

og at

$$\underbrace{Var[\,\overline{X}\,]}_{=\ 0.002668} \quad \ll \quad \underbrace{Var[X]}_{=\ 0.9739} \tag{7.65}$$

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

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

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

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

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

$$\overset{\text{standardiser}}{=} \quad 1 \ - \ P\Bigg(\underbrace{\frac{\overline{X} - E[\,\overline{X}\,]}{\sigma[\,\overline{X}\,]}}_{=\overline{Z}} \ \leq \ \frac{\frac{700}{n} - E[\,\overline{X}\,]}{\sigma[\,\overline{X}\,]} \Bigg) \tag{7.69}$$

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

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

$$= \quad 1 \ - \ G(2.09) \tag{7.72}$$

$$= \quad 1 \ - \ 0.9817 \tag{7.73}$$

$$= \quad \underline{\underline{0.0183}} \tag{7.74}$$

Kommentar:

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

■

Oppgave 4: (økonomi)

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

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

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

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

$$\underline{\underline{R_{xy}}} = \frac{S_{xy}}{S_x \cdot S_y} \tag{7.75}$$

$$= \frac{-110.83}{\sqrt{116.67} \cdot \sqrt{109.54}} = \underline{\underline{-0.98}} \tag{7.76}$$

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

$$\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta} x \tag{7.77}$$

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

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

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

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

$$\underline{\underline{\hat{y} \;=\; 54.68 \;-\; 0.95 \cdot x}} \tag{7.79}$$

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

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

Regresjonslinjen predikerer at kvadratmeterprisen for en leilighet som er 17 km utenfor sentrum er $\underline{\underline{\;38\,530 \text{ NOK}/m^2}}$.

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

$$35 \quad = \quad 54.68 \; - \; 0.95 \cdot x \tag{7.81}$$

Løser med hensyn på x:

$$\underline{x} \; = \; \frac{54.68 \; - \; 35}{0.95} \; = \; \underline{20.72} \tag{7.82}$$

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]

$$\underline{\underline{R^2 \; = \; 0.9613}} \tag{7.83}$$

h) Kommentar 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.

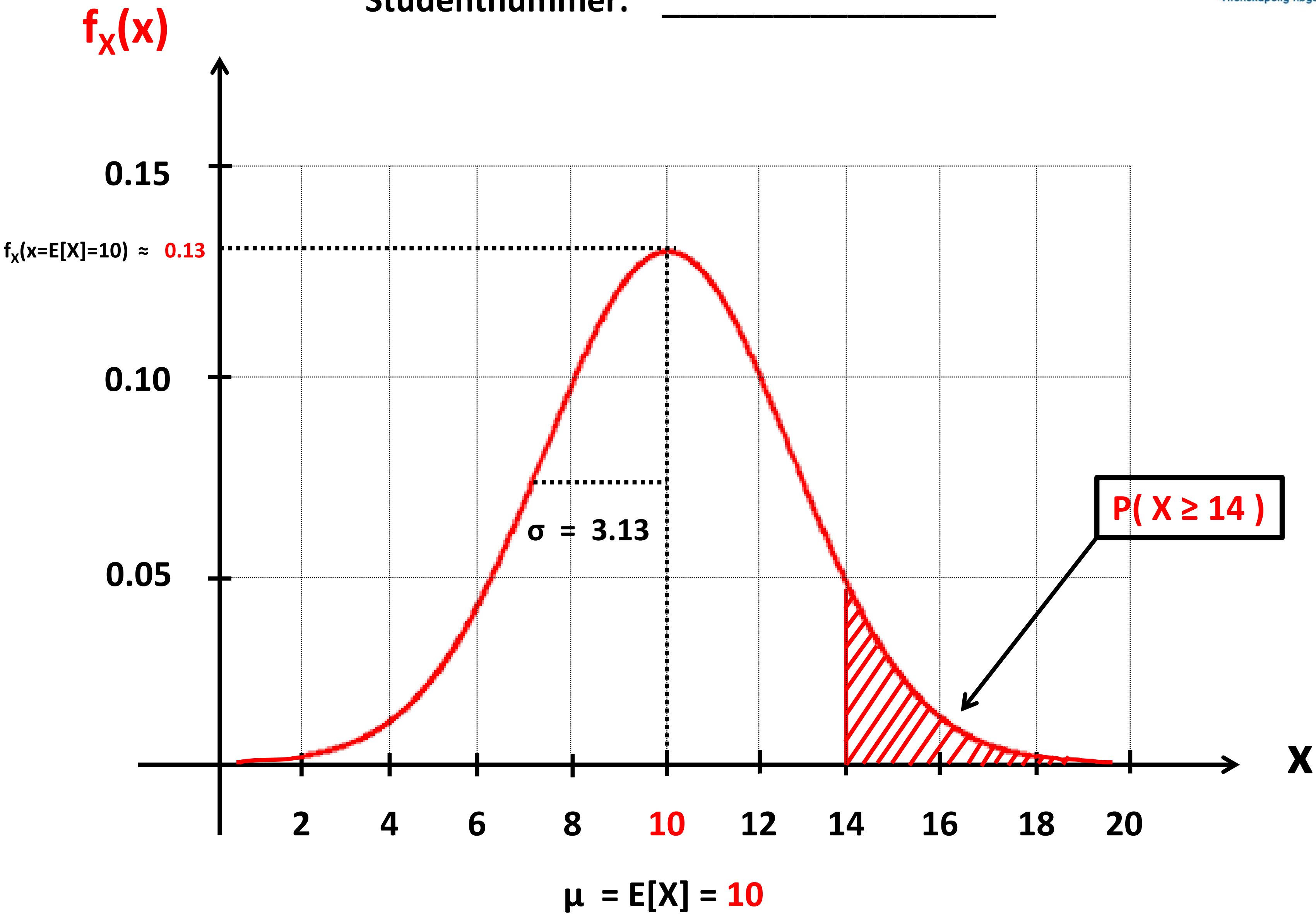
■

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

Vedlegg

Høgskolen i Molde
Vitenskapelig høgskole i logistikk

Studentnummer: ____________________



Kapittel 8

LØSNING: Eksamen 8. januar 2016

"MAT110 Statistikk 1"

Oppgave 1: (logistikk)

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

$$\underline{\underline{P_1}} = \frac{n_1}{n} = \frac{5}{10} = \underline{\underline{0.5}} \tag{8.1}$$

$$\underline{\underline{P_2}} = \frac{n_2}{n} = \frac{3}{10} = \underline{\underline{0.3}} \tag{8.2}$$

$$\underline{\underline{P_3}} = \frac{n_3}{n} = \frac{2}{10} = \underline{\underline{0.2}} \tag{8.3}$$

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

$$\underline{\underline{P_1 + P_2 + P_3}} = 0.5 + 0.3 + 0.2 = \underline{\underline{1}} \tag{8.4}$$

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

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

$$= 10 \cdot \underbrace{P(X = 10)}_{=0.5} + 20 \cdot \underbrace{P(X = 20)}_{=0.3} + 30 \cdot \underbrace{P(X = 30)}_{=0.2}$$

$$= 10 \cdot 0.5 \;+\; 20 \cdot 0.3 \;+\; 30 \cdot 0.2 \;=\; \underline{17} \tag{8.6}$$

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

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

$$\underline{\underline{Var[X]}} \overset{\text{def.}}{=} \sum_{i=1}^{3} \Big(x_i - E[X] \Big)^2 \cdot P(X = x_i) \tag{8.7}$$

$$= (10 - 17)^2 \cdot \underbrace{P(X = 10)}_{=0.5} + (20 - 17)^2 \cdot \underbrace{P(X = 20)}_{=0.3} + (30 - 17)^2 \cdot \underbrace{P(X = 30)}_{=0.2}$$

$$= (10 - 17)^2 \cdot 0.5 \;+\; (20 - 17)^2 \cdot 0.3 \;+\; (30 - 17)^2 \cdot 0.2 \;=\; \underline{\underline{61}} \tag{8.8}$$

med tilhørende standardavvik $\sigma[X]$:

$$\underline{\underline{\sigma[X]}} \overset{\text{def.}}{=} \sqrt{Var[X]} \;=\; \sqrt{61} \;=\; \underline{\underline{7.81}} \tag{8.9}$$

e) Fortventet ventetid $E[V]$:

$$\underline{E[V]} = E[X_1 + X_2 + X_3] \tag{8.10}$$

$$= E[X_1] + E[X_2] + E[X_3] \tag{8.11}$$

$$= E[X] + E[X] + E[X] \tag{8.12}$$

$$= 3\,E[X] = 3 \cdot 17 = \underline{51} \tag{8.13}$$

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

f) Fortventet varians $Var[V]$:

$$\underline{\underline{Var[V]}} = Var[X_1 + X_2 + X_3] \tag{8.14}$$

$$\overset{\text{uavh.}}{=} Var[X_1] + Var[X_2] + Var[X_3] \tag{8.15}$$

$$= Var[X] + Var[X] + Var[X] \tag{8.16}$$

$$= 3\,Var[X] = 3 \cdot 61 = \underline{\underline{183}} \tag{8.17}$$

med tilhørende standardavvik $\sigma[V]$:

$$\underline{\underline{\sigma[V]}} \overset{\text{def.}}{=} \sqrt{Var[V]} = \sqrt{183} = \underline{\underline{13.5}} \tag{8.18}$$

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

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

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

$$= \quad \Big(\, \underbrace{P(X = 30)}_{=\,0.2} \,\Big)^3 \;=\; 0.2^3 \;=\; \underline{\underline{0.008}} \tag{8.21}$$

h) Tolkning:

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

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

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

$$= \quad 1 - \; 0.008 \;=\; \underline{0.992} \tag{8.23}$$

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

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

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

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

$$= \; 3\Big(p(10,10,30) \;+\; p(20,20,10)\Big) \qquad (8.26)$$

$$= \; 3\Big(0.5\cdot 0.5\cdot 0.2 \;+\; 0.3\cdot 0.3\cdot 0.5\Big) \;=\; \underline{0.285} \qquad (8.27)$$

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

■

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:

$$\underline{\underline{P(A)}} \overset{\text{total}}{=} P(A \cap H) \; + \; P(A \cap \overline{H}) \tag{8.28}$$

$$= \; 0.04 \; + \; 0.21 \; = \; \underline{\underline{0.25}} \tag{8.29}$$

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:

$$\underline{\underline{P(H)}} \overset{\text{oppspl.}}{=} P(H \cap R) \; + \; P(H \cap I) \; + \; P(H \cap A) \tag{8.30}$$

$$= \; 0.18 \; + \; 0.10 \; + \; 0.04 \; = \; \underline{\underline{0.32}} \tag{8.31}$$

d) Den generelle multiplikasjonssetningen gir:

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

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

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

eller Bayes lov

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

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

f) Tolkning:

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

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

$$P\big(H|\overline{R}\big) = \frac{P\big(H \cap \overline{R}\big)}{P(\overline{R})} \tag{8.36}$$

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

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

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

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

Tabellen i oppgaven gir videre at: [1]

$$\underline{P\big(H \cap (A \cup I)\big)} \; = \; P\Big((H \cap A) \cup (H \cap I)\Big) \overset{\text{add.}}{=} P(H \cap A) \; + \; P(H \cap I) \tag{8.39}$$

$$= \; 0.04 \; + \; 0.10 \; = \; \underline{0.14} \tag{8.40}$$

Innsatt i lign.(8.37):

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

■

[1] Man 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 Gjemnes på følgende måte:

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

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

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

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

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

$$= \quad 1 \; - \; 0.1496 \; - \; 0.2842 \; - \; 0.2700 \tag{8.46}$$

$$= \quad \underline{\underline{0.2962}} \tag{8.47}$$

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

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

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

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

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 forventing $\lambda_Y = E[Y] = 5.7$.

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

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

$$= \quad 1 \ - \ \Big(P(Y = 0) \ + \ P(Y = 1) \ + \ P(Y = 2) \Big) \tag{8.52}$$

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

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

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

$$= \quad \underline{\underline{0.9232}} \tag{8.56}$$

e) Siden

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

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:

$$P(Y > 2) \;\gg\; P(X > 2) \tag{8.58}$$

<u>NB</u>:

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

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

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

$$\underline{\underline{E[\,X_{\text{år}}\,]}} \;=\; E\,[\,X_1 + X_2 + \;\ldots\; + X_n\,] \;=\; \overbrace{E[X_1] + E[X_2] + \;\ldots\; + E[X_n]}^{\text{n stk.}} \tag{8.60}$$

$$= \;\; n\,\underbrace{E[X]}_{=\lambda_G = 1.9} \;=\; 52 \cdot 1.9 \;=\; \underline{\underline{98.8}} \tag{8.61}$$

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?

$$\underline{\underline{Var[\,X_{\text{år}}\,]}} \quad = \quad Var\Big[\,X_1 + X_2 + \;...\; + X_n\,\Big] \tag{8.62}$$

$$\overset{\text{uavh.}}{=} \quad \overbrace{Var[X_1] + Var[X_2] + \;...\; + Var[X_n]}^{\text{n stk.}} \tag{8.63}$$

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

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 $X_{\text{år}}$ **normal**fordelt:

$$\underline{\underline{X_{\text{år}} \quad \sim \quad \text{N}\Big[\,E[\,X_{\text{år}}\,]\,,\, Var[\,X_{\text{år}}\,]\,\Big]}} \tag{8.65}$$

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

$$\underline{\underline{n \;\gtrsim\; 30}} \tag{8.66}$$

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]

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

$$= P\Bigg(Z > \frac{110 - 98.8}{\sqrt{98.8}} \Bigg) \tag{8.68}$$

$$= P(\,Z > 2.13\,) \tag{8.69}$$

$$= 1 - P(\,Z \leq 1.13\,) \tag{8.70}$$

$$= 1 - \underbrace{G(\,1.13\,)}_{=\,0.8708} \tag{8.71}$$

$$\stackrel{\text{tabell}}{=} 1 - 0.8708 \tag{8.72}$$

$$= \underline{0.1292} \tag{8.73}$$

Sannsynligheten for at det blir mer enn 110 akutte utrykninger i året i Gjemnes kommune er $\underline{\underline{12.92\ \%}}$.

■

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

Oppgave 4: (økonomi)

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

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

iii) R_{xy} er enhetsuavhengig, dvs. ingen enhet. [3]

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

$$\underline{\underline{R_{xy}}} = \frac{S_{xy}}{S_x \cdot S_y} \tag{8.74}$$

$$= \frac{10.36}{\sqrt{6} \cdot \sqrt{19.7}} = \underline{\underline{0.95}} \tag{8.75}$$

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.

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

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

$$\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta} x \tag{8.76}$$

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

$$\underline{\hat{\beta}} \;=\; \frac{S_{xy}}{S_x^2} \;=\; \frac{10.36}{6} \;=\; \underline{1.73} \tag{8.77}$$

$$\underline{\hat{\alpha}} \;=\; \overline{y} \;-\; \beta \overline{x} \;=\; 33.625 - 1.73 \cdot 4.5 \;=\; \underline{25.84} \tag{8.78}$$

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

$$\underline{\underline{\hat{y} \;=\; 25.84 \;+\; 1.73 \cdot x}} \tag{8.79}$$

e) Bruker regresjonslinjen $\hat{y} = \hat{y}(x)$ fra oppgave **4d**, dvs. lign.(8.79):

$$\underline{\hat{y}(12)} \;=\; 25.84 \;+\; 1.73 \cdot 12 \;=\; \underline{46.56} \tag{8.80}$$

Dersom den lineære treden holder seg også i 2016 så predikerer regresjonslinjen at Brunvoll vil få omtrent <u>47 serviceordrer</u>.

f) Med 50 serviceordrer i kvartalet er $\hat{y}(x) = 50$. Ifølge regresjonslinjen er da:

$$\hat{y}(x) \overset{\text{Eq.(8.80)}}{=} 25.84 + 1.73 \cdot x \tag{8.81}$$

$$50 = 25.84 + 1.73 \cdot x \tag{8.82}$$

Løser med hensyn på x:

$$\underline{x} = \frac{50 - 25.84}{1.74} = 13.97 \approx \underline{14} \tag{8.83}$$

Ifølge regresjonslinjen så vil antall serviceordrer bli 50 i kvartalet,
2. kvartal i 2017.

g) Forklaringsstyrken R^2 er: (se formelsamling)

$$\underline{\underline{R^2}} \overset{\text{def.}}{=} 1 - \frac{SSE}{SST} \tag{8.84}$$

$$= 1 - \frac{12.73}{137.88} = \underline{\underline{0.9077}} \tag{8.85}$$

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.

■