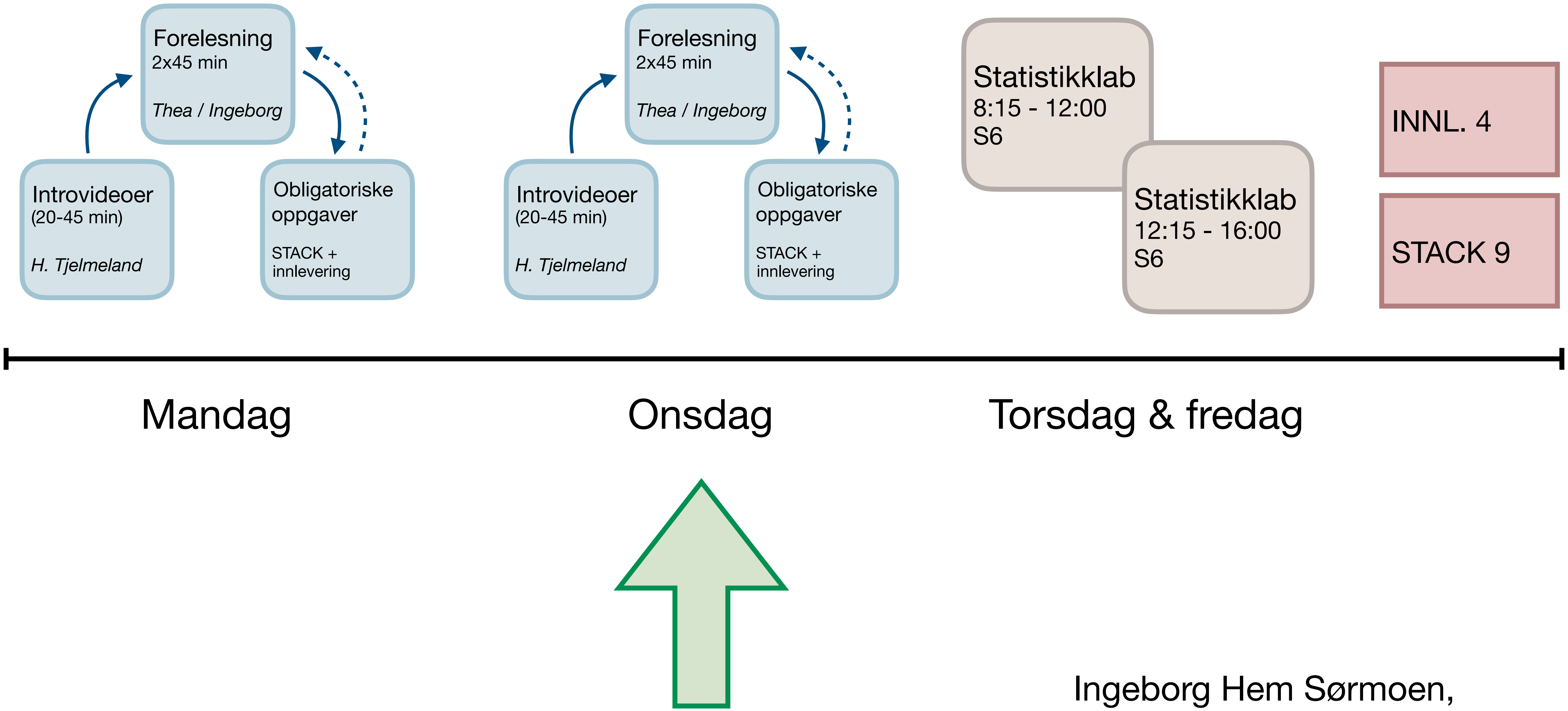
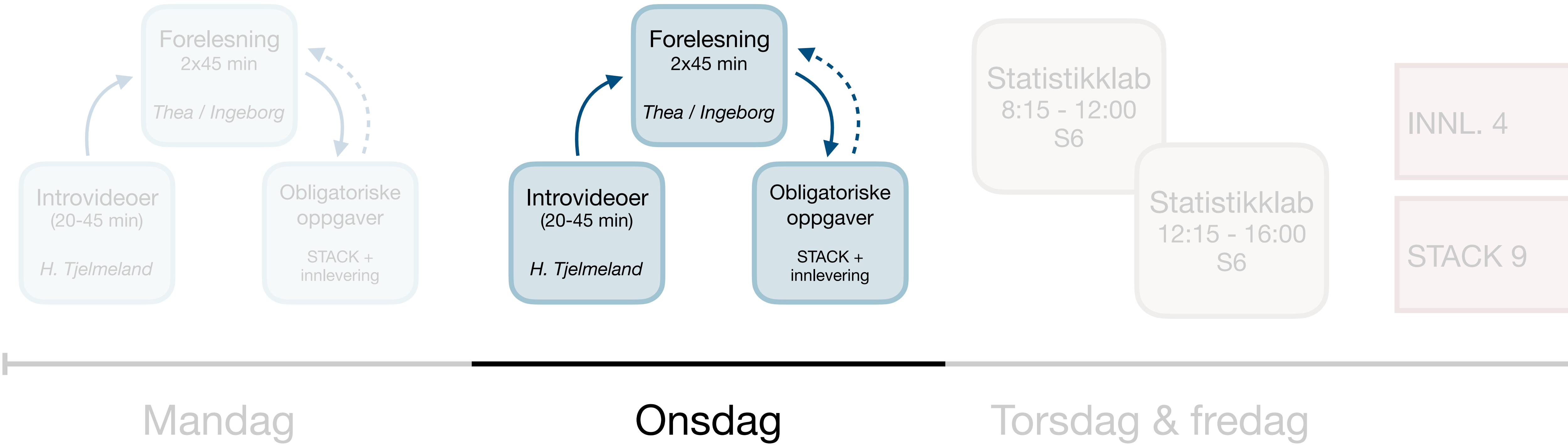
TMA4245 Statistikk

Uke 10 - onsdag



Ingeborg Hem Sørmoen,
Institutt for matematiske fag, NTNU

Uke 10 (mandag 4. mars - fredag 8. mars)



Introvideoer

Konfidensintervall og stokastisk simulering

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland
Institutt for matematiske fag
Norges teknisk-naturvitenskapelige universitet

Prediksjonsintervall

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland
Institutt for matematiske fag
Norges teknisk-naturvitenskapelige universitet

Konfidensintervall og valg av antall observasjoner

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland
Institutt for matematiske fag
Norges teknisk-naturvitenskapelige universitet

Oppgaver

STACK 9: Oppg. 4
INNL. 4: Oppg. 4c

Definisjon (Prediksjonsintervall)

Anta at vi har stokastiske variabler $X_1, X_2, \ldots, X_n$*, og at sannsynlighetsfordelingen til disse inneholder en parameter* θ *og at verdien til* θ *er ukjent. La* $x_1, x_2, \ldots, x_n$ *være observerte verdier for de stokastiske variabler* $X_1, X_2, \ldots, X_n$*. Anta videre at* $X^\star$ *representerer en fremtidig observasjon og at sannsynlighetsfordelingen til denne også avhenger av verdien av parameteren* θ*. Anta så at vi for to observatorer* $\widehat{X}_L^\star(X_1, X_2, \ldots, X_n)$ *og* $\widehat{X}_U^\star(X_1, X_2, \ldots, X_n)$ *har at*

$$P\left(\widehat{X}_L^\star(X_1, X_2, \ldots, X_n) \leq X^\star \leq \widehat{X}_U^\star(X_1, X_2, \ldots, X_n)\right) = 1 - \alpha,$$

for $\alpha \in (0, 1)$*. Det numeriske intervallet*

$$\left[\widehat{X}_L^\star(x_1, x_2, \ldots, x_n), \widehat{X}_U^\star(x_1, x_2, \ldots, x_n)\right]$$

kalles da et $(1 - \alpha) \cdot 100\%$*-prediksjonsintervall for* $X^\star$

$\star$ Prosedyre for å utlede et prediksjonsintervall for en ny observasjon $X^\star$ basert på $x_1, x_2, \ldots, x_n$

1. Bestemme en stokastisk variabel, $Z = h(X^\star, X_1, X_2, \ldots, X_n)$
   - inneholder ingen ukjente parametre
   - har en kjent fordeling

2. Finne kvantiler $z_{1-\frac{\alpha}{2}}$ og $z_{\frac{\alpha}{2}}$ slik at

   $$P\left(z_{1-\frac{\alpha}{2}} \leq h(X^\star, X_1, X_2, \ldots, X_n) \leq z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

Merk: $z_{\alpha/2}$ tilhører ikke nødvendigvis normalfordelingen

3. Løse hver ulikhet med hensyn på $X^\star$

4. Sette ulikhetene sammen igjen med $X^\star$ i midten

   $$P\left(\widehat{X}_L^\star(X_1, X_2, \ldots, X_n) \leq X^\star \leq \widehat{X}_U^\star(X_1, X_2, \ldots, X_n)\right) = 1-\alpha$$

5. Konkludere ved å skrive opp $(1-\alpha) \cdot 100\%$-prediksjonsintervallet for $X^\star$

   $$\left[\widehat{X}_L^\star(x_1, x_2, \ldots, x_n), \widehat{X}_U^\star(x_1, x_2, \ldots, x_n)\right]$$

Oppgave ~~4~~ 6 Bordtennisball

En produsent produserer bordtennisballer til profesjonelt bruk. Et av kravene for at en bordtennisball skal kunne brukes i konkurranse er at diameteren er 40 millimeter. Anta at diameteren, i millimeter, til en tilfeldig valgt bordtennisball er normalfordelt med ukjent forventningsverdi μ og kjent standardavvik $\sigma = 0.5$.

I de siste konkurransene har flere bordtennisspillere hevdet at bordtennisballen har hatt feil diameter.

Vi har et tilfeldig utvalg $X_1, X_2, \ldots, X_{10}$. Det er gitt at $\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 40.2$.

a) Utled et uttrykk for et 95 %-prediksjonsintervall for diameteren X_0 til en ny bordtennisball, uavhengig av $X_1, X_2, \ldots, X_{10}$. Regn så ut intervallet.

b) Hvis oppgaven isteden hadde bedt om et 95%-konfidensintervall for μ, hadde det blitt større eller mindre?

⋆ Prosedyre for å utlede et konfidensintervall for parameter θ basert på $x_1, x_2, \ldots, x_n$

1. Bestemme pivotal, $Z = h(\theta, X_1, X_2, \ldots, X_n)$
   - inneholder ingen ukjente parametre unntatt θ
   - har en kjent fordeling (som ikke avhenger av θ)

2. Finne kvantiler $z_{1-\frac{\alpha}{2}}$ og $z_{\frac{\alpha}{2}}$ slik at

   $$P\Big(z_{1-\frac{\alpha}{2}} \leq h(\theta, x_1, X_2, \ldots, X_n) \leq z_{\frac{\alpha}{2}}\Big) = 1 - \alpha$$

   Merk: $z_{\alpha/2}$ tilhører ikke nødvendigvis normalfordelingen

3. Løse hver ulikhet med hensyn på θ

4. Sette ulikhetene sammen igjen med θ i midten

   $$P\Big(\widehat{\theta}_L(X_1, X_2, \ldots, X_n) \leq \theta \leq \widehat{\theta}_U(X_1, X_2, \ldots, X_n)\Big) = 1-\alpha$$

5. Konkludere ved å skrive opp $(1-\alpha)\cdot 100\%$-konfidensintervallet for θ

   $$\Big[\widehat{\theta}_L(x_1, x_2, \ldots, x_n), \widehat{\theta}_U(x_1, x_2, \ldots, x_n)\Big]$$

~~13C~~ 4

Innledning: Anta at $Y_1, Y_2, \ldots, Y_n$ er uavhengige stokastiske variabler og at

$$Y_i \sim \mathrm{N}(\theta x_i(1 - x_i), \sigma^2 x_i),$$

der θ og σ^2 er parametre, mens $x_1, x_2, \ldots, x_n$ er kjente tall. Vi skal i denne oppgaven anta at verdien til θ er ukjent, mens verdien til σ^2 er kjent.

Det oppgis at sannsynlighetsmaksimeringsestimatoren for θ da er gitt som

$$\hat{\theta} = \frac{\sum_{i=1}^n Y_i(1-x_i)}{\sum_{i=1}^n x_i(1-x_i)^2}.$$

La $y_1, y_2, \ldots, y_n$ betegne verdiene vi har observert for de stokastiske variablene $Y_1, Y_2, \ldots, Y_n$.

Oppgave: Utled, med utgangspunkt i estimatoren $\hat{\theta}$, et 95%-konfidensintervall for θ. Uttrykk svaret som en funksjon av $x_1, x_2, \ldots, x_n$, de observerte verdiene $y_1, y_2, \ldots, y_n$ og den antatt kjente verdien σ^2.

Oppgave ~~1~~ 5 Elektriske komponenter

En bedrift produserer elektriske komponenter. Komponentene kan ha feil.
Vi er ~~på~~ kun interessert i om en komponent er feilfri eller ikke. Ledelsen i bedriften har over mange år overvåket produksjonen, og er sikre på at sannsynligheten for at en tilfeldig valgt komponent er feilfri er 0.9. Vi velger tilfeldig ut 20 komponenter fra produksjonen, og undersøker om komponentene er feilfrie. La X være en stokastisk variabel som angir antall feilfrie komponenter.

b) Hvilken fordeling har X? Begrunn svaret.

Ledelsen i bedriften har innført noen endringer i produksjonsprosessen og håper at det har ført til en økt andel feilfrie komponenter. Kall denne ukjente andelen av feilfrie komponenter for p. Vi trekker et tilfeldig utvalg på n komponenter fra den nye produksjonsprosessen og lar X være antall feilfrie komponenter.

En intuitiv estimator for p er andelen feilfrie komponenter i utvalget, dvs. $\hat{P} = \frac{X}{n}$. Når vi har observert $X = x$ feilfrie komponenter kan vi regne ut et estimat for p som $\hat{p} = \frac{x}{n}$. Det tilfeldige utvalget av størrelse n er så stort at vi kan anta at $\frac{X-np}{\sqrt{n\hat{p}(1-\hat{p})}}$ er tilnærmet standard normalfordelt.

c) Utled et 90% konfidensintervall for p.

Regn ut konfidensintervallet når $n = 500$ og $x = 470$.

Gi en kort tolkning av intervallet.

Ukas forelesninger er ferdige!

Oppgaver

STACK 9: Hele
INNL. 4: Hele

Torsdag

Statistikklab
8:15 - 12:00 S6

Fredag

Statistikklab
12:15 - 16:00 S6

INNL. **Frist 18:00**

STACK **Frist 18:00**

Torsdag & fredag