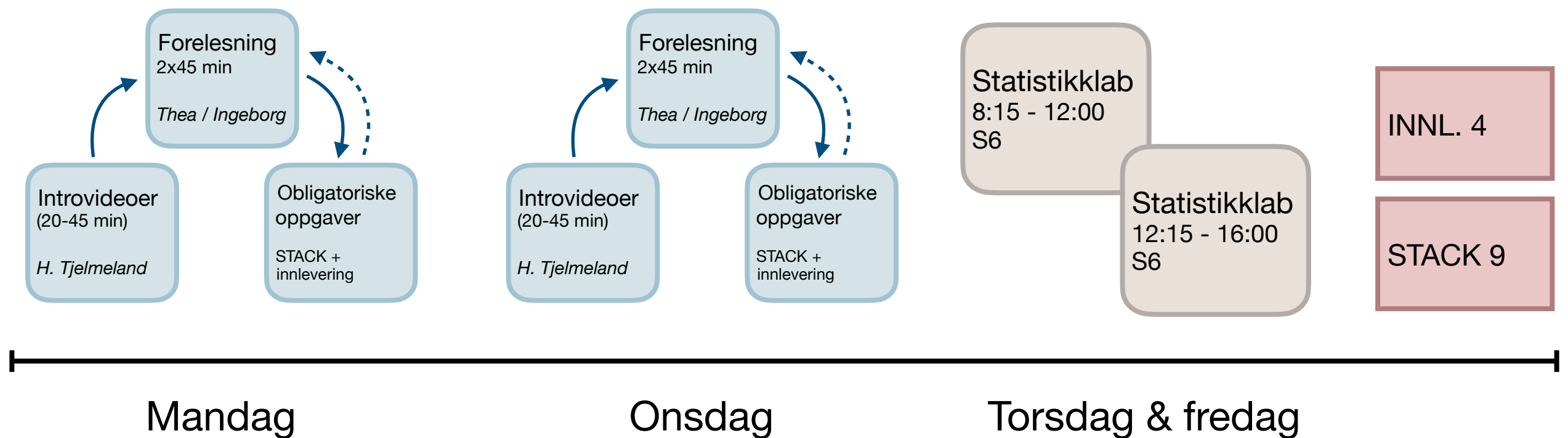


TMA4245 Statistikk

Uke 10 - mandag



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



Introvideoer

Konfidensintervall

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland

Institutt for matematiske fag

Norges teknisk-naturvitenskapelige universitet

Utgangspunkt for konfidensintervall i en del situasjoner

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland

Institutt for matematiske fag

Norges teknisk-naturvitenskapelige universitet

Oppgaver

STACK 9: Oppg. 1-3

INNLL. 4: Oppg. 4ab, 5

Definisjon (Konfidensintervall)

Anta at vi har stokastiske variabler $X_1, X_2, \dots, X_n$, og at sannsynlighetsfordelingen til disse inneholder en parameter θ og at verdien til θ er ukjent. La $x_1, x_2, \dots, x_n$ være observerte verdier for de stokastiske variabler $X_1, X_2, \dots, X_n$. Anta videre at for to observatorer $\hat{\theta}_L(X_1, X_2, \dots, X_n)$ og $\hat{\theta}_U(X_1, X_2, \dots, X_n)$ har at

$$P\left(\hat{\theta}_L(X_1, X_2, \dots, X_n) \leq \theta \leq \hat{\theta}_U(X_1, X_2, \dots, X_n)\right) = 1 - \alpha,$$

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

$$\left[\hat{\theta}_L(x_1, x_2, \dots, x_n), \hat{\theta}_U(x_1, x_2, \dots, x_n)\right]$$

kalles da et $(1 - \alpha) \cdot 100\%$ -konfidensintervall for θ

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

1. Bestemme pivotal, $Z = h(\theta, X_1, X_2, \dots, X_n)$
 - 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

$$P\left(z_{1-\frac{\alpha}{2}} \leq h(\theta, x_1, x_2, \dots, x_n) \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

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

3. Løse hver ulikhet med hensyn på θ
4. Sette ulikhetene sammen igjen med θ i midten

$$P\left(\hat{\theta}_L(X_1, X_2, \dots, X_n) \leq \theta \leq \hat{\theta}_U(X_1, X_2, \dots, X_n)\right) = 1 - \alpha$$

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

$$\left[\hat{\theta}_L(x_1, x_2, \dots, x_n), \hat{\theta}_U(x_1, x_2, \dots, x_n)\right]$$

Oppgave 1

Kritiske verdier i
standard
normalfordelingen

$$P(Z > z_{\alpha}) = \alpha$$

I Norge er promillegrensen for bilkjøring 0.2. I en kontroll ble en bilfører målt med et måleapparat. Målingene kan antas normalfordelte med μ = førerens faktiske promille og $\sigma = 0.044$.

- a) Finn sannsynligheten for at promillemålingen hans havner i intervallet $[0.13, 0.31]$, dersom vi antar at mannens faktiske promille er 0.2.
- b) Politiet måler promillen til 0.22. Utled og regn ut et 95%-konfidensintervall for den faktiske promillen.
- c) Politiet måler på nytt, men nå måler de fem ganger, og får et gjennomsnitt på 0.22. Regn ut et 95%-konfidensintervall for den faktiske promillen.

α	z_{α}
.2	0.842
.15	1.036
.1	1.282
.075	1.440
.05	1.645
.04	1.751
.03	1.881
.025	1.960
.02	2.054
.01	2.326
.005	2.576
.001	3.090
.0005	3.291
.0001	3.719
.00005	3.891
.00001	4.265
.000005	4.417
.000001	4.753

Oppgave 2 Huskjøp

Gustav og Margrethe er nyutdannede sivilingeniører fra NTNU og er nå på boligjakt i Trondheim. Begge ser etter en leilighet i en bestemt bydel.

Vi antar at prisen per kvadratmeter (kvadratprisen) for denne bydelen er normalfordelt.

- c) Gustav og Margrethe har samlet inn data $(x_1, x_2, \dots, x_n)$ for kvadratpris (i kkr) fra de siste $n = 15$ boligsalgene i bydelen, og ønsker basert på disse å finne et 95% konfidensintervall for forventet kvadratpris. Utled et uttrykk for konfidensintervallet (du kan ta utgangspunkt i en kjent observator). Regn ut konfidensintervallet numerisk når gjennomsnittet av kvadratprisene er $\bar{x} = 32$ kkr og $\sum_{i=1}^n (x_i - \bar{x})^2 = 74.1$.

Spesielt gir dette at dersom $X_1, \dots, X_n$ er uavhengige og normalfordelte med forventning μ og varians σ^2 har vi at

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \text{ er } t\text{-fordelt med } n - 1 \text{ frihetsgrader,}$$

$$\text{der } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Kritiske verdier i t -fordelingen

$$P(T > t_{\alpha, \nu}) = \alpha$$

$\nu \backslash \alpha$	.150	.100	.075	.050	.025	.010	.005	.001	.0005
1	1.963	3.078	4.165	6.314	12.706	31.821	63.657	318.309	636.619
2	1.386	1.886	2.282	2.920	4.303	6.965	9.925	22.327	31.599
3	1.250	1.638	1.924	2.353	3.182	4.541	5.841	10.215	12.924
4	1.190	1.533	1.778	2.132	2.776	3.747	4.604	7.173	8.610
5	1.156	1.476	1.699	2.015	2.571	3.365	4.032	5.893	6.869
6	1.134	1.440	1.650	1.943	2.447	3.143	3.707	5.208	5.959
7	1.119	1.415	1.617	1.895	2.365	2.998	3.499	4.785	5.408
8	1.108	1.397	1.592	1.860	2.306	2.896	3.355	4.501	5.041
9	1.100	1.383	1.574	1.833	2.262	2.821	3.250	4.297	4.781
10	1.093	1.372	1.559	1.812	2.228	2.764	3.169	4.144	4.587
11	1.088	1.363	1.548	1.796	2.201	2.718	3.106	4.025	4.437
12	1.083	1.356	1.538	1.782	2.179	2.681	3.055	3.930	4.318
13	1.079	1.350	1.530	1.771	2.160	2.650	3.012	3.852	4.221
14	1.076	1.345	1.523	1.761	2.145	2.624	2.977	3.787	4.140
15	1.074	1.341	1.517	1.753	2.131	2.602	2.947	3.733	4.073
16	1.071	1.337	1.512	1.746	2.120	2.583	2.921	3.686	4.015

Oppgave 3 Levetid

Anta at levetiden T , målt i timer (behøver ikke være et heltall), til en tilfeldig valgt elektronisk komponent er eksponentialfordelt med sannsynlighetstetthet

$$f(t; \beta) = \begin{cases} \frac{1}{\beta} e^{-t/\beta} & t \geq 0 \\ 0 & t < 0, \end{cases}$$

$$f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad x \geq 0.$$

$$E(X) = \beta, \quad \text{Var}(X) = \beta^2,$$

der $\beta > 0$ er en parameter.

La $T_1, T_2, \dots, T_n$ være uavhengige observasjoner av levetiden til n elektroniske komponenter.

- c) Grunngi kort hvorfor $\frac{2}{\beta} \sum_{i=1}^n T_i$ er khikvadratfordelt med $2n$ frihetsgrader. (Vink: du kan bruke resultatene fra oppgave b) uten bevis.)

Bruk dette til å utlede et $(1 - \alpha)100$ %-konfidensintervall for β .

Hva blir 95 %-konfidensintervallet dersom $n = 20$ og $\sum_{i=1}^{20} t_i = 30$?

Vis at $\frac{2}{\beta} T$ er khikvadratfordelt med 2 frihetsgrader. (Vink: du kan bruke at $\Gamma(1) = 1$ i formelen i formelsamlingen for sannsynlighetstettheten til en khikvadratfordelt variabel.)

Resultat fra 2b)

Teorem: Sum av uavhengige kjikvadratfordelte variabler

Anta at $X_1, X_2, \dots, X_n$ er uavhengige stokastiske variabler og, for $i = 1, 2, \dots, n$ anta at X_i er kjivadratfordelt med ν_i frihetsgrader. La en ny stokastisk variabel Y være gitt ved

$$Y = \sum_{i=1}^n X_i.$$

Da er Y kjikvadratfordelt med $\sum_{i=1}^n \nu_i$ frihetsgrader.

Fra temasidene. Bevises med f.eks. momentgenererende funksjoner.

Definisjon: Kjivadratfordelingen

La $Z_1, Z_2, \dots, Z_\nu$ være $\nu \in \{1, 2, \dots\}$ uavhengige og standard normalfordelte variabler, og la

$$X = Z_1^2 + Z_2^2 + \dots + Z_\nu^2 = \sum_{i=1}^{\nu} Z_i^2.$$

Vi sier da at X er kjikvadratfordelt med ν frihetsgrader, alternativt at X har en kjikvadratfordeling med ν frihetsgrader.

Det følger også naturlig fra definisjonen:

$$\sum_{i=1}^1 Z_i^2 \sim \chi_1^2 \text{ (sum av én } \chi^2\text{-variabel)}$$

$$\sum_{i=1}^2 Z_i^2 \sim \chi_2^2 \text{ (sum av to } \chi^2\text{-variabel)}$$

Oppgave 3 Levetid

Anta at levetiden T , målt i timer (behøver ikke være et heltall), til en tilfeldig valgt elektronisk komponent er eksponentialfordelt med sannsynlighetstetthet

$$f(t; \beta) = \begin{cases} \frac{1}{\beta} e^{-t/\beta} & t \geq 0 \\ 0 & t < 0, \end{cases} \quad (1)$$

der $\beta > 0$ er en parameter.

La $T_1, T_2, \dots, T_n$ være uavhengige observasjoner av levetiden til n elektroniske komponenter.

- c) Grunngi kort hvorfor $\frac{2}{\beta} \sum_{i=1}^n T_i$ er khikvadratfordelt med $2n$ frihetsgrader. (Vink: du kan bruke resultatene fra oppgave b) uten bevis.)

Bruk dette til å utlede et $(1 - \alpha)100$ %-konfidensintervall for β .

Hva blir 95 %-konfidensintervallet dersom $n = 20$ og $\sum_{i=1}^{20} t_i = 30$?

Kritiske verdier i χ^2 -fordelingen

$$P(\chi^2 > \chi_{\alpha, \nu}^2) = \alpha$$

$\nu \backslash \alpha$	.995	.990	.975	.950	.050	.025	.010	.005
29	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	90.531	95.023	100.425	104.215

