

# TMA4245 Statistikk

Uke 17 - mandag

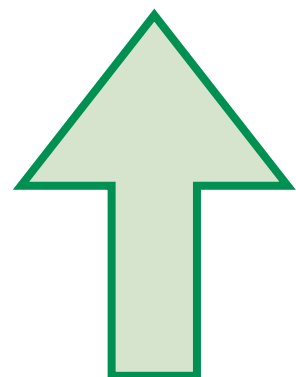
Forelesning  
2x45 min

*Thea / Ingeborg*



Eksamen

Mandag



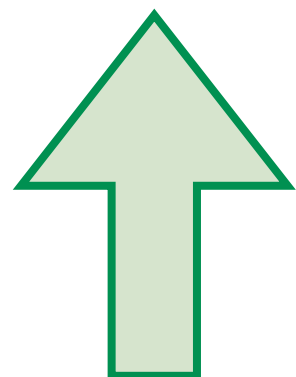
# TMA4245 Statistikk

Uke 17 - mandag

Forelesning  
2x45 min

*Thea / Ingeborg*

Mandag



## TMA4268 - Statistisk læring

Om emnet

Timeplan

Eksamensinfo

### Vurderingsordning

Vurderingsordning: Skriftlig skoleeksamen

Karakter: Bokstavkarakterer

| Vurdering              | Vekting | Varighet | Hjelpemidler |
|------------------------|---------|----------|--------------|
| Skriftlig skoleeksamen | 100/100 | 4 timer  | G            |

### Faglig innhold

Statistisk læring, multipl lineær regresjon, klassifikasjon, resamplingmetoder, modellseleksjon/regularisering, ikke-linearitet, trebaserte metoder, nevralt nettverk

# Plan

I dag:

Informasjon om eksamen

Grundig gjennomgang av gamle eksamensoppgaver

Hjelp før eksamen:

Digitalt forum åpent 24/7 (besvares i vanlig arbeidstid) (e-post til faglærere om oppgaver og pensum blir ikke besvart!)

Eksamenslab:

Tirsdag 21/5 10:15-14:00 i S3

Torsdag 23/5 10:15-14:00 i S3

27. mai klokka 09:00: Eksamen

# Tillatte hjelpemidler på eksamen

Tabeller og formler i statistikk



Ikke skriv noe annet enn navn i heftet!

Bestemt, enkel kalkulator

Ett stemplet gult A5-ark med egne håndskrevne (skrevet rett på arket) formler og notater (skriv på begge sider)

# Obs!

Noe nytt pensum fra høsten 2023 (Python, stokastisk simulering, feilforplantning, fortegnstest, permutasjonstest).

Få/ingen gamle eksamensoppgaver om disse temaene, og ingenting på temasidene, men det er like fullt en del av pensum!

**Husk å lese på disse temaene i eksamensperioden også!**

Vi har laget en side på wikisidene (“Nytt pensum”) der vi skriver utfyllende om dette. Vi henviser også til relevante oppgaver i øvingsopplegget/forelesninger.

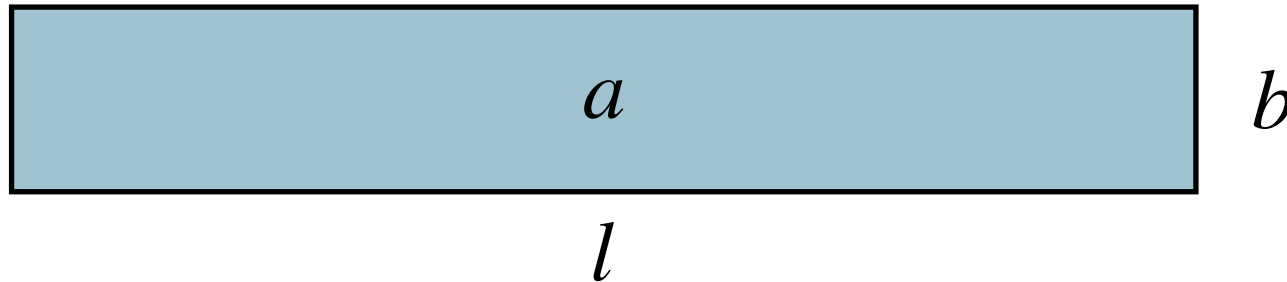
# Format på eksamen

Eksamen er på formen "digital skoleeksamen" (Inspira)

Hovedsakelig oppgaver der du skriver for hånd og scanner inn besvarelsen

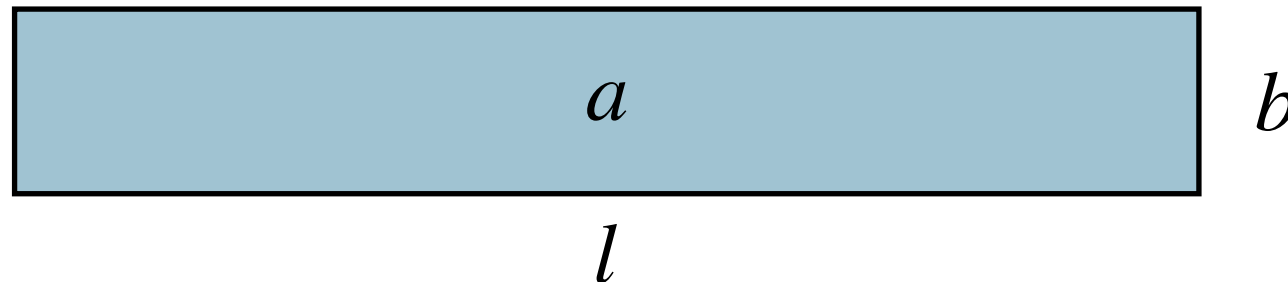
En liten andel flervalg/skriv inn tall

# Enkel oppgave om feilforplantning



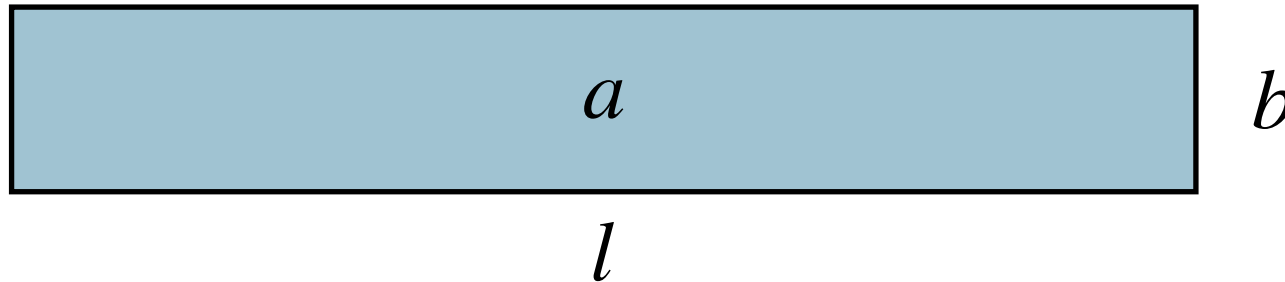
a) Vi skal finne arealet  $a$  av et rektangel som har ukjent lengde  $l$ , og ukjent bredde  $b$ . Sett opp målefunksjonen for arealet og bruk feilforplantningsformelen til å finne et tilnærmet uttrykk for forventning og varians til målingen av areal. Uttrykket skal inneholde  $l$ ,  $b$ ,  $\sigma_l$  og  $\sigma_b$ , der  $\sigma_l$  og  $\sigma_b$  er standard usikkerhet i målinger av lengde og bredde, henholdsvis.

# Enkel oppgave om feilforplantning



- a) Vi skal finne arealet  $a$  av et rektangel som har ukjent lengde  $l$ , og ukjent bredde  $b$ . Sett opp målefunksjonen for arealet og bruk feilforplantningsformelen til å finne et tilnærmet uttrykk for forventning og varians til målingen av areal. Uttrykket skal inneholde  $l$ ,  $b$ ,  $\sigma_l$  og  $\sigma_b$ , der  $\sigma_l$  og  $\sigma_b$  er standard usikkerhet i målinger av lengde og bredde, henholdsvis.
- b) I et rektangel som er langt og smalt, som i figuren over, vil usikkerheten i lengden eller bredden gi størst påvirkning på usikkerheten til arealet til rektanget? Begrunn svaret i maks 2 setninger.

# Enkel oppgave om feilforplantning



- a) Vi skal finne arealet  $a$  av et rektangel som har ukjent lengde  $l$ , og ukjent bredde  $b$ . Sett opp målefunksjonen for arealet og bruk feilforplantningsformelen til å finne et tilnærmet uttrykk for forventning og varians til målingen av areal. Uttrykket skal inneholde  $l$ ,  $b$ ,  $\sigma_l$  og  $\sigma_b$ , der  $\sigma_l$  og  $\sigma_b$  er standard usikkerhet i målinger av lengde og bredde, henholdsvis.
- b) I et rektangel som er langt og smalt, som i figuren over, vil usikkerheten i lengden eller bredden gi størst påvirkning på usikkerheten til arealet til rektanget? Begrunn svaret i maks 2 setninger.
- c) Standard usikkerhet i måling av lengde og bredde er henholdsvis  $\sigma_l = 0.05$  meter, og  $\sigma_b = 0.01$  meter. Måling av lengde og bredde ble  $x_l = 10$  meter, og  $x_b = 0.8$  meter. Skriv ned måleresultatet for arealet.

# Oppgave 4, desember 2010

På en forsøksgård utføres forsøk med produksjon av biomasse. Anta at biomassen,  $Y$ , av en plante er normalfordelt (gaussisk fordelt) med forventningsverdi  $E(Y) = 5$  og varians  $\text{Var}(Y) = 4$ .

a) Regn ut følgende tre sannsynligheter:

$$P(Y > 6)$$

$$P(4 < Y \leq 6)$$

$$P(Y > 6 \mid Y > 4)$$

# Oppgave 4, desember 2010

Anta nå at biomassen  $Y$  av en bestemt plante er avhengig av kultiveringsperioden  $x$ . For planten defineres kultiveringsperioden som tiden fra planten kan observeres over jordsmonnet til tidspunktet for biomassemåling.

Anta videre at sammenhengen mellom biomasse  $Y$  og en gitt kultiveringsperiode  $x$  kan modelleres som en lineær regresjon, uten konstantledd og med feilledd som er avhengig av kultiveringsperioden,

$$Y = \beta x + \varepsilon(x) \text{ for } x > 0,$$

der  $\varepsilon(x)$  er normalfordelt (gaussisk fordelt) med forventningsverdi  $E(\varepsilon) = 0$  og varians  $\text{Var}(\varepsilon) = \tau^2 x^2$ . Det betyr at standardavviket til feilleddet er proporsjonalt med kultiveringsperioden  $x$ . Biomassemålingen gjøres kun for positive kultiveringsperioder,  $x > 0$ .

Modellparameterene  $\beta$  og  $\tau$  ansees som ukjente.

# Oppgave 4, desember 2010

Anta nå at biomassen  $Y$  av en bestemt plante er avhengig av kultiveringsperioden  $x$ . For planten defineres kultiveringsperioden som tiden fra planten kan observeres over jordsmonnet til tidspunktet for biomassemåling.

Anta videre at sammenhengen mellom biomasse  $Y$  og en gitt kultiveringsperiode  $x$  kan modelleres som en lineær regresjon, uten konstantledd og med feilledd som er avhengig av kultiveringsperioden,

$$Y = \beta x + \varepsilon(x) \text{ for } x > 0,$$

der  $\varepsilon(x)$  er normalfordelt (gaussisk fordelt) med forventningsverdi  $E(\varepsilon) = 0$  og varians  $\text{Var}(\varepsilon) = \tau^2 x^2$ . Det betyr at standardavviket til feilleddet er proporsjonalt med kultiveringsperioden  $x$ . Biomassemålingen gjøres kun for positive kultiveringsperioder,  $x > 0$ .

Modellparameterene  $\beta$  og  $\tau$  ansees som ukjente.

Et forsøksopplegg med  $n = 5$  uavhengige målinger ved kultiveringsperioder  $x_1, x_2, \dots, x_5$  og tilhørende biomasser  $Y_1, Y_2, \dots, Y_5$  resulterte i følgende observasjoner:

|       |     |     |     |     |      |
|-------|-----|-----|-----|-----|------|
| $i$   | 1   | 2   | 3   | 4   | 5    |
| $x_i$ | 3   | 6   | 7   | 10  | 14   |
| $y_i$ | 1.0 | 5.0 | 3.0 | 3.0 | 10.0 |

Det oppgis at  $\sum_{i=1}^5 \frac{y_i}{x_i} = 2.61$  og  $\sum_{i=1}^5 \frac{y_i^2}{x_i^2} = 1.59$ .

b) Hvilken sannsynlighetsfordeling har  $Y_i$  gitt  $x_i$ ?

Utleid uttrykk for estimatorer for  $\beta$  og  $\tau^2$ , for eksempel ved å bruke sannsynlighetsmaksimeringsmetoden (maximum likelihood method).

Bruk tallverdier fra tabellen over til å beregne estimatorer for  $\beta$  og  $\tau^2$ .

# Oppgave 4, desember 2010

Anta nå at biomassen  $Y$  av en bestemt plante er avhengig av kultiveringsperioden  $x$ . For planten defineres kultiveringsperioden som tiden fra planten kan observeres over jordsmonnet til tidspunktet for biomassemåling.

Anta videre at sammenhengen mellom biomasse  $Y$  og en gitt kultiveringsperiode  $x$  kan modelleres som en lineær regresjon, uten konstantledd og med feilledd som er avhengig av kultiveringsperioden,

$$Y = \beta x + \varepsilon(x) \text{ for } x > 0,$$

der  $\varepsilon(x)$  er normalfordelt (gaussisk fordelt) med forventningsverdi  $E(\varepsilon) = 0$  og varians  $\text{Var}(\varepsilon) = \tau^2 x^2$ . Det betyr at standardavviket til feilleddet er proporsjonalt med kultiveringsperioden  $x$ . Biomassemålingen gjøres kun for positive kultiveringsperioder,  $x > 0$ .

Modellparameterene  $\beta$  og  $\tau$  ansees som ukjente.

Et forsøksopplegg med  $n = 5$  uavhengige målinger ved kultiveringsperioder  $x_1, x_2, \dots, x_5$  og tilhørende biomasser  $Y_1, Y_2, \dots, Y_5$  resulterte i følgende observasjoner:

|       |     |     |     |     |      |
|-------|-----|-----|-----|-----|------|
| $i$   | 1   | 2   | 3   | 4   | 5    |
| $x_i$ | 3   | 6   | 7   | 10  | 14   |
| $y_i$ | 1.0 | 5.0 | 3.0 | 3.0 | 10.0 |

Det oppgis at  $\sum_{i=1}^5 \frac{y_i}{x_i} = 2.61$  og  $\sum_{i=1}^5 \frac{y_i^2}{x_i^2} = 1.59$ .

- b) Hvilken sannsynlighetsfordeling har  $Y_i$  gitt  $x_i$ ?  
Utled uttrykk for estimatorer for  $\beta$  og  $\tau^2$ , for eksempel ved å bruke sannsynlighetsmaksimeringsmetoden (maximum likelihood method).  
Bruk tallverdier fra tabellen over til å beregne estimatorer for  $\beta$  og  $\tau^2$ .

Følgende estimator for  $\beta$  skal benyttes i resten av oppgaven:

$$B = \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{x_i}$$

- c) Anta i dette punktet at  $\tau^2 = 0.04$  er kjent.

Vi ønsker å utføre hypotesetesten

$$H_0: \beta = 0.50 \text{ mot } H_1: \beta > 0.50$$

Utled en forkastningsregel med signifikansnivå 0.05.

Bruk tallverdiene i tabellen over til å utføre testen.

Utled et uttrykk for styrken til testen over når  $\beta = 0.7$ , og regn ut tallsvar når  $n = 5$ .