Norges teknisk–naturvitenskapelige universitet
Institutt for matematiske fag

ST0103 2015S – løsningsskisse

Oppgave 1

a) Forventningsverdien er $EX = \sum_{x=1}^{6} xP(X=x) = 1 \cdot 0{,}09 + 2 \cdot 0{,}38 + 3 \cdot 0{,}25 + 4 \cdot 0{,}20 + 5 \cdot 0{,}06 + 6 \cdot 0{,}02 = 2{,}82$.

Variansen er $\operatorname{Var} X = E(X-EX)^2 = \sum_{x=1}^{6} (x-2{,}82)^2 P(X=x) = (1-2{,}82)^2 \cdot 0{,}09 + (2-2{,}82)^2 \cdot 0{,}38 + (3-2{,}82)^2 \cdot 0{,}25 + (4-2{,}82)^2 \cdot 0{,}20 + (5-2{,}82)^2 \cdot 0{,}06 + (6-2{,}82)^2 \cdot 0{,}02 = 1{,}3276$, eventuelt $\operatorname{Var} X = EX^2 - (EX)^2 = \sum_{x=1}^{6} x^2 P(X=x) - (EX)^2 = 1^2 \cdot 0{,}09 + 2^2 \cdot 0{,}38 + 3^2 \cdot 0{,}25 + 4^2 \cdot 0{,}20 + 5^2 \cdot 0{,}06 + 6^2 \cdot 0{,}02 - 2{,}82^2 = 1{,}3276$, slik at standardavviket til X er $\sqrt{\operatorname{Var} X} = \sqrt{1{,}3276} = 1{,}15$.

b) Sannsynligheten for 4 eller flere egg er $P(X \geq 4) = P(X=4) + P(X=5) + P(X=6) = 0{,}20 + 0{,}06 + 0{,}02 = 0{,}28$.

Den betingede sannsynligheten for 6 egg gitt minst 4 egg er $P(X=6 \mid X \geq 4) = P(X=6 \cap X \geq 4)/P(X \geq 4) = P(X=6)/P(X \geq 4) = 0{,}02/0{,}28 = 0{,}071$.

c) Det er rimelig å anta at Y er binomisk fordelt med parametre $n = 50$ og $p = 0{,}28$. Forventningsverdien er $EY = np = 50 \cdot 0{,}28 = 14$ og variansen $\operatorname{Var} Y = np(1-p) = 50 \cdot 0{,}28 \cdot 0{,}72 = 10{,}08$, som gir standardavvik $\sqrt{\operatorname{Var} Y} = \sqrt{10{,}08} = 3{,}17$.

Siden $\operatorname{Var} Y \geq 5$, kan vi bruke normaltilnærming, som med heltallskorreksjon gir

$$
\begin{aligned}
P(Y \geq 10) = P(Y \geq 9{,}5) &= P\left(\frac{Y-14}{3{,}17} \geq \frac{9{,}5-14}{3{,}17}\right) \\
&\approx P(Z \geq -1{,}42) = 1 - P(Z < -1{,}42) = 1 - 0{,}0778 = 0{,}922,
\end{aligned}
$$

der Z er standard normalfordelt og den nest siste likheten følger fra oppslag i tabell. Hvis heltallskorreksjon ikke brukes, får vi i stedet

$$
\begin{aligned}
P(Y \geq 10) &= P\left(\frac{Y-14}{3{,}17} \geq \frac{10-14}{3{,}17}\right) \\
&\approx P(Z \geq -1{,}26) = 1 - P(Z < -1{,}26) = 1 - 0{,}1038 = 0{,}896,
\end{aligned}
$$

mens det eksakte svaret er $1 - \sum_{y=0}^{9} \binom{n}{y} p^y (1-p)^{n-y} = 1 - \sum_{y=0}^{9} \binom{50}{y} 0{,}28^y \cdot 0{,}72^{50-y} = 0{,}926$.

d) La A være hendelsen at er tilfeldig valgt reir tilhører underart A. Andelen av reirene som tilhører underart A er da $P(A)$, og $P(X=1 \mid A) = 0{,}05$ og $P(X=1 \mid \overline{A}) = 0{,}15$. Loven om total sannsynlighet gir at $0{,}09 = P(X=1) = P(X=1 \mid A)\, P(A) + P(X=1 \mid \overline{A})\, P(\overline{A}) = 0{,}05\, P(A) + 0{,}15\,(1-P(A))$. Dette er en likning i $P(A)$, som gir $(0{,}15 - 0{,}05)\, P(A) = 0{,}15 - 0{,}09$, og $P(A) = 0{,}6$.

Oppgave 2

a) Hvis $\mu = 7{,}45$, er testobservatoren $Z = (\bar{X} - 7{,}45)/(0{,}05/\sqrt{n})$ standard normalfordelt. En stor verdi av Z tyder på at det er den alternative hypotesen som er riktig, slik at nullhypotesen skal forkastes. Med signifikansnivå på 0,05, blir kritisk verdi 0,95-kvantilen til standard normalfordeling, som er 1,645 (fra tabell). Forkastningsområdet er altså fra 1,645 og oppover. Med de oppgitte tallene får vi verdien $(7{,}47 - 7{,}45)/(0{,}05/\sqrt{20}) = 1{,}79$. Vi forkaster altså nullhypotesen, og påstår at $\mu > 7{,}45$. (Hvis $\mu < 7{,}45$, som også dekkes av nullhypotesen, er $P(Z \geq 1{,}645) < 0{,}05$, slik at vi har sikret oss at sannsynligheten for feilaktig forkastning av nullhypotesen i alle tilfeller høyst er 0,05.)

Alternativt kan vi bruke p-verdimetoden, som gir sannsynligheten for at testobservatoren skal bli så stor som den ble, eller større, når $\mu = 7{,}45$. P-verdien blir $P(Z \geq 1{,}79) = 1 - P(Z < 1{,}79) = 1 - 0{,}9633 = 0{,}037 < 0{,}05$ (fra tabell), og vi får samme konklusjon som over. (Sannsynligheten blir bare mindre hvis $\mu < 7{,}45$, som også dekkes av nullhypotesen.)

b) Testobservatoren er fremdeles $Z = (\bar{X} - 7{,}45)/(0{,}05/\sqrt{n})$ hvis $\mu = 7{,}47$, og vi forkaster nullhypotesen hvis den får verdi 1,645 eller større. Men Z er nå ikke standard normalfordelt – det er derimot $(\bar{X} - 7{,}47)/(0{,}05/\sqrt{n})$. Når $n = 20$ er sannsynligheten for forkastning av nullhypotesen – teststyrken i $\mu = 7{,}47$ –

$$\begin{aligned}
P(Z \geq 1{,}645) &= P\left(\frac{\bar{X} - 7{,}45}{0{,}05/\sqrt{20}} \geq 1{,}645\right) = P\left(\frac{\bar{X} - 7{,}47}{0{,}05/\sqrt{20}} \geq 1{,}645 - \frac{0{,}02}{0{,}05/\sqrt{20}}\right) \\
&= P(U \geq -0{,}14) = 1 - P(U < -0{,}14) = 1 - 0{,}4443 = 0{,}56,
\end{aligned}$$

der U er standard normalfordelt og den nest siste likheten følger fra oppslag i tabell.

c) Som i forrige punkt, er sannsynligheten for forkastning

$$\begin{aligned}
P(Z \geq 1{,}645) &= P\left(\frac{\bar{X} - 7{,}45}{0{,}05/\sqrt{n}} \geq 1{,}645\right) = P\left(\frac{\bar{X} - 7{,}47}{0{,}05/\sqrt{n}} \geq 1{,}645 - \frac{0{,}02}{0{,}05/\sqrt{n}}\right) \\
&= P\left(U \geq 1{,}645 - \frac{0{,}02}{0{,}05/\sqrt{n}}\right),
\end{aligned}$$

der U er standard normalfordelt. Vi ønsker denne sannsynligheten større enn 0,8 – det vil si at $1{,}645 - 0{,}02/(0{,}05/\sqrt{n})$ må være mindre enn 0,2-kvantilen i standard normalfordeling. Tabell over kritiske verdier i standardnormalfordelingen viser at sannsynligheten for at en standard normalfordelt variabel er større enn 0,842, er 0,2, slik at 0,2-kvantilen er $-0{,}842$. Vi krever dermed $1{,}645 - 0{,}02/(0{,}05/\sqrt{n}) < -0{,}842$, som gir $0{,}02\sqrt{n}/0{,}05 > 1{,}645 + 0{,}842$, og $n > ((1{,}645 + 0{,}842) \cdot 0{,}05/0{,}02)^2 = 38{,}6$. Minst 39 prøver må altså tas for at sannsynligheten for at nullhypotesen forkastes skal være større enn 0,8 når $p\mathrm{H} = 7{,}47$.

Oppgave 3

a) Ved å ta logaritmen av $h = a|d|^b$, får vi $\ln h = \ln a + b\ln|d|$. Hvis vi setter $y = \ln h$, $\alpha = \ln a$, $\beta = b$ og $x = \ln|d|$, får vi $y = \alpha + \beta x$.

b) Estimatet av β er $\hat{\beta} = \sum_{i=1}^{20}(x_i - \bar{x})(y_i - \bar{y}) / \sum_{i=1}^{20}(x_i - \bar{x})^2 = 19{,}25/9{,}67 = 1{,}99$. Estimatet av α er $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 3{,}96 - 1{,}99 \cdot 6{,}12 = -8{,}22$.

c) Et 95 %-konfidensintervall for β har grenser $\hat{\beta} \pm t_{0{,}025}\sqrt{(\mathrm{SS_E}/(n-2))/\sum_{i=1}^{n}(x_i - \bar{x})^2}$, der $n = 20$ er antall observasjoner og $t_{0{,}025} = 2{,}101$ (fra tabell) er 0,975-kvantilen i en t-fordeling med $n-2 = 18$ frihetsgrader. Med våre tall får vi $1{,}99 \pm 2{,}101\sqrt{(0{,}156/18)/9{,}67} = 1{,}99 \pm 0{,}063$, som gir konfidensintervallet $[1{,}93, 2{,}05]$.

Dette konfidensintervallet bygger på forutsetningene som ligger til grunn for inferens i en lineær regresjonsmodell: At feilleddene $Y_i - \alpha - \beta x_i$ er uavhengige og normalfordelte, alle med samme varians og forventningsverdi 0, når vi ser på Y_i-ene som stokastiske variabler.

Siden konfidensintervallet for β (og dermed for b) er smalt og inneholder 2, er det rimelig å klassifisere dalen som en U-dal.