

# LØSNING: Eksamen 22. mai 2018

“MAT110 Statistikk 1”, vår 2018

## Oppgave 1: ( logistikk )

- a) Sannsynlighetene  $p_i$ , med  $i = 1, 2, 3, \dots, 8$  utgjør en gyldig sannsynlighetsfordeling fordi:

$$\sum_{i=1}^8 p_i = \frac{5}{350} + \frac{40}{350} + \frac{25}{350} + \frac{80}{350} \quad (1)$$

$$+ \frac{8}{350} + \frac{68}{350} + \frac{22}{350} + \frac{102}{350} \quad (2)$$

$$= \frac{150 + 200}{350} = \frac{350}{350} = \underline{\underline{1}} \quad (3)$$

- b) Fra figuren i oppgaven ser vi at de to begivenhetene  $S_1$  og  $S_2$  ikke overlapper.<sup>1</sup> Derfor er de disjunkte.

---

<sup>1</sup>I likhet med brikkene i et puslespill så overlapper  $S_1$  og  $S_2$  akkurat ikke.

- c) Siden sannsynlighetene  $p_1 = \frac{5}{350}$ ,  $p_2 = \frac{40}{350}$  osv. er alle ulike,  
altså ikke uniformt utfallsrom, så gjelder ikke urnemodellen for vår sannsynlighetsmodell. <sup>2</sup>

- d) Fra første ligningen i oppgaveteksten finner vi:

$$P(S_2 \cap H \cap A) = p_5 = \frac{8}{350} \quad (4)$$

og

$$P(S_2 \cap H) = p_5 + p_7 = \frac{8}{350} + \frac{22}{350} = \frac{30}{350} \quad (5)$$

Dermed har vi det vi trenger for å regne ut den betingede sannsynligheten:

$$\underline{\underline{P(A|S_2 \cap H)}} = \frac{P(S_2 \cap H \cap A)}{P(S_2 \cap H)} = \frac{\frac{8}{\cancel{350}}}{\frac{30}{\cancel{350}}} = \frac{8}{\underline{\underline{30}}} \quad (= 0.2667) \quad (6)$$

- e) Tolkning: <sup>3</sup>

$P(A|S_2 \cap H)$  = sannsynligheten for at et tilbud blir akseptert dersom det er laget av selger 2  
og verdien av tilbudet er høy

---



---

<sup>2</sup>Fra figuren i oppgaven ser vi også at kulene har forskjellige størrelse. Det visualiserer at sannsynlighetene  $p_i$  er forskjellige.

<sup>3</sup>Det er flere måter å formulere seg på her. Løsningsforslaget er bare én mulig formulering.

Oppgave 2: ( logistikk - prognostisering )

a)  $Y$  beskriver en forsøksserie som har følgende egenskaper:

1. Hvert tilbud har kun 2 mulige utfall, **akseptert** (suksess) eller **ikke akseptert** (fiasko).
2. Alle tilbudene har  **samme sannsynlighet**   $p$  for å bli akseptert, hvor  $p = 0.2667$  i vårt tilfelle.
3. I oppgaven antas det at tilbudene er **uavhengig** av hverandre.
4. Det gjennomføres et bestemt antall forsøk,  $n = 4$  i vårt tilfelle.

Forsøksserien oppfyller dermed kravene til en **binomisk** forsøksserie.

Den stokastiske variabelene  $Y$  er da binomisk fordelt:  $Y \sim \text{Bin}[n = 4, p = 0.2667]$ .

b) Siden  $Y$  binomisk fordelt,  $Y \sim \text{Bin}[n = 4, p = 0.2667]$ , så finner vi formelen for sannsynlighet i formelsamlingen:

$$\underline{\underline{P(Y \geq 2)}} = 1 - P(Y = 0) - P(Y = 1) \quad (7)$$

$$= 1 - \binom{n}{0} p^0 (1-p)^{n-0} - \binom{n}{1} p^1 (1-p)^{n-1} \quad (8)$$

$$= 1 - \binom{4}{0} 0.2667^0 (1 - 0.2667)^{4-0} - \binom{4}{1} 0.2667^1 (1 - 0.2667)^{4-1} \quad (9)$$

$$= 1 - 0.2892 - 0.4207 \quad (10)$$

$$= \underline{\underline{0.2901}} \quad (11)$$

- c) Siden  $Y$  er binomisk fordelt,  $Y \sim \text{Bin}[n = 4, p = 0.2667]$ , så finner vi formelen for forventning

$$\underline{\underline{E[Y]}} = n \cdot p = 4 \cdot 0.2667 = \underline{\underline{1.0668}} \quad (12)$$

Tolkning:

$$\underline{\underline{E[Y]}} = \underline{\underline{\text{forventet antall tilbud som blir akseptert av de 4 tilbudene}}}$$

- d) Bruker regnereglene for forventningen i formelsamlingen og finner forventningen av  $V$ :

$$E[V] = E[v_1Y_1 + v_2Y_2 + v_3Y_3 + v_4Y_4] \quad (13)$$

$$= v_1E[Y_1] + v_2E[Y_2] + v_3E[Y_3] + v_4E[Y_4] \quad (14)$$

hvor  $E[Y_1] = E[Y_2] = E[Y_3] = E[Y_4] = p$  (oppgitt i en fotnote i oppgaven). Dermed:

$$\underline{\underline{E[V]}} = v_1p + v_2p + v_3p + v_4p \quad (15)$$

$$= \underline{\underline{(v_1 + v_2 + v_3 + v_4)p}} \quad , \quad \text{q.e.d.} \quad (16)$$

hvor  $p$  er en felles faktor som vi faktoriserer utenfor.

e) Bruker lign.(16) og setter inn tallene for  $v_i$  og  $p$  som oppgitt i oppgaven:

$$\underline{\underline{E[V]}} = (v_1 + v_2 + v_3 + v_4) p \quad (17)$$

$$= (40\,000 + 33\,000 + 105\,000 + 75\,000) 0.2667 \text{ NOK} = \underline{\underline{67\,475 \text{ NOK}}} \quad (18)$$

f) Bruker regnereglene for varians i formelsamlingen og finner variansen av  $V$ :

$$\text{Var}[V] = \text{Var}[v_1 Y_1 + v_2 Y_2 + v_3 Y_3 + v_4 Y_4] \quad (19)$$

$$\stackrel{\text{uavh.}}{=} v_1^2 \text{Var}[Y_1] + v_2^2 \text{Var}[Y_2] + v_3^2 \text{Var}[Y_3] + v_4^2 \text{Var}[Y_4] \quad (20)$$

siden kovariansen er null da alle  $Y_i$ 'ene er uavhengige.

I en fotnote i oppgaven er det oppgitt at:  $(Y_i \sim \text{Bin}[n_i = 1, p])$

$$\text{Var}[Y_1] = \text{Var}[Y_2] = \text{Var}[Y_3] = \text{Var}[Y_4] = p(1-p) \quad (21)$$

Variansen blir da:

$$\underline{\underline{\text{Var}[V]}} = v_1^2 p(1-p) + v_2^2 p(1-p) + v_3^2 p(1-p) + v_4^2 p(1-p) \quad (22)$$

$$= \underline{\underline{(v_1^2 + v_2^2 + v_3^2 + v_4^2) p(1-p)}} \quad , \quad \text{q.e.d.} \quad (23)$$

hvor  $p(1-p)$  er en felles faktor som vi faktoriserer utenfor.

- g) Siden alle verdiene  $v_i$  er forskjellige så er også forventingsverdiene til leddene i lineærkombinasjonen forskjellige.

Dermed er en av forutsetningene til sentralgrenseteoremet brutt.

For at sentralgrenseteoremet skal gjelde må forventingsverdiene til leddene i lineær kombinasjonen være like.

Dermed er ikke summen  $V = v_1Y_1 + v_2Y_2 + \dots + v_{200}Y_{200}$  normalfordelt selv om tilbudene, dvs.  $Y_i$ 'ene, er uavhengige og aksept-sannsynlighetene  $p$  er like.



**Oppgave 3:** ( logistikk og økonomi )

a) Forventet etterspørsel av antall kurver med jordbær per dag på dagtid:

$$\underline{\underline{E[D_{\text{tot}}]}} = E[D_1 + D_2 + D_3 + D_4] \quad (24)$$

$$= E[D_1] + E[D_2] + E[D_3] + E[D_4] \quad (25)$$

$$= 20 + 50 + 30 + 10 \quad (26)$$

$$= \underline{\underline{110}} \quad , \quad \text{q.e.d.} \quad (27)$$

Merk:

Overgangen i lign.(25) gjelder **alltid**, uansett om  $D_i$ 'ene er uavhengige eller ikke.

b) Siden de stokastiske variablene  $D_i$  er **uavhengige**, så er tilhørende kovarians lik null. Variansen blir dermed: ( se formelsamlingen )

$$\underline{\underline{Var[D_{\text{tot}}]}} = Var[D_1 + D_2 + D_3 + D_4] \quad (28)$$

$$\stackrel{\text{uavh.}}{=} Var[D_1] + Var[D_2] + Var[D_3] + Var[D_4] \quad (29)$$

$$= 16 + 36 + 88 + 4 \quad (30)$$

$$= \underline{\underline{144}} \quad (31)$$

- c) Det er oppgitt i oppgaven at alle  $D_i$  er normalfordelte:

$$D_i \sim N[ E[D_i], Var[D_i] ] \quad (32)$$

I tillegg er det oppgitt i oppgaven at alle  $D_i$ 'ene er uavhengige. <sup>4</sup>

Fra læresetningen i formelsamlingen <sup>5</sup> vet vi da at summen  $D_{\text{tot}} = D_1 + D_2 + D_3 + D_4$  også er normalfordelt, dvs.:

$$\underline{\underline{D_{\text{tot}} \sim N[ E[D_{\text{tot}}], Var[D_{\text{tot}}] ]}} \quad (33)$$

hvor  $E[D_{\text{tot}}] = 110$  og  $Var[D_{\text{tot}}] = 144$  fra oppgave **3a** og **3b**, henholdsvis.

---

<sup>4</sup>At  $D_i$ 'ene er uavhengige er det samme som å si at etterspørselen av jordbær i de 4 butikkene er uavhengige.

<sup>5</sup>Se side 108 i formelsamlingen fra 2018.



d) Med innkjøpspris (kostnad for Bunnpris)  $c = 30$  NOK og utslagspris  $p = 40$  NOK fås:

$$P(D_{\text{tot}} \leq q^*) = 1 - \frac{c}{p} \quad (34)$$

$$P(D_{\text{tot}} \leq q^*) = 1 - \frac{30}{40} \quad (35)$$

$$P(D_{\text{tot}} \leq q^*) = 0.25 \quad (36)$$

$$P\left(\underbrace{\frac{D_{\text{tot}} - E[D_{\text{tot}}]}{\sigma[D_{\text{tot}}]}}_{\equiv Z} \leq \underbrace{\frac{q^* - E[D_{\text{tot}}]}{\sigma[D_{\text{tot}}]}}_{\equiv z_0}\right) \stackrel{\text{standardiser}}{=} 0.25 \quad (37)$$

$$P(Z \leq z_0) = 0.25 \quad (38)$$

$z_0$  finnes ved “[omvendt tabelloppslag](#)”.

Men sannsynligheten  $P(Z \leq z_0) = 0.25$  finner ikke i tabellen i formelsamlingen.

Sannsynlighetene i formelsamlingen ”stopper” ved 0.5000.

Men av symmetri grunner så finner man  $z_0$  er negativ og gitt ved:

$$z_0 = -0.675 \quad (39)$$

Vi løser:

$$z_0 = \frac{q^* - E[D_{\text{total}}]}{\sigma[D_{\text{tot}}]} \quad (40)$$

med hensyn på  $q^*$ :

$$\underline{q^*} = E[D_{\text{tot}}] + z_0 \cdot \sigma[D_{\text{tot}}] \quad (41)$$

$$= 110 - 0.675 \cdot \sqrt{144} = \underline{101.9} \quad (42)$$

Bunnpris må bestille  $q^* \approx 102$  kurver med jordbær per gang for å få størst mulig forventet fortjeneste.

- e) For å finne  $p_{\text{red}}$  må vi først regne ut sannsynligheten.  
Deretter kan vi regne ut  $p_{\text{red}}$  via den oppgitte ligningen i oppgaveteksten.

Fra oppgave **3c** vet vi at  $D_{\text{tot}}$  er normalfordelt.

Fra oppgave **3a** og **3b** vet vi at  $E[D_{\text{tot}}] = 110$  og  $\sigma[D_{\text{tot}}] = \sqrt{\text{Var}[D_{\text{tot}}]} = \sqrt{144} = 12$ .

Dermed kan vi regne ut sannsynligheten  $P(D_{\text{tot}} \leq q^*)$  via standardisering og tabelloppslag:

$$\underline{P(D_{\text{tot}} \leq q^*)} = P\left(\underbrace{\frac{D_{\text{tot}} - E[D_{\text{tot}}]}{\sigma[D_{\text{tot}}]}}_{=Z} \leq \frac{q^* - E[D_{\text{tot}}]}{\sigma[D_{\text{tot}}]}\right) \quad (43)$$

$$= P\left(Z \leq \frac{116 - 110}{12}\right) \quad (44)$$

$$= P(Z \leq 0.5) \quad (45)$$

$$\stackrel{\text{tabell}}{=} \underline{0.6915} \quad (46)$$

Ligningen oppgitt i oppgaveteksten er:

$$P(D_{\text{tot}} \leq q^*) = \frac{p - c}{p - p_{\text{red}}} \quad (47)$$

$$p - p_{\text{red}} = \frac{p - c}{P(D_{\text{tot}} \leq q^*)} \quad (48)$$

som gir:

$$\underline{p_{\text{red}}} = p - \frac{p - c}{P(D_{\text{tot}} \leq q^*)} \quad (49)$$

$$= \left(40 - \frac{40 - 30}{0.6915}\right) \text{ NOK} = \underline{25.5 \text{ NOK}} \quad (50)$$

Bunnpris må sette den reduserte prisen til  $p_{\text{red}} = 25.5 \text{ NOK}$   
for å maksimere den forventede fortjenesten.

■

Oppgave 4: (økonomi)

- a) i)  $R_{xy}$  er normalisert og ligger mellom  $-1$  og  $1$ , dvs.:  $\underline{\underline{-1 < R_{xy} < 1}}$ .
- ii)  $R_{xy}$  er et mål på lineær korrelasjon mellom observasjonene  $x$  og  $y$ .  
Det er et mål på i hvor stor grad det er en lineær sammenheng mellom observasjonene  $x$  og  $y$ .
- iii)  $R_{xy}$  er enhetsuavhengig, dvs. ingen enhet.<sup>6</sup>

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

$$\underline{\underline{R_{xy}}} = \frac{S_{xy}}{S_x \cdot S_y} \quad (51)$$

$$= \frac{2\,084\,821}{\sqrt{37\,364} \cdot \sqrt{133\,388\,393}} = \underline{\underline{0.93}} \quad (52)$$

- c) At  $R_{xy} = 0.93$  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.

---

<sup>6</sup>Dette betyr at  $R_{xy}$  har samme verdi uansett hva slags enhet man bruker for å regne ut  $R_{xy} = S_{xy}/(S_x \cdot S_y)$ .

- d) Vi skal finne regresjonslinjen. Formelen for minste kvadraters lineære [regresjonslinje](#) står i formelsamlingen:

$$\hat{y} = \hat{\alpha} + \hat{\beta}x \quad (53)$$

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

$$\hat{\beta} = \frac{S_{xy}}{S_x^2} = \frac{2084821}{37364} = \underline{55.8} \quad (54a)$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 43437.5 - 55.8 \cdot 607.5 = \underline{9539} \quad (54b)$$

Minste kvadraters lineære [regresjonslinje](#)  $\hat{y} = \hat{\alpha} + \hat{\beta}x$  blir dermed:

$$\underline{\underline{\hat{y} = 9539 + 55.8x}} \quad (55)$$

- e) At en person tjener 1 million NOK tilsvarer  $x = 1000$  i antall 1000 NOK:  
Bruker regresjonslinjen  $\hat{y} = \hat{y}(x)$  fra oppgave [4d](#), dvs. lign.(55):

$$\underline{\hat{y}(1000)} = \left( 9539 + 55.8 \cdot 1000 \right) \text{ NOK/m}^2 = \underline{65338 \text{ NOK/m}^2} \quad (56)$$

Ifølge regresjonslinjen så kommer kjøperent til å en leilighet med en kvadratmeterpris på utenfor sentrum er 65 338 NOK/m<sup>2</sup>.

- f) Gjennomsnittlig kvadratmeterpris i Norge er 30 000 NOK/ $m^2$ . Ifølge regresjonslinjen, når  $\hat{y}(x) = 30\,000$ , er da:

$$30\,000 = 9\,539 + 55.8x \quad (57)$$

Løser med hensyn på  $x$ :

$$\underline{x} = \frac{30\,000 - 9\,539}{55.8} = \underline{367} \quad (58)$$

Ifølge regresjonslinjen har kjøperen av leiligheten en årsinntekt på 367 000 NOK.

- g) Forklaringskraften  $R^2$  kan leses direkte fra Excel-utskriften: ( Se cellen som heter “*R Square*” i Excel-utskriften ): <sup>7</sup>

$$\underline{\underline{R^2 = 0.8721}} \quad (59)$$

- h) Kommentaar til svaret i oppgave 4g:

At  $R^2 = 0.8721$  betyr at regresjonslinjen vil i “i stor grad” **forutsi** kvadratmeterprisen på en leilighet for en person med en gitt årslønn.

Modellen har stor forklaringskraft, tilsvarende  $R^2 = 0.8721$  .

■

---

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