V-2017

MAT110

Statistikk 1

Eksamensoppgaver 2012 - 2016

Per Kristian Rekdal

Høgskolen i Molde
Vitenskapelig høgskole i logistikk

Innhold

Forord

Eksamensoppgaver:

Dette er en samling av gamle eksamensoppgaver i emnet "*MAT110 Statistikk 1*" ved Høgskolen i Molde. Samlingen inneholder totalt 10 eksamensoppgaver, i perioden fra og med 2012 til og med 2016.

Det finnes også en tilhørende samling med komplette løsningsforslag til disse eksamensoppgavene. Samlingen med løsningsforslag finnes i et eget hefte, separert fra dette oppgaveheftet.

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?:

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

Per Kr. Rekdal

Høgskolen i Molde

Eksamen i

MAT110 Statistikk 1

(Molde og Kristiansund)

Eksamensdag	**:**	**Fredag 1. juni 2012**
Tid	**:**	**09:00 – 13:00**
Faglærer/telefonnummer	**:**	**Molde:** **Per Kristian Rekdal / 924 97 051** **Kristiansund:** **Knut Nedal / 408 45 257**
Hjelpemidler	**:**	**KD + formelsamling (del 1 & del 2)**
Antall sider inkl. forsiden	**:**	**11 + vedlegg (1 side)**
Målform	**:**	**Norsk (bokmål)**

Noen generelle råd:

- **Kladdark skal ikke leveres. Disse blir ikke sensurert.**

Oppgave 1: (sannsynlighetsregning)

La oss se på begivenhetene A og B. Nedenfor ser du fire matematiske uttrykk for disse begivenhetene. To og to av disse hører sammen og danner to forskjellige *setninger*.

$$P(A) \cdot P(B) \tag{1.1}$$

$$P(A) + P(B) \tag{1.2}$$

$$P(A \cap B) \tag{1.3}$$

$$P(A \cup B) \tag{1.4}$$

a) Sett sammen to og to av disse uttrykkene som hører sammen slik at de danner to setninger.

b) Hvilke navn har disse to setningene?

c) For de to setningene i oppgavene foran, hva er forutsetningen for at de skal gjelde?

■

Figur 1.1: Statistikk.

Oppgave 2: (økonomi)

Sparebanken Møre har erfaringsmessig 20 % kunder med betalingsproblemer. De resterende 80 % av kundene er gode kunde uten betalingsproblemer. Dersom en kunde har inntekt under en viss grense så tilhører kunden det banken klassifiserer som lavinntektsgruppe. I en kartlegging av kundemassen finner banken ut at:

- blant kundene med $\overbrace{\text{betalingsproblemer}}^{=B}$ er 75 % i $\overbrace{\text{lavinntektsgruppen}}^{=L}$, dvs.
  $P(L|B) = 0.75$
- blant kundene som **ikke** $\overbrace{\text{har betalingsproblemer}}^{=\overline{B}}$ er det 30 % i $\overbrace{\text{lavinntektsgruppen}}^{=L}$, dvs.
  $P(L|\overline{B}) = 0.30$

Figur 1.2: Sparebanken Møre.

La

$$\begin{aligned} B &= \text{begivenheten at en tilfeldig valgt kunde har } \textbf{betalingsproblemer} \\ L &= \text{begivenheten at en tilfeldig valgt kunde er i } \textbf{lavinntektsgruppen} \end{aligned}$$

a) i) Hva er $P(B)$? [1]

ii) Hva er $P(\overline{B})$?

[1] Her behøves ingen regning. Svaret finner du ved å se på opplysningene gitt i oppgaven.

b) Tegn et Venn-diagram som viser begivenhetene B og L.

c) Hvor stor andel av bankens kunder er klassifisert som en del av lavinntektsgruppen og, i tillegg, har betalingsproblemer? [2]

d) Hvor stor andel av bankens kunder er i lavinntektsgruppen? [3]

e) Vis at sannsynligheten for at en kunde i lavinntektsgruppen har betalingsproblemer er 38 %, dvs. vis at:

$$\boxed{P(B|L) \;=\; 0.38 \qquad (1.5)}$$

La oss nå **kun** se på nye kunder som er i lavinntektsgruppen. Anta at opplysningene gitt tidligere i oppgaven også gjelder for disse nye kundene. Anta som en forenklet modell at banken *tjener* 8 000 NOK på kunder som ikke har betalingsproblemer, og at banken *taper* 10 000 NOK på kunder som har betalingsproblemer. I denne sammenheng er det hensiktsmessig å definere den stokastiske variabelen:

$$X \;=\; \text{fortjenesten (NOK) til Sparebanken Møre på en ny kunde i lavinntektsgruppen}$$

Denne stokastiske variabelen har følgende sannsynlighetsfordeling:

x	**- 10 000**	**8 000**
P(X=x)	**P(B|L)**	**P($\overline{\text{B}}$|L)**

hvor $P(B|L) = P(X = -10\,000) = 0.38$ fra oppgave **2e**.

[2]Dvs. finn $P(B \cap L)$. Hvilken setning tror du kan være lurt å bruke her?
[3]Dvs. regn ut $P(L)$.

f) Finn den betingede sannsynligheten $P(\overline{B}|L)$.

g) Er det lønnsomt å gi lån til de nye kundene i lavinntektsgruppen? Begrunn svaret ved regning.

■

Oppgave 3: (logistikk)

Du har nylig blitt ansatt som “revenue-analytiker”, dvs. booking-ansvarlig, i SAS ved Kvernberget flyplass i Kristiansund. Denne jobben fikk du mye pga. din nylig avlagte bachelor i logistikk ved Høgskolen i Molde. Du skal tilnærme deg arbeidsoppgavene du har fått ved hjelp av statistikk.

Definer den stokastiske variabelen:

X = antall personer med billett som **faktisk møter opp** til sin flyavgang

La videre n være antall billetter som selges til en bestemt flyavgang. Og la p være sannsynligheten for at en tilfeldig valgt person som har kjøpt billett, faktisk møter opp. For enkelhets skyld, anta at billettkjøperne møter opp uavhengige av hverandre.

a) Forklar hvorfor X er **binomisk** fordelt, dvs. forklar hvorfor $X \sim \text{Bin}[n, p]$. [4]

Basert på historiske data viser det seg at “bare” 95 % av billettkjøperne møter opp, dvs. $p = 0.95$. Til Kvernberget har SAS satt inn Boeing 737-500 maskiner. Disse flyene har en kapasitet på 120 seter. Anta i første omgang at alle plassene blir solgt, dvs. fullt fly. Da er $n = 120$.

b) i) Finn $E[X]$. [5]

ii) Hva betyr $E[X]$ på “godt norsk” i vårt tilfelle? (Dvs. gi en tolkning av $E[X]$.)

Figur 1.3: Kristiansund lufthavn, Kvernberget.

[4]Hvilke 4 kriterier må være oppfylt for at en forsøksserie skal være binomisk? Er disse oppfylt i vårt tilfelle?

[5]Siden $X \sim \text{Bin}[n, p] = \text{Bin}[120, 0.95]$, finnes det da en enkel formel for $E[X]$? Se formelsamling, del II.

c) i) Finn $Var[X]$.

ii) Hva betyr $Var[X]$ på "godt norsk" i vårt tilfelle? (Dvs. gi en tolkning av $Var[X]$.)

Pga. gode ordninger med avbestillingsforsikring så tjener SAS kun penger på passasjerer som møter opp til sin flyavgang. For passasjerer som ikke møter opp blir pengene refundert. Anta at SAS tjener 800 NOK per passasjer på en reise Oslo-Kristiansund for alle passasjerer som møter opp. Dermed kan vi definere en diskret stokastisk variabel I som beskriver **inntekten** til SAS for vår flyavgang:

$$I = a \cdot X \tag{1.6}$$

hvor

$$a = 800 \text{ NOK} \qquad (\text{“a” er bare et konstant tall}) \tag{1.7}$$

d) Hva er forventet inntekt til SAS når alle billetter er solgt, dvs. når $n = 120$? [6]

For at flyavgangen skal gå med færrest mulig tomme seter blir fly ofte "overbooket" - det selges flere billetter enn det er plass til i flyet. Dette fordi SAS regner med at noen passasjerer ikke møter opp. Anta i resten av oppgaven at SAS selger 123 billetter til vårt fly som kun har 120 seter, dvs. en "overbooking" på 3 seter.

e) Hva er forventet inntekt til SAS ved en slik "overbooking", dvs. når $n = 123$?

f) Hva er sannsynligheten for at det møter opp flere passasjerer enn flyet har kapasitet til? [7]

[6] Dvs. finn $E[I]$, hvor I er gitt ved lign.(1.6). (Bruk f.eks. regneregelen i lign.(5.11) i formelsamlingen, del I.)

[7] Dvs. hva er $P(X \geq 121)$? Se side 26 i formelsamlingen, del II.

Dersom en billettkjøper møter opp og ikke får plass fordi flyet er fullt så regner SAS med en utgift på 5000 NOK. Dermed kan vi definere en diskret stokastisk variabel U som beskriver **utgiften** til SAS for vår flyavgang:

$$U = b \cdot Y \tag{1.8}$$

hvor

$$b = 5000 \text{ NOK} \qquad (\text{“b” er bare et konstant tall}) \tag{1.9}$$

$$Y = \text{antall personer med billett som møter opp til sin flyavgang men } \textbf{ikke får plass} \tag{1.10}$$

g) Hva er forventet utgift til SAS ved en slik “overbooking”, dvs. når $n = 123$? [8]

h) Lønner det seg for SAS å “overbooke”? [9]

■

[8] Dvs. finn $E[U]$, hvor U er gitt ved lign.(1.8). Bruk f.eks. regneregelen i lign.(5.11) i formelsamlingen, del I. Bruk deretter gjerne også delsvarene fra oppgave f, dvs. bruk $P(Y = 1) = P(X = 121)$, $P(Y = 2) = P(X = 122)$ og $P(Y = 3) = P(X = 123)$ hvor $X \sim \text{Bin}[n, p] = \text{Bin}[123, 0.95]$.

[9] Gi en kort begrunnelse for svaret med å sammenligne oppgavene **3d**, **3e** og **3h**.

Oppgave 4: (normalfordeling)

Normalfordelingen $\mathrm{N}[\mu, \sigma]$ spiller en svært sentral rolle i statistikk. Blant annet så inngår den i sentralgrensesetningen. Dessuten kan mange sannsynlighetsfordelinger, f.eks. Poisson, hypergeometriske og binomiske fordelinger, med god tilnærmelse beskrives av en normalfordeling under visse betingelser.

Dersom en stokastisk variabel X er normalfordelt, $X \sim \mathrm{N}[\mu, \sigma]$, så er fordelingen gitt ved følgende tetthetsfunksjon $f_X(x)$:

$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi\,\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\,\sigma^2}} \qquad (1.11)}$$

hvor

$$\mu \;=\; E[X] \;=\; \text{forventingen av } X \qquad (\text{ lokalisering av tyngdepunkt }) \quad (1.12)$$

$$\sigma \;=\; \sqrt{Var[X]} \;=\; \text{standardavviket til } X \qquad (\text{ halvbredden til kurven }) \quad (1.13)$$

Via substitusjonen $Z = \frac{X-\mu}{\sigma}$ så kan tetthetsfunksjonen $f_X(x)$ omskaleres til en *standard* normalfordeling $f_Z(z)$. Denne nye *standardiserte* tetthetsfunksjonen $f_Z(z)$ har $\mu = 0$ og $\sigma = 1$, og er gitt ved:

$$\boxed{f_Z(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}} \qquad (1.14)}$$

a) Er normalfordelingen en diskret eller kontinuerlig sannsynlighetsfordeling?

b) I vedlegget til denne eksamensoppgaven ser du to koordinatsystem: Ett med X-variabler og ett med Z-variabler.

i) Tegn inn for hånd tetthetsfunksjon $f_X(x)$ for tilfellet $\mu = 6$ og $\sigma = 2$. [10]

ii) Tegn inn for hånd tilhørende tetthetsfunksjon $f_Z(z)$. [11]

c) Det totale arealet under $f_X(x)$ og $f_Z(z)$ har samme verdi. Hva slags verdi?

■

[10]Bare en enkel håndtegnet skisse er nok. Bruk at toppunktet for $f_X(x)$ er $f_X(x = \mu = 6) \approx 0.20$.

[11]Bare en enkel håndtegnet skisse er nok. Bruk at toppunktet for $f_Z(z)$ er $f_Z(z = \mu = 0) \approx 0.40$.

Oppgave 5: (sentralgrensesetningen, økonomi og logistikk)

Avisen Tidens Krav i Kristiansund ønsker å forbedre logistikken i forbindelse med levering av aviser. For å redusere antall feilleveringer så ønsker de ansatte å se på situasjonen ved hjelp av statistikk. I den sammenheng defineres den stokastiske variabelen:

$$X \;=\; \text{antall feilleveringer som et avisbud gjør per dag}$$

Basert på historiske data finner de ansatte i avisen ut at X har følgende sannsynlighetsfordeling:

x	0	1	2	3
P(X=x)	0.55	0.25	0.15	0.05

Figur 1.4: Sannsynlighetsfordeling $P(X = x)$ basert på historiske data.

Figur 1.5: Tidens Krav i Kristiansund.

a) Vis at sannsynlighetsfordelingen i tabellen i figur 1.4 er en gyldig sannsynlighetsfordeling.

b) i) Finn forventet antall feilleveringer per dag, dvs. $E[X]$.

ii) Finn variansen til antall feilleveringer per dag, dvs. $Var[X]$. [12]

La $X_1, X_2, X_3, \ ... \ , X_n$ være antall feilleveringer for dag nr. 1, 2, 3, ... , n. Gjennomsnittet av antall feilleveringer per dag er da

$$\overline{X} \ = \ \frac{X_1 + X_2 + X_3 + \ ... \ + X_n}{n} \tag{1.15}$$

Anta videre at:

1. antall feilleveringer på ulike dager er <u>u</u>avhengige:

   $X_i \ \sim$ er <u>u</u>avhengige for alle $i = 1, 2, 3, ..., n$

2. alle dager antas å ha samme sannsynlighetsfordeling (gitt ved tabell i figur 1.4):

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

Tidens Krav kommer ut $n = 300$ dager i året.

c) i) Finn forventet antall feilleveringer per dag i *gjennomsnitt* over ett år, dvs. $E[\,\overline{X}\,]$. [13]

ii) Finn variansen til *gjennomsnittet* over ett år av antall feilleveringer per dag, dvs. $Var[\,\overline{X}\,]$. [14]

---

[12]Bruk 4 desimalers nøyaktighet både på $Var[X]$.

[13]Helt *generelt* gjelder regnereglen:

$$E[aX_1 + bX_2] = a\,E[X_1] + b\,E[X_2] \qquad \text{(generelt)} \tag{1.16}$$

hvor a og b er konstanter. Dette gjelder helt generelt, uansett om X_1 og X_2 er uavhengig eller ikke.

[14]Helt *generelt* gjelder:

$$Var[aX_1 + bX_2] = a^2\,Var[X_1] + b^2\,Var[X_2] + 2\,ab\,Cov[X_1, X_2] \qquad \text{(generelt)} \tag{1.17}$$

hvor a og b er konstanter. Dersom X_1 og X_2 er <u>u</u>avhengige så er $Cov[X_1, X_2] = 0$, og lign.(1.17) reduserer seg til

$$Var[aX_1 + bX_2] = a^2\,Var[X_1] + b^2\,Var[X_2] \qquad \text{(gjelder kun dersom } Cov[X_1, X_2] = 0\text{)} \tag{1.18}$$

d) Bruk svarene fra foregående oppgaver og fyll ut tabellen i vedlegget. Kommenter resultatet. [15]

e) i) Med forutsetningene som på forrige side, hvilken setning gjelder da?

ii) Hvilken sannsynlighetsfordeling har da gjennomsnittet $\overline{X}$?

iii) Hvor stor må n ($n =$ antall "forsøk") være, **omtrent**, for at setningen fra oppgave i) skal gjelde? [16]

f) Hva er sannsynligheten for at et avisbud har mer enn 180 feilleveringer i året? Du behøver ikke å bruke heltallskorreksjon. [17]

$\blacksquare$

[15] Hvordan er tyngdepunktet til $P(X = x)$, dvs. $E[X]$, sammenlignet med tyngdepunktet til $P(\overline{X} = \overline{x})$, dvs. $E[\overline{X}]$? Hvordan er spredningen/usikkerheten til $P(X = x)$, dvs. $Var[X]$, sammenlignet med spredningen/usikkerheten til $P(\overline{X} = \overline{x})$, dvs. $Var[\overline{X}]$?

[16] Kun en tommelfingerregel er godt nok her.

[17] Tips:

$$P\big(X_1 + X_2 + \ ... \ + X_n > 180\big) = P\left(\frac{X_1 + X_2 + \ ... \ + X_n}{n} > \frac{180}{n}\right) \tag{1.19}$$

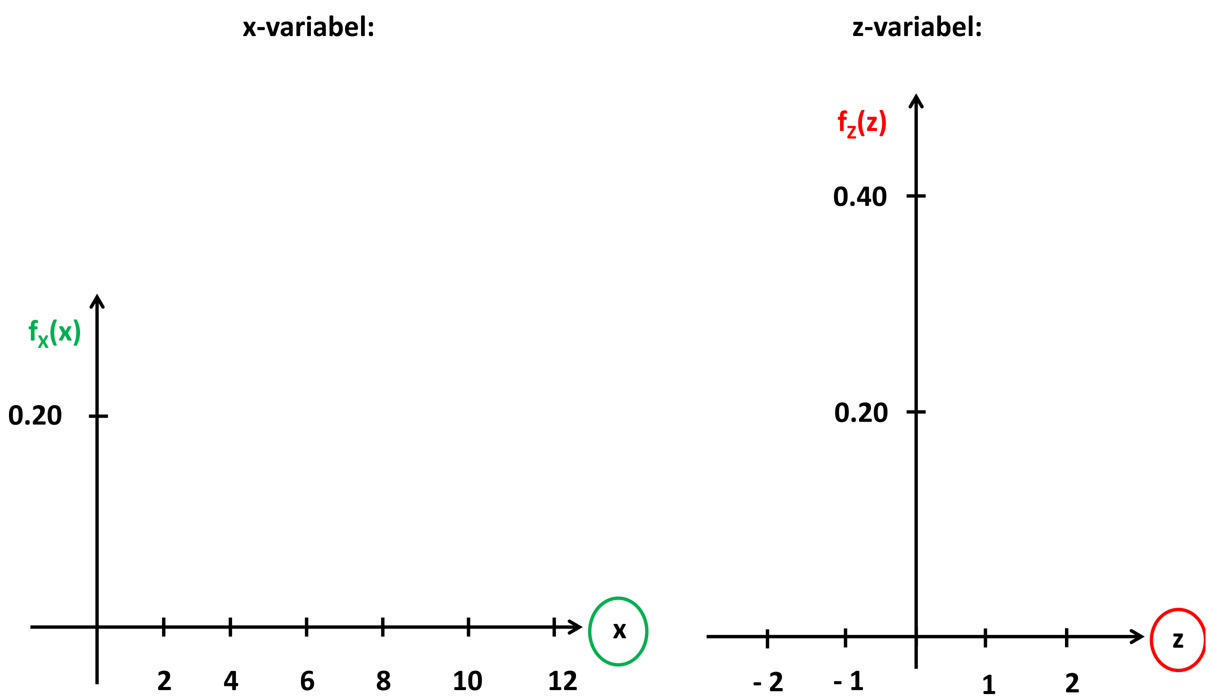
$$= P\Big(\overline{X} > \frac{180}{n}\Big) \tag{1.20}$$

Deretter kan du standardisere lign.(1.20). Bruk 4 desimalers nøyaktighet.

VEDLEGG

(Lever inn arket sammen med besvarelsen din.)

Oppgave 4 b:



Figur 1.6: Tegn inn for hånd tetthetsfunksjone $f_X(x)$ og $f_Z(z)$.

Oppgave 5 d:

tyngdepunkt

E[X]	
E[$\bar{X}$]	

spredning

Var[X]	
Var[$\bar{X}$]	

Figur 1.7: Fyll ut tabellen.

Høgskolen i Molde

Eksamen i

MAT110 Statistikk 1

Eksamensdag	**:**	**Torsdag 10. januar 2013**
Tid	**:**	**09:00 – 13:00**
Faglærer/telefonnummer	**:**	**Per Kristian Rekdal / 924 97 051**
Hjelpemidler	**:**	**KD + formelsamling (del 1 & del 2)**
Antall sider inkl. forsiden	**:**	**12 + vedlegg (2 sider)**
Målform	**:**	**Norsk (bokmål)**

Noen generelle råd:

- **Skriv rett inn. Ikke bruk så mye tid på kladding.**
- **Kladdark skal ikke leveres inn. De blir ikke sensurert.**
- **Det er totalt 4 oppgaver. I *gjennomsnitt* har du èn time per oppgave.**

Oppgave 1: (sannsynlighetsregning)

De ansatte i et finansforetak har studert tre forskjellige aksjer, E, F og G. De har funnet ut at sannsynligheten for at aksje E skal stige, er 52 %. Sannsynligheten for at aksje F skal stige er 46 %, og at G skal stige er 38 %. Matematisk betyr dette:

$$P(E) = 0.52 \tag{2.1}$$
$$P(F) = 0.46 \tag{2.2}$$
$$P(G) = 0.38 \tag{2.3}$$

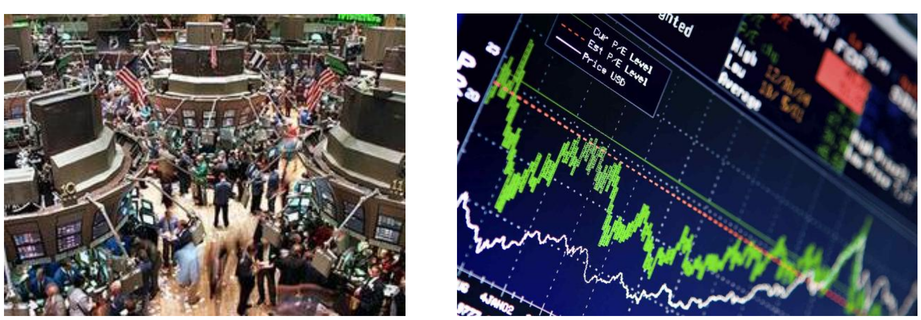
I tillegg har de ansatte funnet ut at sannsynligheten for at både aksje E og F skal stige samtidig er 42 %. Sannsynligheten for at aksje E og G skal stige samtidig er 19.76 %. Matematisk betyr dette:

$$P(E \cap F) = 0.42 \tag{2.4}$$
$$P(E \cap G) = 0.1976 \tag{2.5}$$

a) Regn ut $P(E) \cdot P(F)$.

b) Er hendelsene E og F uavhengige? [1]

c) Er hendelsene E og G uavhengige?



Figur 2.1: Aksjer.

[1] Hvilken setning kan du bruke for å avgjøre dette?

Finansforetaket har også en aksjeportefølje i et helt annet aksjemarked. La oss se på to aksjer i dette markedet, A og B. Anta at sannsynligheten for at aksje B stiger, er 45 %. Dersom B stiger, så er sannsynligheten for at A stiger 33 %. Sannsynligheten for at A stiger dersom B ikke stiger, er 82 %. Matematisk betyr dette:

$$P(B) = 0.45 \tag{2.6}$$
$$P(A|B) = 0.33 \tag{2.7}$$
$$P(A|\overline{B}) = 0.82 \tag{2.8}$$

d) Hva er sannsynligheten for at aksje B ikke stiger? [2]

e) Hva er sannsynligheten for at aksje A stiger? [3]

■

[2] Dvs. hva er $P(\overline{B})$?

[3] Dvs. finn $P(A)$. Bruk setningen for oppsplitting av utfallsrom Ω.

Oppgave 2: (logistikk)

Anta at du skal være transport- og logistikkansvarlig under Åpningsuka 2013 for studentene ved Høgskolen i Molde. En av oppgave du får i den sammenheng er å sørge for at alle studentene kommer seg til Hjertøya med båt. Totalt er det $n = 150$ personer som har takket "*ja*" til invitasjonen. Basert på tidligere åpningsuker viser det seg at det er $p = 90\,\%$ sannsynlighet for at en gitt person som har takket "*ja*", faktisk kommer.

Definer den stokastiske variabelen:

$$X \;=\; \text{antall person som kommer på utflukten til Hjertøya}$$

a) Begrunn hvorfor det er rimelig å anta at X er binomisk fordelt, dvs. begrunn hvorfor: [4]

$$X \;\sim\; \text{Bin}[\,n = 150, p = 0.9\,] \tag{2.9}$$

b) Hva er forventet antall personer som kommer på utflukten til Hjertøya? [5]

Figur 2.2: Åpningsuka ved Høgskolen i Molde. Transport til Hjertøya.

[4] Hvilke 4 krav må være oppfylt for at X skal være binomisk fordelt? Er disse kravene oppfylt i vårt tilfelle?
[5] Dvs. finn $E[X]$.

c) i) Hva er variansen til antall personer som kommer på utflukten, dvs. hva er $Var[X]$?

ii) Hva er det tilhørende standardavviket $\sigma[X]$?

d) i) Dersom en betingelse er oppfylt, så kan en binomisk fordeling tilnærmes med en normalfordeling. Hvilken betingelse er det? [6]

ii) Er denne betingelsen oppfylt i vårt tilfelle?

Hjertøybåten har en kapasitet på 140 passasjerer. Siden $n = 150$ personer er invitert, så må båten kjøre en eller maksimalt to turer, avhengig av hvor mange som faktisk møter opp.

e) Hvor stor sannsynlighet er det for at alle får plass i båten, slik at båten kun trenger å kjøre *en* tur? [7]

f) Hvor stor sannsynlighet er det for at ikke alle får plass i båten på første tur, slik at det må kjøres to turer? [8]

[6] Se formelsamling.
[7] Dvs. finn $P(\text{en tur}) = P(X \leq 140)$. Bruk at $\text{Bin}[\,n, p\,]$ kan tilnærmes med en normalfordeling.
[8] Kan komplementsetningen brukes her?

Utgifter

La oss nå se på utgiftene som eieren av båten pådrar seg ved å frakte passasjerer til Hjertøya. Det viser seg at det koster 950 NOK å kjøre en tur med båten til øya. Her er alle utgifter som lønn, drivstoffutgifter og andre utgifter inkludert.

La Y være en diskret stokastisk variabel som beskriver antall turer som må kjøres for å frakte alle studentene til øya. Den diskrete stokastiske variabelen U definert ved:

$$\boxed{U \;=\; c \cdot Y \quad , \qquad (2.10)}$$

hvor $c = 950$ NOK, beskriver da **utgiften** til båteieren ifm. transport av studentene.

g) Hva er forventet utgift til båteieren ved å frakte studentene til øya? [9]

Fortjeneste

La oss til slutt se på fortjenesten til båteieren ved transport av studentene til Hjertøya. Billettprisen er 35 NOK per person. Den diskrete stokastiske variabelen F definert ved:

$$\boxed{F \;=\; a \cdot X - c \cdot Y \quad , \qquad (2.11)}$$

hvor $a = 35$ NOK, beskriver **fortjenesten** til båteieren forbundet med å transportere studentene til øya.

h) Hva er forventet fortjeneste ved å kjøre studentene til øya? [10]

■

[9] Dvs. finn $E[U]$, hvor U er gitt ved lign.(2.10). Bruk gjerne definisjonen av forventning gitt ved lign.(5.3) i formelsamlingen, del I. Bruk også delsvarene fra oppgave **2e** og **2f** , dvs. bruk $P(Y = 1) = P(X \leq 140)$ og $P(Y = 2) = P(X > 140)$.

[10] Dvs. finn $E[F]$.

Oppgave 3: (normalfordelingen)

I dette kurset har vi lært om 4 fordelinger: binomisk, hypergeometriske, Poisson og normalfordeling. Kortnotasjonsmessig kan disse skrives:

$$\text{Bin}[n, p] \tag{2.12}$$
$$\text{Hyp}[\mu, \sigma] \tag{2.13}$$
$$\text{Poi}[\lambda] \tag{2.14}$$
$$\text{N}[\mu, \sigma] \tag{2.15}$$

Det er sammenhenger mellom disse fordelingene. Under visse betingelser kan en fordeling med god tilnærmelse være lik en annen. Ved hjelp av vedlegg A bakerst i dette oppgavesettet skal det lages en oversikt over disse sammenhengene. Det er 5 store bokser som skal fylles ut med ett av alternativene i lign.(2.12)-(2.15). Til hver av disse store boksene hører det liten boks. I den lille boksen skal det fylles ut om fordelingen er

$$\text{kontinuerlig} \tag{2.16}$$

eller

$$\text{diskret} \tag{2.17}$$

a) Fyll ut alle boksene i vedlegg A.
Både de 5 store boksene (Bin$[n, p]$, Hyp$[\mu, \sigma]$, Poi$[\lambda]$, N$[\mu, \sigma]$) og de 5 små boksene (kontinuerlig, diskret) skal fylles ut.

Dersom en stokastisk variabel X er normalfordelt, $X \sim \text{N}[\mu, \sigma]$, så er fordelingen gitt ved følgende tetthetsfunksjon $f_X(x)$:

$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi\,\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\,\sigma^2}} \qquad (2.18)}$$

hvor

$$\begin{aligned} \mu &= E[X] && = \text{forventingen av } X && (\text{ lokalisering av tyngdepunkt }) && (2.19)\\ \sigma &= \sqrt{Var[X]} && = \text{standardavviket til } X && (\text{ halvbredden til kurven }) && (2.20) \end{aligned}$$

Via substitusjonen $Z = \frac{X-\mu}{\sigma}$ kan tetthetsfunksjonen $f_X(x)$ omskaleres til en **standard** normalfordeling $f_Z(z)$. Denne nye **standardiserte** tetthetsfunksjonen $f_Z(z)$ har $\mu = 0$ og $\sigma = 1$, og er gitt ved:

$$\boxed{f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad (2.21)}$$

b) I vedlegg B ser du to koordinatsystem:
ett med X-variabler og ett med Z-variabler.

i) Tegn inn for hånd tetthetsfunksjon $f_X(x)$ for tilfellet $\mu = 12$ og $\sigma = 0.5$. [11]

ii) Tegn inn for hånd tilhørende tetthetsfunksjon $f_Z(z)$. [12]

NB: være nøye med å tegne bredden på grafene riktig.

c) i) Hva er arealet under grafen $f_X(x)$?

ii) Hva er arealet under grafen $f_Z(z)$?

■

[11] Bare en enkel håndtegnet skisse er nok. Bruk at toppunktet for $f_X(x)$ er $f_X(x = \mu = 6) \approx 0.80$.

[12] Bare en enkel håndtegnet skisse er nok. Bruk at toppunktet for $f_Z(z)$ er $f_Z(z = \mu = 0) \approx 0.40$.

Oppgave 4: (økonomi og logistikk)

Bring er et norsk post- og logistikkselskap med virksomhet i Norden. Det viser seg at tjenestene som Bring leverer, har større hyppighet av feilleveranser av pakker i store byer enn i mindre. Disse feilleveransene koster både tid og penger for Bring. Ledelsen i Bring bestemmer seg derfor for å gjøre tiltak for å forbedre leveringspresisjonen.

Du er ansatt i Bring og har ansvaret for lede en gruppe som skal finne ut mer om feilleveransenes hyppighet. Ledelsen i Bring sier at du skal se nærmere på leveranser av pakker i Stockholm. Dette fordi markedet er stort i den svenske hovedstaden, og gevinsten ved forbedringer av leveringspresisjonen er derfor tilsvarende stor.

Du og den gruppen som du leder bestemmer dere for å definere følgende stokastiske variabel:

$$X = \text{antall feilleveringer som et tilfeldig valgt bud i Stockholm gjør en tilfeldig valgt dag} \tag{2.22}$$

Basert på historiske data for budene i Stockholm finner dere ut at variabelen X har følgende sannsynlighetsfordeling:

x	0	1	2	3
P(X=x)	0.70	0.15	0.10	0.05

Figur 2.3: Sannsynlighetsfordeling $P(X = x)$.

Figur 2.4: Stockholm.

a) Vis at sannsynlighetsfordelingen i tabellen i figur 2.3 er en **gyldig** sannsynlighetsfordeling.

b) i) Finn $E[X]$.

ii) Gi en tolkning av $E[X]$. [13]

c) i) Finn $Var[X]$. [14]

ii) Gi en tolkning av $Var[X]$.

La $X_1, X_2, X_3,$ ... $, X_n$ være antall feilleveringer for et tilfeldig valgt bud for dag nr. 1, 2, 3, ... , n. Gjennomsnittet av antall feilleveringer per dag for et bud hos Bring er da:

$$\boxed{\overline{X} \ = \ \frac{X_1 + X_2 + X_3 + \ ... \ + X_n}{n} \qquad (2.23)}$$

Anta videre at:

1. antall feilleveringer på ulike dager er uavhengige:

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

2. alle dager antas å ha samme sannsynlighetsfordeling (gitt ved tabell i figur 2.3):

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

Pakker leveres også på lørdager, så totalt er det $n = 312$ dager i året hvor Bring leverer pakker.

[13] Dvs. hva betyr $E[X]$ på "godt norsk" i vårt tilfelle?
[14] Bruk 4 desimalers nøyaktighet på $Var[X]$.

d) i) Hva er forventet antall feilleveringer per dag i *gjennomsnitt* over ett år for et bud hos Bring? [15]

ii) Hva er variansen til *gjennomsnittet* over ett år av antall feilleveringer per dag for et bud hos Bring? [16]

e) Sammenlign $E[X]$ fra oppgave **4b** med $E[\,\overline{X}\,]$ fra oppgave **4d**.
Kommenter resultatene. [17]

f) i) Med forutsetningene som på forrige side, hvilken setning gjelder da?

ii) Hvilken sannsynlighetsfordeling har da gjennomsnittet $\overline{X}$?

iii) Hvor stor må n ($n =$ antall "forsøk") være, **omtrent**, for at setningen fra oppgave i) skal gjelde? [18]

---

[15]Dvs. finn $E[\,\overline{X}\,]$. Bruk gjerne regnereglen:

$$E[aX_1 + bX_2] = a\,E[X_1] + b\,E[X_2] \qquad \text{(generelt)} \tag{2.24}$$

hvor a og b er konstanter. Dette gjelder helt generelt, uansett om X_1 og X_2 er uavhengig eller ikke.

[16]Dvs. finn $Var[\,\overline{X}\,]$. Helt *generelt* gjelder:

$$Var[aX_1 + bX_2] = a^2\,Var[X_1] + b^2\,Var[X_2] + 2\,ab\,Cov[X_1, X_2] \qquad \text{(generelt)} \tag{2.25}$$

hvor a og b er konstanter. Dersom X_1 og X_2 er uavhengige så er $Cov[X_1, X_2] = 0$, og lign.(2.25) reduserer seg til

$$Var[aX_1 + bX_2] = a^2\,Var[X_1] + b^2\,Var[X_2] \qquad \text{(gjelder kun dersom } Cov[X_1, X_2] = 0\text{)} \tag{2.26}$$

[17]Hvordan er tyngdepunktet til $P(X = x)$, dvs. $E[X]$, sammenlignet med tyngdepunktet til $P(\overline{X} = \overline{x})$, dvs. $E[\,\overline{X}\,]$? Hvordan er spredningen/usikkerheten til $P(X = x)$, dvs. $Var[X]$, sammenlignet med spredningen/usikkerheten til $P(\overline{X} = \overline{x})$, dvs. $Var[\,\overline{X}\,]$?

[18]Kun en tommelfingerregel er godt nok her.

g) Den lokale ledelsen i Bring sin avdeling i Stockholm ser på det som uakseptabelt at det gjøres mer enn 200 feilleveranser per bud i året.

Hva er sannsynligheten for at et slikt uakseptabelt nivå på feilleveringene inntreffer? Du behøver ikke å bruke heltallskorreksjon. [19]

■

[19] Tips, du skal finne:

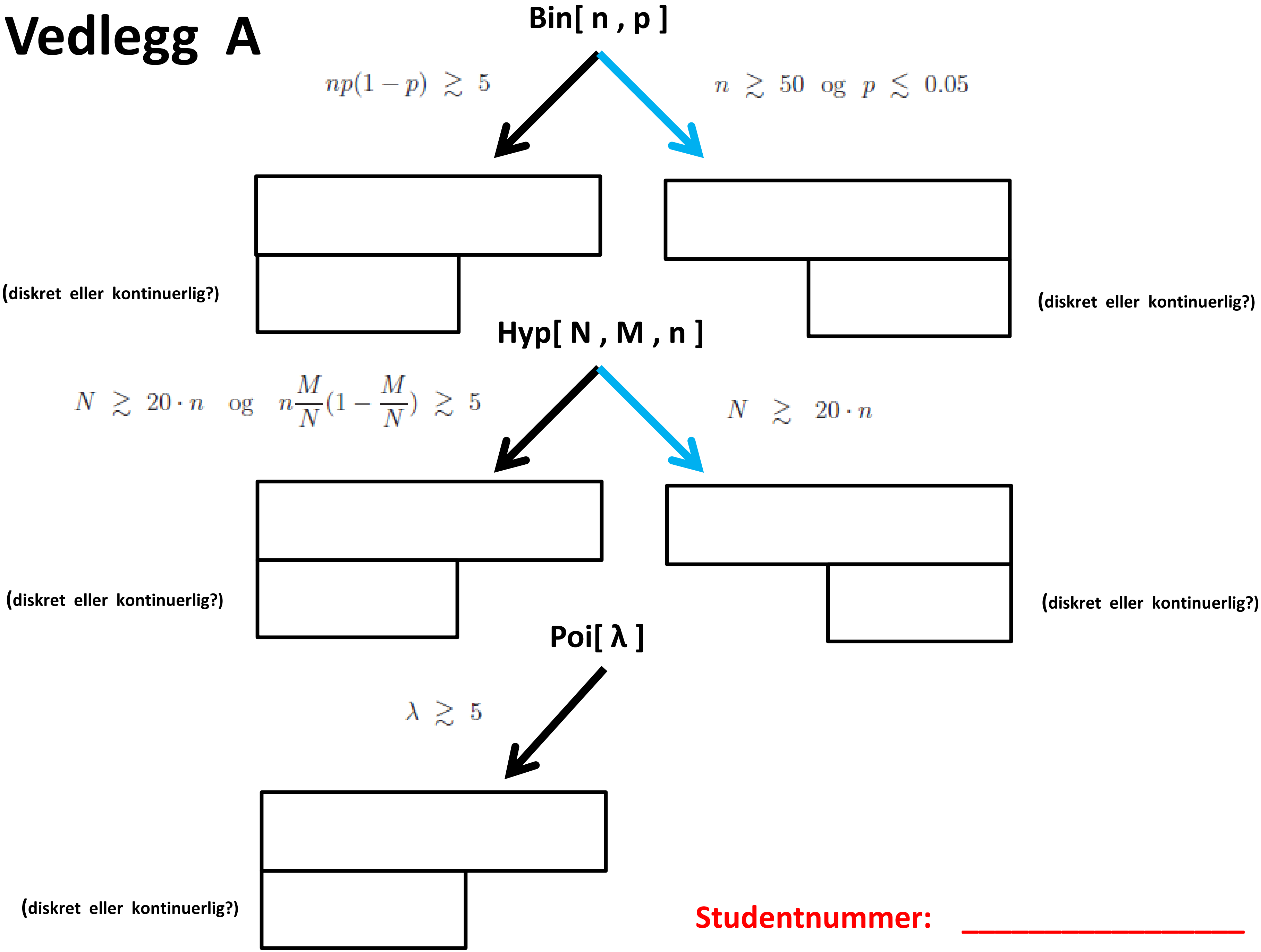
$$P\Big(X_1 + X_2 + \ ... \ + X_n > 200\Big) = P\left(\frac{X_1 + X_2 + \ ... \ + X_n}{n} > \frac{200}{n}\right) \tag{2.27}$$

$$= P\Big(\overline{X} > \frac{200}{n}\Big) \tag{2.28}$$

Deretter kan du standardisere lign.(2.28). Bruk 4 desimalers nøyaktighet.

Vedlegg A

Vedlegg A

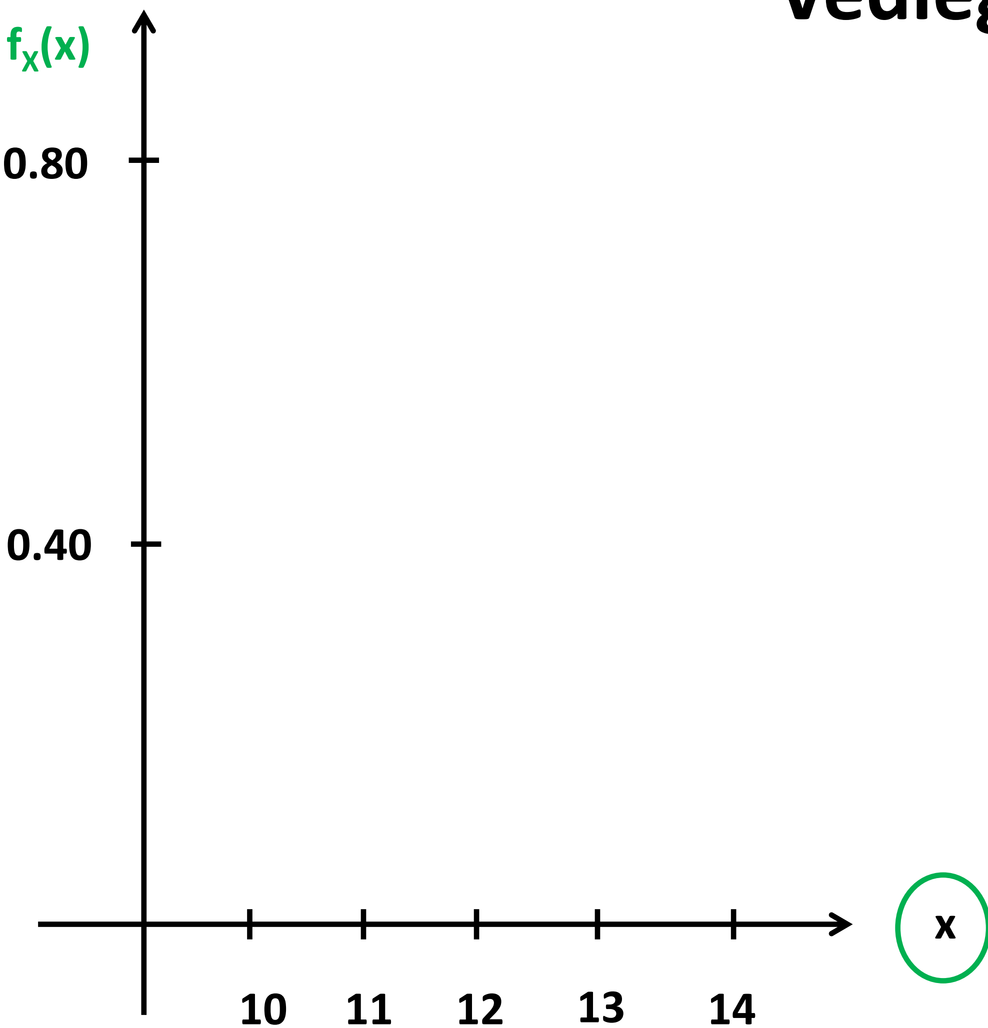


Studentnummer: ________________

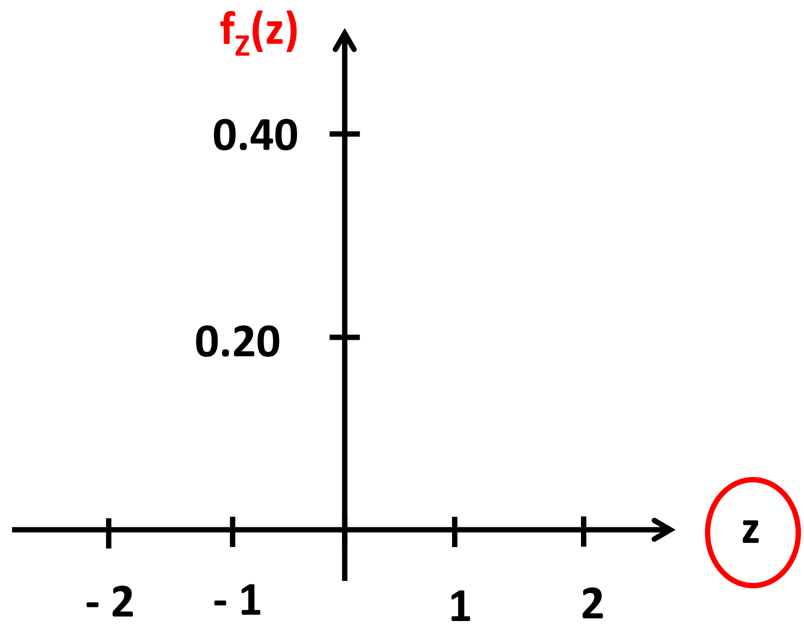
Vedlegg B

Vedlegg B

x-variabel:



z-variabel:



Studentnummer: ____________________

Høgskolen i Molde

Eksamen i

MAT110 Statistikk 1

(Molde og Kristiansund)

Eksamensdag	**:**	**Torsdag 30. mai 2013**
Tid	**:**	**09:00 – 13:00**
Faglærer/telefonnummer	**:**	**Molde:** **Per Kristian Rekdal / 924 97 051** **Kristiansund:** **Terje Bach / 932 55 838**
Hjelpemidler	**:**	**KD + formelsamling**
Antall sider inkl. forsiden	**:**	**10 + vedlegg (2 sider)**
Målform	**:**	**Norsk (bokmål)**

Noen generelle råd:

- **Kladdeark skal ikke leveres inn. Disse blir ikke sensurert.**
- **Ikke gå før tiden. Bruk alle 4 timene. Sjekk svarene dersom det er tid til overs.**
- **Det er totalt 4 oppgaver. I gjennomsnitt har du en time per oppgave.**

Oppgave 1: (sentrale formler, oversikt)

I vedlegg A (2 sider) finner du noen tabeller med ledige ruter. I noen av disse rutene skal det skrives inn formler. I andre ledige ruter skal det skrives inn kommentarer, 6 kommentarer i alt. De 6 kommentarene som skal skrives inn i de ledige rutene finner du nedenfor. Formlene som skal skrives inn finner du i formelsamlingen. [1]

Fyll ut tabellen i vedlegg A (2 sider).

$$\text{Et mål for lineær samvariasjon. Ikke normalisert.} \tag{3.1}$$

$$\text{Sterk negativ korrelasjon.} \tag{3.2}$$

$$\begin{gathered}\text{Et mål på lineær samvariasjon, korrelasjon.} \\ \text{Normalisert, ligger i intervallet} \quad -1 \leq \text{korr. koeff.} \leq 1.\end{gathered} \tag{3.3}$$

$$\text{Spredningsmål:} \quad \text{varians} \tag{3.4}$$

$$\text{Sterk positiv korrelasjon.} \tag{3.5}$$

$$\text{Lokaliseringsmål:} \quad \text{tyngdepunkt} \tag{3.6}$$

■

[1] Tips: Se side 10, 58 og 62 i formelsamlingen.

Oppgave 2: (logistikk)

Du jobber som logistikkoordinator i drillingselskapet Transocean. I forbindelse med en liten ombygging på riggen "*Searcher*" skal en supplybåt legge inntil riggen i to påfølgende dager, dag 1 og dag 2. Av sikkerhetsmessige grunner er det ikke lov å legge inntil riggen med supplybåt dersom bølgehøyden er for stor. Derfor bestemmer du deg for å definere begivenhetene:

$$B_1 = \text{begivenheten at bølgehøyden dag 1 er } \underline{\text{for høy}} \text{ for at supplybåten kan legge til} \quad (3.7)$$

$$B_2 = \text{begivenheten at bølgehøyden dag 2 er } \underline{\text{for høy}} \text{ for at supplybåten kan legge til} \quad (3.8)$$

Ved hjelp av værvarslingstjenesten storm.no har du funnet ut at:
$P(B_1) = P(B_2) = 0.05$ og $P(B_2|B_1) = 0.70$

a) Hva betyr $P(B_2|B_1)$ på "godt norsk" i vårt tilfelle? [2]

b) Er begivenhetene B_1 og B_2 uavhengige? Begrunn svaret. [3]

Figur 3.1: Riggen "*Searcher*".

[2]Dvs. gi en tolkning av dette.

[3]Bruk gjerne definisjonen av uavhengighet på side 40 i formelsamlingen for å begrunne svaret.

c) Hva er sannsynligheten for at bølgehøyden er for stor både dag 1 og dag 2? [4]

d) Hva er sannsynligheten for at bølgehøyden er for stor på dag 1 eller dag 2?

e) Regn ut $P(B_1 \cap \overline{B}_2)$.

f) Hva betyr $P(B_1 \cap \overline{B}_2)$ på "godt norsk" i vårt tilfelle?

g) Regn ut sannsynligheten for at bølgehøyden ikke er for høy dag 2 **gitt** at den ikke var for høy dag 1. [5]

■

[4]Du skal finne $P(B_1 \cap B_2)$. Hvilken setning kan være hensiktsmessig å bruke her? Se side 28 i formelsamlingen.

[5]Oppgaven går ut på å finne $P(\overline{B}_2|\overline{B}_1)$. Bruk f.eks. definisjonen av betinget sannsynlighet:

$$P(\overline{B}_2|\overline{B}_1) \;=\; \frac{P(\overline{B}_2 \cap \overline{B}_1)}{P(\overline{B}_1)} \tag{3.9}$$

Bruk deretter den ene "*tvillingsetningen*" i telleren. Hvilken setning kan være hensiktsmeesig å bruke i nevneren?

Oppgave 3: (økonomi)

Sparebanken Møre ønsker å finne ut hvor ofte privatkundene i banken gjør overtrekk på sine lønnskontoer. Anta at du er ansatt i banken og at du har fått ansvaret for å finne ut mer av dette. Du bestemmer deg for å tilnærme deg problemet ved hjelp av statistikk. I den sammenheng defineres den stokastiske variabelen:

X = antall overtrekk på en tilfeldig valgt lønnskonto per måned

Basert på historiske data finner du at sannsynlighetsfordelingen til X, dvs. $P(X = x)$, er gitt ved følgende tabell:

x	0	1	2	3	4
P(X=x)	0.57	0.13	0.18	0.10	0.02

Figur 3.2: Sannsynlighetsfordeling $P(X = x)$.

Figur 3.3: Sparebanken Møre.

a) Hva er sannsynligheten for at en tilfeldig valgt konto har minst ett overtrekk per måned?

b) i) Hva er forventet antall overtrekk for en tilfeldig valgt konto, dvs. hva er $E[X]$?

ii) Hva er variansen til antall overtrekk for en tilfeldig valgt konto, dvs. hva er $Var[X]$?

Det er totalt $n = 500$ privatkunder i Sparebanken Møre som har opprettet en lønnskonto i banken. La X_i være antall overtrekk per måned for lønnskonto nr. i, hvor $i = 1, 2, ..., n$. Anta at alle variablene X_i har samme sannsynlighetsfordeling (gitt ved tabellen i figur 3.2). Vi definerer gjennomsnittet $\overline{X}$:

$$\boxed{\overline{X} \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i \;=\; \frac{1}{n}\Big(X_1 + X_2 + \;...\; + X_n\Big) \qquad (3.10)}$$

c) i) Hva betyr $E[\,\overline{X}\,]$ på "godt norsk" i vårt tilfelle? [6]

ii) Finn $E[\,\overline{X}\,]$.

Anta at antall overtrekk for de forskjellige lønnskontoene er uavhengige.

d) i) Hva betyr $Var[\,\overline{X}\,]$ på "godt norsk" i vårt tilfelle?

ii) Finn $Var[\,\overline{X}\,]$.

e) Hvilken fordeling (tilnæremet) er det rimelig å anta at $\overline{X}$ har? Begrunn svaret.

[6]Dvs. gi en tolkning av $E[\,\overline{X}\,]$.

f) Sammenlign $E[X]$ med $E[\overline{X}]$ og $Var[X]$ med $Var[\overline{X}]$ fra oppgavene foran. Kommentèr svaret.

g) Hva er sannsynligheten for at samlet antall overtrekk per måned er større enn 400? [7]

Dersom en konto blir overtrukket blir kunden ofte belastet med en relativt høy rente. Finanstilsynet har derfor retningslinjer for hvor mange overtrekk kundene i en bank bør ha. Disse retningslinjene går ut på at det skal være 95 % sannsynlighet for at det samlede antall overtrekk i en gitt måned ikke skal overstige en øvre grense X_{grense}. Denne grensen kan variere fra bank til bank.

h) Finn X_{grense} for Sparebanken Møre. [8]

■

---

[7] Tips:

$$P\Big(X_1 + X_2 + \ ... \ + X_n > 400\Big) = P\left(\frac{X_1 + X_2 + \ ... \ + X_n}{n} > \frac{400}{n}\right) \tag{3.11}$$

$$= P\Big(\overline{X} > \frac{400}{n}\Big) \tag{3.12}$$

Deretter kan du standardisère lign.(3.12). Heltallskorreksjon behøves **ikke**.

[8] Tips:

$$P\Big(X_1 + X_2 + \ ... \ + X_n \leq X_{\text{grense}}\Big) = 0.95 \tag{3.13}$$

$$P\left(\frac{X_1 + X_2 + \ ... \ + X_n}{n} \leq \underbrace{\frac{X_{\text{grense}}}{n}}_{=\overline{X}_{\text{grense}}}\right) = 0.95 \tag{3.14}$$

$$P\Big(\overline{X} \leq \overline{X}_{\text{grense}}\Big) = 0.95 \tag{3.15}$$

Deretter kan du standardisère lign.(3.15):

$$P\Big(\overline{Z} \leq \overline{Z}_{\text{grense}}\Big) = 0.95 \tag{3.16}$$

Heltallskorreksjon behøves **ikke**. Finn så $\overline{Z}_{\text{grense}}$, deretter $\overline{X}_{\text{grense}}$ og til slutt X_{grense}.

Oppgave 4: (logistikk og økonomi)

I 2010 fikk belysningsprodusenten Glamox i Molde en stor kontrakt med det "*danske sygehusvæsen*" for leveranse av spesiallys til operasjonsstuer i alle de statlige sykehusene i Danmark. De skal ikke bare levere selve lysarmaturen, de skal også levere lysrørene til disse spesiallampene. Pga. begrenset holdbarhet til lysrørene skal det gjøres små og frekvente forsendelser til Danmark. Glamox og sykehusene i Danmark gjør en avtale om èn forsendelse i uken, hvor både lysarmatur og tilhørende lysrør er inkludert. Det sendes $\mathbf{n} = \mathbf{25}$ lysarmaturer med lysrør i hver forsendelse.

Noen av lysrørene kan være defekte. Sannsynligheten for at et tilfeldig valgt lysrør er defekt er p_d. I denne sammenheng er det hensiktsmessig å definère den stokastiske variabelen:

$$D \;=\; \underline{\text{antall}} \text{ defekte lysrør i en forsendelse}$$

Noen av lysrørene kan bli ødelagt under transport. Sannsynligheten for at dette skjer er p_t. I denne sammenheng er det hensiktsmessig å definère den stokastiske variabelen:

$$T \;=\; \underline{\text{antall}} \text{ lysrør som blir ødelagt under transport}$$

Anta at lysrør som blir defekte under produksjon er uavhengig. Anta også, for enkelhets skyld, at det er uavhengighet om lysrør blir ødelagt under transport.

Figur 3.4: Spesiallys fra Glamox.

a) Forklar hvorfor D og T er **binomisk** fordelte, dvs. forklar hvorfor
$D \sim \text{Bin}[n, p_d]$ og $T \sim \text{Bin}[n, p_t]$. [9]

Anta at $p_d = 0.05$.

b) i) Finn $E[D]$.

ii) Hva betyr $E[D]$ på "godt norsk" i vårt tilfelle? [10].

c) i) Finn $Var[D]$.

ii) Hva betyr $Var[D]$ på "godt norsk" i vårt tilfelle?

d) Hva er sannsynligheten for at **mer enn 2** lysrør er defekte i en forsendelse? [11]

Glamox ønsker å finne ut hvor mye de tjener på hver forsendelse. Avtalen med det "*danske sygehusvæsen*" er slik at de kun får betalt for lamper med lysrør som er levert i Danmark og som fungerer. I forhold til fortjenesten til Glamox må de også ta hensyn til både produksjonskonstnad k og transportkostnad k_t. Alt i alt, fortjenesten F for en gitt forsendelse er da:

$$\boxed{F = \underbrace{(n - D - T) \cdot i}_{\text{inntekt}} - \underbrace{n \cdot (k + k_t)}_{\text{utgift}} \qquad (3.17)}$$

hvor

$$\begin{aligned}
i &= \text{inntekt per lampe med lysrør som fungerer,} && (\text{"}i\text{" er bare et konstant tall}) && (3.18)\\
k &= \text{produksjonskostnad per lampe med lysrør,} && (\text{"}k\text{" er bare et konstant tall}) && (3.19)\\
k_t &= \text{transportkostnad per lampe med lysrør,} && (\text{"}k_t\text{" er bare et konstant tall}) && (3.20)
\end{aligned}$$

[9]Hvilke 4 kritèrier må være oppfylt for at en forsøksserie skal være binomisk? Svar så kort som mulig på denne oppgaven.

[10]Dvs. gi en tolkning av $E[D]$.

[11]Skriv ned formelen først. Sett inn tall til slutt. Du får en del poeng selv om du ikke finner tallene, dersom du har rett formel.

e) Vis at forventet fortjeneste for Glamox per forsendelse til Danmark er gitt ved:

$$\boxed{E[\,F\,] \;=\; n\Big[(1 - p_d - p_t)\cdot i \;-\; (k + k_t)\Big] \qquad (3.21)}$$

hvor

$$p_d \;=\; \text{sannsynligheten for at et lysrør er defekt} \qquad (3.22)$$

$$p_t \;=\; \text{sannsynligheten for at et lysrør blir ødelagt under transport} \qquad (3.23)$$

Glamox har invitert Bring og DHL til å gi anbud på leveransene til Danmark. Tilbudene til disse to budfirmaene er:

- Bring: $p_t = 0.15$ og $k_t = 275$ NOK
- DHL: $p_t = 0.04$ og $k_t = 750$ NOK

Inntekt per lampe med lysrør som fungerer er $i = 1\,700$ NOK.

f) Hvilket budfirma bør Glamox velge for å oppnå størst forventet fortjeneste? [12]

■

Figur 3.5: Bring og DHL.

[12] Ved hjelp av rett frem algebra kan man vise at lign.(3.21) kan skrives:

$$E[\,F\,] \;=\; n\cdot\Big[\,(1 - p_d)\cdot i - k \;-\; \underbrace{(\,p_t\cdot i + k_t\,)}_{\text{NB !}}\,\Big] \qquad (3.24)$$

Trenger man n, p_d og k for å kunne svare på spørsmålet?

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)** Stud.nr.: ____________

Empirisk gjennomsnitt: (formel)	Forventning: (diskret) (formel)
Kommentar:	

Empirisk varians: (formel)	Varians: (diskret) (formel)
Kommentar:	

Empirisk kovarians: (formel)	Kovarians: (formel)
Kommentar:	

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

Side 2 **(av 2)**

Stud.nr.: ____________

<u>Empirisk korrelasjonskoeffisient</u>:	<u>Korrelasjonskoeffisient</u>: (formel)
<u>Kommentar</u>:	

<u>Empirisk korrelasjonskoeffisient</u>:	<u>Korrelasjonskoeffisient</u>: (formel)
<u>Kommentar</u>:	

<u>Empirisk korrelasjonskoeffisient</u>:	<u>Korrelasjonskoeffisient</u>: (formel)
<u>Kommentar</u>:	

Avdeling for økonomi, informatikk og samfunnsfag

Høgskolen i Molde

Eksamen i

MAT110 Statistikk 1

Eksamensdag	**:**	**Mandag 6. januar 2014**
Tid	**:**	**09:00 – 13:00**
Faglærer/telefonnummer	**:**	**Per Kristian Rekdal / 924 97 051**
Hjelpemidler	**:**	**KD + formelsamling**
Antall sider inkl. forsiden	**:**	**10**
Målform	**:**	**Norsk (bokmål)**

Noen generelle råd:

- **Skriv rett inn. Ikke bruk så mye tid på kladding.**
- **Kladdark skal ikke leveres inn. De blir ikke sensurert.**
- **Det er totalt 4 oppgaver. I *gjennomsnitt* har du èn time per oppgave.**

Oppgave 1: (revisjon)

Et firma som startet opp i 2011 er sikre på at de i sitt andre driftsår, 2012, hadde mye færre feil i sine bilag enn i oppstartsåret 2011. Du er ansatt i revisjonsfirmaet PwC og skal gjøre revisjon av dette nyoppstartede firmaet. Du vurderer to alternative strategier:

Strategi A: 2000 stikkprøver i 2011 gav 12 bilag med feil,
8000 stikkprøver i 2012 gav 24 bilag med feil,
dvs. "få" stikkprøver i 2011 og "mange" stikkprøver i 2012.

Strategi B: 4000 stikkprøver i 2011 gav 20 bilag med feil,
1000 stikkprøver i 2012 gav 2 bilag med feil,
dvs. omvendt i forhold til strategi A:
"mange" stikkprøver i 2011 og "få" stikkprøver i 2012.

Anta at tallene er representative i den forstand at de viser sannsynligheten for at selskapene finner feil i de to aktuelle årene.

a) Dersom du velger strategi A, hva er sannsynligheten for å finne feil i 2011, P_{A11}? Og for 2012, P_{A12}?

b) Dersom du velger strategi B, hva er sannsynligheten for å finne feil i 2011, P_{B11}? Og for 2012, P_{B12}?

Den strategien som gir størst sannsynlighet over en gitt periode anses som best i den aktuelle perioden.

c) i) Hvilken er strategi er best for året 2011?

ii) Hvilken er strategi er best for året 2012?

Istedet for å se på ett år om gangen, som i oppgavene foran, la oss nå se på **begge årene under ett**.

d) Hvilken strategi er best når man ser begge årene under ett? [1]

e) Sammenlign svarene i oppgave **1c** og **1d**. Kommenter sammenligningen.

f) Gi en kort konklusjon/forklaring på sammenligningen i oppgave **1e**. [2]

■

Figur 4.1: Revisjon.

[1] Regn ut sannsynligheten for strategi A når du ser begge årene under ett. Gjør det samme for strategi B.

[2] Firmaet visste at det var mye flere bilag med feil i 2011 enn i 2012. Bør man da velge en strategi hvor man tar flere stikkprøver i 2011 enn 2012?

Oppgave 2: (logistikk)

En gutt jobber som avisselger. Han selger aviser for løssalg på gata. Etterspørselen av aviser en gitt dag kan beskrives av en stokastisk variabel D ("*demand*"), hvor:

$$D = \text{antall aviser som etterspørres en gitt dag} \tag{4.1}$$

Anta videre at sannsynlighetsfordelingen til denne stokastiske variabelen D er gitt ved fordelingen i figur 4.2: [3]

d_i	0	1	2	3	4	5
$P(D=d_i)$	0.10	0.05	0.15	0.30	0.25	0.15

Figur 4.2: Sannsynlighetsfordeling $P(D = d_i)$, for $i = 0, 1, \ldots 4, 5$.

a) Hva er forventet etterspørsel av aviser for en gitt dag, $E[D]$?

Hver morgen må avisgutten bestemme seg for hvor mange aviser han ønsker å prøve å selge. La oss si at avisgutten bestiller q antall ("*order quantity*") aviser fra distributøren en gitt morgen. Dersom han bestiller for få aviser så taper han salg. Dersom han bestiller for mange aviser så blir han sittende igjen med aviser som han ikke får solgt. La oss derfor introdusere en variabel S, hvor

$$S = \text{antall aviser som faktisk selges en gitt dag} \tag{4.2}$$

Denne variabelen er gitt ved

$$S = \min(D, q) \ , \tag{4.3}$$

hvor "$\min(D, q)$" betyr den minste størrelsen av D og q.

b) Forklar *kort* hvorfor $S = \min(D, q)$ også er en stokastisk variabel. [4]

[3] Ut fra denne sannsynlighetsfordelingen ser vi at etterspørselen er maksimalt 5 aviser per dag. Denne begrensningen er introdusert for å unngå for mye repeterende regning.

[4] Variabelen D er stokastisk. Variabelen S er avhengig av D, se lign.(4.2).

En morgen bestemmer avisgutten seg for å bestille $q = 3$ aviser. Anta da at sannsynlighetsfordelingen til den stokastiske variabelen S da er gitt ved fordelingen i figur 4.3:

s_i	0	1	2	3	4	5
$P(S=s_i)$	0.10	0.05	0.15	0.70	0	0

Figur 4.3: Sannsynlighetsfordeling $P(S = s_i)$, for $i = 0, 1, \ldots 4, 5$ når $q = 3$.

c) Vis at sannsynlighetsfordelingen i figur 4.3 er en gyldig sannsynlighetsfordeling.

d) i) Finn $E[S]$.

ii) Gi en tolkning av $E[S]$, dvs. forklar *kort* hva det betyr på "godt norsk".

e) Sammenlign $E[D]$ fra oppgavene **2a** og $E[S]$ fra **2d**.
Er det rimelig at den ene verdien er større enn den andre? Begrunn svaret.

Ved bestilling kjøpes avisene inn for prisen w ("*wholesale*") per avis. Avisgutten selger dem videre på gata for utslagsprisen r ("*revenue*") per avis. Fortjenesten blir da:

$$\boxed{\pi(q) = r\,S - w\,q\ , \qquad (4.4)}$$

hvor, som tidligere angitt, $S = \min(D, q)$. Her er r, q og w konstanter.

f) Anta at innkjøpspris er $w = 5$ NOK og utslagspris er $r = 20$ NOK.
Hva er da forventet fortjeneste $E[\pi(q)]$ dersom avisgutten bestiller $q = 3$ aviser? [5]

[5] Tips: Bruk regneregelen: (a og b er konstanter)

$$E[aX + bY] = aE[X] + bE[Y]\ . \qquad (4.5)$$

Bruk gjerne resultatet fra oppgave **2d**.

Ovenfor har vi behandlet D som en diskret stokastisk variabel. Dersom vi nå istedet behandler D som en **kontinuerlig** variabel så kan man vise at maksimal fortjeneste oppnås når [6]

$$\boxed{P(D \leq q^*) = 1 - \frac{w}{r} \ , \qquad (4.6)}$$

hvor q^* er det antall aviser som avisgutten må kjøpe inn om morgenen for å maksimere sin fortjeneste. Sannsynligheten $P(D \leq q^*)$ i lign.(4.6) er altså den kumulative fordelingen til D. Anta videre at D er **normalfordelt** med forventning $\mu = 3$ og standardavvik $\sigma = 1.5$, dvs.

$$D \ \sim \ \mathrm{N}\big[\,\mu = 3, \sigma = 1.5\,\big] \ . \qquad (4.7)$$

g) Med verdiene $w = 5$ NOK og $r = 20$ NOK, finn det antall aviser q^* som avisgutten må bestille for å få størst fortjeneste.

h) Med fordelingen som i lign.(4.7) er forventet etterspørsel av aviser en gitt dag like tre, $\mu = 3$. Denne forventningsverdien $\mu = 3$ og verdien på q^* fra oppgave **2g** er ikke er sammenfallende. Gi en *kort* forklaring på hvorfor den ene verdien er større enn den andre.

■

Figur 4.4: Avisgutt.

[6] Du skal ikke vise lign.(4.6). Ta den for gitt.

Oppgave 3: (økonomi)

Økonomisjefen ved Tusten Skiheiser er svært opptatt av været. Dette fordi hun ønsker snø slik at de kan få flest mulig driftsdager. I den sammenheng defineres begivenhetene

$$S_i \;=\; \text{det snør dag nr. i}$$

hvor $i = 1, 2, 3$ er tre påfølgende dager i november. Anta at sannsynligheten for at det snør en tilfeldig dag i november er $5/30 = 1/6$, dvs.

$$P(S_1) \;=\; P(S_2) \;=\; P(S_3) \;=\; \frac{1}{6} \tag{4.8}$$

Anta videre at været de ulike dagene er uavhengige av hverandre.

a) Finn sannsynligheten for at det snør dag nr. 1 **og** dag nr. 2.

b) Finn sannsynligheten for at det snør dag nr. 1 **eller** dag nr. 2.

Antagelsen om at været de ulike dagene er uavhengig av hverandre er en sterk forenkling av virkeligheten. La oss derfor anta at været fra en dag til en annen **er** avhengig av hverandre på en slik måte at $P(S_1 \cap S_2) = 2/30$.

c) Finn sannsynligheten for at det snør dag nr. 2 gitt at det snødde dag nr. 1. [7]

d) Beskriv med ord hva begivenheten $S_1 \cap S_2 \cap S_3$ betyr.

[7] Dvs. finn $P(S_2|S_1)$.

Anta at $P(S_3|S_2 \cap S_1) = 0.6$.

e) Finn $P(S_1 \cap S_2 \cap S_3)$. [8]

■

Figur 4.5: Tusten.

[8]Bruk gjerne **multiplikasjonssetningen**. Se formelsamlingen. Denne setningen kan brukes gjentatte ganger. Bruk også den oppgitte verdien for $P(S_3|S_2 \cap S_1)$ øverst på denne siden.

Oppgave 4: (økonomi)

De ansatte på eksamenskontoret ved Høgskolen i Molde er interessert i å vite mer om studiepoengproduksjon. I den sammenheng defineres følgende stokastiske variabel:

$$X \;=\; \text{antall studiepoeng (sp) bestått i løpet av første studieår}$$

Basert på erfaring fra de 10 siste årene har de funnet at sannsynlighetsfordelingen $P(X = x)$ er som gitt ved tabellen i figur 4.6:

x	0	5	10	15	20	25	30
P(X=x)	0.13	0.08	0.05	0.05	0.03	0.03	0.04

x	35	40	45	50	55	60
P(X=x)	0.04	0.06	0.06	0.10	0.10	0.23

Figur 4.6: Sannsynlighetsfordeling $P(X = x)$.

Denne sannsynlighetsfordelingen har forventning

$$E[X] \;=\; 35 \tag{4.9}$$

Denne størrelsen skal ikke vises. Den kan tas for gitt.

a) Hva er sannsynligheten for at en tilfeldig valgt førsteårsstudent består 40 sp eller mer, dvs. hva er $P(X \geq 40)$?

b) Av 250 førsteårsstudenter, omtrent hvor mange vil bestå 40 sp eller mer? [9]

c) Finn $P(25 \leq X \leq 35)$.

d) Gi en tolkning av oppgave **4c**, dvs. forklar *kort* hva $P(25 \leq X \leq 30)$ betyr på "godt norsk".

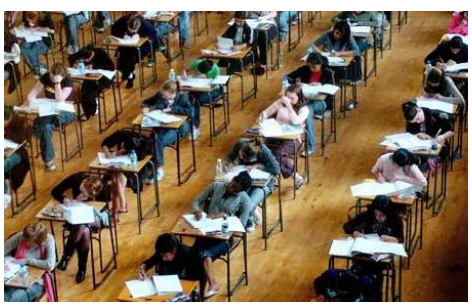
Anta at det er $n = 250$ førsteårsstudenter. La X_i være antall studiepoeng bestått av student nr. i, hvor $i = 1, 2, 3 \ldots n$. Anta videre at alle X_i har **samme fordeling**. La oss videre definere den stokastiske variabelen

$$\boxed{\overline{X} \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i \qquad (4.10)}$$

e) Finn $E[\,\overline{X}\,]$.

f) Gi en tolkning av $E[\,\overline{X}\,]$ for vår situasjon, dvs. forklar *kort* hva det betyr på "godt norsk".

■



Figur 4.7: Eksamen.

[9]Bruk svaret fra oppgave **4a**.

Avdeling for økonomi, informatikk og samfunnsfag

Høgskolen i Molde

Eksamen i

MAT110 Statistikk 1

(Molde og Kristiansund)

Eksamensdag	**:**	**Fredag 9. mai 2014**
Tid	**:**	**09:00 – 13:00**
Faglærer/telefonnummer	**:**	**Molde:** **Per Kristian Rekdal / 924 97 051** **Kristiansund:** **Terje Bach / 932 55 838**
Hjelpemidler	**:**	**KD + formelsamling**
Antall sider inkl. forsiden	**:**	**12**
Målform	**:**	**Norsk (bokmål)**

<u>Noen generelle råd:</u>

- **Kladdeark skal ikke leveres inn. Disse blir ikke sensurert.**
- **Ikke gå før tiden. Bruk alle 4 timene. Sjekk svarene dersom det er tid til overs.**
- **Det er totalt 4 oppgaver. I *gjennomsnitt* har du en time per oppgave.**

Oppgave 1: (revisjon)

KPMG i Møre & Romsdal har utviklet et dataverktøy for å forutsi hvilke bedrifter som vil gå konkurs. Selv om dette verktøyet er konservativt så gir det 5 % falske indikasjoner for bedrifter som ikke går konkurs, dvs. 5 % av de bedriftene som ikke går konkurs blir likevel klassifisert som $\underbrace{\text{konkursobjekter}}_{\text{objekt}}$:

$$P(\text{objekt}|\overline{K}) \;=\; 0.05 \tag{5.1}$$

a) Hva er sannsynligheten for at en bedrift som $\overbrace{\text{ikke går konkurs}}^{\overline{\text{K}}}$ ei heller blir kategorisert som et konkursobjekt? [1]

Det viser seg også at av de bedriftene som faktisk gikk $\overbrace{\text{konkurs}}^{\text{K}}$ ble 80 % klassifisert som konkurs-objekter med KPMG sitt dataverktøy, dvs.:

$$P(\text{objekt}|K) \;=\; 0.80 \tag{5.2}$$

Anta videere at 0.1 % av bedriftene i Møre & Romsdal går konkurs, dvs.:

$$P(K) \;=\; 0.001 \tag{5.3}$$

Figur 5.1: KPMG.

[1] Dvs. finn $P(\overline{\text{objekt}}|\overline{K})$. Bruk komplementsetningen.

b) Hvor stor sannsynlighet er det for at KPMG-verktøyet klassifiserer en bedrift som $\underbrace{\text{konkursobjekt}}_{\text{objekt}}$? [2]

c) En bedrift i Møre & Romsdal blir klassifisert som $\overbrace{\text{konkursobjekt}}^{\text{objekt}}$ ifølge KPMG-verktøyet. Bruk Bayes formel og vis at sannsynligheten for at denne bedriften $\underbrace{\textit{ikke}\text{ går konkurs}}_{\overline{\text{K}}}$ er: (med 4 desimaler nøyaktighet)

$$P(\overline{K}|\text{objekt}) \;=\; 0.9842 \tag{5.4}$$

d) Kommenter svaret i oppgave **c**:

Hva betyr $P(\overline{K}|\text{objekt})$ i lign.(5.4)?
I lys av dette, bør revisjonsselskapet KPMG trekke konklusjoner kun basert på dette verktøyet?

■

[2] Dvs. finn $P(\text{objekt})$. Tips: Bruk formelen for oppsplitting av utfallsrommet Ω.

Oppgave 2: (logistikk)

Transportselskapet *Fjord1* har sett nærmere på antall personbiler som ankommer Molde fergekai. *Fjord1* har funnet ut at det i gjennomsnitt kommer 920 personbiler alle hverdager i tidsrommet 10:00 - 14:00, altså i et tidsrom på 4 timer. Anta, for enkelhets skyld, at det kun er personbiler som ankommer fergekaien i dette tidsrommet, og at raten per tid er konstant. Anta også at alle bilene som ankommer fergekaien er uavhengige av hverandre.

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

$$X = \text{antall personbiler som ankommer fergekaien i en } \underline{\text{30 minutters periode}}, \text{ dvs. mellom to ferger} \tag{5.5}$$

a) Hva slags diskret sannsynlighetsfordeling har den stokastiske variabelen X?

b) Vis, ved en *kort* utregning, at raten λ som beskriver antall biler **per halvtime** er

$$\lambda = 115 \tag{5.6}$$

c) i) Hva er forventet antall biler $E[X]$ som ankommer fergekaien mellom to ferger?

ii) Hva er standardavviket $\sigma[X]$ av antall biler som ankommer fergekaien mellom to ferger?

Figur 5.2: Ferge.

d) i) En diskret Poisson-sannsynlighetsfordeling Poi[λ] kan, under bestemte betingelser, med god tilnærmelse beskrives av en kontinuerlig sannsynlighetsfordeling. Hvilken sannsynlighetsfordeling?

ii) Hva slags betingelse må raten λ oppfylle for at tilnærmelsen som beskrevet ovenfor skal gjelde? [3]

iii) Er denne betingelsen på λ fra oppgave ii oppfylt for vårt tilfelle?

Fergene som går strekningen Molde - Vestnes har alle en kapasitet på $X_0 = 125$ personbiler.

e) Anta at det i utgangspunktet er 20 biler i fergekø i tillegg til de bilene som ankommer fergekaien med en konstant rate λ.
Hva er da sannsynligheten for at ikke alle bilene får plass i fergen? [4]

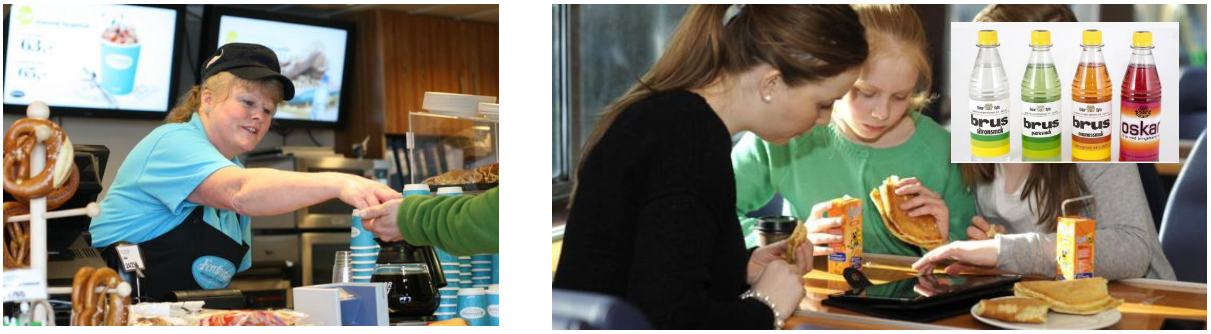
f) Anta nå at det i utgangspunktet *ikke* er noen biler i fergekø.
Dersom *Fjord 1* ønsker at det skal være 95 % sannsynlighet for at alle bilene kommer med, hvor stor kapasitet X_{kap} må fergene ha da? [5]

■

[3]Se formelsamlingen.
[4]Dvs. finn $P(X > X_0 - 20)$. Benytt resultatet fra oppgave **2d** i.
[5]Finn X_{kap} i ligningen $P(X \leq X_{\text{kap}}) = 0.95$.

Oppgave 3: (økonomi)

I tillegg til fergebilletter er også kiosksalg en viktig inntektskilde for *Fjord1*. I den sammenheng skal vi se nærmere på salg av brus. Vi skal bruke to forskjellige statistiske tilnærmelser, modell 1 og modell 2.



Figur 5.3: Kiosk.

Modell 1:

For modell 1 defineres følgende stokastiske variabel:

$$Y \;=\; \text{antall passasjerer som kjøper } \textbf{èn} \text{ brus på en gitt overfart}$$

Basert på historiske data har *Fjord1* funnet ut at det er 11 % av fergepassasjerene som kjøper en brus i løpet av en gitt overfart. Anta at fergepassasjerene kjøper brus uavhengige av hverandre. Anta også at det i gjennomsnitt er 3 passasjerer i hver bil, at det i utgangspunktet ikke er noen biler i fergekø og at alle bilene som ønsker å bli med fergen får plass. Selv om antall ankomne biler egentlig er en stokastisk størrelse så antas det i denne oppgaven at det faktisk kommer $\lambda = 115$ antall biler til fergekaien mellom to fergeravganger.
(Parameteren λ er den samme som i oppgave 2).

a) Begrunn hvorfor det er rimelig å anta at Y er binomisk fordelt,
dvs. begrunn hvorfor [6]

$$Y \;\sim\; \text{Bin}[\, n = 3\lambda, p = 0.11 \,] \tag{5.7}$$

Kommenter *kort* hvorfor $n = 3\lambda$.

[6] Hvilke 4 forutsetninger må være oppfylt for at Y skal være binomisk fordelt? Er disse forutsetningene oppfylt i vårt tilfelle? (Se formelsamling).

b) Finn **forventet** antall solgte brus på en gitt overfart ifølge modell 1,
dvs. finn $E[Y]$.

c) i) Finn **variansen** til antall solgte brus for en overfart ifølge modell 1,
dvs. finn $Var[Y]$.

ii) Hva er det tilhørende **standardavviket** $\sigma[Y]$?

d) i) Dersom en betingelse er oppfylt så kan en **bin**omisk fordeling tilnærmes med en **normal**fordeling. Hvilken betingelse er det? [7]

ii) Er denne betingelsen oppfylt i vårt tilfelle?

e) Hva er sannsynligheten for at det selges mer enn 45 brus på en overfart ifølge modell 1? [8]

[7] Se f.eks. side 60 i formelsamlingen.

[8] Dvs. finn $P(Y > 45)$.

Modell 2:

I modell 1 beskrevet ovenfor kjøper kunden enten èn brus eller ingen brus. Dette er en forenkling av virkeligheten siden en kunde kan kjøpe mer enn en brus på en gitt overfart. Modellen som vi nå skal introdusere, modell 2, tar høyde for at fergepassasjerer kan kjøpe mer enn en brus. For denne modellen defineres derfor følgende stokastiske variabel:

$$B \;=\; \text{antall brus som kjøpes } \textit{per passasjer} \text{ på en gitt overfart}$$

Sannsynlighetsfordelingen $P(B = b_i)$, hvor $i = 0, 1, 2, 3$, er gitt ved:

b_i	0	1	2	3
$P(B=b_i)$	0.93	0.04	0.02	0.01

Figur 5.4: Sannsynlighetsfordeling $P(B = b_i)$.

Tilhørende forventning og standardavvik er:

$$E[B] = 0.11 \quad , \quad \sigma[B] = 0.4449 \tag{5.8}$$

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

f) Hva er sannsynligheten at en tilfeldig valgt passasjer kjøper mer enn en brus? [9]

Det *totale* antallet brus som selges på en gitt overfart er:

$$\boxed{B_{\text{tot}} \;=\; \sum_{i=1}^{n} B_i \;=\; B_1 + B_2 + \;...\; + B_n \qquad (5.9)}$$

hvor $n = 3\lambda$ er antall passasjerer på en gitt overfart.

[9] Dvs. finn $P(B > 1)$.

g) Finn $E[B_{\text{tot}}]$.

Anta at passasjerene kjøper brus **uavhengige** av hverandre.

h) Finn $Var[B_{\text{tot}}]$. [10]

i) Hvilken **fordeling** er det rimelig å anta at B_{tot} har? Begrunn svaret. [11]

j) Hva er sannsynligheten, ifølge modell 2, for at det selges mer enn 45 brus på en overfart? [12]

k) Sammenlign og kommenter svarene fra oppgave **3c** i og **3h**,
dvs. sammenlign og kommenter $Var[Y]$ og $Var[B_{\text{tot}}]$. [13]

Anta at prisen på brus er $p = 20$ NOK. Innkjøpskostnaden er $k = 6$ NOK. Anta videre at *Fjord1* sine faste kostnader forbundet med kioskdriften som er relatert til brussalget er $k_{\text{fast}} = 175$ NOK. Fortjenesten F på brussalget for en gitt overfart er da gitt ved:

$$\boxed{F = (p - k)B_{\text{tot}} - k_{\text{fast}} \qquad (5.11)}$$

l) Finn forventet fortjeneste $E[F]$ på brussalget til *Fjord1* for en overfart.

■

[10] Tips: Se side 44 i formelsamlingen.

[11] Hint: Sentralgrenseteoremet. Det er særlig 3 ting du bør nevne for å begrunne svaret ditt.

[12] Dvs. finn $P(B_{\text{tot}} > 45)$. Bruk:

$$P(B_{\text{tot}} > 45) \quad = \quad 1 \; - \; P(B_{\text{tot}} \leq 45) \tag{5.10}$$

Deretter kan du standardisere lign.(5.10). Heltallskorreksjon behøves **ikke**.

[13] Er det rimelig at den ene variansen er større enn den andre?

Oppgave 4: (økonomi)

I fjor opplevde flyselskapet Norwegian at aksjekursen økte i en periode. La oss se på 5 omsetningsdager i denne perioden.

$$x = \text{omsetningsdag nr. } x, \text{ hvor } x = 1, 2, 3, 4, 5 \quad (5.12)$$
$$y = \text{aksjekurs (i NOK)} \quad (5.13)$$

Aksjekurs, dvs. aksjepris, for omsetningsdag nr. x ($x = 1, 2, 3, 4, 5$) er gitt ved følgende tabell:

x (dag nummer)	**1**	**2**	**3**	**4**	**5**
y (aksjekurs i NOK)	**268**	**279**	**285**	**289**	**299**

Figur 5.5: Omsetningsdag nr. x og aksjekurs (pris) y.

Ut fra tabellen i figur 5.5 kan man regne ut den empiriske variansen for aksjekursen y og den empiriske variansen for x. Resultatet er:

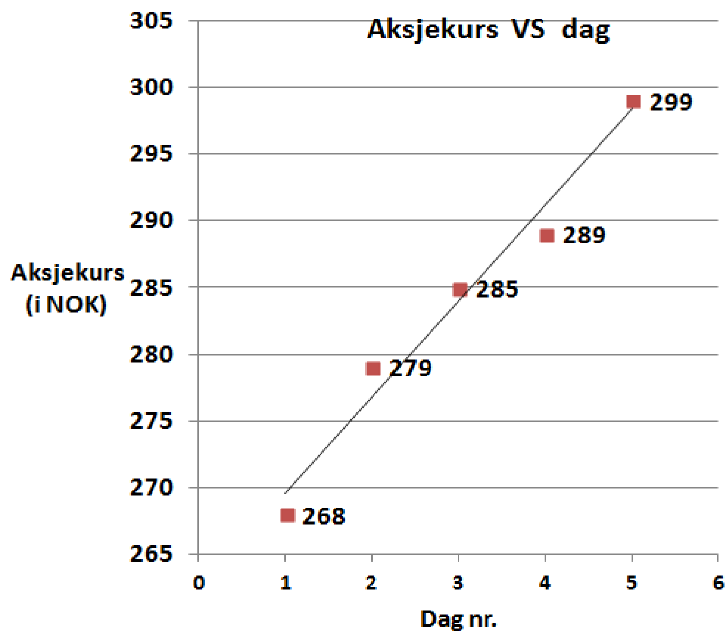
$$S_y^2 = 133 \text{ NOK}^2 \quad , \quad S_x^2 = 2.5 \quad (5.14)$$

Den empiriske **ko**variansen mellom x og y er:

$$S_{xy} = 18 \text{ NOK} \quad (5.15)$$

(Størrelsene i lign.(5.14) og (5.15) trenger du å ikke regne ut. Bare ta dem for gitt).

Figur 5.6: Norwegian.



Figur 5.7: Plott av dataene fra tabellen i figur 5.5.

a) Grafen i figur 5.7 viser minste kvadraters regresjonslinje for x og y.
Finn et analytisk uttrykk for denne **regresjonslinjen**. [14]

b) i) Forklar *kort*, på generelt grunnlag, hva en regresjonslinje mellom variablene x og y beskriver.

ii) På generelt grunnlag, hvordan tolker du parameteren $\hat{\beta}$ i en *lineær* regresjonslinje? [15]

iii) Spesielt, hvordan tolker du parameteren $\hat{\beta}$ i regresjonslinjen fra oppgave **4a**?

c) Anta trenden på aksjemarkedet for Norwegian fortsetter på samme måte som de 5 omsetningsdagene i en periode fremover.
Dersom Norwegian ønsker å *estimere* hvor mye aksjekursen blir for dag nummer 12 så kan de bruke modellen fra oppgave **4a**, altså regresjonslinjen.

Hvor mye predikerer regresjonslinjen at aksjekursen vil være for dag nummer 12?

[14] Dvs. finn formelen for linjen. Se gjerne formelsamlingen for formel for minste kvadraters regresjonslinje.
[15] Tips: stigningstall.

d) Forklaringskraften R^2 blir ofte brukt i forbindelse med lineær regresjonsanalyse.

i) Hva slags mulige verdier har R^2?

ii) Hva slags benevning/enhet har R^2?

iii) Hva er R^2 et mål på?

For å finne forklaringskraften R^2 kan man bruke et dataprogram, f.eks. Excel, i stedet for å regne ut R^2 "for hånd" via definisjonen.

e) Finn forklaringskraften uten å gjøre noe regning "for hånd".
Bare les av fra Excel-utskriften i figur 5.8 nedenfor.
(Bruk 4 desimalers nøyaktighet).

f) Kommenter svaret i oppgave **4e**. [16]

A	B	C	D	E	F	G	H
SUMMARY OUTPUT							
Regression Statistics							
Multiple R	0,987135295						
R Square	0,97443609						
Adjusted R Square	-1,666666667						
Standard Error	2,12916259						
Observations	1						
ANOVA							
	df	*SS*	*MS*	*F*	*Significance F*		
Regression	5	518,4	103,68	114,3529	#NUM!		
Residual	3	13,6	4,533333				
Total	8	532					
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*
X Variable 4	262,4	2,233084563	117,5056	1,36E-06	255,2933283	269,5066717	255,2933283
X Variable 5	7,2	0,673300329	10,69359	0,001748	5,057257855	9,342742145	5,057257855

Figur 5.8: Utskrift fra Excel.

[16] For en gitt dag, vi du si at regresjonslinjen predikerer aksjekursen i stor eller liten grad? Med stort eller lite presisjonsnivå?

Høgskolen i Molde

Eksamen i

MAT110 Statistikk 1

Eksamensdag	**:**	**Torsdag 8. januar 2015**
Tid	**:**	**09:00 – 13:00 (4 timer)**
Faglærer/telefonnummer	**:**	**Molde:** **Per Kristian Rekdal / 924 97 051** **Kristiansund:** **Terje Bach / 932 55 838**
Hjelpemidler	**:**	**KD + formelsamling**
Antall sider inkl. forsiden	**:**	**12**
Målform	**:**	**Norsk (bokmål)**

Noen generelle råd:

- **Skriv rett inn. Ikke bruk så mye tid på kladding.**
- **Kladdark skal ikke leveres inn. De blir ikke sensurert.**
- **Det er totalt 4 oppgaver. Dvs. i *gjennomsnitt* èn time per oppgave.**

Oppgave 1: (økonomi)

Sparebanken Møre i Kristiansund ønsker å gjøre en kredittvurdering av kundemassen. Erfaringstall har vist at banken har 75 % kunder med god betalingsevne og 25 % kunder med dårlig betalingsevne.

a) Bruk notasjonen

$$\begin{aligned} \text{G} &= \text{kunde har } \textbf{god} \text{ betalingsevne} && (6.1) \\ \text{D} &= \text{kunde har } \textbf{dårlig} \text{ betalingsevne} && (6.2) \end{aligned}$$

til å formulere de **to** opplysningene i teksten ovenfor på en matematisk måte via to sannsynligheter $P(\cdot)$.

I en undersøkelse av kundene med god betalingsevne viste det seg at 10 % av disse har lav inntekt. Videre viste det seg at blant kundene med dårlig betalingsevne hadde 50 % lav inntekt.

b) Bruk notasjonen

$$\text{L} = \text{kunde med } \textbf{lav inntekt} \qquad (6.3)$$

til å formulere de **to** opplysningene i teksten ovenfor på en matematisk måte via to betingede sannsynligheter $P(\cdot|\cdot)$.

Figur 6.1: Sparebanken Møre i Kristiansund.

c) Hvor stor del av kundemassen er kunder med lav inntekt? Dvs. finn $P(L)$.

d) Hva er sannsynligheten for at en kunde med lav inntekt har dårlige betalingsevne? Dvs. finn den betingede sannsynligheten $P(D|L)$.

e) Hva er sannsynligheten for at en kunde med lav inntekt har god betalingsevne? Dvs. finn den betingede sannsynligheten $P(G|L)$.

Banken $\overbrace{\text{tjener 6000 NOK}}^{\textbf{god}\text{ betalingsevne}}$ på å gi lån til kunder med god betalingsevne. Og $\overbrace{\text{taper 4000 NOK}}^{\textbf{dårlig}\text{ betalingsevne}}$ på å gi lån til kunder med dårlig betalingsevne. La

$$\text{F} = \text{fortjeneste på en kunde med } \textbf{lav inntekt} \tag{6.4}$$

f) Finn bankens forventede fortjeneste $E[F]$ for en kunde med lav inntekt.

g) Er det lønnsomt for Sparebanken Møre å gi lån til kunder med lav inntekt? Begrunn svaret ut fra svaret i oppgave **1f**.

■

Oppgave 2: (petroleumslogistikk)

Anta at du jobber som logistikk-koordinator ved olje- og gass firmaet *British Petroleum* (BP). En del av jobben din er å få oversikt over nestenulykker og andre uønskede "*hendelser*" som skjer på plattformer som BP opererer i.

Du ønsker å se nærmere på antall "hendelser" ved hjelp av statistikk. Siden "hendelser" skjer relativt sjelden og siden det er hensiktsmessig å se på antall "hendelser" per tidsenhet så foreslår du å bruke "*loven om sjeldne begivenheter*", dvs. **Poisson**fordelingen. Du definerer den stokastiske variabelen:

$$X \;=\; \text{antall ``hendelser'' } \underline{\text{per uke}}$$

som vi altså antar er Poisson fordelt:

$$X \quad \sim \quad \text{Poi}[\lambda] \tag{6.5}$$

Basert på erfaring vet man at det er ca. 0.6 "hendelser" per uke på BP-plattformen Valhall i Nordsjøen, dvs.:

$$\lambda \;=\; 0.6 \tag{6.6}$$

Figur 6.2: BP-plattformen Valhall i Nordsjøen.

a) Hva er sannsynligheten for at det skjer 1 "hendelse" i løpet av en uke, dvs. $P(X = 1)$? [1]

b) Hva er sannsynligheten for at det skjer **mer enn** 1 "hendelse" i løpet av en uke, dvs. $P(X > 1)$?

La oss nå istedet se på antall "hendelser" per måned, altså 4 uker. I den sammenheng defineres den stokastiske variabelen Y:

$$\boxed{Y \;=\; X_1 \;+\; X_2 \;+\; X_3 \;+\; X_4 \qquad\qquad (6.7)}$$

dvs. Y = antall "hendelser" per måned.

c) Hva er forventet antall "hendelser" per måned, dvs. $E[Y]$?

d) Man kan vise at summen av Poisson fordelinger også er Poisson fordelt.
(Du skal ikke vise dette. Bare ta dette for gitt.)
Det betyr at siden X_1, X_2, X_3 og X_4 er Poisson fordelt, så er også summen av dem, $Y = X_1 + X_2 + X_3 + X_4$, Poisson fordelt med forventning $E[Y]$: [2]

$$Y \sim \text{Poi}\big[\, E[Y] \,\big] \tag{6.8}$$

Hva er sannsynligheten for at det skjer **mer enn** 1 "hendelse" i løpet av en måned, $P(Y > 1)$?

e) Sammenlign svaret i oppgave **2d** med svaret i oppgave **2b**,
dvs. sammenlign sannsynlighetene $P(X > 1)$ og $P(Y > 1)$.
Er det rimelig at det ene svaret er større enn det andre?

■

[1] Bruk 4 desimalers nøyaktighet.

[2] Dette gjelder kun for Poisson fordelinger. Det gjelder ikke generelt. Det er ikke slik at f.eks. summen av binomial fordelinger er en ny binomial fordeling.

Oppgave 3: (økonomi og logistikk)

Oskar Sylte er en mineralvannprodusent i Molde. Deres produksjonsanlegg kan i noen tilfeller gi produksjonsfeil på brusen som produseres. Basert på erfaring viser det seg at det er 95 % sannsynlighet for at en brus ikke har produksjonsfeil. Det er 3 % sannsynlighet for at en brus har èn produksjonsfeil. Og det er 2 % sannsynlighet for at en brus har to produksjonsfeil. Den stokastiske variabelen [3]

$$X \quad = \quad \text{antall produksjonsfeil for en tilfeldig valgt brus}$$

har dermed fordelingenen $P(X)$ som vist i tabellen nedenfor i figur 6.3:

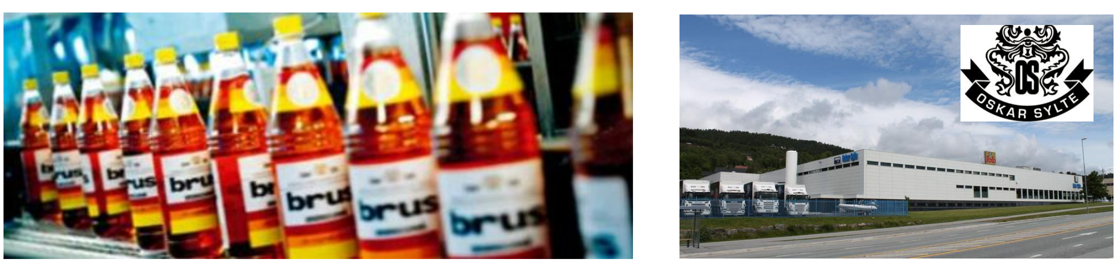
x	0	1	2
P(X=x)	0.95	0.03	0.02

Figur 6.3: Sannsynlighetsfordeling $P(X = x)$.

a) Vis at sannsynlighetsfordelingen i tabellen i figur 6.3 er en **gyldig** sannsynlighets-fordeling.

b) Finn $E[X]$ og gi en tolkning av $E[X]$. [4]

c) Finn $Var[X]$ og gi en tolkning av $Var[X]$.



Figur 6.4: Produksjon av ananasbrus.

[3]Prouksjonsfeil kan være at toppen ikke er rett påskrudd og etiketten ikke er rett påsatt.
[4]Dvs. hva betyr $E[X]$ på "godt norsk" i vårt tilfelle?

La oss se på en dagsproduksjon av ananasbrus. Sylte produserer $n = 4500$ ananasbrus per dag. For hver produksjonsfeil som inntreffer så taper fabrikken $b = 23$ NOK. Den stokastiske variabelen I som beskriver inntekten per dag av ananasbrus er da gitt ved:

$$\boxed{I \quad = \quad p\,n\,0.95 \;-\; b\,n\,X \qquad (6.9)}$$

hvor utslagsprisen på brusen er $p = 15$ NOK.

d) Hva er forventet inntekt per dag på ananasbrus?

La $X_1, X_2, X_3,\ \dots\ , X_n$ være antall produksjonsfeil for ananasbrus nr. 1, 2, 3, ... , n. Det totale antall produksjonsfeil per dag for ananasbrus er da:

$$\boxed{Y \;=\; X_1 + X_2 + X_3 + \ \dots\ + X_n \qquad\qquad (6.10)}$$

Anta videre at:

1. antall produksjonsfeil for de ulike ananasbrusene er uavhengige:

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

2. alle ananasbrus antas å ha samme sannsynlighetsfordeling (gitt ved tabell i figur 6.3):

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

e) i) Hva er forventet antall produksjonsfeil av ananasbrus per dag? [5]

ii) Hva er variansen til antall produksjonsfeil av ananasbrus per dag? [6]

---

[5] Dvs. finn $E[Y]$.
[6] Dvs. finn $Var[Y]$.

f) i) Med forutsetningene som på forrige side, hvilken setning gjelder da?

ii) Hvilken sannsynlighetsfordeling har da den stokastiske variabelen Y?

iii) Hvor stor må n ($n =$ antall "forsøk") være, **omtrent**, for at setningen fra oppgave i) skal gjelde? [7]

g) Ledelsen i brusfabrikken synes det er uakseptabelt at det er mer enn 350 produksjonsfeil per dag av ananasbrus.

Hva er sannsynligheten for at et slikt uakseptabelt nivå inntreffer, dvs. hva er $P(Y > 350)$? [8]

■

[7] Kun en tommelfingerregel er godt nok her.
[8] Du behøver ikke å bruke heltallskorreksjon.

Oppgave 4: (logistikk)

Denne oppgaven handler om **regresjonsanalyse**. Først noen generelle spørsmål. Deretter skal vi anvende teorien på et konkret eksempel innen logistikk.

a) Forklar *kort* hva regresjonsanalyse er. [9]

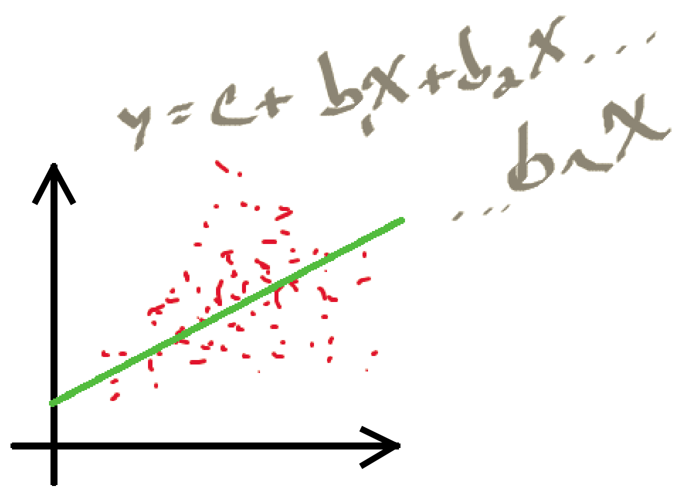
b) La $(x_1, y_2), (x_2, y_2), ..., (x_n, y_n)$ være observasjonspar/datasett.
Formuler setningen for minste kvadraters lineære regresjonslinje slik vi har gjort det i forelesningene og kompendiet. [10]

c) Forklaringskraften R^2 blir ofte brukt i forbindelse med **lineær** regresjonsanalyse.

i) Hva slags mulige verdier har R^2?

ii) Hva slags benevning/enhet har R^2?

iii) Hva er R^2 et mål på?



Figur 6.5: Regresjon.

[9] En setning er nok.
[10] Tips: se formelsamling.

Hustadmarmor AS er en bedrift som leverer "*slurry*" til industrien. Slurry er en hvit flytende masse som blant annet brukes til bleking av papir (se figur 6.7).

Slurry blir levert til flere papirmøller i Europa. Etterspørselen per år av slurry til en av disse møllene ble registrert i tidsrommet 2009-2013. Resultatet ser du i tabellen i figur 6.6. I denne sammenheng er det hensiktsmessig å definere følgende variabler:

$$x = \text{år nr. } x, \text{ hvor } x = 2009, 2010, 2011, 2012, 2013 \tag{6.11}$$
$$y = \text{etterspørsel av slurry per år (i tusen tonn)} \tag{6.12}$$

Etterspørsel av slurry i perioden 2009-2013:

x (år)	2009	2010	2011	2012	2013
y (etterspørsel i tusen tonn)	228	240	245	248	258

Figur 6.6: År x og etterspørsel y.

Ut fra tabellen i figur 6.6 kan man regne ut den empiriske variansen til etterspørselen av slurry y, dvs. S_y^2, og den empiriske variansen til x, dvs. S_x^2. Resultatet er:

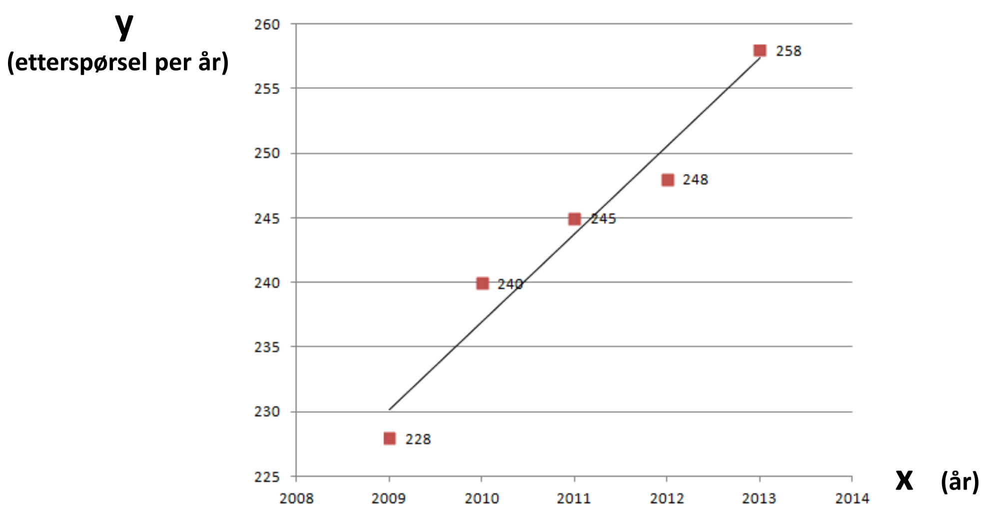
$$S_y^2 = 121.2 \quad , \quad S_x^2 = 2.5 \tag{6.13}$$

Den empiriske **ko**variansen mellom x og y er:

$$S_{xy} = 17 \tag{6.14}$$

(Størrelsene i lign.(6.13) og (6.14) trenger du å ikke regne ut. Bare ta dem for gitt. Benevningen til disse størrelsene er utelatt).

Figur 6.7: Hustadmarmor. Slurry.



Figur 6.8: Plott av dataene fra tabellen i figur 6.6.

d) Den rette linjen i figur 6.8 viser minste kvadraters regresjonslinje for x og y. Regn ut gjennomsnittene $\overline{x}$ og $\overline{y}$ og finn et analytisk uttrykk for denne lineære **regresjonslinjen**. [11]

e) Hvor mye predikerer regresjonslinjen at etterspørselen vil være i år 2017?

f) Hvor mye predikerer regresjonslinjen at den gjennomsnittlige årlige etterspørselen vil øke med? [12]

g) Alle papirmøller har en eller flere lagertanker. I en lagertank blir slurryen lagret. Anta at den aktuelle møllen som vi ser på har en lagertank med kapasitet slik at den kan håndtere 305 tusen tonn slurry i året.

Hva slags årstall må den aktuelle papirmøllen ha større tank for at kapasiteten på tanken ikke skal være en begrensende faktor?

[11] Dvs. finn formelen for linjen via setningen for minste kvadraters lineære regresjonslinje.
[12] Denne oppgaven krever ingen regning.

For å finne forklaringskraften R^2 kan man bruke et dataprogram, f.eks. Excel, i stedet for å regne ut R^2 "for hånd" via definisjonen.

h) Finn forklaringskraften <u>uten</u> å gjøre noe regning "for hånd".
Bare les av fra Excel-utskriften i figur 6.9 nedenfor.
(Bruk 4 desimalers nøyaktighet).

i) Kommenter svaret i oppgave **4h**. [13]

A	B	C	D	E	F	G	H	I
SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0,976624482							
R Square	0,95379538							
Adjusted R Square	-1,666666667							
Standard Error	2,732520204							
Observations	1							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	5	462,4	92,48	61,92857	#NUM!			
Residual	3	22,4	7,466667					
Total	8	484,8						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	2,732520204	3	0,91084	0,429536	-6,81481871	12,27985912	-6,81481871	12,27985912
X Variable 4	-13431	1737,703036	-7,72917	0,004503	-18961,1466	-7900,853395	-18961,1466	-7900,853395
X Variable 5	6,8	0,86409876	7,869471	0,004275	4,050052095	9,549947905	4,050052095	9,549947905

Figur 6.9: Utskrift fra Excel.

■

[13] For ett gitt år, vi du si at regresjonslinjen predikerer etterspørselen av slurry i stor eller liten grad? Med stort eller lite presisjonsnivå?

Høgskolen i Molde

Eksamen i

MAT110 Statistikk 1

Eksamensdag	**:**	**Torsdag 28. mai 2015**
Tid	**:**	**09:00 – 13:00 (4 timer)**
Faglærer/telefonnummer	**:**	**Molde:** **Per Kristian Rekdal / 924 97 051** **Kristiansund:** **Terje Bach / 932 55 838**
Hjelpemidler	**:**	**KD + formelsamling**
Antall sider inkl. forsiden	**:**	**13 + vedlegg (1 side)**
Målform	**:**	**Norsk (bokmål)**

Noen generelle råd:

- **Skriv rett inn. Ikke bruk så mye tid på kladding.**
- **Kladdark skal ikke leveres inn. De blir ikke sensurert.**
- **Det er totalt 4 oppgaver. Dvs. i *gjennomsnitt* èn time per oppgave.**

Oppgave 1: (revisjon)

Revisjonsfirmaet BDO skal ta stikkprøver av regnskapet til en bedrift. Regnskapet består av totalt $N = 100\,000$ bilag. Anta at det er M antall bilag med feil i dette regnskapet.

BDO bestemmer seg for å trekke $n = 1000$ tilfeldige bilag og sjekker disse for feil, se figur 7.1. De bestemmer seg også for å definere den stokastiske variabelen X hvor:

$$X \;=\; \text{antall stikkprøver som inneholder feil}$$

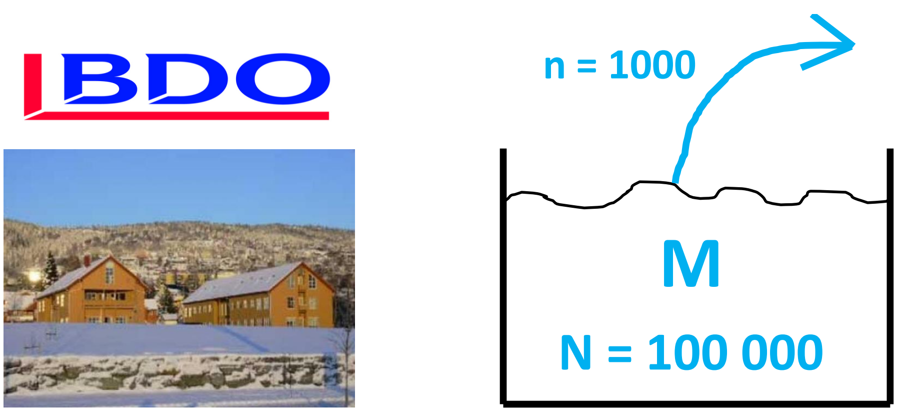
a) Hva slags sannsynlighetsfordeling har X? Gi en kort begrunnelse. [1]

b) Dersom forventet antall stikkprøver som inneholder feil er 10, dvs. $E[X] = 10$, vis da med en kort regning at $M = 1000$.

c) Vis at standardavviket til antall stikkprøver som inneholder feil er:

$$\sigma[X] \;=\; \sqrt{9.8} \;=\; 3.13 \tag{7.1}$$

Bruk at $M = 1000$.



Figur 7.1: BDO revisjon. Det trekkes $n = 1000$ bilag.

[1]Se f.eks. side 71 i formelsamlingen.

d) Dersom noen betingelser er oppfylt så kan en hypergeometrisk fordeling med god tilnærmelse beskrives av en normalfordeling.

i) Hva slags betingelser må da være oppfylt for at dette skal gjelde? [2]

ii) Er disse kriteriene oppfylt i vårt tilfelle?

Anta i resten av oppgaven at X er normalfordelt:

$$X \quad \sim \quad \mathrm{N}\big[\, E[X]\,,\, \sigma[X]\,\big] \quad , \tag{7.2}$$

hvor $E[X] = 10$ og med variansen som ble regnet ut i oppgave **1c**.

e) I vedlegget helt bakerst ser du et koordinatsystem.
Tegn inn for hånd tetthetsfunksjonen $f_X(x)$ for den stokastiske variabelen i lign.(7.2).

Anta at tetthetsfunksjonen sitt toppunkt er $f_X(x = E[X] = 10) \approx 0.13$.
Vær nøye med å tegne toppunktet på rett sted.
Vær også nøye med å tegne noenlunde riktig bredde på grafen. [3]

f) Revisoren underkjenner regnskapet dersom han finner at antall stikkprøver med feil er større enn eller lik en viss grense g, kalt forkastningsgrensen.
Denne forkastningsgrensen er bestemt ved at sannsynligheten for at $X \geq g$ er $10\,\%$, dvs.:

$$P(X \geq g) = 0.10 \tag{7.3}$$

Finn forkastningsgrensen g.

g) Arealet under tetthetsfunksjonen $f_X(x)$ beskriver en sannsynlighet.
Marker på figuren i vedlegget fra oppgave **1e**, gjerne med å skravere, arealet tilsvarende $P(X \geq g) = 0.10$.

■

[2] Se f.eks. side 79 i formelsamlingen.
[3] Husk: vedlegg skal leveres inn sammen med resten av din besvarelse.

Oppgave 2: (økonomi og logistikk)

Du er ansatt i tabloidavisen Dagbladet. Din jobb er å analysere kundemassen til avisen. I den sammenheng velger du å dele aviskjøperne i to kategorier:

1) Trofaste aviskjøpere
2) Spesielle aviskjøpere

Trofaste aviskjøpere er aviskjøpere som kjøper Dagbladet uansett. Uansett om det har skjedd en stor nyhet eller ikke.

Spesielle aviskjøpere er aviskjøpere som kun kjøper Dagbladet dersom det har skjedd en spesiell nyhet, f.eks. dersom det har skjedd en stor nyhet av internasjonal interesse.

La oss først se på situasjonen når en spesiell nyhet **ikke** har skjedd. Da har Dagbladet kun de trofaste aviskjøperne.

La D_1 være en stokastisk variabel som beskriver etterspørselen av aviser hos de trofaste aviskjøperne per dag. Anta at denne stokastiske variabelen er normalfordelt med:

$$D_1 \quad \sim \quad \mathrm{N}\big[\, \mu_1 = 15\,000 \,, \; \sigma_1 = 3000 \,\big] \tag{7.4}$$

a) Hva er sannsynligheten for at mer enn 21 000 trofaste aviskjøpere kjøper avisen en gitt dag? [4]

-USA i krig
Dagbladet

Figur 7.2: Spesielle nyheter. Salg.

[4] Dvs. finn $P(D_1 > 21\,000)$.

La oss nå derimot anta at det **har** skjedd en spesiell nyhet.
La D_2 være en stokastisk variabel som beskriver etterspørselen av aviser hos de spesielle aviskjøperne. Anta at også denne stokastiske variabelen er normalfordelt, men med annen forventing og annen varians:

$$D_2 \quad \sim \quad \mathrm{N}\big[\,\mu_2 = 18\,000\,,\, \sigma_2 = 4000\,\big] \tag{7.5}$$

Den totale etterspørselen D_{tot} av aviser per dag dersom det har skjedd en spesiell nyhet er da:

$$\boxed{D_{tot} = D_1 + D_2 \qquad (7.6)}$$

b) Hva er den forventede totale etterspørselen av aviser $E[D_{tot}]$ per dag dersom en spesiell nyhet har skjedd?

Anta at de to kategoriene aviskjøpere er uavhengige av hverandre.

c) Hva er variansen til den totale etterspørselen av aviser $Var[D_{tot}]$ per dag dersom en spesiell nyhet har skjedd?

Prisen på et eksemplar av Dagbladet er $p = 25$ NOK. Anta at gjennomsnittskostnaden per avis, hvor faste kostnader ikke er inkludert, er $k = 12$ NOK. I tillegg kommer faste kostnader på $c = 400\,000$ NOK. Fortjenesten per dag for Dagbladet er da beskrevet av den stokastiske variabelen:

$$\boxed{F \;=\; (p-k)D_{tot} \;-\; c \qquad (7.7)}$$

dersom det har skjedd en spesiell nyhet.

d) Finn den forventede fortjenesten til Dagbladet $E[F]$ per dag dersom det har skjedd en spesiell nyhet.

Man kan vise at en lineærkombinajon av normalfordelte stokastiske variabler fortsatt er normalfordelt dersom de stokastiske variablene er uavhengige. For vårt tilfelle betyr det at $D_{tot} = D_1 + D_2$ er normalfordelt med:

$$D_{tot} \quad \sim \quad \mathrm{N}\big[\, E[D_{tot}]\, ,\, \sigma[D_{tot}] \,\big] \quad , \tag{7.8}$$

hvor forventing og varians ble regnet ut i oppgavene **2b** og **2c**.

e) La oss nå regne ut samme sannsynlighet som i oppgave **2a**, men denne gangen når en spesiell nyhet har skjedd:

Hva er sannsynligheten for at den totale etterspørselen overstiger 21 000 aviser per dag dersom en spesiell nyhet har skjedd, dvs. hva er $P(D_{tot} > 21\,000 | S)$? [5]

Her betegner S begivenheten at det har skjedd en spesiell nyhet.

Anta at det er 20 % sannsynlighet for at det en gitt dag skjer en spesiell nyhet, dvs. anta at $P(S) = 0.20$.

f) Hva er 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, dvs. hva er $P(D_{tot} > 21\,000)$? [6]

■

[5] Bruk det faktum at D_{tot} er normalfordelt, jfr. lign.(7.8).

[6] Tips: Man kan skrive $P(D_1 > 21\,000) = P(D_{tot} > 21\,000 | \overline{\mathbf{S}})$. Bruk gjerne dette faktum samt resultatene fra oppgave **2a** og **2e** til å regne ut $P(D_{tot} > 21\,000)$ via formelen for oppsplitting av utfallsrommet Ω.

Oppgave 3: (petroleumslogistikk)

Ansatte som jobber på oljeplattform skal rapportere om avvik, ulykker, nestenulykker og forbedringsforslag knyttet til helse, miljø og sikkerhet. La oss se på antall innleverte rapporter per dag på oljeplattformen "Gullfaks C". I den sammenheng defineres den stokastiske variabelen:

$$X \;=\; \text{antall innleverte rapporter per dag på Gullfaks C}$$

Basert på historiske data finner man at sannsynlighetsfordelingen til X, dvs. $P(X = x)$, er gitt ved følgende tabell:

x	0	1	2	3	4
P(X=x)	0.09	0.28	0.41	0.17	0.05

Figur 7.3: Sannsynlighetsfordeling $P(X = x)$.

a) Vis at sannsynlighetsfordelingen $P(X = x)$ i tabellen i figur 7.3 er en **gyldig** sannsynlighetsfordeling.

Figur 7.4: Oljeplattformen "Gullfaks C" i Nordsjøen.

b) Hva er sannsynligheten for at det leveres inn 3 rapporter eller mer per dag?

c) i) Hva er forventet antall innleverte rapporter per dag, dvs. hva er $E[X]$?

ii) Hva er variansen til antall innleverte rapporter per dag, dvs. hva er $Var[X]$?

Istedet for å se på antall innleverte rapporter per dag så ønsker HMS ansvarlig på Gullfaks C å se på antall innleverte rapporter per dag i gjennomsnitt over et helt når, altså over $n = 365$ dager.

La X_i være en stokastisk variabel som beskriver antall innleverte rapporter for dag nr. i, hvor $i = 1, 2, ..., n$. Anta at alle de stokastiske variablene X_i har samme sannsynlighetsfordeling gitt ved tabellen i figur 7.3. Vi definerer gjennomsnittet $\overline{X}$:

$$\boxed{\overline{X} \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i \;=\; \frac{1}{n}\Big(X_1 + X_2 + \;...\; + X_n \Big) \qquad (7.9)}$$

d) i) Hva betyr $E[\overline{X}]$ på "godt norsk" i vårt tilfelle? [7]

ii) Finn $E[\overline{X}]$.

Anta at antall innleverte rapporter per dag er uavhengige.

e) i) Hva betyr $Var[\overline{X}]$ på "godt norsk" i vårt tilfelle?

ii) Finn $Var[\overline{X}]$.

f) Hvilken fordeling (tilnærmet) er det rimelig å anta at $\overline{X}$ har? Begrunn svaret.

[7] Dvs. gi en tolkning av $E[\overline{X}]$.

g) Sammenlign $E[X]$ med $E[\,\overline{X}\,]$ og $Var[X]$ med $Var[\,\overline{X}\,]$ fra oppgavene **3c** og **3e** foran. Kommenter svaret.

h) Hva er sannsynligheten for at samlet antall innleverte rapporter i løpet av et år er større enn 700? [8]

■

[8] Tips:

$$P\Big(X_1 + X_2 + \ \ldots \ + X_n > 700\Big) \ = \ P\left(\frac{X_1 + X_2 + \ \ldots \ + X_n}{n} > \frac{700}{n}\right) \tag{7.10}$$

Deretter kan du standardisere og bruke resultatene for $E[\,\overline{X}\,]$ og $\sigma[\,\overline{X}\,]$. Heltallskorreksjon behøves **ikke**.

Oppgave 4: (økonomi)

I kapittel 1 i kurset "*MAT110 Statistikk 1*" så vi på statistiske størrelser med både èn og to variabler. For statistiske størrelser med to variabler så vi blant annet på graden av samvariasjon mellom variablene.

a) For observasjonene x_1, x_2, ..., x_n har vi tilhørende observasjoner y_1, y_2, ..., y_n. I denne sammenheng kan man snakke om korrelasjonskoeffisienten R_{xy} til disse obervasjonene.

i) Hva slags mulige verdier kan R_{xy} ha?

ii) Hva er R_{xy} et mål på?

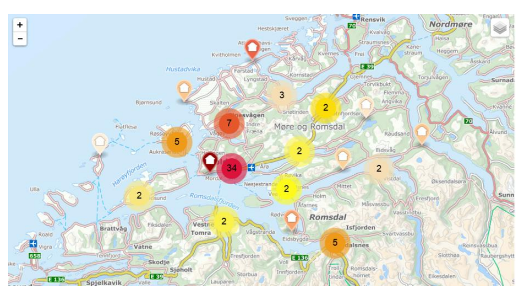
iii) Hva slags enhet har R_{xy}?

Eiendomsmegler "DnB Eiendom" ønsker å se nærmere på sammenhengen mellom kvadratmeterpris y på leiligheter og avstanden x som leilighetene har fra sentrum.

$$x = \text{avstand til sentrum (km)} \tag{7.11}$$

$$y = \text{kvadratmeterpris} \quad \text{(1000 NOK per } m^2 \text{)} \tag{7.12}$$

Observasjonene x måles i antall kilometer (km) fra sentrum. Observasjonene y er gjennomsnittlig kvadratmeterpris målt i 1000 NOK per m^2 for den gitte avstanden fra sentrum.



Figur 7.5: Leiligheter. Avstand fra sentrum.

Basert på salgene av leiligheter så lagt i år fant DnB Eiendom følgende resultat:

x (km fra sentrum)	0	5	10	15	20	25	30
y (gj. kvadratmeterpris)	57.5	48	45.5	39	33	32	28

Figur 7.6: Avstand fra sentrum x og gjennomsnittlig kvadratmeterpris y.

Ut fra tabellen i figur 7.6 kan man regne ut gjennomsnittene $\overline{x}$ og $\overline{y}$:

$$\overline{x} = 15 \qquad , \qquad \overline{y} = 40.43 \tag{7.13}$$

Man kan også regne ut den empiriske variansen til x, dvs. S_x^2, og den empiriske variansen til y, dvs. S_y^2. Resultatet er:

$$S_x^2 \;=\; 116.67 \qquad , \qquad S_y^2 \;=\; 109.54 \tag{7.14}$$

Den empiriske **ko**variansen mellom x og y er:

$$S_{xy} \;=\; -110.83 \tag{7.15}$$

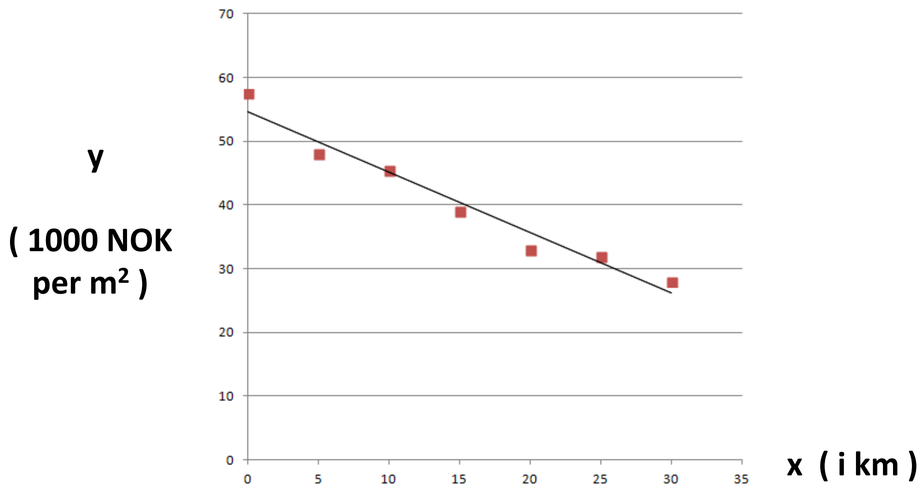
(Størrelsene i lign.(7.13), (7.14) og (7.15) trenger du å ikke regne ut. Bare ta dem for gitt. Benevningen, altså enheten, til disse størrelsene er utelatt).

b) Regn ut korrelasjonskoeffisienten R_{xy} for observasjonene i tabellen i figur 7.6.

c) Tolk svaret du fikk for R_{xy} i oppgave **4b**. [9]

[9] Hva sier den numeriske (tallmessige) verdien av R_{xy} fra oppgave **4b** om graden av korrelasjon mellom observasjonene x og y?

Pga. resultatene fra de foregående deloppgavene innser DnB Eiendom at man kan finne en eksplisitt lineær sammenhengen mellom x og y. Eiendomsmegleren bestemmer seg for å bruke lineær regresjonsanalyse.



Figur 7.7: Plott av dataene fra tabellen i figur 7.6.

d) Den rette linjen i figur 7.7 viser minste kvadraters regresjonslinje for observasjonene x og y. Finn et analytisk uttrykk for denne lineære **regresjonslinjen**.

e) DnB Eiendom har fått i oppdrag å selge ei leilighet som ligger 17 km utenfor sentrum. Hvor mye vil kvadratmeterprisen for denne leiligheten være ifølge regresjonslinjen?

f) Anta at gjennomsnittsprisen per kvadratmeter for leiligheter i Norge er 35 000 NOK/m^2.

Hvor lagt ut fra sentrum, i følge regresjonslinjen, ligger leiligheter som er på landsgjennomsnittet i kvadratmeterpris? [10]

[10] Husk at kvadratmeterpris er i antall 1000 NOK, dvs. gjennomsnittlig kvadratmeterpris i Norge er 35.

For å finne forklaringskraften R^2 kan man bruke et dataprogram, f.eks. Excel, i stedet for å regne ut R^2 "for hånd" via definisjonen.

g) Finn forklaringskraften uten å gjøre noe regning "for hånd".
Bare les av fra Excel-utskriften i figur 7.8 nedenfor.
(Bruk 4 desimalers nøyaktighet).

h) Kommenter svaret i oppgave **4g**. [11]

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	*Regression Statistics*								
4	Multiple R	0,9804							
5	R Square	0,9613							
6	Adjusted R Square	0,9535							
7	Standard Error	2,2567							
8	Observations	7							
9									
10	ANOVA								
11		*df*	*SS*	*MS*	*F*	*Significance F*			
12	Regression	1	631,75	631,75	124,0462833	0,000101753			
13	Residual	5	25,46428571	5,092857					
14	Total	6	657,2142857						
15									
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
17	y (pris)	54,68	1,537706349	35,55853	3,31064E-07	50,72577142	58,63137144	50,72577142	58,63137144
18	x (avstand)	-0,95	0,085296601	-11,1376	0,000101753	-1,169261894	-0,730738106	-1,169261894	-0,730738106

Figur 7.8: Utskrift fra Excel.

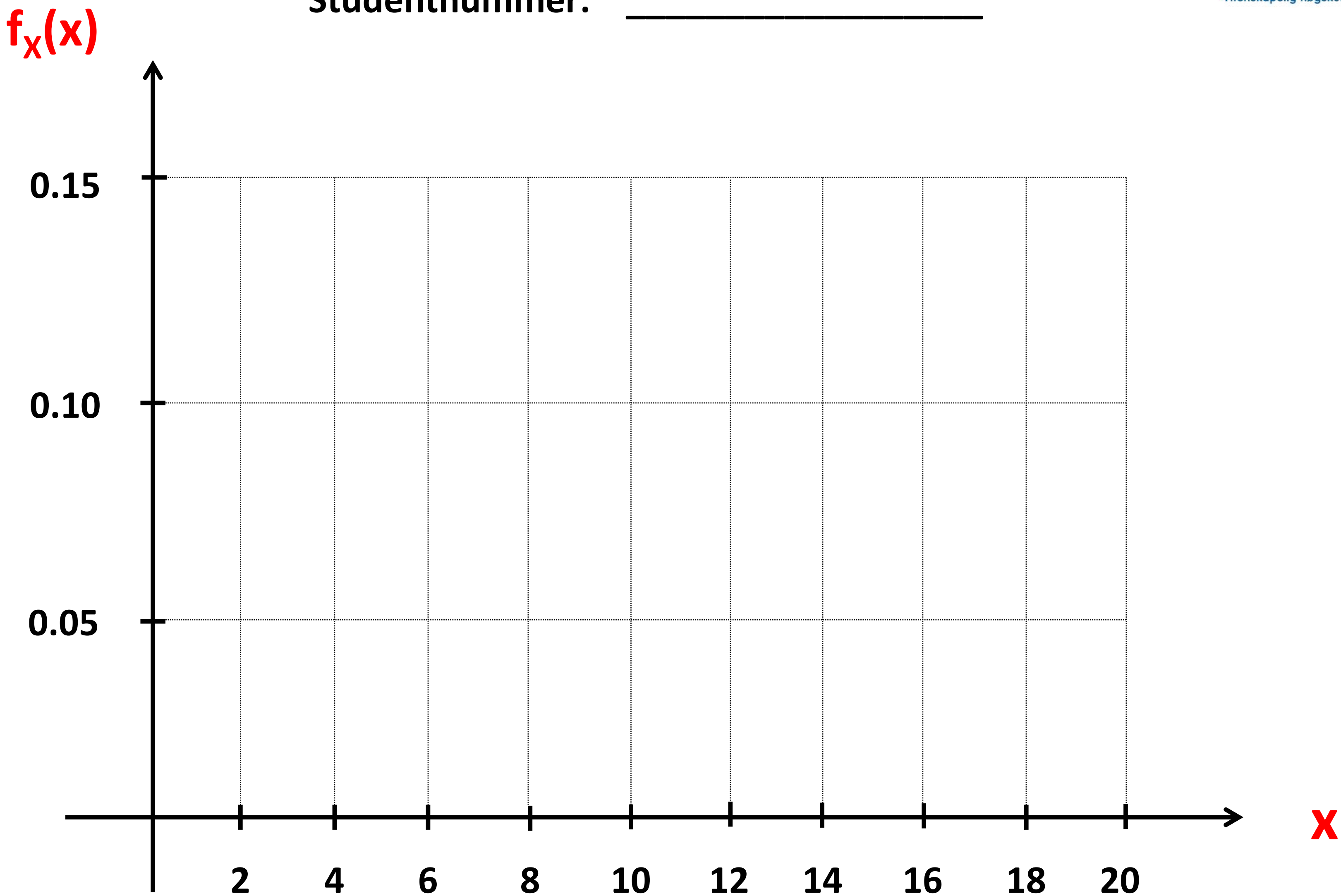
■

[11] For en leilighet med en gitt avstand fra sentrum, vi du si at regresjonslinjen predikerer kvadratmeterprisen i stor eller liten grad? Med stort eller lite presisjonsnivå?

Vedlegg

Høgskolen i Molde
Vitenskapelig høgskole i logistikk

Studentnummer: ____________________



Husk: dette vedlegg skal leveres inn sammen med resten av din besvarelse.

Høgskolen i Molde

Eksamen i

MAT110 Statistikk 1

Eksamensdag	**:**	**Fredag 8. januar 2016**
Tid	**:**	**09:00 – 13:00 (4 timer)**
Faglærer/telefonnummer	**:**	**Molde:** **Per Kristian Rekdal / 924 97 051** **Kristiansund:** **Terje Bach / 932 55 838**
Hjelpemidler	**:**	**KD + formelsamling**
Antall sider inkl. forsiden	**:**	**15**
Målform	**:**	**Norsk (bokmål)**

Noen generelle råd:

- **Skriv rett inn. Ikke bruk så mye tid på kladding.**
- **Kladdark skal ikke leveres inn. De blir ikke sensurert.**
- **Det er totalt 4 oppgaver. Dvs. i *gjennomsnitt* èn time per oppgave.**

Oppgave 1: (logistikk)

Et transportfirma har et varemottak for lastebiler med spesialgods.
Anta at firmaet har totalt 10 lastebiler, og det er tre forskjellige typer:
små, mellomstore og store lastebiler.

Av de totalt $n = 10$ lastebilene er det $n_1 = 5$ små lastebiler, $n_2 = 3$ mellomstore lastebiler og $n_3 = 2$ store lastebiler.

Anta at en tilfeldig lastebil kommer inn til varemottaket.
Anta videre at lastebilene ankommer varemottaket uavhengige av hverandre.

La oss definere sannsynlighetene:

$$P_1 = \text{sannsynligheten for at en liten lastebil ankommer varemottaket} \tag{8.1}$$
$$P_2 = \text{sannsynligheten for at en mellomstor lastebil ankommer varemottaket} \tag{8.2}$$
$$P_3 = \text{sannsynligheten for at en stor lastebil ankommer varemottaket} \tag{8.3}$$

a) Finn sannsynlighetene P_1, P_2 og P_3.

b) Vis at sannsynlighetene P_1, P_2 og P_3 utgjør en **gyldig** sannsynlighetsfordeling.

Figur 8.1: Vogntog og varemottak.

La oss definere den stokastiske variabelen X, hvor:

$$X = \text{antall minutter behandlingstid for å laste av lasten på en tilfeldig valgt lastebil} \quad (8.4)$$

Tiden det tar å laste av de tre typene lastebiler er:

$$\begin{aligned} &\text{laste av en liten lastebil:} && 10 \text{ minutter} && (8.5)\\ &\text{laste av en mellomstor lastebil:} && 20 \text{ minutter} && (8.6)\\ &\text{laste av en stor lastebil:} && 30 \text{ minutter} && (8.7) \end{aligned}$$

Anta at X har følgende sannsynlighetsfordeling:

x_i	10	20	30
$P(X=x_i)$	0.5	0.3	0.2

Figur 8.2: Sannsynlighetsfordeling $P(X = x_i)$, hvor $i = 1, 2, 3$.

c) Finn forventet behandlingstid $E[X]$ for å laste av lasten til en tilfeldig valgt lastebil. [1]

d) Finn $Var[X]$ for å laste av lasten til en tilfeldig valgt lastebil. [2]
Hva er tilhørende standardavvik $\sigma[X]$?

[1]Bruk gjerne definisjonen av forventning $E[X]$. Se formelsamling.
[2]Bruk gjerne definisjonen av varians $Var[X]$. Se formelsamling.

La oss nå se på en situasjon hvor det står 3 tilfeldige lastebiler foran deg i kø. La X_1, X_2 og X_3 være den tiden det tar for å laste av lasten på de tre lastebilene som står foran deg i køen. Ventetiden V for at din lastebil skal bli tømt er da:

$$\boxed{V \;=\; X_1 + X_2 + X_3 \qquad\qquad (8.8)}$$

e) Hva er fortventet ventetid $E[V]$?

f) Finn $Var[V]$, dvs. finn variansen til ventetiden.
Hva er tilhørende standardavvik $\sigma[V]$?

La oss nå innføre følgende kortnotasjon for den simultane sannsynligheten:

$$\boxed{p(x_1, x_2, x_3) \;\equiv\; P\big(\,X_1 = x_1 \;\text{ og }\; X_2 = x_2 \;\text{ og }\; X_3 = x_3\,\big) \qquad\qquad (8.9)}$$

g) Finn den simultane sannsynligheten $p(30, 30, 30)$. [3]

h) Gi en *kort* tolkning av sannsynligheten $p(30, 30, 30)$.

[3] Husk at lastebilene ankommer varemottaket uavhengige av hverandre.

Siden sannsynlighetsfordelingen til V også er normalisert, dvs. summerer seg til 1, så er:

$$\boxed{P(V \leq 80) \ + \ P(V = 90) \ = \ 1 \qquad (8.10)}$$

i) Bruk resultatet fra oppgave **1g** samt lign.(8.10) til finne sannsynligheten for at du må vente 80 minutter eller mindre, dvs. finn $P(V \leq 80)$.

For å finne sannsynligheten for at ventetiden V er akkurat 50 minutter så må vi summere alle kombinasjoner av de simultane sannsynlighetene som gir 50 minutters ventetid:

$$\boxed{\begin{aligned} P(V = 50) \ &= \ p(10,10,30) \ + \ p(10,30,10) \ + \ p(30,10,10) \\ &+ \ p(20,20,10) \ + \ p(20,10,20) \ + \ p(10,20,20) \qquad (8.11) \end{aligned}}$$

j) Finn sannsynligheten $P(V = 50)$,
dvs. finn sannsynligheten for at du må vente i 50 minutter før din lastebil blir ekspedert.

■

Figur 8.3: Kø. Ventetid.

Oppgave 2: (økonomi / aksjer)

"*Econa*" er en interesse- og arbeidstakerorganisasjon for siviløkonomer og masterutdannede innen økonomisk-administrative fag. *Econa* ønsker å finne ut om medlemmene i organisasjonen som har handlet aksjer på aksjemarkedet det siste året, også har en tendens til å lese næringslivsaviser som f.eks. "*Finansavisen*" eller "*Dagens Næringsliv*".

De ansatte i *Econa* definerer følgende begivenheter:

$$H = \text{medlem som handler aksjer} \tag{8.12}$$
$$\overline{H} = \text{medlem som IKKE handler aksjer} \tag{8.13}$$

$$R = \text{medlem som leser næringslivsaviser regelmessig} \tag{8.14}$$
$$I = \text{medlem som leser næringslivsaviser iblant} \tag{8.15}$$
$$A = \text{medlem som aldri leser næringslivsaviser} \tag{8.16}$$

Econa har funnet at disse sannsynlighetene gjelder for medlemsmassen:
(Tabellen viser "og"-sannsynlighetene, f.eks. $P(A \cap H) = 0.04$ og $P(A \cap \overline{H}) = 0.21$ osv.)

	(regelmessig)	(iblant)	(aldri)
Handle aksjer	**R**	**I**	**A**
H	0.18	0.10	0.04
$\overline{\textbf{H}}$	0.16	0.31	0.21

Figur 8.4: Sannsynligheter.

Figur 8.5: Econa og næringslivsaviser.

a) Hva er sannsynligheten for at en tilfeldig valgt medlem av *Econa* aldri leser næringslivsaviser?

Dvs. finn $P(A)$. Bruk gjerne setningen for total sannsynlighet (se formelsamling).

b) Gi en *kort* begrunnelse for hvorfor begivenhetene i lign.(8.12)-(8.16) er disjunkte. [4]

c) Hva er sannsynligheten for at en tilfeldig valgt medlem av *Econa* har handlet aksjer på aksjemarkedet det siste året?

d) Hva er sannsynligheten for at et medlem av *Econa* som aldri leser næringslivsaviser, har handlet aksjer på aksjemarkedet det siste året?

Dvs. finn $P(H|A)$. Bruk gjerne den generelle multiplikasjonssetningen.

e) Finn $P(A|H)$.

f) Gi en *kort* tolkning av hva $P(A|H)$ fra oppgave **2e** betyr. [5]

g) Beregn sannsynligheten $P\big(H|\overline{R}\big) = P\big(H|(A \cup I)\big)$ for at et medlem av *Econa* som ikke regelmessig leser næringslivsaviser, har handlet aksjer på aksjemarkedet det siste året. [6]

■

[4] En eller to setninger er nok. Se gjerne formelsamlingen for definisjon av disjunkte begivenheter.

[5] Èn setning er nok.

[6] Denne oppgaven er vanskelig. Hopp over den dersom du ikke får den til med en gang. Og kom heller tilbake til den senere. Tips: Bruk den generelle multiplikasjonssetningen.

Oppgave 3: (logistikk)

Helse Møre og Romsdal (Helse M&R) har ansvaret for ambulansetjenesten i fylket. Logistikerne i Helse M&R ønsker å se nærmere på antall utrykninger og responstid for å fordele kapasiteten av utrykningsbiler. De bestemmer seg for å begrense studien til 3 kommuner:
Eide, Fræna og Gjemnes.

Logistikerne i Helse M&R ønsker å modellere dynamikken ved hjelp av statistikk. De definerer derfor følgende stokastiske variabeler:

$$\begin{aligned}
X_E &= \text{antall } \textit{akutte} \text{ utrykninger per uke i Eide} && (8.17)\\
X_F &= \text{antall } \textit{akutte} \text{ utrykninger per uke i Fræna} && (8.18)\\
X_G &= \text{antall } \textit{akutte} \text{ utrykninger per uke i Gjemnes} && (8.19)
\end{aligned}$$

Siden utrykninger skjer relativt sjelden og siden vi ser på antall utrykninger per tid, altså en rate, så foreslår de å bruke "loven om sjeldne begivenheter", dvs. **Poisson**fordelingen:

$$X_i \sim \text{Poi}[\lambda_i] \qquad (8.20)$$

hvor $i = E, F, G$. Basert på erfaring fra tidligere år så vet Helse M&R følgende:

$$\begin{aligned}
\lambda_E &= 1.5 && (8.21)\\
\lambda_F &= 2.3 && (8.22)\\
\lambda_G &= 1.9 && (8.23)
\end{aligned}$$

Her er λ_i rater, dvs. antall akutte utrykninger per uke.

Figur 8.6: Ambulanse.

a) Hva er sannsynligheten for at det skjer 2 **utrykninger** i løpet av en uke i Gjemnes? [7]

b) Hva er sannsynligheten for at det skjer **mer enn** 2 utrykninger i løpet av en uke, dvs. $P(X_G > 2)$? Oppgitt: $P(X_G = 0) = 0.1496$ og $P(X_G = 1) = 0.2842$.

Definer den stokastiske variabelen:

$$\boxed{Y \;=\; X_E \;+\; X_F \;+\; X_G \qquad\qquad (8.24)}$$

dvs. Y = antall akutte utrykninger i Eide, Fræna og Gjemnes til sammen per uke.

c) Hva er forventet antall akutte utrykninger i Eide, Fræna og Gjemnes tilsammen per uke? [8]

Anta at de stokastiske variablene X_E, X_F og X_G er uavhengige.

d) Man kan vise at summen av Poissonfordelinger også er Poissonfordelt dersom Poissonfordelingene er uavhengige. [9]

Det betyr at siden X_E, X_F og X_G er uavhengige og Poissonfordelt, så er også summen av dem, $Y = X_E + X_F + X_G$, Poissonfordelt med forventning $E[Y]$:

$$Y \sim \text{Poi}\big[\,E[Y]\,\big] \tag{8.25}$$

Hva er sannsynligheten for at det skjer **mer enn** 2 utrykninger i Eide, Fræna og Eide tilsammen per uke, $P(Y > 2)$?

[7] Finn $P(X_G = 2)$. Se gjerne formelsamling angående formel for Poissonfordeling.
[8] Finn $E[Y]$. Bruk gjerne en av regnereglene på side 40 i formelsamlingen.
[9] Dette skal du ikke vise. Bare ta det for gitt.

e) Sammenlign svaret i oppgave **3d** med svaret i oppgave **3b**. Kommenter resultatet. [10]

Det er $n = 52$ uker i året. La oss nummerere disse ukene, $i = 1, 2, 3, ..., 52$, og slik at:

$$X_i \;=\; \text{antall } \textit{akutte} \text{ utrykninger i Gjemnes i uke nr. } i \tag{8.26}$$

Antall akutte utrykninger i året i Gjemnes kommune er da:

$$\boxed{X_{\text{år}} \;=\; X_1 + X_2 + X_3 + \;...\; + X_n \qquad (8.27)}$$

hvor $n = 52$. Anta videre at:

1. antall akutte utrykninger for de forskjellige ukene i Gjemnes er uavhengige:
   X_i er uavhengige for alle $i = 1, 2, 3, ..., n$
2. alle X_i er Poissonfordelte med samme rate $\lambda_i = \lambda_G = 1.9$
   X_i har samme sannsynlighetsfordeling for alle $i = 1, 2, 3, ..., n$

f) Hva er forventet antall akutte utrykninger i året i Gjemnes kommune, dvs. $E[\,X_{\text{år}}\,]$?

g) Hva er variansen til antall akutte utrykninger i året for Gjemnes kommune, dvs. $Var[\,X_{\text{år}}\,]$?

[10] Er det rimelig at det ene svaret et større enn det andre?

h) i) Med forutsetningene som nevnt på forrige side, hvilken setning gjelder da?

ii) Hvilken sannsynlighetsfordeling har, med god tilnærming, den stokastiske variabelen $X_{\text{år}}$?

iii) Hvor stor må n ($n =$ antall "forsøk") være, **omtrent**, for at setningen fra oppgave 4h i skal gjelde? [11]

i) For Gjemnes kommune er 110 akutte utrykninger i året ansett for å være mye.

Hva er sannsynligheten for at det er mer enn 110 akutte utrykninger i året i Gjemnes, dvs. hva er $P(X_{\text{år}} > 110)$? [12]

■

[11] Kun en tommelfingerregel er godt nok her.

[12] Du behøver ikke å bruke heltallskorreksjon.

Oppgave 4: (logistikk)

I kapittel 1 i kurset "*MAT110 Statistikk 1*" så vi på statistiske størrelser med både èn og to variabler. For statistiske størrelser med to variabler så vi blant annet på graden av samvariasjon/korrelasjon mellom variablene.

a) For observasjonene x_1, x_2, ..., x_n har vi tilhørende observasjoner y_1, y_2, ..., y_n. I denne sammenheng kan man snakke om korrelasjonskoeffisienten R_{xy} til disse obervasjonene.

i) Hva slags mulige verdier kan R_{xy} ha?

ii) Hva er R_{xy} et mål på?

iii) Hva slags enhet har R_{xy}?

Brunvoll AS er en leverandør av propeller for verftsindustrien.

Brunvoll får flere typer ordrer. En av disse ordretypene er *serviceordrer*. Serviceordrer er ordrer fra kunder som ønsker service på sine propeller.

For å planlegge produksjonen så ønsker logistikerne hos Brunvoll å få oversikt over antall serviceordrer. I den sammenheng har de samlet antall serviceordrer for hvert kvartal i 2014 og 2015, se tabellen i figur 8.8.

Figur 8.7: Brunvoll AS.

Logistikerne hos Brunvoll nummererer kvartalene i 2014 og 2015 fra 1 til 8, med x_i hvor $i = 1..8$. De innfører også notasjonen y_i for antall motatte serviceordrer tilhørende kvartal nummer i:

$$x_i = \text{kvartal nr. } i \quad (i = 1..8) \tag{8.28}$$
$$y_i = \text{antall serviceordrer mottatt i kvartal nr. } i \quad (i = 1..8) \tag{8.29}$$

Målingene fra 2014 og 2015 er oppsummert i følgende tabell:

	2014				2015			
x_i (kvartal nr. *i*)	1	2	3	4	5	6	7	8
y_i (antall serviceordrer)	28	30	29	34	33	38	37	40

Figur 8.8: Kvartal og serviceordrer.

Ut fra tabellen i figur 8.8 kan man regne ut gjennomsnittene $\overline{x}$ og $\overline{y}$:

$$\overline{x} = 4.5 \qquad , \qquad \overline{y} = 33.625 \tag{8.30}$$

Man kan også regne ut den empiriske variansen til x, dvs. S_x^2, og den empiriske variansen til y, dvs. S_y^2. Resultatet er:

$$S_x^2 = 6 \qquad , \qquad S_y^2 = 19.7 \tag{8.31}$$

Den empiriske **ko**variansen mellom x og y er:

$$S_{xy} = 10.36 \tag{8.32}$$

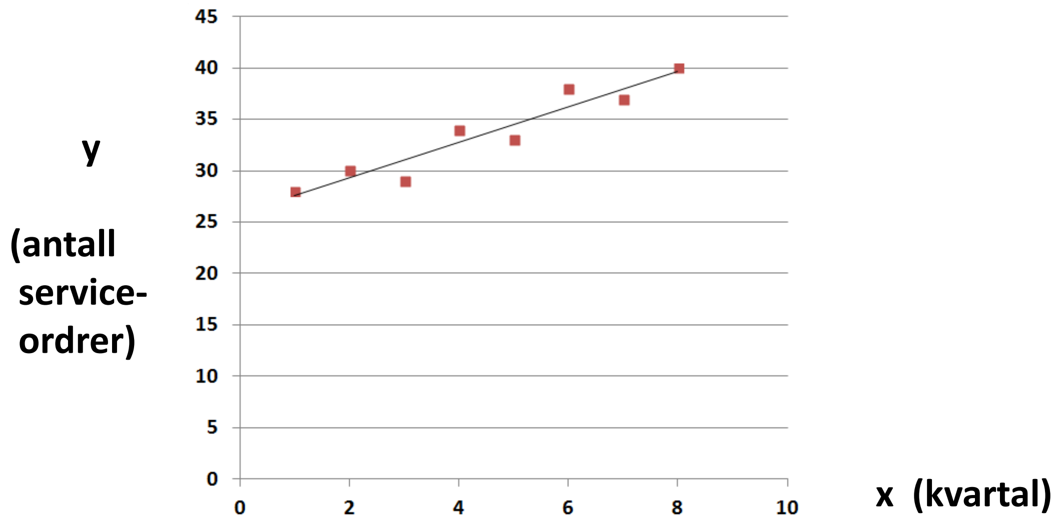
(Størrelsene i lign.(8.30), (8.31) og (8.32) trenger du ikke å regne ut. Bare ta dem for gitt).

b) Regn ut korrelasjonskoeffisienten R_{xy} for observasjonene i tabellen i figur 8.8.

c) Tolk svaret du fikk for R_{xy} i oppgave **4b**. [13]

[13] Hva sier den numeriske (tallmessige) verdien av R_{xy} fra oppgave **4b** om graden av korrelasjon mellom observasjonene x og y?

Pga. resultatene fra de foregående deloppgavene innser logistikerne hos Brunvoll at man kan finne en eksplisitt lineær sammenhengen mellom x og y. Logistikerne bestemmer seg for å bruke lineær regresjonsanalyse.



Figur 8.9: Plott av dataene fra tabellen i figur 8.8.

d) Den rette linjen i figur 8.9 viser minste kvadraters regresjonslinje for observasjonene x og y.

Regn ut de nødvendige parametrene og finn et analytisk uttrykk for den lineære **regresjonslinjen**.

e) Anta at trenden som den lineære **regresjonslinjen** i oppgave **4d** representerer også holder deg for 2016.

Hvor mange serviceordrer vil Brunvoll få i siste kvartal i 2016, dvs. for $x = 12$? [14]

f) Når vil, ifølge den lineære **regresjonslinjen** fra oppgave **4d**, Brunoll få 50 ordrer i kvartalet?

[14] Her er det meningen at man skal regne seg frem til svaret. Man skal **ikke** bare lese av fra figur 8.9.

Ut fra tabellen i figur 8.8 kan man regne ut SSE, "*sum squared error*", og SST, "*sum square total*". Resultatene er:

$$SSE = 12.73 \qquad , \qquad SST = 137.88 \tag{8.33}$$

(Disse verdiene for SSE og SST trenger du ikke å regne ut. Bare ta dem for gitt.)

g) Finn forklaringskraften R^2.

h) Kommenter svaret i oppgave **4g**. [15]

■

[15]Vi du si at regresjonslinjen predikerer antall serviceordrer som Brunvoll får per kvartal, i stor eller liten grad? Med stort eller lite presisjonsnivå?

Høgskolen i Molde

Eksamen i

MAT110 Statistikk 1

Eksamensdag	**:**	**Fredag 27. mai 2016**
Tid	**:**	**09:00 – 13:00 (4 timer)**
Faglærer/telefonnummer	**:**	**Molde:** **Per Kristian Rekdal / 924 97 051** **Kristiansund:** **Terje Bach / 932 55 838**
Hjelpemidler	**:**	**KD + formelsamling**
Antall sider inkl. forsiden	**:**	**17**
Målform	**:**	**Norsk (bokmål)**

Noen generelle råd:

- **Skriv rett inn. Ikke bruk så mye tid på kladding.**
- **Kladdark skal ikke leveres inn. De blir ikke sensurert.**
- **Det er totalt 4 oppgaver. Dvs. i *gjennomsnitt* èn time per oppgave.**

Oppgave 1: (logistikk , økonomi)

Logistikkselskapet Bring transporterer gods for bedrifter mellom Molde og Volda på Sunnmøre. Det er to fergeforbindelser mellom Molde og Volda. Noen ganger er fergene fulle, og da må sjåføren av lastebilen til Bring vente til neste ferge. Slik venting koster både tid og penger for Bring.

Anta at vi har følgende tre opplysninger:

- sannsynligheten for at det ikke er ledig plass på ferge 1 er: 0.15
- sannsynligheten for at det ikke er ledig plass på ferge 2 er: 0.25
- sannsynligheten for at det er ledig plass på ferge 1 og ferge 2 er: 0.70

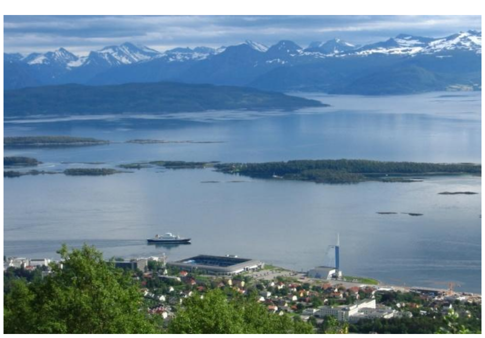
La oss definere følgende begivenheter:

$$L_1 \stackrel{\text{def.}}{=} \text{ledig plass på ferge 1} \tag{9.1}$$

$$L_2 \stackrel{\text{def.}}{=} \text{ledig plass på ferge 2} \tag{9.2}$$

Dersom vi konverterer den første opplysningen i oppgaven til et matematisk uttrykk, dvs. vi oversetter "*sannsynligheten for at det ikke er ledig plass på ferge 1 er 0.15*" til et matematisk uttrykk, så får vi:

$$P(\overline{L}_1) = 0.15 \tag{9.3}$$



Figur 9.1: Ferge og fergekø.

a) Skriv ned, uten utregninger, de matematiske uttrykkene for de to andre opplysningene oppgitt i oppgaven.

b) Hva er sannsynligheten for at det er ledig plass på minst èn av fergene, dvs. $P(L_1 \cup L_2)$? [1]

c) Hva er sannsynligheten for at det er ledig plass på ferge 1, men ikke på ferge 2, dvs. $P(L_1 \cap \overline{L}_2)$? [2]

d) Hva er sannsynligheten for at begge fergene er fulle, dvs. $P(\overline{L}_1 \cap \overline{L}_2)$? [3]

Dersom Bring ikke får plass på en ferge slik at de må vente til neste så får de ekstrautgifter pga. venting.

Dersom ferge 1 er full, så får Bring utgifter på 1 200 NOK.
Tilsvarende får Bring utgifter på 800 NOK dersom ferge 2 er full.

Istedet for å se på begivenheter, som vi har regnet på så langt i oppgaven, la oss nå introdusere en stokastisk variabel. Vi definerer den stokastiske variabelen U:

$$U \quad = \quad \text{utgift for Bring på fergestrekningene mellom Molde og Volda}$$

Utfallsrommet til U er:

$$\Omega \; = \; \{ \; 0 \; , \; 800 \; , \; 1\,200 \; , \; 2\,000 \; \} \tag{9.4}$$

[1] Tips: bruk den generelle addisjonssetningen.
[2] Tips: bruk setningen om total sannsynlighet.
[3] Tips: bruk en av tvillingsetningene.

Bring sine forventede utgifter på fergestrekningene mellom Molde og Volda er dermed:

$$E[U] = \sum_{i=1}^{4} u_i \cdot P(U = u_i) \tag{9.5}$$

$$\begin{aligned} &= \Big(0 \cdot P(L_1 \cap L_2) \ + \ 1200 \cdot \overbrace{P(\overline{L}_1 \cap L_2)}^{=\ 0.05} \\ &+ \ 800 \cdot P(L_1 \cap \overline{L}_2) \ + \ 2000 \cdot P(\overline{L}_1 \cap \overline{L}_2) \Big) \text{ NOK} \end{aligned} \tag{9.6}$$

Sannsynligheten for at første ferge er full, men ikke den andre, er $P(\overline{L}_1 \cap L_2) = 0.05$. (Dette faktum skal du ikke vise. Bare ta det for gitt.)

e) Hva er forventet utgift $E[U]$ for Bring? [4]

■

[4]Tips: bruk svarene for sannsynlighetene fra noen av de foregående deloppgavene samt lign.(9.6).

Oppgave 2: (logistikk , økonomi , "*avisguttens problem*")

En gutt jobber som avisselger. Han selger aviser for løssalg på gata. Etterspørselen av aviser en gitt dag kan beskrives av en stokastisk variabel D ("*demand*"), hvor:

$$D \; = \; \text{antall aviser som etterspørres en gitt dag} \tag{9.7}$$

Anta videre at sannsynlighetsfordelingen til denne stokastiske variabelen D er Poisson fordelt, dvs.:

$$\boxed{D \;\sim\; \text{Poi}[\,\lambda\,] \; . \qquad\qquad (9.8)}$$

hvor $\lambda = 81$.

a) Hva er forventet etterspørsel av aviser for en gitt dag, dvs. $E[D]$? [5]

b) Hva er standardavviket til etterspørselen av aviser for en gitt dag, dvs. $\sigma[D]$?

Figur 9.2: Avisgutt.

[5] Se f.eks. side 58 i formelsamlingen.

c) Begrunn *kort* hvorfor D er tilnærmet normalfordelt. [6]

d) Bruk at D er tilnærmet normalfordelt og finn sannsynligheten for at det etterspørres mer enn 80 aviser en tilfeldig dag. [7]

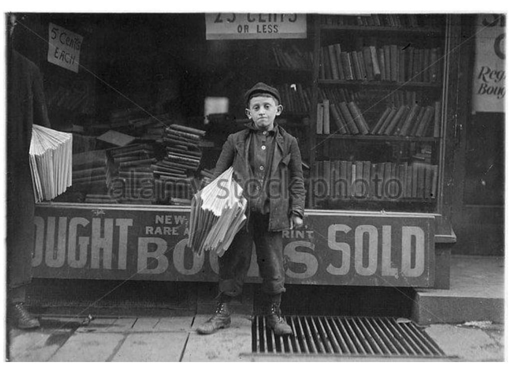
Hver morgen må avisgutten bestemme seg for hvor mange aviser han ønsker å prøve å selge. La oss si at avisgutten bestiller q antall ("*order quantity*") aviser fra distributøren en gitt morgen. Dersom han bestiller for få aviser så taper han salg. Dersom han bestiller for mange aviser så blir han sittende igjen med aviser som han ikke får solgt. La oss derfor introdusere en variabel S, hvor

$$S \;=\; \text{antall aviser som faktisk selges en gitt dag} \tag{9.9}$$

Denne variabelen er gitt ved

$$S = \min(D, q) \;, \tag{9.10}$$

hvor "$\min(D, q)$" betyr den minste størrelsen av D og q.
En morgen bestemmer avisgutten seg for å bestille $\underbrace{q = 85}_{\text{bestiller}}$ aviser.



Figur 9.3: Avisgutten bestiller $q = 85$ aviser.

[6] Se f.eks. side 75 i formelsamlingen.
[7] Dvs. finn $P(D > 80)$.

Anta at sannsynlighetsfordelingen til den stokastiske variabelen S er gitt ved:

$$P(S = s) \quad = \quad \begin{cases} P(D = d) & , \quad s \leq 84 \\ 0.3429 & , \quad s = 85 \\ 0 & , \quad s \geq 86 \end{cases} \tag{9.11}$$

hvor $P(D = d)$ er Poissonfordelt, dvs. $D \sim \text{Poi}[\,\lambda\,]$, jfr. lign.(9.8).

e) Forventningen $E[S]$ er gitt ved

$$E[S] \quad = \quad \sum_{s=1}^{84} sP(S = s) \;+\; 85 \cdot P(S = 85) \tag{9.12}$$

hvor

$$\sum_{s=1}^{84} sP(S = s) \quad = \quad 49.89 \tag{9.13}$$

Finn E[S]. [8]

f) Gi en tolkning av $E[S]$, dvs. forklar *kort* hva det betyr på "godt norsk".

g) Sammenlign $E[D]$ fra oppgave **2a** med $E[S]$ fra **2e**.
Er det rimelig at den ene verdien er større enn den andre?
Gi en *kort* begrunnelse.

[8] Altså du skal bare finne tallverdien for $E[S]$. Du skal ikke bevise lign.(9.12) og (9.13)). Bare ta dem for gitt.

Ved bestilling kjøpes avisene inn for prisen w ("*wholesale*") per avis. Avisgutten selger dem videre på gata for utslagsprisen r ("*revenue*") per avis. Fortjenesten per dag blir da:

$$\boxed{\pi(q) = r\,S - w\,q\ , \qquad (9.14)}$$

hvor, som tidligere angitt, $S = \min(D, q)$. Her er r, q og w konstanter. Anta at:

$$w = 5 \text{ NOK} \qquad \text{(innkjøpspris)} \tag{9.15}$$
$$r = 20 \text{ NOK} \qquad \text{(utsalgspris)} \tag{9.16}$$

i resten av oppgaven.

h) Hva er forventet fortjeneste $E[\pi(q)]$ per dag dersom avisgutten bestiller $q = 85$ aviser?

Ovenfor har vi behandlet D som en diskret stokastisk variabel. Dersom vi nå istedet behandler D som en **kontinuerlig** variabel så kan man vise at maksimal fortjeneste oppnås når [9]

$$P(D \leq q^*) = 1 - \frac{w}{r}\ , \tag{9.17}$$

hvor q^* er det antall aviser som avisgutten må kjøpe inn om morgenen for å maksimere sin fortjeneste. Sannsynligheten $P(D \leq q^*)$ i lign.(9.17) er altså den kumulative fordelingen til D. Anta videre at D er **normalfordelt** med forventning $\mu = 81$ og standardavvik $\sigma = \sqrt{81} = 9$, dvs.

$$D \;\sim\; \mathrm{N}\big[\,\mu = 81, \sigma = 9\,\big]\ . \tag{9.18}$$

[9] Du skal ikke vise lign.(9.17). Bare ta den for gitt.

i) Finn det antall aviser q^* som avisgutten må bestille for å få størst fortjeneste.

j) Med fordelingen som i lign.(9.18) er forventet etterspørsel av aviser en gitt dag lik 81, dvs. $E[D] = \mu = 81$. Denne forventningsverdien og verdien på q^* fra oppgave **2i** er *ikke* er sammenfallende.

Gi en *kort* forklaring på hvorfor den ene verdien er større enn den andre.

■

Oppgave 3: (logistikk)

Du jobber med logistikk og vegplanlegging for Statens Vegvesen i Møre og Romsdal. I forbindelse med planleggingen av den nye innfartsvegen til Molde så ønsker du å få mer kunnskap om trafikkulykkene på denne vegstrekningen. Du bestemmer deg derfor for å definere følgende stokastiske variabel:

$$X \quad = \quad \text{antall trafikkulykker på innfartsvegen til Molde } \textbf{per år}$$

Basert på erfaring har Statens Vegvesen funnet at X har følgende sannsynlighetsfordeling:

x	0	1	2	3	4	≥ 5
P(X=x)	0.25	0.30	?	0.15	0.10	0

Figur 9.4: Sannsynlighetsfordeling $P(X = x)$.

a) Er den stokastiske variabelen X diskret eller kontinuerlig?
Gi en *kort* forklaring. [10]

Figur 9.5: Innfartsvegen til Molde.

[10]Èn setning er nok.

b) Fra tabellen i figur 9.4 ser vi at alle sannsynlighetene $P(X = x)$ er kjent bortsett fra $P(X = 2)$. Begrunn hvorfor $P(X = 2) = 0.20$.

c) Finn $E[X]$.

La $X_1, X_2, X_3, ..., X_n$ være antall ulykker på den aktuelle vegstrekningen for år nr. 1, 2, 3,...,n. Det totale antall ulykker på den aktuelle vegstrekningen over n år er da:

$$\boxed{Y \;=\; X_1 + X_2 + X_3 + \;\ldots\; + X_n \qquad (9.19)}$$

Anta videre at:

1. antall ulykkene er er uavhengige slik at:

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

2. hvert år har samme sannsynlighetsfordeling (gitt ved tabell i figur 9.4):

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

d) La oss først se på en periode på to år, dvs. $n = 2$.
Da er: $Y = X_1 + X_2$.

Hva er sannsynligheten for at det skjer høyst èn trafikkulykke på den aktuelle vegstrekningen over en periode på to år? [11]

[11]Dvs. finn $P(Y \leq 1)$. Bruk formelen:

$$\begin{aligned} P(Y \leq 1) \;&=\; \overbrace{P(X_1 = 0)}^{0 \text{ ulykker år nr.1}} \cdot \overbrace{P(X_2 = 0)}^{0 \text{ ulykker år nr.2}} \;+\; \overbrace{P(X_1 = 1)}^{1 \text{ ulykke år nr.1}} \cdot \overbrace{P(X_2 = 0)}^{0 \text{ ulykker år nr.2}} \\ &+\; \underbrace{P(X_1 = 0)}_{0 \text{ ulykker år nr.1}} \cdot \underbrace{P(X_2 = 1)}_{1 \text{ ulykke år nr.2}} \end{aligned} \qquad (9.20)$$

hvor $P(X_1 = 0) = P(X_2 = 0) = 0.25$ osv., se tabellen i figur 9.4.

La oss i resten av oppgaven se på en periode på 30 år, dvs. $n = 30$.

e) Hva er forventet antall trafikkulykker på den aktuelle vegstrekningen over en periode på 30 år, dvs. $E[\,Y\,]$?

f) Anta at variansen til antall trafikkulykker per år, dvs. $Var[X]$, er: $Var[X] = 1.6475$. (Dette faktum skal du ikke vise. Bare ta det for gitt.)

Bruk denne opplysningen og finn variansen til antall trafikkulykker over en periode på 30 år, dvs. finn $Var[\,Y\,]$.

g) i) Med forutsetningene som på forrige side og dersom vi ser på et tilstrekkelig antall år, hvilken læresetning gjelder da for den stokastiske variabelen Y?

ii) Hvilken sannsynlighetsfordeling har da variabelen Y, med god tilnærmelse?

iii) Hvor stor må n ($n =$ antall år, dvs. antall "forsøk") være, **omtrent**, for at læresetningen fra deloppgave g i) skal gjelde? [12]

h) Hva er sannsynligheten at det skjer mer enn 50 trafikkulykker i løpet av 30 år, dvs. $P(Y > 50)$?

$\blacksquare$

[12] Kun en tommelfingerregel er godt nok her.

Oppgave 4: (økonomi)

Denne oppgaven handler om samvariasjon mellom variabler og lineær regresjonsanalyse med to variabler. Først noen generelle spørsmål. Deretter skal vi anvende teorien på et konkret eksempel innen økonomi.

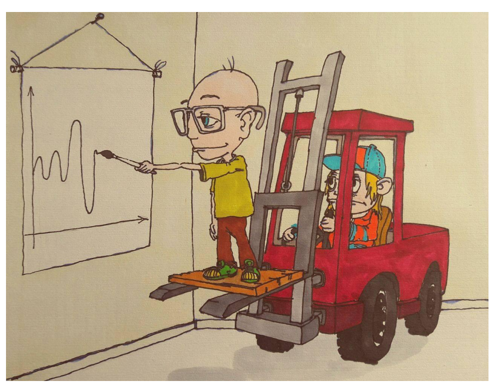
Anta at vi har observasjonene x_1, x_2, ... , x_n av en variabel x med tilhørende observasjoner y_1, y_2, ... , y_n av en variabel y. Når man har to slike tilhørende variabler så ønsker man ofte å se på graden av samvariasjon/korrelasjon mellom disse.

I denne sammenheng kan man snakke om den empiriske kovariansen S_{xy} til observasjonene og tilhørende korrelasjonskoeffisient R_{xy}.

a) La oss først se på den empiriske kovariansen S_{xy}.

i) Hva slags mulige verdier kan S_{xy} ha?

ii) Selv om fortegnet til S_{xy} har enkle tolkninger så er tallverdien til S_{xy} vanskeligere å tolke.
Hvorfor er tallverdien til S_{xy} vanskelig å tolke? [13]



Figur 9.6: Samvariasjon.

[13] Et par grunner er nok. Svar *kort* på denne oppgaven.

b) La oss nå se på den tilhørende korrelasjonskoeffisienten R_{xy}.

i) Hva slags mulige verdier kan R_{xy} ha?

ii) Hva er R_{xy} et mål på?

iii) Hva slags enhet har R_{xy}?

Bilprodusenten *Volvo Car Corporation* har hovedkontor utenfor Göteborg i Sverige.

Anta at du jobber som en del av teamet i Volvo som har med prissettingen av bilene å gjøre. Dere regner med at salget av biler vil øke dersom prisen senkes. I den sammenheng har dere foretatt en markedsundersøkelse ut fra 5 forskjellige prisnivåer x_1, x_2, x_3, x_4 og x_5 på bestselgeren Volvo V70. Tilhørende antall solgte biler per dag på global basis er y_1, y_2, y_3, y_4 og y_5.

$$x_i = \text{pris nr. } i \quad (\ i = 1, 2, ..., 5 \ ) \tag{9.21}$$
$$y_i = \text{antall solge biler per dag på global basis tilhørende pris nr. } i \quad (\ i = 1, 2, .., 5 \ ) \tag{9.22}$$

Figur 9.7: Volvo i Göteborg.

Markedsundersøkelsen gav følgende resultat:

x_i (prisnivå nr. i , i 100 000 NOK)	4	3.75	3.5	3.25	3
y_i (antall solgte biler per dag, i 1 000 stk.)	0.3	0.45	0.65	1	1.25

Figur 9.8: Pris x_i og salg y_i.

Ut fra tabellen i figur 9.8 kan man regne ut gjennomsnittene $\overline{x}$ og $\overline{y}$:

$$\overline{x} = 3.5 \qquad , \qquad \overline{y} = 0.73 \tag{9.23}$$

Man kan også regne ut den empiriske variansen til x, dvs. S_x^2, og den empiriske variansen til y, dvs. S_y^2. Resultatet er:

$$S_x^2 \;=\; 0.15625 \qquad , \qquad S_y^2 \;=\; 0.15325 \tag{9.24}$$

Den empiriske **ko**variansen mellom x og y er:

$$S_{xy} \;=\; -0.1531 \tag{9.25}$$

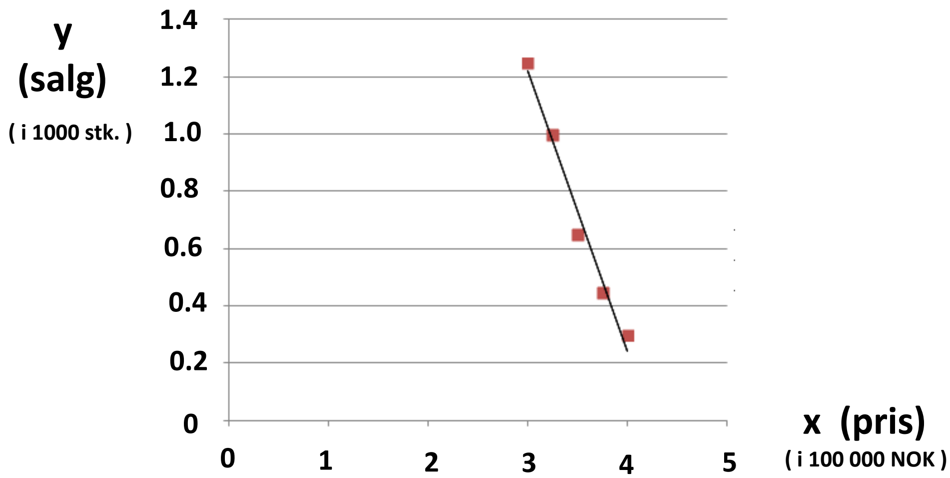
(Størrelsene i lign.(9.23), (9.24) og (9.25) trenger du ikke å regne ut. Bare ta dem for gitt).

c) Regn ut korrelasjonskoeffisienten R_{xy} for observasjonene i tabellen i figur 9.8.

d) Tolk svaret du fikk for R_{xy} i oppgave **4c**. [14]

[14]Hva sier den numeriske (tallverdien) av R_{xy} fra oppgave **4c** om graden av korrelasjon mellom observasjonene x og y?

Ut fra resultatene i markedsundersøkelsen så ønsker dere å finne en eksplisitt lineær sammenhengen mellom x og y. Dere bestemmer dere for å bruke lineær regresjonsanalyse.



Figur 9.9: Plott av dataene fra tabellen i figur 9.8.

e) Den rette linjen i figur 9.9 viser minste kvadraters regresjonslinje for observasjonene x og y.

Regn ut de nødvendige parametrene og finn et analytisk uttrykk for den lineære **regresjonslinjen** $\hat{y}$.

f) Anta at prisnivået som Volvo V70 ligger på i 2016 er 340 000 NOK, dvs. $x = 3.4$.

Ifølge regresjonslinjen du fant i oppgave 4e, hvor mange biler av typen V70 vil Volvo selge per dag dersom prisen er $x = 3.4$? [15]

g) Ledelsen i Volvo ønsker å ligge på et salgsvolum på 1 000 biler per dag. Hva slags pris x må de sette på sin bestselger Volvo V70, ifølge regresjonslinjen, for å oppnå salgsvolumet $\hat{y} = 1$?

[15] Her er det meningen at man skal regne seg frem til svaret. Man skal **ikke** bare lese av fra figur 9.9.

Ut fra tabellen i figur 9.8 kan man regne ut SSE, "*sum squared error*", og SST, "*sum square total*". Resultatene er:

$$SSE = 0.01275 \qquad , \qquad SST = 0.613 \tag{9.26}$$

(Disse verdiene for SSE og SST trenger du <u>ikke</u> å regne ut. Bare ta dem for gitt.)

h) Finn forklaringskraften R^2.

i) Kommenter svaret i oppgave **4h**. [16]

■

[16] Vi du si at regresjonslinjen predikerer antall solgte biler per dag, i stor eller liten grad? Med stort eller lite presisjonsnivå?

Høgskolen i Molde

Eksamen i

MAT110 Statistikk 1

Eksamensdag	**:**	**Torsdag 5. jan. 2017**
Tid	**:**	**09:00 – 13:00 (4 timer)**
Faglærer/telefonnummer	**:**	**Molde:** **Per Kristian Rekdal / 924 97 051** **Kristiansund:** **Terje Bach / 932 55 838**
Hjelpemidler	**:**	**KD + formelsamling** **+ generell ordbok** **morsmål/norsk/engelsk i papirformat**
Antall sider inkl. forsiden	**:**	**19**
Målform	**:**	**Norsk (bokmål)**

Noen generelle råd:

- **Skriv rett inn. Ikke bruk så mye tid på kladding.**
- **Kladdark skal ikke leveres inn. De blir ikke sensurert.**
- **Det er totalt 4 oppgaver.**

Oppgave 1: (produksjonsplanlegging , økonomi)

Malingsprodusenten Jotun er en industribedrift som produserer maling og pulverlakk.

Malingens vei fra pulver til ferdig produkt innbærer en del blandingsprosesser samt justeringer av fordelingen mellom pulver og andre stoffer.

Før malingen kan legges ut for salg må den tilfredsstille visse blandings- og kvalitetkrav. Malingen anses som "*defekt*" dersom den ikke tilfredsstiller disse kravene.

Jotun blander malingen i maksimalt 4 timer i en gitt tank. Etter hver time gjør de en test for å se om malingen er innenfor blandings- og kvalitetkravene. Dersom malingen ikke er det, så forsetter de blandings- og justeringsprosessen en time til. Første måling skjer etter èn time.

Blanding av maling koster tid og penger. Derfor ønsker Jotun å se nærmere på hvor lenge de bør blande malingen for å komme innenfor blandings- og kvalitetkravene.
I den sammenheng definerer Jotun følgende begivenheter:

$$D_t \stackrel{\text{def.}}{=} \text{malingen er fortsatt "}\textit{defekt}\text{" etter } t \text{ timer med blanding og justeringer} \tag{10.1}$$

hvor $t = 1, 2, 3, 4$.

Figur 10.1: Jotun.

Anta i denne oppgaven at:

$$P(D_1) \;=\; 0.30 \tag{10.2}$$

$$P(D_2|D_1) \;=\; 0.10 \tag{10.3}$$

$$P(D_3|D_2 \cap D_1) \;=\; 0.02 \tag{10.4}$$

$$P(D_4|D_3 \cap D_2 \cap D_1) \;=\; 0 \tag{10.5}$$

a) Finn $P(\overline{D}_1)$. [1]

b) Finn $P(\overline{D}_2|D_1)$. [2]

c) Gi en *kort* tolkning av sannsynligheten $P(\overline{D}_2 \cap D_1)$.

d) Bruk resultatet fra oppgave **1b** samt den generelle multiplikasjonssetningen til å vise at:

$$\boxed{P(\overline{D}_2 \cap D_1) \;=\; 0.27 \qquad (10.6)}$$

[1] Bruk komplementsetningen.

[2] Husk at komplementsetningen for en betinget sannsynlighet er: $P(\overline{B}|A) = 1 - P(B|A)$.

Sannsynligheten $P(\overline{D}_3 \cap D_2 \cap D_1)$, dvs. sannsynligheten for at malingen er OK etter 3 timer, kan finnes ved å bruke sannsynlighetstreet i figur 10.2.
Dersom man følger den rosa, stiplede linjen så innser man at:

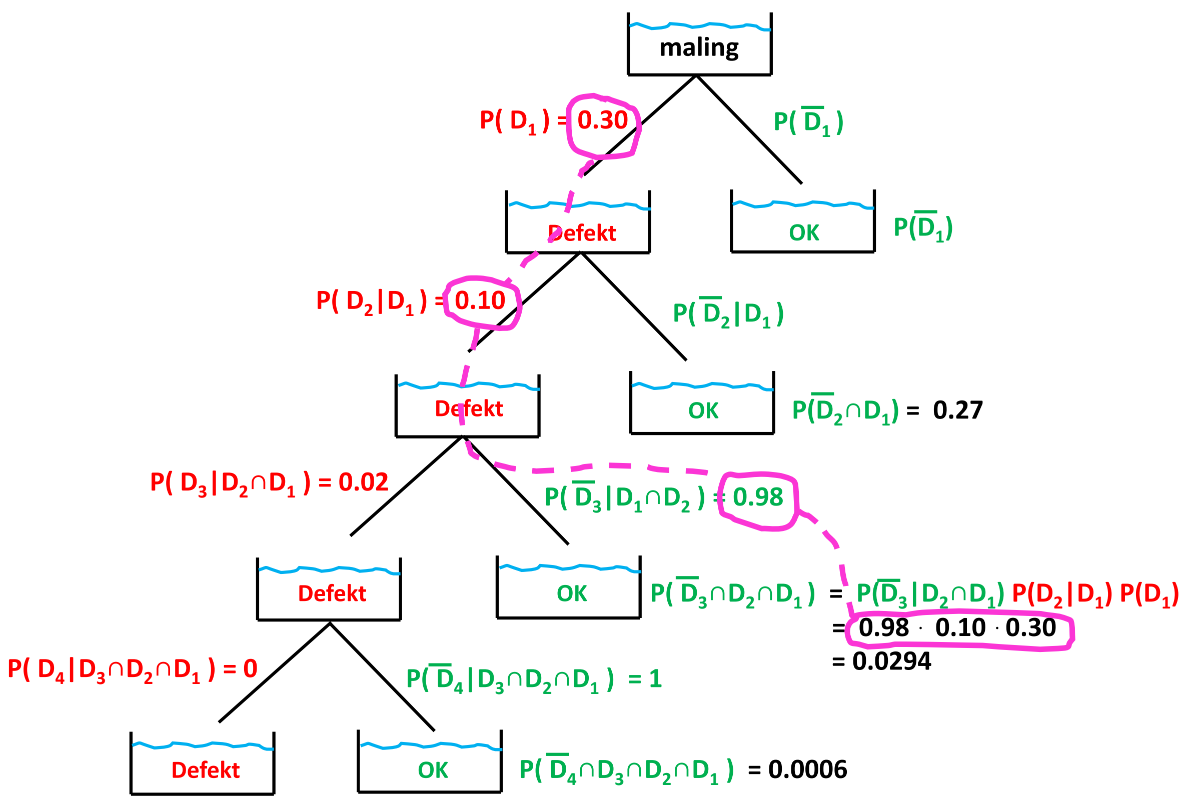
$$\underline{P(\overline{D}_3 \cap D_2 \cap D_1)} = {\color{green}P(\overline{D}_3|D_2 \cap D_1)}{\color{red}P(D_2|D_1)P(D_1)} \tag{10.7}$$

$$= 0.98 \cdot 0.10 \cdot 0.30 = \underline{0.0294} \tag{10.8}$$

e) Bruk sannsynlighetstreet i figur 10.2 og vis, slik som illustrert i lign.(10.7) og (10.8), at:

$$\boxed{P(\overline{D}_4 \cap D_3 \cap D_2 \cap D_1) = 0.0006. \qquad (10.9)}$$

hvor $P(\overline{D}_4 \cap D_3 \cap D_2 \cap D_1)$ er sannsynligheten for at malingen er OK etter 4 timer.



Figur 10.2: Sannsynlighetstre.

Istedet for å se på begivenheter, som vi har regnet på så langt i oppgaven, la oss nå introdusere en stokastisk variabel. Vi definerer den stokastiske variabelen X:

$$
\begin{aligned}
X \quad = \quad & \underline{\text{antall timer}} \text{ som Jotun bruker på blanding og justeringer av malingen} \\
& \text{før malingen kommer innenfor blandings- og kvalitetskravene}
\end{aligned}
\tag{10.10}
$$

med utfallsrom $\Omega = \{1, 2, 3, 4\}$. Forventet antall timer som Jotun bruker på blanding og justeringer av malingen før de kommer innenfor kravene er dermed:

$$
E[X] = \sum_{i=1}^{4} x_i \cdot P(X = x_i) \tag{10.11}
$$

$$
\begin{aligned}
= \quad & 1 \cdot P(\overline{D}_1) \; + \; 2 \cdot P(\overline{D}_2 \cap D_1) \\
+ \quad & 3 \cdot P(\overline{D}_3 \cap D_2 \cap D_1) \; + \; 4 \cdot P(\overline{D}_4 \cap D_3 \cap D_2 \cap D_1)
\end{aligned}
\tag{10.12}
$$

f) Bruk resultatene fra tidligere deloppgaver samt lign.(10.12) til å finne $E[X]$.

g) Vis at $P(X = x_i)$ er en gyldig sannsynlighetsfordeling. [3]

■

[3]Tips: vis at de 4 sannsynlighetene i lign.(10.12) summeres til 1.

Oppgave 2: (økonomi , aksjer)

Anta at du jobber i investeringsselskapet Gjelsten Holding som skal investere i tre ulike selskaper: $\underbrace{\text{Statoil}}_{\text{selskap } A}$, $\underbrace{\text{Seadrill}}_{\text{selskap } B}$ og $\underbrace{\text{Yara}}_{\text{selskap } C}$.

Prisen på aksjene i dag er 100 NOK for selskap A, 105 NOK for selskap B og 130 NOK for selskap C. Du bestemmer deg for å kjøpe

$$N \;=\; 100\,000 \tag{10.13}$$

antall aksjer, men du er usikker på hvor mange aksjer du skal kjøpe i de respektive selskapene. Derfor definerer du konstantene a, b og c:

$$N \cdot a \;=\; \text{antall aksjer som investeres i selskap } A\text{, Statoil} \tag{10.14}$$

$$N \cdot b \;=\; \text{antall aksjer som investeres i selskap } B\text{, Seadrill} \tag{10.15}$$

$$N \cdot c \;=\; \text{antall aksjer som investeres i selskap } C\text{, Yara} \tag{10.16}$$

hvor

$$\boxed{a \;+\; b \;+\; c \;=\; 1 \qquad (10.17)}$$

dvs. a er den brøkdelen av de $N = 100\,000$ aksjene som investeres i selskap A. Tilsvarende for b og c.

Figur 10.3: Selskap A, B og C.

For å avgjøre hvor mange aksjer du skal kjøpe i de respektive selskapene så ønsker du å finne ut mer om den potensielle fortjenesten av aksjene ved et eventuelt salg om ett år.

Derfor defineres følgende stokastiske variabler:

$$X = \text{pris på èn aksje (altså aksjekurs) for selskap } A\text{, Statoil (i NOK), om ett år} \tag{10.18}$$

$$Y = \text{pris på èn aksje (altså aksjekurs) for selskap } B\text{, Seadrill (i NOK), om ett år} \tag{10.19}$$

$$Z = \text{pris på èn aksje (altså aksjekurs) for selskap } C\text{, Yara (i NOK), om ett år} \tag{10.20}$$

Anta videre at forventet pris (altså aksjekurs) om ett år er: [4]

$$E[X] = 90 \quad , \quad E[Y] = 125 \quad , \quad E[Z] = 180 \tag{10.21}$$

med tilhørende varianser: [5]

$$Var[X] = 100 \quad , \quad Var[Y] = 200 \quad , \quad Var[Z] = 600 \tag{10.22}$$

Figur 10.4: Gjelsten Holding.

[4]Forventningene er i NOK, men dropper benevningen her for enkelhets skyld.
[5]Variansene er i NOK^2.

a) Vis at fortjenesten F om ett år er gitt ved: [6]

$$\boxed{F \;=\; Na\big(X-100\big) \;+\; Nb\big(Y-105\big) \;+\; Nc\big(Z-130\big) \qquad (10.24)}$$

b) Vis at den forventede fortjenesten $E[F]$ ved et eventuelt salg av aksjene om ett år er: [7]

$$\boxed{E[F] \;=\; 10N\Big(-a \;+\; 2b \;+5c\Big) \qquad (10.25)}$$

c) Anta nå at aksjene er uavhengige.
Bruk lign.(10.24) samt lign.(10.22) og vis at variansen $Var[F]$ til fortjenesten ved et eventuelt salg om ett år er:

$$\boxed{Var[F] \;=\; 100N^2\Big(a^2 \;+\; 2b^2 \;+\; 6c^2\Big) \qquad (10.26)}$$

d) Du ønsker å minimere risikoen i din investering,
dvs. du ønsker å minimere variansen $Var[F]$.
Man kan vise at variansen $Var[F]$ i lign.(10.26) er minst mulig når

$$a = 0.6 \qquad , \qquad b = 0.3 \qquad , \qquad c = 0.1 \tag{10.27}$$

(Du skal ikke vise lign.(10.27). Bare ta dette faktum for gitt.)

Hva blir fordelingen av antall aksjer for de tre selskapene dersom du ønsker å minimere risikoen i din investering? [8]

---

[6]Bruk gjerne at:

$$\boxed{F \;=\; \text{antall aksjer} \cdot \text{pris per aksje om ett år - antall aksjer} \cdot \text{pris per aksje i dag} \qquad (10.23)}$$

for hver av de tre selskapene. Bruk også lign.(10.14)-(10.16) samt lign.(10.18)-(10.20).

[7]Bruk lign.(10.24).

[8]Bruk lign.(10.14)-(10.16) sammen med lign.(10.27).

e) Dersom du bestemmer deg for å maksimere den forventede fortjenetsen $E[F]$ i lign.(10.25) istedet for å minimere risikoen $Var[F]$, hva slags selskap bør du investere i da?

Hva er verdien på den forventede fortjenesten $E[F]$ i det tilfellet?

Anta i resten av oppgaven at aksjene er avhengige på en slik måte at:

$$X = 90 + 10\epsilon \qquad , \qquad Y = 125 + 10\sqrt{2}\epsilon \qquad , \qquad Z = 180 + 10\sqrt{6}\epsilon \tag{10.28}$$

hvor ϵ er en stokastisk variabel som er standard normalfordelt, dvs.:

$$\epsilon \sim N\Big(E[\epsilon], \sigma[\epsilon] \Big) \tag{10.29}$$

med $E[\epsilon] = 0$ og $Var[\epsilon] = \sigma^2[\epsilon] = 1$.

f) Finn $E[X]$, $E[Y]$ og $E[Z]$.

g) Finn $Var[X]$, $Var[Y]$ og $Var[Z]$.

h) Sammenlign svarene i oppgavene **2f** og **2g** med lign.(10.21) og (10.22)
Gi en *kort* kommentar.

i) Siden den stokastiske variabelen ϵ er normalfordelt så er også X normalfordelt, dvs.:

$$X \sim N\Big(E[X], \sigma[X] \Big) \tag{10.30}$$

Hva er sannsynligheten for at aksjekursen for selskap A ligger mellom 80 NOK og 100 NOK om ett år, dvs. hva er $P\big(E[X] - \sigma[X] \le X \le E[X] + \sigma[X]\big)$? [9]

[9] Her behøves ingen utregninger. Se formelsamlingen.

Ved å sette lign.(10.28) inn i uttrykket for fortjenesten F i lign.(10.24) så får man:

$$F \tag{10.31}$$

hvor k er en konstant. (Lign.(10.31) skal du ikke vise. Bare ta den for gitt.)

j) Bruk lign.(10.31) og vis at variansen til fortjenesten F om ett år er

$$\boxed{Var[F] \;=\; (10N)^2\left(a+\sqrt{2}b+\sqrt{6}c\right)^2 \qquad (10.32)}$$

nå når de stokastiske variablene X, Y og Z er avhengige.

k) Man innser at variansen $Var[F]$ i lign.(10.32) er minimal dersom:

$$a = 1 \quad , \quad b = 0 \quad , \quad c = 0 \tag{10.33}$$

(Du skal ikke vise lign.(10.33). Bare ta dette faktum for gitt.)

Sammenlign resultatet i $\overbrace{\text{lign.(10.33)}}^{\text{avh. aksjer}}$ med $\overbrace{\text{lign.(10.27)}}^{\underline{\text{u}}\text{avh. aksjer}}$.
Gi en *kort* kommentar. [10]

■

[10] Dersom du ønsker minst mulig risiko i dine investeringer, dvs. minst mulig varians $Var[F]$, hvordan bør da fordelingen av aksjene på de tre selskapene være når aksjene er avhengige, jfr. lign.(10.33)? Og når de er uavhengige, jfr. lign.(10.27)?

Oppgave 3: (logistikk)

Helse Møre og Romsdal (Helse M&R) har ansvaret for ambulansetjenesten i fylket. Logistikerne i Helse M&R ønsker å se nærmere på antall utrykninger. De bestemmer seg for å begrense studien til 5 kommuner:

Gjemnes, Eide, Averøy, Sunndal, Fræna

Logistikerne i Helse M&R ønsker å modellere dynamikken til utrykningene ved hjelp av statistikk. De definerer derfor følgende stokastiske variabeler:

$$\boxed{X_i \;=\; \text{antall } \textit{akutte} \text{ utrykninger per uke i kommune nr. } i \qquad (10.34)}$$

hvor indeksen $i = 1, 2, .., 5$ betegner de 5 kommunene som vist i tabellen i figur 10.6.

Siden akutte utrykninger skjer relativt sjelden og siden vi ser på antall utrykninger per tid, altså en rate, mener Helse M&R at "loven om sjeldne begivenheter", dvs. **Poisson**fordelingen, beskriver fordelingen til variablene X_i:

$$\boxed{X_i \quad \sim \quad \text{Poi}[\lambda_i] \qquad (10.35)}$$

hvor $i = 1, 2, .., 5$.

Figur 10.5: Fem kommuner.

Basert på erfaring fra tidligere år så finner Helse M&R gjennomsnittlig antall akutte utrykninger per uke for de 5 aktuelle kommunene:

	i = 1 Gjemnes	i = 2 Eide	i = 3 Averøy	i = 4 Sunndal	i = 5 Fræna
λ_i	0.97	2.43	4.29	5.02	7.21

Figur 10.6: Parameteren λ_i.

hvor λ_i er rater, altså gjennomsnittlig antall akutte utrykninger per uke i kommune nr. i.
For eksempel er $\lambda_2 = 2.43$ for Eide kommune osv.
Helse M&R definerer nå den stokastiske variabelen:

$$\boxed{Y = X_1 + X_2 + X_3 + X_4 + X_5 \qquad (10.36)}$$

hvor Y = *summen* av antall akutte utrykninger i de 5 akutelle kommunene per uke.

a) Finn forventingen $E[Y]$. [11]

b) Finn standardavviket $\sigma[Y]$.

[11] Bruk gjerne en av regnereglene på side 39 i formelsamlingen.

Anta nå at de stokastiske variablene X_i er uavhengige.
Man kan vise at summen av Poissonfordelinger også er Poissonfordelt dersom Poissonfordelingene er uavhengige. [12]

Det betyr at siden X_i er uavhengige og Poissonfordelt, så er også summen av dem, $Y = X_1 + X_2 + X_3 + X_4 + X_5$, Poissonfordelt med forventning $E[Y]$:

$$Y \sim \text{Poi}\big[\, E[Y] \,\big] \tag{10.37}$$

hvor $E[Y]$ er svaret fra **3a**.

c) Begrunn *kort* hvorfor Y med god tilnærmelse er normalfordelt. [13]

d) Hva er sannsynligheten for at det til sammen skjer mer enn 25 akutte **utrykninger** i løpet av en uke i de 5 aktuelle kommunene? [14]

[12] Dette skal du ikke vise. Bare ta det for gitt.
[13] Tips: se f.eks. formelsamlingen på side 75.
[14] Dvs. finn $P(Y > 25)$.

La oss i resten av denne oppgaven kun se næremene på Eide kommune.

e) Hva er sannsynligheten for at det skjer 2 akutte **utrykninger** i løpet av en uke i Eide, dvs. hva er $P(X_2 = 2)$? [15]

Det er $n = 52$ uker i året. La oss nummerere disse ukene, $j = 1, 2, 3, ..., 52$, og slik at:

$$Z_j \;=\; \text{antall } \textit{akutte} \text{ utrykninger i Eide kommune i uke nr. } j \tag{10.38}$$

Antall akutte utrykninger i året i Eide kommune er da:

$$\boxed{Z_{\text{år}} \;=\; Z_1 + Z_2 + Z_3 + \;\ldots\; + Z_n \qquad (10.39)}$$

hvor $n = 52$. Anta videre at:

1. antall akutte utrykninger for de forskjellige ukene i Eide er uavhengige:
   Z_j er uavhengige for alle $j = 1, 2, 3, ..., n$
2. alle Z_j er Poissonfordelte med samme rate $\lambda_j = 2.43$ for alle $j = 1, 2, 3, ..., n$

f) Hva er forventet antall akutte utrykninger i året i Eide kommune, dvs. $E[\, Z_{\text{år}}\,]$?

g) Hva er variansen til antall akutte utrykninger i året i Eide kommune, dvs. $Var[\, Z_{\text{år}}\,]$?

[15]Se gjerne formelsamling angående formel for en Poissonfordeling.

h) i) Med forutsetningene som nevnt på forrige side, hvilken setning gjelder da?

ii) Hvilken sannsynlighetsfordeling har, med god tilnærming, den stokastiske variabelen $Z_{\text{år}}$?

iii) Hvor stor må n ($n =$ antall "forsøk") være, **omtrent**, for at setningen fra oppgave 3h i skal gjelde? [16]

i) For Eide kommune er 115 akutte utrykninger i året ansett for å være lite.

Hva er sannsynligheten for at det er mindre enn 115 akutte utrykninger i året i Eide, dvs. hva er $P(Z_{\text{år}} < 115)$?

■

[16] Kun en tommelfingerregel er godt nok her.

Oppgave 4: (logistikk)

Logistikerne i Helse M&R ønsker nå å se nærmere på antall utrykninger i forhold til antall innbyggere i kommunen. Derfor innfører de følgende notasjon for **observasjonene** x_i og y_i:

$$x_i = \text{antall innbyggere i kommune nr. } i \tag{10.40}$$

$$y_i = \text{gj.snittlig antall utrykninger } \underline{\text{per uke}} \text{ i kommune nr. } i \tag{10.41}$$

hvor, som i forrige oppgave, indeksen $i = 1, 2, .., 5$ betegner de 5 kommunene, se figur 10.7.

Logistikerne hos Helse M&R bruker fortsatt dataene λ_i fra tabellen i figur 10.6 fra oppgave 3, dvs. $y_i = \lambda_i$. Tilhørende x_i for de aktuelle kommunene er samlet i denne tabellen:

	i = 1 Gjemnes	i = 2 Eide	i = 3 Averøy	i = 4 Sunndal	i = 5 Fræna
x_i (antall innbyggere i kommune nr. i)	2557	3467	5826	7160	9717
y_i (gj.snittlig antall utrykninger per uke i kommune nr. i)	0.97	2.43	4.29	5.02	7.21

Figur 10.7: Observasjoner.

Ut fra tabellen i figur 10.7 kan man regne ut gjennomsnittene $\overline{x}$ og $\overline{y}$:

$$\overline{x} = 5745.4 \qquad , \qquad \overline{y} = 3.984 \tag{10.42}$$

Figur 10.8: Fem kommuner.

Man kan også regne ut den empiriske variansen til x, dvs. S_x^2, og den empiriske variansen til y, dvs. S_y^2. Resultatet er:

$$S_x^2 = 8\,284\,549 \qquad , \qquad S_y^2 = 5.76828 \tag{10.43}$$

Den empiriske **ko**variansen mellom x og y er:

$$S_{xy} = 6863.26 \tag{10.44}$$

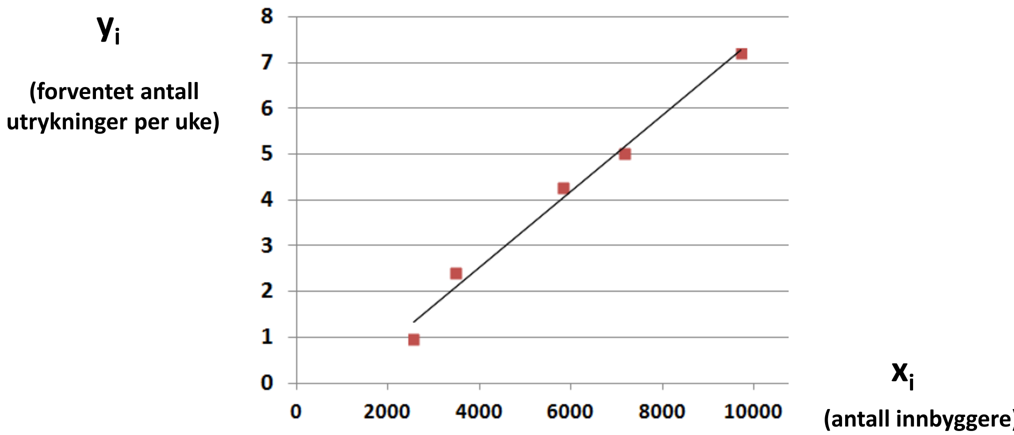
(Størrelsene i lign.(10.42), (10.43) og (10.44) trenger du <u>ikke</u> å regne ut. Bare ta dem for gitt).

a) Regn ut korrelasjonskoeffisienten R_{xy} for observasjonene i tabellen i figur 10.7.

b) Tolk svaret du fikk for R_{xy} i oppgave **4a**. [17]

[17] Hva sier den numeriske (tallverdien) av R_{xy} fra oppgave **4a** om graden av korrelasjon mellom observasjonene x og y?

Ut fra observasjonene og resulatene foran så ønsker Helse M&R å finne en eksplisitt sammenhengen mellom x og y. De bestemmer seg for å bruke lineær regresjonsanalyse.



Figur 10.9: Plott av dataene fra tabellen i figur 10.7.

c) Den rette linjen i figur 10.9 viser minste kvadraters regresjonslinje for observasjonene x og y.

Regn ut de nødvendige parametrene $\hat{\alpha}$ og $\hat{\beta}$ og finn et analytisk uttrykk for den lineære **regresjonslinjen** $\hat{y}$.

d) Rauma kommune har 7492 innbyggere, dvs. $x = 7492$.

Ifølge regresjonslinjen du fant i oppgave **4c**, hvor mange utrykninger per uke vil Rauma kommune ha? [18]

e) Anta at dersom en kommune har 6 aktutte utrykninger eller mer per uke så må Helse M&R ha èn ekstra person i telefonberedskap.

Hvor mange innbyggere må en kommune minst ha, ifølge regresjonslinjen fra oppgave **4c**, dersom Helse M&R må ha èn ekstra person i telefonberedskap? [19]

■

[18] Her er det meningen at man skal regne seg frem til svaret. Man skal **ikke** bare lese av fra figur 10.9.
[19] Bruk ligningen for regresjonslinjen til å finne x når $\hat{y} = 6$. Man skal **ikke** bare lese av fra figur 10.9.