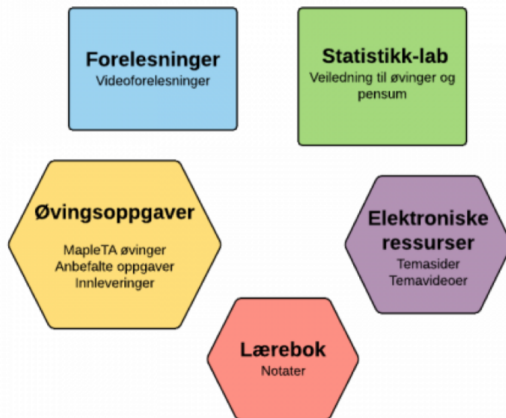


Velkommen til TMA4240

- Praktisk info
- Hva er statistikk
- Hva skal dere lære i dette kurset
- Deskriptiv statistikk

# TMA4240: Tilgjengelig resurser

- Faglærer: Sara Martino og Torstein Fjeldstad
- Hjemmeside:  
<https://wiki.math.ntnu.no/tma4240/2018h/start>



- To Parallell:
  - Mandag-Fredag 10:15-12:00
  - Tirsdag 14:15-16:00, Torsdag 16:15-18:00
- Du kan velge hvilken forelesning du ønsker
- Begge parallel har samme pensum og de er delt ”verticalt” mellom to foreleser
  - 1 Blok: Sara Martino
  - 2 Block: Torstein Fjeldstad

Elektronisk  
øving

Anbefalt  
øving

Skriftlig  
øving

## Statistikklab

Krav til godkjent øvingsopplegg

- Elektronisk (Maple TA) øving
  - Minst 4 godkjente øvinger fra Blokk 1
  - Minst 4 godkjente øvinger fra Blokk 2
- Skriftlige innleveringer
  - 3 godkjente innleveringer, hvorav minst 1 må være fra Blokk 1 og minst 1 fra Blokk 2

Her kan du samarbeide med andre og få hjelp av studentassistenter

Dag	Tid	Rom
Mandag	14:15-18:00	S4
Torsdag	12:15-16:00	S1
Fredag	12:15-16:00	S4 *

(\* ) I uke 36 er det flyttet i S1

- **Walpole, Myers, Myers and Ye (2012): Probability and Statistics for engineers (9. utgave), Pearson / Prentice Hall.**
  - Det finnes en egen utgave for NTNU (ISBN 978-1-78726-081-8) som selges på Akademika.
  - Det finnes en NTNU-versjon av 9. utgave med ISBN 978-1-78448-014-1 som også kan brukes. Originalversjonen (internasjonal versjon) av 9. utgave kan også brukes.
  - Det finnes en utgave med **blå skrift** (ISBN 10:1-292-02392-9. ISBN 13: 978-1-292-02392-2), som **vi fraråder** da flere har gitt tilbakemelding om at den er vanskelig å lese og å finne frem i.
- **Tabeller of formler i statistikk. Akademika**

- Temasider
- Temavideo



- Vi trenger 2/3 studenter som kan være i referansegruppe
  - 2/3 møter i løpet av semester
  - skrive en kort rapport til slutt

# Hva er Statistikk??

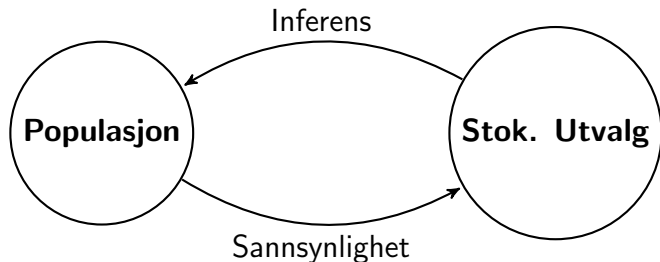
# Hva er Statistikk??

**Statistikk er** en gren av matematikk som tar for seg analyse av tallfestede dataobservasjoner med den hensikt å kunne trekke konklusjoner basert på disse dataene

**Statistikk er** en verktøy for å ta beslutnings under usikkerhet og variasjon.

**Statistikk** handler om å tallfeste vår usikkerhet

# Hva skal vi lære i dette kurset?



- Deskriptiv Statistikk, Kap 1 (Idag)
  - Presentere og beskrive data i tall, tabeller, figurer
- Sannsynlighet, Kap 2-7 (Fra neste gang)
  - regne men sannsynlighet
- Statistikk inferens, Kap 8-11 (Senere)
  - trekke slutninger om en hel populasjon fra observasjoner på et utvalg

**Sannsynlighetseksempel** Du har en rettferdig mynt (like sannsynlighet for hoder eller haler). Hva er sannsynligheten for 60 eller flere hoder hvis du kaster mynt 100 ganger?

**Statistikk eksempel** Du har en ukjent mynt. Du kaster du det 100 ganger og teller 60 hoder. Er mynten rettferdig?

**Sannsynlighetseksempel** Du har en rettferdig mynt (like sannsynlighet for hoder eller haler). Hva er sannsynligheten for 60 eller flere hoder hvis du kaster mynt 100 ganger?

Det er bare ett svar (ca. 0.028444), og vi skal lære å beregne det.

**Statistikk eksempel** Du har en ukjent mynt. Du kaster du det 100 ganger og teller 60 hoder. Er mynten rettferdig?

**Sannsynlighetseksempel** Du har en rettferdig mynt (like sannsynlighet for hoder eller haler). Hva er sannsynligheten for 60 eller flere hoder hvis du kaster mynt 100 ganger?

Det er bare ett svar (ca. 0.028444), og vi skal lære å beregne det.

**Statistikk eksempel** Du har en ukjent mynt. Du kaster du det 100 ganger og teller 60 hoder. Er mynten rettferdig? Det er mange måter å gå videre. Faktisk kan forskjellige statistikere trekke forskjellige konklusjoner.

**Sannsynlighetseksempel** Tilfeldige prosessen fullt kjent (sannsynlighet for hodene = .5).

Målet er å finne sannsynligheten for et bestemt utfall (minst 60 hoder).

**Statistikk eksempel** Resultatet kjent (60 hoder) og målet er å belyse den ukjente tilfeldige prosessen (sannsynligheten for hoder).



Vi skal se på to datasett fra

<http://www.math.hope.edu/swanson/statlabs/data.html>:

- Høyde til 50 mannlige og 50 kvinnlige studenter gitt i meter
- Kroppstemperatur for 65 menn og 65 kvinner gitt i grader Celsius

Vi skal se på to datasett fra

<http://www.math.hope.edu/swanson/statlabs/data.html>:

- Høyde til 50 mannlige og 50 kvinnlige studenter gitt i meter
- Kroppstemperatur for 65 menn og 65 kvinner gitt i grader Celsius

Vi ønsker å finne ut om det er en forskjell mellom høyden og kroppstemperaturen til kvinner og menn

Kvinner: 170.18 170.18 170.18 152.4 172.72 162.56 175.26  
180.34 170.18 170.18 167.64 160.02 170.18 157.48 167.64  
177.8 170.18 154.94 172.72 170.18 172.72 162.56 175.26  
170.18 177.8 182.88 154.94 170.18 175.26 172.72 175.26  
182.88 167.64 170.18 167.64 170.18 175.26 162.56 162.56  
160.02 172.72 167.64 165.1 152.4 177.8 165.1 172.72 167.64  
154.94 165.1

Manner: 182.88 187.96 190.5 180.34 180.34 172.72 190.5  
190.5 187.96 177.8 187.96 187.96 200.66 182.88 182.88  
190.5 182.88 193.04 193.04 187.96 185.42 182.88 165.1  
180.34 187.96 175.26 185.42 175.26 180.34 170.18 187.96  
177.8 177.8 177.8 187.96 187.96 187.96 177.8 187.96 185.42  
175.26 182.88 175.26 187.96 180.34 182.88 187.96 190.01  
195.58 182.88

# Hvor begynner man??

Datasettene er for store til å bare skrive ut tallene for menn og for kvinner og sammenligne dem

# Hvor begynner man??

Datasettene er for store til å bare skrive ut tallene for menn og for kvinner og sammenligne dem

Men vi kan

- Regne ut tall som beskriver sentrene til de observerte verdiene for menn og kvinner
- Regne ut tall som beskriver spredningene til de observerte verdiene for menn og kvinner
- Lage plott som viser hvordan de observerte verdiene er spredt i forhold til hverandre

# Hvor begynner man??

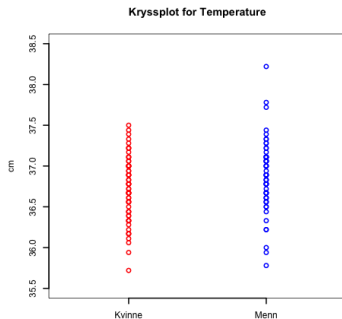
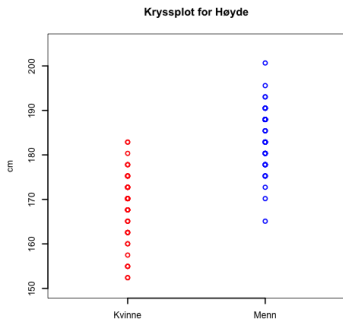
Datasettene er for store til å bare skrive ut tallene for menn og for kvinner og sammenligne dem

Men vi kan

- Regne ut tall som beskriver sentrene til de observerte verdiene for menn og kvinner
- Regne ut tall som beskriver spredningene til de observerte verdiene for menn og kvinner
- Lage plott som viser hvordan de observerte verdiene er spredt i forhold til hverandre

Dette er metoder for å beskrive de observerte verdiene og bruk av disse kalles **deskriptiv statistikk**

Et enkelt startpunkt er å plote verdiene for hvert av kjønnene og sammenligne



## Definition (Empirisk Middelerdi)

Den empirisk middelerdi til observasjonene  $x_1, x_2, \dots, x_n$  er gitt ved

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$



## Definition (Empirisk Middelerdi)

Den empirisk middelerdi til observasjonene  $x_1, x_2, \dots, x_n$  er gitt ved

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

## Definition (Empirisk Median)

Den empirisk median til observasjonene  $x_1, x_2, \dots, x_n$  er gitt ved

$$\tilde{x} = \begin{cases} x_{(n+1)/2}, \\ \frac{1}{2}(x_{(n/2)} + x_{(n/2+1)}) \end{cases}$$

- 1,2,3,4,5
- 1,1,1,2,2,2
- 1,2,3,4,20

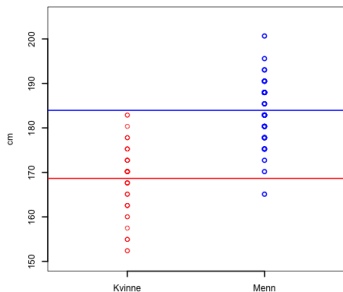
# Sammenligning av senterer

Sentralmål	Høyde		Kroppstemperatur	
	Kvinne	Mann	Kvinne	Mann
Empirisk middelverdi	168.6	183.9	36.7	36.8
Empirisk median	170.1	184.1	36.7	36.8

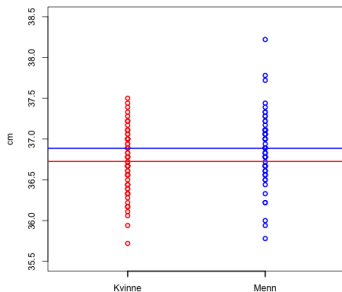
Er dette nok for å si at menn og kvinne har forskjellige høyder og temperaturer?

# Kryssplot med empirisk middelværdi

Kryssplot for Høyde



Kryssplot for Temperature



Vi må også vurdere variabilitetene til de observerte verdiene for å kunne trekke konklusjoner om forskjeller for menn og kvinner

For høyde ser forskjellen ut til å være stor sammenlignet med variabiliteten, men for kroppstemperatur ser forskjellen ut til å være liten sammenlignet med variabiliteten

Vi må også vurdere variabilitetene til de observerte verdiene for å kunne trekke konklusjoner om forskjeller for menn og kvinner

For høyde ser forskjellen ut til å være stor sammenlignet med variabiliteten, men for kroppstemperatur ser forskjellen ut til å være liten sammenlignet med variabiliteten

Men vi er nødt til å tallfeste forskjellene!

## Definition (Empirisk varians)

Den empirisk varians til observasjonene  $x_1, x_2, \dots, x_n$  er gitt ved

$$s^2 = \frac{1}{n-1} [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2] = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2$$

# Sammenligning av menn og kvinner

	Høyde			Kroppstemperatur		
	Kvinne	Mann	Diff	Kvinne	Mann	Diff
Empirisk middelverdi	168.6	183.9	15.3	36.7	36.8	0.1
Empirisk std	7.4	6.7		0.4	0.4	



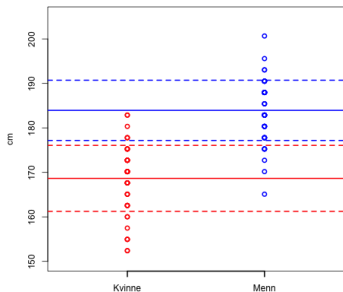
# Sammenligning av menn og kvinner

	Høyde			Kroppstemperatur		
	Kvinne	Mann	Diff	Kvinne	Mann	Diff
Empirisk middelverdi	168.6	183.9	15.3	36.7	36.8	0.1
Empirisk std	7.4	6.7		0.4	0.4	

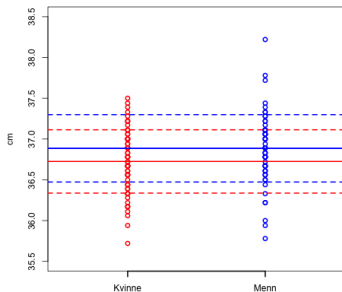
Forskjellen i kroppstemperatur er mindre enn ett standardavvik, men forskjellen i høyde er større enn to standardavvik

# Kryssplot med empirisk middelværdi og standarddeviasjon

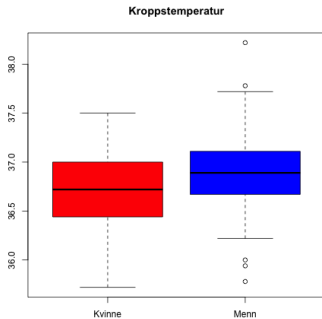
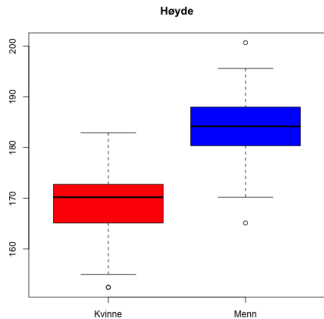
Kryssplot for Høyde



Kryssplot for Temperature

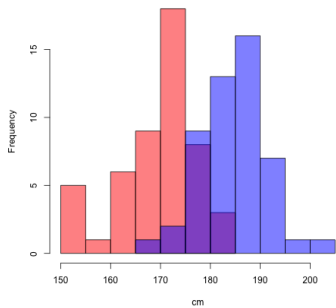


# Boxplot

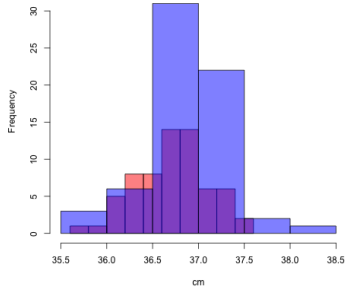


# Histogram

Høyde



Kroppstemperatur



Deskriptiv statistikk er en viktig startpunkt for å utforske de observerte verdiene.

Hjelper å skaffe seg en “bilde” av datane.

Deskriptiv statistikk er en viktig startpunkt for å utforske de observerte verdiene.

Hjelper å skaffe seg en “bilde” av datane.

Men:

- Ingen bruk av sannsynlighet. Vi har ingen tall på hvor sikker vi er på at menn er høyere enn kvinner.
- Veldig subjektivt om man anser det for å være en forskjell eller ikke

Deskriptiv statistikk er en viktig startpunkt for å utforske de observerte verdiene.

Hjelper å skaffe seg en “bilde” av datane.

Men:

- Ingen bruk av sannsynlighet. Vi har ingen tall på hvor sikker vi er på at menn er høyere enn kvinner.
- Veldig subjektivt om man anser det for å være en forskjell eller ikke

Neste gang begynner vi reisen mot statistisk inferens hvor vi gjennom sannsynlighetsteori kan tallfeste vår sikkerhet til konklusjonene vi gjør