

Statistikk for ingeniører

Digital plenumstime - uke 7 (04.10.2021)

Mette Langaas, Institutt for matematiske fag, NTNU

Fellesmodul

7-9: Statistisk inferens

Uke 7-8: Estimering og
hypotesetesting

Uke 9: Lineær
regresjon

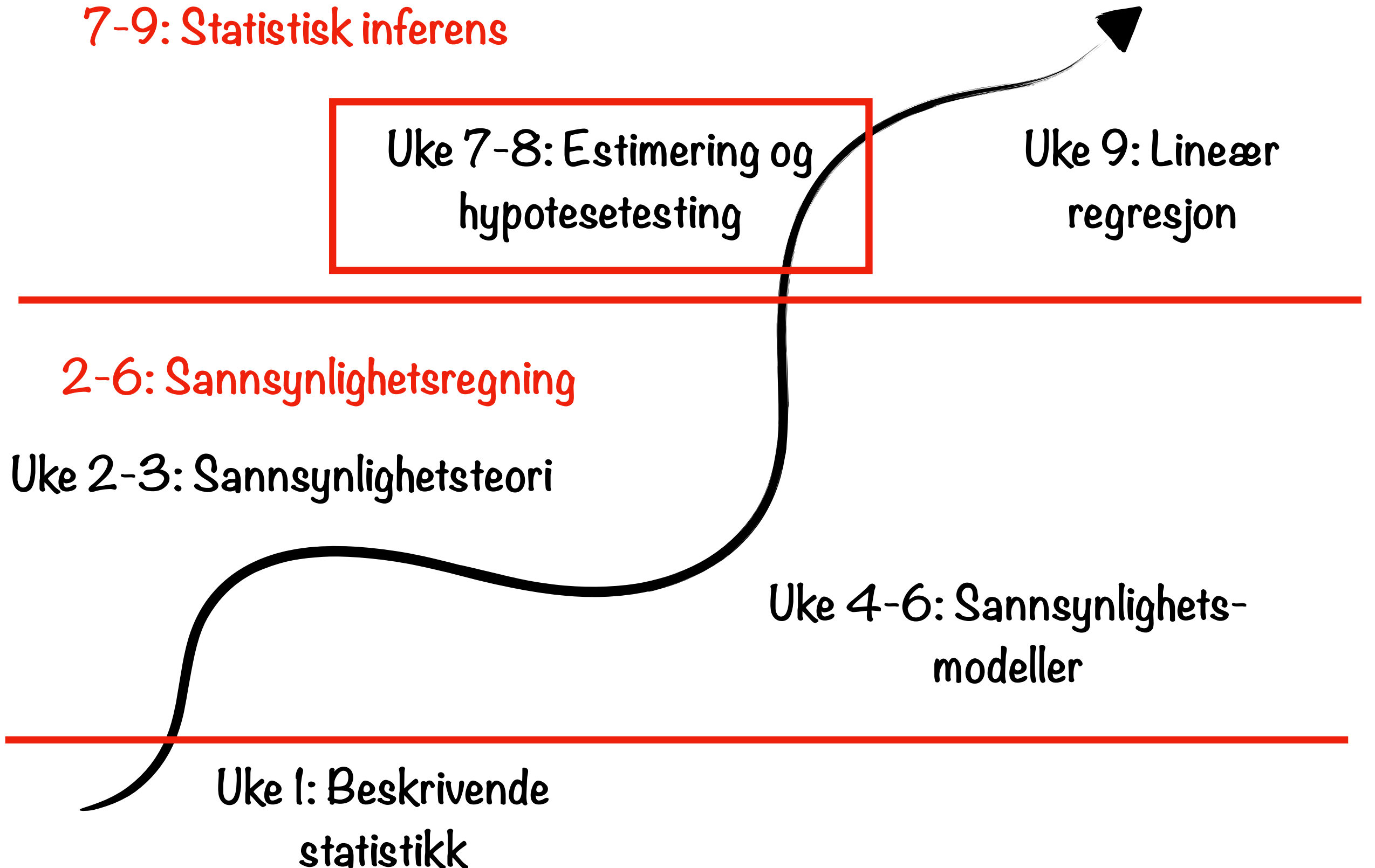
Uke 10: Oppsummering
og skoleeksamen

2-6: Sannsynlighetsregning

Uke 2-3: Sannsynlighetsteori

Uke 4-6: Sannsynlighets-
modeller

Uke 1: Beskrivende
statistikk



Plan for timen

Uke 6: Normal-, standard normalfordeling og sentralgrenseteoremet



Uke 7: Punkt- og intervallestimering



Prosjektmodulen: starte med å finne hvem du skal jobbe sammen med



Prosjektmodulen: 5 minutters intro til hver av de 3 modulene

Eksempler fra eksamensoppgaver

Litt om uka

Blackboard

Informasjon

Informasjon om emnet

Arbeidskrav, pensum og
ressurserFagteam og
referansegruppe

Digital plenumstime

ISTA: Campus Ålesund

ISTG: Campus Gjøvik

ISTT1001/3: Campus
TrondheimISTT1002: Campus
Trondheim

Undervisning og øvinger

Fellesmodul

STACK-øvinger

Python/Jupyter

Digitalt forum

Eksamen

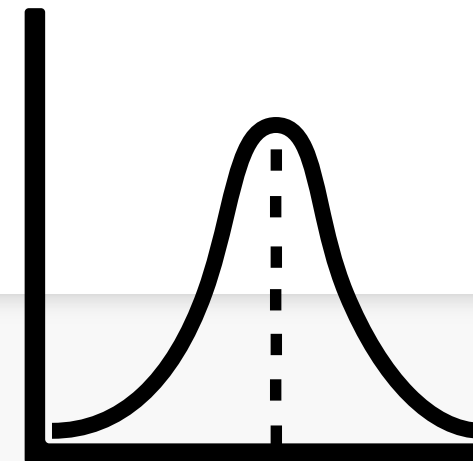
Informasjon om eksamen

Formelark og tabeller

Tidligere
eksamensoppgaverPrøveeksamen og
eksamensforberedelser

Uke 6

Uke 6: Normalfordeling, standard normalfordeling og sentralgrenseteoremet



Introduksjon

Normalfordelingen er kanskje den viktigste av alle sannsynlighetsfordelingene fordi det er mye som i praksis kan antas å være *tilnærmet* normalfordelt. Faktisk kan både binomisk fordeling og poissonfordelingen i noen tilfeller tilnærmes med en normalfordeling.

Normalfordelingen er også veldig fleksibel. Både lineærtransformasjoner ($aX + b$) av normalfordelte variabler, og summen av uavhengige normalfordelte variabler ($aX + bY$) blir også normalfordelte.

Desverre er det ikke helt rett frem å regne på sannsynligheter i normalfordeling, fordi dette må gjøres med numerisk integrasjon. Hvis vi bruker programvare slik som python trenger vi ikke å bry oss om det - python integrerer for oss.

Hvis vi skal løse problemer uten tilgang til en datamaskin så trenger vi tabeller. Heldigvis trenger vi bare tabeller for det som kalles *standard normalfordeling*, fordi alle oppgaver med normalfordeling kan "oversettes" til et uttrykk basert på standard normalfordelingen. Tabeller finner du i fanen Formelark og tabeller her i Blackboard.

Ukas aktiviteter:

1. Du har *muligheten* for å starte uka med å følge den digitale plenumsforelesningen som zoom-webinar på mandag 27.september kl 14.15-15.00 (se info under - samme lenke hele semestret).
2. I løpet av uka skal du se fire temavideoer. Disse finner du i Panopto, se lenke til mappe under.
3. Ukas øving i STACK: stack.math.ntnu.no
4. Denne uke er det en Jupyter-notatbøker du trenger for å gjøre øvingen. Den finner du under Uke 6 på s.ntnu.no/isthub.
5. Du får hjelp med øvingen i øvingsveiledningen (se din tid under din campus) eller ved å stille spørsmål i forumet <https://mattelab.math.ntnu.no/> under kategorien ISTx100y og så underkategori for øving 6.
6. Campusforelesning er på fredag 01.10 kl 12.15-14.00 (se sted under din campus). Husk at du må se på temavideoene (og/eller lese de relevante kapitlene i læreboka) før du kommer i campusforelesningen.
7. Husk at innleveringsfristen for STACK øving 5 er fredag 01.10 kl 23.59 - og at du må ha minst 8 av 10 poeng for å få øvingen godkjent, fristen for øving 6 er 08.10 kl 23.59.



Digital plenumstime

Zoom-webinar mandag 27.september kl 14.15-15.00.

Samme zoom webinar lenke i hele fellesdelen: <https://NTNU.zoom.us/j/99568542340?pwd=aGU2MUJFbZzE0VFJ0b2ZlUT2flZU9nZz09>

Over til Mentimeter

Hvilken fordeling?

Røde: ikke normalfordelt

Antall frø som spirer

Antall mål for hjemmelaget i en fotballkamp

Høyden til en tilfeldig valgt person fra en populasjon

IQen til en tilfeldig valgt person fra en populasjon

Levetiden til en lyspære

Vekten av en aluminiumsplate

Hvilken fordeling?

Røde: ikke normalfordelt

Diskret eller kontinuert?

Antall frø som spirer

D

Antall mål for hjemmelaget i en fotballkamp

D

Høyden til en tilfeldig valgt person fra en populasjon

K

IQen til en tilfeldig valgt person fra en populasjon

K

Levetiden til en lyspære

K

Vekten av en aluminiumsplate

K



Diskret uniform

Kontinuerlig uniform (a, b)

Normal $N(\mu, \sigma)$

Poisson (λt)

Poisson-prosess

Standard normal $N(0, 1)$

Eksponential (λ)

t (ν)

Weibull (λ, α)

Eksempler på normalfordelte stokastisk variabel

Målefeil for et vanlig måleinstrument (mer i prosjektmodulen 1001 “Usikkerhet og støy i målinger”)

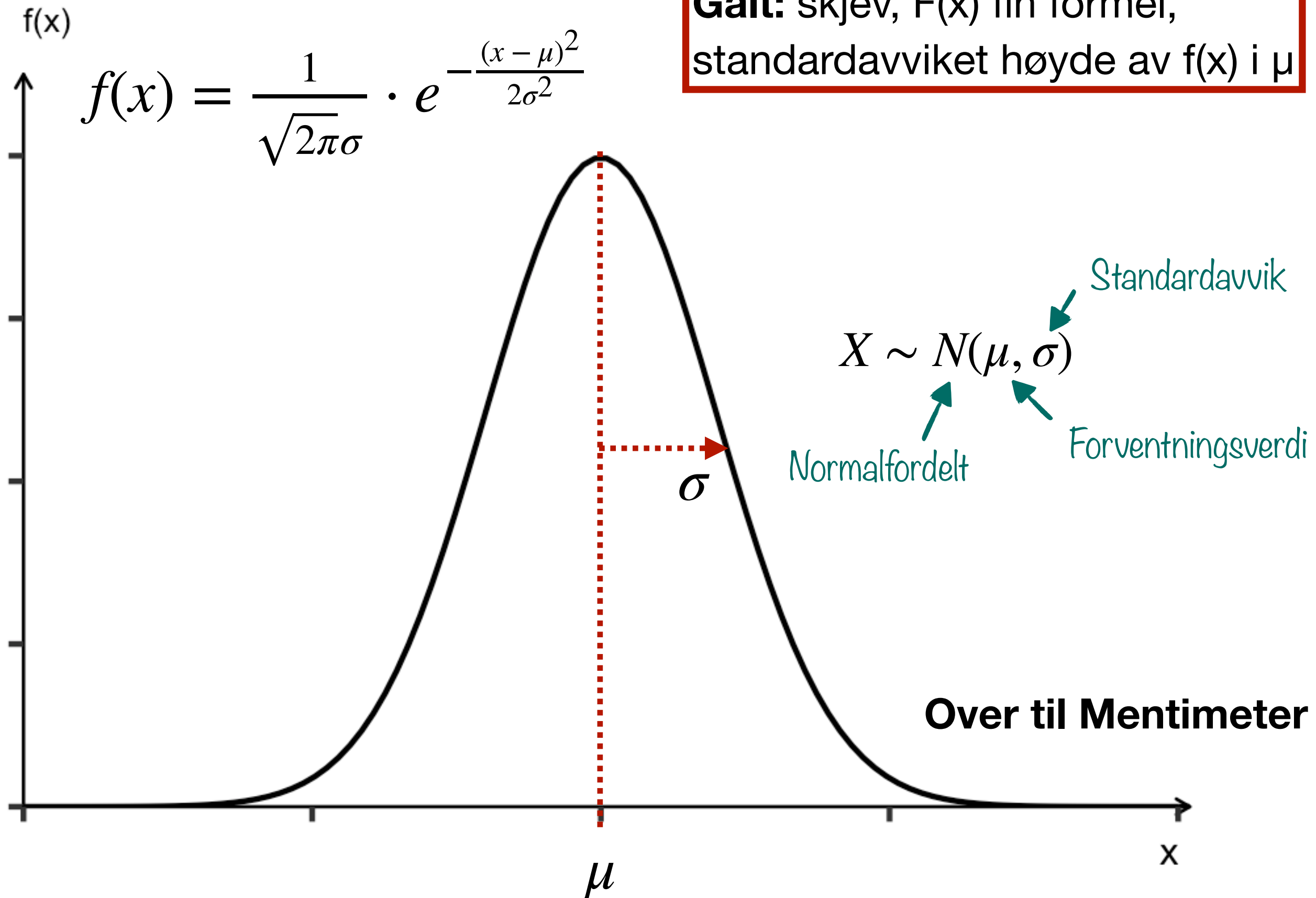
Kontrollvariabler i en produksjonsprosess (mer i prosjektmodulen 1002 “Industriell statistikk”)

Boligpris til leilighet i en by (mer i prosjektmodulen 1003 “Statistisk læring og data science”)

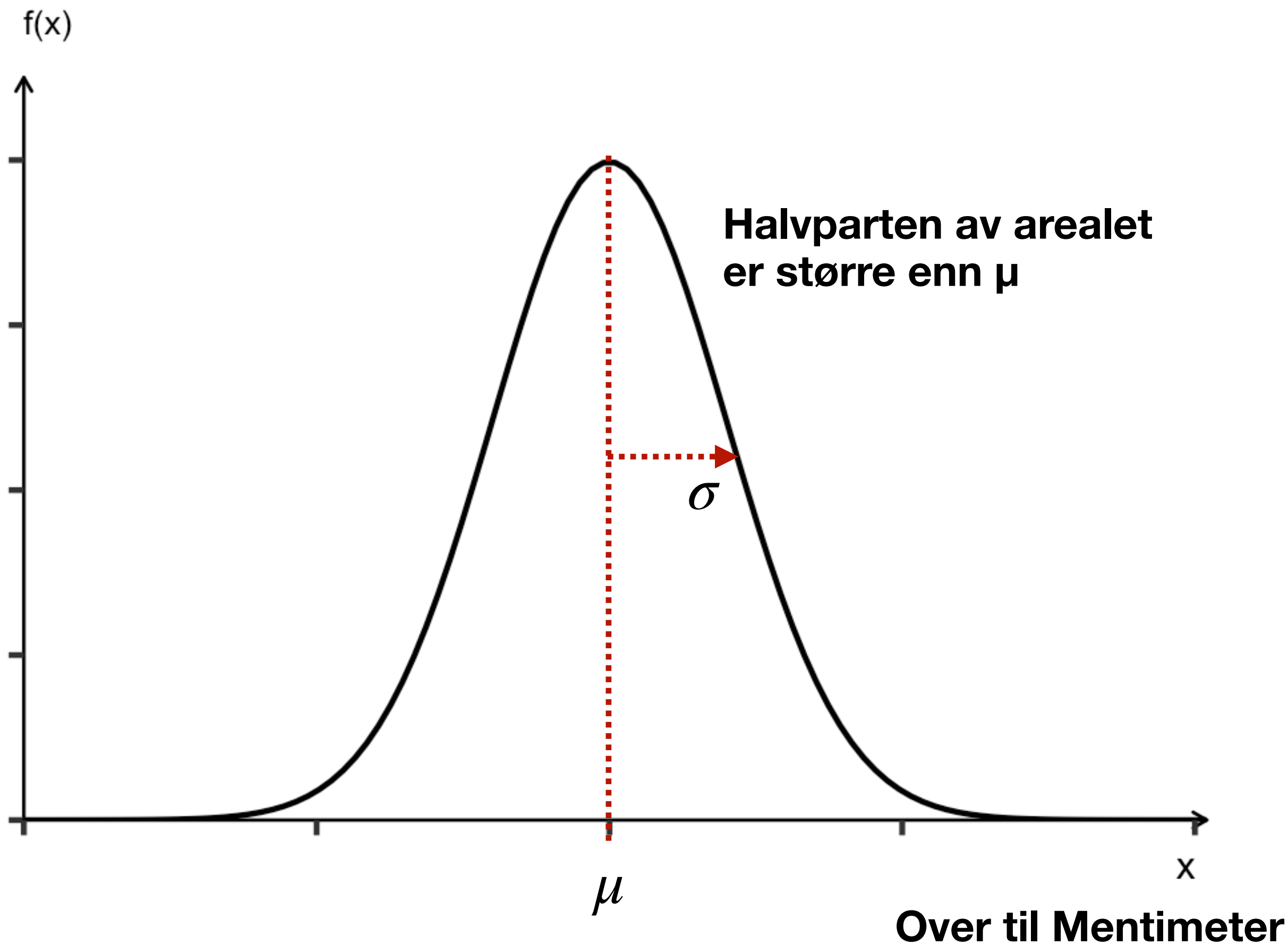
Høyden til en tilfeldig valgt person

Variabler som er en sum av mange delvariabler

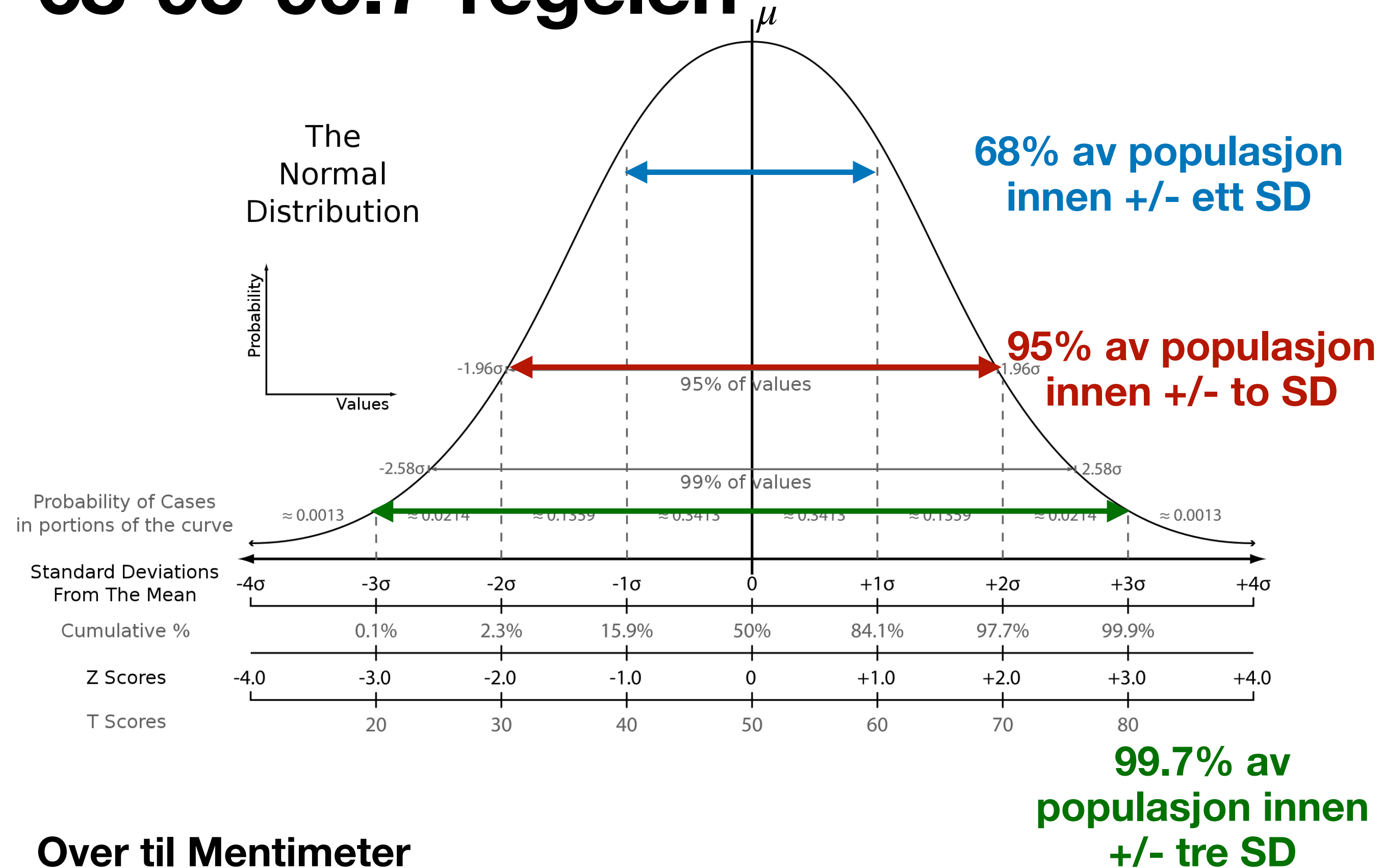
Galt: skjev, F(x) fin formel,
standardavviket høyde av f(x) i μ



F(x) krever numerisk integrasjon: bruk tabell eller Python



68-95-99.7-regelen



Over til Mentimeter

https://commons.wikimedia.org/wiki/File:The_Normal_Distribution.svg

Standardisering

Hvis X har $E(X) = \mu$ og $SD(X) = \sigma$ så vil Z ha $E(Z) = 0$ og $SD(Z) = 1$ dersom:

$$Z = \frac{X - \mu}{\sigma}$$

Fordi:

$$E(Z) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma}E(X) - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$$

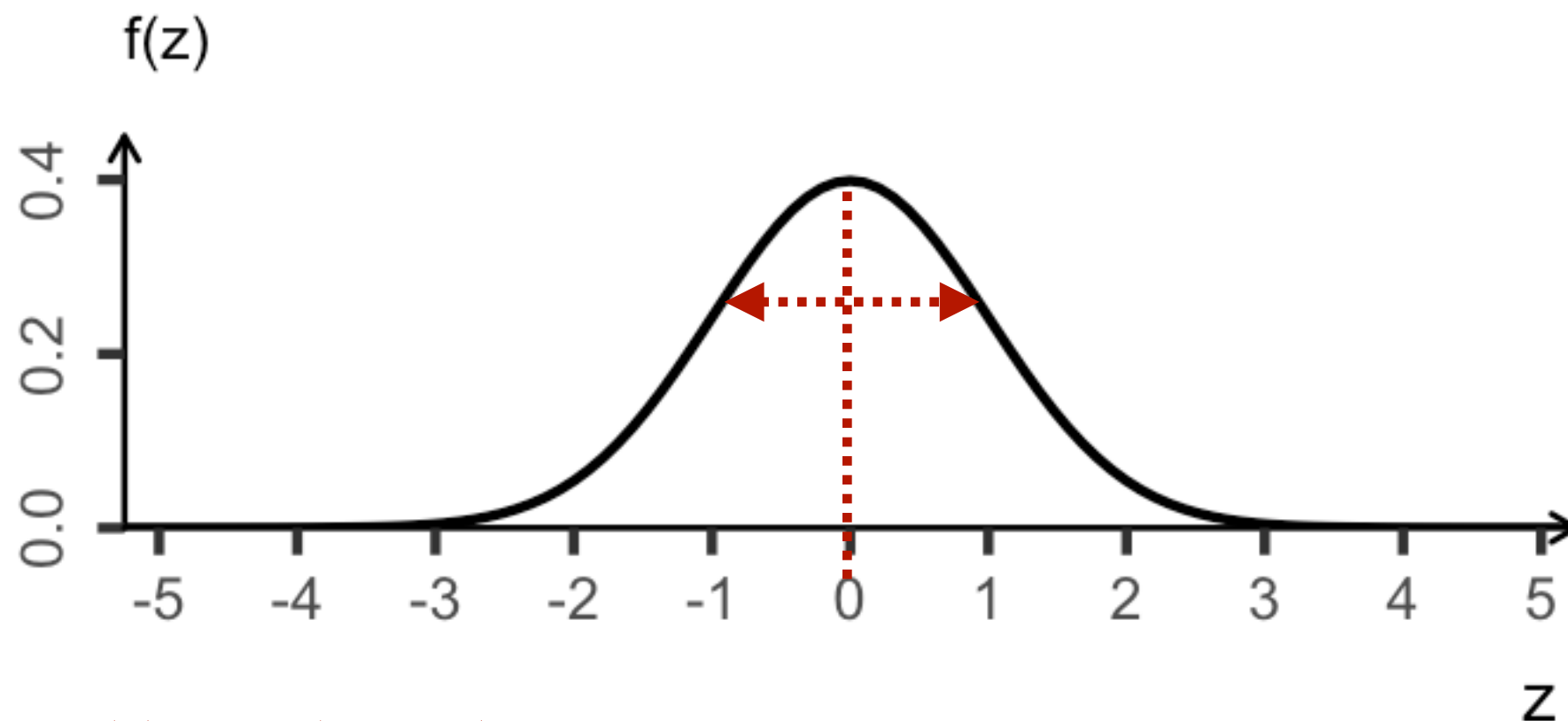
$$Var(Z) = Var\left(\frac{X - \mu}{\sigma}\right) = \left(\frac{1}{\sigma}\right)^2 Var(X) + 0 = \left(\frac{1}{\sigma}\right)^2 \sigma^2 = 1$$

$$SD(Z) = 1$$

Hvis X er normalfordelt vil

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1) \text{ være standard normalfordelt}$$

$$Z \sim N(0, 1) \quad f(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}$$



$\Phi(z) = P(Z \leq z)$ finnes i tabell

ISTx100yH20

4 Oppgave 2 [30%]

Instruksjoner: Oppgave 2 består av deloppgaver a, b og c, som besvares direkte i Inspira.

c) i. La X være en normalfordelt stokastisk variabel med forventningsverdi 8 og standardavvik 4, altså $X \sim N(8, 4)$. Hva er $P(X \leq 6)$?

Velg ett alternativ:

☐ 0.309

☐ 0.369

☐ 0.227

☐ 0.212

☐ 0.345

☐ 0.401

☐ 0.421

☐ 0.252

Over til Mentimeter

La $X_1, X_2, \dots, X_n$ være n **uavhengige** stokastiske variabler der

$$E(X_i) = \mu \quad \text{og} \quad Var(X_i) = \sigma^2 \quad \text{for alle } i = 1, \dots, n$$

Gjennomsnitt:

$$Y = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

$$E(Y) = \frac{1}{n}(\mu + \mu + \dots + \mu) = \mu$$

$$Var(Y) = \left(\frac{1}{n}\right)^2 (\sigma^2 + \sigma^2 + \dots + \sigma^2) = \left(\frac{1}{n}\right)^2 n\sigma^2 = \frac{\sigma^2}{n}$$

$$SD(Y) = \frac{\sigma}{\sqrt{n}}$$

Normalfordeling

$$X \sim N(\mu, \sigma)$$

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \text{for } -\infty < x < \infty$$

$$F(x) = P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

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

Standard normalfordeling

$$Z \sim N(0, 1)$$

$$f(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad \text{for } -\infty < z < \infty$$

$$F(z) = P(Z \leq z) = \Phi(z)$$

$$E(Z) = 0, \quad \text{Var}(Z) = 1$$

4.1 Regneregler normalfordeling

Vi ser på n uafhængige stokastiske variable $X_1, X_2, \dots, X_n$ slik at $X_i \sim N(\mu, \sigma)$, for $i = 1, \dots, n$.
Da er

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

og

$$X_1 + X_2 + \dots + X_n \sim N(n\mu, \sqrt{n} \cdot \sigma)$$

4.1.1 Normaltilnærminger

$\text{Binom}(n, p) \approx N(np, \sqrt{np(1-p)})$ hvis $np(1-p) \geq 5$,
 $\text{Poisson}(\lambda t) \approx N(\lambda t, \sqrt{\lambda t})$ hvis $\lambda t > 10$

4.1.2 Sentralgrenseteoremet

Dersom $X_1, X_2, \dots, X_n$ er uafhængige stokastiske variable med forventning $E(X_i) = \mu$ og varians $\text{Var}(X_i) = \sigma^2$, for $i = 1, \dots, n$, og dersom $n > 30$, så er

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \text{ tilnærmet } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

og

$$X_1 + X_2 + \dots + X_n \text{ tilnærmet } N(n\mu, \sqrt{n} \cdot \sigma)$$

ii. La $\bar{X}$ være gjennomsnittet av ni uavhengige stokastiske variabler $X_1, X_2, \dots, X_9$, der $X_i \sim N(8, 4)$ for $i = 1, \dots, 9$. Hva er $P(\bar{X} \leq 6)$?

Velg ett alternativ

- ☐ 0.023
- ☐ 0.012
- ☐ 0.115
- ☐ 0.227
- ☐ 0.274
- ☐ 0.067
- ☐ 0.008
- ☐ 0.159

Over til Mentimeter

Sentralgrenseteoremet

La $X_1, X_2, \dots, X_n$ være n **uavhengige** ~~normalfordelte~~ stokastiske variabler der

$$n > 30 \quad E(X_i) = \mu \quad \text{og} \quad \text{Var}(X_i) = \sigma^2 \quad \text{for alle } i = 1, \dots, n$$

Sum:

$$Y = X_1 + X_2 + \dots + X_n \quad E(Y) = n\mu \quad \text{Var}(Y) = n\sigma^2$$
$$SD(Y) = \sqrt{n}\sigma$$

$$Y \sim N(n\mu, \sqrt{n}\sigma)$$

Tilnærmet

“Tilsvarende” resultat for gjennomsnittet

Normalfordeling

$$X \sim N(\mu, \sigma)$$

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \text{for } -\infty < x < \infty$$

$$F(x) = P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

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

Standard normalfordeling

$$Z \sim N(0, 1)$$

$$f(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad \text{for } -\infty < z < \infty$$

$$F(z) = P(Z \leq z) = \Phi(z)$$

$$E(Z) = 0, \quad \text{Var}(Z) = 1$$

4.1 Regneregler normalfordeling

Vi ser på n uafhængige stokastiske variable $X_1, X_2, \dots, X_n$ slik at $X_i \sim N(\mu, \sigma)$, for $i = 1, \dots, n$.
Da er

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

og

$$X_1 + X_2 + \dots + X_n \sim N(n\mu, \sqrt{n} \cdot \sigma)$$

4.1.1 Normaltilnærminger

$\text{Binom}(n, p) \approx N(np, \sqrt{np(1-p)})$ hvis $np(1-p) \geq 5$,
 $\text{Poisson}(\lambda t) \approx N(\lambda t, \sqrt{\lambda t})$ hvis $\lambda t > 10$

4.1.2 Sentralgrenseteoremet

Dersom $X_1, X_2, \dots, X_n$ er uafhængige stokastiske variable med forventning $E(X_i) = \mu$ og varians $\text{Var}(X_i) = \sigma^2$, for $i = 1, \dots, n$, og dersom $n > 30$, så er

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \text{ tilnærmet } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

og

$$X_1 + X_2 + \dots + X_n \text{ tilnærmet } N(n\mu, \sqrt{n} \cdot \sigma)$$

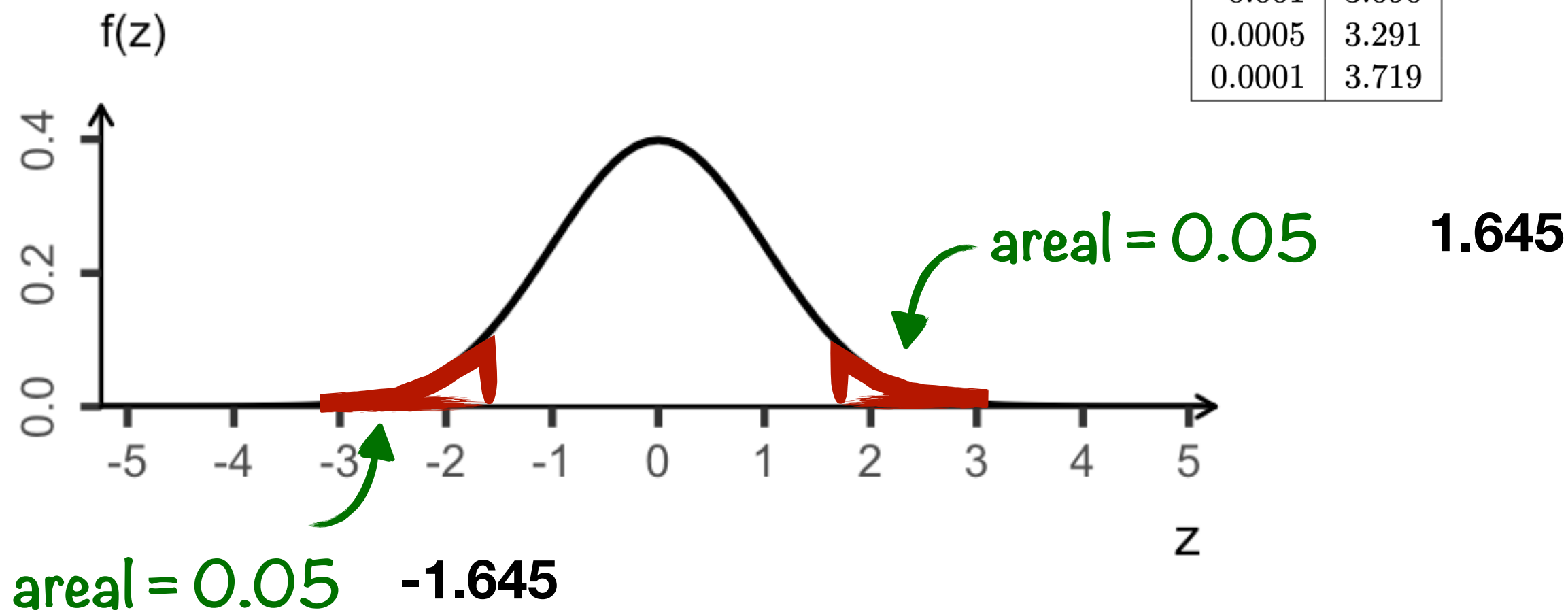
Over til Mentimeter

Hvilket tall i standard normalfordelingen har areal 5% til venstre for seg?

Kritiske verdier i standard normalfordelingen

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

α	z_{α}
0.20	0.842
0.10	1.282
0.05	1.645
0.025	1.960
0.02	2.054
0.01	2.326
0.005	2.576
0.001	3.090
0.0005	3.291
0.0001	3.719



Uke 7: Punkt- og intervallestimering

Punktestimering i normalfordelingen

Thea Bjørnland
Institutt for matematiske fag
NTNU

Konfidensintervaller i normalfordelingen (del 1)

Thea Bjørnland
Institutt for matematiske fag
NTNU

Estimering og konfidensintervaller for normalfordelingen

Konfidensintervaller i normalfordelingen (del 2)

Thea Bjørnland
Institutt for matematiske fag
NTNU

t-fordeling

t -fordelingen er en ekstremt spesialisert fordeling. I dette emne dukker behovet for t -fordeling opp når vi har observasjoner som kan antas *normalfordelte*, men med både ukjent

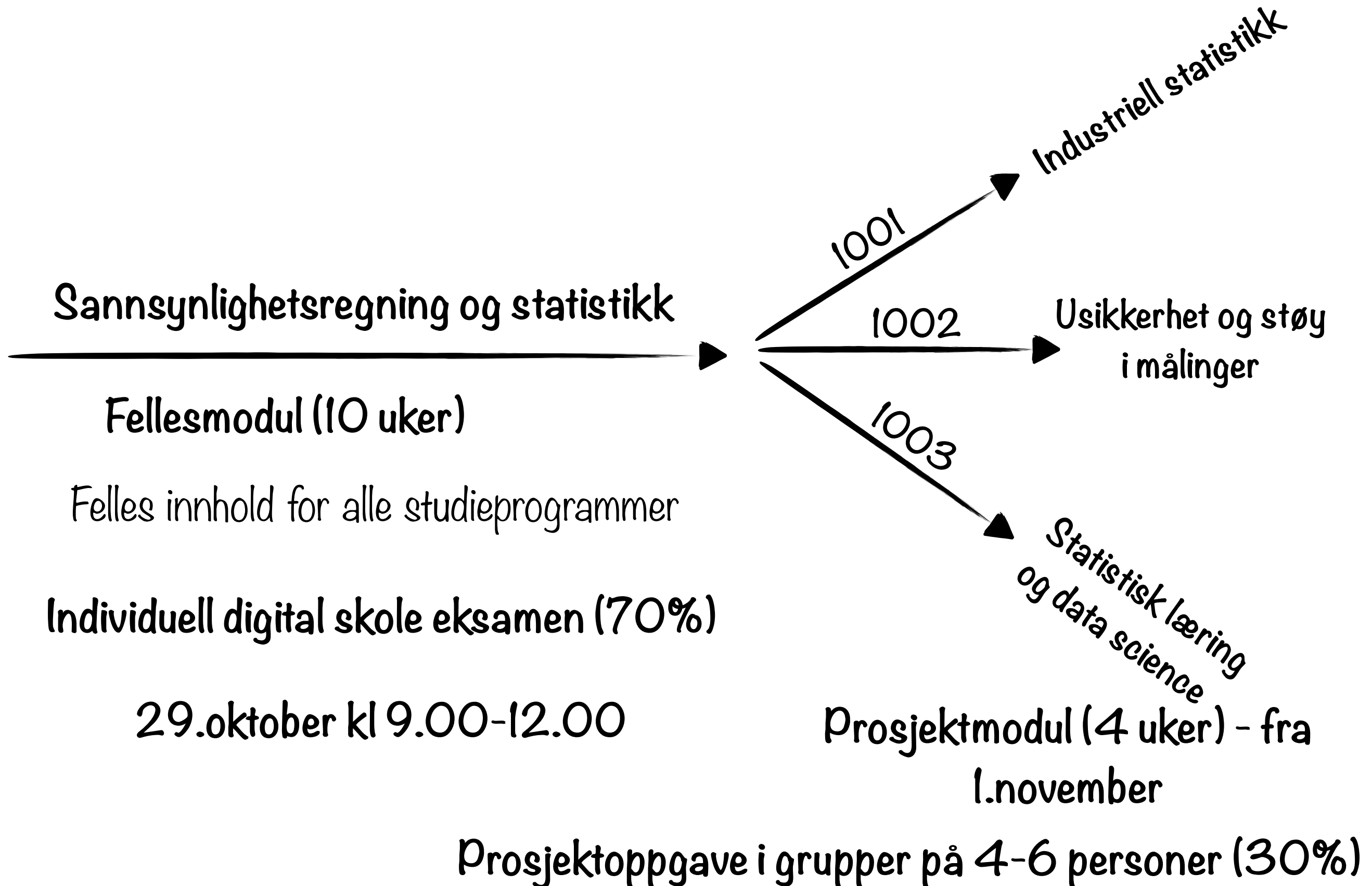
Estimering og konfidensintervaller for sannsynligheten p og raten λ

Thea Bjørnland
Institutt for matematiske fag
NTNU

STACK

Campus-forelesning

ISTx100y Statistikk for ingeniører



Informasjon

Informasjon om emnet

Arbeidskrav, pensum og
ressurser

Referansegruppe og
fagteam

Digital plenumstime

ISTA: Campus Ålesund

ISTG: Campus Gjøvik

ISTT1001/3: Campus
Trondheim

ISTT1002: Campus
Trondheim

Undervisning og øvinger

Fellesmodul

STACK-øvinger

Python/Jupyter

Digitalt forum

Prosjektgrupper:
påmelding og kontrakt

Prosjektgrupper: påmelding og kontrakt



Informasjon om prosjektgrupper og innlevering A↓

- Det må være **4-6 studenter** på hver gruppe, og alle studenter må være oppmeldt i den samme emnekode
- Trenger du **hjelp med å finne gruppemedlemmer** så ta kontakt med ditt studieprogram - se liste over res
- First for å melde seg på grupper er **mandag 8. november!**
- Etter gruppen er etablert skal det skrives en gruppekontrakt - se under for ressurser og mal (kommer).
- Gruppekontrakten leveres på Blackboard innen **fredag 12.november**.
- Hvis noe går **galt under gruppearbeidet** så ta kontakt med ditt studieprogram (se ressurspersoner nede p
- **Innleveringsfrist: Prosjektrapporten skal leveres i Inspira senest mandag 29.november kl 09.00.**
- Se informasjon under om hvordan man lager gruppeinnlevering i Inspira (samme fremgangsmøte som for



Gruppearbeid og gruppekontrakt A↓

Hvordan lykkes med gruppe (team)-arbeid? <https://www.youtube.com/watch?v=4x28G36LSZs>

Forslag til mal for gruppekontrakt (som egen fil under her).

Datoer:

- **1. november: oppstart undervisning prosjektmodul**
- **8. november: frist opprette gruppe**
- **12. november: frist levere inn gruppekontrakt**
- **29. november kl 09.00: frist innlevere prosjektrapport i Inspira**

Husk at alle må melde seg inn i gruppen via sin egen Inspira pålogging før gruppen leverer.

	<u>Prosjektgrupper ISTA1001 Industriell statistikk Ålesund</u>  Tilgjengelighet: Elementet er tilgjengelig, men det kan hende noen studenter er påmelding for prosjektgrupper.
	<u>Prosjektgrupper ISTA1002 Usikkerhet og støy i målinger Ålesund</u> Tilgjengelighet: Elementet er tilgjengelig, men det kan hende noen studenter er påmelding for prosjektgrupper.
	<u>Prosjektgrupper ISTA1003 Statistisk læring og data science Ålesund</u> Tilgjengelighet: Elementet er tilgjengelig, men det kan hende noen studenter er påmelding for prosjektgrupper.
	<u>Prosjektgrupper ISTG1001 Industriell statistikk Gjøvik</u>  Tilgjengelighet: Elementet er tilgjengelig, men det kan hende noen studenter er påmelding for prosjektgrupper.
	<u>Prosjektgrupper ISTG1002 Usikkerhet og støy i målinger Gjøvik</u> Tilgjengelighet: Elementet er tilgjengelig, men det kan hende noen studenter er påmelding for prosjektgrupper.
	<u>Prosjektgrupper ISTG1003 Statistisk læring og data science Gjøvik</u> Tilgjengelighet: Elementet er tilgjengelig, men det kan hende noen studenter er påmelding for prosjektgrupper.
	<u>Prosjektgrupper ISTT1001 Industriell statistikk Trondheim</u> Tilgjengelighet: Elementet er tilgjengelig, men det kan hende noen studenter er påmelding for prosjektgrupper.
	<u>Prosjektgrupper ISTT1002 Usikkerhet og støy i målinger Trondheim</u> Tilgjengelighet: Elementet er tilgjengelig, men det kan hende noen studenter er påmelding for prosjektgrupper.
	<u>Prosjektgrupper ISTT1003 Statistisk læring og data science Trondheim</u> Tilgjengelighet: Elementet er tilgjengelig, men det kan hende noen studenter er påmelding for prosjektgrupper.

	Sammenheng mellom studieprogram og prosjektmodul  Hvilken prosjektmodul du skal ha er bestemt av studieprogrammet ditt. Det kan du også lese fra emnekoden din. ISTx1001 med Industriell statistikk Byggingeniør (BIBYGG, BIBYG-F) Geomatikkingeniør (BIGEOMAT) Havbruksingeniør (BIHAV) Logistikkingeniør (FTHINGLOG) Skipesdesign (699D) ISTx1002 med Usikkerhet og støy i målinger Elektroingeniør (BIELEKTRO) Fornybar energi, ingeniør (BIFOREN) Kjemiingeniør (FTHINGKJ) Maskiningeniør (BIMAS-F), Maskiningeniør (BIMASKIN) Materialteknologi (FTHINGMAT) Olje- og gassteknologi (FTHINGOG) ISTx1003 med Statistisk læring og data science Dataingeniør (BIDATA) Digital infrastruktur og cybersikkerhet (BDIGSEC)
--	---

	Liste over ressurspersoner på studieretningene  Har du ikke noen du kan jobbe sammen med i prosjektoppgaven? Disse ressurspersonene kan hjelpe til med å danne grupper og i gruppeprosessen. Antallet oppmeldte studenter for hvert studieprogram og campus står i parentes. Liste over kontaktpersoner for studieprogram og campus: Skipsdesign (699D) Ålesund (9 oppmeldt): havard.v.lien@ntnu.no BDIGSEC Trondheim (16 oppmeldt): geir.o.rosvold@ntnu.no BDIGSEC Gjøvik (46 oppmeldte): erik.hjelmas@ntnu.no BIBYG-F Gjøvik (10 oppmeldte): tron.h.torneby@ntnu.no BIBYGG Ålesund (56 oppmeldte): robin.satre@ntnu.no BIBYGG Gjøvik (62 oppmeldte): tron.h.torneby@ntnu.no BIBYGG Trondheim (119 oppmeldte): robert.mortensen@ntnu.no BIDATA Ålesund (59 oppmeldte): gist@ntnu.no BIDATA Gjøvik (35 oppmeldte): kiran.raja@ntnu.no BIDATA Trondheim (97 oppmeldte): grethe.sandstrak@ntnu.no BIELEKTRO Ålesund (40 oppmeldte): siebe.b.v.albada@ntnu.no BIELEKTRO Gjøvik (49 oppmeldte): knut.wold@ntnu.no BIELEKTRO Trondheim (139 oppmeldte): Arne.Midjo@ntnu.no BIFOREN Ålesund (23 oppmeldte): ann.r.nerheim@ntnu.no BIFOREN Gjøvik (29 oppmeldte): alemayehu.gebremedhin@ntnu.no BIFOREN Trondheim (69 oppmeldte): jacob.j.lamb@ntnu.no BIGEOMAT Gjøvik (9 oppmeldte): torbjoern.kravdal@ntnu.no BIHAV Trondheim (27 oppmeldte): rolf.e.olsen@ntnu.no BIMAS-F Gjøvik (19 oppmeldte): thomas.haavi@ntnu.no BIMASKIN Ålesund (19 oppmeldte): lars.p.bryne@ntnu.no BIMASKIN Gjøvik (21 oppmeldte): thomas.haavi@ntnu.no BIMASKIN Trondheim (91 oppmeldte): anna.olsen@ntnu.no FTHINGKJ Trondheim (41 oppmeldte): ina.m.stuen@ntnu.no FTHINGLOG Trondheim (48 oppmeldte): alireza.ashrafian@ntnu.no FTHINGMAT Trondheim (33 oppmeldte): roald.lilletvedt@ntnu.no
---	---

**1001: Industriell
statistikk - faglærer
Ketil Arnesen**



**1002: Usikkerhet og
støy i målinger -
faglærer Charles Curry**



**1003: Statistisk læring
og data science -
faglærer Stefanie Muff**



Siste tre mandagstimer

Uke 8 (11.oktober) Hypotesetesting

Prøveeksamen blir tilgjengelig i løpet av uken

Uke 9 (18.oktober) Lineær regresjon

Info og spørsmål rundt eksamen og prosjekt

Andre møte med referansegruppa - innspill til uke 10

Uke 10 (25.oktober) Oppsummering - og forberedt Q&A for faglige spørsmål