

LØSNING: **Test**eksamen mai 2020

“MAT110 Statistikk 1”, vår 2020

Versjon 01

Oppgave 1: (sannsynlighetsteori - koronatest)

a) 2.5 % Utfallsrommet Ω består av fire 2-tuppler:

$$\Omega = \left\{ (t, s) , (\bar{t}, s) , (t, \bar{s}) , (\bar{t}, \bar{s}) \right\} \quad (1)$$

b) 5 % Multiplikasjonssetningen og komplementsetningen:

$$P(t, \bar{s}) = P(T | \bar{S})P(\bar{S}) \quad (2)$$

$$= \underbrace{P(T | \bar{S})}_{= 0.01} \left(1 - \underbrace{P(S)}_{= 0.01}\right) \quad (3)$$

$$= 0.01 \cdot (1 - 0.01) = 0.099 \quad (4)$$

c) 5 % Tolkning:

$P(t, \bar{s})$ = sannsynligheten for at test sier **positiv** og personen er **ikke smittet**
(falsk positiv test)

d) 5 % Betinget sannsynlighet:

$$P(\textcolor{red}{S} | \overline{\textcolor{green}{T}}) = \frac{\overbrace{P(\textcolor{red}{S} \cap \overline{\textcolor{green}{T}})}^{=0}}{P(\overline{\textcolor{green}{T}})} \quad (5)$$

$$= \frac{0}{P(\overline{\textcolor{green}{T}})} = 0 \quad (6)$$

PS: Legg merke til at vi ikke behøver verdien på $P(\overline{\textcolor{green}{T}}) > 0$ for å regne ut den betingende sannsynligheten.

e) 2.5 % Komplementsetningen:

$$P(\overline{\textcolor{green}{T}} | \textcolor{red}{S}) = 1 - \overbrace{P(\textcolor{red}{T} | \textcolor{red}{S})}^{=1} \quad (7)$$

$$= 1 - 1 = 0 \quad (8)$$

hvor $P(\textcolor{red}{T} | \textcolor{red}{S}) = 1$ er oppgitt i oppgaven.

f) 5 % Siden $P(\overline{\textcolor{green}{T}} | \textcolor{red}{S}) = 0$ så gir testen aldri falskt svar dersom man tester på smittede personer.

■

Oppgave 2: (stokastiske variabler og forventningverdi - aldersfordeling)

a) 5% Forventet levealder på en person som tester positivt på korona:

$$E[X] = \sum_{i=1}^{10} x_i P(X = x_i) \quad (9)$$

$$= \sum_{i=1}^{10} x_i p_i \quad (10)$$

$$\begin{aligned} &= 5 \cdot \frac{77}{6791} + 15 \cdot \frac{295}{6791} + 25 \cdot \frac{1002}{6791} + 35 \cdot \frac{1100}{6791} + 45 \cdot \frac{1244}{6791} + \\ &+ 55 \cdot \frac{1331}{6791} + 65 \cdot \frac{789}{6791} + 75 \cdot \frac{530}{6791} + 85 \cdot \frac{323}{6791} + 95 \cdot \frac{102}{6791} \end{aligned} \quad (11)$$

$$= 47.95 \text{ år} \quad , \quad \text{q.e.d.} \quad (12)$$

Varianssentingen: (oppgitt i år²)

$$Var[X] = E[X^2] - E[X]^2 \quad (13)$$

$$= 2672.36 - 47.95^2 = 373.16 \quad , \quad \text{q.e.d.} \quad (14)$$

med tilhørende standardavvik (oppgitt i år)

$$\sigma[X] = \sqrt{Var[X]} = \sqrt{373.16} = 19.32 \quad (15)$$

b) 5% Siden

$$Y \sim N \left[E[Y], \sigma[Y] \right] \quad (16)$$

hvor $E[Y] = E[X] = 47.95$ og $\sigma[Y] = \sigma[X] = 19.32$, så er:

$$\underline{P(Y < 20)} \stackrel{\text{standardiser}}{=} P \left(\frac{Y - E[X]}{\sigma[X]} < \frac{20 - E[X]}{\sigma[X]} \right) \quad (17)$$

$$= P \left(\underbrace{\frac{Y - E[X]}{\sigma[X]}}_{\equiv Z} < \frac{20 - E[X]}{\sigma[X]} \right) \quad (18)$$

$$= P \left(Z < \frac{20 - 47.95}{19.32} \right) \quad (19)$$

$$= P \left(Z < -1.45 \right) \quad \text{minustegn} \quad (20)$$

$$= 1 - P \left(Z < 1.45 \right) \quad (21)$$

Tabeloppslag:

$$P(Y < 20) = 1 - 0.9265 = 0.0735 \quad (22)$$

Tolkning:

$$P(X < 20) = \text{sannsynligheten for at en tilfeldig valgt person som} \\ \text{har testet positivt på korona i eksperiment } E' \quad (23)$$

$$\text{er under 20 år, er 7.35 \%} \quad (24)$$

- c) 5% La n_i antall smittende i aldergruppe nr. $i = 1, \dots, 10$.
 “Korrigerte” forventningen $E'[X]$:

$$E'[X] = \sum_{i=1}^{10} x_i P(X = x_i) \quad (25)$$

$$= 5 \frac{n_1}{N+m} + 15 \frac{n_2 + m}{N+m} + 25 \frac{n_3}{N+m} + \dots + 95 \frac{n_{10}}{N+m} \quad (26)$$

$$= \left(5n_1 + 15n_2 + 15m + 25n_3 + \dots + 95n_{10} \right) \frac{1}{N+m} \quad (27)$$

$$= \left(5n_1 + 15n_2 + 25n_3 + \dots + 95n_{10} \right) \frac{1}{N+m} + \frac{15m}{N+m} \quad (28)$$

$$= \underbrace{\frac{5n_1 + 15n_2 + 25n_3 + \dots + 95n_{10}}{N}}_{= E[X]} \frac{N}{N+m} + \frac{15m}{N+m} \quad (29)$$

$$= E[X] \frac{N}{N+m} + \frac{15m}{N+m} \quad (30)$$

$$= \frac{E[X] N + 15m}{N+m}, \quad \text{q.e.d.} \quad (31)$$

- d) 5% Bruker lign.(31) og løser med hensyn på m alene:

$$E'[X] = \frac{E[X] N + 15m}{N+m} \quad \Big| \cdot (N+m) \quad (32)$$

$$\begin{aligned} & \Updownarrow \\ (N+m)E'[X] &= E[X] N + 15m \end{aligned} \quad (33)$$

$$\begin{aligned} & \Updownarrow \\ NE'[X] + mE'[X] &= E[X] N + 15m \end{aligned} \quad (34)$$

$$\begin{aligned} & \Updownarrow \\ mE'[X] - 15m &= E[X] N - NE'[X] \end{aligned} \quad (35)$$

$$\begin{aligned} & \Updownarrow \\ m(E'[X] - 15) &= N(E[X] - E'[X]) \quad \Big| \cdot \frac{1}{E'[X] - 15} \end{aligned} \quad (36)$$

$$\begin{aligned} & \Updownarrow \\ m &= \frac{N(E[X] - E'[X])}{E'[X] - 15} \end{aligned} \quad (37)$$

Sett til slutt inn tall: ($N = 6791$, $\overbrace{E[X] = 47.95}^{\text{oppg. 2a}}$ og $E'[X] = 46$)

$$m = \frac{N(E[X] - E'[X])}{E'[X] - 15} \quad (38)$$

$$= \frac{6791(47.95 - 46)}{46 - 15} = 427 \quad (39)$$

Altså, mørketallet er $m = 427$ dersom $E'[X] = 46$ år.

e) 5% Med den nye modellen er

$$E'[Y] = E'[X] = 47.95 \quad (40)$$

$$\sigma'[Y] = \sigma'[X] = 20.30 \quad (41)$$

og Y er fortsatt normalfordelt, så standardiserer vi:

$$P(Y < a) = 0.0735 \quad (42)$$

$$\Updownarrow \quad (43)$$

$$P\left(\underbrace{\frac{Y - E'[X]}{\sigma'[X]}}_{= Z} < \underbrace{\frac{a - E'[X]}{\sigma'[X]}}_{= z_0}\right) = 0.0735 \quad (44)$$

$$\Updownarrow \quad (45)$$

$$P(Z < z_0) = 0.0735 \quad (46)$$

z_0 finnes ved “omvendt tabeloppslag”.

Men sannsynligheten $P(Z < z_0) = 0.0735$ finnes ikke i tabellen i kompendiet.

Sannsynlighetene i kompendiet “stopper” ved 0.5000.

Men av symmetrigrunner så må z_0 være negativ og absoluttverdi tilsvarende sannsynligheten $1 - 0.0735 = 0.9265$. Dermed:

$$z_0 = -1.45 \quad (47)$$

Løser

$$z_0 = \frac{a - E'[X]}{\sigma'[X]} \quad (48)$$

med hensyn på a og setter inn tall

$$a = E'[X] + z_0 \sigma'[X] \quad (49)$$

$$= 46 + (-1.45)20.30 = 16.57 \quad (50)$$

Siden det nå er flere personer i aldersgruppen 10-19 år sammenlignet med oppgave **2b**, og siden $E'[X] = 46 < E[X] = 47.95$, så er det rimelig at alderen a er mindre enn 20 år.

■

Oppgave 3: (25 % - statistisk modellering - bremse koronaspredning)

a) 2.5 % Fra oppgaven vet vi at

$$I_1 \mid B_0(N-1, 1, 0) \sim \text{Bin}[\overbrace{N-1}^{NB!}, p^1 = p] \quad (51)$$

som har forventningsverdi $(N-1)p$, dvs.

$$E[I_1|B_0] = (N-1)p \quad (52)$$

som skulle forklares.

Tolkning:

Forventningsverdien $E[I_1|B_0]$ er forventet antall smittede i slutten av uke 1 når startverdiene er $B_0 = B_0(N-1, 1, 0)$.¹

b) 2.5 % Matematisk modell: (oppgitt i oppgaven)

$$R_0 = \beta c D \quad (53)$$

hvor R_0 er forventningsverdien fra oppgave 3a, dvs. $R_0 = E[I_1|B_0] = (N-1)p$.
Dermed:

$$E[I_1|B_0] = R_0 \quad (54)$$

$$\begin{aligned} &\Updownarrow \\ (N-1)p &= \beta c D \end{aligned} \quad (55)$$

Løser med hensyn på p alene og setter inn tall til slutt:

$$p = \frac{\beta c D}{N-1} \quad (56)$$

$$= \frac{0.25 \cdot 6 \cdot 1}{3-1} = 0.75 \quad , \quad \text{q.e.d.} \quad (57)$$

¹Man kan også si at $E[I_1|B_0]$ forteller oss hvor mange personer pasient zero forventes å smitte.

c) 2.5 % La $B_0 = B_0(2, 1, 0)$. Dermed:

$$I_1 \mid B_0(\overbrace{N-1}^{=2}, 1, 0) \sim \text{Bin}[\overbrace{2, p^1=p}^{NB!}] \quad (58)$$

som gir at

$$P(I_1 = 1 \mid B_0) = \binom{2}{1} p^1 (1-p)^{2-1} = 2p(1-p) = 2p - 2p^2, \quad \text{q.e.d.} \quad (59)$$

d) 2.5 % Med $B_0 = B_0(2, 1, 0)$ og $I_1 = 1$:

$$\left. \begin{array}{l} S_1 = S_0 - I_1 = 2 - 1 = 1 \\ I_1 = 1 \\ R_1 = R_0 + I_0 = 0 + 1 = 1 \end{array} \right\} B_1(1, 1, 1) \quad (60)$$

hvor vi har brukt de oppgitte ligningene i oppgaven. ²

Fra definisjonen av betinget sannsynlighet:

$$P(I_1 = 1, I_2 = 1 \mid B_0) = P(I_2 = 1 \mid I_1 = 1, B_0) \cdot \overbrace{P(I_1 = 1 \mid B_0)}^{= 2p-2p^2 \text{ lign. (59)}} \quad (61)$$

$$= P(I_2 = 1 \mid B_1, \cancel{B_0}) (2p - 2p^2) \quad (62)$$

$$= P(\underbrace{I_2 = 1}_{\sim \text{Bin}[s_1=1, p^1=p]} \mid B_1) (2p - 2p^2) \quad (63)$$

$$= \binom{1}{1} p^1 (1-p)^{1-1} (2p - 2p^2) \quad (64)$$

$$= 1 \cdot p \cdot 1 \cdot (2p - 2p^2) \quad (65)$$

$$= 2p^2 - 2p^3, \quad \text{q.e.d.} \quad (66)$$

hvor vi i lign.(62) kan fjerne $\cancel{B_0}$ i betingelsen pga. Markov tilnærmelsen, dvs. prosessen er kun avhengig av hva som skjedde bare én uke før, altså “korttidshukommelse”.

I lign.(63) har vi brukt binomialfordelingen med $B_1 = B_1(1, 1, 1)$.

²Man kan også innse lign.(60) ved hjelp av figuren tilhørende oppgave 3 og antagelsene i oppgaven.

e) 5% La $B_0 = B_0(2, 1, 0)$ som før.

Addisjonssetningen: ³ (den spesielle addisjonssetningen)

$$P\left(\overbrace{I_1 > 0, I_2 = 0}^{\text{dør ut i uke 2}} \mid B_0 \right) = P\left(I_1 = 1, I_2 = 0 \vee I_1 = 2, I_2 = 0 \mid B_0 \right) \quad (67)$$

$$= \underbrace{P\left(I_1 = 1, I_2 = 0 \mid B_0 \right)}_{\text{første}} + \underbrace{P\left(I_1 = 2, I_2 = 0 \mid B_0 \right)}_{\text{andre}} \quad (68)$$

Med $B_0 = B_0(2, 1, 0)$ og $I_1 = 1$: (slik som i oppgave 3d)

$$\left. \begin{array}{l} S_1 = S_0 - I_1 = 2 - 1 = 1 \\ I_1 = 1 \\ R_1 = R_0 + I_0 = 0 + 1 = 1 \end{array} \right\} B_1(1, 1, 1) \quad (\text{første}) \quad (69)$$

Med $B_0 = B_0(2, 1, 0)$ og $I_1 = 2$:

$$\left. \begin{array}{l} S_1 = S_0 - I_1 = 2 - 2 = 0 \\ I_1 = 2 \\ R_1 = R_0 + I_0 = 0 + 1 = 1 \end{array} \right\} B_1(0, 2, 1) \quad (\text{andre}) \quad (70)$$

hvor vi har brukt de oppgitte ligningene i oppgaven.

Vi tar for oss de to sannsynlighetene i lign.(68) hver for seg:

³Med $I_1 > 0$ og $I_2 = 0$ og vår B_0 så er det to muligheter: $I_1 = 1$ og $I_1 = 2$.

i) Definisjon betinget sannsynlighet: (den første sannsynligheten i lign.(68))

$$P\left(I_1 = 1 , I_2 = 0 \mid B_0 \right) = P\left(I_2 = 0 \mid I_1 = 1 , B_0 \right) \cdot \overbrace{P\left(I_1 = 1 \mid B_0 \right)}^{= 2p-2p^2 \text{ lign.(59)}} \quad (71)$$

$$= P\left(I_2 = 0 \mid B_1, \cancel{B_0} \right) (2p - 2p^2) \quad (72)$$

$$= \underbrace{P\left(I_2 = \textcolor{red}{0} \mid B_1(\textcolor{brown}{1}, \textcolor{red}{1}, 1) \right)}_{= \text{Bin}[s_1=\textcolor{brown}{1}, p^{\textcolor{red}{1}}=p]} (2p - 2p^2) \quad (73)$$

$$= \binom{\textcolor{brown}{1}}{\textcolor{red}{0}} p^{\textcolor{red}{0}} (1-p)^{\textcolor{brown}{1}-\textcolor{red}{0}} (2p - 2p^2) \quad (74)$$

$$= 1 \cdot 1 \cdot (1-p) (2p - 2p^2) \quad (75)$$

$$= 2p - 2p^2 \quad - \quad 2p^2 + 2p^3 \quad (76)$$

$$= 2p^3 - 4p^2 + 2p \quad (77)$$

ii) Definisjon betinget sannsynlighet: (den andre sannsynligheten i lign.(68))

$$P\left(I_1 = 2, I_2 = 0 \mid B_0 \right) = P\left(I_2 = 0 \mid I_1 = \textcolor{red}{2}, B_0 \right) \cdot P\left(\overbrace{I_1 = 2 \mid B_0}^{\sim \text{Bin}[s_0=\textcolor{red}{2}, p^1=p]} \right) \quad (78)$$

$$= P\left(I_2 = 0 \mid B_1, \textcolor{blue}{B}_0 \right) \binom{\textcolor{red}{2}}{\textcolor{red}{2}} p^{\textcolor{red}{2}} (1-p)^{\textcolor{red}{2}-\textcolor{red}{2}} \quad (79)$$

$$= P\left(\underbrace{I_2 = \textcolor{red}{0} \mid B_1(\textcolor{red}{0}, \textcolor{red}{2}, 1)}_{\sim \text{Bin}[\textcolor{red}{0}, p^{\textcolor{red}{2}}]} \right) \cdot 1 \cdot p^2 \cdot 1 \quad (80)$$

$$= \binom{\textcolor{red}{0}}{\textcolor{red}{0}} (p^2)^0 (1-p^2)^{\textcolor{red}{0}-\textcolor{red}{0}} p^2 \quad (81)$$

$$= 1 \cdot 1 \cdot 1 \cdot p^2 \quad (82)$$

$$= p^2 \quad (83)$$

Vi setter inn lign.(83) og lign.(77) inn i lign.(68):

$$P\left(I_1 > 0, I_2 = 0 \mid B_0 \right) = \underbrace{P\left(I_1 = 1, I_2 = 0 \mid B_0 \right)}_{\text{første}} + \underbrace{P\left(I_1 = 2, I_2 = 0 \mid B_0 \right)}_{\text{andre}} \quad (84)$$

$$= 2p^3 - \textcolor{red}{4}p^{\textcolor{red}{2}} + 2p \quad + \quad \textcolor{red}{p}^{\textcolor{red}{2}} \quad (85)$$

$$= 2p^3 - 3p^2 - 2p \quad , \quad \text{q.e.d.} \quad (86)$$

f) 10 % Vi skal finne den nye c' bestemt ved:

$$P(M = 1 \mid B_0) = 0.01 \quad (87)$$

Ut fra lign.(87) finner vi først smitte-sannsynligheten p .
Deretter finner vi c' via lign.(55), dvs.: ⁴

$$(N - 1)p = R_0 \quad (88)$$

$$\Updownarrow$$

$$(N - 1)p = \beta c' D \quad (89)$$

$$\Updownarrow$$

$$c' = \frac{(N - 1)p}{\beta D} \quad (90)$$

Oppgitt:

$$M = \max_{n \geq 0} I_n = \max [I_0, I_1, I_2, I_3, I_4, \dots] \quad (91)$$

Fra figuren i oppgave **3**, som har samme $B_0 = B_0(2, 1, 0)$ som i denne oppgaven, så ser vi at $I_3 = 0$.

Med disse startverdiene, $N = 3$ samt antagelsene i oppgaven så kan vi se på de forskjellige mulige forløpene. Da finner vi at det forløpet som vist i figuren i oppgaveteksten er det som gir max n , dvs. max levelengde på epedemien. Epedemien ut da ut i uke 3, dvs. $I_3 = 0$, og dermed $I_3 = I_4 = I_5 = \dots = 0$. ⁵

Dermed:

$$M = \max_{n \geq 0} [I_0, I_1, I_2, \overbrace{I_3, I_4, \dots}^{=0}] = \max [I_0, I_1, I_2] \quad (92)$$

⁴Husk at vi kjenner $\beta = 0.25$, $D = 6$ og $N = 3$. Så den eneste størrelsen vi mangler er p .

⁵Her er et eksempel å et mulig forløp hvor epedemien før ut tidligere:

Dersom begge $S_0 = 2$ hadde blitt smittet i uke 1, så er det ikke flere som kan smittes. Begge hadde blitt immune påfølgende uke, og epedemien hadde dør ut i uke $n = 2$, og $I_2 = I_3 = I_4 = \dots = 0$.

Innsatt i lign.(87):

$$P\left(\overbrace{\max[I_0, I_1, I_2]}^{=M} = 1 \mid B_0\right) = 0.01 \quad (93)$$

bvor $B_0 = B_0(2, 1, 0)$ som før, dvs. $I_0 = 1$.

Med $I_0 = 1$:

$$P\left(\max[1, I_1, I_2] = 1 \mid B_0\right) = 0.01 \quad (94)$$

$\Updownarrow$

$$P\left(\max[I_1, I_2] = 1 \mid B_0\right) = 0.01 \quad (95)$$

$\Updownarrow$

$$P\left(I_1 \leq 1, I_2 \leq 1 \mid B_0\right) = 0.01 \quad (96)$$

Tre kombinasjoner hvor $I_1 \leq 1$, $I_2 \leq 1$: (jfr. figuren tilhørende oppgave **3**)

$$\text{Hvis } I_1 = 0 \text{ så må } I_2 = 0 \quad (I_2 \text{ kan da ikke være } 1) \quad (97)$$

$$\text{Hvis } I_1 = 1 \text{ så må } I_2 = 0 \text{ eller } I_2 = 1 \quad (98)$$

Addisjonssetningen:

$$P(I_1 \leq 1, I_2 \leq 1 \mid B_0) = \overbrace{P(I_1 = 0, I_2 = 0 \mid B_0)}^{= 1-2p+p^2, \text{ se kommentar under}} \quad (99)$$

$$+ \overbrace{P(I_1 = 1, I_2 = 0 \mid B_0)}^{= 2p^3-4p^2+2p \text{ lign. (77)}} \quad (100)$$

$$+ \overbrace{P(I_1 = 1, I_2 = 1 \mid B_0)}^{= 2p^2-2p^3 \text{ lign. (66)}} \quad (101)$$

$$= (1 - \cancel{2p} + p^2) + (\cancel{2p^3} - 4p^2 + \cancel{2p}) + (2p^2 - \cancel{2p^3}) \quad (102)$$

$$= 1 - p^2 - 4p^2 - 2p^2 \quad (103)$$

$$= 1 - p^2 \quad (104)$$

Kommentar:

Med $B_0 = B_0(\textcolor{brown}{2}, \textcolor{red}{1}, \textcolor{blue}{0})$ og $I_1 = \textcolor{red}{0}$:

$$\left. \begin{array}{l} S_1 = S_0 - I_1 = \textcolor{brown}{2} - \textcolor{red}{0} = 2 \\ I_1 = 0 \\ R_1 = R_0 + I_0 = \textcolor{blue}{0} + \textcolor{red}{1} = 1 \end{array} \right\} B_1(2, \textcolor{red}{0}, 1) \quad (105)$$

og fra definisjonen av betinget sannsynlighet:

$$P(I_1 = 0, I_2 = 0 \mid B_0) = P(I_2 = 0 \mid \overbrace{I_1 = 0, B_0}^{\Rightarrow B_1(\textcolor{brown}{2}, \textcolor{red}{0}, 1)}) P(I_1 = 0 \mid B_0) \quad (106)$$

$$= P(\underbrace{I_2 = 0 \mid B_1(2, \textcolor{red}{0}, 1)}_{\text{ikke} \sim \text{Bin siden } s_1 = \textcolor{red}{0}}) P(I_1 = 0 \mid B_0) \quad (107)$$

altså $P(I_2 = 0 \mid B_1(2, \textcolor{red}{0}, 1))$ er ikke binomisk fordelt siden $i_1 = 0$.

Fra oppgaven ser vi at den betingede stokastiske variabelen er binomisk fordelt kun dersom $i_n > 0$.

Derfor må vi regne ut $P(I_1 = 0, I_2 = 0 \mid B_0)$ på en annen måte, argumentasjon og deretter regning:

Dersom $I_1 = 0$ så må også $I_2 = 0$, siden ingen nye kan bli smittet i uke 2 når det er ingen smittede i uke 1. Vi kan dermed fjerne $I_2 = 0$ i lign(99):

$$P(I_1 = 0, I_2 = 0 \mid B_0) = P(I_1 = 0 \mid B_0(2, 1, 0)) \quad (108)$$

$$= \binom{2}{0} p^0 (1-p)^{2-0} = (1-p)^2 = 1 - 2p + p^2 \quad (109)$$

Dermed:

$$P(I_1 \leq 1, I_2 \leq 1 \mid B_0) = 0.01 \quad (110)$$

$$\begin{array}{c} \Updownarrow \\ 1 - p^2 = 0.01 \end{array} \quad (111)$$

$$\begin{array}{c} \Updownarrow \\ p^2 = 1 - 0.01 = 0.99 \end{array} \quad (112)$$

$$\begin{array}{c} \Updownarrow \\ p = \sqrt{0.99} = 0.995 \end{array} \quad (113)$$

Innsatt i lign.(90): ($p = 0.995$, $\beta = 0.25$, $D = 1$ og $N = 3$)

$$c' = \frac{(N-1)p}{\beta D} \quad (114)$$

$$= \frac{(3-1)0.995}{0.25 \cdot 1} = 7.96 \quad (115)$$

Vi må ha en isolasjon som tilsvarer at man treffer i gjennomsnitt $c' = 7.96$ personer i uken for å sikre med sannsynlighet på 0.01 at ikke flere enn én person er smittet samtidig.

■

Nedenfor skal vi vise en kortere måte å løse oppgave **3f** på. Grunnen til at vi løste oppgave **3f** på den “lange” måten ovenfor er at det gir mye god innsikt i problemet som kan være til nytte på hovedeksamen.

Med startbetingelsen $B_0 = B_0(2, 1, 0)$ og antagelsene som i oppgaven, så er det 4 mulige forløp. Tre av dem gir $M = 1$, mens ett gir $M = 2$. Istedet for å løse lign.(87), så kan vi heller løse

$$P(M = 2 \mid B_0) = 0.99 \quad (116)$$

Men $M = 2$ tilsvarer $I_1 = 2 \mid B_0$. Dermed:

$$P(I_1 = 2 \mid B_0) = 0.99 \quad (117)$$

Med $B_0 = B_0(\textcolor{brown}{2}, \textcolor{red}{1}, \textcolor{blue}{0})$ og $I_1 = \textcolor{red}{2}$:

$$\left. \begin{array}{l} S_1 = S_0 - I_1 = \textcolor{brown}{2} - \textcolor{red}{2} = 0 \\ I_1 = 2 \\ R_1 = R_0 + I_0 = \textcolor{blue}{0} + \textcolor{red}{1} = 1 \end{array} \right\} B_1(0, \textcolor{red}{2}, 1) \quad (118)$$

som gir

$$P(I_1 = 2 \mid B_0) = \underbrace{P(I_1 = \textcolor{blue}{2} \mid B_1(\textcolor{brown}{0}, \textcolor{red}{2}, 1))}_{= \text{Bin}[2, p^2]} \quad (119)$$

$$= \binom{\textcolor{red}{2}}{\textcolor{blue}{2}} p^{\textcolor{red}{2}} (1-p)^{\textcolor{red}{2}-\textcolor{blue}{2}} = p^2 = 0.99 \quad (120)$$

som er det samme som lign.(113).

⁶Takk til Snorre Himberg for gode diskusjoner og god løsning til oppgaven.

Oppgave 4: (25 % - regresjonsanalyse - antall smittede)

a) 2.5 % Siden

$$r_{xy} = \frac{s_{xy}}{s_x s_y} \quad (121)$$

og løser med hensyn på s_{xy} alene

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

$$= 0.9763 \cdot \sqrt{188} \cdot \sqrt{4\,808\,757.10} \quad (123)$$

$$= \underline{\underline{29\,354.77}} \quad (124)$$

b) 2.5 % At $r_{xy} = 0.9763$ betyr at det er en sterk positiv korrelasjon mellom x og y , dvs. store x hører sammen med store y og omvendt.
Det er altså en sterk lineær sammenheng mellom x og y med positivt stigningstall.

c) 7.5 % Vi skal finne regresjonslinjen.
Formelen for minste kvadraters lineære regresjonslinje finner man i kompendiet:

$$\hat{y} = \hat{\alpha} + \hat{\beta}x \quad (125)$$

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

$$\underline{\hat{\beta}} = \frac{s_{xy}}{s_x^2} = \frac{29\,354.77}{188} = \underline{156.14} \quad (126)$$

$$\underline{\hat{\alpha}} = \bar{y} - \hat{\beta}\bar{x} = 4959.34 - 156.14 \cdot 23 = \underline{1368.13} \quad (127)$$

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

$$\underline{\underline{\hat{y} = 1368.13 + 156.14x}} \quad (128)$$

- d) 2.5 % Datoen 10. mai 2020 tilsvarer $x = 61$.
Bruker regresjonslinjen $\hat{y} = \hat{y}(x)$ fra oppgave 4c, dvs. lign.(128):

$$\underline{\hat{y}(61)} = \left(1368.13 + 156.14 \cdot 61 \right) = \underline{10\,892.67} \quad (129)$$

Ifølge regresjonslinjen kommer det til å være 10 893 personer som er smittet per 10. mai 2020.

- e) 5 % Ifølge regresjonslinjen i lign.(128), når $\hat{y}(x) = 10\,000$, er da:

$$10\,000 = \hat{\alpha} + \hat{\beta}x \quad (130)$$

Løser med hensyn på x : ($\alpha = 1368.13$ and $\hat{\beta} = 156.14$)

$$\underline{x} = \frac{10\,000 - \hat{\alpha}}{\hat{\beta}} \quad (131)$$

$$= \frac{10\,000 - 1368.13}{156.14} = \underline{55.28} \quad (132)$$

Ifølge regresjonslinjen er det 10 000 personer i Norge som er smittet av korona dag nr. 55.⁷

⁷Man behøver ikke å skrive at dag nr. 55 tilsvarer 7. mai 2020.

- f) 2.5 % Forklaringskraften r^2 kan leses direkte fra Excel-utskriften:
(Se cellen som heter “*R Square*” i Excel-utskriften): ⁸

$$\underline{\underline{r^2 = 0.9531}} \quad (133)$$

Kommentar:

At $r^2 = 0.9531$ betyr at 95.31 % av variasjonen til y , antall smittede, kan forklares av x , dvs. tiden.

■

⁸Man kan også regne ut forklaringskraften R^2 “for hånd” via definisjonen $r^2 = 1 - sse/sst$. Men det er mye mer arbeidskrevende. Fint at dataprogrammer (som f.eks. Excel) kan hjelpe oss med slikt.