

LØSNING: Eksamen 10. sept. 2018

“MAT110 Statistikk 1”, vår 2018

Oppgave 1: (logistikk)

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

$$\sum_{i=1}^8 p_i = \frac{5}{350} + \frac{40}{350} + \frac{25}{350} + \frac{80}{350} + \frac{8}{350} + \frac{68}{350} + \frac{22}{350} + p_8 = 1 \quad (1)$$

Løser med hensyn på p_8 :

$$\underline{p_8} = 1 - \left(\frac{5}{350} + \frac{40}{350} + \frac{25}{350} + \frac{80}{350} + \frac{8}{350} + \frac{68}{350} + \frac{22}{350} \right) \quad (2)$$

$$= 1 - \frac{248}{350} \quad (3)$$

$$= \frac{350}{350} - \frac{248}{350} \quad (4)$$

$$= \frac{350 - 248}{350} \quad (5)$$

$$= \frac{102}{\underline{350}}, \quad \text{q.e.d.} \quad (6)$$

- b) Fra figuren tilhørende oppgaven (utfallsrom) ser vi at utfallene i utfallsrommet ikke overlapper.¹
Derfor er utfallene disjunkte.
- 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.²
- d) Definisjonen av begivenheten for aksepterte tilbud A :

$$\underline{\underline{A = \left\{ (1, h, a), (1, \bar{h}, a), (2, h, a), (2, \bar{h}, a) \right\}}}$$

Definisjonen av begivenheten for høye tilbud H :

$$\underline{\underline{H = \left\{ (1, h, a), (1, h, \bar{a}), (2, h, a), (2, h, \bar{a}) \right\}}} \quad (7)$$

¹I likhet med brikkene i et puslespill så overlapper utfallene i utfallsrommet akkurat ikke.

²Fra figuren tilhørende oppgave 3 i oppgaveteksten (urne) ser vi også at kulene har forskjellige størrelse. Det visualiserer at sannsynlighetene p_i er forskjellige.

- e) Ved å bruke sannsynlighetene oppgitt i lign.(1) i oppgaveteksten og samme fremgangsmåten som i lign.(5) i oppgaveteksten, så innser man at begivenheten H for høye tilbud er gitt ved:

$$\begin{aligned}\underline{\underline{P(H)}} &= p_1 + p_3 + p_5 + p_7 \\ &= \frac{5}{350} + \frac{25}{350} + \frac{8}{350} + \frac{22}{350} = \frac{60}{350} = \underline{\underline{0.1714}}\end{aligned}\tag{8}$$

- f) Fra første ligningen i oppgaveteksten finner vi:

$$P(S_1 \cap H \cap A) = p_1 = \frac{5}{350}\tag{9}$$

og

$$P(S_1 \cap H) = p_1 + p_3 = \frac{5}{350} + \frac{25}{350} = \frac{30}{350}\tag{10}$$

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

$$\underline{\underline{P(A|S_1 \cap H)}} = \frac{P(S_1 \cap H \cap A)}{P(S_1 \cap H)} = \frac{\frac{5}{\cancel{350}}}{\frac{30}{\cancel{350}}} = \frac{5}{30} \quad (= 0.1667)\tag{11}$$

- e) Tolkning: ³

$$\underline{\underline{P(A|S_1 \cap H)}} = \text{sannsynligheten for at et tilbud blir akseptert gitt at tilbudet er}$$

$$\underline{\underline{\text{laget av selger 1 og verdien av tilbudet er høyt}}}$$

■

³Det 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 $p = 0.1667$ 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 variablene Y er da binomisk fordelt: $Y \sim \text{Bin}[n = 4, p = 0.1667]$.

b) Siden Y binomisk fordelt, $Y \sim \text{Bin}[n = 4, p = 0.1667]$, så finner vi formelen for sannsynlighet i formelsamlingen:

$$\underline{\underline{P(Y = 2)}} = \binom{n}{2} p^1 (1 - p)^{n-2} \quad (12)$$

$$= \binom{4}{2} 0.1667^1 (1 - 0.1667)^{4-2} = \underline{\underline{0.1158}} \quad (13)$$

- c) Siden Y er binomisk fordelt, $Y \sim \text{Bin}[n = 4, p = 0.1667]$, så finner vi formelen for forventning

$$\underline{\underline{E[Y]}} = n \cdot p = 4 \cdot 0.1667 = \underline{\underline{0.6668}} \quad (14)$$

Tolkning:

$$\begin{aligned} \underline{\underline{E[Y]}} &= \text{forventet antall tilbud som blir akseptert} \\ &\quad \underline{\underline{\text{av de 4 høyre tilbudene laget av selger 1}}} \end{aligned} \quad (15)$$

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

$$E[V] = E[v_1 Y_1 + v_2 Y_2 + v_3 Y_3 + v_4 Y_4] \quad (16)$$

$$= v_1 E[Y_1] + v_2 E[Y_2] + v_3 E[Y_3] + v_4 E[Y_4] \quad (17)$$

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

$$\underline{\underline{E[V]}} = v_1 p + v_2 p + v_3 p + v_4 p \quad (18)$$

$$= \underline{\underline{(v_1 + v_2 + v_3 + v_4) p}}, \quad \text{q.e.d.} \quad (19)$$

hvor p er en felles faktor som vi faktoreriserer utenfor.

- e) Bruker lign.(19) og setter inn tallene for v_i og p som oppgitt i oppgaven:

$$\underline{\underline{E[V]}} = (v_1 + v_2 + v_3 + v_4) p \quad (20)$$

$$= (40\,000 + 33\,000 + 105\,000 + 75\,000) \cdot 0.1667 \text{ NOK} = \underline{\underline{42\,175 \text{ NOK}}} \quad (21)$$

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

$$\text{Var}[V] = \text{Var}[v_1 Y_1 + v_2 Y_2 + v_3 Y_3 + v_4 Y_4] \quad (22)$$

$$\stackrel{\text{uavh.}}{=} v_1^2 \text{Var}[Y_1] + v_2^2 \text{Var}[Y_2] + v_3^2 \text{Var}[Y_3] + v_4^2 \text{Var}[Y_4] \quad (23)$$

siden alle Y_i 'ene er uavhengige.

I en fotnote i oppgaven er det oppgitt at: $(Y_i \sim \text{Bin}[n_i = 1, p])$ ⁴

$$\text{Var}[Y_1] = \text{Var}[Y_2] = \text{Var}[Y_3] = \text{Var}[Y_4] = p(1-p) \quad (24)$$

Variansen blir da:

$$\underline{\underline{\text{Var}[V]}} = v_1^2 p(1-p) + v_2^2 p(1-p) + v_3^2 p(1-p) + v_4^2 p(1-p) \quad (25)$$

$$= \underline{\underline{(v_1^2 + v_2^2 + v_3^2 + v_4^2) p(1-p)}} \quad , \quad \text{q.e.d.} \quad (26)$$

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

⁴En *bernulfordeling* er bare et spesialtilfelle av en binomialfordelingen: en bernulfordeling er en binomialfordelingen med $n_i = 1$. Siden bernulfordeling ikke er nevnt eksplisitt i forelesningene eller kompendiet så brukes navnet "binomialfordelingen" både i oppgaven og løsningen.

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

Dermed er en av forutsetningene til sentralgrenseteoremet brutt ⁵ og sentralgrenseteoremet kan ikke garantere at V blir normalfordelt.

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.

■

⁵For at sentralgrenseteoremet skal gjelde så må sannsynlighetsfordelingene til de stokastiske variablene V_i i en lineærkombinasjon være uavhengige og identiske slik at forventningene og variansene i lineær kombinasjonen er like. De stokastiske variablene Y_i er faktisk identiske og uavhengige, men forventingsverdiene og variansene til de ulike leddene, dvs. til V_i , i lineær kombinasjonen er forskjellige siden v_i er forskjellige.

Oppgave 3: (logistikk og økonomi)

- a) Forventet etterspørsel av antall kg med bananer per dag på dagtid:

$$\underline{\underline{E[D_{\text{tot}}]}} = E[D_1 + D_2 + D_3 + D_4] \quad (27)$$

$$= E[D_1] + E[D_2] + E[D_3] + E[D_4] \quad (28)$$

$$= 40 + 100 + 60 + 20 \quad (29)$$

$$= \underline{\underline{220}}, \quad \text{q.e.d.} \quad (30)$$

Merk:

Overgangen i lign.(28) 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)

$$\underline{\underline{Var[D_{\text{tot}}]}} = Var[D_1 + D_2 + D_3 + D_4] \quad (31)$$

$$\stackrel{\text{uavh.}}{=} Var[D_1] + Var[D_2] + Var[D_3] + Var[D_4] \quad (32)$$

$$= 64 + 144 + 352 + 16 \quad (33)$$

$$= \underline{\underline{576}} \quad (34)$$

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

$$D_i \sim N[E[D_i], Var[D_i]] \quad (35)$$

I tillegg er det oppgitt i oppgaven at alle D_i 'ene er uavhengige.⁶

Fra læresetningen i formelsamlingen⁷ vet vi da at summen

$D_{\text{tot}} = D_1 + D_2 + D_3 + D_4$ også er normalfordelt, dvs.:

$$\underline{\underline{D_{\text{tot}} \sim N[E[D_{\text{tot}}], Var[D_{\text{tot}}]]}} \quad (36)$$

hvor $E[D_{\text{tot}}] = 220$ og $Var[D_{\text{tot}}] = 576$ fra oppgave **3a** og **3b**, henholdsvis.

⁶At D_i 'ene er uavhengige er det samme som å si at etterspørselen av bananer i de 4 butikkene er uavhengige.

⁷Se side 108 i formelsamlingen fra 2018.

- d) Med innkjøpspris (kostnad for Kiwi) $c = 8$ NOK og utslagspris $p = 12.9$ NOK fås:

$$P(D_3 \leq q_3^*) = 1 - \frac{c}{p} \quad (37)$$

$$P(D_3 \leq q_3^*) = 1 - \frac{8}{12.9} \quad (38)$$

$$P(D_3 \leq q^*) = 0.38 \quad (39)$$

$$P\left(\underbrace{\frac{D_3 - E[D_3]}{\sigma[D_3]}}_{\equiv Z} \leq \underbrace{\frac{q_3^* - E[D_3]}{\sigma[D_3]}}_{\equiv z_0}\right) \stackrel{\text{standardiser}}{=} 0.38 \quad (40)$$

$$P(Z \leq z_0) = 0.38 \quad (41)$$

z_0 finnes ved “[omvendt tabeloppslag](#)”. Men z_0 var oppgitt i oppgaven. Dermed slipper man å gjøre dette omvendte tabeloppslaget. ⁸

$$z_0 = -0.30 \quad (42)$$

Vi løser:

$$z_0 = \frac{q_3^* - E[D_3]}{\sigma[D_3]} \quad (43)$$

med hensyn på q^* :

$$\underline{q_3^*} = E[D_3] + z_0 \cdot \sigma[D_3] \quad (44)$$

$$= 60 - 0.30 \cdot \sqrt{352} = \underline{54.4} \quad (45)$$

Bunnpris må bestille $q_3^* \approx 54.4$ kg med banan per gang for å få størst mulig forventet fortjeneste.

⁸Følgende kommentar behøver man ikke ha med i eksamensbesvarelsen fordi $z_0 = -0.30$ er oppgitt: Sannsynligheten $P(Z \leq z_0) = 0.38$ finnes ikke direkte i tabellen i formelsamlingen fordi sannsynlighetene i formelsamlingen ikke viser sannsynligheter under 0.5000. Siden $P(Z = z_0)$ er symmetrisk om origo så er $P(Z \leq -z_0) = P(Z \geq z_0)$. (Helt generelt er en funksjon $f(x)$ symmetrisk om y -aksen dersom $f(x) = f(-x)$.) Arealet under en sannsynlighetsfunksjon $P(Z = z_0)$ er en sannsynlighet som er normalisert, dvs. $P(Z \leq z_0) + P(Z \geq z_0) = 1$. Ved å kombinere disse sammenhengene gir: $P(Z \leq -z_0) = 1 - 0.38 = 0.62$.

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

Fra oppgaveteksten vet vi at D_i er normalfordelt.

Fra tabellen i oppgave 3 vet vi at $E[D_3] = 60$ og $\sigma[D_3] = \sqrt{\text{Var}[D_3]} = \sqrt{352}$.

Dermed kan vi regne ut sannsynligheten $P(D_3 \leq q_3^*)$ via standardisering og tabelloppslag:

$$\underline{P(D_3 \leq q^*)} = P\left(\underbrace{\frac{D_3 - E[D_3]}{\sigma[D_3]}}_{=Z} \leq \frac{q_3^* - E[D_3]}{\sigma[D_3]}\right) \quad (46)$$

$$= P\left(Z \leq \frac{70 - 60}{\sqrt{352}}\right) \quad (47)$$

$$= P(Z \leq 0.53) \quad (48)$$

$$\stackrel{\text{tabell}}{=} \underline{0.7019} \quad (49)$$

Ligningen oppgitt i oppgaveteksten er:

$$P(D_3 \leq q_3^*) = \frac{p - c}{p - p_{\text{red}}} \quad (50)$$

$$p - p_{\text{red}} = \frac{p - c}{P(D_3 \leq q_3^*)} \quad (51)$$

som gir:

$$\underline{p_{\text{red}}} = p - \frac{p - c}{P(D_3 \leq q_3^*)} \quad (52)$$

$$= \left(12.9 - \frac{12.9 - 8}{0.7019}\right) \text{ NOK} = \underline{5.92 \text{ NOK}} \quad (53)$$

Bunnpris må sette den reduserte prisen til $p_{\text{red}} = 5.92 \text{ NOK}$
for å maksimere den forventede fortjeningen.

■

Oppgave 4: (økonomi)

- a) Vi bruker læresetningen på side 66 i formelsamlingen for å finne minste kvadraters **regresjonslinje** for x og y . Parametrene $\hat{\beta}$ og $\hat{\alpha}$ er da: (dropper benevning her)

$$\underline{\hat{\beta}} = \frac{S_{xy}}{S_x^2} = \frac{18}{2.5} = \underline{7.2} \quad (54)$$

$$\underline{\hat{\alpha}} = \bar{y} - \beta\bar{x} = 284 - 7.2 \cdot 3 = \underline{262.4} \quad (55)$$

Minste kvadraters lineære regresjonslinje $\hat{y} = \hat{\alpha} + \hat{\beta}x$ blir dermed:
(se lign.(11.7) i formelsamlingen fra 2018)

$$\underline{\underline{\hat{y} = 262.4 + 7.2x}} \quad (56)$$

- b) i) En regresjonslinje mellom variablene x og y sier at for en gitt verdi av x så kan y estimeres/predikeres via regresjonslinjen.
- ii) Parameteren $\hat{\beta}$ er stigningstallet for en lineær regresjonslinje.
Parameteren angir estimert/predikert endring $\hat{y}$ når x endres med en enhet.
- iii) Parameteren $\hat{\beta} = 7.2$ i oppgave 4a.
Det betyr at aksjeprisen til Telenor øker med 7.2 NOK per dag.

- c) Regresjonslinjen i lign.(56) predikerer at aksjekursen til Telenor er:

$$\underline{\underline{\hat{y}(12)}} = (262.4 + 7.2 \cdot 12) \text{ NOK} = \underline{\underline{348.8 \text{ NOK}}} \quad (57)$$

etter 12 dager.

d) i) Forklaringsstyrken R^2 er normalisert til: $0 \leq R^2 \leq 1$.

ii) Forklaringsstyrken R^2 er enhetsuavhengig, den er benevningsløs.

iii) Forklaringsstyrken R^2 er et mål på:

- i hvor stor grad x_i kan brukes til å **predikere** y_i
- hvor godt de faktiske observasjonene y_i **passer** med lineær regresjonen $\hat{y}_i$
- styrken i **samvariasjon** mellom prediksjonene $\hat{y}_i$ og de faktiske observasjonene y_i

Det er nok at man nevner én eller lignende av disse kulepunkt-kommentarene.

e) Forklaringsstyrken R^2 kan leses direkte fra Excel-utskriften:
(Se cellen som heter “*R Square*” i Excel-utskriften): ⁹

$$\underline{\underline{R^2 = 0.9744}} \quad (58)$$

f) Kommentar til svaret i oppgave 4e:

At $R^2 = 0.9744$ betyr at for en gitt dag innenfor den aktuelle perioden hvor den lineære trenden gjelder så ligger observasjonene svært tett på regresjonslinjen. ¹⁰

⁹Man kan også regne ut forklaringsstyrken 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.

¹⁰Litt mer presist og litt mer teknisk: At $R^2 = 0.9744$ betyr at 97.44 % av variansen til observasjonene ligger innenfor regresjonslinjen. (Denne fotnoten er bare en ekstrainformasjon som ikke kreves at studentene har med i sin eksamensbesvarelse.

At $R^2 = 0.9744$ betyr at for en gitt dag innenfor perioden hvor den lineære trenden gjelder så kan vi “i stor grad”, tilsvarende 97.44 %, **forutsi** aksjekursen. Vi sier at modellen har stor forklaringsstyrke.

