LØSNING: Eksamen 22. mai 2018

"*MAT110 Statistikk 1*", vår 2018

Oppgave 1: (logistikk)

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

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

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

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

b) Fra figuren i oppgaven ser vi at de to begivenhetene S_1 og S_2 ikke overlapper. [1] Derfor er de disjunkte.

[1] 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 $\underline{\underline{\text{ulike}}}$,
altså $\underline{\underline{\text{ikke uniformt utfallsrom}}}$, så gjelder ikke urnemodellen for vår sannsynlighetsmodell. [2]

d) Fra første ligningen i oppgaveteksten finner vi:

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

og

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

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

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

e) Tolkning: [3]

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

■

[2] Fra figuren i oppgaven ser vi også at kulene har forskjellige størrelse. Det visualiserer at sannsynlighetene p_i er forskjellige.

[3] 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 variablene Y er da binomisk fordelt: $Y \sim \text{Bin}\big[\, n = 4, p = 0.2667 \,\big]$.

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

$$\underline{\underline{P(X \geq 2)}} = 1 - P(X = 0) - P(X = 1) \tag{7}$$

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

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

$$= 1 - 0.2892 - 0.4207 \tag{10}$$

$$= \underline{\underline{0.2901}} \tag{11}$$

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

$$\underline{\underline{E[Y]}} \;=\; n \cdot p \;=\; 4 \cdot 0.2667 \;=\; \underline{\underline{1.0668}} \tag{12}$$

Tolkning:

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

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

$$E[V] \;=\; E\big[\, v_1Y_1 \;+\; v_2Y_2 \;+\; v_3Y_3 \;+\; v_4Y_4 \,\big] \tag{13}$$

$$\;=\; v_1E[Y_1] \;+\; v_2E[Y_2] \;+\; v_3E[Y_3] \;+\; v_4E[Y_4] \tag{14}$$

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

$$\underline{\underline{E[V]}} \;=\; v_1p + v_2p + v_3p + v_4p \tag{15}$$

$$\;=\; \underline{\underline{\big(v_1 + v_2 + v_3 + v_3\big)\, p}} \qquad , \qquad \text{q.e.d.} \tag{16}$$

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:

$$\underline{\underline{E[V]}} = \big(v_1 + v_2 + v_3 + v_3\big)\, p \tag{17}$$

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

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

$$Var[V] = Var\big[\, v_1Y_1 + v_2Y_2 + v_3Y_3 + v_4Y_4 \,\big] \tag{19}$$

$$\overset{\text{uavh.}}{=} v_1^2 Var[Y_1] \ + \ v_2^2 Var[Y_2] \ + \ v_3^2 Var[Y_3] \ + \ v_4^2 Var[Y_4] \tag{20}$$

siden covariansen er null da alle Y_i'ene er uavhengige.
I en fotnote i oppgaven er det oppgitt at: ($Y_i \ \sim \ \text{Bin}[\, n_i = 1, p\,]$)

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

Variansen blir da:

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

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

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 forventingensverdiene til leddene i lineær kombinasjonen være like.

Dermed er ikke summen $V = v_1Y_1 + v_2Y_2 + ... + v_{200}Y_{200}$ 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:

$$\underline{\underline{E[D_{\text{tot}}]}} = E\big[\, D_1 + D_2 + D_3 + D_4 \,\big] \tag{24}$$

$$= E[D_1] + E[D_2] + E[D_3] + E[D_4] \tag{25}$$

$$= 20 + 50 + 30 + 10 \tag{26}$$

$$= \underline{\underline{110}} \qquad , \qquad \text{q.e.d.} \tag{27}$$

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)

$$\underline{\underline{Var[D_{\text{tot}}]}} = Var\big[\, D_1 + D_2 + D_3 + D_4 \,\big] \tag{28}$$

$$\overset{\text{uavh.}}{=} Var[D_1] + Var[D_2] + Var[D_3] + Var[D_4] \tag{29}$$

$$= 16 + 36 + 88 + 4 \tag{30}$$

$$= \underline{\underline{144}} \tag{31}$$

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

$$D_i \;\sim\; N\big[\, E[D_i]\,,\, Var[D_i]\,\big] \tag{32}$$

I tillegg er det oppgitt i oppgaven at alle D_i'ene er uavhengige. [4]
Fra læresetningen i formelsamlingen [5] vet vi da at summen
$D_{\text{tot}} = D_1 + D_2 + D_3 + D_4$ også er normalfordelt, dvs.:

$$\underline{\underline{D_{\text{tot}} \;\sim\; N\big[\, E[D_{\text{tot}}]\,,\, Var[D_{\text{tot}}]\,\big]}}7 \tag{33}$$

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

[4] At D_i'ene er uavhengige er det samme som å si at etterspørselen av jordbær i de 4 butikkene er uavhengige.
[5] Se side 108 i formelsamlingen fra 2018.

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

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

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

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

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

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

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 symmetrigrunner så finner man z_0 er negativ og gitt ved:

$$z_0 \quad = \quad -0.675 \tag{39}$$

Vi løser:

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

med hensyn på q^*:

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

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

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

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 oppgave **3c** vet vi at D_{tot} er normalfordelt.
Fra oppgave **3a** og **3b** vet vi at $E[D_{\text{tot}}] = 110$ og $\sigma[D_{\text{tot}}] = \sqrt{Var[D_{\text{tot}}]} = \sqrt{144} = 12$.
Dermed kan vi regne ut sannsynligheten $P(D_{\text{tot}} \leq q^*)$ via standardisering og tabelloppslag:

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

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

$$= P(Z \leq 0.5) \tag{45}$$

$$\overset{\text{tabell}}{=} \underline{0.6915} \tag{46}$$

Ligningen oppgitt i oppgaveteksten er:

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

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

som gir:

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

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

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

■

Oppgave 4: (økonomi)

a) **i**) R_{xy} er normalisert og ligger mellom -1 og 1, dvs.: $\underline{\underline{-1 \leq R_{xy} \leq 1}}$.

ii) R_{xy} er et mål på <u>lineær korrelasjon</u> mellom obervasjonene x og y.
Det er et mål på i hvor stor grad det er en lineær sammmenheng mellom obervasjonene x og y.

iii) R_{xy} er <u>enhetsuavhengig</u>, dvs. ingen enhet. [6]

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

$$\underline{\underline{R_{xy}}} = \frac{S_{xy}}{S_x \cdot S_y} \tag{51}$$

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

c) At $R_{xy} = 0.93$ betyr at det er en <u>sterk positiv korrelasjon</u> mellom x og y, dvs. store x hører sammen med store y og omvendt.
Det er altså en <u>sterk lineær sammenheng</u> mellom x og y med positivt stigningstall.

[6] 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:

$$\hat{y} \;=\; \hat{\alpha} \;+\; \hat{\beta}x \tag{53}$$

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

$$\underline{\hat{\beta}} \;=\; \frac{S_{xy}}{S_x^2} \;=\; \frac{2\,084\,821}{133\,388\,393} \;=\; \underline{55.8} \tag{54a}$$

$$\underline{\hat{\alpha}} \;=\; \overline{y} \;-\; \beta\overline{x} \;=\; 43\,437.5 - 55.8 \cdot 607.5 \;=\; \underline{9\,539} \tag{54b}$$

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

$$\underline{\underline{\hat{y} \;=\; 9\,539 \;+\; 55.8\,x}} \tag{55}$$

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):

$$\underline{\hat{y}(1000)} \;=\; \Big(9\,539 \;+\; 55.8 \cdot 1000\Big)\;\;\text{NOK/m}^2 = \;\underline{65\,338\;\;\text{NOK/m}^2} \tag{56}$$

Ifølge regresjonslinjen så kommer kjøperent til å en leilighet med en kvadratmeterpris på utenfor sentrum er $\underline{\underline{65\,338\ \text{NOK/m}^2}}$.

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

$$30\,000 \quad = \quad 9\,539 \;+\; 55.8\,x \tag{57}$$

Løser med hensyn på x:

$$\underline{x} \;=\; \frac{30\,000 \;-\; 9\,539}{55.8} \;=\; \underline{367} \tag{58}$$

Ifølge regresjonslinjen har kjøperen av leiligheten en årsinntekt på $\underline{\underline{367\,000 \text{ NOK}}}$.

g) Forklaringskraften R^2 kan leses direkte fra Excel-utskriften:
(Se cellen som heter "*R Square*" i Excel-utskriften): [7]

$$\underline{\underline{R^2 \;=\; 0.8721}} \tag{59}$$

h) Kommentar til svaret i oppgave **4g**:

At $R^2 = 0.8721$ betyr at regresjonslinjen vil i "i stor grad" forutsi kvadrameterprisen på en leilighet for en person med en gitt årslønn.
Modellen har $\underline{\underline{\text{stor forklaringskraft}}}$, tilsvarende $R^2 = 0.8721$.

■

[7] 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.