

Velkommen til TMA4240

- TMA4240 Statistikk
- Jeg er Sara Martino
- Dere er MTDT, MTKJ, MTNANO, MTPETR
- Vi had forelesning:
 - Tirsdager kl 14.15-16.00 i F1
 - Torsdager kl 08.15-10.00 i A1 (Rustbygget, Handelshøyskolen, 2. etasje).

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.

Hvor finner dere statistikk i hverdagen??

Info om kurset finner dere her:

<https://wiki.math.ntnu.no/tma4240/2017h/start>

Hva skal vi gjøre 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

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

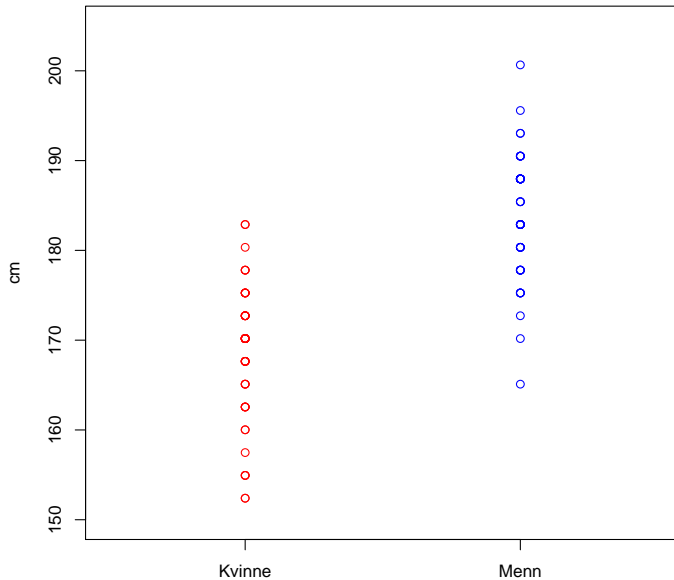
Men vi kan

- 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

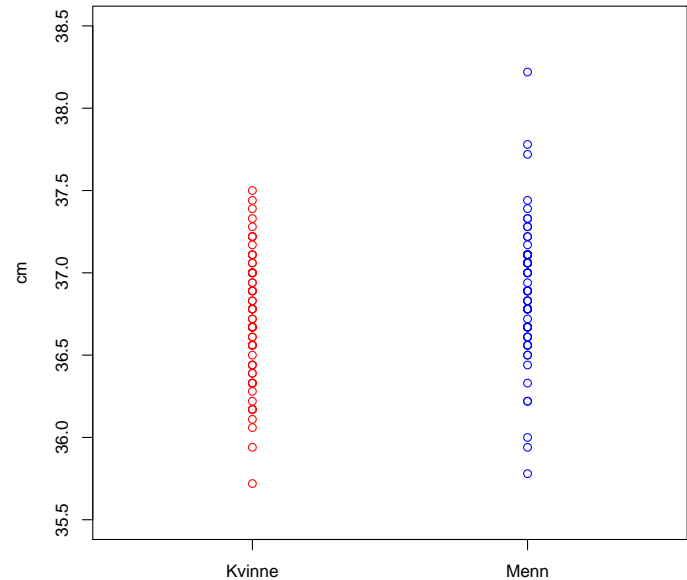
Dette er metoder for å beskrive de observerte verdiene og bruk av disse kalles **deskriptiv statistikk**

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

Kryssplot for Høyde



Kryssplot for Temperature



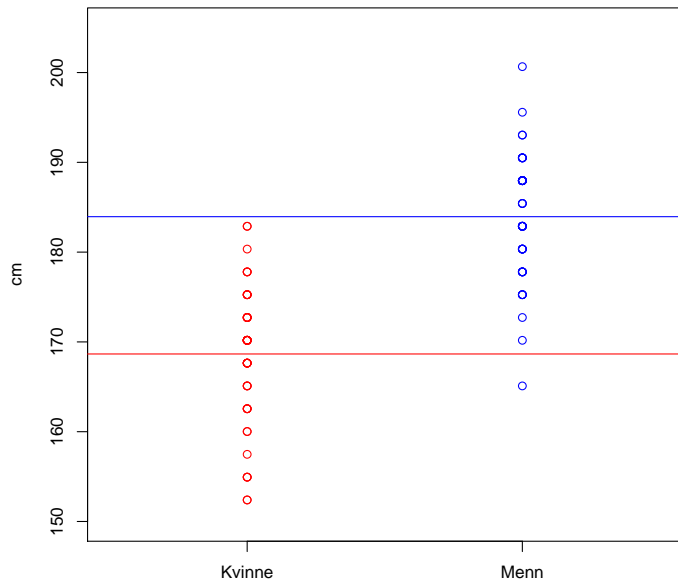
Sammenligning av senterer

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

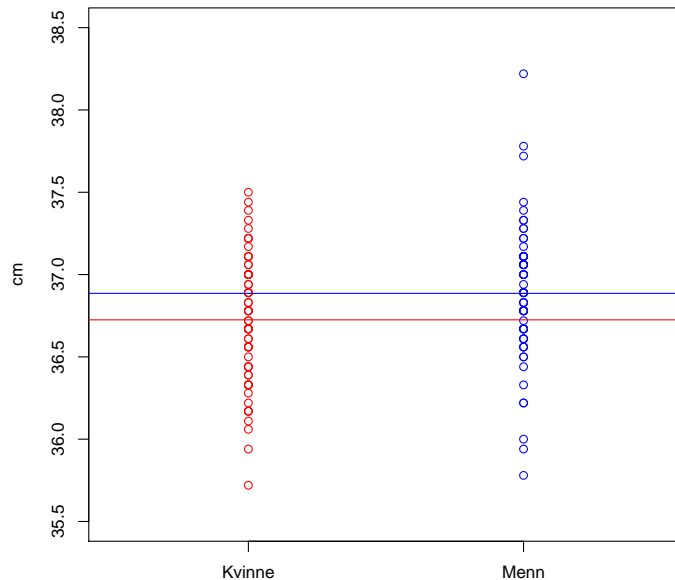
Er dette nok for å si at menn og kvinne har forskjellige høyder og temperaturer?

Kryssplot med empirisk middelværdi

Kryssplot for Højde



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 variabiliten, 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 variabiliten, men for kroppstemperatur ser forskjellen ut til å være liten sammenlignet med variabiliteten

Men vi er nødt til å tallfeste forskjellene!

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	

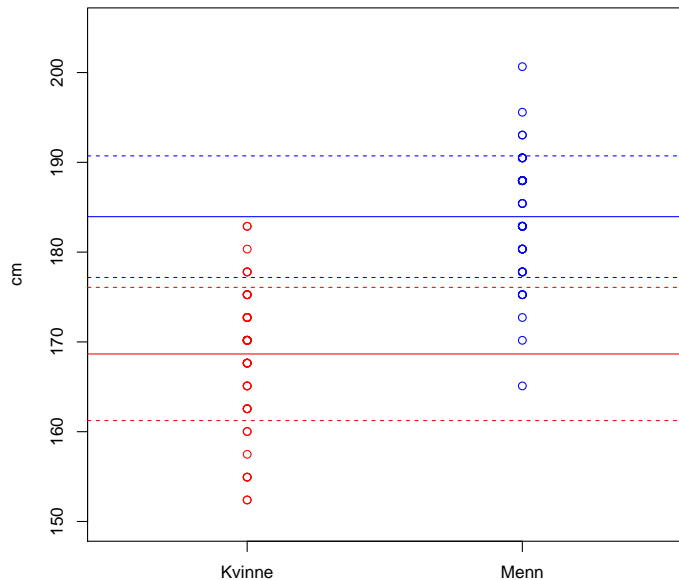
Sammenligning av menn og kvinner

	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	

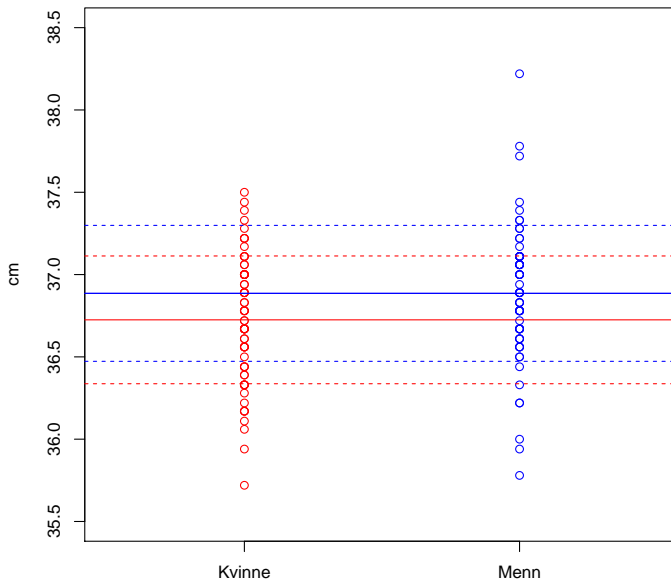
Forskjellen i kroppstemperatur er mindre enn ett standardavvik, men forskjellen i høyde er større enn to standardavvik

Kryssplot med empirisk middelværdi og standarddeviation

Kryssplot for Høyde

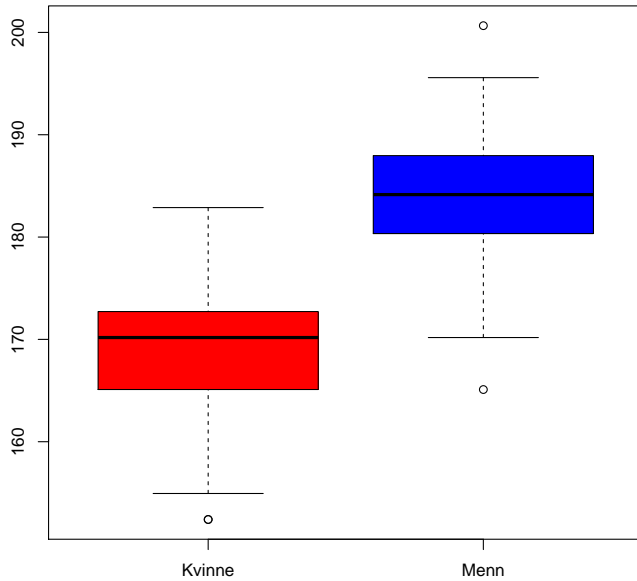


Kryssplot for Temperature

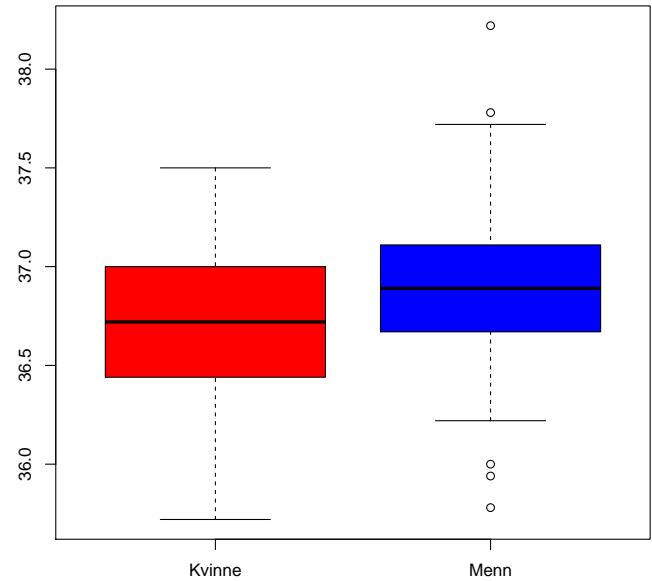


Boxplot

Høyde

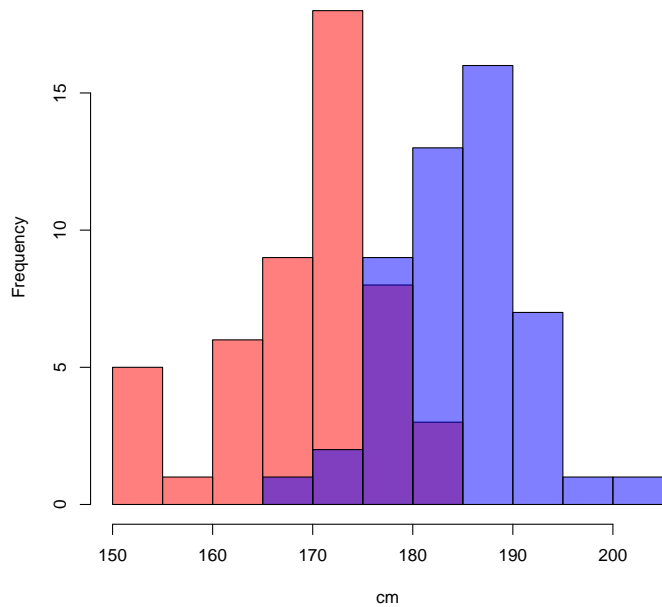


Kroppstemperatur

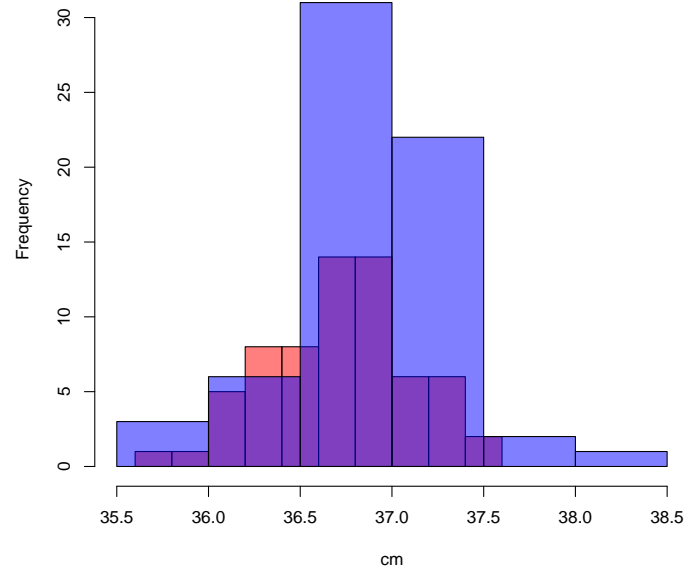


Histogram

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