

Statistics for Engineering

Thobias Høivik

Fall 2025

Contents

1	Introduction	3
2	Lecture 1	3
2.1	Sentralkmål	3
2.2	Spredningsmål.	3
2.3	Tsjebytsjevs regel	4
3	Lecture 2	5
3.1	Stokastisk forsøk, utfallsrom og hending	5
3.2	Sannsyn for handling	6
3.3	Mengdelære	7
4	Lecture 3	8
4.1	Grunnreglar for sannsynlighet	8
4.2	Betinget sannsynlighet	8

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 eksemplet 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

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

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 høyeste 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 .

$$X = \{80, 90, 100, 110, 120\}$$

$$\bar{X} = 100$$

$$Y = \{20, 60, 100, 140, 180\}$$

$$\bar{Y} = 100$$

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.

Varsans og standardavik.

Varsans:

$$\text{var}(X) = s_X^2 = \frac{1}{|X| - 1} \sum_{i=1}^{|X|} (x_i - \bar{X})^2$$

Standardavik:

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

Viss vi ser på Example 2.3 får vi

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

og

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

mens for Y

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

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

2.3 Tsjebytsjevs regel

"Sammenhengen mellom ei datamengde, gjennomsnitt og standardavik."

Minimum 75% av observasjonane har verdi i intervallet

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

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 forhånd kva vi får.

Vi kaster 2 terninger. Dei mulige utfalla er

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

som utgør $6^2 = 36$ mulige utfall.

Example 3.2

Kaster eit kronestykke 3 gongar. Dei mulige utfalla er

<i>KKK</i>	<i>KKM</i>	<i>KMK</i>	<i>KMM</i>
<i>MKK</i>	<i>MKM</i>	<i>MMK</i>	<i>MMM</i>

som utgør $2^3 = 8$ 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 $S = \{\text{alle reelle tal mellom 0.5 og 2.5}\}$.

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 eksemplet 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

$$A = \{\text{sum} = 4\}, P(A) = \frac{1}{12}.$$

$$B = \{\text{minst eitt partal}\}, P(B) = \frac{3}{4}.$$

$$C = \{\text{sum} \geq 9\}, P(C) = \frac{5}{18}.$$

$P(\text{sum} = 4 \text{ og minst eitt partal})?$

$P(\text{sum} \leq 4 \text{ og sum} \geq 9)?$

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. $P(\bar{A}) = 1 - P(A)$

Example 4.1

Betrakt kastingen av to terninger.

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

$B = \{\text{minst 1 terning er partall}\}$

$C = \{\text{sum} \geq 9\}$

$P(A) = \frac{3}{36}, P(B) = \frac{27}{36}, P(C) = \frac{10}{36}$

$P(A \cup C) = P(A) + P(C) = \frac{13}{36}$ 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 partall, 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).

$P(\bar{A} \cup \bar{B}) = 1 - P(A \cap B) = \frac{7}{36}$

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

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

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

Theorem 4.1: Bayes teorem

La A, B være hendinger da har vi at

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

Kort bevis.

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

□

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

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