



MAT110

Statistikk 1

Eksamensoppgaver 2012 - 2019

Per Kristian Rekdal
Bård-Inge Pettersen



Forord

Eksamensoppgaver:

Dette er en [samling av eksamensoppgaver](#) i emnet “*MAT110 Statistikk 1*” ved Høgskolen i Molde fra 2015 – 2019.

Det finnes også en tilhørende samling med komplette løsningsforslag til disse eksamensoppgavene. Samlingen med løsningsforslag finnes i en egen PDF-fil, separert fra dette oppgaveheftet.

Gratis:

Både samlingen med eksamensoppgaver 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 gamle eksamensoppgaver?:

Man blir ikke god i logistikk kun ved å se på video. Man må løse oppgaver.

Bård-Inge Pettersen
Per Kristian Rekdal

Copyright © Høgskolen i Molde, januar 2020.

Innhold

1	Hovedeksamen 2012	7
2	Kontinuasjoneksamen 2012	21
3	Hovedeksamen 2013	39
4	Kontinuasjoneksamen 2013	54
5	Hovedeksamen 2014	66
6	Kontinuasjoneksamen 2014	80
7	Hovedeksamen 2015	94
8	Kontinuasjoneksamen 2015	110
9	Hovedeksamen 2016	127
10	Kontinuasjoneksamen 2016	146
11	Hovedeksamen 2017	167
12	Kontinuasjoneksamen 2017	186
13	Hovedeksamen 2018	204
14	Kontinuasjoneksamen 2018	227
15	Testeksamen 2019	249
16	Hovedeksamen 2019	267
17	Kontinuasjoneksamen 2019	285

Kapittel 1

Hovedeksamen 2012



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*.

$$P(A) \cdot P(B) \tag{1}$$

$$P(A) + P(B) \tag{2}$$

$$P(A \cap B) \tag{3}$$

$$P(A \cup B) \tag{4}$$

- 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?



Oppgave 2: (betingede sannsynligheter, 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 $\overbrace{\text{ikke har betalingsproblemer}}^{= \bar{B}}$ er det 30 % i $\overbrace{\text{lavinntektsgruppen}}^{= L}$, dvs.
 $P(L|\bar{B}) = 0.30$



Figur 1: Sparebanken Møre.

La

B = begivenheten at en tilfeldig valgt kunde har **betalingsproblemer**

L = begivenheten at en tilfeldig valgt kunde er i **lavinntektsgruppen**

- a) i) Hva er $P(B)$? ¹
- ii) Hva er $P(\bar{B})$?

¹Her 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

$$P(B|L) = 0.38 \quad (5)$$

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 = fortjenesten (NOK) til Sparebanken Møre på en ny kunde i lavinntektsgruppen

Denne stokastiske variabelen har følgende sannsynlighetsfordeling:

x	- 10 000	8 000
$P(X=x)$	$P(B L)$	$P(\bar{B} L)$

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

- f) Finn den betingede sannsynligheten $P(\bar{B}|L)$.
- g) Er det lønnsomt å gi lån til de nye kundene i lavinntektsgruppen? Begrunn svaret ved regning. ■

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

³Dvs. regn ut $P(L)$.

Oppgave 3: (binomisk fordeling, 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.

Definèr 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 2: Kristiansund lufthavn, Kvernberget.

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

⁵Siden $X \sim \text{Bin}[n, p] = \text{Bin}[120, 0.95]$, 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 passasjører som møter opp til sin flyavgang. For passasjører som ikke møter opp blir pengene refundert. Anta at SAS tjener 800 NOK per passasjær på en reise Oslo-Kristiansund for alle passasjører som møter opp. Dermed kan vi definere en diskret stokastisk variabel I som beskriver **inntekten** til SAS for vår flyavgang:

$$I = a \cdot X \quad (6)$$

hvor

$$a = 800 \text{ NOK} \quad (\text{“a” er bare et konstant tall}) \quad (7)$$

- 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 passasjører 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 passasjører enn flyet har kapasitet til? ⁷

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

⁷Dvs. hva er $P(X \geq 121)$?

Dersom en billettgjøper 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:

$$U = b \cdot Y \quad (8)$$

hvor

$$b = 5000 \text{ NOK} \quad (\text{“}b\text{” er bare et konstant tall}) \quad (9)$$

$$Y = \text{antall personer med billett som møter opp til sin flyavgang} \\ \text{men } \mathbf{ikke \text{ får plass}} \quad (10)$$

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



⁸Dvs. finn $E[U]$, hvor U er gitt ved lign.(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 $X \sim \text{Bin}[n, p] = \text{Bin}[123, 0.95]$.

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

Oppgave 4: (normalfordeling)

Normalfordelingen $N[\mu, \sigma]$ 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 tilnærming beskrives av en normalfordeling under visse betingelser.

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

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (11)$$

hvor

$$\mu = E[X] \quad = \text{forventingen av } X \quad (\text{lokalisering av tyngdepunkt}) \quad (12)$$

$$\sigma = \sqrt{\text{Var}[X]} \quad = \text{standardavviket til } X \quad (\text{halvbredde til kurven}) \quad (13)$$

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:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (14)$$

- 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?

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

¹¹Bare 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 3: Sannsynlighetsfordeling $P(X = x)$ basert på historiske data.



Figur 4: Tidens Krav i Kristiansund.

- a) Vis at sannsynlighetsfordelingen i tabellen i figur (3) er en gyldig sannsynlighetsfordeling.

- b) i) Finn forventet antall feillevèringer per dag, dvs. $E[X]$.
 ii) Finn variansen til antall feillevèringer per dag, dvs. $Var[X]$.¹²

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

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n} \quad (15)$$

Anta videre at:

1. antall feillevèringer på ulike dager er uavhengige:
 $X_i \sim$ er uavhengige for alle $i = 1, 2, 3, \dots, n$
2. alle dager antas å ha samme sannsynlighetsfordeling (gitt ved tabell i figur 3):
 $X_i \sim$ samme sannsynlighetsfordeling for alle $i = 1, 2, 3, \dots, n$

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

- c) i) Finn forventet antall feillevèringer per dag i *gjennomsnitt* over ett år, dvs. $E[\bar{X}]$.¹³
 ii) Finn variansen til *gjennomsnittet* over ett år av antall feillevèringer per dag, dvs. $Var[\bar{X}]$.¹⁴

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

¹³Helt *generelt* gjelder regnereglen:

$$E[aX_1 + bX_2] = aE[X_1] + bE[X_2] \quad (\text{generelt}) \quad (16)$$

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

¹⁴Helt *generelt* gjelder:

$$Var[aX_1 + bX_2] = a^2 Var[X_1] + b^2 Var[X_2] + 2ab Cov[X_1, X_2] \quad (\text{generelt}) \quad (17)$$

hvor a og b er konstanter. Dersom X_1 og X_2 er uavhengige så er $Cov[X_1, X_2] = 0$, og lign.(17) reduserer seg til

$$Var[aX_1 + bX_2] = a^2 Var[X_1] + b^2 Var[X_2] \quad (\text{gjelder kun dersom } Cov[X_1, X_2] = 0) \quad (18)$$

- 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 \bar{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 feilleværing i året?
Du behøver ikke å bruke heltallskorreksjon. ¹⁷



God sommer!

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

¹⁶Kun en tommelfingerregel er godt nok her.

¹⁷Tips:

$$P(X_1 + X_2 + \dots + X_n > 180) = P\left(\frac{X_1 + X_2 + \dots + X_n}{n} > \frac{180}{n}\right) \quad (19)$$

$$= P\left(\bar{X} > \frac{180}{n}\right) \quad (20)$$

Deretter kan du standardisere lign.(20). Bruk 4 desimalers nøyaktighet.

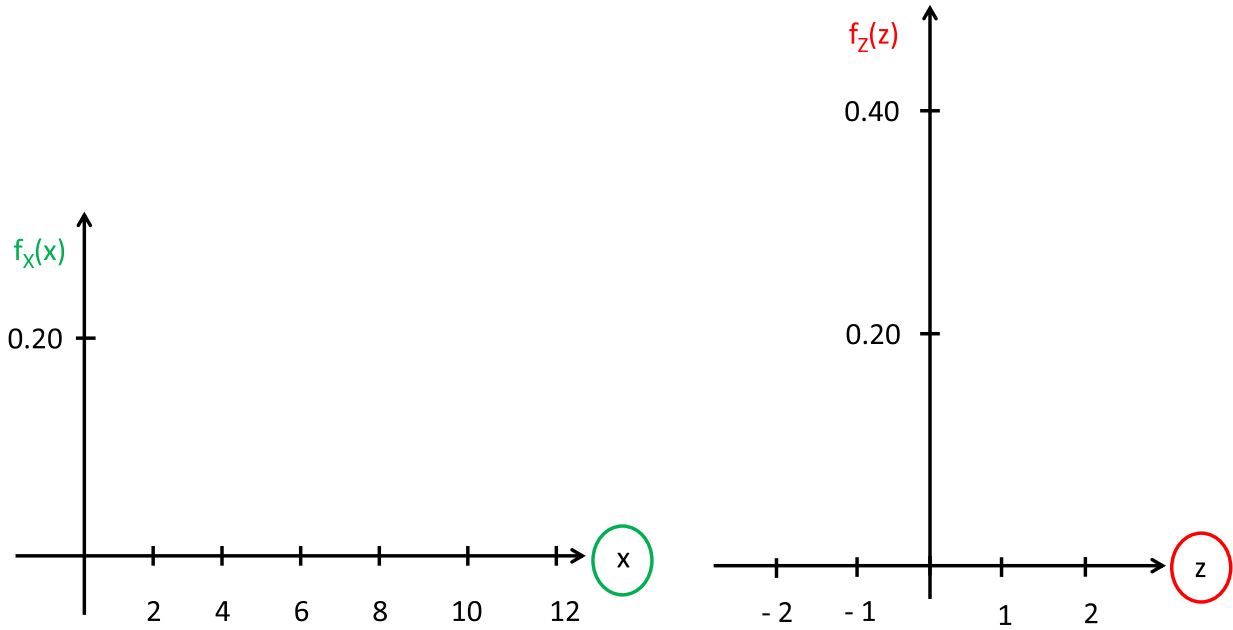
VEDLEGG

(Levèr inn arket sammen med besvarelsen din.)

Oppgave 4 b:

x-variabel:

z-variabel:



Figur 5: Tegn inn for hånd tetthetsfunksjone $f_x(x)$ og $f_z(z)$.

Oppgave 5 d:

tyngdepunkt

spredning

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

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

Figur 6: Fyll ut tabellen.

Kapittel 2

Kontinuasjonseksamen 2012



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:

$$P(E) = 0.52 \quad (1)$$

$$P(F) = 0.46 \quad (2)$$

$$P(G) = 0.38 \quad (3)$$

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:

$$P(E \cap F) = 0.42 \quad (4)$$

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

- a) Regn ut $P(E) \cdot P(F)$.
- b) Er hendelsene E og F uavhengige? ¹
- c) Er hendelsene E og G uavhengige?



Figur 1: Aksjer.

¹Hvilken 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:

$$P(B) = 0.45 \quad (6)$$

$$P(A|B) = 0.33 \quad (7)$$

$$P(A|\bar{B}) = 0.82 \quad (8)$$

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

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



²Dvs. hva er $P(\bar{B})$?

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

Oppgave 2: (binomisk- og normalfordeling, 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

$$X \sim \text{Bin}[n = 150, p = 0.9] \quad (9)$$

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



Figur 2: Åpningsuka ved Høgskolen i Molde. Transport til Hjertøya.

⁴Hvilke 4 krav må være oppfylt for at X skal være binomisk fordelt? Er disse kravene oppfylt i vårt tilfelle?

⁵Dvs. 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

⁶Se formelsamling.

⁷Dvs. finn $P(\text{en tur}) = P(X \leq 140)$. Bruk at $\text{Bin}[n, p]$ kan tilnærmes med en normalfordeling.

⁸Kan 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:

$$U = c \cdot Y, \quad (10)$$

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:

$$F = a \cdot X - c \cdot Y, \quad (11)$$

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



⁹Dvs. finn $E[U]$, hvor U er gitt ved lign.(10). Bruk gjerne definisjonen av forventning gitt ved lign.(5.3) i formelsamlingen, del I. Bruk også delsvarene fra oppgave **e** og **f**, dvs. bruk $P(Y = 1) = P(X \leq 140)$ og $P(Y = 2) = P(X > 140)$.

¹⁰Dvs. finn $E[F]$.

Oppgave 3: (sannsynlighetsfordelinger, normalfordelingen)

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

$$\text{Bin}[n, p] \quad (12)$$

$$\text{Hyp}[\mu, \sigma] \quad (13)$$

$$\text{Poi}[\lambda] \quad (14)$$

$$\text{N}[\mu, \sigma] \quad (15)$$

Det er sammenhenger mellom disse fordelingene. Under visse betingelser kan en fordeling med god tilnærming 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.(12)-(15). Til hver av disse store boksene hører det liten boks. I den lille boksen skal det fylles ut om fordelingen er

$$\text{kontinuerlig} \quad (16)$$

eller

$$\text{diskret} \quad (17)$$

a) Fyll ut alle boksene i vedlegg A.

Både de 5 store boksene ($\text{Bin}[n, p]$, $\text{Hyp}[\mu, \sigma]$, $\text{Poi}[\lambda]$, $\text{N}[\mu, \sigma]$) 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 $f_X(x)$:

$$f_X(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (18)$$

hvor

$$\mu = E[X] \quad = \text{forventingen av } X \quad (\text{lokalisering av tyngdepunkt}) \quad (19)$$

$$\sigma = \sqrt{\text{Var}[X]} \quad = \text{standardavviket til } X \quad (\text{halvbredde til kurven}) \quad (20)$$

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:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (21)$$

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 $f_Z(z)$?

■

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

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

Oppgave 4: (sentralgrensesetningen, økonomi og logistikk)

Bring er et norsk post- og logistikksselskap 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:

$$X = \begin{array}{l} \text{antall feilleveringer som et tilfeldig valgt bud i Stockholm gjør} \\ \text{en tilfeldig valgt dag} \end{array} \quad (22)$$

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 3: Sannsynlighetsfordeling $P(X = x)$.



Figur 4: Stockholm.

- a) Vis at sannsynlighetsfordelingen i tabellen i figur (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 $X_1, X_2, X_3, \dots, X_n$ være antall feilleveringer for et tilfeldig valgt bud for dag nr. 1, 2, 3, \dots , n . Gjennomsnittet av antall feilleveringer per dag for et bud hos Bring er da:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n} \quad (23)$$

Anta videre at:

1. antall feilleveringer på ulike dager er avhengige:
 X_i er avhengige for alle $i = 1, 2, 3, \dots, n$
2. alle dager antas å ha samme sannsynlighetsfordeling (gitt ved tabell i figur 3):
 X_i har samme sannsynlighetsfordeling for alle $i = 1, 2, 3, \dots, n$

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

¹³Dvs. hva betyr $E[X]$ på "godt norsk" i vårt tilfelle?

¹⁴Bruk 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 **b** med $E[\bar{X}]$ fra oppgave **d**. Kommenter resultatene. ¹⁷
- f) i) Med forutsetningene som på forrige side, hvilken setning gjelder da?
- ii) Hvilken sannsynlighetsfordeling har da gjennomsnittet \bar{X} ?
- iii) Hvor stor må n ($n =$ antall “forsøk”) være, **omtrent**, for at setningen fra oppgave i) skal gjelde? ¹⁸

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

$$E[aX_1 + bX_2] = aE[X_1] + bE[X_2] \quad (\text{generelt}) \quad (24)$$

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

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

$$Var[aX_1 + bX_2] = a^2 Var[X_1] + b^2 Var[X_2] + 2ab Cov[X_1, X_2] \quad (\text{generelt}) \quad (25)$$

hvor a og b er konstanter. Dersom X_1 og X_2 er uavhengige så er $Cov[X_1, X_2] = 0$, og lign.(25) reduserer seg til

$$Var[aX_1 + bX_2] = a^2 Var[X_1] + b^2 Var[X_2] \quad (\text{gjelder kun dersom } Cov[X_1, X_2] = 0) \quad (26)$$

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

¹⁸Kun 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



¹⁹Tips, du skal finne:

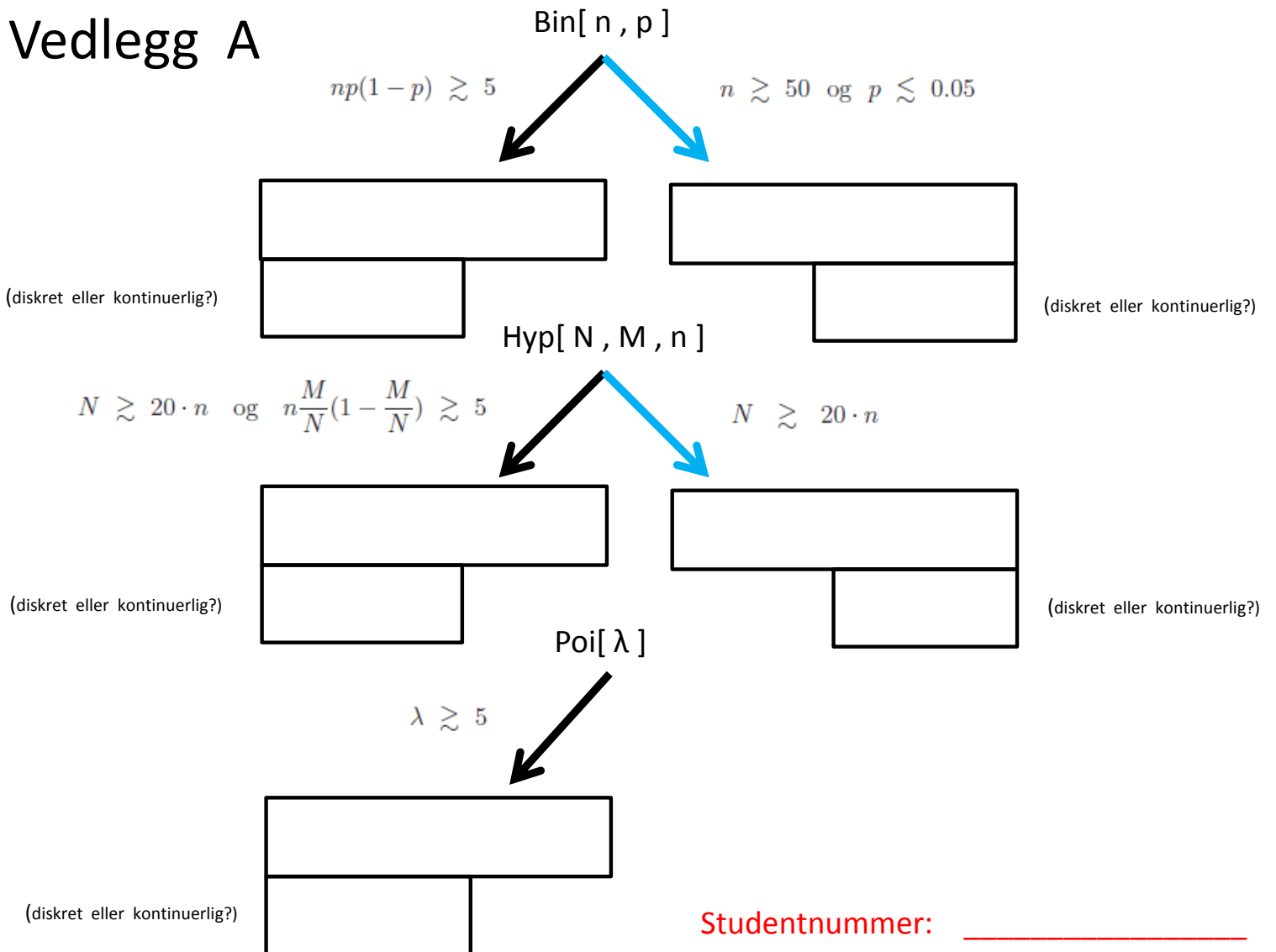
$$P(X_1 + X_2 + \dots + X_n > 200) = P\left(\frac{X_1 + X_2 + \dots + X_n}{n} > \frac{200}{n}\right) \quad (27)$$

$$= P(\bar{X} > \frac{200}{n}) \quad (28)$$

Deretter kan du standardisere lign.(28). Bruk 4 desimalers nøyaktighet.

Vedlegg A

Vedlegg A

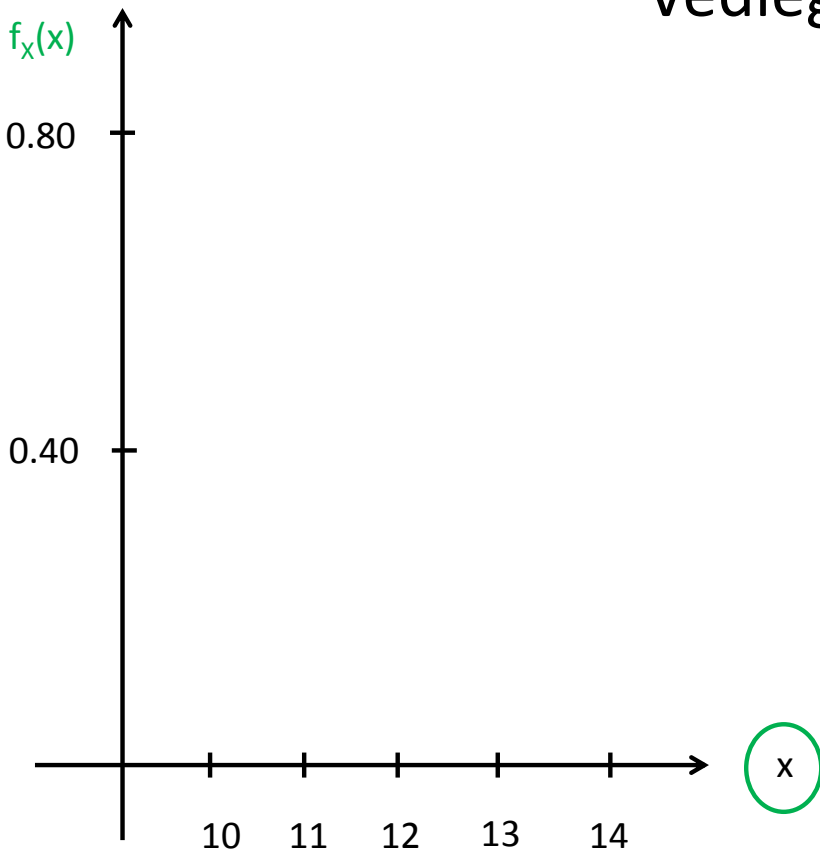


Studentnummer: _____

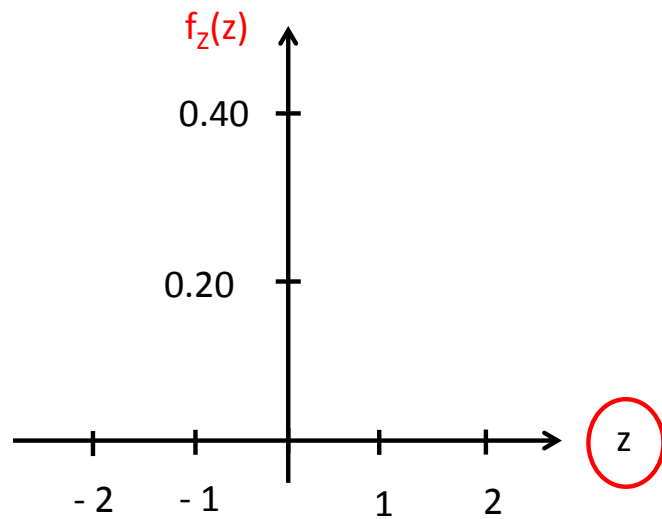
Vedlegg B

Vedlegg B

x-variabel:



z-variabel:



Studentnummer: _____

Kapittel 3

Hovedeksamen 2013

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).

Et mål for lineær samvariasjon. **Ikke** normalisert. (1)

Sterk negativ korrelasjon. (2)

Et mål på lineær samvariasjon, korrelasjon.
Normalisert, ligger i intervallet $-1 \leq \text{korr. koeff.} \leq 1$. (3)

Spredningsmål: varians (4)

Sterk positiv korrelasjon. (5)

Lokaliseringsmål: tyngdepunkt (6)



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

Oppgave 2: (sannsynlighetsregning, 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 = begivenheten at bølgehøyden **dag 1** er for høy for at supplybåten kan legge til (7)

B_2 = begivenheten at bølgehøyden **dag 2** er for høy for at supplybåten kan legge til (8)

Ved hjelp av værvarslings-tjenesten storm.no har du funnet ut at:

$$P(B_1) = P(B_2) = 0.05 \quad \text{og} \quad P(B_2|B_1) = 0.70$$

- 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 1: Riggen “*Searcher*”.

²Dvs. gi en tolkning av dette.

³Bruk 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



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

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

$$P(\overline{B}_2|\overline{B}_1) = \frac{P(\overline{B}_2 \cap \overline{B}_1)}{P(\overline{B}_1)} \quad (9)$$

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

Oppgave 3: (normalfordeling, sentralgrensesetningen, 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:

$$X = \text{antall overtrekk på en tilfeldig valgt lønnskonto per måned}$$

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 2: Sannsynlighetsfordeling $P(X = x)$.



Figur 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 $i = 1, 2, \dots, n$. Anta at alle variablene X_i har **samme sannsynlighetsfordeling** (gitt ved tabellen i figur 2). Vi definerer gjennomsnittet \bar{X} :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} (X_1 + X_2 + \dots + X_n) \quad (10)$$

- c) i) Hva betyr $E[\bar{X}]$ på “godt norsk” i vårt tilfelle? ⁶
- ii) Finn $E[\bar{X}]$.

Anta at antall overtrekk for de forskjellige lønnskontoene er **uavhengige**.

- d) i) Hva betyr $Var[\bar{X}]$ på “godt norsk” i vårt tilfelle?
- ii) Finn $Var[\bar{X}]$.
- e) **Hvilken fordeling** (tilnæremet) er det rimelig å anta at \bar{X} har? Begrunn svaret.

⁶Dvs. gi en tolkning av $E[\bar{X}]$.

f) Sammenlign $E[X]$ med $E[\bar{X}]$ og $Var[X]$ med $Var[\bar{X}]$ fra oppgavene foran. Kommenter 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



⁷Tips:

$$P(X_1 + X_2 + \dots + X_n > 400) = P\left(\frac{X_1 + X_2 + \dots + X_n}{n} > \frac{400}{n}\right) \quad (11)$$

$$= P(\bar{X} > \frac{400}{n}) \quad (12)$$

Deretter kan du standardisere lign.(12). Heltallskorreksjon behøves **ikke**.

⁸Tips:

$$P(X_1 + X_2 + \dots + X_n \leq X_{\text{grense}}) = 0.95 \quad (13)$$

$$P\left(\frac{X_1 + X_2 + \dots + X_n}{n} \leq \underbrace{\frac{X_{\text{grense}}}{n}}_{= \bar{X}_{\text{grense}}}\right) = 0.95 \quad (14)$$

$$P(\bar{X} \leq \bar{X}_{\text{grense}}) = 0.95 \quad (15)$$

Deretter kan du standardisere lign.(15):

$$P(\bar{Z} \leq \bar{Z}_{\text{grense}}) = 0.95 \quad (16)$$

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

Oppgave 4: (binomisk fordeling, logistikk og økonomi)

I 2010 fikk belysningsprodusenten Glamox i Molde en stor kontrakt med det “*danske sygehus-væsen*” 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 én forsendelse i uken, hvor både lysarmatur og tilhørende lysrør er inkludert. Det sendes $n = 25$ 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 definere den stokastiske variabelen:

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

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

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

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 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 $\text{Var}[D]$.
 ii) Hva betyr $\text{Var}[D]$ 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 produksjonskostnad k og transportkostnad k_t . Alt i alt, fortjenesten F for en gitt forsendelse er da:

$$F = \underbrace{(n - D - T) \cdot i}_{\text{inntekt}} - \underbrace{n \cdot (k + k_t)}_{\text{utgift}} \quad (17)$$

hvor

$$i = \text{inntekt per lampe med lysrør som fungerer, ("i" er bare et konstant tall)} \quad (18)$$

$$k = \text{produksjonskostnad per lampe med lysrør, ("k" er bare et konstant tall)} \quad (19)$$

$$k_t = \text{transportkostnad per lampe med lysrør, ("k_t" er bare et konstant tall)} \quad (20)$$

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

¹⁰Dvs. gi en tolkning av $E[D]$

¹¹Skriv 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:

$$E[F] = n \left[(1 - p_d - p_t) \cdot i - (k + k_t) \right] \quad (21)$$

hvor

$$p_d = \text{sannsynligheten for at et lysrør er defekt} \quad (22)$$

$$p_t = \text{sannsynligheten for at et lysrør blir ødelagt under transport} \quad (23)$$

Glamox har invitert Bring og DHL til å gi tilbud 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 5: Bring og DHL.

¹²Ved hjelp av rett frem algebra kan man vise at lign.(21) kan skrives:

$$E[F] = n \cdot \left[(1 - p_d) \cdot i - k - \underbrace{(p_t \cdot i + k_t)}_{\text{NB!}} \right] \quad (24)$$

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.: _____

Empirisk korrelasjonskoeffisient:

Korrelasjonskoeffisient: (formel)

.....
Kommentar:

Empirisk korrelasjonskoeffisient:

Korrelasjonskoeffisient: (formel)

.....
Kommentar:

Empirisk korrelasjonskoeffisient:

Korrelasjonskoeffisient: (formel)

.....
Kommentar:

Kapittel 4

Kontinuasjonseksamen 2013



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 **c** og **d**. Kommenter sammenligningen.
- f) Gi en kort konklusjon/forklaring på sammenligningen i oppgave **e**. ²



Figur 1: Revisjon.

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

²Firmaet 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:

$$D = \text{antall aviser som etterspørres en gitt dag} \quad (1)$$

Anta videre at sannsynlighetsfordelingen til denne stokastiske variabelen D er gitt ved fordelingen i figur (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 2: Sannsynlighetsfordeling $P(D = d_i)$, for $i = 0, 1, \dots, 4, 5$.

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

$$S = \text{antall aviser som faktisk selges en gitt dag} \quad (2)$$

Denne variabelen er gitt ved

$$S = \min(D, q) , \quad (3)$$

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

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

⁴Variabelen D er stokastisk. Variabelen S er avhengig av D , se lign.(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 (3):

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

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

- c) Vis at sannsynlighetsfordelingen i figur (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 **a** og $E[S]$ fra **d**.
 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:

$$\pi(q) = r S - w q, \quad (4)$$

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

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

$$E[aX + bY] = aE[X] + bE[Y]. \quad (5)$$

Bruk gjerne resultatet fra oppgave **d**.

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

$$P(D \leq q^*) = 1 - \frac{w}{r}, \quad (6)$$

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.(6) er altså den kumulative fordelingen til D . Anta videre at D er **normalfordelt** med forventning $\mu = 3$ og standardavvik $\sigma = 1.5$, dvs.

$$D \sim N[\mu = 3, \sigma = 1.5]. \quad (7)$$

- 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.(7) er forventet etterspørsel av aviser en gitt dag like tre, $\mu = 3$. Denne forventningsverdien $\mu = 3$ og verdien på q^* fra oppgave **g** er ikke sammenfallende. Gi en *kort* forklaring på hvorfor den ene verdien er større enn den andre. ■



Figur 4: Avisgutt.

⁶Du skal ikke vise lign.(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.

$$P(S_1) = P(S_2) = P(S_3) = \frac{1}{6} \quad (8)$$

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.

⁷Dvs. 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 5: Tusten.

⁸Bruk 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 (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 6: Sannsynlighetsfordeling $P(X = x)$.

Denne sannsynlighetsfordelingen har forventning

$$E[X] = 35 \tag{9}$$

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

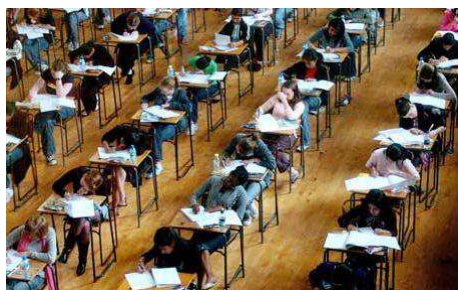
- a) Hva er sannsynligheten for at en tilfeldig valgt førsteårsstudent består 40 sp eller mer, dvs. hva er $P(X \geq 40)$?

- 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 c, 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 $i = 1, 2, 3 \dots n$. Anta videre at alle X_i har **samme fordeling**. La oss videre definere den stokastiske variabelen

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (10)$$

- e) Finn $E[\bar{X}]$.
- f) Gi en tolkning av $E[\bar{X}]$ for vår situasjon, dvs. forklar *kort* hva det betyr på “godt norsk”. ■



Figur 7: Eksamen.

⁹Bruk svaret fra oppgave a.

Kapittel 5

Hovedeksamen 2014



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)

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: (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 konkursobjekter:

$$P(\text{objekt}|\bar{K}) = 0.05 \quad (1)$$

- a) Hva er sannsynligheten for at en bedrift som ikke går konkurs ei heller blir kategorisert som et konkursobjekt? ¹

Det viser seg også at av de bedriftene som faktisk gikk konkurs ble 80 % klassifisert som konkursobjekter med KPMG sitt dataverktøy, dvs.:

$$P(\text{objekt}|K) = 0.80 \quad (2)$$

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

$$P(K) = 0.001 \quad (3)$$



Figur 1: KPMG.

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

b) Hvor stor sannsynlighet er det for at KPMG-verktøyet klassifiserer en bedrift som konkursobjekt? ²

$\underbrace{\hspace{10em}}_{\text{objekt}}$

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 $\underbrace{\text{ikke går konkurs}}_{\bar{K}}$ er: (med 4 desimaler nøyaktighet)

$$P(\bar{K}|\text{objekt}) = 0.9842 \quad (4)$$

d) Kommenter svaret i oppgave c):

Hva betyr $P(\bar{K}|\text{objekt})$ i lign.(4)?

I lys av dette, bør revisjonsselskapet KPMG trekke konklusjoner kun basert på dette verktøyet?

■

²Dvs. 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:

$$X = \text{antall personbiler som ankommer fergekaien i en } \underline{\text{30 minutters periode}}, \quad (5)$$

dvs. mellom to ferger

- 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
- $$\lambda = 115 \quad (6)$$
- 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 2: Ferge.

- d) i) En diskret Poisson-sannsynlighetsfordeling $\text{Poi}[\lambda]$ kan, under bestemte betingelser, med god tilnærming 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

³Se formelsamlingen.

⁴Dvs. finn $P(X > X_0 - 20)$. Benytt resultatet fra oppgave 2d i.

⁵Finn 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 3: Kiosk.

Modell 1:

For modell 1 defineres følgende stokastiske variabel:

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

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 fergeren 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

$$Y \sim \text{Bin}[n = 3\lambda, p = 0.11] \quad (7)$$

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

⁶Hvilke 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 **binomisk** fordeling tilnærmes med en **normalfordeling**. 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

⁷Se f.eks. side 60 i formelsamlingen.

⁸Dvs. finn $P(Y > 45)$.

Modell 2:

I modell 1 beskrevet ovenfor kjøper kunden enten en 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:

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

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 4: Sannsynlighetsfordeling $P(B = b_i)$.

Tilhørende forventning og standardavvik er:

$$E[B] = 0.11 \quad , \quad \sigma[B] = 0.4449 \quad (8)$$

(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:

$$B_{\text{tot}} = \sum_{i=1}^n B_i = B_1 + B_2 + \dots + B_n \quad (9)$$

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

⁹Dvs. 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 3 c i og 3 h, 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 kioskdirften som er relatert til brussalget er $k_{\text{fast}} = 175$ NOK. Fortjenesten F på brussalget for en gitt overfart er da gitt ved:

$$F = (p - k)B_{\text{tot}} - k_{\text{fast}} \quad (11)$$

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

¹⁰Tips: Se side 44 i formelsamlingen.

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

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

$$P(B_{\text{tot}} > 45) = 1 - P(B_{\text{tot}} \leq 45) \quad (10)$$

Deretter kan du standardisere lign.(10). Heltallskorreksjon behøves **ikke**.

¹³Er 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.

$$x = \text{omsetningsdag nr. } x, \text{ hvor } x = 1, 2, 3, 4, 5 \quad (12)$$

$$y = \text{aksjekurs (i NOK)} \quad (13)$$

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

Ut fra tabellen i figur (5) kan man regne ut den empiriske variansen for aksjekursen y og den empiriske variansen for x . Resultatet er:

$$S_y^2 = 133 \text{ NOK}^2, \quad S_x^2 = 2.5 \quad (14)$$

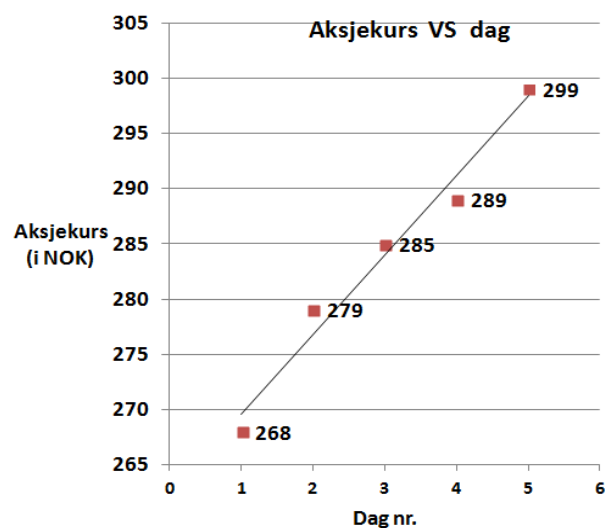
Den empiriske kovariansen mellom x og y er:

$$S_{xy} = 18 \text{ NOK} \quad (15)$$

(Størrelsene i lign.(14) og (15) trenger du å ikke regne ut. Bare ta dem for gitt).



Figur 6: Norwegian.



Figur 7: Plott av dataene fra tabellen i figur (5).

- a) Grafen i figur (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 a)?
- 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 a, altså regresjonslinjen.

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

¹⁴Dvs. finn formelen for linjen. Se gjerne formelsamlingen for formel for minste kvadraters regresjonslinje.

¹⁵Tips: 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 (8) nedenfor. (Bruk 4 desimalers nøyaktighet).
- f) Kommenter svaret i oppgave e.¹⁶

A	B	C	D	E	F	G	H
SUMMARY OUTPUT							
<i>Regression Statistics</i>							
Multiple R	0,987135295						
R Square	0,97443609						
Adjusted R Square	-1,666666667						
Standard Error	2,12916259						
Observations	1						
<i>ANOVA</i>							
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>		
Regression	5	518,4	103,68	114,3529	#NUM!		
Residual	3	13,6	4,533333				
Total	8	532					
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95,0%</i>
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 8: Utskrift fra Excel.

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

Kapittel 6

Kontinuasjonseksamen 2014

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* en time per oppgave.**

Oppgave 1: (økonomi)

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

a) Bruk notasjonen

$$G = \text{kunde har } \mathbf{god} \text{ betalingsevne} \quad (1)$$

$$D = \text{kunde har } \mathbf{dårlig} \text{ betalingsevne} \quad (2)$$

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

$$L = \text{kunde med } \mathbf{lav} \text{ inntekt} \quad (3)$$

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



Figur 1: Sparebanken Møre.

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

$$F = \text{fortjeneste på en kunde med lav inntekt} \quad (4)$$

- 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 f.



Oppgave 2: (petroleumslogistikk)

Anta at du jobber som logistikk-koordinator ved olje- og gass firmaet 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:

$$X = \text{antall "hendelser" per uke}$$

som vi altså antar er Poisson fordelt:

$$X \sim \text{Poi}[\lambda] \tag{5}$$

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

$$\lambda = 0.6 \tag{6}$$



Figur 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 :

$$Y = X_1 + X_2 + X_3 + X_4 \quad (7)$$

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

$$Y \sim \text{Poi}[E[Y]] \quad (8)$$

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

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



¹Bruk 4 desimalers nøyaktighet.

²Dette 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 en produksjonsfeil. Og det er 2 % sannsynlighet for at en brus har to produksjonsfeil. Den stokastiske variabelen ³

X = antall produksjonsfeil for en tilfeldig valgt brus

har dermed fordelingen $P(X)$ som vist i tabellen nedenfor i figur 3:

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

Figur 3: Sannsynlighetsfordeling $P(X = x)$.

- a) Vis at sannsynlighetsfordelingen i tabellen i figur 3 er en **gyldig** sannsynlighetsfordeling.
- b) Finn $E[X]$ og gi en tolkning av $E[X]$. ⁴
- c) Finn $Var[X]$ og gi en tolkning av $Var[X]$.



Figur 4: Produksjon av ananasbrus.

³Prouksjonsfeil kan være at toppen ikke er rett påskrudd og etiketten ikke er rett påsatt.

⁴Dvs. 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:

$$I = pn0.95 - bnX$$

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, \dots , n . Det totale antall produksjonsfeil per dag for ananasbrus er da:

$$Y = X_1 + X_2 + X_3 + \dots + X_n \quad (9)$$

Anta videre at:

1. antall produksjonsfeil for de ulike ananasbrusene er uavhengige:
 X_i er uavhengige for alle $i = 1, 2, 3, \dots, n$
2. alle ananasbrus antas å ha samme sannsynlighetsfordeling (gitt ved tabell i figur 3):
 X_i har samme sannsynlighetsfordeling for alle $i = 1, 2, 3, \dots, n$

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

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

⁵Dvs. finn $E[Y]$.

⁶Dvs. 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



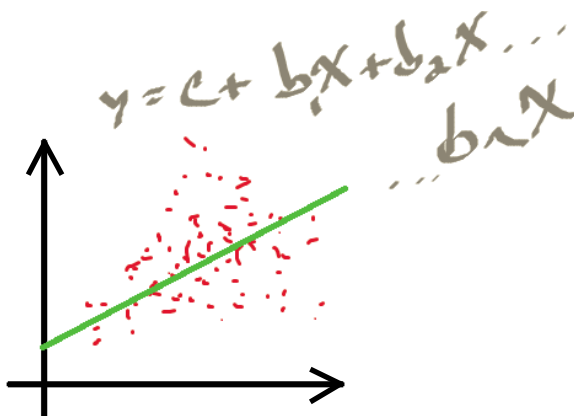
⁷Kun en **tommelfingerregel** er godt nok her.

⁸Du 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 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 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.
- Hva slags mulige verdier har R^2 ?
 - Hva slags benevning/enhet har R^2 ?
 - Hva er R^2 et mål på?



Figur 5: Regresjon.

⁹En setning er nok.

¹⁰Tips: 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 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. I denne sammenheng er det hensiktsmessig å definere følgende variabler:

$$x = \text{år nr. } x, \text{ hvor } x = 2009, 2010, 2011, 2012, 2013 \quad (10)$$

$$y = \text{etterspørsel av slurry per år (i tusen tonn)} \quad (11)$$

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: År x og etterspørsel y .

Ut fra tabellen i figur 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:

$$S_y^2 = 121.2 \quad , \quad S_x^2 = 2.5 \quad (12)$$

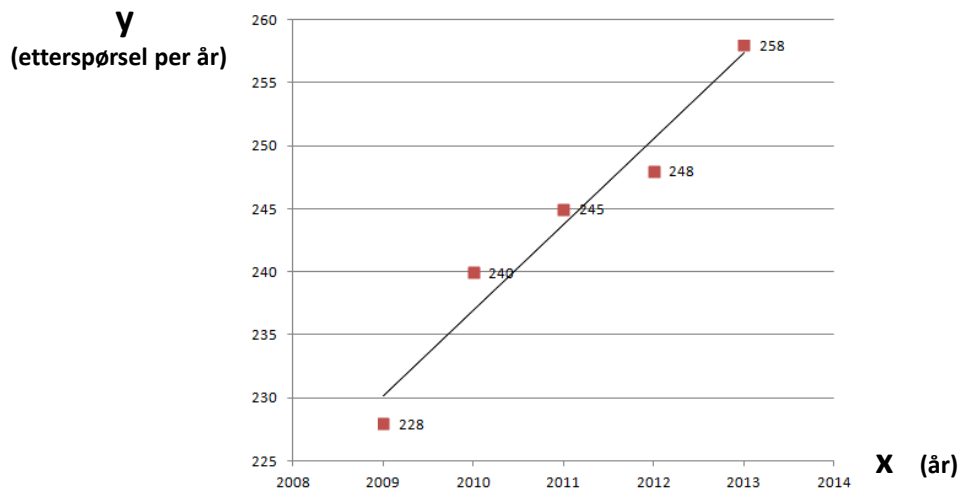
Den empiriske kovariansen mellom x og y er:

$$S_{xy} = 17 \quad (13)$$

(Størrelsene i lign.(12) og (13) trenger du å ikke regne ut. Bare ta dem for gitt. Benevnningen til disse størrelsene er utelatt).



Figur 7: Hustadmarmor. Slurry.



Figur 8: Plott av dataene fra tabellen i figur 6.

- d) Den rette linjen i figur 8 viser minste kvadraters regresjonslinje for x og y . Regn ut gjennomsnittene \bar{x} og \bar{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?

¹¹Dvs. finn formelen for linjen via setningen for minste kvadraters lineære regresjonslinje.

¹²Denne 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 uten å gjøre noe regning “for hånd”. Bare les av fra Excel-utskriften i figur 9 nedenfor. (Bruk 4 desimalers nøyaktighet).

i) Kommenter svaret i oppgave h.¹³

A	B	C	D	E	F	G	H	I
SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0,976624482							
R Square	0,95379538							
Adjusted R Square	-1,666666667							
Standard Error	2,732520204							
Observations	1							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	5	462,4	92,48	61,92857	#NUM!			
Residual	3	22,4	7,466667					
Total	8	484,8						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
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 9: Utskrift fra Excel.



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

Kapittel 7

Hovedeksamen 2015

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* en 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 1. De bestemmer seg også for å definere den stokastiske variabelen X hvor:

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

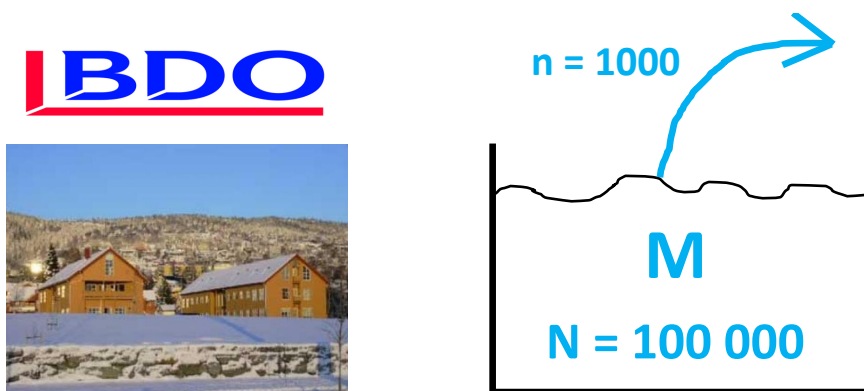
- a) Hva slags sannsynlighetsfordeling har X ? Gi en kort begrunnelse. ¹

- 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:

$$\sigma[X] = \sqrt{9.8} = 3.13 \tag{1}$$

Bruk at $M = 1000$.



Figur 1: BDO revisjon. Det trekkes $n = 1000$ bilag.

¹Se f.eks. side 71 i formelsamlingen.

d) Dersom noen betingelser er oppfylt så kan en hypergeometrisk fordeling med god tilnærming 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:

$$X \sim N[E[X], \sigma[X]] \quad , \quad (2)$$

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.(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.:

$$P(X \geq g) = 0.10 \quad (3)$$

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 $P(X \geq g) = 0.10$.

■

²Se f.eks. side 79 i kompendiet.

³Husk: 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:

$$D_1 \sim N[\mu_1 = 15\,000, \sigma_1 = 3000] \quad (4)$$

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



Figur 2: Spesielle nyheter. Salg.

⁴Dvs. 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 avis-kjøperne. Anta at også denne stokastiske variabelen er normalfordelt, men med annen forventning og annen varians:

$$D_2 \sim N[\mu_2 = 18\,000, \sigma_2 = 4000] \quad (5)$$

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

$$D_{tot} = D_1 + D_2 \quad (6)$$

- 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:

$$F = (p - k)D_{tot} - c \quad (7)$$

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 lineærkombinasjon 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:

$$D_{tot} \sim N[E[D_{tot}], \sigma[D_{tot}]] \quad , \quad (8)$$

hvor forventning 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

■

⁵Bruk det faktum at D_{tot} er normalfordelt, jfr. lign.(8).

⁶Tips: Man kan skrive $P(D_1 > 21\,000) = P(D_{tot} > 21\,000|\bar{S})$. 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 3: Sannsynlighetsfordeling $P(X = x)$.

- a) Vis at sannsynlighetsfordelingen $P(X = x)$ i tabellen i figur 3 er en **gyldig** sannsynlighetsfordeling.



Figur 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 år, altså over $n = 365$ dager.

La X_i være en stokastisk variabel som beskriver antall innleverte rapporter for dag nr. i , hvor $i = 1, 2, \dots, n$. Anta at alle de stokastiske variablene X_i har **samme sannsynlighetsfordeling** gitt ved tabellen i figur 3. Vi definerer gjennomsnittet \bar{X} :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} (X_1 + X_2 + \dots + X_n) \quad (9)$$

- d) i) Hva betyr $E[\bar{X}]$ på “godt norsk” i vårt tilfelle? ⁷
- ii) Finn $E[\bar{X}]$.

Anta at antall innleverte rapporter per dag er **uavhengige**.

- e) i) Hva betyr $Var[\bar{X}]$ på “godt norsk” i vårt tilfelle?
- ii) Finn $Var[\bar{X}]$.
- f) **Hvilken fordeling** (tilnærmet) er det rimelig å anta at \bar{X} har? Begrunn svaret.

⁷Dvs. gi en tolkning av $E[\bar{X}]$.

- g) Sammenlign $E[X]$ med $E[\bar{X}]$ og $Var[X]$ med $Var[\bar{X}]$ 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

■

⁸Tips:

$$P(X_1 + X_2 + \dots + X_n > 700) = P\left(\frac{X_1 + X_2 + \dots + X_n}{n} > \frac{700}{n}\right) \quad (10)$$

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

Oppgave 4: (økonomi)

I kapittel 1 i kurset “*MAT110 Statistikk 1*” så vi på statistiske størrelser med både én og to variabler. For statistiske størrelser med to variabler så vi blant annet på graden av [samvariasjon](#) mellom variablene.

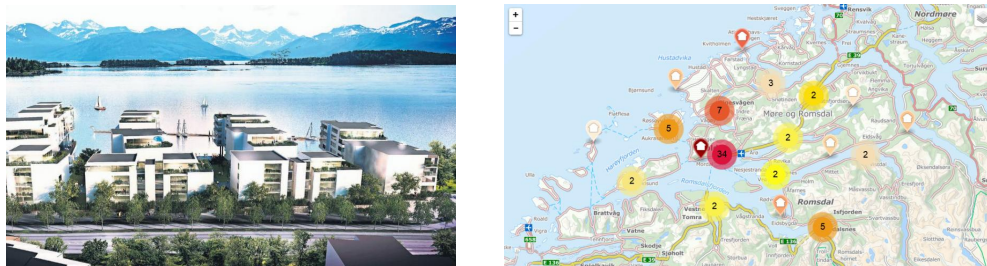
- a) For observasjonene x_1, x_2, \dots, x_n har vi tilhørende observasjoner y_1, y_2, \dots, y_n . I denne sammenheng kan man snakke om korrelasjonskoeffisienten R_{xy} til disse observasjonene.
- i) Hva slags mulige verdier kan R_{xy} ha?
 - ii) Hva er R_{xy} et mål på?
 - iii) Hva slags enhet har R_{xy} ?

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

$$x = \text{avstand til sentrum (km)} \quad (11)$$

$$y = \text{kvadratmeterpris (1000 NOK per } m^2 \text{)} \quad (12)$$

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 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 6: Avstand fra sentrum x og gjennomsnittlig kvadratmeterpris y .

Ut fra tabellen i figur 6 kan man regne ut gjennomsnittene \bar{x} og \bar{y} :

$$\bar{x} = 15 \quad , \quad \bar{y} = 40.43 \quad (13)$$

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:

$$S_x^2 = 116.67 \quad , \quad S_y^2 = 109.54 \quad (14)$$

Den empiriske **k**ovariansen mellom x og y er:

$$S_{xy} = -110.83 \quad (15)$$

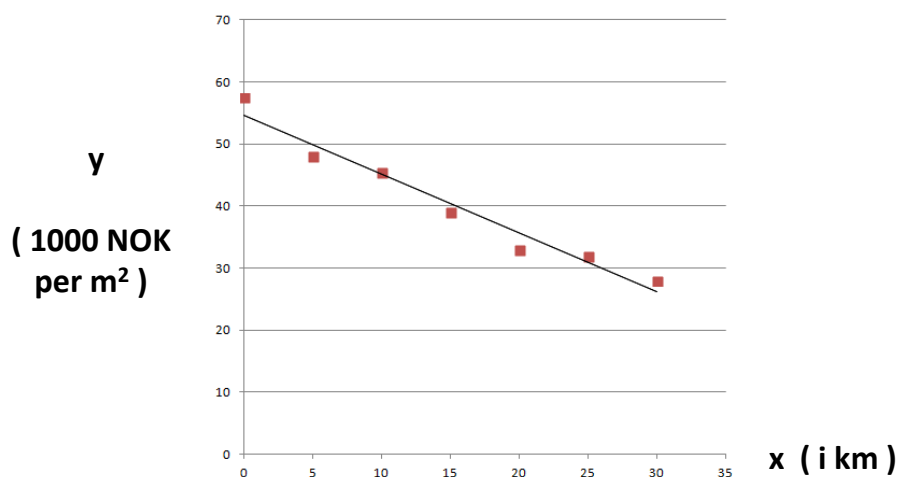
(Størrelsene i lign.(13), (14) og (15) trenger du å ikke regne ut. Bare ta dem for gitt. Benevningen, altså enheten, til disse størrelsene er utelatt).

b) Regn ut korrelasjonskoeffisienten R_{xy} for observasjonene i tabellen i figur 6.

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

⁹Hva 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 sammenheng mellom x og y . Eiendomsmegleren bestemmer seg for å bruke lineær regresjonsanalyse.



Figur 7: Plott av dataene fra tabellen i figur 6.

- d) Den rette linjen i figur 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/m².

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

¹⁰Husk 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 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	<i>Regression Statistics</i>								
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		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
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		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
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 8: Utskrift fra Excel.

¹¹For 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?

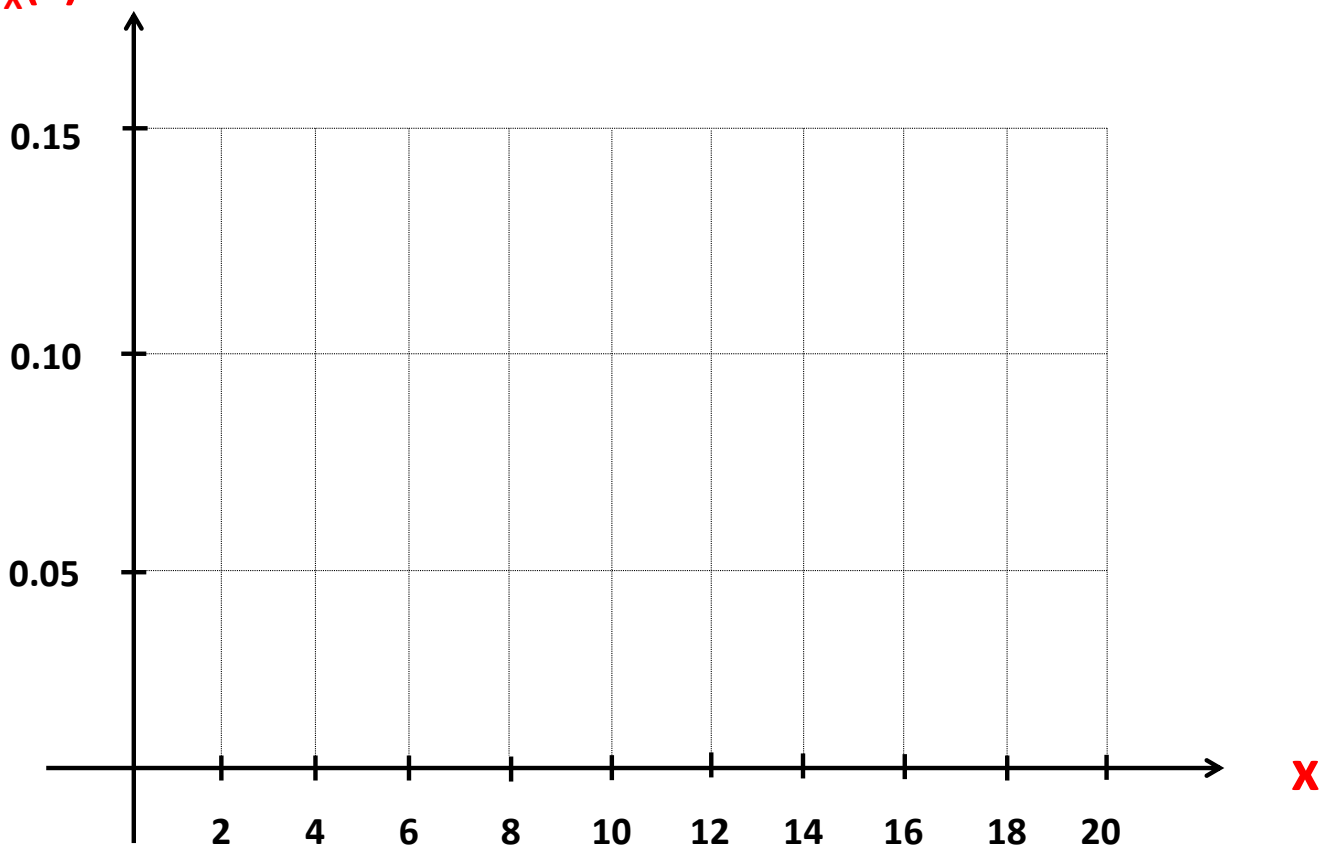
Vedlegg



Høgskolen i Molde
Vitenskapelig høgskole i logistikk

Studentnummer: _____

$f_x(x)$



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

Kapittel 8

Kontinuasjonseksamen 2015

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* en 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:

$$P_1 = \text{sannsynligheten for at en liten lastebil ankommer varemottaket} \quad (1)$$

$$P_2 = \text{sannsynligheten for at en mellomstor lastebil ankommer varemottaket} \quad (2)$$

$$P_3 = \text{sannsynligheten for at en stor lastebil ankommer varemottaket} \quad (3)$$

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 1: Vogntog og varemottak.

La oss definere den stokastiske variabelen X , hvor:

$$X = \text{antall minutter behandlingstid for å laste av lasten på en tilfeldig valgt lastebil} \quad (4)$$

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

$$\text{laste av en liten lastebil:} \quad 10 \text{ minutter} \quad (5)$$

$$\text{laste av en mellomstor lastebil:} \quad 20 \text{ minutter} \quad (6)$$

$$\text{laste av en stor lastebil:} \quad 30 \text{ minutter} \quad (7)$$

Anta at X har følgende sannsynlighetsfordeling:

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

Figur 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 $\sigma[X]$?

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

²Bruk 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:

$$V = X_1 + X_2 + X_3 \quad (8)$$

- e) Hva er forventet 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:

$$p(x_1, x_2, x_3) \equiv P(X_1 = x_1 \text{ og } X_2 = x_2 \text{ og } X_3 = x_3) \quad (9)$$

- g) Finn den simultane sannsynligheten $p(30, 30, 30)$.³
- h) Gi en *kort* tolkning av sannsynligheten $p(30, 30, 30)$.

³Husk at lastebilene ankommer varemottaket uavhengige av hverandre.

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

$$P(V \leq 80) + P(V = 90) = 1 \quad (10)$$

- i) Bruk resultatet fra oppgave **1g** samt lign.(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:

$$\begin{aligned} P(V = 50) &= p(10, 10, 30) + p(10, 30, 10) + p(30, 10, 10) \\ &+ p(20, 20, 10) + p(20, 10, 20) + p(10, 20, 20) \end{aligned} \quad (11)$$

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

■



Figur 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:

$$H = \text{medlem som handler aksjer} \tag{12}$$

$$\bar{H} = \text{medlem som IKKE handler aksjer} \tag{13}$$

$$R = \text{medlem som leser næringslivsaviser regelmessig} \tag{14}$$

$$I = \text{medlem som leser næringslivsaviser iblant} \tag{15}$$

$$A = \text{medlem som aldri leser næringslivsaviser} \tag{16}$$

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

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

Figur 4: Sannsynligheter.



Figur 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.(12)-(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 $P(H|\bar{R}) = P(H|(A \cup I))$ for at et medlem av *Econa* som ikke regelmessig leser næringslivsaviser, har handlet aksjer på aksjemarkedet det siste året. ⁶



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

⁵En setning er nok.

⁶Denne 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 variabler:

$$X_E = \text{antall akutte utrykninger per uke i Eide} \quad (17)$$

$$X_F = \text{antall akutte utrykninger per uke i Fræna} \quad (18)$$

$$X_G = \text{antall akutte utrykninger per uke i Gjemnes} \quad (19)$$

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. **Poissonfordelingen**:

$$X_i \sim \text{Poi}[\lambda_i] \quad (20)$$

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

$$\lambda_E = 1.5 \quad (21)$$

$$\lambda_F = 2.3 \quad (22)$$

$$\lambda_G = 1.9 \quad (23)$$

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



Figur 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:

$$Y = X_E + X_F + X_G \quad (24)$$

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 $E[Y]$:

$$Y \sim \text{Poi}[E[Y]] \quad (25)$$

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

⁷Finn $P(X_G = 2)$. Se gjerne formelsamling angående formel for Poissonfordeling.

⁸Finn $E[Y]$. Bruk gjerne en av regnereglene på side 40 i formelsamlingen.

⁹Dette 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, $i = 1, 2, 3, \dots, 52$, og slik at:

$$X_i = \text{antall akutte utrykninger i Gjemnes i uke nr. } i \quad (26)$$

Antall akutte utrykninger i året i Gjemnes kommune er da:

$$X_{\text{år}} = X_1 + X_2 + X_3 + \dots + X_n \quad (27)$$

hvor $n = 52$. Anta videre at:

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

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. $Var[X_{\text{år}}]$?

¹⁰Er det rimelig at det ene svaret er 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 **4hi** 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



¹¹Kun en **tommelfingerregel** er godt nok her.

¹²Du behøver ikke å bruke heltallskorreksjon.

Oppgave 4: (logistikk)

I kapittel 1 i kurset “*MAT110 Statistikk 1*” så vi på statistiske størrelser med både én 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, \dots, x_n har vi tilhørende observasjoner y_1, y_2, \dots, y_n . I denne sammenheng kan man snakke om korrelasjonskoeffisienten R_{xy} til disse observasjonene.
- 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.



Figur 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 :

$$x_i = \text{kvartal nr. } i \quad (i = 1..8) \quad (28)$$

$$y_i = \text{antall serviceordrer mottatt i kvartal nr. } i \quad (i = 1..8) \quad (29)$$

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: Kvartal og serviceordrer.

Ut fra tabellen i figur 8 kan man regne ut gjennomsnittene \bar{x} og \bar{y} :

$$\bar{x} = 4.5 \quad , \quad \bar{y} = 33.625 \quad (30)$$

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:

$$S_x^2 = 6 \quad , \quad S_y^2 = 19.7 \quad (31)$$

Den empiriske **k**ovariansen mellom x og y er:

$$S_{xy} = 10.36 \quad (32)$$

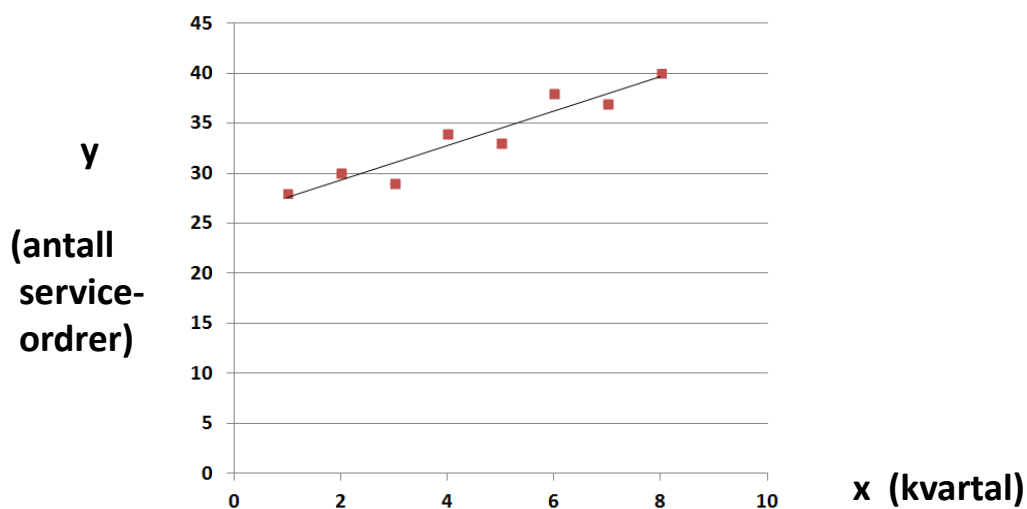
(Størrelsene i lign.(30), (31) og (32) trenger du ikke å regne ut. Bare ta dem for gitt).

b) Regn ut korrelasjonskoeffisienten R_{xy} for observasjonene i tabellen i figur 8.

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

¹³Hva 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 sammenheng mellom x og y . Logistikerne bestemmer seg for å bruke lineær regresjonsanalyse.



Figur 9: Plott av dataene fra tabellen i figur 8.

- d) Den rette linjen i figur 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?

¹⁴Her er det meningen at man skal regne seg frem til svaret. Man skal **ikke** bare lese av fra figur 9.

Ut fra tabellen i figur 8 kan man regne ut SSE , “*sum squared error*”, og SST , “*sum square total*”. Resultatene er:

$$SSE = 12.73 \quad , \quad SST = 137.88 \quad (33)$$

(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



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

Kapittel 9

Hovedeksamen 2016

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* én 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

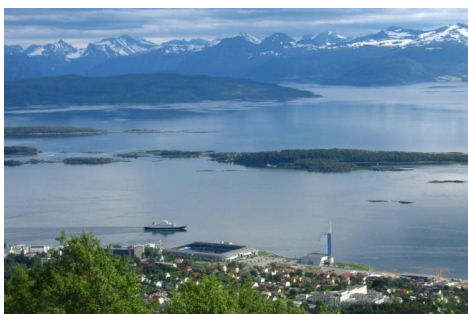
La oss definere følgende begivenheter:

$$L_1 \stackrel{\text{def.}}{=} \text{ledig plass på ferge 1} \quad (1)$$

$$L_2 \stackrel{\text{def.}}{=} \text{ledig plass på ferge 2} \quad (2)$$

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:

$$P(\bar{L}_1) = 0.15 \quad (3)$$



Figur 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 én** 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. $P(L_1 \cap \bar{L}_2)$? ²
- d) Hva er sannsynligheten for at begge fergene er fulle, dvs. $P(\bar{L}_1 \cap \bar{L}_2)$? ³

Dersom Bring ikke får plass på en ferge slik at de må vente til neste så får de ekstraavgifter 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 = utgift for Bring på fergestrekningene mellom Molde og Volda

Utfallsrommet til U er:

$$\Omega = \{ 0, 800, 1200, 2000 \} \quad (4)$$

¹Tips: bruk den generelle addisjonssetningen.

²Tips: bruk setningen om total sannsynlighet.

³Tips: bruk en av tvillingsetningene.

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

$$E[U] = \sum_{i=1}^4 u_i \cdot P(U = u_i) \quad (5)$$

$$= \left(0 \cdot P(L_1 \cap L_2) + 1200 \cdot \overbrace{P(\bar{L}_1 \cap L_2)}^{= 0.05} \right. \\ \left. + 800 \cdot P(L_1 \cap \bar{L}_2) + 2000 \cdot P(\bar{L}_1 \cap \bar{L}_2) \right) \text{ NOK} \quad (6)$$

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

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

■

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

$$D = \text{antall aviser som etterspørres en gitt dag} \quad (7)$$

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

$$D \sim \text{Poi}[\lambda] . \quad (8)$$

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 2: Avisgutt.

⁵Se 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

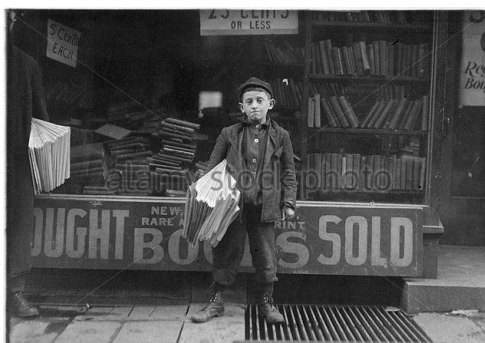
$$S = \text{antall aviser som faktisk selges en gitt dag} \quad (9)$$

Denne variabelen er gitt ved

$$S = \min(D, q), \quad (10)$$

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 3: Avisgutten bestiller $q = 85$ aviser.

⁶Se f.eks. side 75 i formelsamlingen.

⁷Dvs. finn $P(D > 80)$.

Anta at sannsynlighetsfordelingen til den stokastiske variabelen S er gitt ved:

$$P(S = s) = \begin{cases} P(D = d) & , s \leq 84 \\ 0.3429 & , s = 85 \\ 0 & , s \geq 86 \end{cases} \quad (11)$$

hvor $P(D = d)$ er Poissonfordelt, dvs. $D \sim \text{Poi}[\lambda]$, jfr. lign.(8).

e) Forventningen $E[S]$ er gitt ved

$$E[S] = \sum_{s=1}^{84} sP(S = s) + 85 \cdot P(S = 85) \quad (12)$$

hvor

$$\sum_{s=1}^{84} sP(S = s) = 49.89 \quad (13)$$

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.

⁸Altså du skal bare finne tallverdien for $E[S]$. Du skal ikke bevise lign.(12) og (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:

$$\pi(q) = r S - w q , \quad (14)$$

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

$$w = 5 \text{ NOK} \quad (\text{innkjøpspris}) \quad (15)$$

$$r = 20 \text{ NOK} \quad (\text{utsalgpris}) \quad (16)$$

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

$$P(D \leq q^*) = 1 - \frac{w}{r} , \quad (17)$$

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.(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.

$$D \sim N[\mu = 81, \sigma = 9] . \quad (18)$$

⁹Du skal ikke vise lign.(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.(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* 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 = antall trafikkulykker på innfartsvegen til Molde **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 4: Sannsynlighetsfordeling $P(X = x)$.

- a) Er den stokastiske variabelen X diskret eller kontinuerlig?
Gi en *kort* forklaring.¹⁰



Figur 5: Innfartsvegen til Molde.

¹⁰En setning er nok.

- b) Fra tabellen i figur 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 $X_1, X_2, X_3, \dots, X_n$ 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:

$$Y = X_1 + X_2 + X_3 + \dots + X_n \quad (19)$$

Anta videre at:

1. antall ulykkene er uavhengige slik at:
 X_i er uavhengige for alle $i = 1, 2, 3, \dots, n$
2. hvert år har samme sannsynlighetsfordeling (gitt ved tabell i figur 4):
 X_i har samme sannsynlighetsfordeling for alle $i = 1, 2, 3, \dots, n$

- 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 en trafikkulykke på den aktuelle vegstrekningen over en periode på to år? ¹¹

¹¹Dvs. finn $P(Y \leq 1)$. Bruk formelen:

$$\begin{aligned}
 P(Y \leq 1) &= \overbrace{P(X_1 = 0)}^{0 \text{ ulykker år nr.1}} \cdot \overbrace{P(X_2 = 0)}^{0 \text{ ulykker år nr.2}} + \overbrace{P(X_1 = 1)}^{1 \text{ ulykke år nr.1}} \cdot \overbrace{P(X_2 = 0)}^{0 \text{ ulykker år nr.2}} \\
 &+ \overbrace{P(X_1 = 0)}^{0 \text{ ulykker år nr.1}} \cdot \overbrace{P(X_2 = 1)}^{1 \text{ ulykke år nr.2}}
 \end{aligned} \quad (20)$$

hvor $P(X_1 = 0) = P(X_2 = 0) = 0.25$ osv., se tabellen i figur 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 tilnærming?

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. $P(Y > 50)$?

■

¹²Kun 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, \dots, x_n av en variabel x med tilhørende observasjoner y_1, y_2, \dots, 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 6: Samvariasjon.

¹³Et 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 = pris nr. i ($i = 1, 2, \dots, 5$) (21)

y_i = antall solgte biler per dag på global basis tilhørende pris nr. i ($i = 1, 2, \dots, 5$) (22)



Figur 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 8: Pris x_i og salg y_i .

Ut fra tabellen i figur 8 kan man regne ut gjennomsnittene \bar{x} og \bar{y} :

$$\bar{x} = 3.5 \quad , \quad \bar{y} = 0.73 \quad (23)$$

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:

$$S_x^2 = 0.15625 \quad , \quad S_y^2 = 0.15325 \quad (24)$$

Den empiriske **kovariansen** mellom x og y er:

$$S_{xy} = -0.1531 \quad (25)$$

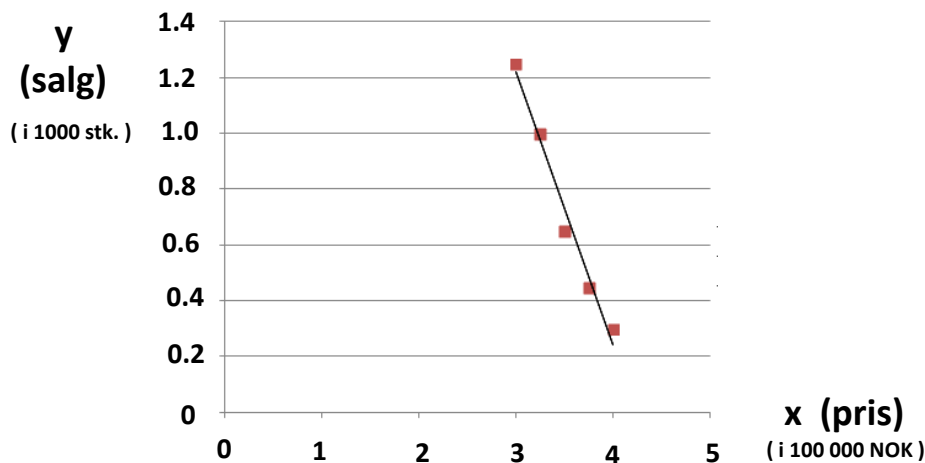
(Størrelsene i lign.(23), (24) og (25) trenger du ikke å regne ut. Bare ta dem for gitt).

c) Regn ut korrelasjonskoeffisienten R_{xy} for observasjonene i tabellen i figur 8.

d) Tolk svaret du fikk for R_{xy} i oppgave 4c.¹⁴

¹⁴Hva 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 sammenheng mellom x og y . Dere bestemmer dere for å bruke lineær regresjonsanalyse.



Figur 9: Plott av dataene fra tabellen i figur 8.

- e) Den rette linjen i figur 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 $\hat{y} = 1$?

¹⁵Her er det meningen at man skal regne seg frem til svaret. Man skal **ikke** bare lese av fra figur 9.

Ut fra tabellen i figur 8 kan man regne ut SSE , “*sum squared error*”, og SST , “*sum square total*”. Resultatene er:

$$SSE = 0.01275 \quad , \quad SST = 0.613 \quad (26)$$

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

h) Finn forklaringskraften R^2 .

i) Kommenter svaret i oppgave [4h](#).¹⁶

■

¹⁶Vi du si at regresjonslinjen predikerer antall solgte biler per dag, i stor eller liten grad? Med stort eller lite presisjonsnivå?

Kapittel 10

Kontinuasjoneksamen 2016

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 innebærer en del blandingsprosesser samt justeringer av fordelingen mellom pulver og andre stoffer.

Før malingen kan legges ut for salg må den tilfredsstillende 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 en 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:

$$D_t \stackrel{\text{def.}}{=} \text{malingen er fortsatt "defekt" etter } t \text{ timer med blanding og justeringer} \quad (1)$$

hvor $t = 1, 2, 3, 4$.



Figur 1: Jotun.

Anta i denne oppgaven at:

$$P(D_1) = 0.30 \quad (2)$$

$$P(D_2|D_1) = 0.10 \quad (3)$$

$$P(D_3|D_2 \cap D_1) = 0.02 \quad (4)$$

$$P(D_4|D_3 \cap D_2 \cap D_1) = 0 \quad (5)$$

- 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:

$$\boxed{P(\overline{D}_2 \cap D_1) = 0.27} \quad (6)$$

¹Bruk komplementsetningen.

²Husk at komplementsetningen for en betinget sannsynlighet er: $P(\overline{B}|A) = 1 - P(B|A)$.

Sannsynligheten $P(\bar{D}_3 \cap D_2 \cap D_1)$, dvs. sannsynligheten for at malingen er OK etter 3 timer, kan finnes ved å bruke sannsynlighetstreet i figur 2.

Dersom man følger den rosa, stiplede linjen så innser man at:

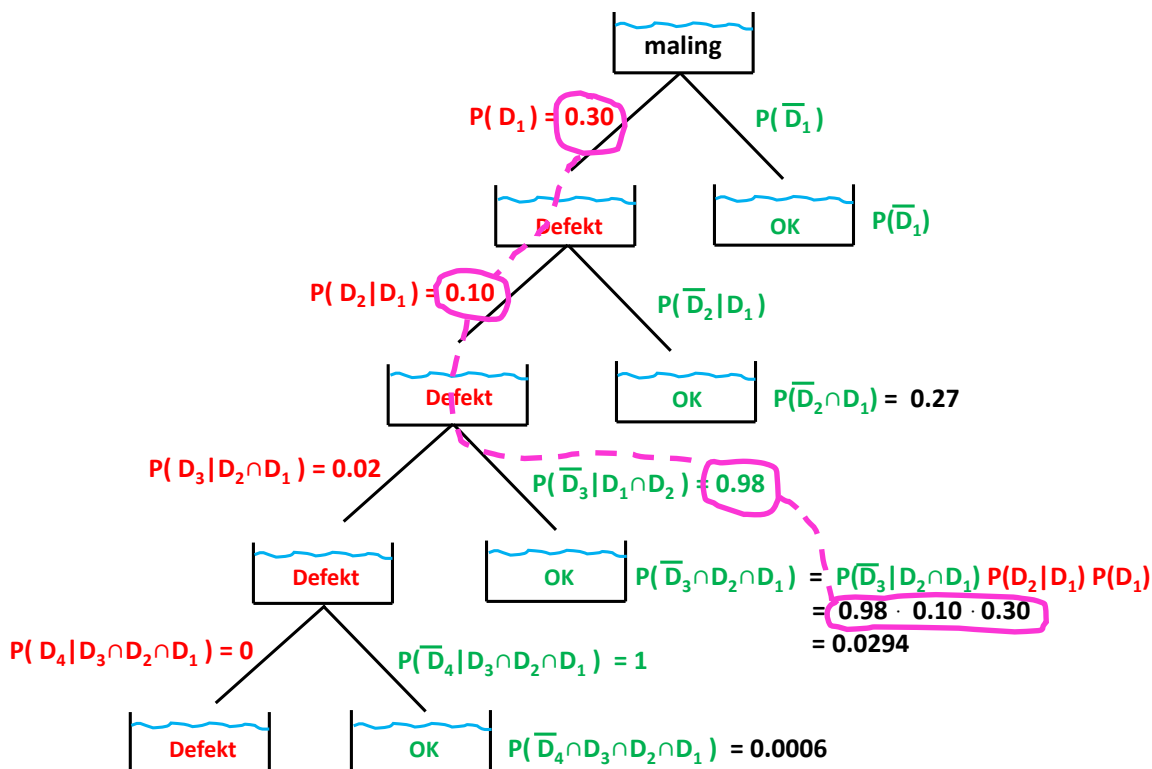
$$P(\bar{D}_3 \cap D_2 \cap D_1) = P(\bar{D}_3|D_2 \cap D_1)P(D_2|D_1)P(D_1) \quad (7)$$

$$= 0.98 \cdot 0.10 \cdot 0.30 = \underline{0.0294} \quad (8)$$

e) Bruk sannsynlighetstreet i figur 2 og vis, slik som illustrert i lign.(7) og (8), at:

$$P(\bar{D}_4 \cap D_3 \cap D_2 \cap D_1) = 0.0006. \quad (9)$$

hvor $P(\bar{D}_4 \cap D_3 \cap D_2 \cap D_1)$ er sannsynligheten for at malingen er OK etter 4 timer.



Figur 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 :

$$X = \text{antall timer som Jotun bruker på blanding og justeringer av malingen før malingen kommer innenfor blandings- og kvalitetskravene} \quad (10)$$

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:

$$E[X] = \sum_{i=1}^4 x_i \cdot P(X = x_i) \quad (11)$$

$$\begin{aligned} &= 1 \cdot P(\overline{D}_1) + 2 \cdot P(\overline{D}_2 \cap D_1) \\ &+ 3 \cdot P(\overline{D}_3 \cap D_2 \cap D_1) + 4 \cdot P(\overline{D}_4 \cap D_3 \cap D_2 \cap D_1) \end{aligned} \quad (12)$$

f) Bruk resultatene fra tidligere deloppgaver samt lign.(12) til å finne $E[X]$.

g) Vis at $P(X = x_i)$ er en gyldig sannsynlighetsfordeling.³

■

³Tips: vis at de 4 sannsynlighetene i lign.(12) summeres til 1.

Oppgave 2: (økonomi , aksjer)

Anta at du jobber i investeringselskapet Gjelsten Holding som skal investere i tre ulike selskaper:

Statoil , Seadrill og Yara .
selskap *A* selskap *B* 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

$$N = 100\,000 \tag{13}$$

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*:

$$N \cdot a = \text{antall aksjer som investeres i selskap } A, \text{ Statoil} \tag{14}$$

$$N \cdot b = \text{antall aksjer som investeres i selskap } B, \text{ Seadrill} \tag{15}$$

$$N \cdot c = \text{antall aksjer som investeres i selskap } C, \text{ Yara} \tag{16}$$

hvor

$$a + b + c = 1 \tag{17}$$

dvs. *a* er den brøkdelen av de $N = 100\,000$ aksjene som investeres i selskap *A*. Tilsvarende for *b* og *c*.



Figur 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 = pris på en aksje (altså aksjekurs) for selskap A , Statoil (i NOK), **om ett år** (18)

Y = pris på en aksje (altså aksjekurs) for selskap B , Seadrill (i NOK), **om ett år** (19)

Z = pris på en aksje (altså aksjekurs) for selskap C , Yara (i NOK), **om ett år** (20)

Anta videre at forventet pris (altså aksjekurs) **om ett år** er: ⁴

$$E[X] = 90 \quad , \quad E[Y] = 125 \quad , \quad E[Z] = 180 \quad (21)$$

med tilhørende varianser: ⁵

$$Var[X] = 100 \quad , \quad Var[Y] = 200 \quad , \quad Var[Z] = 600 \quad (22)$$



Figur 4: Gjelsten Holding.

⁴Forventningene er i NOK, men dropper benevningen her for enkelhets skyld.

⁵Variansene er i NOK².

- a) Vis at fortjenesten F om ett år er gitt ved: ⁶

$$F = Na(X - 100) + Nb(Y - 105) + Nc(Z - 130) \quad (24)$$

- b) Vis at den forventede fortjenesten $E[F]$ ved et eventuelt salg av aksjene om ett år er: ⁷

$$E[F] = 10N(-a + 2b + 5c) \quad (25)$$

- c) Anta nå at aksjene er uavhengige.
Bruk lign.(24) samt lign.(22) og vis at variansen $Var[F]$ til fortjenesten ved et eventuelt salg om ett år er:

$$Var[F] = 100N^2(a^2 + 2b^2 + 6c^2) \quad (26)$$

- 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.(26) er minst mulig når

$$a = 0.6 \quad , \quad b = 0.3 \quad , \quad c = 0.1 \quad (27)$$

(Du skal ikke vise lign.(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

⁶Bruk gjerne at:

$$F = \text{antall aksjer} \cdot \text{pris per aksje om ett år} - \text{antall aksjer} \cdot \text{pris per aksje i dag} \quad (23)$$

for hver av de tre selskapene. Bruk også lign.(14)-(16) samt lign.(18)-(20).

⁷Bruk lign.(24).

⁸Bruk lign.(14)-(16) sammen med lign.(27).

- e) Dersom du bestemmer deg for å maksimere den forventede fortjenesten $E[F]$ i lign.(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:

$$X = 90 + 10\epsilon \quad , \quad Y = 125 + 10\sqrt{2}\epsilon \quad , \quad Z = 180 + 10\sqrt{6}\epsilon \quad (28)$$

hvor ϵ er en stokastisk variabel som er standard normalfordelt, dvs.:

$$\epsilon \sim N\left(E[\epsilon], \sigma[\epsilon]\right) \quad (29)$$

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.(21) og (22)
Gi en *kort* kommentar.

i) Siden den stokastiske variabelen ϵ er normalfordelt så er også X normalfordelt, dvs.:

$$X \sim N\left(E[X], \sigma[X]\right) \quad (30)$$

Hva er sannsynligheten for at aksjekursen for selskap A ligger mellom 80 NOK og 100 NOK om ett år, dvs. hva er $P(E[X] - \sigma[X] \leq X \leq E[X] + \sigma[X])$?⁹

⁹Her behøves ingen utregninger. Se formelsamlingen.

Ved å sette lign.(28) inn i uttrykket for fortjenesten F i lign.(24) så får man:

$$F = 10N(a + \sqrt{2}b + \sqrt{6}c)\epsilon + k \quad (31)$$

hvor k er en konstant. (Lign.(31) skal du ikke vise. Bare ta den for gitt.)

j) Bruk lign.(31) og vis at variansen til fortjenesten F om ett år er

$$\boxed{Var[F] = (10N)^2(a + \sqrt{2}b + \sqrt{6}c)^2} \quad (32)$$

nå når de stokastiske variablene X , Y og Z er avhengige.

k) Man innser at variansen $Var[F]$ i lign.(32) er minimal dersom:

$$a = 1 \quad , \quad b = 0 \quad , \quad c = 0 \quad (33)$$

(Du skal ikke vise lign.(33). Bare ta dette faktum for gitt.)

Sammenlign resultatet i $\overbrace{\text{lign.(33)}}^{\text{avh. aksjer}}$ med $\overbrace{\text{lign.(27)}}^{\text{uavh. aksjer}}$.
Gi en *kort* kommentar.¹⁰

■

¹⁰Dersom 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.(33)? Og når de er uavhengige, jfr. lign.(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 variabler:

$$X_i = \text{antall akutte utrykninger per uke i kommune nr. } i \quad (34)$$

hvor indeksen $i = 1, 2, \dots, 5$ betegner de 5 kommunene som vist i tabellen i figur 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. **Poissonfordelingen**, beskriver fordelingen til variablene X_i :

$$X_i \sim \text{Poi}[\lambda_i] \quad (35)$$

hvor $i = 1, 2, \dots, 5$.



Figur 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 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:

$$Y = X_1 + X_2 + X_3 + X_4 + X_5 \quad (36)$$

hvor $Y = \text{summen}$ av antall akutte utrykninger i de 5 aktuelle kommunene per uke.

- a) Finn forventingen $E[Y]$.¹¹
- b) Finn standardavviket $\sigma[Y]$.

¹¹Bruk 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 $E[Y]$:

$$Y \sim \text{Poi}[E[Y]] \quad (37)$$

hvor $E[Y]$ er svaret fra **3a**.

- c) Begrunn *kort* hvorfor Y med god tilnærming 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

¹²Dette skal du ikke vise. Bare ta det for gitt.

¹³Tips: se f.eks. formelsamlingen på side 75.

¹⁴Dvs. 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, $j = 1, 2, 3, \dots, 52$, og slik at:

$$Z_j = \text{antall akutte utrykninger i Eide kommune i uke nr. } j \quad (38)$$

Antall akutte utrykninger i året i Eide kommune er da:

$$Z_{\text{år}} = Z_1 + Z_2 + Z_3 + \dots + Z_n \quad (39)$$

hvor $n = 52$. Anta videre at:

1. antall akutte utrykninger for de forskjellige ukene i Eide er uavhengige:
 Z_j er uavhengige for alle $j = 1, 2, 3, \dots, n$
2. alle Z_j er Poissonfordelte med samme rate $\lambda_j = 2.43$ for alle $j = 1, 2, 3, \dots, n$

- 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. $Var[Z_{\text{år}}]$?

¹⁵Se 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 **3hi** 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 $P(Z_{\text{år}} < 115)$?



¹⁶Kun 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 :

$$x_i = \text{antall innbyggere i kommune nr. } i \quad (40)$$

$$y_i = \text{gj.snittlig antall utrykninger per uke i kommune nr. } i \quad (41)$$

hvor, som i forrige oppgave, indeksen $i = 1, 2, \dots, 5$ betegner de 5 kommunene, se figur 7.

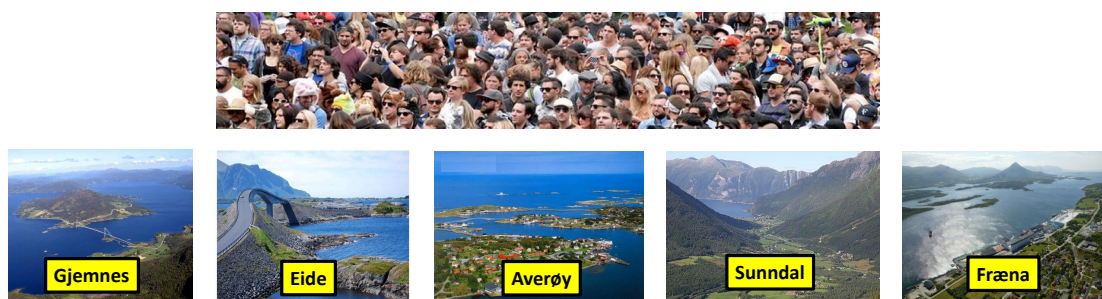
Logistikerne hos Helse M&R bruker fortsatt dataene λ_i fra tabellen i figur 6 fra [oppgave 3](#), dvs. $y_i = \lambda_i$. Tilhørende x_i for de aktuelle kommunene er samlet i denne tabellen:

	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 7: Observasjoner.

Ut fra tabellen i figur 7 kan man regne ut gjennomsnittene \bar{x} og \bar{y} :

$$\bar{x} = 5745.4 \quad , \quad \bar{y} = 3.984 \quad (42)$$



Figur 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:

$$S_x^2 = 8\,284\,549 \quad , \quad S_y^2 = 5.76828 \quad (43)$$

Den empiriske **kovariansen** mellom x og y er:

$$S_{xy} = 6863.26 \quad (44)$$

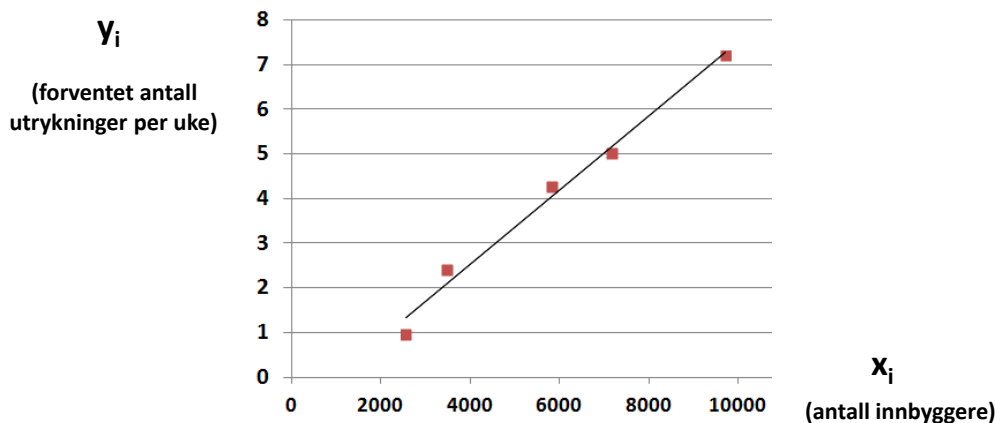
(Størrelsene i lign.(42), (43) og (44) trenger du ikke å regne ut. Bare ta dem for gitt).

a) Regn ut korrelasjonskoeffisienten R_{xy} for observasjonene i tabellen i figur 7.

b) Tolk svaret du fikk for R_{xy} i oppgave [4a](#).¹⁷

¹⁷Hva sier den numeriske (tallverdien) av R_{xy} fra oppgave [4a](#) om graden av korrelasjon mellom observasjonene x og y ?

Ut fra observasjonene og resultatene foran så ønsker Helse M&R å finne en eksplisitt sammenheng mellom x og y . De bestemmer seg for å bruke lineær regresjonsanalyse.



Figur 9: Plott av dataene fra tabellen i figur 7.

- c) Den rette linjen i figur 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 en ekstra person i telefonberedskap.

Hvor mange innbyggere må en kommune minst ha, ifølge regresjonslinjen fra oppgave 4c, dersom Helse M&R må ha en ekstra person i telefonberedskap? ¹⁹



¹⁸Her er det meningen at man skal regne seg frem til svaret. Man skal **ikke** bare lese av fra figur 9.

¹⁹Bruk ligningen for regresjonslinjen til å finne x når $\hat{y} = 6$. Man skal **ikke** bare lese av fra figur 9.

Kapittel 11

Hovedeksamen 2017



Eksamen i

MAT110 Statistikk 1

Eksamensdag	: Onsdag 24. mai 2017
Tid	: 09:00 – 13:00 (4 timer)
Faglærer/telefonnummer	: Molde + Kristiansund + Hardanger: Per Kristian Rekdal / 924 97 051
Hjelpemidler	: KD + formelsamling + generell ordbok (morsmål/norsk/engelsk i papirformat)
Antall sider inkl. forsiden	: 17
Målform	: Norsk (bokmål)

Noen generelle råd:

- **Skriv rett inn. Ikke bruk så mye tid på kladding.**
- **Det er totalt 4 oppgaver.**

Oppgave 1: (logistikk)

Logistikk- og budfirmaet JetPak leverer blant annet pakker fra dør til dør.

De ansatte i firmaet ønsker å se nærmere på pakker som blir skadet eller forsinket. Derfor definerer de følgende begivenheter:

$$S = \text{pakke som blir skadet under transport} \quad (1)$$

$$F = \text{pakke som blir forsinket ved transport} \quad (2)$$

Det er da totalt 4 forskjellige kombinasjoner av begivenheter som pakkene kan deles inn i: $\bar{S} \cap \bar{F}$, $\bar{S} \cap F$, $S \cap \bar{F}$, $S \cap F$

Anta at JetPak leverer $N = 100$ pakker i løpet av en dag.

Resultatet av disse leveransene er vist i figur 1.

I denne oppgaven skal vi bruke dette som grunnlag for beregningene.

begivenhet	$\bar{S} \cap \bar{F}$	$\bar{S} \cap F$	$S \cap \bar{F}$	$S \cap F$
antall pakker	70	15	10	5

Figur 1: Pakker.

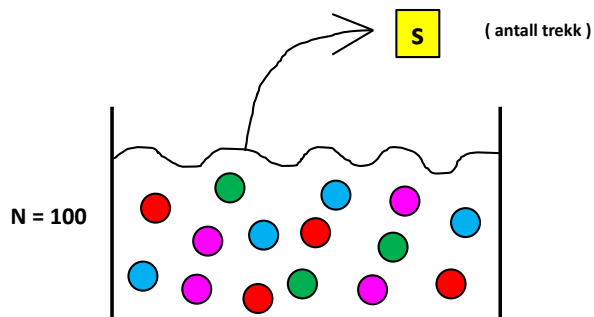
Anta at alle begivenhetene S og F er uavhengige av hverandre.



Figur 2: JetPak.

La oss beskrive problemet til JetPak med en **urne**.

Anta anta at vi trekker tilfeldige pakker urnen og at det er lik sannsynlighet for å trekke de forskjellige pakkene.



Figur 3: Urne.

- a) Hvilke to betingelser må være oppfylt for at urnemodellen skal gjelde? ¹
Gjelder urnemodellen for for vårt tilfelle med pakker?

¹Se gjerne formelsamlingen.

Ett trekk ($s = 1$)

La oss først se på problemet når vi kun trekker – altså leverer – en tilfeldig valgt pakke av de $N = 100$ pakkene.

- b) Vis – ved svært korte utregninger basert på tallene i figur 1 – at vi har følgende sannsynligheter:

$$P(\bar{S} \cap \bar{F}) = 0.70 \quad (3)$$

$$P(\bar{S} \cap F) = 0.15 \quad (4)$$

$$P(S \cap \bar{F}) = 0.10 \quad (5)$$

$$P(S \cap F) = 0.05 \quad (6)$$

JetPak taper penger på pakker som blir skadet eller forsinket.

For å se nærmere på utgiftene forbundet med disse uønskede begivenhetene så introduseres en stokastisk variabel U :

$$U = \text{utgift for JetPak forbundet med skade eller forsinkelse av en pakke} \quad (7)$$

Dersom en pakke kommer frem uskadet til rett tid så har ikke JetPak noen utgift.

Anta videre at utgiften til JetPak er 500 NOK dersom en pakke er forsinket, men ikke skadet.

Dersom en pakke er skadet, men ikke forsinket, så er utgiften 900 NOK.

Og – til slutt – dersom en pakke er både skadet og forsinket så er utgiften 1 500 NOK.

Utfallsrommet til U er dermed $\Omega = \{ 0, 500, 900, 1\,500 \}$.

JetPak sine forventede utgifter per pakke er da:

$$E[U] = \sum_{i=1}^4 u_i \cdot P(U = u_i) \quad (8)$$

hvor

$$P(U = 0) = P(\bar{S} \cap \bar{F}) \quad (9)$$

$$P(U = 500) = P(\bar{S} \cap F) \quad (10)$$

$$P(U = 900) = P(S \cap \bar{F}) \quad (11)$$

$$P(U = 1\,500) = P(S \cap F) \quad (12)$$

- c) Hva er forventet utgift $E[U]$ for JetPak?

d) Figur 4 viser Venn-diagrammet for de 4 aktuelle kombinasjonene av begivenheter.

Er disse 4 kombinasjonene av begivenheter **disjunkte**?

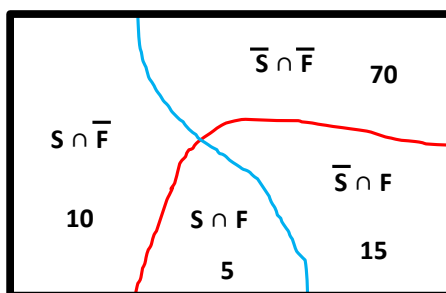
Gi en kort begrunnelse.

$P((\bar{S} \cap F) \cup (S \cap \bar{F}))$ er sannsynligheten for at en tilfeldig valgt pakke er enten bare skadet eller bare forsinket.

e) Vis – ved å bruke addisjonssetningen – at sannsynligheten for at en tilfeldig valgt pakke er enten bare skadet eller bare forsinket, er:

$$P((\bar{S} \cap F) \cup (S \cap \bar{F})) = 0.25 \quad (13)$$

f) Finn samme sannsynlighet som i oppgave 1e ved å bruke enkel kombinatorikk og tallene i figur 1 direkte.²



Figur 4: Venn-diagram.

²I likhet med oppgave 1b så er utregningen i denne oppgaven svært kort. Bruk grunnleggende kombinatorikk:

$$P((\bar{S} \cap F) \cup (S \cap \bar{F})) = \frac{\# \text{ gunstige kombinasjoner}}{\# \text{ mulige kombinasjoner}} \quad (14)$$

og tallene i figur 1.

g) Hva er sannsynligheten for at en tilfeldig valgt pakke er skadet? ³

h) Man kan vise at sannsynligheten for at en tilfeldig valgt pakke er uskadet eller kommer til riktig tid, er:

$$P(\overline{S} \cup \overline{F}) = 0.95 \quad (15)$$

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

Bruk lign.(15) og den ene tvillingsetningen ⁴ til å finne sannsynligheten for at en tilfeldig valgt pakke er både skadet og forsinket, dvs. finn $P(S \cap F)$.

Stemmer denne sannsynligheten med oppgave **1b**?
Gi en *kort* begrunnelse.

³Finn $P(S)$. Bruk gjerne *oppsplitting av utfallsrom*.

⁴Se gjerne formelsamlingen.

Flere trekk ($s > 1$)

La oss nå se på problemet når vi trekker – altså leverer – flere (dvs. flere enn én) tilfeldig valgte pakker av de $N = 100$ pakkene. Rekkefølgen pakkene trekkes i betyr ikke noe.

i) I dette kurset har vi lært om 4 forskjellige situasjoner i forbindelse med kombinatorikk. Gi en kort begrunnelse for hvorfor dette – å trekke fra de utvalgte pakkene – tilsvarer situasjon 3.⁵

j) Anta at JetPak leverer $s = 12$ tilfeldige pakker en dag.

Hva er sannsynligheten P_{mix} for at JetPak leverer 3 pakker fra hver av de 4 aktuelle begivenhetene i figur 1?⁶



⁵Se gjerne formelsamlingen angående hva situasjon 3 er.

⁶Denne oppgaven er litt vanskelig. Dersom dere greier litt så får dere poeng. Prøv å ikke lever inn blankt.

Oppgave 2: (økonomi)

I denne oppgaven skal vi se på små- og mellomstore bedrifter i IKT-bransjen i Møre og Romsdal (M&R).

Anta at den generelle trenden er – over en periode på 10 år – at sannsynligheten for at en slik bedrift går konkurs i løpet et gitt år er:

$$p = 5\% \quad (16)$$

Her er p bare gjennomsnittet av konkurs-sannsynlighetene over de 10 aktuelle årene.

La oss definere følgende stokastiske variabel:

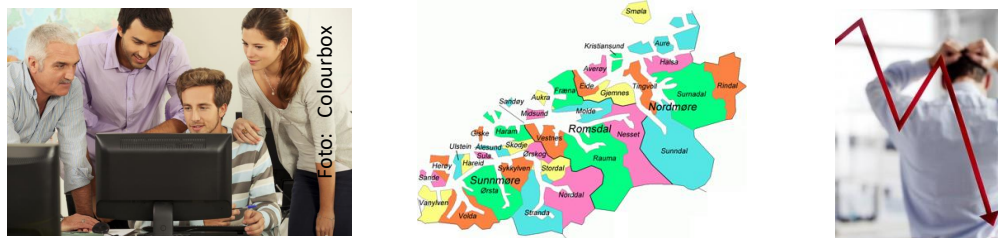
$$X = \text{antall små- og mellomstore IKT-bedrifter i M\&R som går konkurs i løpet av ett år} \quad (17)$$

Anta videre at det er $n = 120$ små- og mellomstore IKT-bedrifter i M&R.

Anta også at konkurs eller ikke i disse bedriftene er uavhengige begivenheter.

- a) Begrunn hvorfor det er rimelig å anta at X er binomisk fordelt, dvs. begrunn hvorfor ⁷

$$X \sim \text{Bin}[n = 120, p = 0.05] \quad (18)$$



Figur 5: IKT, Møre og Romsdal, konkurs.

⁷Se gjerne formelsamlingen.

- b) Hva er sannsynligheten for at nøyaktig 2 av de 120 bedriftene går konkurs i løpet av et gitt år? ⁸
- c) Hva er sannsynligheten for at høyst 2 av de 120 bedrifter går konkurs i løpet av et gitt år? ⁹

En journalist i nettavisen Panorama har skrevet en sak om sannsynligheten for å gå konkurs for IKT-bedrifter i Møre og Romsdal (M&R), se figur 6.

Journalisten bygger saken sin på observasjonen fra 2016 – at det dette året bare var 2 av 120 IKT-bedrifter som gikk konkurs – dvs. kun

$$\frac{2}{120} \cdot 100 \% \approx 1.7 \% \quad (19)$$

av bedriftene i M&R gikk konkurs i 2016.

På dette grunnlaget konkluderer journalisten med at da vil trolig også konkurs-sannsynligheten for 2017 være det samme, dvs. 1.7 % også i 2017. Journalisten sin prediksjon for 2017 er altså i strid med lign.(16) som sier at konkurs-sannsynligheten er 5 %.

- d) Synes du at journalisten baserer sin konklusjon på riktig grunnlag når han predikerer at konkurs-sannsynligheten trolig også blir 1.7 % i 2017? Gi en *kort* begrunnelse.



Figur 6: Nettavisen Panorama.

⁸Dvs. finn $P(X = 2)$.

⁹Dvs. finn $P(X \leq 2)$.

- e) *i)* Dersom en betingelse er oppfylt så kan en **binomisk** fordeling tilnærmes med en **normalfordeling**. Hvilken betingelse er det? ¹⁰
- ii)* Er denne betingelsen oppfylt i vårt tilfelle?
- f) Hva er sannsynligheten for at høyst 2 av de 120 bedrifter går konkurs dersom vi bruker tilnærmelsen om at X er normalfordelt? ¹¹
- g) Sammenlign svaret i oppgave **2c** med **2f**.
Kommenter svaret. ¹²

La oss nå ta utgangspunkt i at de aktuelle IKT-bedriftene er relativt like slik at konkurs-sannsynligheten hovedsakelig styres av **markedet** for denne typen bedrifter. Somregel er det slik at markedsutsiktene for bedriftene styres av de **generelle trendene i økonomien**. I dette bildet er det altså **ytrefaktorer** – opp- og nedgangskonjukturer – som styrer sannsynligheten for konkurs.

- h) Er det fremdeles rimelig å anta at X er **binomisk** fordelt?
Begrunn svaret. ¹³



¹⁰Se formelsamlingen.

¹¹Her må man ha $E[X]$ og $\sigma[X]$ for å løse oppgaven.

¹²Er det rimelig at disse to sannsynlighetene er omtrent like?

¹³Se på oppgave **2a** og vurder om noen av forutsetningene er brutt.

Oppgave 3: (økonomi)

Du driver AVIS bilutleie i Molde. AVIS tjener penger på å leie ut biler.

Du ønsker å se nærmere på økonomien til firmaet.

I den sammenheng defineres den stokastiske variabelen

$$X = \text{antall biler som leies ut en tilfeldig valgt dag}$$

For enkelhet skyld, anta at firmaet kun disponerer 5 biler.

Basert på historikk har man funnet sannsynlighetsfordelingen til X , se tabellen i figur 7 nedenfor:

x_i	0	1	2	3	4	5
$P(X=x_i)$	0.03	0.08	0.15	?	0.25	0.20

Figur 7: Sannsynlighetsfordeling $P(X = x)$.

- a) Er den stokastiske variabelen X diskret eller kontinuertlig?
Gi en *kort* forklaring.¹⁴



Figur 8: AVIS bilutleie i Molde.

¹⁴En setning er nok.

b) Fra tabellen i figur 7 ser vi at alle sannsynlighetene $P(X = x_i)$ er kjent bortsett fra $P(X = 3)$. Begrunn hvorfor $P(X = 3) = 0.29$.

c) Finn $E[X]$.

d) Man kan vise at $E[X^2] = 12.29$.
(Dette faktum skal du ikke vise. Bare ta det for gitt.)

Finn standavviket $\sigma[X]$ og gi en kort tolkning av $\sigma[X]$.

Anta at AVIS tjener 620 NOK per dag for hver bil de leier ut. Anta videre at firmaet har faste utgifter på 1880 NOK per dag. Utleiefirmaet sin **fortjeneste** F per dag er dermed:

$$F = aX - b \quad (20)$$

hvor

$$a = 620 \text{ NOK/bil} \quad (21)$$

$$b = 1880 \text{ NOK} \quad (22)$$

e) Hva er sannsynligheten for at firmaet går med **overskudd** en gitt dag? ¹⁵

f) Finn forventet fortjeneste $E[F]$ en gitt dag.

g) Finn variansen til fortjenesten $Var[F]$ en gitt dag.

¹⁵Dvs. finn $P(F > 0)$.

Tips: $F > 0$ når $X > b/a = 3.03$. Dermed vet vi at:

$$P(F > 0) = P(X \geq 4) \quad (23)$$

Anta at antall biler som leies ut en dag er uavhengig av antall biler som leies ut andre dager. For å spare skrivearbeid, la oss bruke kortnotasjonen:

$$p_i = P(X = x_i) \quad (24)$$

hvor $i = 0, 1, 2, 3, 4, 5$.

- h)** La Z være en stokastisk variabel som sier noe om antall biler som leies ut i løpet av en periode på **tre dager**.

Vis at $P(Z = 2)$ – dvs. sannsynligheten for at AVIS leier ut to biler i løpet av en periode på tre dager – er gitt ved: ¹⁶

$$P(Z = 2) = 3 p_0 p_1^2 + 3 p_0^2 p_2 \quad (26)$$

- i)** Regn ut den numeriske verdien, altså tallverdien, av $P(Z = 2)$ i lign.(26).

¹⁶Det er totalt 6 måter å leie ut **to biler** på i løpet av en periode på **tre dager**. Bruk derfor formelen:

$$\begin{aligned} P(Z = 2) &= p_0 \cdot p_1 \cdot p_1 + p_1 \cdot p_0 \cdot p_1 + p_1 \cdot p_1 \cdot p_0 \\ &+ p_2 \cdot p_0 \cdot p_0 + p_0 \cdot p_2 \cdot p_0 + p_0 \cdot p_0 \cdot p_2 \end{aligned} \quad (25)$$

La $F_1, F_2, F_3, \dots, F_n$ være fortjenesten for firmaet for dag nr. 1, 2, 3, ..., n . Den totale fortjenesten for AVIS etter n dager er da:

$$Y = F_1 + F_2 + F_3 + \dots + F_n \quad (27)$$

hvor, som nevnt tidligere, alle F_i -ene er uavhengige og alle F_i -ene har samme sannsynlighetsfordeling (som gitt i tabellen i figur 7).

- j) *i*) Med forutsetningene som nevnt ovenfor og dersom vi ser på et tilstrekkelig antall dager n , hvilken læresetning gjelder da for den stokastiske variabelen Y ?
- ii*) Hvilken sannsynlighetsfordeling har da Y , med god tilnærming?
- iii*) Hvor stor må n ($n =$ dager, dvs. antall "forsøk") være, **omtrent**, for at læresetningen fra deloppgave **3j i** skal gjelde? ¹⁷

AVIS har $n = 250$ åpningsdager i året.

- k) Hva er forventet fortjeneste i løpet av et helt år, dvs. hva er $E[Y]$?
- l) Hva er variansen til fortjenesten over et helt år, dvs. hva er $Var[Y]$?

■

¹⁷Kun en **tommelfingerregel** er godt nok her.

Oppgave 4: (økonomi)

Helse Møre og Romsdal (Helse M&R) har ansvaret for ambulansetjenesten i fylket. Logistikerne i Helse M&R har en mistanke om at antall utrykninger øker fra år til år. Derfor bestemmer de seg for å observere antall utrykninger i perioden 2011 – 2016 og innfører følgende variabler:

$$x_i = \text{år nr. } i \quad (i = 1, \dots, 6) \quad (28)$$

$$y_i = \text{antall utrykninger for år nr. } i \quad (i = 1, \dots, 6) \quad (29)$$

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

Figur 9: År x_i og antall utrykninger y_i .

Et Excel-plott av dataene ovenfor er vist i figur 11.

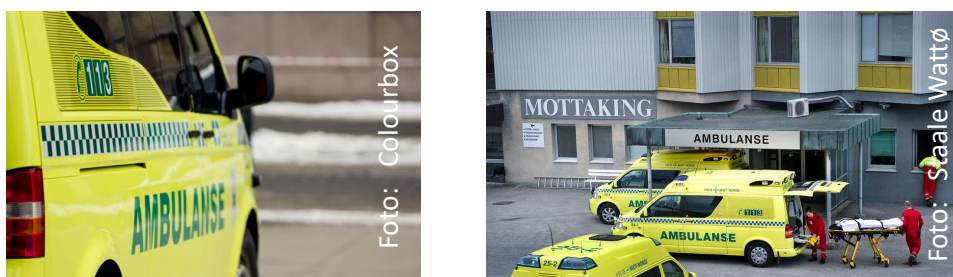
Dataprogrammet Excel kan også regne ut minste kvadraters regresjonslinje. Resultatet er:

$$\hat{y} = 5226.5 + 1168.5x \quad (30)$$

hvor altså

$$\hat{\alpha} = 5226.1 \quad (31)$$

$$\hat{\beta} = 1168.5 \quad (32)$$

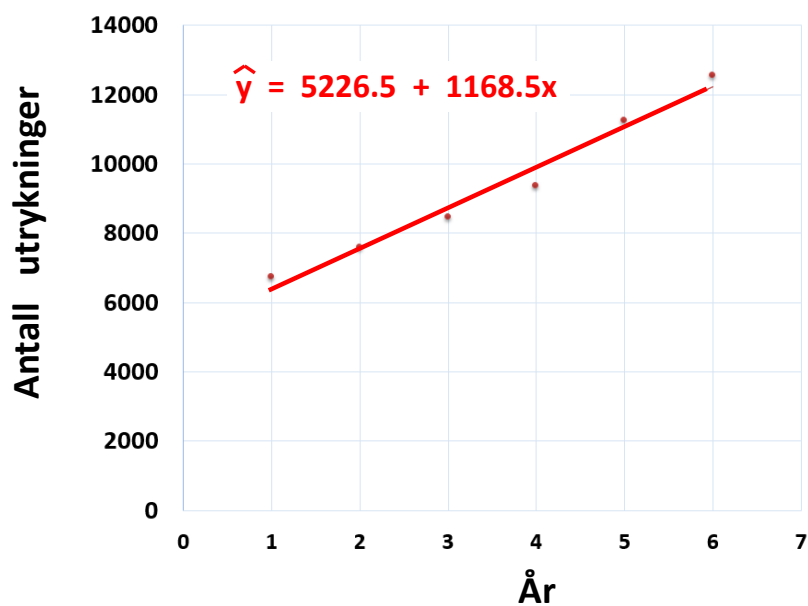


Figur 10: Ambulanse.

- a) *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? ¹⁸
- iii)* Spesielt, hvordan tolker du parameteren $\hat{\beta}$ i regresjonslinjen i lign.(30)?

- b) Anta at den lineære trenden i antall utrykninger fortsetter helt frem til 2020.

Hvor mange utrykninger predikerer regresjonslinjen at det vil være i år 2020? ¹⁹



Figur 11: Plott av dataene i figur 9.

¹⁸Tips: stigningstall.

¹⁹År 2020 tilsvarer $x = 10$.

- c) Vis at korrelasjonskoeffisienten R_{xy} er gitt ved

$$R_{xy} = \hat{\beta} \frac{S_x}{S_y} \quad (33)$$

Tips: Bruk definisjonen av $\hat{\beta}$ i regresjonslinjen, se formelsamlingen.
Bruk deretter definisjonen av R_{xy} .

Ut fra tabellen i figur 9 kan man regne ut gjennomsnittet \bar{x} :

$$\bar{x} = 3.5 \quad (34)$$

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

$$S_x = 1.87 \quad , \quad S_y = 2213.9 \quad (35)$$

(Størrelsene i lign.(34) og (35) trenger du ikke å regne ut. Bare ta dem for gitt).

- d) Regn ut korrelasjonskoeffisienten R_{xy} for observasjonene ved å bruke lign.(33).
- e) Tolk svaret du fikk for R_{xy} i oppgave 4d.²⁰
- f) Finn gjennomsnittlig antall utrykninger per år \bar{y} uten å bruke definisjonen av gjennomsnitt.²¹



²⁰Hva sier den numeriske (tallverdien) av R_{xy} fra oppgave 4d om graden av korrelasjon mellom observasjonene x og y ?

²¹Bruk definisjonen av $\hat{\alpha}$ i regresjonslinjen, se formelsamlingen.

Kapittel 12

Kontinuasjoneksamen 2017

Eksamen i

MAT110 Statistikk 1

Eksamensdag	:	Onsdag 3. jan. 2018
Tid	:	09:00 – 13:00 (4 timer)
Faglærer/telefonnummer	:	Molde: Per Kristian Rekdal / 924 97 051
Hjelpemidler	:	KD + formelsamling fra 2017 + generell ordbok (morsmål/norsk/engelsk i papirformat)
Antall sider inkl. forsiden	:	16
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: (økonomi , aksjer)

For økonomer som investerer i aksjer så er pris på aksjene, altså aksjekursen, interessant.

Anta at du jobber i et investeringsselskap og at du har spesialisert deg på aksjer til handelshus innen dagligvarer. Du ønsker å se nærmere på aksjekursen til NorgesGruppen ASA - Norges største handelshus.

Dagens aksjekurs er kjent, men du usikker på hva aksjekursen vil bli **om ett år**.

For å finne mer ut av hva forventning og varians (= usikkerhet) til aksjekursen blir om ett år så defineres følgende stokastiske variabel:

$$X = \overbrace{\text{aksjekurs}}^{\text{pris}} \text{ for én aksje i NorgesGruppen (i NOK) om ett år} \quad (1)$$

Anta at den stokastiske variabelen X har følgende sannsynlighetsfordeling: ¹

x_i	< 180	180	200	220	240	260	> 260
$P(X=x_i)$	0	0.05	0.05	?	0.20	0.10	0

Figur 1: Sannsynlighetsfordeling $P(X = x)$.



Figur 2: Norgesgruppen.

¹Enheten til verdiene x_i er norske kroner, altså NOK.

- a) Er den stokastiske variabelen X diskret eller kontinuerlig?
Gi en *kort* forklaring. ²
- b) Fra tabellen i figur 1 ser vi at alle sannsynlighetene $P(X = x_i)$ er kjent bortsett fra $P(X = 220)$.

Begrunn hvorfor $P(X = 220) = 0.60$.
- c) Hva er sannsynligheten for at prisen på aksjene, altså aksjekursen, er 200 NOK eller mindre? ³
- d) Finn forventet aksjekurs $E[X]$.
- e) Finn variansen til aksjekursen $Var[X]$.



²Én setning er nok.

³Tips: Bruk de relevante tallene fra tabellen i figur 1 og finn $P(X \leq 200)$.

Oppgave 2: (økonomi , aksjer)

La oss fortsette å se på investeringselskapet fra oppgave 1.

Istedet for å investere i kun ett selskap så ønsker du nå heller å spre risikoen ved å investere i tre ulike selskap, nemlig de tre største grossistgrupperingene i Norge:

$$\underbrace{\text{Norgesgruppen}}_{\text{selskap A}}, \quad \underbrace{\text{Coop}}_{\text{selskap B}}, \quad \underbrace{\text{Reitangruppen}}_{\text{selskap C}} \quad (2)$$

Du bestemmer deg for å kjøpe

$$N = 50\,000 \quad (3)$$

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 :

$$N \cdot a = \text{antall aksjer som investeres i selskap A, Norgesgruppen} \quad (4)$$

$$N \cdot b = \text{antall aksjer som investeres i selskap B, Coop} \quad (5)$$

$$N \cdot c = \text{antall aksjer som investeres i selskap C, Reitangruppen} \quad (6)$$

hvor

$$a + b + c = 1 \quad (7)$$

dvs. a er den brøkdelen av de $N = 50\,000$ aksjene som investeres i selskap A. Tilsvarende for b og c .



Figur 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 forventning og varians til fortjenesten av aksjene ved et eventuelt salg **om ett år**.

Den stokastiske variabelen X er den samme som i oppgave **1**.

I tillegg defineres to nye stokastiske variabler Y og Z . Alt i alt:

$$X = \overbrace{\text{aksjekurs}}^{\text{pris}} \text{ for én aksje i selskap } A, \text{ Norgesgruppen (i NOK), om ett år} \quad (8)$$

$$Y = \text{aksjekurs for én aksje i selskap } B, \text{ Coop (i NOK), om ett år} \quad (9)$$

$$Z = \underbrace{\text{aksjekurs}}_{\text{pris}} \text{ for én aksje i selskap } C, \text{ Reitangruppen (i NOK), om ett år} \quad (10)$$

- a) Anta at prisen på aksjene i dag er **230** NOK for selskap A , **205** NOK for selskap B og **235** NOK for selskap C . Bruk disse opplysningene samt formelen

$$F = \text{antall aksjer} \cdot \text{pris per aksje om ett år} - \text{antall aksjer} \cdot \text{pris per aksje i dag} \quad (11)$$

til å vise at fortjenesten F **om ett år** er gitt ved: ⁴

$$F = Na(X - 230) + Nb(Y - 205) + Nc(Z - 235) \quad (12)$$



Foto: Colourbox



Figur 4: Investering.

⁴I tillegg til lign.(11), bruk også lign.(4)-(6).

Anta videre at forventet pris (altså aksjekurs) om ett år er: ⁵

$$E[X] = 225 \quad , \quad E[Y] = 215 \quad , \quad E[Z] = 250 \quad (13)$$

- b) Bruk lign.(12) samt lign.(13) og vis at den forventede fortjenesten $E[F]$ ved et salg av aksjene om ett år er: ⁶

$$E[F] = N(-5a + 10b + 15c) \quad (14)$$

- c) I lign.(14) er det et ledd som er negativt, nemlig leddet $-5a$.

Hva kan man da si om investeringen i selskap A dersom man selger etter ett år? ⁷

- d) Dersom du bestemmer deg for å maksimere den forventede fortjenesten $E[F]$ i lign.(14) uten å tenke på usikkerheten $Var[F]$, hva slags selskap bør du investere i da? ⁸

Hva er verdien på den forventede fortjenesten $E[F]$ i det tilfellet?

⁵Forventningene er i NOK, men dropper benevningen her for enkelhets skyld.

⁶Husk regnereglene for forventning i formelsamlingen (se side 50 i 2017-versjonen).

⁷Denne oppgaven krever ingen regning. Kun et kort svar.

⁸Denne oppgaven krever ingen regning. Kun en kort begrunnelse. Referer gjerne til lign.(14).

Anta at det nasjonale rentenivået i Norge påvirker aksjekursen til alle grossistgruppene i landet. Vi skal nå modellere hvordan en slik felles påvirkningsfaktor kan inkluderes på to forskjellige måter, modell 1 og modell 2.

Modell 1:

I modell 1 antar vi at aksjekursene er uavhengige, dvs. at X , Y og Z er uavhengige. Vi tar nå hensyn til rentenivået sin påvirkning av aksjekursene ved å inkludere dette i variansene, på følgende måte:

$$\text{Var}[X] = \sigma^2 \quad , \quad \text{Var}[Y] = 1.5 \sigma^2 \quad , \quad \text{Var}[Z] = 3 \sigma^2 \quad (15)$$

hvor σ^2 er en felles faktor for variansene.

- e) Bruk lign.(12) samt lign.(15) og vis at variansen $\text{Var}[F]$ til fortjenesten ved et eventuelt salg **om ett år** er: ⁹

$$\text{Var}[F] = \sigma^2 N^2 (a^2 + 1.5b^2 + 3c^2) \quad (16)$$

- f) Du ønsker å minimere risikoen i din investering, dvs. du ønsker å **minimere** $\text{Var}[F]$. Man kan vise at variansen $\text{Var}[F]$ i lign.(16) er minst mulig når

$$a = 0.5 \quad , \quad b = 0.33 \quad , \quad c = 0.17 \quad (17)$$

(Du skal ikke vise lign.(17). 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? ¹⁰

⁹Tips: Bruk gjerne setningen om kovarians i formelsamlingen (se side 59 i 2017-versjonen). Husk at X , Y og Z er uavhengige. Hva kan man da si om kovariansen $\text{Cov}[X, Y, Z]$?

¹⁰Bruk lign.(4)-(6) sammen med lign.(17).

Modell 2:

I modell 2 inkluderes rentenivåets påvirkning på en annen måte enn i den forrige modellen.

Her, i motsetning til den forrige modellen, er de stokastiske variablene **avhengige**.

De er avhengige på en slik måte at de, per konstruksjon, har samme forventning og varians som den forrige modellen:

$$X = 225 + R \quad , \quad Y = 215 + \sqrt{1.5}R \quad , \quad Z = 250 + \sqrt{3}R \quad (18)$$

hvor “renteindikatoren” R er en stokastisk variabel som er normalfordelt:

$$R \sim N[0, \sigma] \quad (19)$$

dvs. med $E[R] = 0$ og $Var[R] = \sigma^2$.

Ved å sette lign.(18) inn i uttrykket for fortjenesten F i lign.(12) så får man:

$$F = RN(a + \sqrt{1.5}b + \sqrt{3}c) + k \quad (20)$$

hvor k er en konstant. (Lign.(20) skal du ikke vise. Bare ta den for gitt.)

g) Bruk lign.(20) og vis at variansen til fortjenesten F om ett år er: ¹¹

$$\boxed{Var[F] = \sigma^2 N^2 (a + \sqrt{1.5}b + \sqrt{3}c)^2} \quad (21)$$

¹¹Tips: I lign.(20) er R en stokastisk variabel mens resten er en konstant. Bruk gjerne dette faktum sammen med setningen om kovarians i formelsamlingen (se side 59 i 2017-versjonen).

Som i modell 1 så ønsker vi nå også i modell 2 å minimere risikoen i investeringen, dvs. vi ønsker å minimere $Var[F]$. Variansen $Var[F]$ i lign.(21) er minst mulig når:

$$a = 1 \quad , \quad b = 0 \quad , \quad c = 0 \quad (22)$$

(Du skal ikke vise lign.(22). Bare ta dette for gitt.)

- h)** I modell 1 så vi at man får minimal risiko når investeringen fordeles over alle tre selskapene A , B og C .

Her i modell 2, derimot, innser vi fra lign.(22) at risikoen i investeringen er minst mulig dersom alt investeres i selskap A og ingenting i de andre selskapene.

Sett i lys av at modell 2 gir en avhengighet mellom aksjekursene X , Y og Z , hvorfor er det rimelig at man skal investere alt i selskap A i modell 2? ¹²



¹²Her behøver man ingen utregninger, kun argumentasjon. Bruk gjerne lign.(18) i din argumentasjon. Hvilket selskap har minst følsomhet for rentenivået ifølge denne ligningen?

Oppgave 3: (logistikk)

Norsk Hydro ASA er en norsk leverandør av aluminium og aluminiumprodukter.

En av fabrikkene til Norsk Hydro ligger i Årdal kommune i Sogn og Fjordane.
En av råvarene som trengs for å lage aluminium er *alumina*.
Alumina fraktes i båter fra Brasil til Norge.

Forsinkelser i transporten kan få store konsekvenser for Norsk Hydro.
Derfor ønsker logistikkplanleggerne i Norsk Hydro å se nærmere dette, og definerer følgende stokastiske variabel:

$$X_i = \text{antall dager forsinkelse for leveranse nr. } i \text{ fra Brasil til Norge} \quad (23)$$

hvor indeksen $i = 1, 2, \dots, n$ er leveranse nr. i .

Her er n antall leveranser Norsk Hydro bestemmer seg for å se på.

Vi antar at det er identiske båter som bruker for de forskjellige leveransene.

Transportleverandøren oppgir at det ikke forventes noen forsinkelser, dvs.

$$E[X_i] = 0 \quad (24)$$

for alle $i = 1, 2, \dots, n$.



Figur 5: Årdal verk. Transport.

Anta at det i løpet av et år trengs n antall leveranser av alumina fra Brasil til Norge. Det totale antall dager forsinkelser Y per år på grunn av leveranser er derfor:

$$Y = \sum_{i=1}^n X_i = X_1 + X_2 + \dots + X_n \quad (25)$$

- a) i) Gi en tolkning av størrelsen $E[Y]$.¹³
ii) Finn $E[Y]$.

Selv om transportleverandøren oppgir at det ikke forventes noen forsinkelser så er leveringstiden selvsagt forbundet med noe usikkerhet.

Anta at standardavviket $\sigma[X_i]$ er gitt ved:

$$\sigma[X_i] = \frac{aL}{M} \quad (26)$$

hvor

$$a = 0.16 \quad (27)$$

$$L = \text{forventet leveringstid i antall dager for en gitt leveranse} \quad (28)$$

$$M = \text{antall identiske båter som settes inn i ruten mellom Brasil og Norge} \quad (29)$$

¹³Dvs., hva betyr $E[Y]$ på “godt norsk” i vårt tilfelle?

Anta at forsinkelsene X_i er uavhengige av hverandre.

- b) i) Gi en tolkning av størrelsen $Var[Y]$.
ii) Vis at variansen $Var[Y]$ er gitt ved:¹⁴

$$Var[Y] = n \left(\frac{aL}{M} \right)^2 \quad (30)$$

- c) Fra deloppgavene foran innser vi at variansene $Var[X_i]$ og $Var[Y]$ er ulike. Hvorfor er det slik?¹⁵

- d) Anta at alle stokastiske variabler X_i har samme ukjente sannsynlighetsfordeling. Anta videre at Norsk Hydro trenger 40 leveranser med alumina per år mellom Brasil og Norge, dvs.:

$$n = 40 \quad (31)$$

Hvilken fordeling (tilnærmet) er det rimelig å anta at Y har?
Begrunn svaret.

¹⁴Husk at: $Var[X_i] = (\sigma[X_i])^2 = \left(\frac{aL}{M}\right)^2$.

¹⁵Her behøver man ingen utregninger, kun en *kort* forklaring. Én setning er nok.

- e) Ledelsen i Norsk Hydro ved Årdal synes at 10 dagers forsinkelse i løpet av et år er mye. Derfor ønsker de ikke at det skal inntreffe.

Dersom det skal være 95 % sannsynlighet for at det er mindre enn 10 dagers forsinkelse i løpet av et år, hvor mange båter M må det da settes inn mellom Brasil og Norge? ¹⁶

Anta at forventet leveringstid er $L = 18$ dager.



¹⁶Finn antall båter M som må settes inn for at det skal være 95 % sannsynlighet for at det er mindre enn 10 dagers forsinkelse i løpet av et år. Det er bestemt av:

$$P(Y < 10) = 0.95 \quad (32)$$

Siden $Y \sim N[E[Y], Var[Y]]$, hvor $E[Y]$ og $Var[Y]$ ble funnet i oppgave **3a** og **3b**, så kan vi standardisere lign.(32) og deretter bruke “*omvendt tabelloppslag*”. Løs så med hensyn på M alene og sett inn tall til slutt.

Oppgave 4: (logistikk)

La oss fortsette med Norsk Hydro fra forrige oppgave.

Logistikkplanleggerne i Norsk Hydro i Årdal ønsker å se nærmere på behovet av alumina - et av råvarene som trengs for å lage aluminium. De ønsker å bruke dataene av alumina-behovet fra og med 2013 til og med 2017. De innfører derfor følgende notasjon for **observasjonene** x_i og y_i :

$$x_i = \text{år nr. } i, \text{ hvor } x_1 = 2013, x_2 = 2014, \dots, x_5 = 2017 \quad (33)$$

$$y_i = \text{behov alumina for år nr. } i \quad (34)$$

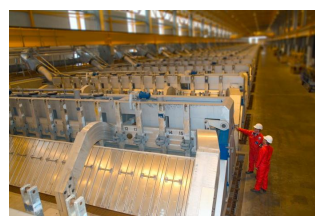
hvor indeksen er $i = 1, 2, \dots, 5$ betegner de 5 årene vi ser på.

x_i (år)	2013	2014	2015	2016	2017
y_i (behov i 1000 tonn)	762	788	794	825	834

Figur 6: Observasjoner.

a) Regn og gjennomsnittene \bar{x} og \bar{y} og vis at disse blir:

$$\bar{x} = 2015, \quad \bar{y} = 800.6 \quad (35)$$



Figur 7: Årdal verk.

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:

$$S_x^2 = 2.5 \quad , \quad S_y^2 = 850.8 \quad (36)$$

Den empiriske **kovariansen** mellom x og y er:

$$S_{xy} = 45.25 \quad (37)$$

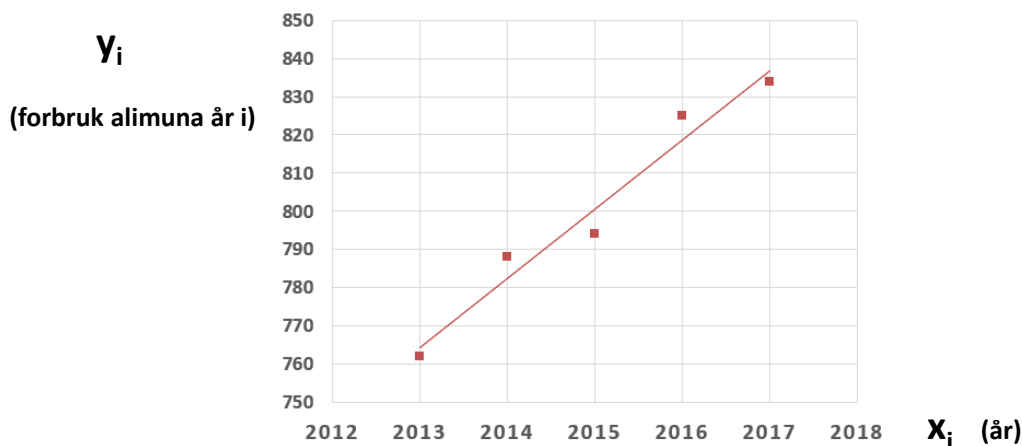
(Størrelsene i lign.(35), (36) og (37) trenger du ikke å regne ut. Bare ta dem for gitt).

b) Regn ut korrelasjonskoeffisienten R_{xy} for observasjonene i tabellen i figur 6.

c) Tolk svaret du fikk for R_{xy} i oppgave **4b**. ¹⁷

¹⁷Hva sier den numeriske (tallverdien) av R_{xy} fra oppgave **4b** om graden av korrelasjon mellom observasjonene x og y ?

Ut fra behovet de 5 foregående årene ønsker Norsk Hydro å finne en eksplisitt sammenhengen mellom x og y . De bestemmer seg for å bruke lineær regresjonsanalyse.



Figur 8: Plott av dataene fra tabellen i figur 6.

- d) Den rette linjen i figur 8 viser minste kvadraters regresjonslinje for observasjonene x og y . Regn ut de nødvendige parametrene $\hat{\alpha}$ og $\hat{\beta}$ og vis at den lineære regresjonslinjen er:

$$\hat{y} = -35\,670.9 + 18.1x \quad (38)$$

- e) Anta at den lineære økningen beskrevet av regresjonslinjen fortsetter de neste 10 årene. Anta videre at kapasiteten til Årdal verk er 900 000 tonn alimuna i året. Dersom forbruket oversiger dette så må de gjøre investeringer for å øke kapasiteten.

Ifølge regresjonslinjen i lign.(38), hvilket år vil kapasiteten til Årdal verk være sprengt? ¹⁸



¹⁸Her er det meningen at man skal regne seg frem til svaret. Man skal **ikke** bare lese av fra figur 8. Husk å bruke 900 i regresjonslinjen siden \hat{y} er i antall 1000 tonn.

Kapittel 13

Hovedeksamen 2018

Eksamen i

MAT110 Statistikk 1

Eksamensdag	:	Tirsdag 22. mai 2018
Tid	:	09:00 – 13:00 (4 timer)
Faglærer/telefonnummer	:	Molde + Kristiansund: Per Kristian Rekdal / 924 97 051
Hjelpemidler	:	KD + formelsamling + generell ordbok (morsmål/norsk/engelsk i papirformat)
Antall sider inkl. forsiden	:	21
Målform	:	Norsk (bokmål)

Noen generelle råd:

- **Skriv rett inn. Ikke bruk så mye tid på kladding.**
- **Det er totalt 4 oppgaver.**

Oppgave 1 : (logistikk)

Glamox er en norsk produsent og leverandør av belysning.

Glamox har til enhver tid mange tilbud ute hos kunder som de venter tilbakemelding på. Hensikten med denne oppgaven er å finne relevante sannsynligheter for at tilbud aksepteres basert på to kriterier: **selger** som har laget tilbudet og **verdien** på tilbudet.

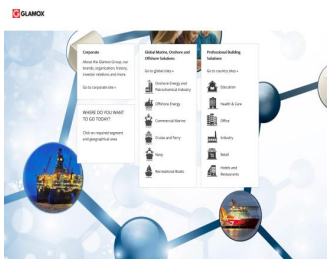
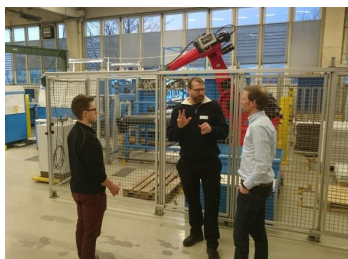
1) Selgere:

La oss kun se på 2 selgere hos Glamox, nemlig selgerne 1 og 2.

2) Verdi:

La oss for enkelhets skyld kun skille mellom store og små tilbud, dvs. tilbud som har høy (h) verdi eller lav (\bar{h}) verdi i kroner og øre. Det kan f.eks. være tilbud som er over eller under 10 000 NOK.

Et tilbud blir enten akseptert (a) eller ikke akseptert (\bar{a}). Utfallsrommet har da 8 utfall, se lign.(1) på neste side.



Figur 1: Glamox.

Glamox bruker 350 historiske data for å estimere relevante sannsynligheter. Resultatene av de historiske dataene er at utfall $(1, h, a)$ inntraff 5 ganger, utfall $(1, \bar{h}, a)$ inntraff 40 ganger osv.

Sannsynlighetene p_1, p_2, \dots, p_8 for de 8 mulige utfallene i utfallsrommet er dermed:

$$\Omega = \left\{ \begin{array}{cccc} \overbrace{(1, h, a)}^{p_1 = \frac{5}{350}}, & \overbrace{(1, \bar{h}, a)}^{p_2 = \frac{40}{350}}, & \overbrace{(1, h, \bar{a})}^{p_3 = \frac{25}{350}}, & \overbrace{(1, \bar{h}, \bar{a})}^{p_4 = \frac{80}{350}} \\ \overbrace{(2, h, a)}^{p_5 = \frac{8}{350}}, & \overbrace{(2, \bar{h}, a)}^{p_6 = \frac{68}{350}}, & \overbrace{(2, h, \bar{a})}^{p_7 = \frac{22}{350}}, & \overbrace{(2, \bar{h}, \bar{a})}^{p_8 = \frac{102}{350}} \end{array} \right\} \quad (1)$$

hvor utfallet $(1, h, a)$ betyr at selger **1** lager et tilbud med høy (h) verdi som blir akseptert (a). Tilsvarende for de øvrige utfallene.

Videre er $N_1 = 150$ tilbud gitt av selger **1** og $N_2 = 200$ gitt av selger **2**. Totalt er da: $N = N_1 + N_2 = 350$.

a) Vis at sannsynlighetfordelingen p_1, p_2, \dots, p_8 i lign.(1) er en gyldig fordeling.

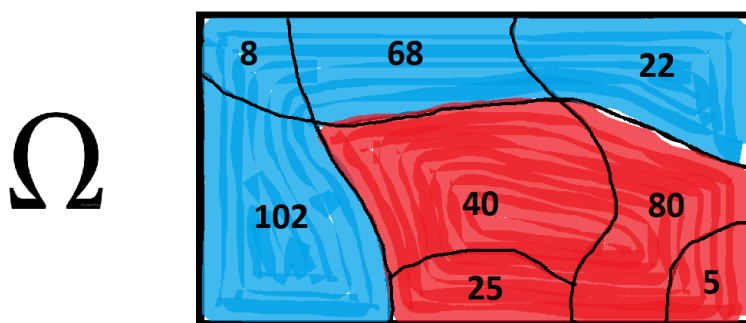
La oss definere følgende begivenhene:

$$S_1 = \left\{ (1, h, a), (1, \bar{h}, a), (1, h, \bar{a}), (1, \bar{h}, \bar{a}) \right\} \quad (2)$$

$$S_2 = \left\{ (2, h, a), (2, \bar{h}, a), (2, h, \bar{a}), (2, \bar{h}, \bar{a}) \right\} \quad (3)$$

- b) Figur 2 viser Venn-diagrammet for de 8 utfallene i utfallsrommet Ω . Tallene i figuren viser antall ganger de respektive utfallene inntreffer, se lign.(1). Den røde og den blå delen av utfallsrommet definerer begivenhetene S_1 og S_2 , henholdsvis.

Er begivenhetene S_1 og S_2 disjunkte? Gi en *kort* begrunnelse. ¹



Figur 2: Utfallsrom Ω .

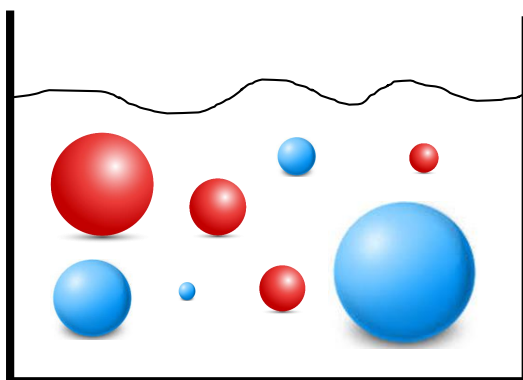
¹Én setning er nok.

- c) La oss si at hvert av de 8 utfallene i utfallsrommet Ω er assosiert med ei kule. Altså totalt 8 kuler, se figur 3.

La oss videre anta at vi har to typer kuler, røde kuler og blå kuler. De røde kulene er assosiert med utfallene til selger 1, og de blå kulene er assosiert med utfallene til selger 2.

La oss anta at tilbudene laget av selger 1 er uavhengige av tilbudene laget av selger 2.

På tross av at tilbudene som de to selgerne har laget er uavhengige av hverandre, hvorfor kan vi ikke bruke urnemodellen til å f.eks. finne sannsynligheten $P(S_2)$ for at et tilbud er gitt av selger 2? ²



Figur 3: Urne.

²Svar kort på denne oppgaven. Én setning er nok. Ingen utregninger behøves. Hva må være oppfylt, i tillegg til uavhengighet, for at urnemodellen skal gjelde? Er det oppfylt i vårt tilfelle?

Man kan altså ikke bruke urnemodellen, men man kan likevel finne sannsynligheten $P(S_2)$ på følgende måte, via lign.(1):

$$\begin{aligned} \underline{P(S_2)} &= p_5 + p_6 + p_7 + p_8 \\ &= \frac{8}{350} + \frac{68}{350} + \frac{22}{350} + \frac{102}{350} = \frac{200}{350} = \underline{0.5714} \end{aligned} \tag{4}$$

Sannsynligheter for andre begivenheter finner man på tilsvarende måte.

La oss også definere følgende begivenheter:

$$A = \left\{ (1, h, a), (1, \bar{h}, a), (2, h, a), (2, \bar{h}, a) \right\}$$

og

$$H = \left\{ (1, h, a), (1, h, \bar{a}), (2, h, a), (2, h, \bar{a}) \right\} \tag{5}$$

La oss nå se på et tilbud med høy (h) verdi laget av selger 2.

Den betingede sannsynligheten for at et slikt tilbud blir akseptert er gitt ved den generelle multiplikasjonssetningen:

$$\boxed{P(A | \overbrace{S_2 \cap H}^{\text{vet}}) = \frac{P(S_2 \cap H \cap A)}{P(S_2 \cap H)}} \tag{6}$$

d) Bruk lign.(6) og vis at $P(A|S_2 \cap H)$ er: ³

$$P(A|S_2 \cap H) = \frac{8}{30} \quad (= 0.2667) \quad (7)$$

e) Gi en tolkning av $P(A|S_2 \cap H)$,
dvs. forklar *kort* hva det betyr på ”godt norsk”.



³Tips: Finn først $P(S_2 \cap H \cap A)$ og $P(S_2 \cap H)$ ved å bruke tilsvarende metode som i lign.(4).

Oppgave 2: (logistikk - prognostisering)

La oss fortsette med Glamox-beskrivelsen fra oppgave 1.

For å unngå mye repeterende regning, la oss kun se på 4 nye tilbud.
For enkelhets skyld, la oss anta at alle disse 4 nye tilbudene er gitt av selger 2
og at alle har høy (h) verdi.

Anta at **verdien** til tilbud nr. i er v_i (i NOK), se tabell 1:

Tilbud nr. i ($i = 1,2,3,4$)	1	2	3	4
Høyt (h) tilbud	h	h	h	h
Selger S_2	2	2	2	2
Verdi v_i ($i=1,2,3,4$) (NOK)	40 000	33 000	105 000	75 000

Tabell 1: Nye tilbud.



Figur 4: Glamox.

La oss introdusere en stokastisk variabel som sier noe om et tilbud aksepteres eller ikke:

$$Y_i = \begin{cases} 1 & , \text{ dersom kunde aksepterer tilbud nr. } i \\ 0 & , \text{ dersom kunde ikke aksepterer tilbud nr. } i \end{cases} \quad (8)$$

hvor $i = 1, 2, 3, 4$ indikerer tilbud nr. i .

Aksept-sannsynligheten for et gitt tilbud laget av selger 2 med høy (h) verdi fant vi i oppgave **1d**. For enkelhets skyld, la oss kalle denne sannsynligheten for p , dvs. $p = P(A|S_2 \cap H)$, altså:

$$p = \frac{8}{30} = 0.2667 \quad (9)$$

Den stokastiske variabelen Y definert ved:

$$Y = Y_1 + Y_2 + Y_3 + Y_4 \quad (10)$$

beskriver da antall tilbud som aksepteres.

Anta at tilbudene er uavhengige av hverandre, dvs. Y_i 'ene er uavhengige.

- a) Forklar hvorfor Y er **binomisk** fordelt, dvs. forklar hvorfor $Y \sim \text{Bin}[n = 4, p = 0.2667]$.⁴
- b) Hva er sannsynligheten for at to eller flere tilbud blir akseptert, dvs. hva er $P(Y \geq 2)$?⁵
- c) Finn tallverdien for $E[Y]$ og gi en *kort* tolkning på hva $E[Y]$ betyr på godt norsk”.⁶

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

⁵Bruk gjerne at: $P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1)$.

⁶Én setning er nok.

Summen av verdien til de aksepterte tilbudene er gitt ved:

$$V = v_1 Y_1 + v_2 Y_2 + v_3 Y_3 + v_4 Y_4 \quad (11)$$

hvor v_i er verdien på tilbud nr. i .

d) Vis at forventningsverdien til verdien til de aksepterte tilbudene er: ⁷

$$E[V] = (v_1 + v_2 + v_3 + v_4) p \quad (13)$$

e) Bruk lign.(13) og regn ut tallverdien til $E[V]$. ⁸

f) Vis da at variansen til verdien til de aksepterte tilbudene er: ⁹

$$Var[V] = (v_1^2 + v_2^2 + v_3^2 + v_4^2) p(1 - p) \quad (15)$$

⁷Tips: Siden $Y_i \sim \text{Bin}[n_i = 1, p]$ så er:

$$E[Y_1] = E[Y_2] = E[Y_3] = E[Y_4] = p \quad (12)$$

⁸Bruk tallene for v_i i tabell 1 på side 8.

⁹Siden $Y_i \sim \text{Bin}[n_i = 1, p]$ så er:

$$Var[Y_1] = Var[Y_2] = Var[Y_3] = Var[Y_4] = p(1 - p) \quad (14)$$

g) Dersom selger 2 hadde lagt 200 tilbud istedet for bare 4 så er

$$V = v_1 Y_1 + v_2 Y_2 + \dots + v_{200} Y_{200} \quad (16)$$

verdien på de aksepterte tilbudene.

La oss si at alle tilbudene har høy verdi (h) slik at aksept-sannsynligheten p til tilbudene er den samme, men at verdien v_i på de $i = 1, 2, 3, \dots, 200$ tilbudene er forskjellig.

Dersom vi i tillegg fortsatt antar at alle tilbudene er uavhengige, hvorfor kan vi ikke da anvende sentralgrenseteoremet og si at V er tilnærmet normalfordelt?



Oppgave 3: (logistikk og økonomi)

La oss anta at det er $n = 4$ Bunnpris-butikker i Molde.

Alle disse butikkene selger jordbær når det er sesong.

Anta at etterspørselen er den samme hver dag, men at etterspørselen i de 4 butikkene er forskjellige.

Anta videre at etterspørselen i all hovedsak skjer på dagtid, ikke kveldstid.

Dermed er det hensiktsmessig å introdusere en stokastisk variabel D_i (“demand”), hvor:

$$D_i = \text{etterspørsel av antall kurver med friske jordbær per dag på dagtid for butikk nr. } i \quad (17)$$

med $i = 1, 2, 3, 4$.

Den totale etterspørselen per dag på dagtid for de 4 butikkene er da beskrevet av:

$$D_{\text{tot}} = D_1 + D_2 + D_3 + D_4 \quad (18)$$



Figur 5: Bunnpris-butikker og jordbær.

Innkjøperne hos Bunnpris har funnet ut at forventet etterspørsel og tilhørende varians per dag på dagtid for de 4 aktuelle butikkene ($i = 1, 2, 3, 4$) er som oppgitt i tabell 2:

i	1	2	3	4
$E[D_i]$	20	50	30	10
$Var[D_i]$	16	36	88	4

Tabell 2: Forventning $E[D_i]$ og varians $Var[D_i]$ for $i = 1, 2, 3, 4$.

- a) Vis at den totale forventede etterspørselen $E[D_{tot}]$ per dag på dagtid for de 4 butikkene er:

$$E[D_{tot}] = 110 \quad (19)$$

kurver med jordbær.

Anta at etterspørselen av jordbær i de 4 butikkene er uavhengige.

- b) Regn ut den totale variansen $Var[D_{tot}]$ per dag på dagtid for de 4 butikkene.

Anta videre at alle de fire stokastiske variablene D_i er normalfordelt, dvs. anta:

$$D_i \sim N[E[D_i], Var[D_i]] \quad (20)$$

hvor $E[D_i]$ og $Var[D_i]$ er som oppgitt i tabell 2 for $i = 1, 2, 3, 4$.

c) Hva slags fordeling har da D_{tot} ? ¹⁰

¹⁰Se side 108 i formelsamlingen fra 2018. Definisjonen at D_{tot} er gitt i lign.(18).

Jordbærene kastes

La oss anta at jordbærene kun har holdbarhet én dag.

Det betyr at dersom Bunnpris ikke får solgt jordbærene på dagtid så må de kaste dem på kvelden når de stenger butikken. Fortjenesten F per dag blir da:

$$F = \underbrace{p S_{\text{dag}}}_{\text{inntekt}} - \underbrace{c q}_{\text{kostnad}}, \quad (21)$$

hvor

$$\begin{aligned} S_{\text{dag}} &= \min(D_{\text{tot}}, q) \\ &= \text{antall kurver med jordbær som selges i de 4 butikkene per dag på dagtid} \end{aligned} \quad (22)$$

$$q = \text{antall kurver med jordbær som innkjøperne hos Bunnpris bestiller per dag for de 4 butikkene (*order quantity*)} \quad (23)$$

$$c = 30 \text{ NOK (kostnaden for en kurv med jordbær for Bunnpris)} \quad (24)$$

$$p = 40 \text{ NOK (pris som Bunnpris selger en kurv med jordbær for)} \quad (25)$$

Det optimale bestillingskvantumet q^* som gir maksimal forventet fortjeneste av lign.(21) er da bestemt av:

$$P(D_{\text{tot}} \leq q^*) = 1 - \frac{c}{p} \quad (26)$$

hvor

$$q^* = \text{antall kurver med jordbær som Bunnpris må bestille per dag for å maksimere forventet fortjeneste} \quad (27)$$

d) Finn q^* . ¹¹

¹¹**Tips 1:** Bruk de oppgitte verdiene for c og p samt lign.(26).

Tips 2: Argumentet z_0 for sannsynligheten $P(Z \leq z_0) = 0.25$ finnes ikke direkte i tabellen i formelsamlingen. Sannsynlighetene i formelsamlingen ”stopper” ved 0.5000. Men av symmetrigrunner så er $z_0 = -0.675$, altså negativ z_0 . Dermed slipper du å gjøre tabelloppslag i denne deloppgaven.

Redusert pris

Istedet for å kaste jordbærene så velger Bunnpris heller å selge jordbærene til redusert pris på kvelden, et par timer før stengetid.

Antall kurver med jordbær som selges på kvelden til redusert pris er da beskrevet av den stokastiske variabelen:

$$S_{\text{kveld}} = \begin{cases} q - D_{\text{tot}} & , \quad q \geq D_{\text{tot}} \\ 0 & , \quad q < D_{\text{tot}} \end{cases} \quad (28)$$

Fortjenesten F per dag blir da en utvidelse av lign.(21), nemlig:

$$F = \underbrace{p S_{\text{dag}} + p_{\text{red}} S_{\text{kveld}}}_{\text{inntekt}} - \underbrace{c q}_{\text{kostnad}} \quad (29)$$

Det nye, optimale bestillingskvantumet q^* som gir maksimal forventet fortjeneste av lign.(29) er da bestemt av: ¹²

$$P(D_{\text{tot}} \leq q^*) = \frac{p - c}{p - p_{\text{red}}} \quad (30)$$

- e) Anta at Bunnpris kjøper inn $q^* = 116$ kurver med jordbær per dag, hvilken redusert pris p_{red} må da Bunnpris sette på en kurv med jordbær for å maksimere den forventede fortjenesten? ¹³



¹²Du skal ikke vise lign.(30). Bare ta den for gitt.

¹³Dette er en litt vanskelig oppgave. Hopper over den dersom du har dårlig tid.

Oppgave 4: (økonomi)

I kapittel 1 i kurset “*MAT110 Statistikk 1*” så vi på statistiske størrelser med både én 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, \dots, x_n har vi tilhørende observasjoner y_1, y_2, \dots, y_n . I denne sammenheng kan man snakke om korrelasjonskoeffisienten R_{xy} til disse observasjonene.
- i) Hva slags mulige verdier kan R_{xy} ha?
 - ii) Hva er R_{xy} et mål på?
 - iii) Hva slags enhet har R_{xy} ?

Eiendomsmegleren *Notar* ønsker å se nærmere på sammenhengen mellom kvadratmeterpris y på leiligheter og årsinntekten til den som kjøper leiligheten.

$$x = \text{årsinntekt} \quad (\text{i antall 1000 NOK}) \quad (31)$$

$$y = \text{kvadratmeterpris} \quad (\text{NOK per } m^2) \quad (32)$$

Observasjonene x måles i antall 1000 NOK per år, brutto årsinntekt.

Observasjonene y er kvadratmeterpris målt i NOK per m^2 for den aktuelle leiligheten.



Foto: Colourbox

Figur 6: Leiligheter. Lønn.

Basert på de $n = 8$ salgene av leiligheter så langt i år fant *Notar* følgende resultat:

x (årsinntekt, 1000 NOK)	320	450	480	560	610	750	810	880
y (kvadratmeterpris, NOK/m ²)	23 500	36 000	33 000	48 000	45 000	52 000	57 000	53 000

Tabell 3: Årsinntekt x og kvadratmeterpris y .

Ut fra tabell 3 kan man regne ut gjennomsnittene \bar{x} og \bar{y} :

$$\bar{x} = 607.5 \quad , \quad \bar{y} = 43\,437.5 \quad (33)$$

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

$$S_x^2 = 37\,364 \quad , \quad S_y^2 = 133\,388\,393 \quad (34)$$

Den empiriske **kovariansen** mellom x og y er: ¹⁴

$$S_{xy} = 2\,084\,821 \quad (35)$$

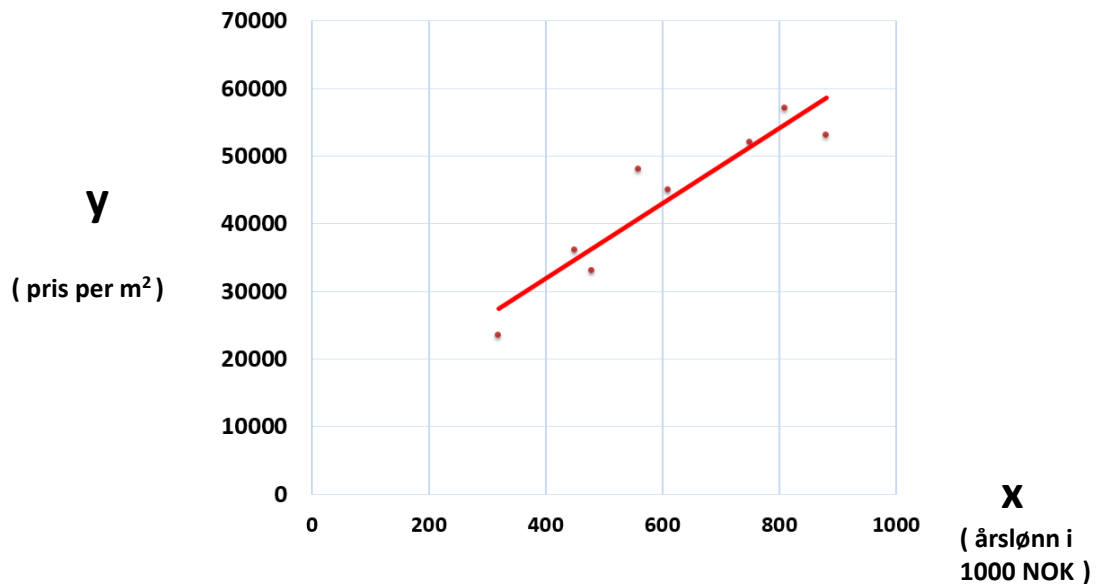
b) Regn ut korrelasjonskoeffisienten R_{xy} for observasjonene i tabell 3.

c) Tolk svaret du fikk for R_{xy} i oppgave **4b**. ¹⁵

¹⁴Størrelsene i lign.(33), (34) og (35) trenger du å ikke regne ut. Bare ta dem for gitt. Benevningen, altså enheten, til disse størrelsene er utelatt.

¹⁵Hva sier den numeriske (tallmessige) verdien av R_{xy} fra oppgave **4b** om graden av korrelasjon mellom observasjonene x og y ?

Pga. resultatene fra oppgave 4a og 4b konkluderer *Notar* med at det kan være hensiktsmessig å finne en eksplisitt lineær sammenheng mellom x og y . Eiendomsmegleren bestemmer seg for å bruke lineær regresjonsanalyse.



Figur 7: Plott av dataene fra tabell 3.

- d) Den rette linjen i figur 7 viser minste kvadraters regresjonslinje for observasjonene x og y . Finn et analytisk uttrykk for denne lineære **regresjonslinjen**.¹⁶
- e) La oss anta at en person med en årinntekt på 1 million NOK skal kjøpe seg en leilighet. Dersom vi antar at personen kjøper en leilighet med en kvadratmeterpris i henhold til regresjonslinjen, hva blir da kvadratmeterpris på leiligheten?¹⁷

¹⁶Dvs. finn formelen for linjen via setningen for minste kvadraters lineære regresjonslinje. Se gjerne formelsamling.

¹⁷Bruk regresjonslinjen som du fant i oppgave 4d for å løse denne oppgaven.

- f) Anta at en person har kjøpt ei leilighet som har en kvadratmeterpris på er 30 000 NOK/ m^2 .

Ifølge regresjonslinjen, hva slags årsinntekt har denne kjøperen?

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 R^2 uten å gjøre noe regning “for hånd”. Bare les av fra Excel-utskriften i figur 8 nedenfor. (Bruk 4 desimalers nøyaktighet).

- h) Kommenter svaret i oppgave 4g.¹⁸

1	SUMMARY OUTPUT								
2									
3	<i>Regression Statistics</i>								
4	Multiple R	0,9339							
5	R Square	0,8721							
6	Adjusted R Square	0,8508							
7	Standard Error	4461,4794							
8	Observations	8							
9									
10	ANOVA								
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
12	Regression	1	814289960,1	814289960	40,90922938	0,000687936			
13	Residual	6	119428789,9	19904798,3					
14	Total	7	933718750						
15									
16		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
17	Intercept	9540,7188	5529,418628	1,725447	0,135202977	-3989,281179	23070,71876	-3989,281179	23070,71876
18	X Variable 1	55,7972	8,723716198	6,39603232	0,000687936	34,45100616	77,14333526	34,45100616	77,14333526

Figur 8: Utskrift fra Excel.



¹⁸For en person med en gitt årslønn som skal kjøpe leilighet, vi du si at regresjonslinjen predikerer kvadratmeterprisen til leiligheten i stor eller liten grad? Med stort eller lite presisjonsnivå?

Kapittel 14

Kontinuasjonseksamen 2018

Eksamen i

MAT110 Statistikk 1

Eksamensdag	: Mandag 10. sept. 2018
Tid	: 13:00 – 17:00 (4 timer)
Faglærer/telefonnummer	: Molde + Kristiansund: Per Kristian Rekdal / 924 97 051
Hjelpemidler	: KD + formelsamling fra 2018 + generell ordbok (morsmål/norsk/engelsk i papirformat)
Antall sider inkl. forsiden	: 20
Målform	: Norsk (bokmål)

Noen generelle råd:

- **Skriv rett inn. Ikke bruk så mye tid på kladding.**
- **Det er totalt 4 oppgaver.**

Oppgave 1: (logistikk)

Glamox er en norsk produsent og leverandør av belysning.

Glamox har til enhver tid mange tilbud ute hos kunder som de venter tilbakemelding på. Hensikten med denne oppgaven er å finne relevante sannsynligheter for at tilbud aksepteres basert på to kriterier: **selger** som har laget tilbudet og **verdien** på tilbudet.

1) Selgere:

La oss kun se på 2 selgere hos Glamox, nemlig selgerne 1 og 2.

2) Verdi:

La oss for enkelhets skyld kun skille mellom store og små tilbud, dvs. tilbud som har høy (h) verdi eller lav (\bar{h}) verdi i kroner og øre. Det kan f.eks. være tilbud som er over eller under 10 000 NOK.

Et tilbud blir enten akseptert (a) eller ikke akseptert (\bar{a}). Utfallsrommet har da 8 utfall, se lign.(1) på neste side.



Figur 1: Glamox.

Glamox bruker 350 historiske data for å estimere relevante sannsynligheter. Resultatene av de historiske dataene er at utfall $(1, h, a)$ inntraff 5 ganger, utfall $(1, \bar{h}, a)$ inntraff 40 ganger osv.

Sannsynlighetene p_1, p_2, \dots, p_8 for de 8 mulige utfallene i utfallsrommet er dermed:

$$\Omega = \left\{ \begin{array}{cccc} \overbrace{(1, h, a)}^{p_1 = \frac{5}{350}}, & \overbrace{(1, \bar{h}, a)}^{p_2 = \frac{40}{350}}, & \overbrace{(1, h, \bar{a})}^{p_3 = \frac{25}{350}}, & \overbrace{(1, \bar{h}, \bar{a})}^{p_4 = \frac{80}{350}} \\ \\ \overbrace{(2, h, a)}^{p_5 = \frac{8}{350}}, & \overbrace{(2, \bar{h}, a)}^{p_6 = \frac{68}{350}}, & \overbrace{(2, h, \bar{a})}^{p_7 = \frac{22}{350}}, & \overbrace{(2, \bar{h}, \bar{a})}^{p_8 = ?} \end{array} \right\} \quad (1)$$

hvor utfallet $(1, h, a)$ betyr at selger **1** lager et tilbud med høy (h) verdi som blir akseptert (a). Tilsvarende for de øvrige utfallene.

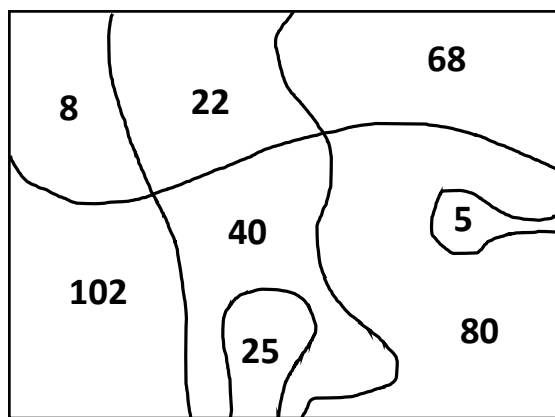
Videre er $N_1 = 150$ tilbud gitt av selger **1** og $N_2 = 200$ gitt av selger **2**. Totalt er da: $N = N_1 + N_2 = 350$ tilbud.

- a) Sannsynlighetfordelingen p_1, p_2, \dots, p_8 i lign.(1) er en gyldig sannsynlighetsfordeling. Vis da at sannsynligheten p_8 er gitt ved:

$$p_8 = \frac{102}{350} \quad (2)$$

- b) Figur 2 viser Venn-diagrammet for de 8 utfallene i utfallsrommet Ω . Tallene i figuren viser antall ganger de respektive utfallene inntreffer, se lign.(1).

Er utfallene i utfallsrommet disjunkte? Gi en *kort* begrunnelse. ¹



Figur 2: Utfallsrom Ω .

¹Én setning er nok.

La oss definere følgende begivenheter:

$$S_1 = \left\{ (1, h, a), (1, \bar{h}, a), (1, h, \bar{a}), (1, \bar{h}, \bar{a}) \right\} \quad (3)$$

$$S_2 = \left\{ (2, h, a), (2, \bar{h}, a), (2, h, \bar{a}), (2, \bar{h}, \bar{a}) \right\} \quad (4)$$

Her er S_1 begivenheten at et tilbud er laget av selger 1.

Her er S_2 begivenheten at et tilbud er laget av selger 2.

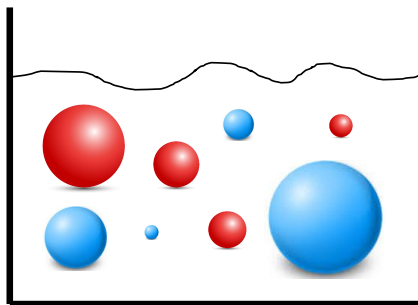
- c) La oss si at hvert av de 8 utfallene i utfallsrommet Ω er assosiert med ei kule. Altså totalt 8 kuler, se figur 3.

La oss videre anta at vi har to typer kuler, røde kuler og blå kuler som tilsvarer begivenhetene S_1 og S_2 henholdsvis.

De røde kulene er assosiert med utfallene til selger 1, og de blå kulene er assosiert med utfallene til selger 2.

La oss anta at tilbudene laget av selger 1 er uavhengige av tilbudene laget av selger 2.

På tross av at tilbudene som de to selgerne har laget er uavhengige av hverandre, hvorfor kan vi ikke bruke urnemodellen til å f.eks. finne sannsynligheten $P(S_1)$ for at et tilbud er gitt av selger 1? ²



Figur 3: Urne.

²Svar kort på denne oppgaven. Én setning er nok. Ingen utregninger behøves. Hva må være oppfylt, i tillegg til uavhengighet, for at urnemodellen skal gjelde? Er det oppfylt i vårt tilfelle?

Begivenhetene S_1 og S_2 definert i lign.(3) and (4).

Begivenhetene er delmengder av det totale utfallsrommet Ω i lign.(1)

Som nevnt på forrige side, S_1 er begivenheten at et tilbud er laget av selger 1.

Tilsvarende for selger 2.

- d) På tilsvarende måte som for begivenhetene S_1 og S_2 i lign.(3) and (4), skriv ned definisjonen av begivenheten for aksepterte tilbud A og begivenheten for høye tilbud H .³

Man kan altså ikke bruke urnemodellen, men man kan likevel finne sannsynligheten $P(A)$ på følgende måte, via lign.(1):

$$\begin{aligned} \underline{P(A)} &= p_1 + p_2 + p_5 + p_6 \\ &= \frac{5}{350} + \frac{40}{350} + \frac{8}{350} + \frac{68}{350} = \frac{121}{350} = \underline{0.3457} \end{aligned} \tag{5}$$

- e) Finn sannsynligheten $P(H)$ ved å bruke tilsvarende metode som i lign.(5).

³Her skal du altså bare skrive ned svaret. Ingen utregninger behøves.

La oss nå se på et tilbud med høy (h) verdi laget av selger **1**. Den betingede sannsynligheten for at et slikt tilbud blir akseptert er gitt ved den generelle multiplikasjonssetningen:

$$P(A | \overbrace{S_1 \cap H}^{\text{vet}}) = \frac{P(S_1 \cap H \cap A)}{P(S_1 \cap H)} \quad (6)$$

f) Bruk lign.(6) og vis at: ⁴

$$P(A | S_1 \cap H) = \frac{5}{30} \quad (= 0.1667) \quad (7)$$

g) Gi en tolkning av $P(A | S_1 \cap H)$, dvs. forklar *kort* hva det betyr på ”godt norsk”.



⁴Tips: Finn først $P(S_1 \cap H \cap A)$ og $P(S_1 \cap H)$ ved å bruke tilsvarende metode som i lign.(5).

Oppgave 2: (logistikk - prognostisering)

La oss fortsette med Glamox-beskrivelsen fra oppgave 1.

For å unngå mye repeterende regning, la oss kun se på 4 nye tilbud.

For enkelhets skyld, la oss anta at alle disse 4 nye tilbudene er gitt av selger 2 og at alle har høy (h) verdi.

Anta at verdien til tilbud nr. i er v_i (i NOK), se tabell 1:

Tilbud nr. i ($i = 1,2,3,4$)	1	2	3	4
Høyt (h) tilbud	h	h	h	h
Selger S_2	2	2	2	2
Verdi v_i ($i=1,2,3,4$) (NOK)	40 000	33 000	105 000	75 000

Tabell 1: Nye tilbud.



Figur 4: Glamox.

La oss introdusere en stokastisk variabel som sier noe om et tilbud aksepteres eller ikke:

$$Y_i = \begin{cases} 1 & , \text{ dersom kunde aksepterer tilbud nr. } i \\ 0 & , \text{ dersom kunde ikke aksepterer tilbud nr. } i \end{cases} \quad (8)$$

hvor $i = 1, 2, 3, 4$ indikerer tilbud nr. i .

Aksept-sannsynligheten for et gitt tilbud er laget av selger 1 og har høy (h) verdi fant vi i oppgave **1f**. For enkelhets skyld, la oss kalle denne sannsynligheten for p , dvs. $p = P(A|S_1 \cap H)$, altså:

$$p = \frac{5}{30} = 0.1667 \quad (9)$$

Den stokastiske variabelen Y definert ved:

$$Y = Y_1 + Y_2 + Y_3 + Y_4 \quad (10)$$

beskriver da antall tilbud som aksepteres.

Anta at tilbudene er uavhengige av hverandre, dvs. at Y_i 'ene er uavhengige.

- a) Forklar hvorfor Y er **binomisk** fordelt, dvs. forklar hvorfor $Y \sim \text{Bin}\left[\underbrace{4}_{n=4}, \underbrace{0.1667}_{p=0.1667}\right]$.⁵
- b) Hva er sannsynligheten for at to tilbud blir akseptert, dvs. hva er $P(Y = 2)$?
- c) Finn tallverdien for $E[Y]$ og gi en *kort* tolkning på hva $E[Y]$ betyr på godt norsk”.⁶

⁵Hvilke 4 kriterier må være oppfylt for et en forsøksserie skal være binomisk? Er disse oppfylt i vårt tilfelle?

⁶Én setning er nok.

Summen av verdien til de aksepterte tilbudene er gitt ved:

$$V = v_1 Y_1 + v_2 Y_2 + v_3 Y_3 + v_4 Y_4 \quad (11)$$

hvor v_i er verdien på tilbud nr. i .

d) Vis at forventningsverdien til verdien til de aksepterte tilbudene er: ⁷

$$E[V] = (v_1 + v_2 + v_3 + v_4) p \quad (13)$$

e) Bruk lign.(13) og regn ut tallverdien til $E[V]$. ⁸

f) Vis da at variansen til verdien til de aksepterte tilbudene er: ⁹

$$Var[V] = (v_1^2 + v_2^2 + v_3^2 + v_4^2) p(1 - p) \quad (15)$$

⁷Tips: Siden $Y_i \sim \text{Bin}[\overbrace{1}^{n_i=1}, p]$ så er:

$$E[Y_1] = E[Y_2] = E[Y_3] = E[Y_4] = p \quad (12)$$

⁸Bruk tallene for v_i i tabell 1 på side 8.

⁹Siden $Y_i \sim \text{Bin}[\overbrace{1}^{n_i=1}, p]$ så er:

$$Var[Y_1] = Var[Y_2] = Var[Y_3] = Var[Y_4] = p(1 - p) \quad (14)$$

g) Dersom selger 1 hadde lagt 200 tilbud istedet for bare 4 så er

$$V = V_1 + V_2 + \dots + V_{200} \quad (16)$$

$$= v_1 Y_1 + v_2 Y_2 + \dots + v_{200} Y_{200} \quad (17)$$

verdien på de aksepterte tilbudene, hvor $V_i = v_i Y_i$ og $i = 1, 2, 3, \dots, 200$.

La oss si at alle tilbudene har høy verdi (h) slik at aksept-sannsynligheten p til tilbudene er den samme, men at verdien v_i på de $i = 1, 2, 3, \dots, 200$ tilbudene er forskjellig.

Dersom vi i tillegg fortsatt antar at alle tilbudene er uavhengige, hvorfor kan vi ikke da anvende sentralgrenseteoremet og si at V er tilnærmet normalfordelt?

■

Oppgave 3: (logistikk og økonomi)

La oss anta at det er $n = 4$ Kiwi-butikker i Molde.

Alle disse butikkene selger bananer når det er sesong.

La oss introdusere en stokastisk variabel D_i (“demand”), hvor:

$$D_i = \text{etterspørsel av antall kg bananer per dag i butikk nr. } i \quad (18)$$

med $i = 1, 2, 3, 4$.

Den totale etterspørselen per dag for de 4 butikkene er da beskrevet av:

$$D_{\text{tot}} = D_1 + D_2 + D_3 + D_4 \quad (19)$$

Anta videre at de stokastiske variablene D_i har samme sannsynlighetsfordeling for alle dagene i sesongen, dvs. fordelingen til etterspørselen er den samme hver dag for en gitt butikk.



Figur 5: Kiwi-butikker og bananer.

Innkjøperne hos Kiwi har funnet ut at forventet etterspørsel og tilhørende varians per dag for de 4 aktuelle butikkene ($i = 1, 2, 3, 4$) er som oppgitt i tabell 2:

i	1	2	3	4
$E[D_i]$	40	100	60	20
$Var[D_i]$	64	144	352	16

Tabell 2: Forventning $E[D_i]$ og varians $Var[D_i]$ for $i = 1, 2, 3, 4$.

- a) Vis at den totale forventede etterspørselen $E[D_{\text{tot}}]$ per dag for de 4 butikkene er:

$$E[D_{\text{tot}}] = 220 \quad (20)$$

kg med bananer.

Anta at etterspørselen av bananer i de 4 butikkene er uavhengige.

- b) Regn ut den totale variansen $Var[D_{\text{tot}}]$ per dag for de 4 butikkene.

Anta videre at alle de fire stokastiske variablene D_i er normalfordelt, dvs. anta:

$$D_i \sim N[E[D_i], Var[D_i]] \quad (21)$$

hvor $E[D_i]$ og $Var[D_i]$ er som oppgitt i tabell 2 for $i = 1, 2, 3, 4$.

c) Hva slags fordeling har da D_{tot} ? ¹⁰

¹⁰Se side 108 i formelsamlingen fra 2018. Definisjonen at D_{tot} er gitt i lign.(19).

Bananene kastes

La oss anta at bananene kun har holdbarhet én dag.

Det betyr at dersom Kiwi ikke får solgt bananene en gitt dag så må de kaste dem når de stenger butikken om kvelden.

La oss nå kun se på butikk nr. 3.

Fortjenesten F_3 per dag for denne butikken er:

$$F_3 = \underbrace{p S_3}_{\text{inntekt}} - \underbrace{c q_3}_{\text{kostnad}}, \quad (22)$$

hvor

$$\begin{aligned} S_3 &= \min(D_3, q_3) \\ &= \text{antall } kg \text{ med bananer som } \underline{\text{selges}} \text{ i butikk nr. 3 per dag} \end{aligned} \quad (23)$$

$$q_3 = \text{antall } kg \text{ bananer som innkjøperne hos Kiwi } \underline{\text{bestiller}} \text{ per dag for butikk nr. 3 (} \underline{\text{order quantity}} \text{)} \quad (24)$$

$$c = 8 \text{ NOK} \quad (\underline{\text{kostnaden}} \text{ per } kg \text{ bananer for Kiwi)} \quad (25)$$

$$p = 12.9 \text{ NOK} \quad (\underline{\text{pris}} \text{ per } kg \text{ som Kiwi selger bananene for)} \quad (26)$$

Det optimale bestillingskvantumet q_3^* som gir maksimal forventet fortjeneste av lign.(22) er da bestemt av:

$$P(D_3 \leq q_3^*) = 1 - \frac{c}{p} \quad (27)$$

hvor

$$q_3^* = \text{antall } kg \text{ med bananer som Kiwi må bestille per dag for å } \underline{\text{maksimere}} \text{ forventet fortjeneste til butikk nr. 3} \quad (28)$$

d) Finn q_3^* .¹¹

¹¹**Tips 1:** Bruk de oppgitte verdiene for c og p samt lign.(27).

Tips 2: I denne oppgaven slipper du å gjøre tabelloppslag. Bruk verdien $z_0 = -0.30$, altså negativ verdi.

Tips 3: Husk at $E[D_3]$ og $\sigma[D_3] = \sqrt{\text{Var}[D_3]}$ finnes i tabell 2 på side 14.

Redusert pris

I stedet for å kaste bananene kan Kiwi gjøre en avtale med et firma som kjøper bananene som Kiwi ikke får solgt en gitt dag. Dette firmaet får tilbud om å kjøpe bananene til redusert pris p_{red} .

Antall kg med bananer som ikke blir solgt en gitt dag for butikk nr. 3, dvs. usolgte bananer, er da beskrevet av følgende stokastiske variabel:

$$U_3 = \begin{cases} q_3 - D_3 & , \quad q_3 \geq D_3 \\ 0 & , \quad q_3 < D_3 \end{cases} \quad (29)$$

Fortjenesten F_3 per dag for butikk 3 blir da en utvidelse av lign.(22), nemlig:

$$F_3 = \underbrace{p S_3 + p_{\text{red}} U_3}_{\text{inntekt}} - \underbrace{c q_3}_{\text{kostnad}} \quad (30)$$

Det nye, optimale bestillingskvantumet q_3^* som gir maksimal forventet fortjeneste av lign.(30) er da bestemt av: ¹²

$$P(D_3 \leq q_3^*) = \frac{p - c}{p - p_{\text{red}}} \quad (31)$$

- e) Anta at Kiwi kjøper inn $q_3^* = 70 \text{ kg}$ bananer per dag til butikk nr. 3, hvilken redusert pris p_{red} må da Kiwi og dette firmaet bli enige om dersom Kiwi skal maksimere den forventede fortjenesten F_3 ? ¹³



¹²Du skal ikke vise lign.(31). Bare ta den for gitt.

¹³ Dette er en litt vanskelig oppgave. Hopp over den dersom du har dårlig tid. Tips: Bruk lign.(31) og løs med hensyn på p_{red} .

Oppgave 4: (økonomi)

I fjor opplevde teleselskapet Telenor at $\overbrace{\text{aksjekursen}}^{\text{pris på aksje}}$ økte i en periode. La oss se på 5 dager med omsetning i denne aktuelle perioden.

$$x = \text{dag nr. } x, \text{ hvor } x = 1, 2, 3, 4, 5 \quad (32)$$

$$y = \text{aksjekurs (i NOK)} \quad (33)$$

Aksjekurs, dvs. aksjepris, for dag 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 6: Omsetningsdag nr. x og aksjekurs (pris) y .

Ut fra tabellen i figur (6) kan man regne ut den empiriske variansen for aksjekursen y og den empiriske variansen for x . Resultatet er:

$$S_y^2 = 133 \text{ NOK}^2, \quad S_x^2 = 2.5 \quad (34)$$

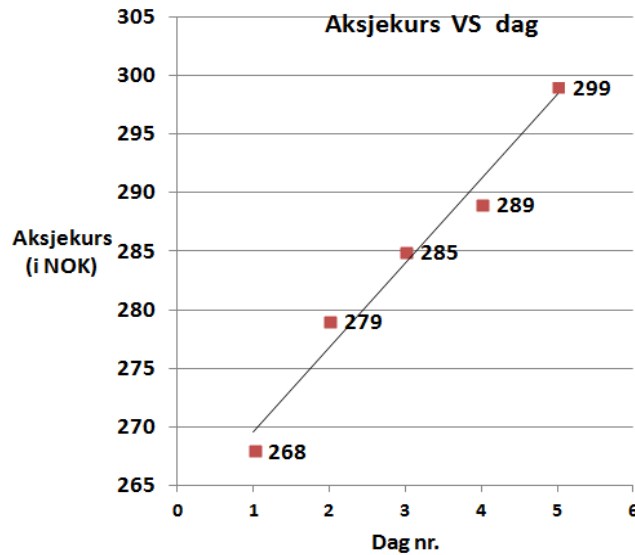
Den empiriske **k**ovariansen mellom x og y er:

$$S_{xy} = 18 \text{ NOK} \quad (35)$$

(Størrelsene i lign.(34) og (35) trenger du å ikke regne ut. Bare ta dem for gitt).



Figur 7: Telenor.



Figur 8: Plott av dataene fra tabellen i figur (6).

- a) Grafen i figur (8) 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.¹⁵
- ii) På generelt grunnlag, hvordan tolker du parameteren $\hat{\beta}$ i en *lineær* regresjonslinje?¹⁶
- iii) Spesielt, hvordan tolker du parameteren $\hat{\beta}$ i regresjonslinjen fra oppgave 4a?
- c) Anta trenden på aksjemarkedet for Telenor fortsetter på samme måte som de 5 dagene i en periode fremover. Dersom Telenor ønsker å *predikere* 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?

¹⁴Dvs. finn formelen for linjen. Se gjerne formelsamlingen for formel for minste kvadraters regresjonslinje.

¹⁵Tips: estimat / prediksjon.

¹⁶Tips: stigningstall.

d) Forklaringsstyrken 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 forklaringsstyrken R^2 kan man bruke et dataprogram, f.eks. Excel, i stedet for å regne ut R^2 “for hånd” via definisjonen.

e) Finn forklaringsstyrken uten å gjøre noe regning “for hånd”. Bare les av fra Excel-utskriften i figur (9) nedenfor. (Bruk 4 desimalers nøyaktighet).

f) Kommenter svaret i oppgave 4e.¹⁷

A	B	C	D	E	F	G	H
SUMMARY OUTPUT							
<i>Regression Statistics</i>							
Multiple R	0,987135295						
R Square	0,97443609						
Adjusted R Square	-1,666666667						
Standard Error	2,12916259						
Observations	1						
<i>ANOVA</i>							
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>		
Regression	5	518,4	103,68	114,3529	#NUM!		
Residual	3	13,6	4,533333				
Total	8	532					
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95,0%</i>
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 9: Utskrift fra Excel.

¹⁷For en gitt dag, vi du si at regresjonslinjen predikerer aksjekursen i stor eller liten grad? Med stort eller lite presisjonsnivå?

Kapittel 15

Testeksamen 2019

Testeksamen i

MAT110 Statistikk 1

Versjon 03

Eksamensdag	: Mai 2019
Tid	: 09:00 – 13:00 (4 timer)
Faglærer/telefonnummer	: Molde + Kristiansund: Per Kristian Rekdal / 924 97 051
Hjelpemidler	: KD + formelsamling + generell ordbok (morsmål/norsk/engelsk i papirformat)
Antall sider inkl. forsiden	: xx
Målform	: Norsk (bokmål)

Noen generelle råd:

- Skriv rett inn. Ikke bruk så mye tid på kladding.
- Det er totalt 4 oppgaver.

Hver uke er alle oppdrettsanlegg i Norge pålagt å rapportere til forvaltningsmyndighetene om hvor mange voksne hunnlus som i gjennomsnitt er på alle fiskene i anlegget. Hensikten med tellingen er tofoldig:

1. Ha en oversikt over spredningen av voksne hunnlus i norske oppdrettsanlegg.
2. Kunne regulere oppdrettsanlegg som har for høye lusetall. ¹

Forvaltningsmyndighetene opererer derfor med en **lusegrense**, a , som det sanne lusetallet må ligge under:

$$\begin{aligned} a &= \text{lusegrense} \\ &= 0.2 \text{ voksne hunnlus i gjennomsnitt per fisk per anlegg} \quad (1) \end{aligned}$$

Veterinærinstituttet har skrevet en rapport med tittelen:

” Telling av lakselus - Hvordan forstå og håndtere usikkerheten i telleresultatene”

som handler om hvordan den ukentlige tellingen av lakselus kan brukes til å forvalte lusegrensa. ²



Figur 1: Oppdrettsanlegg. Lakselus. Rapport.

¹Lusetall = forventet antall voksne hunnlus per fisk per oppdrettsanlegg.

²Denne rapporten finnes [her](#). Den er også lastet opp på Canvas under ”materiale for emnet”. Vi anbefaler at man tar en titt på denne rapporten. Mye av eksamen er basert på denne rapporten med hensikt å aktualisere eksamen.

Oppgave 1: (sannsynlighetsteori - telling av lakselus)

Anta nå at vi ser på et oppdrettsanlegg med kun **3 merder**.

En merd er en innhengning for oppdrettsfisk, se figur 2.

Vi gjør følgende *eksperiment*:

1. Nummerer de 3 merdene fra 1 til 3.
2. Én fisk velges tilfeldig fra hver av de 3 merdene.
La i representere den valgte fisken fra merd nr. i , for $i = 1, 2, 3$.
3. For hver valgte fisk registreres det om den har lakselus på seg eller ikke.
Notasjon:

$$l_i = \text{en tilfeldig valgt fisk fra merd nr. } i \text{ har lakselus} \quad (2)$$

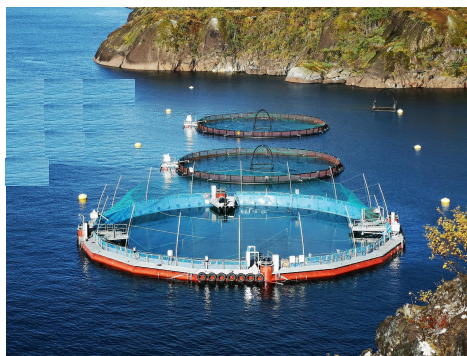
$$\bar{l}_i = \text{en tilfeldig valgt fisk fra merd nr. } i \text{ har } \underline{\text{ikke}} \text{ lakselus} \quad (3)$$

I vårt tilfelle er altså:

$$\text{eksperiment} = \text{sekvens med 3 forsøk} \quad (4)$$

Anta at vi kjenner:

$$p = \underline{\text{sannsynligheten}} \text{ for at en vilkårlig valgt fisk har } \underline{\text{lakselus}} \quad (5)$$



Figur 2: 3 merder.

Definer et utfall u av eksperimentet som:

$$u = (a_1, a_2, a_3) \tag{6}$$

hvor $a_i \in \{l_i, \bar{l}_i\}$ for $i = 1, 2, 3$.

- a) Definer utfallsrommet Ω_3 for eksperimentet. ³
Bruk notasjonen som i lign.(6).

- b) Er utfallsrommet Ω_3 diskret eller kontinuert?
Begrunn svaret kort.

Definer begivenheten:

$$S = \text{minst to av de trekte fiskene har lakselus} \tag{7}$$

- c) Beskriv begivenheten S som en delmengde av utfallsrommet Ω_3 .

- d) Finn et uttrykk for sannsynligheten for begivenheten S beskrevet i lign.(7).
Anta $p = 0.5$. Regn ut sannsynligheten for begivenheten.

³Skriv gjerne opp utfallene eksplisitt.

For de resterende oppgavene ser vi på et annet oppdrettsanlegg som har 10 merder. Anta de gjør det samme eksperimentet men at de trekker 10 fisker istedet for 3, en fra hver merd, og registrerer om fiskene har lakselus eller ikke. Definer et utfall u av eksperimentet som:

$$u = (a_1, a_2, \dots, a_{10}) \quad (8)$$

hvor $a_i \in \{l_i, \bar{l}_i\}$ for $i = 1, 2, \dots, 10$.

Anta $p = 0.5$ for resten av oppgavene.

e) Definer utfallsrommet for Ω_{10} for dette eksperimentet. ⁴

f) Er sannsynlighetsmodellen uniform? Begrunn svaret.

Hadde sannsynlighetsmodellen vært uniform dersom vi hadde hatt en annen verdi enn $p = 0.5$? Begrunn svaret.

g) Regn ut sannsynligheten for at nøyaktig 7 av de 10 trekte fiskene har lakselus. ⁵

h) Anta vi får vite at fisk nr. 1 og fisk nr. 10 har lakselus.

Bruk betinget sannsynlighet for å regne ut sannsynligheten for at minst 8 av de 10 trekte fiskene har lakselus.



⁴Hint: *Ikke* skriv opp utfallene eksplisitt - bruk en formel for å definere utfallsrommet, dvs. $\Omega_{10} = \{ \text{formel} \}$.

⁵Hint: Bruk urnemodellen.

Oppgave 2: (stokastiske variabler og forventningverdi - telling av lakselus)

Avsnitt 4.1 i rapporten til Veterinærinstituttet:

”For å beskrive fordelingen av lus mellom fiskene brukte vi et datasett fra 20 oppdrettsanlegg i Rogaland, med lusetall på enkeltfisknivå fra lusetellinger utført som ledd i daglig drift. Tellingene var utført fra og med uke 1 i 2012 til og med uke 37 i 2017. Til sammen hadde vi lusetellinger fra 352 228 enkeltfisker.”

Fra dette datasettet ble det trekt ut et tilfeldig utvalg av fisker som kom fra merder med gjennomsnittlig antall hunnlus per fisk mellom 0.1 og 0.3. Det tilfeldige utvalget utgjorde 44 001 fisker. Gjennomsnittlig antall hunnluser per fisk \bar{x} for dette utvalget samt tilhørende empirisk varians s_x^2 er vist i tabell 1. ⁶

Antall fisker N	\bar{x}	s_x^2
44 001	0.1970	0.2970

Tabell 1: Nøkkeltall for utvalgte merder.



Figur 3: Rapport.

⁶Disse nøkkeltallene kommer fra rapporten og derfor faktiske nøkkeltall.

Populasjonsvariabel:

$$X = \text{antall hunn lus på en tilfeldig valgt fisk fra et vilkårlig oppdrettsanlegg} \quad (9)$$

med lusetall mellom 0.1 og 0.3

Forsøksvariabler:

$$X_i = \text{antall hunn lus på fisk nr. } i \text{ fra utvalget} \quad (10)$$

hvor

$$i = 1, 2, \dots, N \quad (11)$$

med $N = 44\,001$ antall fisker.

I denne oppgaven antar vi en statistisk modell for forsøksvariablene basert på den *empiriske* fordelingen, dvs. vi antar:

$$X_1, X_2, \dots, X_N \stackrel{\text{i.i.d.}}{\sim} X \sim P(X = j) \quad (12)$$

hvor

$$j = 0, 1, 2, 3, 4, 5 \quad (\text{antall lakseluser}) \quad (13)$$

med fordelingen: ⁷

	$j = 0$	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
$P(X = j)$	0.85	0.12	0.0185	0.0070	0.0035	0.0010

Tabell 2: Sannsynlighetsfordeling $P(X = j)$.

⁷Vi antar her for enkelthetsskyld at det ikke ble registrert fisker med flere enn 5 lakseluser. Tallene i tabell 2 er fiktive da selve rådataene ikke er tilgjengelige.

- a) Forklar kort med ord hva sannsynligheten $P(X \leq j)$ betyr. ⁸
- b) Vis at $P(X = j)$ er en gyldig sannsynlighetsfordeling.
- c) Regn ut forventningsverdien $E[X]$.
- d) Regn ut variansen $Var[X]$ og standardavviket $\sigma[X]$.

⁸En setning er nok.

Et av de vanligste virkemidlene for å redusere lusetallet på er å gi fiskene medisin i fôret.⁹ Anta at vi har funnet følgende sammenheng mellom mengden M av legemiddelet, som må blandes i fôret, og antall hunnlus per fisk:

$$M = aX^2 + bX + c \quad (14)$$

hvor M er antall gram medisin per 100 gram fôr, og

$$a = 1.2 \quad (15)$$

$$b = 2.3 \quad (16)$$

$$c = 1.4 \quad (17)$$

- e) Regn ut $E[M]$.
Forklar med ord hva $E[M]$ betyr.



Figur 4: Fôr.

⁹Fôr = mat for fisken.

Hver uke må hvert oppdrettsanlegg telle antall voksne hunnlus på 50 tilfeldig valgte fisker og rapportere lusetallet til Mattilsynet. Basert på gjennomsnittet av antall hunnlus per fisk for disse 50 fiskene må Mattilsynet vurdere om hvorvidt lusegrensa er overskredet eller ikke.

Anta vi ser på ett av disse oppdrettsanleggene for en vilkårlig uke t .

Forsøksvariablene er

$$X_{kt} = \text{antall hunnlus på valgt fisk nr. } k \text{ i uke } t \quad (18)$$

hvor

$$k = 1, 2, \dots, n \quad (\text{antall fisk}) \quad (19)$$

med $n = 50$ antall fisk.

Anta videre at:

$$X_{1t}, X_{2t}, \dots, X_{50t} \stackrel{i.i.d.}{\sim} X \sim P(X = j) \quad (20)$$

hvor *i.i.d.* betyr uavhengige identiske fordelte variabler og hvor X er definert ved lign.(9) med fordeling som vist i tabell 2.

- f) Regn ut sannsynligheten for at det målte lusetallet \bar{X}_t til oppdrettsanlegget i uke t er lik eller ligger under lusegrensa a , dvs. regn ut:

$$P(\bar{X}_t \leq a) \quad (21)$$

hvor $a = 0.2$.

■

Oppgave 3: (statistisk modell og estimatorer)

I rapporten formulerte forfatterne to kandidater som mulig statistisk modell for forsøksvariablene X_i fra lign.(10), den ene var den velkjente Poisson-familien mens den andre var en annen diskret fordeling kalt *Negativ Binomial*:

Alternativ 1: Poisson (X og X_i er definert i lign.(9) og lign.(10))

$$X_1, X_2, \dots, X_N \stackrel{\text{i.i.d.}}{\sim} X \sim \text{Poi}[\lambda] \quad (22)$$

hvor

$$\lambda = \text{forventningsverdien} \quad (23)$$

Dersom $X \sim \text{Poi}[\lambda]$ så gjelder:

$$\begin{aligned} E[X] &= \lambda \\ \text{Var}[X] &= \lambda \end{aligned} \quad (24)$$

hvor $\lambda > 0$.

Alternativ 1: Negativ binomial

$$X_1, X_2, \dots, X_N \stackrel{\text{i.i.d.}}{\sim} X \sim \text{NBin}[\lambda, \theta] \quad (25)$$

hvor

$$\lambda = \text{forventningsverdien} \quad (26)$$

$$\theta = \text{spredningsparameteren} \quad (27)$$

Dersom $X \sim \text{NBin}[\lambda, \theta]$ så gjelder:

$$E[X] = \lambda \quad (28)$$

$$\text{Var}[X] = \lambda + \frac{\lambda^2}{\theta} \quad (29)$$

hvor $\lambda > 0$ og $\theta > 0$.

Poissonfordelingen, også kalt ”loven om sjeldne begivenheter”, er en statistisk fordeling som kan *utledes* dersom en del *forutsetninger* er oppfylt.

En viktig forutsetning som bør være oppfylt i den sammenheng er det som kalles *homogenitet* eller *ekvispredning*, dvs. at forventningsverdien og variansen er lik. Med $X \sim \text{Poi}[\lambda]$ har vi da:

$$E[X] = \text{Var}[X] = \lambda \quad (30)$$

Dersom:

$$E[X] < \text{Var}[X] : \quad \text{overspredning} \quad (31)$$

$$E[X] > \text{Var}[X] : \quad \text{underspredning} \quad (32)$$

Det kan være flere grunner til *overspredning* i et datasett, men den mest vanlige er at det er *heterogenitet* i dataene. Heterogenitet betyr i prinsippet at dataene ikke er fremskaffet under like betingelser/omgivelser.

Dersom dataene viser tegn til *overspredning*, kan Negativ Binomial modellen være en passende modell å bruke siden den har en ekstra parameter i *spredningsparameteren* θ definert i lign.(27).

Fra lign.(28) og (29) ser vi at θ gjør at Negativ Binomial fordelingen har rom for *overspredning*.

Siden datasettet i rapporten faktisk viste tegn til *overspredning*, valgte forfatterne å bruke Negativ Binomial modellen.

Vi skal videre i denne oppgaven finne estimatorer for parameterne λ og θ og et asymptotisk $(1 - \alpha) 100\%$ konfidensintervall for λ for Negativ Binomial modellen definert i lign.(25).

Anta vi definerer følgende estimator $\hat{\lambda}$ for parameteren λ :

$$\hat{\lambda} = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \quad (33)$$

- a) Vis at $\hat{\lambda}$ i lign.(33) er forventningsrett.
Hva er realiseringen av $\hat{\lambda}$?

Vi skal nå konstruere en estimator $\hat{\theta}$ for θ .
Først definerer vi parameteren:

$$\sigma^2 = \lambda + \frac{\lambda^2}{\theta} \quad (34)$$

Siden $Var[X] = \sigma^2$, kan vi bruke S_X^2 som estimator $\hat{\sigma}^2$ for σ^2 .

- b) Finn et uttrykk for $\hat{\theta}$ basert på lign.(34).
Regn ut realiseringen av $\hat{\theta}$?

c) Vis at ¹⁰

$$[LB_N^\lambda, UB_N^\lambda] = \left[\bar{X} - z_{1-\alpha/2} \sqrt{\frac{\bar{X} + \frac{\bar{X}^2}{\hat{\theta}}}{N}}, \bar{X} - z_{\alpha/2} \sqrt{\frac{\bar{X} + \frac{\bar{X}^2}{\hat{\theta}}}{N}} \right] \quad (35)$$

er et asymptotisk $(1 - \alpha)$ 100 % konfidensintervall for λ ,
hvor $z_{\alpha/2}$ og $z_{1-\alpha/2}$ er de tilhørende kvantilene til normalfordelingen.

d) Vis at konfidensintervallet i lign.(35) kan skrives om på formen: ¹¹

$$[LB_N^\lambda, UB_N^\lambda] = \left[\bar{X} - z_{1-\alpha/2} \frac{S_X}{\sqrt{N}}, \bar{X} - z_{\alpha/2} \frac{S_X}{\sqrt{N}} \right] \quad (36)$$

■



Figur 5: Lakselus.

¹⁰Se definisjon av asymptotiske $(1 - \alpha)$ 100 % konfidensintervall i formelsamlingen.

¹¹Bruke uttrykket for $\hat{\theta}$ fra oppgave 2b.

Oppgave 4: (krav som indikerer overskridelse av lusegrensa)

Vi skal i denne oppgaven se på kravene foreslått i rapporten som indikerer overskridelse av lusegrensa $a = 0.2$.

I oppgave 2 definerte vi forsøksvariablene fra en telling av 50 fisker på et vilkårlig oppdrettsanlegg en vilkårlig uke t :

$$X_{kt} = \text{antall hunnlus på valgt fisk nr. } k \text{ i uke } t \quad (37)$$

hvor

$$k = 1, 2, \dots, n \quad (\text{antall fisk}) \quad (38)$$

med $n = 50$ fisker.

Anta nå at vi bruker Negativ Binomial-modellen i lign.(25) fra oppgave 3 som statistisk modell for forsøksvariablene i lign.(37), dvs. anta:

$$X_{1t}, X_{2t}, \dots, X_{nt} \stackrel{\text{i.i.d.}}{\sim} X \sim \text{NBin}[\lambda, \theta] \quad (39)$$

Oppdrettsanlegget regner ut så gjennomsnittet \bar{X}_t for uke t og sender resultatet til Mattilsynet. Mattilsynet vurderer deretter, på bakgrunn av den målte verdien for \bar{X}_t , om lusegrensa er overskredet eller ikke.

Veterinærinstituttet har foreslått to krav på \bar{X}_t som Mattilsynet kan bruke når de skal vurdere om lusegrensa er overskredet eller ikke. Hovedmålet med disse kravene er å unngå feile konklusjoner av typen ”*falske positive*”, dvs.

konkludere med at lusetallet er over lusegrensa, gitt at det sanne lusetallet ligger under

siden det å overskride lusegrensa har store økonomiske konsekvenser for oppdretterne.

De to kravene som Veterinærinstituttet foreslår er:

1. For å unngå *falske positiviteter* så ønsker man å teste \bar{X}_t mot et kriterium på formen

$$\bar{X}_t > b \tag{40}$$

som er slik at det kun er $(1 - 0.96) 100\% = 4\%$ sannsynlighet for at \bar{X}_t er ligger over lusegrensen $a = 0.2$ gitt at det sanne lusetallet er under grensen, dvs. gitt at $E[X] < 0.2$.

2. Lusegrensa $a = 0.2$ anses som overskredet dersom lusetellingen ligger over lusegrensa 4 uker på rad, gitt at det sanne lusetallet ligger under, dvs. gitt at $E[X] < 0.2$.

a) Formuler krav nr. 1 matematisk. ¹²

b) Formuler et matematisk uttrykk for *sannsynligheten* for at krav 2 inntreffer.



¹²Dette er en A-oppgave.

Kapittel 16

Hovedeksamen 2019



Eksamen i

MAT110 Statistikk 1

Eksamensdag	: Tirsdag 21. mai 2019
Tid	: 09:00 – 13:00 (4 timer)
Faglærer/telefonnummer	: Molde + Kristiansund: Per Kristian Rekdal / 924 97 051
Hjelpemidler	: KD + formelsamling + generell ordbok (morsmål/norsk/engelsk i papirformat)
Antall sider inkl. forsiden	: 16
Målform	: Norsk (bokmål)

Noen generelle råd:

- **Skriv rett inn. Ikke bruk så mye tid på kladding.**
- **Det er totalt 4 oppgaver.**

Hver uke er alle oppdrettsanlegg i Norge pålagt å rapportere til forvaltningsmyndighetene om hvor mange voksne hunnlus som i gjennomsnitt er på alle fiskene i anlegget. Hensikten med tellingen er tofoldig:

1. Ha en oversikt over spredningen av voksne hunnlus i norske oppdrettsanlegg.
2. Kunne regulere oppdrettsanlegg som har for høye lusetall. ¹

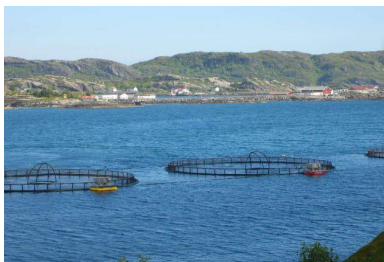
Forvaltningsmyndighetene opererer derfor med ei lusegrense, a , som det sanne lusetallet må ligge under:

a	=	lusegrense	
	=	0.5 voksne hunnlus i gjennomsnitt per fisk per anlegg	(1)

Veterinærinstituttet har skrevet en rapport med tittelen:

” Telling av lakselus - Hvordan forstå og håndtere usikkerheten i telleresultatene”

som handler om hvordan den ukentlige tellingen av lakselus kan brukes til å forvalte lusegrensa.



Figur 1: Oppdrettsanlegg. Lakselus. Rapport.

¹Lusetall = forventet antall voksne hunnlus per fisk per oppdrettsanlegg.

Oppgave 1: (sannsynlighetsteori - telling av lakselus)

Anta nå at vi ser på et oppdrettsanlegg med kun **4 merder**. En merd er en innhengning for oppdrettsfisk, se figur 2. Vi gjør følgende *eksperiment*:

1. Nummerer de 4 merdene fra 1 til 4.
2. Én fisk velges tilfeldig fra hver av de 4 merdene.
La i representere den valgte fisken fra merd nr. i , for $i = 1, 2, 3, 4$.
3. For hver valgte fisk registreres det om den har lakselus på seg eller ikke.
Notasjon:

$$l_i = \text{en tilfeldig valgt fisk fra merd nr. } i \text{ har lakselus} \quad (2)$$

$$\bar{l}_i = \text{en tilfeldig valgt fisk fra merd nr. } i \text{ har } \underline{\text{ikke}} \text{ lakselus} \quad (3)$$

I vårt tilfelle er altså:

$$\text{eksperiment} = \text{sekvens med 4 forsøk} \quad (4)$$

Anta at vi kjenner:

$$p = \underline{\text{sannsynligheten}} \text{ for at en vilkårlig valgt fisk har } \underline{\text{lakselus}} \quad (5)$$



Figur 2: 4 merder.

Utfallsrommet Ω for eksperimentet er:

$$\Omega = \left\{ \begin{array}{l} (\bar{l}_1, \bar{l}_2, \bar{l}_3, \bar{l}_4) , (l_1, l_2, l_3, l_4) , \\ (l_1, \bar{l}_2, \bar{l}_3, \bar{l}_4) , (\bar{l}_1, l_2, l_3, l_4) , \\ (\bar{l}_1, l_2, \bar{l}_3, \bar{l}_4) , (l_1, \bar{l}_2, l_3, l_4) , \\ (\bar{l}_1, \bar{l}_2, l_3, \bar{l}_4) , (l_1, l_2, \bar{l}_3, l_4) , \\ (\bar{l}_1, \bar{l}_2, \bar{l}_3, l_4) , (l_1, l_2, l_3, \bar{l}_4) , \\ \\ (\bar{l}_1, \bar{l}_2, l_3, l_4) , (l_1, l_2, \bar{l}_3, \bar{l}_4) , \\ (\bar{l}_1, l_2, \bar{l}_3, l_4) , (l_1, \bar{l}_2, l_3, \bar{l}_4) , \\ (\bar{l}_1, l_2, l_3, \bar{l}_4) , (l_1, \bar{l}_2, \bar{l}_3, l_4) \end{array} \right\} \quad (6)$$

- a) Er utfallsrommet Ω diskret eller kontinuertlig?
Begrunn svaret kort.
- b) Fra Ω i lign.(6) ser vi at det er 16 utfall i utfallsrommet.
Dersom det hadde vært n merder istedet for 4, hvor mange utfall hadde det vært i utfallsrommet da?
- c) Anta at eksperimentet gir følgende resultat:

$$l_1 , l_2 , l_3 , \bar{l}_4 \quad (7)$$

Skriv kort med ord hva resultatet betyr.

Anta $p = 0.5$.

Regn ut sannsynligheten for utfallet.

Begivenheten:

$$S = \text{minst halvparten av de trekte fiskene har lakselus} \quad (8)$$

er gitt ved

$$S = \left\{ \begin{array}{l} (l_1, l_2, l_3, l_4) , \\ (\bar{l}_1, l_2, l_3, l_4) , \\ (l_1, \bar{l}_2, l_3, l_4) , \\ (l_1, l_2, \bar{l}_3, l_4) , \\ (l_1, l_2, l_3, \bar{l}_4) , \\ \\ (\bar{l}_1, \bar{l}_2, l_3, l_4) , (l_1, l_2, \bar{l}_3, \bar{l}_4) , \\ (\bar{l}_1, l_2, \bar{l}_3, l_4) , (l_1, \bar{l}_2, l_3, \bar{l}_4) , \\ (\bar{l}_1, l_2, l_3, \bar{l}_4) , (l_1, \bar{l}_2, \bar{l}_3, l_4) \end{array} \right\} \quad (9)$$

- d) Finn et uttrykk for sannsynligheten for begivenheten S beskrevet i lign.(9), dvs. finn $P(S)$ uttrykt i ” p ”.
- e) Anta at vi vet at minst halvparten av de trekte fiskene har lakselus. Regn ut sannsynligheten for at minst to av de trekte fiskene *ikke* har lakselus. Bruk at $p = 0.5$.²



²Med begivenheten:

$$A = \text{minst to av de trekte fiskene } \textit{ikke} \text{ har lakselus} \quad (10)$$

kan man vise at

$$P(A \cap S) = 6 \cdot 0.5^4 \quad (11)$$

Bruk dette for å løse oppgave **1e**.

Oppgave 2: (stokastiske variabler og forventningverdi - telling av lakselus)

Avsnitt 4.1 i rapporten til Veterinærinstituttet:

”For å beskrive fordelingen av lus mellom fiskene brukte vi et datasett fra 20 oppdrettsanlegg i Rogaland, med lusetall på enkeltfisknivå fra lusetellinger utført som ledd i daglig drift. Tellingene var utført fra og med uke 1 i 2012 til og med uke 37 i 2017. Til sammen hadde vi lusetellinger fra 352 228 enkeltfisker.”

Fra dette datasettet ble det trekt ut et tilfeldig utvalg av fisker som kom fra merder med gjennomsnittlig antall hunnlus per fisk mellom 0.4 og 0.6. Det tilfeldige utvalget utgjorde 7971 fisker. Gjennomsnittlig antall hunnluser per fisk \bar{x} for dette utvalget samt tilhørende empirisk varians s_x^2 er vist i tabell 1. ³

Antall fisker N	\bar{x}	s_x^2
7971	0.499	0.894

Tabell 1: Nøkkeltall for utvalgte merder.



Figur 3: Rapport.

³Disse nøkkeltallene kommer fra rapporten og derfor faktiske nøkkeltall.

Populasjonsvariabel:

$$X = \text{antall hunnlus på en tilfeldig valgt fisk fra et vilkårlig oppdrettsanlegg} \quad (12)$$

med lusetall mellom 0.4 og 0.6

Forsøksvariabler:

$$X_i = \text{antall hunnlus på fisk nr. } i \text{ fra utvalget} \quad (13)$$

hvor

$$i = 1, 2, \dots, N \quad (14)$$

med $N = 7971$ antall fisker.

I denne oppgaven antar vi en statistisk modell for forsøksvariablene basert på den *empiriske* fordelingen, dvs. vi antar:

$$X_1, X_2, \dots, X_N \stackrel{\text{i.i.d.}}{\sim} X \sim P(X = j) \quad (15)$$

hvor

$$j = 0, 1, 2, 3 \quad (\text{antall lakseluser}) \quad (16)$$

med fordelingen: ⁴

	$j = 0$	$j = 1$	$j = 2$	$j = 3$
$P(X = j)$	0.72	0.13	0.08	0.07

Tabell 2: Sannsynlighetsfordeling $P(X = j)$.

⁴Vi antar her for enkelthetsskyld at det ikke ble registrert fisker med flere enn 3 lakseluser. Tallene i tabell 2 er fiktive da selve rådataene ikke er tilgjengelige.

- a) Forklar kort med ord hva sannsynligheten $P(X = j)$ betyr. ⁵
- b) Vis at $P(X = j)$ er en gyldig sannsynlighetsfordeling.
- c) Regn ut forventningsverdien $E[X]$.
- d) Regn ut variansen $Var[X]$ og standardavviket $\sigma[X]$.

⁵En setning er nok.

Hver uke må hvert oppdrettsanlegg telle antall voksne hunnlus på 50 tilfeldig valgte fisker og rapportere lusetallet til Mattilsynet. Basert på gjennomsnittet av antall hunnlus per fisk for disse 50 fiskene må Mattilsynet vurdere om hvorvidt lusegrensa er overskredet eller ikke.

Anta vi ser på ett av disse oppdrettsanleggene for en vilkårlig uke t .

Forsøksvariablene er

$$X_{kt} = \text{antall hunnlus på valgt fisk nr. } k \text{ i uke } t \quad (17)$$

hvor

$$k = 1, 2, \dots, n \quad (\text{antall fisk}) \quad (18)$$

med $n = 50$ antall fisk.

Anta videre at:

$$X_{1t}, X_{2t}, \dots, X_{50t} \stackrel{i.i.d.}{\sim} X \sim P(X = j) \quad (19)$$

hvor *i.i.d.* betyr uavhengige identiske fordelte variabler og hvor X er definert ved lign.(12) med fordeling som vist i tabell 2.

- e) Bruk sentralgrensesetningen til å vise at sannsynligheten for at det målte lusetallet \bar{X}_t til oppdrettsanlegget i uke t ligger over lusegrensa a , er 50 %, dvs. vis at:

$$P(\bar{X}_t > a) = 0.5 \quad (20)$$

hvor $a = 0.5$.

- f) Bør Mattilsynet konkludere at lusegrensa er overskredet dersom $\bar{X}_t > 0.5$?
Begrunn svaret kort.



Oppgave 3: (statistisk modell og estimatorer)

I rapporten bestemte forfatterne seg for å formulere en statistisk modell for forsøksvariablene X_i fra lign.(13) basert på Poisson-familien av fordelinger, dvs. de antok at:

$$X_1, X_2, \dots, X_N \stackrel{\text{i.i.d.}}{\sim} X \sim \text{Poi}[\lambda] \quad (21)$$

hvor

$$N = 7971 \text{ antaller fisker} \quad (22)$$

$$\lambda = \text{forventet antall hunnlus per fisk} \quad (23)$$

(ukjent parameter som er positiv, dvs. $\lambda > 0$)

Anta vi definerer følgende estimator $\hat{\lambda}$ for parameteren λ :

$$\hat{\lambda} = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \quad (24)$$

a) Vis at $\hat{\lambda}$ i lign.(24) er forventningsrett.



Figur 4: Laxselus.

b) Vis at ⁶ ⁷

$$[LB_N^\lambda, UB_N^\lambda] = \left[\bar{X} - z_{1-\alpha/2} \sqrt{\frac{\bar{X}}{N}}, \bar{X} - z_{\alpha/2} \sqrt{\frac{\bar{X}}{N}} \right] \quad (27)$$

er et asymptotisk $(1 - \alpha)$ 100% konfidensintervall for λ ,
hvor $z_{\alpha/2}$ og $z_{1-\alpha/2}$ er de tilhørende kvantilene til normalfordelingen.

⁶Bruk at forventningen og standardavviket til $\hat{\lambda} = \bar{X}$ er, jfr. lign.(21), gitt ved:

$$E[\bar{X}] = \lambda \quad (25)$$

$$\sigma[\bar{X}] = \sqrt{\frac{\lambda}{N}} \quad (26)$$

⁷Se definisjon av asymptotiske $(1 - \alpha)$ 100% konfidensintervall i formelsamlingen.

Poissonfordelingen, også kalt ”*loven om sjeldne begivenheter*”, er en statistisk fordeling som kan *utledes* dersom en del *forutsetninger* er oppfylt.

En viktig forutsetning som bør være oppfylt i den sammenheng er det som kalles *homogenitet* eller *ekvispredning*, dvs. at forventningsverdien og variansen er lik. Med $X \sim \text{Poi}[\lambda]$ har vi da:

$$E[X] = \text{Var}[X] = \lambda \quad (28)$$

Dersom:

$$E[X] < \text{Var}[X] : \quad \text{overspredning} \quad (29)$$

$$E[X] > \text{Var}[X] : \quad \text{underspredning} \quad (30)$$

Det kan være flere grunner til *overspredning* i et datasett, men den mest vanlige er at det er *heterogenitet* i dataene. Heterogenitet betyr i prinsippet at dataene ikke er fremskaffet under like betingelser/omgivelser.

- c) Viser dataene fra rapporten en tendens til *overspredning*?
Begrunn svaret kort. ⁸

⁸Tips: Bruk tabell 1.

For å ta høyde for overspredningen i datasettet foreslo forfatterne en ny statistisk modell basert på en annen diskret fordeling kalt *Negativ Binomial*:

$$X_1, X_2, \dots, X_N \stackrel{\text{i.i.d.}}{\sim} X \sim \text{NBin}[\lambda, \theta] \quad (31)$$

hvor

$$\begin{aligned} \lambda &= \text{forventningsverdien} \\ \theta &= \text{spredningsparameteren} \end{aligned} \quad (32)$$

Dersom $X \sim \text{NBin}[\lambda, \theta]$ så gjelder:

$$\begin{aligned} E[X] &= \lambda \\ \text{Var}[X] &= \lambda + \frac{\lambda^2}{\theta} \end{aligned} \quad (33)$$

hvor $\lambda > 0$ og $\theta > 0$.

- d)** Synes du negativ binomial fordelingen er en passende modell dersom en har **overspredning** i dataene?

Begrunn svaret kort.



Oppgave 4: (krav som indikerer overskridelse av lusegrensa)

Vi skal i denne oppgaven se på kravene foreslått i rapporten som indikerer overskridelse av lusegrensa $a = 0.5$.

I oppgave 2 definerte vi forsøksvariablene fra en telling av 50 fisker på et vilkårlig oppdrettsanlegg en vilkårlig uke t :

$$X_{kt} = \text{antall hunnlus på valgt fisk nr. } k \text{ i uke } t \quad (34)$$

hvor

$$k = 1, 2, \dots, n \quad (\text{antall fisk}) \quad (35)$$

med $n = 50$ fisker.

Anta nå at vi bruker Poisson-modellen fra oppgave 3 som statistisk modell for forsøksvariablene i lign.(34), dvs. anta:

$$X_{1t}, X_{2t}, \dots, X_{nt} \stackrel{\text{i.i.d.}}{\sim} X \sim \text{Poi}[\lambda] \quad (36)$$

Oppdrettsanlegget regner ut så gjennomsnittet \bar{X}_t for uke t og sender resultatet til Mattilsynet. Mattilsynet vurderer deretter, på bakgrunn av den målte verdien for \bar{X}_t , om lusegrensa er overskredet eller ikke.

Veterinærinstituttet har foreslått noen krav på \bar{X}_t som Mattilsynet kan bruke når de skal vurdere om lusegrensa er overskredet eller ikke. Hovedmålet med disse kravene er å unngå feile konklusjoner av typen ”*falske positive*”, dvs.

konkludere med at lusetallet er over lusegrensa, gitt at det sanne lusetallet ligger under

siden det å overskride lusegrensa har store økonomiske konsekvenser for oppdretterne.

I rapporten foreslås *to* krav:

1. For å unngå *falske positiviteter* så ønsker man å teste \bar{X}_t mot et kriterium som er slik at det kun er $(1 - 0.96) \cdot 100\% = 4\%$ sannsynlighet for at \bar{X}_t er ligger over lusegrensen $a = 0.5$ gitt at det sanne lusetallet er under grensen, dvs. $E[X] < 0.5$. Man må da finne ei grense, si $b_{0.96}$, som oppfyller kravet:

$$P(\bar{X}_t > b_{0.96}) = 0.04 \quad (37)$$

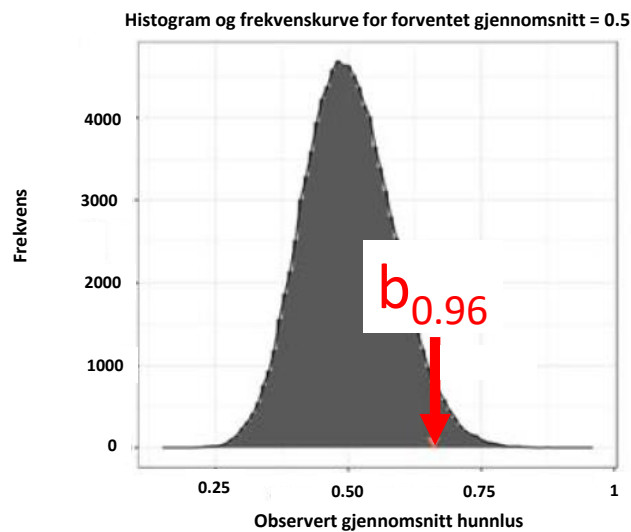
gitt at det sanne lusetallet er $E[X] < 0.5$, hvor

$$b_{0.96} = 96\% \text{-kvantilen til } \bar{X}_t \quad (38)$$

2. Lusegrensa $a = 0.5$ ansees som overskredet dersom lusetellingen ligger over lusegrensa 4 uker på rad, gitt at det sanne lusetallet ligger under, dvs. $E[X] < 0.5$. Begivenheten B som beskriver dette er:

$$B = \left\{ \bar{X} \mid \bar{X}_t > a \text{ og } \bar{X}_{t+1} > a \text{ og } \bar{X}_{t+2} > a \text{ og } \bar{X}_{t+3} > a \right\} \quad (39)$$

hvor $a = 0.5$ er lusegrensa.



Figur 5: Kvantilen $b_{0.96}$ til \bar{X}_t .

- a) Anta at ei lusetelling for et vilkårlig oppdrettsanlegg med $E[X] = \lambda = 0.49$ for ei vilkårlig uke t , gav $\bar{x}_t = 0.68$.

Vil Mattilsynet konkludere med at lusegrensa er overskredet dersom de legger krav 1 til grunn, dvs. lign.(37)?

Begrunn svaret ved regning. ⁹

- b) Anta at $E[X] = 0.49$, dvs. anta det sanne lusetallet ligger rett under lusegrensa. Anta også at lusetellingene er uavhengige av hverandre fra uke til uke.

Regn ut sannsynligheten $P(B)$ for at lusetellingen ligger over lusegrensa 4 uker på rad. ¹⁰

■

⁹Dette er en A-oppgave.

¹⁰Dvs. regn ut:

$$P(B) = P(\bar{X}_t > a \text{ og } \bar{X}_{t+1} > a \text{ og } \bar{X}_{t+2} > a \text{ og } \bar{X}_{t+3} > a) \quad (40)$$

Anta at $P(\bar{X}_t > a) = 0.5$.

Kapittel 17

Kontinuasjonseksamen 2019



Eksamen i

MAT110 Statistikk 1

Eksamensdag	: Mandag 9. sep. 2019
Tid	: 13:00 – 17:00 (4 timer)
Faglærer/telefonnummer	: Molde + Kristiansund: Per Kristian Rekdal / 924 97 051
Hjelpemidler	: KD + formelsamling + generell ordbok (morsmål/norsk/engelsk i papirformat)
Antall sider inkl. forsiden	: 17
Målform	: Norsk (bokmål)

Noen generelle råd:

- **Skriv rett inn. Ikke bruk så mye tid på kladding.**
- **Det er totalt 4 oppgaver.**

Hver uke er alle oppdrettsanlegg i Norge pålagt å rapportere til forvaltningsmyndighetene om hvor mange voksne hunnlus som i gjennomsnitt er på alle fiskene i anlegget. Hensikten med tellingen er tofoldig:

1. Ha en oversikt over spredningen av voksne hunnlus i norske oppdrettsanlegg.
2. Kunne regulere oppdrettsanlegg som har for høye lusetall. ¹

Forvaltningsmyndighetene opererer derfor med ei lusegrense, a , som det sanne lusetallet må ligge under:

a	=	lusegrense	
	=	0.5 voksne hunnlus i gjennomsnitt per fisk per anlegg	(1)

Veterinærinstituttet har skrevet en rapport med tittelen:

” Telling av lakselus - Hvordan forstå og håndtere usikkerheten i telleresultatene”

som handler om hvordan den ukentlige tellingen av lakselus kan brukes til å forvalte lusegrena.



Figur 1: Rapport. Lakselus. Oppdrettsanlegg.

¹Lusetall = forventet antall voksne hunnlus per fisk per oppdrettsanlegg.

Oppgave 1: (sannsynlighetsteori - telling av lakselus)

Anta nå at vi ser på et oppdrettsanlegg med kun **3 merder**. En merd er en innhengning for oppdrettsfisk, se figur 2. Vi gjør følgende *eksperiment*:

1. Nummerer de 3 merdene fra 1 til 3.
2. Én fisk velges tilfeldig fra hver av de 3 merdene.
La i representere den valgte fisken fra merd nr. i , for $i = 1, 2, 3$.
3. For hver valgte fisk registreres det om den har lakselus på seg eller ikke.
Notasjon:

$$l_i = \text{en tilfeldig valgt fisk fra merd nr. } i \text{ har lakselus} \quad (2)$$

$$\bar{l}_i = \text{en tilfeldig valgt fisk fra merd nr. } i \text{ har } \underline{\text{ikke}} \text{ lakselus} \quad (3)$$

I vårt tilfelle er altså:

$$\text{eksperiment} = \text{sekvens med 3 forsøk} \quad (4)$$

Anta at vi kjenner:

$$p = \underline{\text{sannsynligheten}} \text{ for at en vilkårlig valgt fisk har } \underline{\text{lakselus}} \quad (5)$$



Figur 2: 3 merder.

Utfallsrommet Ω for eksperimentet er:

$$\Omega = \left\{ \begin{array}{l} (\bar{l}_1, \bar{l}_2, \bar{l}_3) , (l_1, l_2, l_3) , \\ (l_1, \bar{l}_2, \bar{l}_3) , (\bar{l}_1, l_2, l_3) , \\ (\bar{l}_1, l_2, \bar{l}_3) , (l_1, \bar{l}_2, l_3) , \\ (\bar{l}_1, \bar{l}_2, l_3) , (l_1, l_2, \bar{l}_3) \end{array} \right\} \quad (6)$$

- a) Er utfallsrommet Ω diskret eller kontinuerlig?
Begrunn svaret kort.
- b) Fra Ω i lign.(6) ser vi at det er 8 utfall i utfallsrommet.
Dersom det hadde vært n merder istedet for 3, hvor mange utfall hadde det vært i utfallsrommet da?
- c) Anta at eksperimentet gir følgende resultat:

$$l_1 , \bar{l}_2 , l_3 \quad (7)$$

Skriv kort med ord hva resultatet betyr.

Anta $p = 0.3$.

Anta også at eksperimentene er uavhengige.

Regn ut sannsynligheten for utfallet.

Begivenheten:

$$S = \text{minst to av de trekte fiskene har lakselus} \quad (8)$$

er en delmengde av utfallsrommet Ω gitt ved

$$S = \left\{ \begin{array}{l} , (l_1, l_2, l_3) , \\ , (\bar{l}_1, l_2, l_3) , \\ , (l_1, \bar{l}_2, l_3) , \\ , (l_1, l_2, \bar{l}_3) \end{array} \right\} \quad (9)$$

- d) Finn et uttrykk for sannsynligheten for begivenheten S beskrevet i lign.(9), dvs. finn $P(S)$ uttrykt i "p".

Definer begivenheten:

$$A = \text{én av de trekte fiskene ikke har lakselus} \quad (10)$$

Begivenheten $A \cap S$ er da en delmengde av S i lign.(9).

- e) Definer begivenheten $A \cap S$.
Bruk samme notasjon som i lign.(9).

Man kan vise at: ²

$$P(A \cap S) = 3p^2(1-p)^1 \quad (11)$$

f) Anta at vi vet at minst to av de trekte fiskene har lakselus, dvs. vi vet S .

Regn ut sannsynligheten for at én av de trekte fiskene *ikke har lakselus*, dvs. A , gitt at vi vet S .

Bruk at $p = 0.3$.

Tips:

Bruk betinget sannsynlighet og lign.(11).



²Du skal ikke vise lign.(11). Bare ta den for gitt.

Oppgave 2: (stokastiske variabler og forventningverdi - telling av lakselus)

Avsnitt 4.1 i rapporten til Veterinærinstituttet:

”For å beskrive fordelingen av lus mellom fiskene brukte vi et datasett fra 20 oppdrettsanlegg i Rogaland, med lusetall på enkeltfisknivå fra lusetellinger utført som ledd i daglig drift. Tellingene var utført fra og med uke 1 i 2012 til og med uke 37 i 2017. Til sammen hadde vi lusetellinger fra 352 228 enkeltfisker.”

Fra dette datasettet ble det trekt ut et tilfeldig utvalg av fisker som kom fra merder med gjennomsnittlig antall hunnlus per fisk mellom 0.01 og 0.20. Det tilfeldige utvalget utgjorde 97 773 fisker. Gjennomsnittlig antall hunnluser per fisk \bar{x} for dette utvalget samt tilhørende empirisk varians s_x^2 er vist i tabell 1. ³

Antall fisker N	\bar{x}	s_x^2
97 773	0.097	0.126

Tabell 1: Nøkkeltall for utvalgte merder.



Figur 3: Rapport.

³Disse nøkkeltallene kommer fra rapporten og derfor faktiske nøkkeltall.

Populasjonsvariabel:

$$X = \text{antall hunnlus på en tilfeldig valgt fisk fra et vilkårlig oppdrettsanlegg} \quad (12)$$

med lusetall mellom 0.01 og 0.20

Forsøksvariabler:

$$X_i = \text{antall hunnlus på fisk nr. } i \text{ fra utvalget} \quad (13)$$

hvor

$$i = 1, 2, \dots, N \quad (14)$$

med $N = 97\,773$ antall fisker.

I denne oppgaven antar vi en statistisk modell for forsøksvariablene basert på den *empiriske* fordelingen, dvs. vi antar:

$$X_1, X_2, \dots, X_N \stackrel{\text{i.i.d.}}{\sim} X \sim P(X = j) \quad (15)$$

hvor

$$j = 0, 1, 2, 3 \quad (\text{antall lakseluser}) \quad (16)$$

med fordelingen: ⁴

	$j = 0$	$j = 1$	$j = 2$	$j = 3$
$P(X = j)$	0.88	0.07	0.04	0.01

Tabell 2: Sannsynlighetsfordeling $P(X = j)$.

⁴Vi antar her for enkelthetskyld at det ikke ble registrert fisker med flere enn 3 lakselus. Tallene i tabell 2 er fiktive da selve rådataene ikke er tilgjengelige.

- a) Forklar kort med ord hva sannsynligheten $P(X \geq j)$ betyr. ⁵
- b) Vis at $P(X = j)$ er en gyldig sannsynlighetsfordeling.
- c) Regn ut forventningsverdien $E[X]$.
- d) Regn ut variansen $Var[X]$ og standardavviket $\sigma[X]$.

⁵En setning er nok.

Hver uke må hvert oppdrettsanlegg telle antall voksne hunnlus på 50 tilfeldig valgte fisker og rapportere lusetallet til Mattilsynet. Basert på gjennomsnittet av antall hunnlus per fisk for disse 50 fiskene må Mattilsynet vurdere om hvorvidt lusegrensa er overskredet eller ikke.

Anta vi ser på ett av disse oppdrettsanleggene for en vilkårlig uke t .

Forsøksvariablene er

$$X_{kt} = \text{antall hunnlus på valgt fisk nr. } k \text{ i uke } t \quad (17)$$

hvor

$$k = 1, 2, \dots, n \quad (\text{antall fisk}) \quad (18)$$

med $n = 50$ antall fisk.

Anta videre at:

$$X_{1t}, X_{2t}, \dots, X_{50t} \stackrel{\text{i.i.d.}}{\sim} X \sim P(X = j) \quad (19)$$

hvor *i.i.d.* betyr uavhengige identiske fordelte variabler og hvor X er definert ved lign.(12) med fordeling som vist i tabell 2.

- e) Bruk sentralgrensesetningen til å vise at sannsynligheten for at det målte lusetallet \bar{X}_t til oppdrettsanlegget i uke t ligger under lusegrensa a , er 60.26 %, dvs. vis at:

$$P(\bar{X}_t \leq a) = 0.6026 \quad (20)$$

hvor $a = 0.2$.

- f) Bør Mattilsynet konkludere at lusegrensa er overholdt dersom $\bar{X}_t \leq 0.2$?
Begrunn svaret kort.



Oppgave 3: (statistisk modell og estimatorer)

I rapporten bestemte forfatterne seg for å formulere en statistisk modell for forsøksvariablene X_i fra lign.(13) basert på Poisson-familien av fordelinger, dvs. de antok at:

$$X_1, X_2, \dots, X_N \stackrel{\text{i.i.d.}}{\sim} X \sim \text{Poi}[\lambda] \quad (21)$$

hvor

$$N = 97773 \text{ antaller fisker} \quad (22)$$

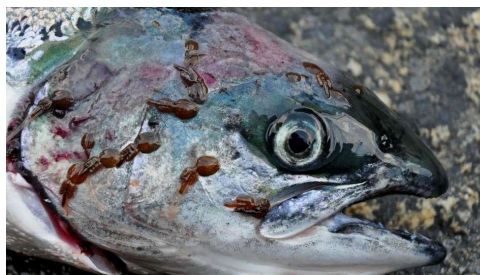
$$\lambda = \text{forventet antall hunnlus per fisk} \quad (23)$$

(ukjent parameter som er positiv, dvs. $\lambda > 0$)

Anta vi definerer følgende estimator $\hat{\lambda}$ for parameteren λ :

$$\hat{\lambda} = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \quad (24)$$

a) Vis at $\hat{\lambda}$ i lign.(24) er forventningsrett.



Figur 4: Lakselus.

b) Vis at ⁶ ⁷

$$[LB_N^\lambda, UB_N^\lambda] = \left[\bar{X} - z_{1-\alpha/2} \sqrt{\frac{\bar{X}}{N}}, \bar{X} - z_{\alpha/2} \sqrt{\frac{\bar{X}}{N}} \right] \quad (27)$$

er et asymptotisk $(1 - \alpha) 100\%$ konfidensintervall for λ ,
hvor $z_{\alpha/2}$ og $z_{1-\alpha/2}$ er de tilhørende kvantilene til normalfordelingen.

⁶Bruk at forventningen og standardavviket til $\hat{\lambda} = \bar{X}$ er, jfr. lign.(21), gitt ved:

$$E[\bar{X}] = \lambda \quad (25)$$

$$\sigma[\bar{X}] = \sqrt{\frac{\lambda}{N}} \quad (26)$$

⁷Se definisjon av asymptotiske $(1 - \alpha) 100\%$ konfidensintervall i formelsamlingen.

Poissonfordelingen, også kalt "loven om sjeldne begivenheter", er en statistisk fordeling som kan *utledes* dersom en del *forutsetninger* er oppfylt.

En viktig forutsetning som bør være oppfylt i den sammenheng er det som kalles *homogenitet* eller *ekvispredning*, dvs. at forventningsverdien og variansen er lik. Med $X \sim \text{Poi}[\lambda]$ har vi da:

$$E[X] = \text{Var}[X] = \lambda \quad (28)$$

Dersom:

$$E[X] < \text{Var}[X] : \quad \text{overspredning} \quad (29)$$

$$E[X] > \text{Var}[X] : \quad \text{underspredning} \quad (30)$$

Det kan være flere grunner til *overspredning* i et datasett, men den mest vanlige er at det er *heterogenitet* i dataene. Heterogenitet betyr i prinsippet at dataene ikke er fremskaffet under like betingelser/omgivelser.

- c) Viser dataene fra rapporten en tendens til *overspredning*?
Begrunn svaret kort. ⁸

⁸Tips: Bruk tabell 1.

For å ta høyde for overspredningen i datasettet foreslo forfatterne en ny statistisk modell basert på en annen diskret fordeling kalt *Negativ Binomial*:

$$X_1, X_2, \dots, X_N \stackrel{\text{i.i.d.}}{\sim} X \sim \text{NBin}[\lambda, \theta] \quad (31)$$

hvor

$$\begin{aligned} \lambda &= \text{forventningsverdien} \\ \theta &= \text{spredningsparameteren} \end{aligned} \quad (32)$$

Dersom $X \sim \text{NBin}[\lambda, \theta]$ så gjelder:

$$\begin{aligned} E[X] &= \lambda \\ \text{Var}[X] &= \lambda + \frac{\lambda^2}{\theta} \end{aligned} \quad (33)$$

hvor $\lambda > 0$ og $\theta > 0$.

- d) Synes du negativ binomial fordelingen er en passende modell dersom en har **overspredning** i dataene?

Begrunn svaret kort.



Oppgave 4: (krav som indikerer overskridelse av lusegrensa)

Vi skal i denne oppgaven se på kravene foreslått i rapporten som indikerer overskridelse av lusegrensa $a = 0.2$.

I oppgave 2 definerte vi forsøksvariablene fra en telling av 50 fisker på et vilkårlig oppdrettsanlegg en vilkårlig uke t :

$$X_{kt} = \text{antall hunnlus på valgt fisk nr. } k \text{ i uke } t \quad (34)$$

hvor

$$k = 1, 2, \dots, n \quad (\text{antall fisk}) \quad (35)$$

med $n = 50$ fisker.

Anta nå at vi bruker Poisson-modellen fra oppgave 3 som statistisk modell for forsøksvariablene i lign.(34), dvs. anta:

$$X_{1t}, X_{2t}, \dots, X_{nt} \stackrel{\text{i.i.d.}}{\sim} X \sim \text{Poi}[\lambda] \quad (36)$$

Oppdrettsanlegget regner ut så gjennomsnittet \bar{X}_t for uke t og sender resultatet til Mattilsynet. Mattilsynet vurderer deretter, på bakgrunn av den målte verdien for \bar{X}_t , om lusegrensa er overskredet eller ikke.

Veterinærinstituttet har foreslått noen krav på \bar{X}_t som Mattilsynet kan bruke når de skal vurdere om lusegrensa er overskredet eller ikke. Hovedmålet med disse kravene er å unngå feile konklusjoner av typen ”*falske positive*”, dvs.

konkludere med at lusetallet er over lusegrensa, gitt at det sanne lusetallet ligger under

siden det å overskride lusegrensa har store økonomiske konsekvenser for oppdretterne.

I rapporten foreslås *to* krav:

1. For å unngå *falske positiviteter* så ønsker man å teste \bar{X}_t mot et kriterium som er slik at det kun er $(1 - 0.95) \cdot 100\% = 5\%$ sannsynlighet for at \bar{X}_t er ligger over lusegrensen $a = 0.2$ gitt at det sanne lusetallet er under grensen, dvs. $E[X] < a = 0.2$. Man må da finne ei grense, si $b_{0.95}$, som oppfyller kravet:

$$P(\bar{X}_t > b_{0.95}) = 0.05 \quad (37)$$

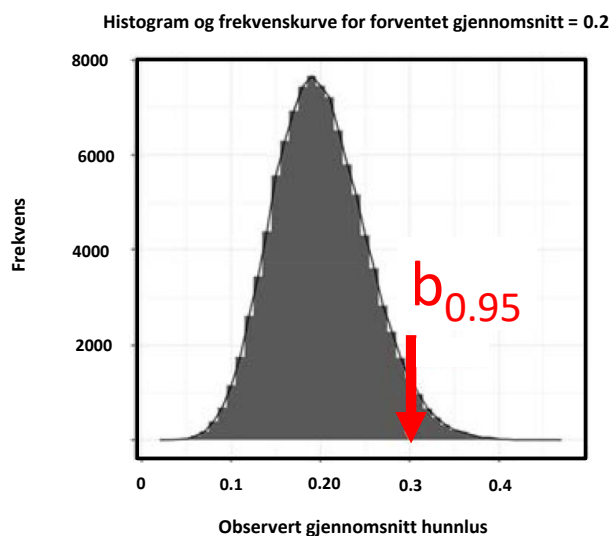
gitt at det sanne lusetallet er $E[X] < 0.5$, hvor

$$b_{0.95} = 95\% \text{-kvantilen til } \bar{X}_t \quad (38)$$

2. Lusegrensa $a = 0.2$ anses som overskredet dersom lusetellingen ligger over lusegrensa 4 uker på rad, gitt at det sanne lusetallet ligger under, dvs. $E[X] < a = 0.2$. Begivenheten B som beskriver dette er:

$$B = \left\{ \bar{X} \mid \bar{X}_t > a \text{ og } \bar{X}_{t+1} > a \text{ og } \bar{X}_{t+2} > a \text{ og } \bar{X}_{t+3} > a \right\} \quad (39)$$

hvor $a = 0.2$ er lusegrensa.



Figur 5: Kvantilen $b_{0.95}$ til \bar{X}_t .

Siden målet er å unngå falske positiviteter, dvs. at et oppdrettsanlegg konkluderer åbrutt llusegrensa når de egentlig er under, så antar vi at vi ser på et oppdrettsanlegg som faktisk har forventningsverdi under lusegrensa. I vår tilfelle:

$$E[X] = \lambda = 0.17 < \underbrace{a = 0.2}_{\text{lusegrensa}} \quad (40)$$

- a) Anta at lusetellingen for ett av disse oppdrettsanleggene ga $\bar{x}_t = 0.25$. for en vilkårlig uke t .

Vil Mattilsynet konkludere med at lusegrensa er overskredet dersom de legger krav 1 til grunn, dvs. lign.(37)?

Begrunn svaret ved regning. ⁹

- b) Anta, for det samme oppdrettsanlegget, at sannsynligheten for å overskride lusegrensa er

$$P(\bar{X}_t > a) = 0.5 \quad (41)$$

for en vilkårlig uke t .

Anta at lusetellingene er uavhengige fra uke til uke.

Regn ut sannsynligheten for at Mattilsynet vil konkludere med at lusegrensa er overskredet dersom de legger krav 2 til grunn, dvs. regn ut:

$$P(B) = P(\bar{X}_t > a \text{ og } \bar{X}_{t+1} > a \text{ og } \bar{X}_{t+2} > a \text{ og } \bar{X}_{t+3} > a) \quad (42)$$

■

⁹Dette er en A-oppgave.