

Løsningsskisse til kontinuasjonseksamen i ISTx100y Statistikk for ingeniører 17.08.2022

Mette Langaas

Fasit

1: 0.85

2: 35%

3: 0.82

4a: 0.180 4b: 0.943

5a: 0.16 5b: 9.49 5c: 368

6: 0.79

7a: 2 7b: 3

8a: 0.27 8b: 0.22

9: 3.9

10: $[0.21, 0.39]$

11: 62

12a: 0.003 12b: 0.98

13a: $\hat{y} = 13.4428 + 0.0796x$ 13b: 37.32

Oppgave 1: Kjemisk laboratorium

Tema: sannsynlighet for union av hendelser.

Vi har to hendelser: A er at forsøk A lykkes og B er at forsøk B lykkes. Vi har oppgitt at $P(A) = 0.7$, $P(B) = 0.4$ og $P(A \cap B) = 0.25$.

Sannsynligheten for at minst ett av de to forsøkene lykkes er gitt som $P(A \cup B)$.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.7 + 0.4 - 0.25 = 0.85$$

Oppgave 2: Vareparti

Tema: setning om total sannsynlighet.

La hendelsen A være at et tilfeldig valgt produkt fra vareparti A er valgt, og tilsvarende for B . Videre la K være hendelsen at et tilfeldig valgt produkt er av ypperste kvalitet.

I oppgaveteksten er følgende sannsynligheter gitt: $P(K \mid A) = 0.6$, $P(K \mid B) = 0.4$ og $P(K) = 0.47$.

Spørsmålet er hvor stor andel av produktene som er fra vareparti A, dvs $P(A)$.

Vi bruker setningen om total sannsynlighet for å skrive $P(K)$:

$$P(K) = P(K \mid A)P(A) + P(K \mid B)P(B)$$

I tillegg vet vi at det bare er to varepartier, så $P(B) = 1 - P(A)$. Da er det bare $P(A)$ som er ukjent når vi kombinerer det vi vet.

$$0.47 = 0.6 \cdot P(A) + 0.4 \cdot (1 - P(A))$$

og vi løser ut for $P(A)$ og får

$$0.2 \cdot P(A) = 0.47 - 0.4 = 0.07$$

slik at $P(A) = 0.35$.

Oppgave 3: Diskret stokastisk variabel

Tema: standardavvik til diskret stokastisk variabel fra punktsannsynlighet.

Standardavviket er gitt som kvadratroten til variansen. Variansen til en diskret stokastisk variabel X er definert som:

$$\text{Var}(X) = \sum_{\text{alle } x} (x - \mu)^2 P(X = x)$$

der $\mu = E(X) = \sum_{\text{alle } x} xP(X = x)$.

Vi har at $\mu = -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0$. Dermed blir variansen

$$(-1 - 0)^2 \cdot \frac{1}{3} + (0 - 0)^2 \cdot \frac{1}{3} + (1 - 0)^2 \cdot \frac{1}{3} = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

og standardavviket blir da $\sqrt{\frac{2}{3}} = 0.82$.

Oppgave 4: Mineralforekomst

Tema: Punktsannsynlighet og kumulativ sannsynlighet i binomisk fordeling.

Antall hull, X , der man finner mineral er binomisk fordelt med $p = 0.4$ og $n = 20$ forsøk gjøres.

a) $P(X = 8) = \binom{20}{8} 0.4^8 (1 - 0.4)^{12} = 0.18$.

Alternativt med tabellen: $P(X = 8) = P(X \leq 8) - P(X \leq 7) = 0.596 - 0.416 = 0.180$.

b) $P(X < 12) = P(X \leq 11) = 0.943$ ved oppslag i tabellen.

Oppgave 5: Vannforbruk på turisthytte

Tema: Sum av normalfordelte stokastiske variabler. Forvening og standardavvik. Kumulativ sannsynlighet. Kritisk verdi i normalfordelingen.

a) La X angi vannforbruk til tilfeldig valgt gjest. $P(X > 39) = 1 - P(X \leq 39) = 1 - \Phi\left(\frac{39 - 36}{3}\right) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587 \approx 0.16$

Videre:

$X_1, X_2, \dots, X_{10}$ er uavhengige normalfordelte stokastiske variabler (med forventningsverdi 36 og standardavvik 3) som angir vannforbruk for hver av 10 tilfeldig valgte gjester som dusjer, og summen av disse er $Y = \sum_{i=1}^{10} X_i$. I resten av oppgaven jobber vi med denne summen, Y .

Vi vet at Y da er normalfordelt (siden den er en sum av uavhengige normalfordelte stokastiske variabler), og forventningsverdi blir: $E(Y) = E(\sum_{i=1}^{10} X_i) = \sum_{i=1}^{10} E(X_i) = \sum_{i=1}^{10} 36 = 360$.

b) Først varians (vi har kun regneregler for varians).

Husk at $\text{Var}(X_i) = 9$ siden variansen er kvadratet av standardavviket.

$$\text{Var}(Y) = \text{Var}(\sum_{i=1}^{10} X_i) = \sum_{i=1}^{10} \text{Var}(X_i) = \sum_{i=1}^{10} 9 = 90.$$

Standardavviket blir da $\sqrt{90} = 9.49$.

Alternativt: Variansen blir $10 \cdot \text{Var}(X_i)$ og standardavviket blir $\sqrt{10} \cdot \text{SD}(X_i)$, som gir samme svar.

c) La v være den ukjente vannmengden vi er ute etter. Det vi skal lagre i vår tank må være minst like mye som det gjestene bruker, så vi må skrive opp sannsynligheten for at gjestene bruker mer enn denne mengden v - og sette den til å være 0.2.

$$P(Y > v) = 0.2$$

$$1 - P(Y \leq v) = P\left(\frac{Y - 360}{9.49} \leq \frac{v - 360}{9.49}\right) = 0.2$$

Fra tabell for kritisk verdi i standard normalfordelingen finner vi at $P(Z > 0.842) = 0.2$.

$$\frac{v - 360}{9.49} = 0.842$$

Løser vi ut for v finner vi $v \geq 0.842 \cdot 9.49 + 360 = 367.99$, og dermed må vi lagre 368 liter vann.

Oppgave 6: Standardavvik til gjennomsnitt

Tema: Varians til gjennomsnitt. Standardavvik fra varians.

Variansen til et gjennomsnitt av n uavhengige stokastiske variabler fra sannsynlighetsfordeling med standardavvik σ er $\text{Var}(\bar{X}) = \text{Var}(\frac{1}{n} \sum_{i=1}^n X_i) = (\frac{1}{n})^2 \sum_{i=1}^n \text{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$, og dermed er standardavviket $\sigma/\sqrt{n}$.

Setter vi inn at $\sigma = 2.5$ og $n = 10$ får vi $\frac{2.5}{\sqrt{10}} = 0.79$.

Oppgave 7: Varians til differanse av korrelerte stokastiske variabler

Tema: Forventningsverdi og varians til lineær kombinasjon av korrelerte variabler. Kovarians fra korrelasjon.

a)

Forventningsverdien til en lineær funksjon av to avhengige stokastiske variabler, $aX + bY$. Denne er gitt som

$$E(aX + bY) = aE(X) + bE(Y)$$

Hos oss er $a = 1$ og $b = -1$, og dermed får vi $E(1 \cdot X + (-1) \cdot Y) = 1 \cdot E(X) + (-1) \cdot E(Y) = 0 + 2 = 2$.

b)

Vi skal se på variansen til en lineær funksjon av to avhengige stokastiske variabler, $aX + bY$. Denne er gitt som

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$$

Hos oss er $a = 1$ og $b = -1$.

Sammenhengen mellom korrelasjon (kalt korrelasjonskoeffisienten på formelarket), varianser og kovarians er (formelarket 3.3):

$$\rho = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

der $\sigma_x = \text{SD}(X) = \sqrt{\text{Var}(X)}$ og $\sigma_y = \text{SD}(Y) = \sqrt{\text{Var}(Y)}$.

Vi har fått oppgitt varianser og korrelasjon, og da har vi alt vi trenger. Vi regner først ut kovariansen.

$\text{Cov}(X, Y) = \rho \cdot \sigma_x \sigma_y = 0.5 \cdot 1 \cdot \sqrt{4} = 1$ som vi setter inn i formelen for variansen:

$\text{Var}(1 \cdot X + (-1) \cdot Y) = 1^2\text{Var}(X) + (-1)^2\text{Var}(Y) + 2 \cdot 1 \cdot (-1) \cdot \text{Cov}(X, Y) = 1 + 4 - 2 \cdot 1 = 3$.

Oppgave 8: Levetid

La T være en stokastisk variabel som angir levetiden til et tilfeldig valgt LED-lys. I eksponentialfordelingen har vi parameter λ der $E(T) = \frac{1}{\lambda}$, slik at hvis $E(T) = 16$ år så er $\lambda = \frac{1}{16} = 0.0625/\text{år}$.

Både kumulativ fordeling $F(t)$ og pålitelighetsfunksjon $R(t) = 1 - F(t)$ er oppgitt på formelarket.

a) $P(T < 5) = P(T \leq 5) = F(5) = 1 - e^{-\lambda 5} = 1 - e^{-0.0625 \cdot 5} = 1 - e^{-0.3125} = 0.27$.

b) Hvis alle 4 LED-lys skal virke i minst 8 år ser vi på påliteligheten til det vi kan tenke på som en seriekobling av 4 LED-lys (neida, de er ikke koblet i serie, det er bare at vi vil at alle skal virke). Kall de 4 levetidene $T_1, \dots, T_4$. Først kan vi regne ut at $P(T_i > 6) = R_i(6) = e^{-6\lambda} = e^{-6 \cdot 0.0625} = e^{-0.375} = 0.6873$.

$$P(T_1 > 6 \cap T_2 > 6 \cap \dots \cap T_4 > 6) = R_1(6) \cdot R_2(6) \cdots R_4(6) = 0.6873^4 = 0.2231.$$

Oppgave 9: Estimat for standardavvik

Tema: Punktestimering av standardavvik.

Vi har et tilfeldig utvalg av $n = 36$ observasjoner fra aldersbestemmelse av bergart. Vi ønsker et estimat for standardavviket i sannsynlighetsfordelingen målingene er tatt fra.

Estimator for varians er $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ der $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ er gjennomsnittet. Estimator for standardavviket blir da kvadratroten av S^2 .

Vi har fått oppgitt at $\sum_{i=1}^{36} (x_i - \bar{x})^2 = 534$. Dermed blir standardavviket gitt som $\sqrt{\frac{534}{35}} = 3.9$.

Kommentar: hvis man heller hadde brukt estimatoren $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ (som ikke er forventningsrett), ville man fått samme numeriske svar (avrundet til 1 desimal), siden $\sqrt{\frac{534}{36}} = 3.851$.

Oppgave 10: Konfidensintervall for andel lyspærer med bestemt defekt

Tema: Konfidensintervall (KI). Binomisk tilnærmet med normal.

Et konfidensintervall for binomisk andel (basert på normaltilnærming) er gitt som

$$\left[\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

Vi har oppgitt at det er $n = 100$ lyspærer, der $x = 30$ er defekte. Estimat for andel er dermed $\hat{p} = \frac{x}{n} = 30/100 = 0.3$. Vi skal ha et 95/%₀₀ konfidensintervall, så $\alpha = 0.025$, som gir fra tabell “kritisk verdi i standard normalfordelingen” $z_{0.025} = 1.960$.

Innsatt gir dette

$$\left[0.3 - 1.96\sqrt{0.3 \cdot 0.7/100}, 0.3 + 1.96\sqrt{0.3 \cdot 0.7/100} \right]$$

som gir $[0.21, 0.39]$

Oppgave 11: Konfidensintervallets lengde og antall observasjoner

Tema: Konfidensintervall for normalfordelte data. Lengde av KI. Antall observasjoner.

Konfidensintervallet for μ med kjent σ har grenser $\bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$.

Lengden til intervallet blir øvre grense minus nedre grense, dvs $2z_{\alpha/2}\sigma/\sqrt{n}$.

I oppgaven er $\sigma = 2$, og $\alpha = 0.05$, som gir $z_{\alpha/2} = 1.960$ fra tabell over kritiske verdier i standard normalfordelingen.

Videre er n antall målinger, som vi skal finne.

Vi ønsker at lengden til intervallet ikke overskrider 1, dermed

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

$$\sqrt{n} \geq 2z_{\alpha/2}\sigma$$

$$n \geq (2z_{\alpha/2}\sigma)^2$$

Setter vi inn tall finner vi

$$n \geq (2 \cdot 1.96 \cdot 2)^2 = 61.47$$

Dermed må vi ha minst 62 målinger.

Oppgave 12: Hypotesetest av massetetthet

Tema: Hypotesetest for normalfordelte data. Ensidig test. P -verdi. Styrkefunksjon.

- a) La X_i være målingen av massetetthet for prøve i . Vi skal utføre en hypotesetest og regne ut p -verdi. Den alternative hypotesen er at $\mu < 3.1$ så vi har en venstresidig test.

Testobservator er $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ der vi har fått oppgitt at $\sigma = 0.12$, $\mu_0 = 3.10$, $\bar{x} = 2.99$ og $n = 9$.

Testobservatoren får verdi $z = \frac{2.99 - 3.10}{0.12/\sqrt{9}} = -2.75$

P -verdien regner vi som sannsynligheten for å observere denne verdien av testobservatoren eller en mer ekstrem verdi, og når nullhypotesen er sann er testobservatoren standard normalfordelt. P -verdien blir da $P(Z < z) = \Phi(Z < -2.75) = 0.003$.

- b) Det første vi gjør er å skrive opp forkastningsregelen til testen når vi bruker signifikansnivå 0.05, og så flytte litt om slik at vi får en forkastningsregel der $\bar{X}$ står alene på venstre side.

Testobservatoren er $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$, og vi forkaster for lave verdier av testobservatoren (venstresidig test). Når vi har signifikansnivå 0.05 vil vi forkaste når $Z \leq -1.645$, der vi har funnet -1.645 i tabellen for kritisk verdi i standard normalfordelingen.

$$\begin{aligned} Z &\leq -1.645 \\ \frac{\bar{X} - 3.10}{0.12/\sqrt{9}} &\leq -1.645 \\ \bar{X} &\leq 3.10 - 1.645 \cdot \frac{0.12}{\sqrt{9}} \\ \bar{X} &\leq 3.0342 \end{aligned}$$

Deretter skal vi regne ut teststyrken når $\mu = 2.95$.

$$\begin{aligned} P(\text{forkaste } H_0 \text{ når alternativ hypotese er sann og } \mu = 2.95) &= P(\bar{X} \leq 3.0342 \text{ når } \mu = 2.95) = \\ P\left(\frac{\bar{X} - 2.95}{0.12/\sqrt{9}} \leq \frac{3.0342 - 2.95}{0.12/\sqrt{9}}\right) &= P\left(\frac{\bar{X} - 2.95}{0.12/\sqrt{9}} \leq 2.105\right) = \Phi(2.10) = 0.98 \end{aligned}$$

Oppgave 13: Trykkfasthet av betong

- a) I kolonnen coef har vi parameterestimer, og Intercept-raden tilhører parameteren β_0 og sement tilhører β_1 . Estimert verdi av β_0 kaller vi $\hat{\beta}_0$ og den er 13.4428. Estimert verdi av β_1 kaller vi $\hat{\beta}_1$ og den er 0.0796. Dermed blir regresjonslinja

$$\hat{y} = 13.4428 + 0.0796x$$

b)

Predikert verdi for sementinnhold 300 får vi ved å sette inn $x = 300$ i formelen til den estimerte regresjonslinja.

$$\hat{y} = 13.4428 + 0.0796 \cdot 300 = 37.32.$$

Informasjon om datasettet:

Dataene er hentet fra <https://archive.ics.uci.edu/ml/datasets/concrete+compressive+strength> og beskrevet i denne publikasjonen: I-Cheng Yeh, "Modeling of strength of high performance concrete using artificial neural networks," Cement and Concrete Research, Vol. 28, No. 12, pp. 1797-1808 (1998).

Infomasjon om betong: <https://snl.no/betong> <https://ndla.no/nb/subject:1:0d67724e-d9fa-4365-9839-4cc91c012855/topic:1:eaf487bb-a2ad-4725-8e57-fd709dd4cade/topic:2:158640/resource:1:142929>