

Oppgave 1 — 25%

Post- og logistikkelskapet Bring leverer blant annet pakker fra dør til dør. Bring ønsker å analysere forsinkede leveranser eller leveranser hvor en eller flere av pakkene er skadet.¹

Eksperiment:

Vi observerer om en gitt leveranse er forsinket og om det er skade ved leveransen.

Notasjon: (2-tupple)

$(\bar{s}, \bar{f})$ = en pakke som ikke er skadet og ikke forsinket (1)

$(\bar{s}, f)$ = en pakke som ikke er skadet, men som er forsinket (2)

$(s, \bar{f})$ = en pakke som er skadet, men ikke forsinket (3)

(s, f) = en pakke er både skadet og forsinket (4)

Utfallsrom:

$$\Omega = \{ (\bar{s}, \bar{f}), (\bar{s}, f), (s, \bar{f}), (s, f) \} \quad (5)$$

Vi definerer følgende begivenheter:

S = en pakke blir skadet under leveransen (6)

F = leveransen er forsinket (7)

Bring antar videre at disse to begivenhetene er uavhengige.²



Figur 1: Bring.

¹En "leveranse" er en levering til en spesifikk kunde. En leveranse kan bestå av en eller flere pakker.

²Dvs. at sannsynligheten for at en pakke blir skadet, ikke påvirker sannsynligheten for at leveransen blir forsinket. Rent matematisk betyr uavhengighet i vårt tilfelle at

$$P(S \cap F) = P(S)P(F) \quad P(\bar{S} \cap F) = P(\bar{S})P(F) = (1 - P(S))P(F) \quad (8)$$

$$P(S \cap \bar{F}) = P(S)P(\bar{F}) = P(S)(1 - P(F)) \quad P(\bar{S} \cap \bar{F}) = P(\bar{S})P(\bar{F}) = (1 - P(S))(1 - P(F)) \quad (9)$$

(Her har vi også brukt komplementsetningen).

Anta at Bring har samlet historiske data fra 100 uavhengige leveranser. Resultatene er gitt i tabell 1.

Begivenhet	S	F
Antall pakker	50	10

Tabell 1: Data fra 100 tidligere leveranser.

På bakgrunn av dataene i tabell 1 og store talls lov, kan vi anta at vi har følgende tilnærmede begivenhetssannsynligheter:

$$P(S) \approx f_S = \frac{50}{100} = 0.5 \quad (10)$$

$$P(F) \approx f_F = \frac{10}{100} = 0.1 \quad (11)$$

hvor f -ene er de eksperimentelle sannsynlighetene.

Anta vi setter $P(S) = 0.5$ og $P(F) = 0.1$. Fra antagelsen om at begivenhetene S og F er uavhengige, se lign.(8) og (9) i fotnoten på forrige side, kan vi nå beregne utfallssannsynlighetene på følgende måte:

$$p_{(\bar{s},f)} = P(\bar{S} \cap F) = (1 - P(S))P(F) = (1 - 0.5) \cdot 0.1 = 0.05 \quad (12)$$

$$p_{(s,\bar{f})} = P(S \cap \bar{F}) = P(S)(1 - P(F)) = 0.5 \cdot (1 - 0.1) = 0.45 \quad (13)$$

$$p_{(s,f)} = P(S \cap F) = P(S)P(F) = 0.5 \cdot 0.1 = 0.05 \quad (14)$$

- a)** (2.5%) Vis, på tilsvarende måte som i lign.(12)-(14) ovenfor, at den resterende utfallssannsynligheten er

$$p_{(\bar{s},\bar{f})} = 0.45 \quad (15)$$

Si gjerne hva slags fundamentale læresetninger du bruker for å komme frem til svaret ditt.

Løsning:

Siden begivenhetene er uavhengige så kan vi bruke den spesielle multiplikasjonssetningen og komplementsetningen slik som vist i lign.(9) i fotnoten i oppgaveteksten. Dermed:

$$P_{(\bar{S}, \bar{F})} = P(\bar{S} \cap \bar{F}) = P(\bar{S})P(\bar{F}) = (1 - P(S))(1 - P(F)) \quad (16)$$

$$= (1 - 0.5) \cdot (1 - 0.1) = 0.45 \quad , \quad \text{q.e.d.} \quad (17)$$

- b) (2.5%) For å regne ut relevante sannsynligheter skal vi modellere eksperimentet E som beskrevet ovenfor via en sannsynlighetsmodell.

- i) Hvilken type sannsynlighetsmodell har vi, diskret eller kontinuerlig?
- ii) Er sannsynlighetsmodellen uniform?

Gi korte begrunnelser for dine svar.

Løsning:

- i) Siden utfallsrommet er tellbart, så er utfallsrommet diskret.
- ii) Siden det er ulik sannsynlighet for å trekke de ulike pakkene, se lign.(10)-(11) så er sannsynlighetsmodellen ikke uniform.

- c) (5%) Regn ut sannsynligheten $P(S \cup F)$ ved hjelp av den generelle addisjonssetningen.

Løsning:

Vi bruker at begivenhetene S og F er uavhengige slik at $P(S \cap F) = P(S)P(F)$. I tillegg brukes den generelle addisjonssetningen. Dermed:

$$P(S \cup F) = P(S) + P(F) - P(S \cap F) \quad (18)$$

$$= P(S) + P(F) - P(S)P(F) \quad (19)$$

$$= 0.5 + 0.1 - 0.5 \cdot 0.1 \quad (20)$$

$$= 0.55 \quad (21)$$

Anta på en gitt dag at Bring har $n = 10$ uavhengige leveranser. Vi introduserer da følgende stokastiske variabel:

X = antall feilfrie leveranser, dvs. som er verken skadet eller forsinket

med utfallsrom $\Omega = \{ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 \}$.

d) (5%)

- i)* Hva slags sannsynlighetsfordeling har den stokastiske variabelen X ?
- ii)* Regn ut sannsynligheten for at 7 av disse pakkene er feilfrie.

Løsning:

- i)* For eksperimentet med å levere en pakke og deretter sjekke om den er feilfri eller ikke, gjelder:
 - For en gitt leveranse: to mulige utfall (leveransene er feilfrie eller ikke feilfrie)
 - det gjennomføres et bestemt antall leveranser n ($n = 10$ i vårt tilfelle)
 - Alle utfallene har samme suksesssannsynlighet p , nemlig $p = p(\bar{s}, \bar{f}) = 0.45$.
 - Det er oppgitt i oppgaven at leveransene er uavhengige av hverandre.

Dermed er konklusjonen at X er binomisk fordelt med:

$$X \sim \text{Bin}[n = 10, p = 0.45] \quad (22)$$

- ii)* Siden X er binomisk fordelt så har vi en formel for sannsynligheten:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad (23)$$

Med våre tall får vi da:

$$P(X = 7) = \binom{10}{7} 0.45^7 (1 - 0.45)^{10-7} = 0.0746 \quad (24)$$

Til hver leveranse er det assosiert en kostnad avhengig av utfallet til leveransen som angitt i tabell 2:

Utfall	$(\bar{s}, \bar{f})$	$(\bar{s}, f)$	$(s, \bar{f})$	(s, f)
Kostnad	0	500	900	1 500

Tabell 2: Kostnader assosiert med utfallene til en leveranse.

Bring har beregnet forventet kostnad ved en leveranse til 505 NOK. Dette synes Bring er ganske mye, så de vurderer å investere i en ny og moderne biltype som vil redusere antall skadde leveranser.³

Denne biltypen er imidlertid dyrere i drift, så de har estimert en ekstrakostnad på $a = 100$ NOK per leveranse.

Før de bestemmer seg for å investere i denne biltypen, har de fått tilbud å låne en slik bil og teste den på 100 uavhengige leveranser. Det har Bring sagt ja til.

Bring antar fremdeles uavhengighet mellom skadde pakker og forsinkede leveranser.

- e) (10%) Bring har bestemt seg for å investere i den nye biltypen dersom den forventede kostnaden blir mindre enn eller lik 400 NOK.

Av de 100 testleveransene, hvor mange pakker kan maksimalt bli skadet for at Bring skal gå for den nye biltypen?^{4 5}

³Biltypen vil *ikke* påvirke sannsynligheten for antall forsinkelser, dvs. anta at $P(F)$ er uendret.

⁴Anta at vi kan tilnærme $P(S)$ med den eksperimentelle sannsynligheten f_S .

⁵Dette er en A-oppgave. Hopp gjerne over denne, og ta den til slutt dersom du har tid.

Løsning:

Vi må først definere en sannsynlighetsmodell:

Eksperiment E :

Vi gjennomfører en leveranse med den nye biltypen og observerer om den er forsinket og om en av pakkene er skadet.

Utfallsrommet er som før:

$$\Omega = \left\{ (\bar{s}, \bar{f}), (\bar{s}, f), (s, \bar{f}), (s, f) \right\} \quad (25)$$

Siden Bring bruker den nye biltypen, vil derimot sannsynlighetsloven endre seg.

Vi definerer variabelen

$$x = \text{antall skadde pakker av de 100 testforsøkene.} \quad (26)$$

Da har vi at:

$$P(S) = \frac{x}{100} \quad (\text{tilnærmet } f_S) \quad (27)$$

$$P(F) = \frac{10}{100} \quad (\text{som før}) \quad (28)$$

Utfallssannsynlighetene:

$$p_{(\bar{s}, \bar{f})} = (1 - P(S))(1 - P(F)) = \left(1 - \frac{x}{100}\right) \left(1 - \frac{10}{100}\right) = 0.9 - 0.009x$$

$$p_{(\bar{s}, f)} = (1 - P(S))P(F) = \left(1 - \frac{x}{100}\right) \frac{10}{100} = 0.1 - 0.001x \quad (29)$$

$$p_{(s, \bar{f})} = P(S)(1 - P(F)) = \frac{x}{100} \left(1 - \frac{10}{100}\right) = 0.009x \quad (30)$$

$$p_{(s, f)} = P(S)P(F) = \frac{x}{100} \frac{10}{100} = 0.001x \quad (31)$$

Vi definerer nå en stokastisk variabel C som beregner kostnaden ved en gitt leveranse under denne nye sannsynlighetsmodellen:

U = kostnad for Bring forbundet med skadet pakke og/eller forsinket leveranse

hvor utfallsrommet er, se tabell 2,

$$\Omega_U = \{ 0, 500, 900, 1500 \} \quad (32)$$

Vi har assosiasjonene

$$U(\bar{s}, \bar{f}) = 0 \quad (33)$$

$$U(\bar{s}, f) = 500 \quad (34)$$

$$U(s, \bar{f}) = 900 \quad (35)$$

$$U(s, f) = 1500 \quad (36)$$

Forventningsverdien $E[U]$ er gitt ved

$$E[U] = 0 \cdot p_{(\bar{s}, \bar{f})} + 500p_{(\bar{s}, f)} + 900p_{(s, \bar{f})} + 1500p_{(s, f)} \quad (37)$$

$$= 0 + 500 \cdot (0.1 - 0.001x) + 900 \cdot 0.009x + 1500 \cdot 0.001x \quad (38)$$

$$= 50 - 0.5x + 8.1x + 1.5x \quad (39)$$

$$= 50 + 9.1x \quad (40)$$

Den totale ekstrakostnaden C ved en leveranse er nå gitt ved

$$C = U + a \quad (41)$$

Bruker lign.(40) samt en regneregel for forventning, og får at den forventet total ekstrakostand er:

$$E[C] = E[U + a] = E[U] + a \quad (42)$$

$$= 50 + 9.1x + 100 = 150 + 9.1x \quad (43)$$

Kravet for at Bring skal investere i den nye biltypen er at forventet kostnad ved en laveranse må være mindre enn eller lik 400 NOK:

$$E[C] \leq 400 \quad (44)$$

$$150 + 9.1x \leq 400 \quad (45)$$

$$9.1x \leq 400 - 150 \quad (46)$$

$$x \leq \frac{250}{9.1} \quad (47)$$

$$x \leq 27.47 \quad (48)$$

Siden x må være et heltall, vil Bring investere i den nye biltypen dersom antall skadde pakker ved de 100 testleveransene blir maksimalt 28 pakker.

■

Oppgave 2 — 25%

Transportselskapet *Boreal* har sett nærmere på antall kjøretøy som ankommer Molde fergekai. I tidsrommet 10:00 - 1400 på hverdagene finner *Boreal* at følgende kjøretøy kommer:

- 392 personbiler
- 64 trailere

Anta, for enkelhets skyld, at det kun er disse to typene kjøretøy som ankommer fergekaien i dette tidsrommet, og at raten er konstant for begge tilfellene.⁶

Det kommer en ny ferge **hver halvtime** i det aktuelle tidsrommet. Derfor definerer vi følgende stokastiske variabler:

$$X_1 = \text{antall personbiler som ankommer mellom to fergeavganger,} \\ \text{dvs. i en 30 minutters periode} \quad (49)$$

$$X_2 = \text{antall trailere som ankommer mellom to fergeavganger,} \\ \text{dvs. i en 30 minutters periode}$$

Anta videre at alle kjøretøy som ankommer Molde fergekai er uavhengige av hverandre slik at det er rimelig å modellere de stokastiske variablene X_1 og X_2 via en Poissonfordeling, dvs.

$$X_i \sim \text{Poi}[\lambda_i] \quad (50)$$

med $i = 1, 2$ og tilhørende rater λ_i for antall kjøretøy per 30 minutter.



Figur 2: Ferge.

⁶En rate i vårt tilfelle er antall kjøretøy som ankommer fergekaien per tidsenhet.

- a) (2.5%) Vis, ved en *korte* utregninger, at raten λ_1 som beskriver antall biler **per halvtime** og λ_2 som beskriver antall trailere **per halvtime**, er gitt ved

$$\lambda_1 = 49 \quad , \quad \lambda_2 = 8 \quad (51)$$

Løsning:

I oppgaveteksten opplyses det at det ankommer 392 personbiler og 64 trailere per 4 timer til fergekaiaen.

Antall biler som kommer hver halvtime λ_1 er derfor:

$$\lambda_1 = 392 \cdot \frac{0.5 \text{ time}}{4 \text{ timer}} = 49 \quad , \quad \text{q.e.d.} \quad (52)$$

Antall trailere som kommer hver halvtime λ_2 er derfor:

$$\lambda_2 = 64 \cdot \frac{0.5 \text{ time}}{4 \text{ timer}} = 8 \quad , \quad \text{q.e.d.} \quad (53)$$

- b) (2.5%) Hva er sannsynligheten $P(X_2 = 6)$ for at det kommer 6 trailere mellom to ferger? ⁷

Løsning:

Siden $X_2 \sim \text{Poi}[\lambda_2]$, så kan vi regne ut sannsynligheten ved å bruke følgende formel: (Se boken dersom du ikke har formelen i hodet).

$$P(X_2 = 6) = \frac{\lambda_2^6}{6!} e^{-\lambda_2} \quad (54)$$

$$\stackrel{\lambda_2=8}{=} \frac{8^6}{6!} e^{-8} \stackrel{\text{kalkis}}{=} 0.1221 \quad (55)$$

⁷Bruk 4 desimalers nøyaktighet.

Sannsynligheten for at det kommer 8 trailere mellom to ferger er 12.21 %.

- c) (5%) Hva er sannsynligheten for at det kommer mer enn 6 trailere mellom to ferger, dvs. $P(X_2 > 6)$?⁸

Løsning:

Komplementsetningen:

$$P(X_2 > 6) = 1 - P(X_2 \leq 6) \quad (59)$$

Siden $P(X_2 \leq 6) = \sum_{i=0}^6 P(X_2 = i) = 0.3134$ er oppgitt, får vi

$$P(X_2 > 6) = 1 - P(X_2 \leq 6) \quad (60)$$

$$= 1 - 0.3134 = 0.6866 \quad (61)$$

Sannsynligheten for at det kommer mer enn én trailer mellom to ferger er 68.66 %.

⁸Bruk 4 desimalers nøyaktighet. Oppgitt:

$$P(X_2 \leq 6) = P(X_2 = 0) + P(X_2 = 1) + P(X_2 = 2) + P(X_2 = 3) \quad (56)$$

$$+ P(X_2 = 4) + P(X_2 = 5) + P(X_2 = 6) \quad (57)$$

$$= \sum_{i=0}^6 P(X_2 = i) = 0.3134 \quad (58)$$

Antall plasser i fergen oppgis i enheter av personbiler. En trailer tar 3 ganger større plass enn en personbil. Det totale behovet for antall plasser i fergen er derfor gitt ved:

$$X_{\text{tot}} = X_1 + 3X_2 \quad (62)$$

- d) (2.5%) Vis at forventet antall plasser som det er behov for i perioden mellom to ferger, er

$$E[X_{\text{tot}}] = 73 \quad (63)$$

Løsning:

Forventningen til en Poissonfordeling er $E[X_i] = \lambda_i$, med $i = 1, 2$ i vårt tilfelle. Ratene λ_i er gitt ved lign.(51). Forventningen av $E[X_{\text{tot}}]$ er dermed:

$$E[X_{\text{tot}}] = E[X_1 + 3X_2] \quad (64)$$

$$= E[X_1] + 3E[X_2] \quad (65)$$

$$= \lambda_1 + 3\lambda_2 \quad (66)$$

$$= 49 + 3 \cdot 8 \quad (67)$$

$$= 73, \quad \text{q.e.d.} \quad (68)$$

Legg merke til at vi i overgangen i lign.(65) har brukt regnereglene for forventning. Disse regnereglene gjelder uansett om X_i er uavhengige av hverandre eller ikke.

- e) (5%) Vis at standardavviket $\sigma[X_{\text{tot}}]$ av antall plasser som det er behov for mellom to ferger, er gitt ved

$$\sigma[X_{\text{tot}}] = 11 \quad (69)$$

Løsning:

Variansen til en Poissonfordelt sannsynlighetfordeling er $\text{Var}[X_i] = \lambda_i$, med $i = 1, 2$ i vår tilfelle. Ratene λ_i er gitt ved lign.(51). Variansen til X_{tot} i lign.(62) er dermed:

$$\text{Var}[X_{\text{tot}}] = \text{Var}[X_1 + 3X_2] \quad (70)$$

$$= \text{Var}[X_1] + 3^2 \text{Var}[X_2] + 2 \cdot 1 \cdot 3 \underbrace{\text{Cov}[X_1, X_2]}_{=0} \quad (71)$$

$$= \lambda_1 + 9\lambda_2 \quad (72)$$

$$= 49 + 9 \cdot 8 \quad (73)$$

$$= 121 \quad (74)$$

Legg merke til at vi i overgangen i lign.(71) har brukt regnereglene for varians og at X_i variablene er uavhengige av hverandre. Dermed er kovariansen lik null, dvs.⁹

$$\text{Cov}[X_1, X_2, X_3] = 0 \quad (75)$$

Tilhørende standardavvik er da:

$$\sigma[X_{\text{tot}}] = \sqrt{\text{Var}[X_{\text{tot}}]} = \sqrt{121} = 11 \quad , \quad \text{q.e.d.} \quad (76)$$

⁹For å få fullt hus på denne oppgaven må man nevne at uavhengighet mellom de stokastiske variablene X_i impliserer at kovariansen er null.

Siden forventningsverdien $E[X_{\text{tot}}]$ er stor kan vi med god tilnærming anta at den stokastiske variabelen X_{tot} i lign.(62) kan tilnærmes med en normalfordeling. Med resultatene fra de to foregående deloppgavene har vi da:

$$X_{\text{tot}} \sim N[\mu_{\text{tot}}, \sigma_{\text{tot}}] \quad (77)$$

hvor $\mu_{\text{tot}} = E[X_{\text{tot}}] = 73$ er forventningsverdien og $\sigma_{\text{tot}} = \sigma[X_{\text{tot}}] = 11$ er standardavviket.

Fergene som går strekningen fra Molde fergekai har alle en kapasitet på $X_0 = 80$ plasser.

- f) (7.5%) Anta at 8 personbiler og 3 trailere står igjen fra forrige fergeavgang.

Hva er da sannsynligheten for at alle kjøretøyene får plass i kommende fergeavgang? ¹⁰

Løsning:

Ved å bruke det oppgitte resultatet at X_{tot} er normalfordelt, se lign.(77),

$$X_{\text{tot}} \sim N[\mu_{\text{tot}}, \sigma_{\text{tot}}] \quad (78)$$

så kan vi finne sannsynligheten ved standardisering og deretter tabelloppslag.
Først standardisering:

$$P(X_{\text{tot}} \leq X_0 - 3 \cdot 3 - 8) = P(X_{\text{tot}} \leq 80 - 9 - 8) \quad (79)$$

$$= P(X_{\text{tot}} \leq 63) \quad (80)$$

$$\stackrel{\text{standardiser}}{=} P\left(\underbrace{\frac{X_{\text{tot}} - \mu_{\text{tot}}}{\sigma_{\text{tot}}}}_{= Z_{\text{tot}}} \leq \frac{63 - \mu_{\text{tot}}}{\sigma_{\text{tot}}}\right) \quad (81)$$

$$= P\left(Z_{\text{tot}} \leq \frac{63 - 73}{11}\right) \quad (82)$$

¹⁰Dvs. finn $P(X_{\text{tot}} \leq X_0 - 3 \cdot 3 - 8)$.

$$P(X_{\text{tot}} \leq X_0 - 5 \cdot 3 - 8) = P\left(Z_{\text{tot}} \leq \frac{63 - 73}{11}\right) \quad (83)$$

$$= P(Z \leq -0.91) \quad (84)$$

$$= 1 - \underbrace{P(Z \leq 0.91)}_{= 0.8186} \quad (85)$$

$$= 1 - 0.8186 \quad (86)$$

$$= 0.1814 \quad (87)$$

Dersom det står igjen 8 personbiler og 3 trailere fra forrige ferge så er det bare 18.14 % sannsynlighet for at alle kjøretøyene får plass i fergen.

■

Oppgave 3 — 25%

I tillegg til fergebilletter er også kiosksalg en viktig inntektskilde for *Boreal*. I den sammenheng skal vi se nærmere på salg av brus for en overfart fra Molde fergekai.

For å modellere hvor mange brus som kjøpes i løpet av en overfart så defineres følgende stokastiske variabel:

$$Y_i = \text{antall brus som passasjer nr. } i \text{ kjøper i løpet av en overfart} \quad (88)$$

hvor $i = 1, 2, 3, \dots, n$ og n er antall passasjerer i en overfart fra Molde fergekai. Totalt antall brus som selges på overfarten er da

$$Y = Y_1 + Y_2 + Y_3 + \dots + Y_n \quad (89)$$

La oss, for enkelhets skyld, anta at det er $n = 208$ passasjerer for en bestemt overfart og at 52 av disse passasjerene kjøper brus.

Anta at passasjerene kjøper brus uavhengige av hverandre.

I oppgavene nedenfor skal vi modellere salg av brus som beskrevet ovenfor med to forskjellige modeller, modell A og modell B, som har forskjellige sannsynlighetsfordelinger for $P(Y = y_i)$. Deretter skal vi sammenligne modellene.



Figur 3: Kiosk.

Modell A:

I modell A forenkler vi virkeligheten litt med å anta at fergepassasjerene enten kjøper én brus eller ingen brus i løpet av den aktuelle overfarten. La p være sannsynligheten for at en fergepassasjer kjøper en brus. For modell A blir da sannsynlighetsfordelingen som gitt i tabell 3:

y_i	0	1
$P(Y_i = y_i)$	$1 - p$	p

Tabell 3: Sannsynlighetsfordeling $P(Y_i = y_i)$ for modell A.

- a) (2.5%) For modell A som beskrevet ovenfor så er den stokastiske variabelen Y i lign.(89) binomisk fordelt, dvs. ¹¹

$$Y \sim \text{Bin}[n, p] \quad (90)$$

Med opplysningene oppgitt i oppgaven kan man finne et estimat for suksess-sannsynligheten p .

Vis, ved en enkel og kort regning, at

$$p = 0.25 \quad (91)$$

Løsning:

Parameteren n er antall "forsøk" som gjennomføres i den binomiske forsøksrekken. I vårt tilfelle er:

$$\begin{aligned} \text{et forsøk} &= \text{en fergepassasjer} \\ \text{siden alle fergepassasjerer er potensielle bruskjøpere} \end{aligned} \quad (92)$$

Siden $n = 208$ fergepassasjerer og 52 av disse kjøper brus, så er suksess-sannsynligheten p er da i vårt tilfelle

$$p = \frac{52}{208} = 0.25 \quad (93)$$

¹¹ Dette faktum skal du ikke vise. Bare ta det for gitt.

Dersom noen betingelser er oppfylt så vet vi via sentralgrensesetningen at summen $Y = Y_1 + Y_2 + \dots + Y_n$, som gitt ved lign.(89), med god tilnærming kan beskrives av en normalfordeling for store n , dvs.

$$Y \sim N[\mu_A, \sigma_A] \quad (94)$$

med forventning μ_A og standardavvik σ_A for modell A.

- b)** (2.5%) Hvilken betingelse må de stokastiske variablene Y_i oppfylle for at sentralgrensesetningen, og dermed også lign.(94), skal gjelde?

Løsning:

Sentralgrensesetningen gjelder dersom Y_i er uavhengige identiske fordelte (i.i.d.).

- c)** (5%) Regn ut verdien på μ_A og vis at σ_A er gitt ved

$$\sigma_A = \sqrt{39} \approx 6.2450 \quad (95)$$

Løsning:

Siden Y i utgangspunktet er binomialfordelt for modell A så er forventningen μ_A gitt ved

$$\mu_A = E[Y] = np = 52 \quad (96)$$

og variansen $Var[Y]$

$$Var[Y] = np(1 - p) \quad (97)$$

$$= 208 \cdot 0.25 (1 - 0.25) = 39 \quad (98)$$

dvs. tilhørende standardavvik er:

$$\sigma_A = \sqrt{\text{Var}[Y]} = \sqrt{39} \approx 6.2450 \quad (99)$$

Modell B:

Modell B er en utvidelse av modell A ved å inkludere muligheten for en passasjer til å kjøpe flere brus.

Anta nå at sannsynlighetsfordelingen $P(Y_i = y_i)$ for modell B er som angitt i tabell 4.

y_i	0	1	2	3
$P(Y_i = y_i)$	0.75	0.15	0.075	0.025

Tabell 4: Sannsynlighetsfordeling $P(Y_i = y_i)$ for modell B .

Tilhørende forventning og standardavvik for modell B er:

$$E[Y_i] = 0.3750 \quad , \quad \sigma[Y_i] = 0.7310 \quad (100)$$

(Du trenger ikke regne ut disse størrelsene. Bare ta dem for gitt).

- d)** (2.5%) Ifølge modell B , hvor mange brus kan en passasjer maksimalt kjøpe for en gitt overfart?

Gi en *kort* begrunnelse for svaret ditt. Ingen utregninger behøves.

Løsning:

Fra tabell 4 for modell B er vi at en passasjer maksimalt kan kjøpe 3 brus per overfart.¹²

- e)** (7.5%) Via sentralgrensesetningen vet vi at også for modell B så er den stokastiske variabelen Y gitt ved lign.(89) er normalfordelt for store n , dvs.

$$Y \sim N[\mu_B, \sigma_B] \quad (101)$$

med forventning μ_B og standardavvik σ_B for modell B .

Regn ut verdiene på μ_B og σ_B .

¹²Vi ser fra tabell 4 at det er null sannsynlighet for at en passasjer kjøper mer enn 3 brus for en gitt overfart.

Løsning:

I oppgaven det oppgitt at de stokastiske variablene Y_i har samme sannsynlighetsfordelingen med forventningsverdi $E[Y_i] = 0.3750$. Ved å bruke regnereglene for forventning får vi da

$$\mu_B = E[B_Y] \quad (102)$$

$$= E[Y_1 + Y_2 + \dots + Y_n] \quad (103)$$

$$= E[Y_1] + E[Y_2] + \dots + E[Y_n] \quad (104)$$

$$= nE[Y_i] = 208 \cdot 0.3750 = 78 \quad (105)$$

I oppgaven det oppgitt at de stokastiske variablene Y_i har samme sannsynlighetsfordelingen med standardavvik $\sigma[Y_i] = 77.6284$. Ved å bruke regnereglene for varians får vi da

$$\sigma_B^2 = \text{Var}[Y] \quad (106)$$

$$= \text{Var}[Y_1 + Y_2 + \dots + Y_n] \quad (107)$$

$$= \text{Var}[Y_1] + \text{Var}[Y_2] + \dots + \text{Var}[Y_n] \quad (108)$$

$$= n\text{Var}[Y_i] \quad (109)$$

$$= n\sigma[Y_i]^2 \quad (110)$$

Legg merke til at vi i overgangen i lign.(108) har brukt at variablene Y_i variablene er uavhengige av hverandre. Dermed er kovariansen lik null. ¹³

I oppgaven er det oppgitt at $\sigma[Y_i] = 0.7310$. Det tilhørende standardavvik er da:

$$\sigma_B = \sqrt{\sigma_B^2} = \sqrt{n\sigma^2[Y_i]} = \sqrt{n}\sigma[Y_i] \quad (111)$$

$$= \sqrt{208} \cdot 0.7310 = 10.5428 \quad (112)$$

¹³ Akkurat som i oppgave 2e): For å få fullt hus på denne oppgaven må man nevne at uavhengighet mellom de stokastiske variablene Y_i impliserer at kovariansen er null.

Dersom du har gjort oppgave 3c) og 3e) rett så vet du at både forventningsverdi og standardavvik for modell B er større enn for modell A :

$$\mu_A < \mu_B \quad , \quad \sigma_A < \sigma_B \quad (113)$$

(Du skal ikke vise lign.(113). Bare ta det for gitt.)

- f) (5%) Gi en *kort* begrunnelse for hvorfor det er slik, dvs. begrunn hvorfor lign.(113) stemmer. ¹⁴

Løsning:

Modell $P(Y_i = 0)$ er den samme for begge modellene og siden modell B har, i motsetning til modell A , mulighet for at en passasjer kan kjøpe flere brus, så følger lign.(113).

■

¹⁴Tips: Ved å sammenligne tabell 3 og tabell 4 ser vi at $P(Y_i = 0)$ er den samme for begge modellene. dvs. sannsynligheten for at en passasjer ikke kjøper brus er den samme for begge modellene. Men ellers er sannsynlighetsfordelingene for disse modellene forskjellige. (Husk: Du trenger ikke skrive langt i denne oppgaven. En eller to setninger er nok.)

Oppgave 4 — 25%

Helse Møre og Romsdal har ansvaret for ambulansetjenesten i fylket. Logistikerne i Helse M&R har en mistanke om at antall utrykninger øker fra år til år. Derfor bestemmer de seg for å observere antall utrykninger for hvert år i perioden 2011 – 2016, og bruker notasjonen

$$\underbrace{(x_1, y_1), (x_2, y_2), \dots, (x_6, y_6)}_{\text{observasjoner}}. \quad (114)$$

hvor

$$x_i = \text{år nr. } i \quad (i = 1, \dots, 6) \quad (115)$$

$$y_i = \text{antall utrykninger for år nr. } i \quad (i = 1, \dots, 6) \quad (116)$$

Helse Møre og Romsdal fant følgende:

	2011	2012	2013	2014	2015	2016
x_i (år nr. i)	1	2	3	4	5	6
y_i (antall utrykninger år nr. i)	6726	7595	8451	9350	11 233	12 543

Tabell 5: År x_i og tilhørende antall utrykninger y_i .



Figur 4: Ambulanse.

Fra tabell 5 kan man regne ut korrelasjonskoeffisienten r_{xy} :

$$r_{xy} = 0.9874 \quad (117)$$

Man kan også regne ut det empiriske standardavviket til x , dvs. s_x , og det empiriske standardavviket til y , dvs. s_y . Resultatene er:

$$s_x = 1.8708, \quad s_y = 2213.9440 \quad (118)$$

Gjennomsnittene for variablene i tabell 5 er

$$\bar{x} = 3.5000, \quad \bar{y} = 9316.3333 \quad (119)$$

(Størrelsene i lign.(117), (118) og (119) trenger du ikke å regne ut. Bare ta dem for gitt).

- a)** (2.5%) Gi kort tolkning av den numeriske verdien (tallverdien) til r_{xy} i lign.(117). Tolk både på fortegnet og absoluttverdien til r_{xy} .

Hva sier den numeriske verdien av r_{xy} om graden av korrelasjon mellom observasjonene x og y ?

Men denne verdien for r_{xy} , er det rimelig å modellere dataene i tabell 5 via en lineær regresjonsanalyse? Gi en *kort* begrunnelse for svaret ditt.

Løsning:

Siden $r_{xy} = 0.9874$, altså veldig nære den maksimale verdien 1, så er det et sterk **positiv** korrelasjon mellom x og y . Det er en nesten perfekt lineær sammenheng mellom x og y , med positivt stigningstall. Ja, det er derfor rimelig å modellere dataene i tabell 5 via lineær regresjonsanalyse.

- b)** (2.5%) Man kan vise at sammenhengen mellom korrelasjonskoeffisienten r_{xy} og stigningstallet $\hat{b}$ i en lineær regresjonslinje er gitt ved: ¹⁵

$$r_{xy} = \hat{b} \frac{s_x}{s_y} \quad (120)$$

Hva sier lign.(120) om sammenhengen mellom fortegnet til r_{xy} og fortegnet til stigningstallet $\hat{b}$ for tilhørende regresjonslinje?

Gi en *kort* begrunnelse for svaret ditt. Ingen utregninger behøves.

¹⁵Du skal ikke vise lign.(120). Bare ligningen for gitt.

Løsning:

De empiriske standardavvikene s_x og s_y er alltid positive eller null, dvs. $s_x \geq 0$ og $s_y \geq 0$.

Dermed ser vi fra lign.(120) at fortegnet til r_{xy} er det samme som fortegnet til stigningstallet $\hat{b}$ til tilhørende regresjonslinje.

- c) (2.5%) Vis, ved en kort regning, at den empiriske korrelasjonen s_{xy} er

$$s_{xy} = 4089.6592 \quad (121)$$

med 4 desimales nøyaktighet.

Løsning:

Korrelasjonskoeffisienten r_{xy} er definert ved:

$$r_{xy} = \frac{s_{xy}}{s_x s_y} \quad (122)$$

Løser med hensyn på s_{xy} alene:

$$s_{xy} = r_{xy} s_x s_y \quad (123)$$

$$= 0.9874 \cdot 1.87 \cdot 2213.9 = 4089.6592 \quad , \quad \text{q.e.d.} \quad (124)$$

hvor vi bruker at de numeriske verdiene for r_{xy} , s_x og s_y alle er oppgitt i oppgaven.

La oss anta at observasjonene (x_i, y_i) samvarierer på en slik måte at det er rimelig å modellene sammenhengen mellom dem via regresjonlinjen

$$y_i = \hat{a} + \hat{b}x_i \quad (125)$$

for $i = 1, 2, \dots, n$. I lign.(125) er $\hat{a}$ og $\hat{b}$ konstanter.

d) (10%) Vis at minste kvadraters regresjonslinje er gitt ved: ¹⁶

$$\hat{y} = 5226.5 + 1168.5x \quad (126)$$

Løsning:

For å vise regresjonslinjen i lign.(126) må vi finne koeffisientene $\hat{a}$ og $\hat{b}$ i ligningen

$$\bar{y} = \hat{a} + \hat{b}\bar{x} \quad (127)$$

Vi starter med å finne $\hat{b}$. Fra læreboken vet vi at denne er:

$$\hat{b} = \frac{s_{xy}}{s_x^2} \quad (128)$$

$$= \frac{4089.6592}{1.8708^2} = 1168.5099 \quad (129)$$

hvor vi har brukt at s_x og s_y er oppgitt i oppgaven, og s_{xy} fant vi i oppgave 4c). Etter at $\hat{b}$ er funnet, kan vi regne ut $\hat{a}$ via formelen i læreboken.

¹⁶I mellomregningene kan du gjerne operere med 4 desimaler. Men i sluttsvaret kan du runde av til kun én desimal slik som i lign.(126).

Dermed:

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \quad (130)$$

$$= 9316.3333 - 1168.5099 \cdot 3.5000 = 5226.5487 \quad (131)$$

Regresjonslinjen i lign.(127) er da:

$$\bar{y} = \hat{a} + \hat{b}\bar{x} \quad (132)$$

$$= 5226.5487 + 1168.5099x \quad (4 \text{ desimaler}) \quad (133)$$

$$\approx 5226.5 + 1168.5x \quad , \quad \text{q.e.d. (1 desimal)} \quad (134)$$

- e) (2.5%) Anta at den lineære trenden i antall utrykninger også gjelder helt tilbake til 2008, dvs. før det ble gjort observasjoner og samlet inn data.

Hvor mange utrykninger var det, ifølge regresjonslinjen, i 2008?

Løsning:

År 2008 tilsvarer $x = -2$. Regresjonslinjen sier da at antall utrykninger var:

$$\hat{y}(-2) = 5226.5 + 1168.5 \cdot (-2) = 2889.2 \quad (135)$$

dvs. det er 2889 utrykninger i 2008.

- f) (5%) Anta nå at den lineære trenden i antall utrykninger faktisk gjelder helt frem til antall utrykninger doubler seg sammenlignet med første året det ble gjort observasjoner.

Hva slags årstall skjer denne doblingen?

Løsning:

Første året med observasjoner var i 2011 med $y_1 = 6726$ utrykninger. Denne blir doblet når:

$$\hat{y} = 5226.5 + 1168.5x = 2y_1 \quad (136)$$

Løser med hensyn på x alene:

$$5226.5 + 1168.5x = 2y_1 \quad \left| -5226.5 \right. \quad (137)$$

$$\Updownarrow$$

$$1168.5x = 2y_1 - 5226.5 \quad \left| \frac{1}{1168.5} \right. \quad (138)$$

$$\Updownarrow$$

$$x = \frac{2y_1 - 5226.5}{1168.5} \quad (139)$$

Med $y_1 = 6726$:

$$x = \frac{2 \cdot 6726 - 5226.5}{1168.5} = 7.0396 \approx 7 \quad (140)$$

I 2017 så er antall utrykninger doblet sammenlignet med antall utrykninger første året med observasjoner.

■