Løsningsskisse til eksamen i ISTx100y Statistikk: 29.10.2021

29 October, 2021

Fasit

1a) 0.01, 1b) 0.0347, 1c) 0.259

2a) 0.050, 2b) 0.323, 2c) 2 liter

3a) 0.1587, 3b) 0.0668, 3c) 0.5222

4a) $\mathrm{E}(X+2Y)=-4$, $\mathrm{Var}(X+2Y)=7$, 4b) $\mathrm{E}(2X-3V)=-3$, $\mathrm{Var}(2X-3V)=36$

5a) λ, 5b) $\frac{\lambda}{\sum_{i=1}^{n} v_i}$, c) Forventningsrett og lav varians

6a) $(91.02, 100.00)$, 6b) alle verdier i intervallet $[50.0, 60.0]$.

7a) $H_1 : \mu > 5$ gram, 7b) $\frac{\bar{X}-5}{1/\sqrt{10}}$, 7c) Testobservatoren er større enn kritisk verdi og vi forkaster nullhypotesen, 7d) 0.023, 7e) 0.936

Oppgave 1: Hendelser

Innledning:

En bedrift produserer solcellepaneler, og det gjøres en kvalitetssjekk av panelene før de sendes ut på markedet. For et tilfeldig valgt solcellepanel definerer vi følgende fire hendelser:

- A_1 : panelet er feilfritt
- A_2 : panelet har en liten feil
- A_3 : panelet har en alvorlig feil
- B : en inspektør har markert at kan være et problem med panelet

Her er hendelsene A_1, A_2, A_3 en oppdeling (partisjon) av utfallsrommet, dvs. at nøyaktig én av dem inntreffer (de er parvis disjunkte og unionen er utfallsrommet).

Erfaring viser at vi kan anta følgende betingede sannsynligheter:

$P(B \mid A_1) = 0.01, P(B \mid A_2) = 0.80, P(B \mid A_3) = 0.90.$
Videre er $P(A_1) = 0.97$ og $P(A_2) = 0.02$.

Spørsmål:

a) Hva er sannsynligheten for at et tilfeldig valgt solcellepanel har en alvorlig feil, $P(A_3)$?

$$P(A_3) = 1 - P(A_1) - P(A_2) = 1 - 0.97 - 0.02 = 0.01$$

b) Hva er sannsynligheten for at kvalitetssjekken har markert at det er et problem med et tilfeldig valgt solcellepanel, $P(B)$?

$$P(B) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + P(B \mid A_3)P(A_3) = 0.01 \cdot 0.97 + 0.80 \cdot 0.02 + 0.90 \cdot 0.01 = 0.0347$$

c) Hvor sannsynlig er det at solcellepanelet har en alvorlig feil, gitt at inspektøren markerte et problem, dvs. $P(A_3 \mid B)$? Oppgi svaret med 3 desimaler, f.eks. 0.123 eller 0.987.

$$\frac{P(B \mid A_3)P(A_3)}{P(B)} = \frac{0.90 \cdot 0.01}{0.0347} = \frac{0.009}{0.0347} = 0.259$$

Kommentar: vi kunne bare spurt om c) direkte, men har valgt å legge inn enkle spørsmål for a) og b) med flervalg, slik at det kan være lettere å få til c).

Oppgave 2: Poissonfordelt stokastisk variabel

Innledning:

Vi antar at antall tarmbakterier, X, i v liter vann fra en drikkevannskilde er poissonfordelt med punktsannsynlighet

$$P(X = x) = \frac{(\lambda v)^x}{x!} e^{-\lambda v}, x = 0, 1, 2, \ldots$$

Vi kaller λ for forurensningsgraden av vannet. I denne oppgaven antar vi at $\lambda = 3$.

Spørsmål:

a) Hva er sannsynligheten for at en prøve på $v = 1$ liter drikkevann ikke inneholder noen tarmbakterier? Skriv inn svaret med 3 desimaler, for eksempel 0.123 eller 0.987.

Svar Nå er $\lambda v = 3 \cdot 1 = 3$.

$$P(X = 0) = e^{-3} = 0.0498$$

b) Gitt at man vet at en prøve på $v = 0.5$ liter drikkevann inneholder minst en tarmbakterie, hva er sannsynligheten for at prøven inneholder akkurat 2 tarmbakterier? Skriv inn svaret med 3 desimaler, for eksempel 0.123 eller 0.987.

Svar Nå er $\lambda v = 3 \cdot 0.5 = 1.5$.

$$P(X = 2 \mid X > 0) = \frac{P(X = 2 \cap X > 0)}{1 - P(X \leq 0)} = \frac{P(X = 2)}{1 - P(X = 0)} = \frac{\frac{1.5^2 e^{-1.5}}{2!}}{1 - e^{-1.5}} = 0.2510/0.7769 = 0.323$$

c) Hvor mange liter vann (v) må en prøve minst inneholde for at denne prøven med sannsynlighet minst 0.9975 skal inneholde en eller flere tarmbakterier?

$$\begin{aligned}
P(X > 0) = 1 - P(X = 0) &\geq 0.9975 \\
1 - e^{-3v} &\geq 0.9975 \\
e^{-3v} &\leq 1 - 0.9975 = 0.0025 \\
-3v &\leq \ln(0.0025) = -5.99 \\
3v &\geq 5.99 \\
v &\geq 2
\end{aligned}$$

Kommentar: Oppgaven består av tre delspørsmål som alle kan besvares separat.

Oppgave 3: Normalfordelte kulediametere

Innledning:

En bedrift produserer kuler som skal brukes i kulelager. Kulelager er en viktig del av mange typer verktøy og husholdningsapparater.

Diameteren til kulene antas å være normalfordelt med forventningsverdi 20 mm og standardavvik 0.1 mm. Diameteren til en kule er uavhengig av diameteren av de andre kulene.

Spørsmål:

a) Hva er sannsynligheten for at en tilfeldig valgt kule har diameter mindre enn 19.9 mm? Skriv inn svaret med 4 desimaler, for eksempel 0.1234 eller 0.9876.

Svar:

$$P(D < 19.9) = P(\frac{D - 20}{0.1} < \frac{19.9 - 20}{0.1}) = \Phi(-1) = 0.1587$$

b) Hva er sannsynligheten for at gjennomsnittlig diameter til 9 tilfeldig valgte kuler er større enn 20.05 mm? Skriv inn svaret med 4 desimaler, for eksempel 0.1234 eller 0.9876.

Svar:

$Y = \frac{D_1+D_2+\cdots+D_9}{9} \sim N(20, \frac{0.1}{3})$,

$$P(Y > 20.05) = P(Z > \frac{20.05 - 20}{\frac{0.1}{3}}) = 1 - \Phi(1.5) = 1 - 0.9332 = 0.0668$$

c) Hva er sannsynligheten for at forskjellen i diameter mellom to tilfeldig valgte kuler er mindre enn 0.1 mm? Skriv inn svaret med 4 desimaler, for eksempel 0.1234 eller 0.9876.

Svar:

Her er fordelingen til $Y = D_1 - D_2 \sim N(0, \sqrt{0.1^2 + 0.1^2} = \sqrt{2} \cdot 0.1)$

$$P(|Y| < 0.1) = P(-0.1 < Y < 0.1) = P(Y < 0.1) - P(Y < -0.1) = P(Z < \frac{0.1 - 0}{\sqrt{2} \cdot 0.1}) - P(Z < \frac{-0.1 - 0}{\sqrt{2} \cdot 0.1} =$$

$$\Phi(\frac{1}{\sqrt{2}}) - \Phi(-\frac{1}{\sqrt{2}}) = \Phi(0.71) - \Phi(-0.71) = 0.7611 - 0.2389 = 0.5222$$

Kommentar: Oppgaven består av tre delspørsmål som alle kan besvares separat.

Oppgave 4: Forventingsverdi og varians

Innledning: La X og Y være to uavhengige stokastiske variabler med $\mathrm{E}(X) = 0$, $\mathrm{E}(Y) = -2$, $\mathrm{Var}(X) = 3$, $\mathrm{Var}(Y) = 1$.

Spørsmål:

a) Finn forventningsverdi og varians for $X + 2Y$. Angi svarene som heltall.

Svar:

$$\mathrm{E}(X + 2Y) = \mathrm{E}(X) + 2\mathrm{E}(Y) = 0 + 2 \cdot (-2) = -4$$
$$\mathrm{Var}(X + 2Y) = \mathrm{Var}(X) + 2^2\mathrm{Var}(Y) = 3 + 4 \cdot 1 = 7$$

Innledning: La X ha forventningsverdi og varians som oppgitt over. La V være en stokastisk variabel med $\mathrm{E}(V) = 1$ og $\mathrm{Var}(V) = 4$. Videre er $\mathrm{Cov}(X, V) = 1$.

Spørsmål:

b) Finn forventningsverdi og varians for $2X - 3V$. Angi svarene som heltall.

Svar:

$$\mathrm{E}(2X - 3V) = 2\mathrm{E}(X) - 3\mathrm{E}(V) = 2 \cdot 0 - 3 \cdot 1 = -3$$

$$\mathrm{Var}(2X-3Y) = 2^2\mathrm{Var}(X)+(-3)^2\mathrm{Var}(V)+2\cdot2\cdot(-3)\cdot\mathrm{Cov}(X,V) = 4\cdot3+9\cdot4+2\cdot2\cdot(-3)\cdot1 = 12+36-12 = 36$$

Kommentar: Oppgaven består av fire delspørsmål som alle kan besvares separat.

Oppgave 5: Punktestimering

Innledning:

Som i oppgave 2 antar vi at antall tarmbakterier, X, i v liter vann fra en drikkevannskilde er poissonfordelt med punktsannsynlighet

$$P(X = x) = \frac{(\lambda v)^x}{x!} e^{-\lambda v}, x = 0, 1, 2, \ldots$$

Vi kaller λ for forurensningsgraden av vannet, og i denne oppgaven er λ ukjent.

Vi vil samle inn data fra n uavhengige prøver, der vi antar at forurensningsgraden er den samme i alle prøvene. For hver prøve bruker vi $v_1, v_2, \ldots, v_n$ for antall liter av drikkevannet, og $X_1, X_2, \ldots, X_n$ for de tilhørende antallene tarmbakterier.

Vi skal se på en estimator for forurensningsgraden λ:

$\widehat{\lambda} = \frac{X_1+X_2+\cdots+X_n}{v_1+v_2+\cdots+v_n} = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n v_i}$

Spørsmål:

a) Hva er forventningsverdien til estimatoren $\widehat{\lambda}$?

Svar:

$$\mathrm{E}(\widehat{\lambda}) = \mathrm{E}(\frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n v_i}) = \frac{1}{\sum_{i=1}^n v_i}\mathrm{E}(\sum_{i=1}^n X_i) = \frac{1}{\sum_{i=1}^n v_i}\sum_{i=1}^n \mathrm{E}(X_i) = \frac{1}{\sum_{i=1}^n v_i}\sum_{i=1}^n \lambda v_i = \lambda$$

b) Hva er variansen til estimatoren $\widehat{\lambda}$?

Svar:

$$\mathrm{Var}(\widehat{\lambda}) = (\frac{1}{\sum_{i=1}^n v_i})^2 \sum_{i=1}^n \mathrm{Var}(X_i) = (\frac{1}{\sum_{i=1}^n v_i})^2 \sum_{i=1}^n \lambda v_i = \frac{\lambda}{\sum_{i=1}^n v_i}$$

c) Hvis du skal vurdere hvor god en estimator er, hvilke egenskaper bør du da se etter?

- Forventingsretthet og lav varians
- Forventningsretthet og høy varians
- Lav forventningsverdi og lav varians
- Lav forventningsverdi og høy varians
- Høy forventningsverdi og lav varians

- Høy forventningsverdi og høy varians

Svar: Forventingsretthet og lav varians

Kommentar: Oppgaven består av tre delspørsmål som alle kan besvares separat.

Oppgave 6: Konfidensintervall

Innledning:

Anta at X er normalfordelt med ukjent forventningsverdi μ og ukjent standardavvik σ. Vi tar et tilfelig utvalg på $n = 10$ observasjoner fra en populasjon, og observerer
$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 95.51$ og $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 354.56$.

Spørsmål:

a) Hva blir nedre og øvre grenser i et 95% konfidensintervall for μ?

Svar: Bruker formelen på formelarket.

Dersom standardavviket σ er *ukjent*, så er

$$\left[\bar{X} - t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}},\ \ \bar{X} + t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}}\right]$$

et $100(1-\alpha)\%$ konfidensintervall for μ.

Vi har fått beskjed om 95% og $n - 1 = 9$, dermed slår vi opp $t_{0.025,9} = 2.262$. Deretter trenger vi å regne ut $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{354.56/9} = 6.2766$.

Så setter vi inn i formelen:

$$d = t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}} = 2.262 \cdot 6.2766/\sqrt{10} = 4.4877$$

Nedre grense: $\bar{x} - d = 95.51 - 4.4877 = 91.02$

Øvre grense: $\bar{x} + d = 95.51 + 4.4877 = 100.00$

Innledning Anta at du har et 95% konfidensintervall for μ med nedre grense 50.0 og øvre grense 60.0. Du vil utføre en hypotesetest med nullhypotese $H_0 : \mu = \mu_0$ og alternativ hypotese $H_1 : \mu \neq \mu_0$.

b) Angi en mulig verdi for μ_0, som ville gjort at du *ikke forkastet* nullhypotesen på signifikansnivå 5%? Skriv svaret med 1 desimal, f.eks 1.2 eller 98.7.

Svar:

For alle verdier av μ_0 utenfor det 95% konfidensintervallet ville man ha forkastet nullhypotesen i en tosidig test med signifikansnivå 5%. Her vil alle verdier av μ_0 innenfor intervallet ikke gi forkastning. I læreboka kan du lese om dette på i 6.4.9 på side 269.

Kommentar: Oppgaven består av to delspørsmål som begge kan besvares separat.

Oppgave 7: Hypotesetest

Innledning:

Steinprøver som er tatt fra et geologisk felt blir analysert kjemisk for å bestemme innholdet av et spesielt grunnstoff. La X være det målte innholdet av grunnstoffet i en prøve på 100 g. Vi antar at X er normalfordelt med forventningsverdi μ og kjent standardavvik $\sigma = 1$ g. Hvis forventningsverdien μ er større enn 5 gram, regnes feltet som drivverdig og man vil ønske å starte utvinning av grunnstoffet. Disse opplysningene gjelder i hele oppgaven.

Spørsmål:

a) Et gruveselskap ønsker å undersøke om feltet er drivverdig. Sett opp en alternativ hypotese for å undersøke om feltet er drivverdig.

Svar: $H_1 : \mu > 5$.

Vi spør ikke etter nullhypotesen, men den er $H_0 : \mu = 5$ eller $H_0 : \mu \leq 5$. Løvås bruker den siste og på temavideoene brukes i hovedsak den første.

Innledning:

For å utføre hypotesetesten vil vi samle inn data på innholdet av grunnstoffet i $n = 10$ uavhengige og tilfeldig valgte steinprøver $X_1, X_2, \ldots, X_{10}$. Vi får verdiene $x_1, x_2, \ldots, x_{10}$ og finner at $\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 5.74$. Vi skal bruke signifikansnivå $\alpha = 0.05$.

Spørsmål:

b) Hvilket av følgende uttrykk er en testobservator for hypotesetesten?

Svar

$$\frac{\bar{X} - 5}{1/\sqrt{10}}$$

c) Hva blir konklusjonen for hypotestetesten?

- Testobservatoren er større enn kritisk verdi og vi forkaster nullhypotesen.
- Testobservatoren er større enn kritisk verdi og vi forkaster ikke nullhypotesen.
- Testobservatoren er mindre enn kritisk verdi og vi forkaster nullhypotesen.
- Testobservatoren er mindre enn kritisk verdi og vi forkaster ikke nullhypotesen.

Svar: Vi setter inn i formelen for testobservatoren og finner $z = \frac{5.74-5}{1/\sqrt{10}} = 2.34$.

Kritisk verdi for høyresidig test for kjent σ er $z_{0.05} = 1.645$. Testobservatoren er større enn kritisk verdi og vi forkaster nullhypotesen.

d) Anta at verdien til testobservatoren i punkt b ble 2.0. Regn ut p-verdien for hypotesetesten for denne verdien av testobservatoren. Oppgi svaret med 3 desimaler, f.eks. 0.123 eller 0.987.

Her oppgir vi 2.0 så ikke det blir følgefeil fra oppgave c), men man må vite hva alternativ hypotese og testobservatoren er.

P-verdien blir $P(Z > 2) = 1 - P(Z \leq 2)$, og da slår vi opp i normaltabellen og finner 0.9772, og vi skal forkaste for store verdier så vi får 1-0.9772=0.0228

(Testobservatoren ble 2.34 som slås opp i tabellen, og gir 0.9904, som gir p-verdi 0.0096, men det spurte vi ikke om.)

e) Regn ut teststyrken til testen når forventningsverdien til innholdet av grunstoffet er $\mu = 6$ gram. Fremdeles er $\sigma = 1$ gram, utvalgsstørresen $n = 10$ og signifikansnivået $\alpha = 0.05$. Det oppgis at den numeriske verdien til testobservatoren ikke inngår i denne utregningen. Oppgi svaret med 3 desimaler, f.eks. 0.123 eller 0.987.

Svar:

Ad nest siste setning: Det er for å unngå spørsmål rundt om det er 2.34 eller 2.0 som skal brukes for testobservatoren - selv om dette tallet ikke inngår i beregning av styrke.

Det første vi gjør er å skrive opp forkastningsregelen til testen og flytte litt om slik at vi får et uttrykk der $\bar{X}$ står alene på venstre side:

$$
\begin{aligned}
Z &\geq 1.645 \\
\frac{\bar{X}-5}{1/\sqrt{10}} &\geq 1.645 \\
\bar{X} &\geq 5 + 1.645 \cdot \frac{1}{\sqrt{10}} \\
\bar{X} &\geq 5.52
\end{aligned}
$$

Deretter skal vi regne ut teststyrken når $\mu = 6$.

P(forkaste H_0 når alternativ hypotese er sann og $\mu = 6) = P(\bar{X} \geq 5.52$ når $\mu = 6) = P(\frac{\bar{X}-6}{1/\sqrt{10}} \geq \frac{5.52-6}{1/\sqrt{10}}) = 1 - P(\frac{\bar{X}-6}{1/\sqrt{10}} \leq -1.52) = 1 - \Phi(-1.52) = 1 - 0.0643 = 0.9357$

Alternativ utregning:

$$
\begin{aligned}
P(\text{forkaste } H_0 \text{ gitt at } \mu = 6) = P(Z > 1.645 \mid \mu = 6) = P((\bar{X}-5)\sqrt{10} > 1.645 \mid \mu = 6) \\
= P((\bar{X}-5) > 1.645/\sqrt{10}) = P((\bar{X}-6) > 1.645/\sqrt{10} - 1) = P((\bar{X}-6)\sqrt{10} > 1.645 - 1\sqrt{10}) \\
= P(Z > -1.52) = 1 - \Phi(-1.52) = 1 - 0.0643 = 0.9357
\end{aligned}
$$

Kommentar: Oppgaven består av fem delspørsmål der oppgave a) regnes som lett og må besvares korrekt for kunne svare på cde). Delspørsmål b) kan besvares uten at a) er besvart. Delspørsmål c) bygger på a og b), delspørsmål d) og e) kan i prinsippet besvares uten at delspørsmål c) er korrekt.