

Løsningsskisse til prøveeksamen i ISTx100y høsten 2021

Oppgave 1: Hendelser

Innledning:

La oss anta at en tilfeldig valgt student som jobber med et spørsmål på en eksamen har en sannsynlighet på 0.75 for å kunne teorien og klare utregningene, som gir korrekt svar på spørsmålet (dvs. studenten trenger ikke å tippe).

Hvis studenten ikke finner frem til korrekt svar ved teori og utregning, så har studenten mulighet til å tippe. Hvis studenten tipper, så antar vi at sannsynligheten for at studenten velger det riktige svaret er 0.2.

Spørsmål:

a) Hva er sannsynligheten for at en tilfeldig valgt student svarer rett på eksamensspørsmålet?

- 0.80
- 0.75
- 0.20
- 0.05
- 0.50

b) Gitt at studenten svarer korrekt på dette spørsmålet, hva er sannsynligheten for at studenten *ikke tippet*? Oppgi svaret med 4 desimaler, f.eks. 0.1234 eller 0.9876.

Fasit

Det stokastiske eksperimentet er at vi tilfeldig trekker en student og en oppgave. Vi definerer to hendelser.

A: studenten trenger ikke tippe (kan teorien og klarer utregningene som gir korrekt svar)

B: studenten svarer rett

Oppgitt er $P(A) = 0.75$ og $P(B | A) = 1$, $P(A^C) = 1 - 0.75 = 0.25$ og $P(B | A^C) = 0.2$

a) Først regner vi ut $P(B)$.

$$P(B) = P(B | A)P(A) + P(B | A^C)P(A^C) = 1 \cdot 0.75 + 0.2 \cdot 0.25 = 0.80$$

b) Vi spør etter $P(A | B)$:

$$P(A | B) = \frac{P(B|A)P(A)}{P(B)} = \frac{1 \cdot 0.75}{0.80} = 0.9375$$

Grunnen til å be om 4 desimaler er at det blir akkurat 4 desimaler når vi deler de to tallene på hverandre.

Oppgave 2: Geometrisk fordelt stokastisk variabel

Innledning: La X være geometrisk fordelt med parameter $p = 0.1$. Dvs at X har punktsannsynlighet $P(X = x) = p(1 - p)^{x-1}$ for $x = 1, 2, \dots$

Spørsmål: Fyll inn riktige verdier for følgende tre sannsynligheter. Oppgi svarene med 2 desimaler, for eksempel 0.12 eller 0.98.

- a) $P(X = 5)$
 b) $P(X \geq 20)$
 c) $P(x < 5 \mid X \geq 3)$

Fasit

a) $P(X = 5) = 0.1 \cdot (1 - 0.1)^{(5-1)} = 0.1 \cdot 0.9^4 = 0.0656$ svar 0.07.

b) Bruk formel fra formelark for $F(x) = 1 - (1 - p)^x$. $P(X \geq 20) = 1 - P(X < 20) = 1 - P(X \leq 19) = (1 - p)^{19} = 0.9^{19} = 0.1351$ avrundet til 0.14.

“Vanlig feil”: regne ut $1 - P(Z \leq 20)$, som vil gi $(1 - p)^{20} = 0.9^{20} = 0.1216$ som ikke er regnet som rett svar.

c)

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) = 1 - P(X \leq 2) = 1 - P(X = 1) - P(X = 2) \\ &= 1 - 0.1 \cdot 0.9^0 - 0.1 \cdot 0.9^1 = 1 - 0.1 - 0.1 \cdot 0.9 = 0.81 \end{aligned}$$

eller

$$P(X \geq 3) = 1 - P(X < 3) = 1 - P(X \leq 2) = (1 - p)^2 = 0.9^2 = 0.81$$

$$P(X < 5 \mid X \geq 3) = \frac{P(X < 5 \cap X \geq 3)}{P(X \geq 3)} = \frac{P(3 \leq X < 5)}{P(X \geq 3)} = \frac{P(X=3) + P(X=4)}{P(X \geq 3)} = \frac{0.1 \cdot 0.9^2 + 0.1 \cdot 0.9^3}{0.81} = 0.19$$

Oppgave 3: Normalfordelte rør

Innledning: En bedrift produserer rør som settes sammen til gassrørledninger. Vi antar at lengden til et rør er normalfordelt med forventningsverdi 12 meter og standardavvik 0.1 meter. Vi antar videre at lengden til rørene som produseres er uavhengige av hverandre.

Spørsmål:

- a) Hva er sannsynligheten for at et tilfeldig valgt rør er lengre enn 12.2 meter? Skriv inn svaret med 4 desimaler, for eksempel 0.1234 eller 0.9876.
- b) Dersom standardavviket til lengdene på rørene kunne justeres til en annen verdi, hvilken verdi måtte det justeres til for at sannsynligheten skal være 0.95 for at lengden på et tilfeldig valgt rør er mellom 11.90 og 12.10 meter?

- 0.01
- 0.02
- 0.05
- 0.10
- 0.12
- 0.15

Innledning:

Rørene som brukes er fremdeles normalfordelte med forventningsverdi 12 meter og standardavvik 0.1 meter. Bedriften har fått i oppdrag å produsere en 10800 meter lang rørledning, og planlegger å produsere 900 rør til denne rørledningen. Totallengden på rørledningen blir da summen av lengdene til de 900 rørene.

Spørsmål:

- c) Hva er forventningsverdi og varians til totallengden av rørledningen? Skriv inn svarene som heltall, for eksempel 1 eller 100.

Forventningsverdi:

Varians:

- d) Hvis rørledningen er kortere enn 10795 meter eller lengre enn 10805 meter vil kunden klage på leveransen fra bedriften. Hvor sannsynlig er det at kunden klager? Skriv inn svaret med 3 desimaler, for eksempel 0.123 eller 0.987.

Fasit

a) $P(X > 12.2) = 1 - P(\frac{X-12}{0.1} \leq \frac{12.2-12.0}{0.1}) = 1 - P(Z \leq 2.00) = 1 - 0.9772 = 0.0228$

b) De som vet at det er 95% sannsynlighet for at en observasjon ligger innen for $\mu \pm 1.96\sigma$, og ser at intervallet som er oppgitt et symmetrisk om forventningsverdien 12 meter, kan resonnerer seg frem til at øvre grense 12.10 da vil være $\mu + 1.96\sigma$ og siden $\mu = 12$ så vil $\sigma = (12.1 - 12)/1.96 = 0.051$ avrundt til 0.05. Vi får samme svar hvis vi bruke den tilnærmede huskereglene med pluss minus to standardavvik.

Det er viktig er at intervallet var sentrert i forventningsverdien. Metoden som ville virket for andre tall enn akkurat 0.95 er:

$$\begin{aligned} P(11.90 < X < 12.10) &= 1 - P(X \geq 12.10) - P(X \leq 11.90) \\ &= 1 - P(Z \geq \frac{12.10 - 12}{\sigma}) - P(Z \leq \frac{11.90 - 12}{\sigma}) \\ &= 1 - P(Z \geq \frac{0.10}{\sigma}) - P(Z \leq -\frac{0.10}{\sigma}) \end{aligned}$$

Videre vet vi at $P(Z \geq \frac{0.10}{\sigma})$ må være lik $P(Z \leq -\frac{0.10}{\sigma})$

$$\begin{aligned} P(11.90 < X < 12.10) &= 1 - 2P(Z \geq \frac{0.10}{\sigma}) = 0.95 \\ P(Z \geq \frac{0.10}{\sigma}) &= \frac{1 - 0.95}{2} = 0.025 \end{aligned}$$

Slår vi opp i tabellen for kritiske verdier i standard normalfordelingen - dvs areal til høyre er 0.025 finner vi at den tilhørende kritiske verdien er $1.96 \frac{0.1}{\sigma} = 1.96$ gir $\sigma = \frac{0.1}{1.96} = 0.051$.

c) La T være en stokastisk variabel som er totallengden til 900 rør. $T = X_1 + X_2 + \dots + X_{900}$

Merk: det er IKKE $900X$ som er at man har 900 rør som er identisk like lange, det har vi ikke.

$$\begin{aligned} E(T) &= E(X_1 + X_2 + \dots + X_{900}) \\ &= E(X_1) + \dots + E(X_{900}) = 12 + 12 + \dots + 12 \\ &= 900 \cdot 12 = 10800 \end{aligned}$$

.

$$\begin{aligned} \text{Var}(T) &= \text{Var}(X_1 + X_2 + \dots + X_{900}) = \text{Var}(X_1) + \dots + \text{Var}(X_{900}) \\ &= 0.1^2 + 0.1^2 + \dots + 0.1^2 = 900 \cdot 0.1^2 = 9 \end{aligned}$$

d)

$$\begin{aligned} P(T < 10795 \cup T > 10805) &= 1 - P(10795 \leq T \leq 10805) \\ &= 1 - [P(T \leq 10805) - P(T \leq 10795)] = \\ 1 - [P(\frac{T - E(T)}{\text{SD}(T)} \leq \frac{10805 - 10800}{\sqrt{9}}) - P(\frac{T - E(T)}{\text{SD}(T)} \leq \frac{10795 - 10800}{\sqrt{9}})] \\ &= 1 - (P(Z \leq 1.67) - P(Z \leq -1.67)) = 1 - [0.9525 - 0.0475] = 0.095 \end{aligned}$$

Oppgave 4: Forventning og varians

Innledning: La X_1 og X_2 være to stokastiske variabler med $E(X_1) = 0$, $E(X_2) = -1$, $\text{Var}(X_1) = 3$, $\text{Var}(X_2) = 2$ og $\text{Cov}(X_1, X_2) = 1$. Anta videre at vi har en stokastisk variabel Y som er uavhengig av X_1 og X_2 og som har $E(Y) = 3$ og $\text{Var}(Y) = 2$.

La stokastiske variabler Z_1 og Z_2 være gitt som

$$Z_1 = 3X_1 + Y \quad \text{og} \quad Z_2 = X_1 - 4X_2.$$

Spørsmål: Finn forventningsverdi og varians for Z_1 og for Z_2 . Angi tallene som heltall, f.eks. 1 eller 100.

Fasit:

$$E(Z_1) = 3$$

$$\text{Var}(Z_1) = 29$$

$$E(Z_2) = 4$$

$$\text{Var}(Z_2) = 27$$

Løsning:

$$E(Z_1) = E(3X_1 + Y) = E(3X_1) + E(Y) = 3 \cdot 0 + 3 = 3$$

$$\text{Var}(Z_1) = \text{Var}(3X_1 + Y) = 3^2 \text{Var}(X_1) + \text{Var}(Y) + 2 \cdot 3 \cdot 1 \cdot \text{Cov}(X_1, Y) = 9 \cdot 3 + 2 + 0 = 29$$

$$E(Z_2) = E(X_1 - 4X_2) = E(X_1) - 4E(X_2) = 0 - 4 \cdot (-1) = 4$$

$$\begin{aligned} \text{Var}(Z_2) &= \text{Var}(X_1 - 4X_2) = \text{Var}(X_1) + (-4)^2 \text{Var}(X_2) + 2 \cdot 1 \cdot (-4) \text{Cov}(X_1, X_2) \\ &= 3 + 16 \cdot 2 - 8 \cdot 1 = 27 \end{aligned}$$

Oppgave 5: Konfidensintervall

Innledning: Vi ser på en lignende problemstilling som i oppgave 3, der en bedrift produserer rør som settes sammen til gassrørledninger. Vi antar også her at lengden til et rør er normalfordelt, men nå antar vi at forventningsverdien er ukjent og at standardavvik er kjent og lik $\sigma = 0.1$ meter. Vi antar fremdeles at lengden til rørene som produseres er uavhengige av hverandre.

Bedriften produserer $n = 12$ rør og måler lengden av rørene. Gjennomsnittlig lengde ble 11.97 meter.

Spørsmål:

a) Regn ut et 95% konfidensintervall for μ og oppgi nedre og øvre grense med to desimaler, f.eks, 1.23 eller 98.34.

Nedre grense:

Øvre grense:

b) Hvor mange målinger må man i denne situasjonen minst gjøre dersom man ønsker et 95% konfidensintervall for μ med lengde maksimalt 0.02 meter?

- 20
- 192
- 193
- 271
- 385

Fasit

a)

Vi har normalfordelte data med kjent standardavvik σ . Fra formelarket ser vi at et $100(1 - \alpha)\%$ konfidensintervall for μ er gitt ved

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

Vi skal ha et 95% konfidensintervall, og da må $\alpha = 0.05$ - fordi da blir $100 \cdot (1 - \alpha) = 100 \cdot 0.95 = 95$.

Da må vi finne kritiske verdier i normalfordelingen fra tabellen “Kritiske verdier i standard normalfordelingen” for $\alpha/2$, som da blir lik 0.025. Slår vi opp finner vi at $z_{0.025} = 1.96$. Da har vi alle størrelser vi trenger og setter inn i formelen for å finne øvre og nedre grense.

$$\text{Først regner vi ut det vi skal legge til og trekke fra (så det blir halve lengden av intervallet)} \quad z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{0.1}{\sqrt{12}} = 0.0566$$

Dermed blir nedre grense:

$$\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 11.97 - 0.0566 = 11.91$$

$$\text{Øvre grense: } \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 11.97 + 0.0566 = 12.03$$

b)

Lengden til intervallet er $2(z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$ og vi vet når vi har 12 observasjoner blir denne lengden $2 \cdot 0.0566 = 0.1132$. Det betyr at vi trenger endel flere observasjoner når vi skal ha et intervall som bare har lengde 0.02. Her kan man selvfølgelig prøve seg frem men meningen er at man skal sette opp en ligning og løse den:

$$L = 2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \text{ og vi skal ha at } L \leq 0.02.$$

$$2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq 0.02$$

$$\frac{2}{0.02} \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq 1$$

$$\frac{2}{0.02} \cdot z_{\alpha/2} \cdot \sigma \leq \sqrt{n}$$

$$\left(\frac{2}{0.02} \cdot z_{\alpha/2} \cdot \sigma \right)^2 \leq n$$

$$\left(\frac{2}{0.02} \cdot 1.96 \cdot 0.1 \right)^2 \leq n$$

$$19.6^2 \leq n$$

$$384.16 \leq n$$

Dermed må n være minst 385 for at konfidensintervallet får ønsket lengde.

Oppgave 6: Hypotesetest for fosfor i utslippsvann

Innledning:

I denne oppgaven skal vi se på utslipp av fosfor fra et kommunalt kloakkrenseanlegg. Renseanlegget har tatt i bruk nytt utstyr og fått oppgitt at mengdene fosfor i like store prøver tatt på ulike dager vil være uavhengige og normalfordelte med forventningsverdi μ mg/l som ikke overskrider 0.15 mg/l, og standardavvik $\sigma = 0.02$ mg/l.

Spørsmål:

a) Vi vil bruke målinger av fosforinnhold i utslippsvannet til å utføre en hypotesetest der vi tester om vi har grunnlag for å konkludere at at gjennomsnittlig mengde fosfor i utslippsvannet i det lange løp kan være større enn 0.15 mg/l, slik at tiltak må iverksettes.

Hvilken alternativ hypotese setter vi opp?

- $\mu \neq 0.15$
- $\mu = 0.15$
- $\mu > 0.15$
- $\mu < 0.15$

Innledning:

b) Vi har samlet inn data og får oppgitt at gjennomsnittet av forforinnholdet i 12 uavhengige prøver av utslippsvannet ble 0.158 mg/l. Bruk signifikansnivå 5% og utfør hypotesetesten. Hva bli konklusjonen?

Velg ett alternativ

- 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.

c) Regn ut p-verdien til testen. Oppgi svaret med 4 desimaler, f.eks. 0.1234 eller 0.9876.

d) Ledelsen ved renseanlegget ønsker å studere testen fra b) nærmere.

Vi antar fremdeles at $\sigma = 0.02$ mg/l og at signifikansnivå 5% benyttes, og at testen utføres basert på 12 målinger. Regn ut styrken til testen for alternativet $\mu = 0.17$ mg/l. Oppgi svaret med 4 desimaler, f.eks. 0.1234 eller 0.9876.

Fasit:

a)

Alternativhypotesen er alltid spørsmålet vi stiller, og derfor er $H_1 : \mu > 0.15$. Vi har ikke spurt om nullhypotesen, men nullhypotesen vil være $\mu = 0.15$ eller om man vil $\mu \leq 0.15$. De to variantene av nullhypotesen vil føre til samme forkastningsområde for hypotesetesten. Hvis vi forkaster nullhypotesen må vi ta en alvorsprat med produsenten av det nye utstyret.

b) Som testobservator for hypotestesten kan vi bruke Z -testobservatoren fra formelarket - fordi vi har at $X_1, X_2, \dots, X_{12}$ er uavhengige og normalfordelte med kjent standardavvik $\sigma = 0.02$ og når nullhypotesen er sann er det også kjent at forventningsverdien er $\mu_0 = 0.15$. Dermed vil $\bar{X} = \frac{1}{12} \sum_{i=1}^n X_i$ være normalfordelt med samme forventningsverdi men med varians $\sigma/\sqrt{12}$. Testobservatoren vår blir $Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

og den kan vi regne ut for våre data til å bli $z = \frac{0.158 - 0.15}{\frac{0.02}{\sqrt{12}}} = 1.3856$. Vi har en høyresidig test og vil forkaste hvis testobservatoren er veldig stor og da større enn og kritisk verdi $z_\alpha = z_{0.05} = 1.645$. Vi skal sammenligne 1.3856 med 1.645, og vår testobservator er mindre enn kritisk verdi.

Riktig svar er da: Testobservatoren er mindre enn kritisk verdi og vi forkaster ikke nullhypotesen.

c) P -verdien regner vi ut direkte fra den observerte testobservatoren $z = 1.3856$, som sannsynligheten for å observere en testobservator som er større eller lik denne verdien når nullhypotesen er sann. $P(Z > 1.3856) = 1 - P(Z \leq 1.39)$ slår opp i tabellen for 1.39 og finner at det er 0.9177, slik at p -verdien blir $1 - 0.9177 = 0.0823$.

d) (Dette skal være det aller vanskeligste punktet på eksamen.)

Teststyrken gir oss sannsynligheten for å forkaste H_0 for ulike verdier av μ , når nullhypotesen er usann.

Teststyrken regner vi $P(\text{forkaste } H_0 \mid \text{sann verdi er } \mu = 0.17)$.

Før vi starter med teststyrken er det enklest å skrive ut forkastningsgrensen uttrykt ved $\bar{X}$:

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geq z_\alpha$$
$$\bar{X} \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

og deretter sette inn alle kjente numeriske verdier. Her er $\mu_0 = 0.15$, $\alpha_{0.05} = 1.645$, $\sigma/\sqrt{n} = 0.02/\sqrt{12} = 0.00577$.

Da får vi

$$\bar{X} \geq 0.15 + 1.645 \cdot 0.00577 = 0.1595$$

Observer at ikke vi har brukt verdien vi har observert for $\bar{x}$, eller for testobservatoren - den inngår ikke i teststyrken heller.

Nå er oppdraget å regne ut teststyrke for denne forkastningsregelen.

$$P(\text{forkaste } H_0 \mid \text{sann verdi er } \mu = 0.17)$$

$$P(\bar{X} \geq 0.1595 \mid \text{sann verdi er } \mu = 0.17)$$

Dette med at “gitt at sann verdi er $\mu = 0.17$ ” betyr at nå er fordelingen til $\bar{X}$ lik

$$\bar{X} \sim N(0.17, 0.00577)$$

og det er disse to verdiene vi må bruke når vi standardiserer for å regne ut sannsynlighet ved hjelp av standard normalfordelingen.

$$\begin{aligned} &P(\bar{X} \geq 0.1595 \mid \text{sann verdi er } \mu = 0.17) \\ &= P\left(\frac{\bar{X} - 0.17}{0.00577} \geq \frac{0.1595 - 0.17}{0.00577}\right) \\ &= P(Z \geq -1.82) = 1 - \Phi(-1.82) = 1 - 0.0344 = 0.9656 \end{aligned}$$

Dermed har denne veldig god styrke for dette alternativet - vi vil svært ofte oppdage et slikt avvik. Verdier for teststyrke rundt 80% og over er veldig bra!

Tillegg: Her er en alternativ løsning når vi dropper å regne ut forkastningsgrensen først. Denne varianten er nyttig hvis vi hadde lyst til å endre antallet observasjoner eller signifikansnivået til testen.

$$\begin{aligned} &P\left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geq 1.645 \mid \text{sann verdi er } \mu = 0.17\right) = P\left(\bar{X} \geq \mu_0 + 1.645 \frac{\sigma}{\sqrt{n}} \mid \text{sann verdi er } \mu = 0.17\right) \\ &= P\left(\bar{X} - 0.17 \geq (0.15 - 0.17) + 1.645 \frac{\sigma}{\sqrt{n}} \mid \text{sann verdi er } \mu = 0.17\right) \\ &= P\left(\frac{\bar{X} - 0.17}{\frac{\sigma}{\sqrt{n}}} \geq \frac{(0.15 - 0.17)}{\frac{\sigma}{\sqrt{n}}} + 1.645\right) \\ &= P\left(Z \geq \frac{(0.15 - 0.17)}{\frac{0.02}{\sqrt{12}}} + 1.645\right) = P(Z \geq -1.82) = 1 - P(Z \leq -1.82) = 1 - 0.0344 = 0.9656 \end{aligned}$$

Oppgave 7: Lineær regresjon

Vi skal studere antallet røde blodceller (per liter blod) til ulike idrettsutøvere. Målet er å undersøke hvordan antallet røde blodceller varierer med høyde.

Datasettet består av informasjon for 105 idrettsutøvere og om inneholder følgende variabler:

- Blodceller: antall røde blodceller, oppgis som antall/ 10^{12} per liter (tera pr liter i plottet under)
- Høyde: høyde i cm

Et kryssplott av datasettene er vist i oppgaven.

Spørsmål:

a) Den empiriske korrelasjonskoeffisienten mellom Høyde og Blodceller kan regnes ut, men du skal fra kryssplottet over velge det korrekte utsagnet om denne empiriske korrelasjonskoeffisienten mellom Høyde og Blodceller:

- den vil være mellom 0 og -1
- den vil være mellom 0 og 1
- den vil være større enn 1
- den vil være mindre enn -1

Innledning: En lineær regresjonsmodell $Y = \beta_0 + \beta_1 x + \varepsilon$ har blitt tilpasset i Python, der responsen Y er Blodceller og forklaringsvariabelen (kovariaten) x er Hoeyde. Deler av utskriften (`modell.summary()`) er vist her. Bruk utskriften til å svare på oppgave b, c og d.

Spørsmål:

b) Hva ble den estimerte regresjonslinja?

- $\hat{y} = 1.069 + 0.0199x$
- $\hat{y} = 1.069 + 0.630x$
- $\hat{y} = 0.0199 + 0.004x$
- $\hat{y} = 0.630 + 0.004x$
- $\hat{y} = 0.630 + 1.692x$
- $\hat{y} = 0.0199 + 5.669x$

c) Hvilken hypotesetest utføres i utskriften i linjen som starter med "Hoeyde" ?

- $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$
- $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 > 0$
- $H_0 : \beta_1 = 0.0199$ mot $H_1 : \beta_1 \neq 0.0199$
- $H_0 : \beta_1 = 1$ mot $H_1 : \beta_1 \neq 1$
- $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 < 0$
- $H_0 : x \leq y$ mot $H_1 : x < y$
- $H_0 : \beta_1 = 0.004$ mot $H_1 : \beta_1 \neq 0.004$

d) Se på linjen som starter med "Hoeyde". Hvilket av følgende utsagn er korrekte?

- Stigningstallet er ikke forskjellig fra 0
- Kovariaten "Hoeyde" har ingen effekt i regresjonen
- Vi beholder nullhypotesen
- Vi forkaster nullhypotesen

Fasit

a) Det er en klar positiv sammenheng, så den empiriske korrelasjonskoeffisienten vil være mellom 0 og 1. (Numerisk er korrelasjonskoeffisienten 0.488 for dette datasettet.)

b) $\hat{y} = 1.069 + 0.0199x$

c) $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$

d) Vi ser at testobservatoren er 5.669, som gir en veldig lav p -verdi. Python oppgir p -verdien til 0, og det er lavere en signifikansnivået på 5%. Dermed forkaster vi nullhypotesen.

Det at vi forkaster nullhypotesen betyr at vi tror at stigningstallet er forskjellig fra 0, og at kovariaten "Hoeyde" har en effekt på Blodceller.

Ekstraoppgave 8: Punktestimering i laboratoriet

Innledning:

En laborant skal bestemme konsentrasjonen av et bestemt stoff i en prøve. La μ betegne denne ukjente konsentrasjonen.

Til å måle konsentrasjonen har laboranten tilgjengelig to måleinstrumenter som begge gir normalfordelte målinger med forventningsverdi μ , men de to måleinstrumentene har ulik nøyaktighet.

Instrument A har standardavvik $\sigma_x = 2.4$ og er ikke så dyrt å bruke, så der gjør laboranten 9 uavhengige målinger $X_1, X_2, \dots, X_9$ av den ukjente konsentrasjonen. La $\bar{X} = \frac{1}{9} \sum_{i=1}^9 X_i$.

Instrument B er dyrere i drift, og har standardavvik $\sigma_y = 0.6$. Laboranten får ikke lov til ta mer enn én måling, Y , med instrument 2. De 9 målingene med instrument 1 er uavhengige av den ene målingen med instrument 2.

Oppgaven til laboranten er å finne et godt estimat for μ . Laboranten vurderer to estimatorene:

$$\hat{\mu}_1 = \frac{1}{2}(\bar{X} + Y) \text{ og } \hat{\mu}_2 = 0.36\bar{X} + 0.64Y$$

Spørsmål:

- a) Finn forventningsverdien til de hver av de to estimatorene.

Her var fire mulige alternativer gitt,

- b) Regn ut numerisk verdi for variansen til $\hat{\mu}_1$, når det oppgis at $\bar{x} = 67.2$ og $y = 66.4$. Oppgi svaret med to desimaler, f.eks. 1.23 eller 34.56.

- c) Det oppgis at $\text{Var}(\hat{\mu}_2) = 0.23$. Hvilken av de to estimatorene vil du da foretrekke? Velg ett alternativ

- $\hat{\mu}_1$
- $\hat{\mu}_2$
- De er like gode.

- d) Anta at laboranten velger å bruke estimatoren $\hat{\mu}_2$. Hvilken fordeling har estimatoren? Velg ett alternativ

- Normalfordelt
- Binomisk fordelt
- Weibullfordelt
- Poisson-fordelt
- Geometrisk fordelt
- Eksponentialfordelt

Fasit:

- a) Begge estimatorene er forventningsrette.

$$E(\hat{\mu}_1) = E\left(\frac{1}{2}(\bar{X} + Y)\right) = \frac{1}{2}(E(\bar{X}) + E(Y)) = \frac{1}{2}(\mu + \mu) = \mu$$

$$E(\hat{\mu}_2) = E(0.36\bar{X} + 0.64Y) = 0.36(E\bar{X}) + 0.64E(Y) = 0.36\mu + 0.64\mu = \mu$$

- b) $\text{Var}(\hat{\mu}_1) = \text{Var}\left(\frac{1}{2}(\bar{X} + Y)\right) = \left(\frac{1}{2}\right)^2 \left(\frac{\sigma_x^2}{n} + \sigma_y^2\right) = 0.25 \cdot (2.4^2/9 + 0.6^2) = 0.25$

- c) Begge estimatorene er forventningsrette, men $\hat{\mu}_2$ har lavest varians. Derfor foretrekker vi $\hat{\mu}_2$.

- d)

Først er hver X_i normalfordelt, og dermed er også gjennomsnittet av uavhengige normalfordelte størrelser også normalfordelt. Deretter er en sum av uavhengige normalfordelte størrelser også normalfordelte. Dermed er korrekt svar "Normalfordelt".

Ekstraoppgave 9: Levetid og pålitelighet

Innledning: Levetiden til en bestemt type LED-lys er eksponentialfordelt med forventning 10 år. La T være levetiden til et tilfeldig LED-lys, og anta levetiden til ulike LED-lys er uavhengige av hverandre.

Spørsmål:

- a) Hva er sannsynligheten for at et LED-lys av denne typen virker lenger enn 1 år. Oppgi svaret med 3 desimaler, f.eks. 0.123 eller 0.987.

b) I et rom installeres 2 LED-lys av den aktuelle typen. Anta at levetiden til de to lysene er uavhengig av hverandre. Finn sannsynligheten for at minst ett LED-lys feiler i løpet av 1 år. Oppgi svaret med 3 desimaler, f.eks. 0.123 eller 0.987.

c) I et annet rom er det 6 LED-lys. Anta at levetiden til de 6 lysene er uavhengig av hverandre. Hva er sannsynligheten for at alle LED-lysene virker i minst 5 år? Oppgi svaret med 4 desimaler, f.eks. 0.1234 eller 0.9876.

Fasit:

I eksponentialfordelingen har vi parameter λ der $E(T) = \frac{1}{\lambda}$, slik at hvis $E(T) = 10$ år så er $\lambda = \frac{1}{10} = 0.1/\text{år}$. Både $F(t)$ og $R(t) = 1 - F(t)$ er oppgitt på formelarket.

a) $P(T > 1) = R(1) = e^{-\lambda \cdot 1} = e^{-0.1} = 0.905$

b) La X være antallet LED-lys som feiler i løpet av et år. Da vil X være binomisk fordelt med $p = P(T < 1) = 1 - 0.905 = 0.095$ og $n = 2$.

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (1 - p)^2 = 1 - 0.905^2 = 0.181$$

c) Hvis alle 6 LED-lys skal virke i minst 5 år ser vi på påliteligheten til det vi kan tenke på som en seriekobling av 6 LED-lys (neida, de er ikke koblet i serie, det er bare at vi vil at alle skal virke). Kall de 6 levetidene $T_1, \dots, T_6$. Først kan vi regne ut at $P(T_i > 5) = R_i(5) = e^{-5\lambda} = e^{-0.5} = 0.6065$.

$$P(T_1 > 5 \cap T_2 > 5 \cap \dots \cap T_6 > 5) = R_1(5) \cdot R_2(5) \cdots R_6(5) = 0.6065^6 = 0.0497 \text{ avrundt til } 0.050.$$