Statistics for Engineering

Thobias Høivik

Fall 2025

Contents

1 Introduction

This course covers introductory statistics and is taught at the Western Norway University of Applied Science. This will be among my most scattered and improvised notes, probably mostly in norwegian.

2 Lecture 1

Mål:

- Finne matematiske størrelser som beskriver data i eit datasett
- Sentralmål
- Spredningsmål

2.1 Sentralmål

- Finne ein verdi som representerer ein "typisk" enhet i ei mengde.

Example 2.1

Lønna til 13 personer oppgitt som: $110, 125, 125, 300, 350, 370, 375, 380, 390, 410, 430, 435, 440$.

Modus/typetal.

Verdien som dukkar opp flest gangar. Fra eksempellet har at vi moduslønn er 125.

Median. Median er den midterste verdien i ei datamengde. I dette tilfellet fra Example 2.1 har vi at medianlønn ligger på 375. I tilfellet hvor det er et partall antall verdier tar vi gjennomsnittet av de to midterste verdiene.

Gjennomsnitt.

Gjennomsnittet av en datamengde X er gitt ved

$$\overline{X} = \frac{1}{|X|} \sum_{i=1}^{|X|} x_i$$

I dette tilfellet, fra Example 2.1, har vi gjennomsnitt på 326.

Anta nå at en 14-ende tjener blir lagt til i mengden som tjener $30,000$. Da er gjennomsnittslønnen på de 14 personene rundt 2445 som ligger langt over den nest høgste tjeneren. Med andre ord drar person nummer 14 gjennomsnittet opp vanvittigt.

2.2 Spredningsmål.

Målet med et spredningsmål er å beskrive hvor stor variasjon det er i datasettet.

Example 2.2

Gitt to mengder X, Y.
$X = \{80, 90, 100, 110, 120\}$
$\overline{X} = 100$
$Y = \{20, 60, 100, 140, 180\}$
$\overline{Y} = 100$

Variasjonsbredde.

Variasjonsbredda er største verdi minus minste verdi:

$$\max X - \min X$$

Fra Example 2.2 har vi at at variasjonsbredda til X er 40 og variasjonsbredda til Y er 160.

Varians og standardavik.

Varians:

$$\mathrm{var}(X) = s_X^2 = \frac{1}{|X|-1} \sum_{i=1}^{|X|} \left(x_i - \overline{X}\right)^2$$

Standardavik:

$$\mathrm{std}(X) = \sqrt{s_X^2} = \sqrt{\mathrm{var}(X)}$$

Viss vi ser på Example 2.3 får vi

$$\mathrm{var}(X) = \frac{(80-100)^2 + (90-100)^2 + \cdots + (120-100)^2}{4} = 250 = 15.8^2$$

og

$$\mathrm{std}(X) = 15.8$$

mens for Y

$$\mathrm{var}(Y) = \frac{(20-100)^2 + (40-100)^2 + \cdots + (180-100)^2}{4} = 4000 = 63.25^2$$

$$\mathrm{std}(Y) = 63.25$$

2.3 Tsjebytsjevs regel

"Sammenhengen mellom ei datamengde, gjennomsnitt og standardavik."

Minimum 75% av observasjonane har verdi i intervallet

$$\left[\overline{X} - 2s_X, \overline{X} + 2s_X\right]$$

3 Lecture 2

Vi begynner med begrepet sannsyn.

Mål:

- kjennskap til sentrale begrep
- rekne ut sannsyn
- kjenne til grunnleggande begrep frå mengdelære

3.1 Stokastisk forsøk, utfallsrom og hending

Definition 3.1: Stokastisk forsøk

- Vi veit kva utfall som er mulig
- Berre eitt utfall kan skje i kvart forsøk
- Veit ikkje utfallet vil bli

Example 3.1

Terningskast er eit eksempel på eit stokastisk forsøk. Vi veit at vi får $1-6$. Vi kan kunn få eit utfall. Vi veit ikkje på førhånd kva vi får.
Vi kaster 2 terninger. Dei mulige utfalla er

$$\begin{array}{cccccc} 1,1 & 1,2 & 1,3 & 1,4 & 1,5 & 1,6 \\ 2,1 & 2,2 & 2,3 & 2,4 & 2,5 & 2,6 \\ 3,1 & 3,2 & 3,3 & 3,4 & 3,5 & 3,6 \\ 4,1 & 4,2 & 4,3 & 4,4 & 4,5 & 4,6 \\ 5,1 & 5,2 & 5,3 & 5,4 & 5,5 & 5,6 \\ 6,1 & 6,2 & 6,3 & 6,4 & 6,5 & 6,6 \end{array}$$

som utgør $6^2 = 36$ mulige utfall.

Example 3.2

Kaster eit kronestykke 3 gongar. Dei mulige utfalla er

$$\begin{array}{cccc} KKK & KKM & KMK & KMM \\ MKK & MKM & MMK & MMM \end{array}$$

som utgjør $2^3 = 8$ mulige utfall.

Definition 3.2: Utfallsrom og hending

Utfallsrommet S er mengda av alle mulige utfall.
Hending er eitt eller fleire utfall som oppfyller ein gitt betingelse.

Example 3.3

Viss vi tar utgangspunkt i Example 3.1 kan vi se på hendinga $A = \{\text{sum er } 4\} = \{(3,1),(2,2),(1,3)\}$.
$B = \{\text{minst eit partal}\} = \{(1,2),(1,4),(1,6),\dots\}$ som gir $|B| = 27$.

Example 3.4

Viss ein målar høgda til personer er tilstandsrommet $S = \{\text{alle reelle tal mellom } 0.5 \text{ og } 2.5\}$.

3.2 Sannsyn for handling

- Bruke modell for sannsyn
- Empiri
- Magefølelse

Notasjon.

La A være ei hending:

- sannsyn for $A : P(A)$
- $0 \leq P(A) \leq 1$

Definition 3.3: Uniform sannsynlighetsmodell

La S være utfallsrommet til eit stokastisk forsøk, og la A være ei hending.
Viss alle utfall er like sannsynlige, då er

$$P(A) = \frac{\text{antal utfall der } A \text{ er gunnstig}}{\text{antal mulige utfall}}$$

Example 3.5

Betrakt terningskast eksempellet fra tidligere.
La S være terningskast med to terninger.

$$A = \{\text{sum er } 4\}$$

$$P(A) = \frac{3}{6^2} = \frac{3}{36} = \frac{1}{12}$$

$$B = \{\text{minst eitt partal}\}$$

$$P(B) = \frac{27}{36} = \frac{3}{4}$$

Empirisk sannsyn.

Kva er sannsynligheita for at ein person er mellom 175cm og 180cm?

Dette er ikkje ein uniform modell siden det ikkje er like stor sannsynligheit for ein person å ha ein gitt høgde.

Det vi kan gjere da er å velge 1000 personer, måle høgda og lage ein frekvenstabell.

Da ville vi hatt

$$P(\text{mellom 175 og 180}) = \text{ relativ frekvens}$$

Store tals lov.

Viss eit stokastisk forsøk blir gjentatt mange gangar så vil den empiriske sannsynligheita nærme seg den teoretiske sannsynligheita.

3.3 Mengdelære

Example 3.6

$A = \{\text{sum} = 4\}$, $P(A) = \frac{1}{12}$.
$B = \{\text{minst eitt partal}\}$, $P(B) = \frac{3}{4}$.
$C = \{\text{sum} \geq 9\}$, $P(C) = \frac{5}{18}$.
$P(\text{sum } = 4 \text{ og minst eitt partal})$?
$P(\text{sum})4 \text{ og sum} \geq 9)$?

Definition 3.4: Mengde

Ei mengde A er ei samling element (fra ei utvalgsmengde S).

Definition 3.5: Union, snitt, komplement, disjunkt

La A, B være mengder med element fra ei utvalgsmengde S.
Union.

$$A \cup B = \{x : x \in A \vee x \in B\}$$

Snitt.

$$A \cap B = \{x : x \in A \wedge x \in B\}$$

Komplement.

$$\overline{A} = A^c = S \setminus A = \{x \in S : x \notin A\}$$

Disjunkt.
A og B er disjunkte viss

$$A \cap B = \emptyset$$

4 Lecture 3

4.1 Grunnreglar for sannsynlighet

La S være utfallsrommet til eit stokastisk forsøk, og la A, B være hendingar.

1. $0 \leq P(A) \leq 1$
2. $P(S) = 1, P(\emptyset) = 0$
3. Viss A og B er disjunkte hendingar ($A \cap B = \emptyset$), då er $P(A \cup B) = P(A) + P(B)$
4. $P(\overline{A}) = 1 - P(A)$

Example 4.1

Betrakt kastingen av to terninger.
$A = \{\text{summen er } 4\}$
$B = \{\text{minst 1 terning er partal}\}$
$C = \{\text{sum } \geq 9\}$
$P(A) = \frac{3}{36}, P(B) = \frac{27}{36}, P(C) = \frac{10}{36}$
$P(A \cup C) = P(A) + P(C) = \frac{13}{36}$ siden A og C er disjunkte (ingen tall x oppfyller $x = 4 \wedge x \geq 9$ samtidig).
$P(A \cup B)$. Her kan vi ikke bruke regel 3, siden $2 + 2 = 4$ og $(2, 2)$ oppfyller at minst en er partal, så $A \cap B = \{(2, 2)\} \neq \emptyset$. I dette tilfellet har vi $P(A \cup B) = P(A \cup B) - P(A \cap B) = \frac{29}{36}$ (som addisjonsregelen for ikke-disjunkte mengder fra MAT210).
$P(\overline{A \cup B}) = 1 - P(A \cup B) = \frac{7}{36}$

4.2 Betinget sannsynlighet

Example 4.2

Eit bilfirma har 2000 tilsette.
La
$A = \{\text{stemmer Ap}\}$
$B = \{\text{er bilmekaniker}\}$
La oss si at vi vet at en person er bilmekaniker, hva er da sjansen for at de stemmer Ap?
Viss vi vet at blant bilmekanikere så stemmer 500 Ap og 380 ikke da er sjansen for at, gitt en bilmekaniker, de stemmer Ap

$$P(A|B))\frac{500}{500 + 380} = \frac{500}{880}$$

Den betinget sannsynligheten for at A skjer viss B har skjedd

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Fra dette følger det et par formler

1. $P(A \cap B) = P(A|B) \cdot P(B)$

2. $P(A) = P(A \cap B) + P(A \cap \overline{B})$

3. $P(A) = P(A|B) \cdot P(B) + p(A|\overline{B}) \cdot P(\overline{B})$

Theorem 4.1: Bayes teorem

La A, B være hendingar da har vi at

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Kort bevis.

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \\ &= \frac{\frac{P(B\cap A)}{P(A)}P(A)}{P(B)} \\ &= \frac{P(B \cap A)}{P(B)} \\ &= \frac{P(A \cap B)}{P(B)} \\ &= P(A|B) \end{aligned}$$

$\square$

Sensitivitet, spesifisitet, basisrate.

La

$B = \{\text{pasient er syk}\}$

$A = \{\text{testmetode er positiv}\}$

Vi har som mål å finne sjansen for at en person er syk når testen er positiv, $P(B|A)$.

- Sensitivitet: $P(A|B)$, sannsynligheten for at test er positiv når pasient er syk
- Spesifisitet: $P(\overline{A}|\overline{B})$, sannsynlighet for at test er negativ når pasient er frisk
- Basisrate: $P(B)$, sannsynlighet for at gitt pasient er syk

$$\begin{aligned} P(B|A) &= P(A|B)\frac{P(B)}{P(A)} \\ P(A) &= P(A|B) \cdot P(B) + P(A|\overline{B})P(\overline{B}) \\ P(\overline{B}) &= 1 - P(B) \\ P(A|\overline{B}) &= 1 - P(\overline{A}|\overline{B}) \\ \Rightarrow P(B|A) &= P(A|B)\frac{P(B)}{P(A|B) \cdot P(B) + (1 - P(\overline{A}|\overline{B}))(1 - P(B))} \end{aligned}$$

5 Lecture 4

5.1 Diskret synnsynlighetsmodell

Mål:

- kva er ein stokastisk variabel
- innholdet i sannsynlighetsmodell
- fordelingsfunksjonar og anvendelse
- sentralmål og spredningsmål for diskret sannsynlighetsmodell

Definition 5.1: Stokastisk variabel

Ein stokastisk variabel X er ein tilfeldig variabel som knytter alle utfall i utfallsrommet til ein verdi.

Definition 5.2: Verdimengde

Dei moglege verdiane av eit stokastisk forsøk kalles verdimengda V_X.

Example 5.1

Anta at vi kaster 2 terninger.
La $X = \{\text{summen av terningene}\}$.
$X = 4 = \{13, 22, 31\}$.
$X \leq 4 = \{11, 12, 21, 13, 22, 31\}$.
$V_X = \{2, 3, 4, \ldots, 9, 10, 11, 12\}$.
$Y = \{\text{antall partall}\}$.
$V_Y = \{0, 1, 2\}$.
I eit pengespill taper du 10 viss sum ≤ 6, vinner 8 viss sum $= 7$, og vinner 1 viss sum ≥ 8.
Viss vi lar Z betegne den stokastiske variabelen i pengespillet får vi:
$V_Z = \{-10, 8, 1\}$.

Definition 5.3: Sannsynlighetsfordeling

Ein sannsynlighetsfordeling til ein stokastisk variabel X inneholder:

1. verdimengden V_X
2. sannsynligheten for alle mulige utfall i V_X

V_X	$P(X=x)$	$F(x)$	$x \cdot P(X=x)$	$x^2 \cdot P(X=x)$
2	1/36	1/36	2/36	4/36
3	2/36	3/36	6/36	18/36
4	3/36	6/36	12/36	43/36
5	4/36	10/36	20/36	100/36
6	5/36	15/36	30/36	180/36
7	6/36	21/36	42/36	294/36
8	5/36	26/36	40/36	320/36
9	4/36	30/36	36/36	324/36
10	3/36	33/36	30/36	300/36
11	2/36	35/36	22/36	242/36
12	1/36	36/36	12/36	144/36
			$E(X) = 252/36 = 7$	$E(X^2) = 1974/36$

er hvordan synnsynlighetsfordelingen kan se ut for $X = \{\text{summen av terningene}\}$. $F(x)$ er definert som sannsynligheten for at $X \le x$. Med andre ord

$$F(x) = P(X \le x) = \sum_{y \in V_X, y \le x} P(X = y)$$

Example 5.2

Finn sannsynlighet for at sum er 7 eller mindre. Da skal vi finne $P(X \le 7)$ som næmlig er

$$F(7) = \frac{21}{36}$$

Finn sannsynlighet for at sum er større enn 8.

$$P(X > 8) = 1 - P(x \le 8) = 1 - F(8) = \frac{10}{36}$$

Sentralmål og spredningsmål.

Forventningsverdi:

$$E(X) = \mu_X = \sum_{x \in V_X} xP(X = x)$$

$$Var(X) = \sigma_X^2 = \left(\sum_{x \in V_X} x^2 P(X = x) \right) - \mu_X^2$$

$$std(X) = \sigma_X$$

6 Lecture 5

6.1 Fleire stokastiske variabler

I eit terningspel kaster ein to terningar. Det er 2 måtar å vinne.

I spel 1 vinner ein 2 viss summen av terningane er 7, vinner 1 viss sum ≥ 8 og ein vinner 0 viss sum ≤ 6.

I spel 2 vinner ein 2 viss begge terningane er partal, 1 viss ein av terningane er partal og 0 viss begge er oddetal.

La $X = \{\text{gevinst spel 1}\}$, $V_X = \{0, 1, 2\}$. $Y = \{\text{gevinst spel 2}\}$, $V_Y = \{0, 1, 2\}$.

	$y=0$	$y=1$	$y=2$	$P(X=x)$
$x=0$	$1/6$	$1/6$	$1/12$	$5/12$
$x=1$	$1/12$	$1/6$	$1/6$	$5/12$
$x=2$	0	$1/6$	0	$1/6$
$P(Y=y)$	$1/4$	$1/2$	$1/4$	1

$E(X), E(Y), Var(X), Var(Y)$

$$E(X) = \sum xP(X = x) = \mu_X$$

$$Var(X) = \sum x^2 P(X = x) - \mu_X^2$$

$$E(X) = 0 \cdot \frac{5}{12} + 1 \cdot \frac{5}{12} + 2 \cdot \frac{1}{6} = \frac{3}{4} = \mu_X$$

$$Var(X) = 0 + \frac{5}{12} + 4 \cdot \frac{1}{6} - \mu_X^2 = \frac{13}{12} - \mu_X^2 = \frac{13}{12} - \frac{9}{16} = \frac{25}{48} = \left(\frac{5\sqrt{3}}{12}\right)^2 = \sigma_X^2$$

$$E(Y) = \frac{1}{2} + \frac{2}{4} = 1 = \mu_Y$$

$$Var(Y) = \frac{1}{2} + \frac{4}{4} - \mu_Y^2 = \frac{1}{2} = \frac{1}{\sqrt{2}} = \sigma_Y^2$$

Interessante spørsmål.

- Viss eg spelar på begge spela, kva er forventa verdi og varians?
- Viss eg vinner på spel 1, aukar eg eller minkar eg sjansen for å vinne på spel 2?
- Er spela uavhengige av kvarandre?

La $Z = X + Y$ (den samla gevinsten ved å spele på begge spela). Verdimengda $V_Z = \{0, 1, 2, 3, 4\}$.

z	$P(Z=z)$	$zP(Z=z)$	$z^2P(Z=z)$
0	1/6	0	0
1	1/4	1/4	1/4
2	1/4	1/2	1
3	1/3	1	3
4	0	0	0
		$\mu_Z = 7/4$	$E(Z^2) = 17/4$

$E(Z) = \mu_Z = 7/4$

$Var(Z) = 17/4 - (7/4)^2 = (\sqrt{19}/4)^2 = \sigma_Z^2$

Theorem 6.1

La X og Y vere stokastiske variablar. La $Z = aX + bY + c$.

$$E(Z) = aE(X) + bE(X) + c$$

6.2 Kovarians og korrelasjon

- Er det samanheng mellom X og Y?
- Er X og Y uavhengige?
- Vil seier i X påvirke Y?

Kovarians

$$Cov(X,Y) = E(X \cdot Y) - \mu_X \cdot \mu_Y = \left(\sum_{x,y} x \cdot y P(X = x \wedge Y = y)\right) - \mu_X \mu_Y$$

Korrelasjon

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

Viss vi betrakter eksempellet fra tidligere får vi

$$E(X \cdot Y) = \frac{1}{6} + 2\frac{1}{6} + 2\frac{1}{6} = \frac{5}{6}$$

så

$$Cov(X,Y) = \frac{5}{6} - \mu_X \mu_Y = \frac{1}{12}$$

Deretter finner vi korrelasjon ved

$$Corr(X,Y) = \frac{1/12}{\sigma_X \sigma_Y} = \frac{1/12}{\left(\frac{5\sqrt{3}}{12}\right)\left(\frac{1}{\sqrt{2}}\right)} = \frac{\sqrt{2}}{5\sqrt{3}} = 0.164$$

For korrelasjon har vi at

$$-1 \le Corr(X,Y) \le 1$$

hvor nærmere -1 betyr mer negativ relasjon og nærmere 1 betyr mer positiv relasjon. Med andre ord ρ nær 1 betyr ein lineær samanheng mellom X og Y (altså en økning i X betyr en økning i Y og vice-versa). På samme måte betyr ρ nær -1 at en økning i en er en minking i den andre. $\rho = 0$ betyr ingen lineær samanheng.

Theorem 6.2

For stokastiske variabler X og Y og konstanter a, b, c har vi:

$$Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$$

Definition 6.1: Uavhengighet

To stokastiske variablar X og Y er uavhengige dersom

$$P(X = a \wedge Y = b) = P(X = a) \cdot P(Y = b)$$

Theorem 6.3

Viss to stokastiske variablar X og Y er uavhengige så er

$$Corr(X, Y) = 0$$

Denne implikasjonen går bare en veg.