

Tema: Deskriptiv statistikk for kontinuerlige data.

Av Kathrine Frey Frøslie, www.statistrikk.no

Kontinuerlige data er målinger som gjøres langs en skala, for eksempel tid, lengde og vekt. Noen ganger kan telldata (diskrete data) også oppsummeres som om de var kontinuerlige. Det gjelder hvis det er stor variasjon i det som telles, for eksempel antall elever på en skole eller antall melkekyr i et fjøs (i motsetning til f.eks. antall trampoliner i en hage.)

Beskrivelse av data gjøres ved hjelp av figurer (diagrammer) og oppsummeringstall.

Før man kan velge riktig oppsummeringstall, må man se på (tegne, plote) fordelingen til dataene og vurdere hvordan den ser ut, fordi symmetriske data og skjeve data (evt data med ekstremverdier) krever ulike oppsummeringstall.

De viktigste oppsummeringstallene for kontinuerlige data beskriver to egenskaper ved dataene: Det som er *typisk* for observasjonene, og *variasjonen* i observasjonene.

I dette dokumentet finner du to eksempler med figurer og repetisjonsoppgaver som er laget for å bygge opp den statistiske magesfølelsen for disse begrepene:

Histogram og Boksplott

Oppsummeringstall for det typiske

Oppsummeringstall for variasjon

Eksemplene og oppgavene er knyttet til blogginnlegget *Variasjon og standardavvik* som ble publisert på bloggen Statistrikk 28. august 2017.

<http://www.statistrikk.no/2017/08/28/variasjon-del-2/>

Tallene i eksemplene er oppdiktet, men basert på faktakunnskap om tenåringer og gravide damer, så de kunne vært reelle.

Du kan gjerne hente inn dine egne data og bruke appen www.tallforteller.no til å lage figurer og oppsummeringstall i stedet.

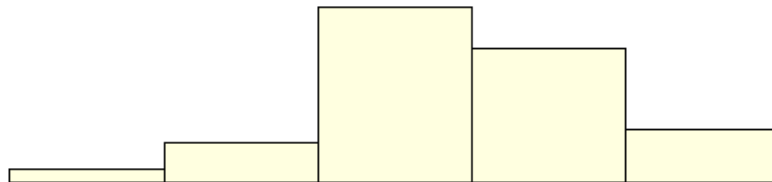
Eksempel 1: Svangerskapslengder for 31 kvinner

A) Tegn en tallinje under de fargerike prikkene og skisser både et histogram og et boksplott.

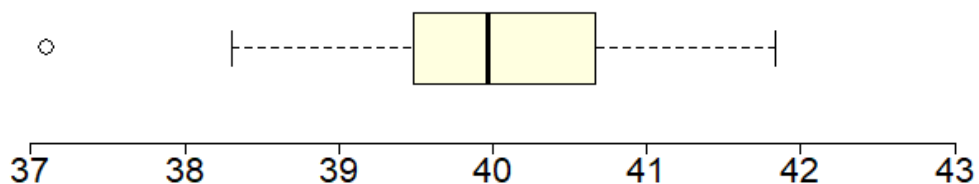
B) Sjekk om det du får ligner på histogrammet og boksplottet under.

C) Kommenter fordelings form.

Observasjonene



Histogram og boksplott



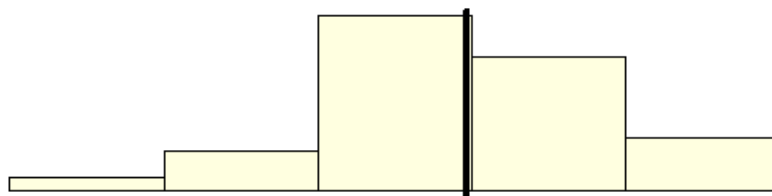
Gjennomsnitt og median er to oppsummeringstall som *kan* beskrive det typiske, det vanlige, senteret eller midten i kontinuerlige data.

A) Tell prikker på hver side av medianen og se at det er like mange på hver side av medianen

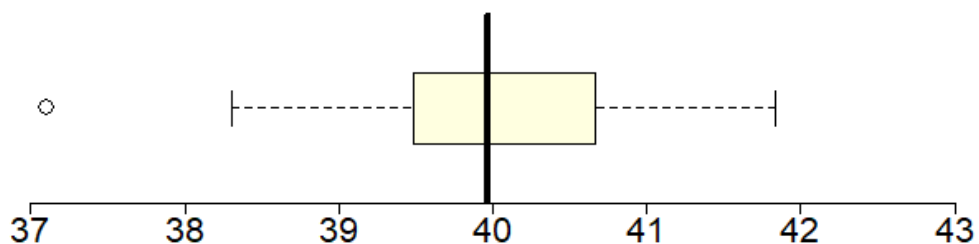
B) Sjekk at gjennomsnittsstreken er ved tyngdepunktet/balansepunktet for fordelingen

C) Hvor stor forskjell er det på median og gjennomsnitt? Hva ville du valgt som oppsummeringstall for det typiske i denne fordelingen?

Observasjonene

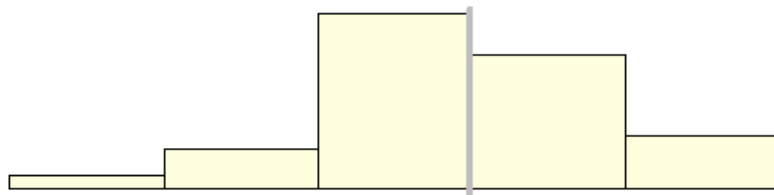


Medianen er markert med svart

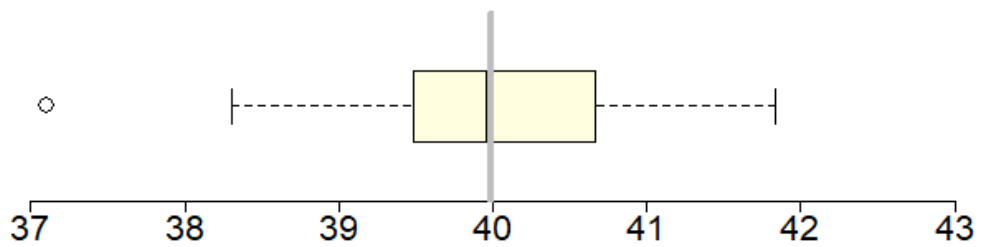


Median = 40,0

Observasjonene



Gjennomsnittet er markert med grå linje

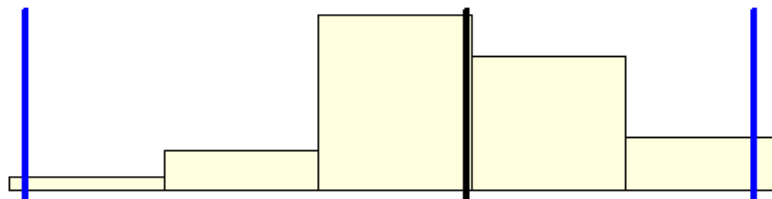


Gjennomsnittsverdien = 40,0

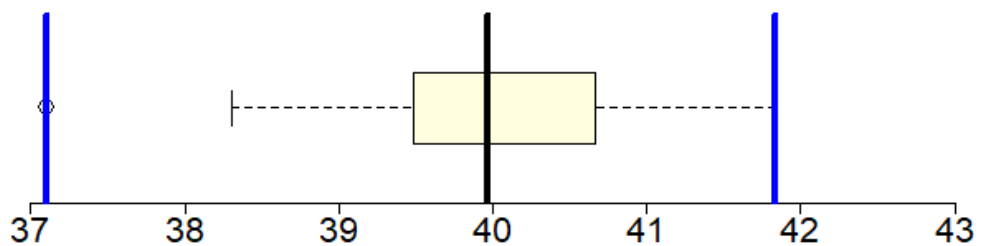
Variasjon kan oppsummeres på mange måter.

A) Hva menes med minimumsverdi, maksimumsverdi, range, nedre og øvre kvartil, og interkvartilbredde?

Observasjonene



Variasjon målt med range: Avstand mellom de blå strekene

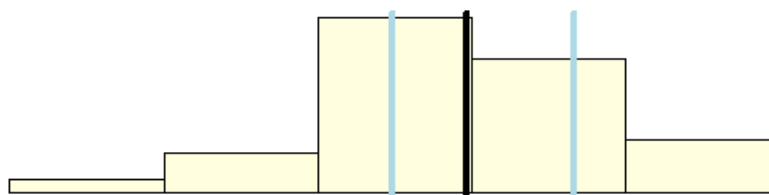


Minimumsverdi = 37,1

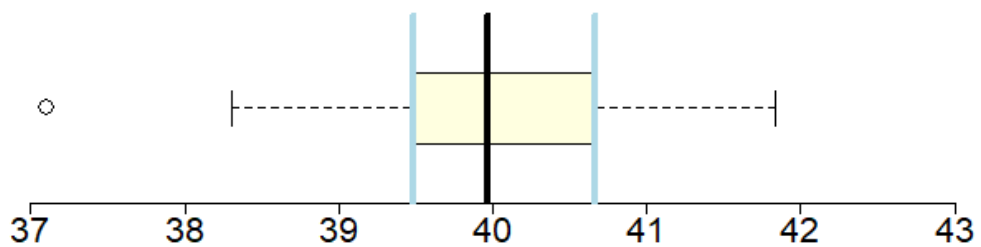
Maksimumsverdi = 41,8

B) Er alle disse tallene meningsfulle oppsummeringstall for variasjon i dette datasettet?

c) Gi en praktisk tolkning av dem (lett hvis du leser innlegget om variasjon og standardavvik på www.statistikk.no)



Variasjon målt med IQR: Avstand mellom de lyseblå strekene

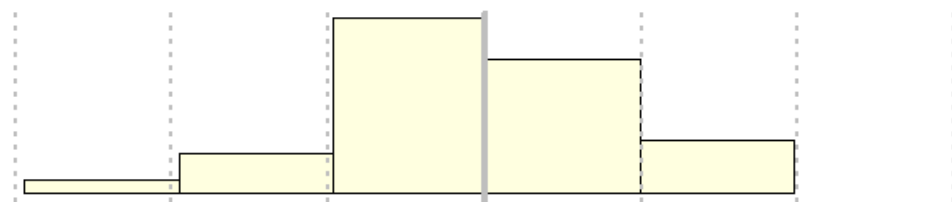


Nedre kvartil = 39,5

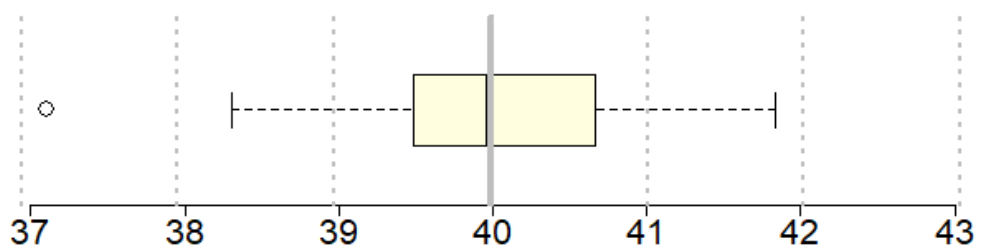
Øvre kvartil = 40,7

- A) Hva kan du oppsummere med standardavviket?
- B) Gi en muntlig beskrivelse eller tolkning av standardavvik
- C) Hva er formelen for standardavvik?
- D) Er standardavviket knyttet til gjennomsnittet eller medianen?

Observasjonene



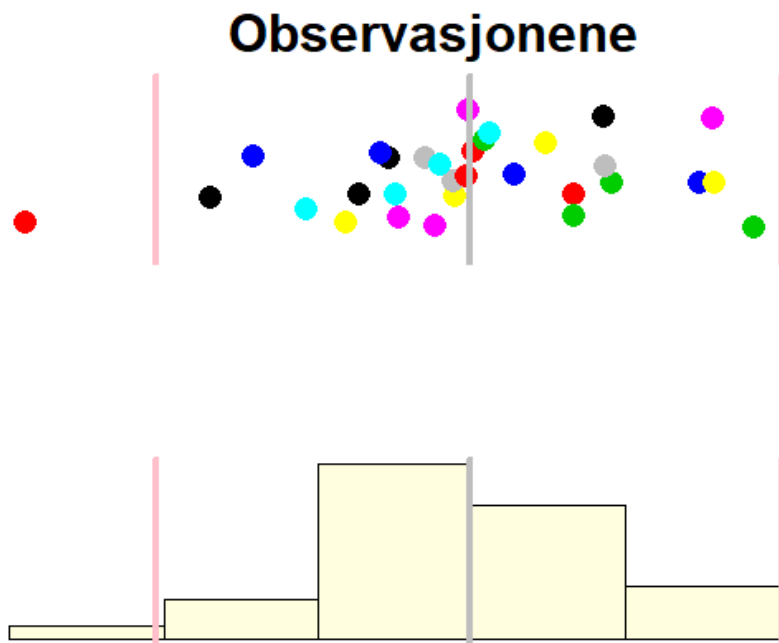
Variasjon målt i 1, 2, 3 standardavvik fra gjennomsnittet



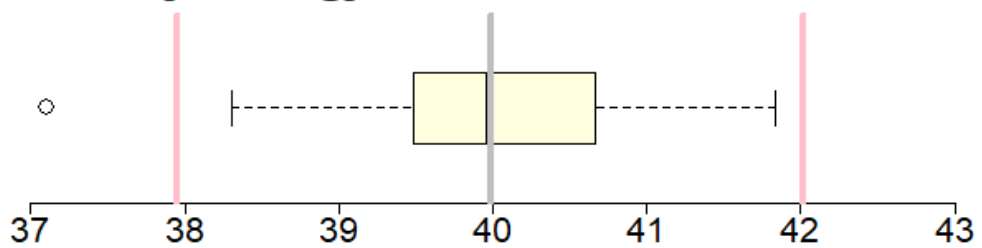
SD = 1,0

A) Hvor mange standardavvik er det mellom de rosa strekene?

B) Hvor stor del av observasjonene havner mellom de røde strekene?

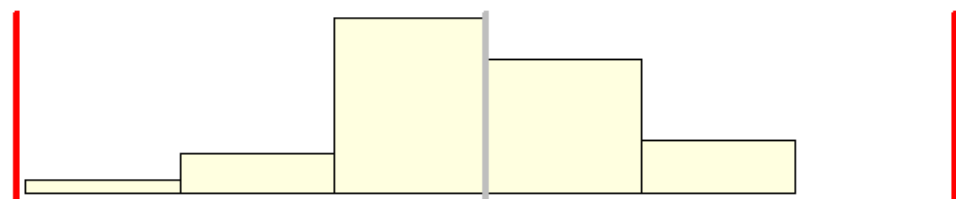


Rosa linjer er gj.sn. +/- 2 standardavvik

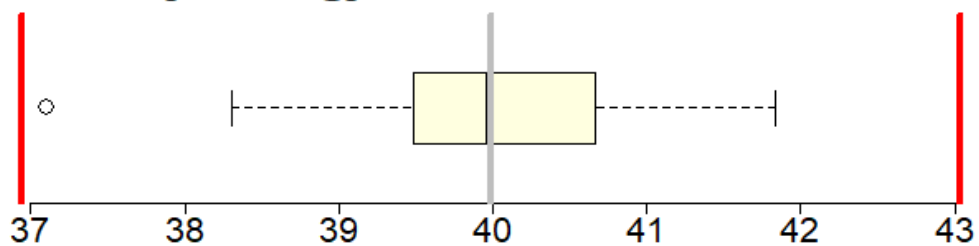


A) Hvor mange standardavvik er det mellom de røde strekene?

B) Hvor stor del av observasjonene havner mellom de røde strekene?



Røde linjer er gj.sn. ± 3 standardavvik



C) Er gjennomsnittet og standardavviket meningsfulle oppsummeringstall i dette datasettet?

Hvis ja: På hvilken måte? Hvis nei: Hvorfor ikke?

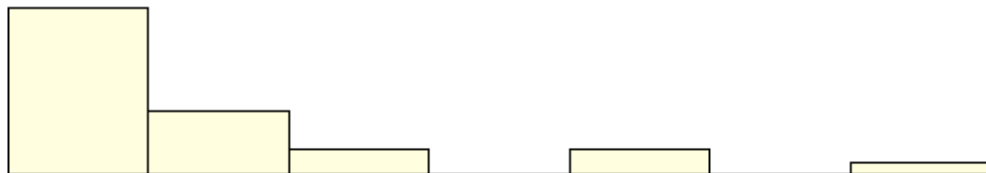
Eksempel 2: Dobesøk i minutter for 23 tenåringer

A) Tegn en tallinje under de fargerike prikkene og skisser både et histogram og et boksploott.

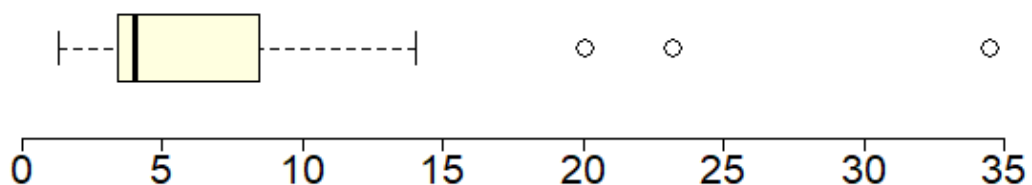
B) Sjekk om det du får ligner på histogrammet og boksplottet under.

C) Kommenter fordelingsens form.

Observasjonene



Histogram og boksploott



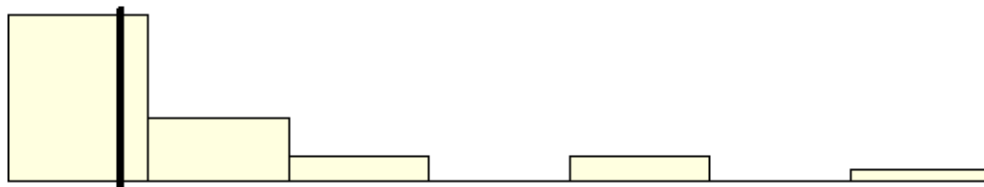
Gjennomsnitt og median er to oppsummeringstall som *kan* beskrive det typiske, det vanlige, senteret eller midten i kontinuerlige data.

A) Tell prikker på hver side av medianen og se at det er like mange på hver side av medianen

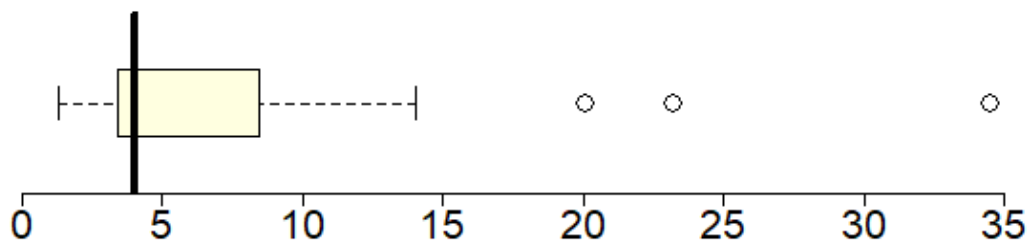
B) Sjekk at gjennomsnittsstreken er ved tyngdepunktet/balansepunktet for fordelingen

C) Hvor stor forskjell er det på median og gjennomsnitt? Hva ville du valgt som oppsummeringstall for det typiske i denne fordelingen?

Observasjonene

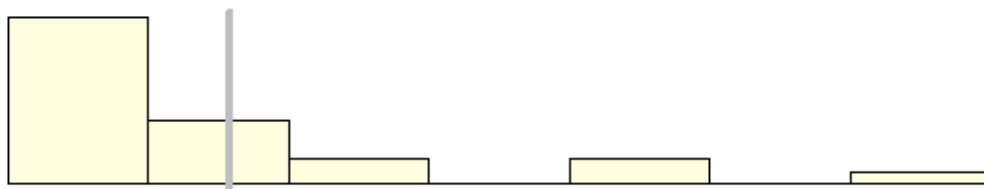


Medianen er markert med svart linje

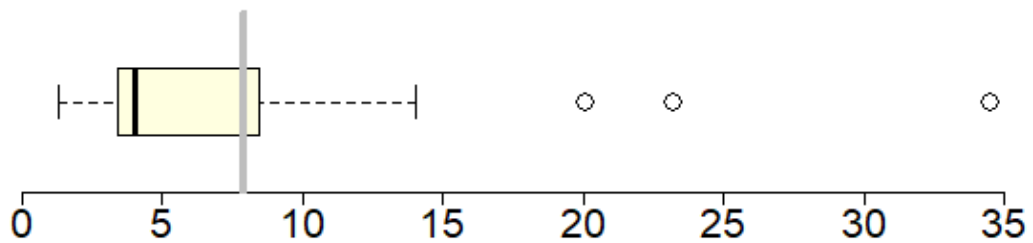


Median = 4,0

Observasjonene



Gjennomsnittet er markert med grå linje

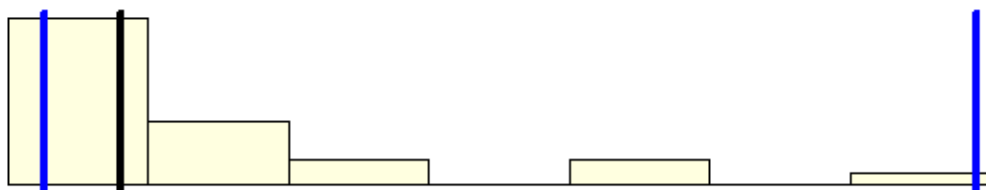


Gjennomsnittsverdien = 7,9

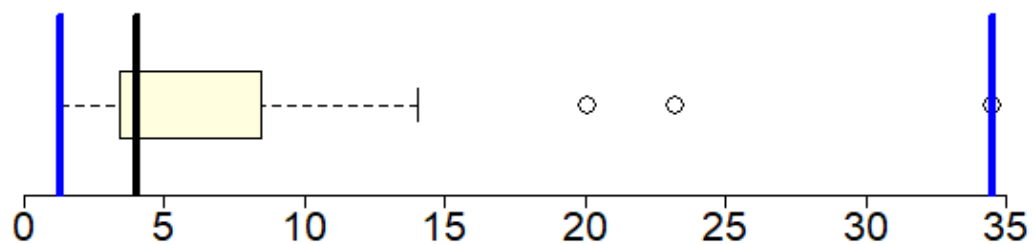
Variasjon kan oppsummeres på mange måter.

A) Hva menes med minimumsverdi, maksimumsverdi, range, nedre og øvre kvartil, og interkvartilbredde?

Observasjonene



Variasjon målt med range: Avstand mellom de blå strekene



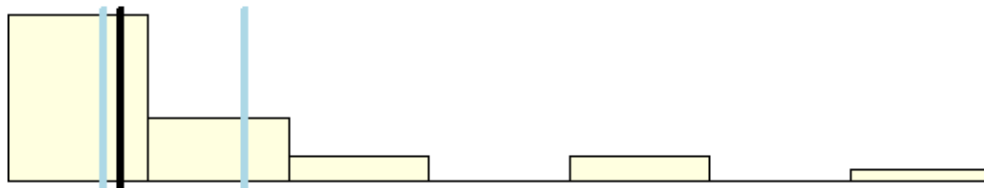
Minimumsverdi = 1,3

Maksimumsverdi = 34,5

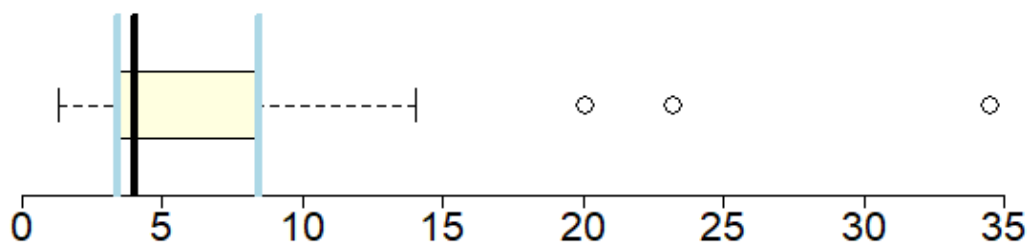
B) Er alle disse tallene meningsfulle oppsummeringstall for variasjon i dette datasettet?

c) Gi en praktisk tolkning av dem (lett hvis du leser innlegget om variasjon og standardavvik på www.statistikk.no)

Observasjonene



Variasjon målt med IQR: Avstand mellom de lyseblå strekene

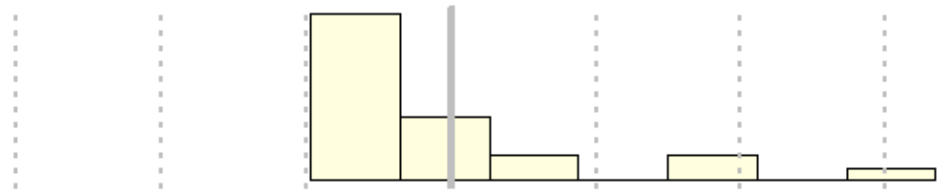
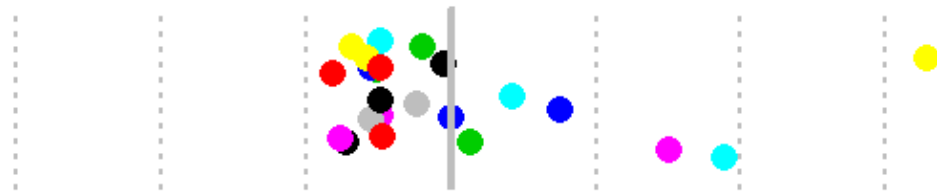


Nedre kvartil = 3,4

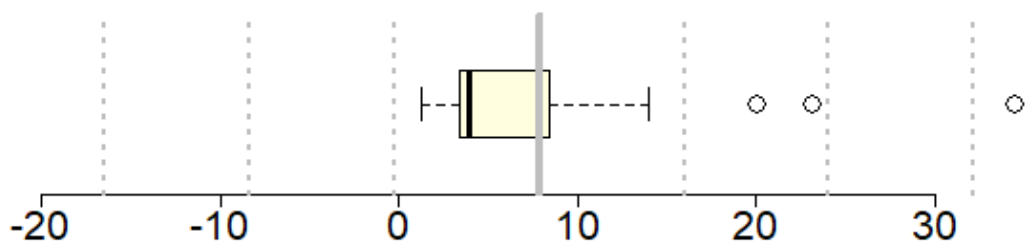
Øvre kvartil = 8,5

- A) Hva kan du oppsummere med standardavviket?
- B) Gi en muntlig beskrivelse eller tolkning av standardavvik
- C) Hva er formelen for standardavvik?
- D) Er standardavviket knyttet til gjennomsnittet eller medianen?

Observasjonene



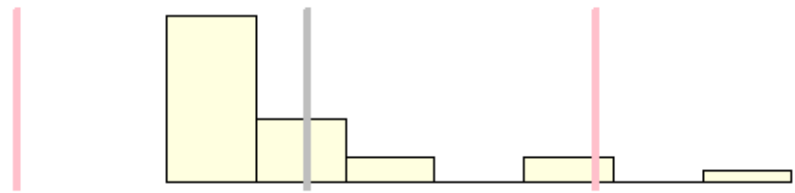
Variasjon målt i 1, 2 og 3 standardavvik fra gjennomsnittet



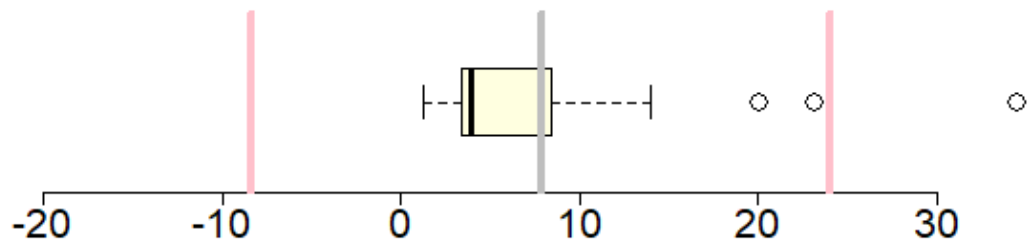
SD = 8,1

- A) Hvor mange standardavvik er det mellom de rosa strekene?
- B) Hvor stor del av observasjonene havner mellom de røde strekene?

Observasjonene



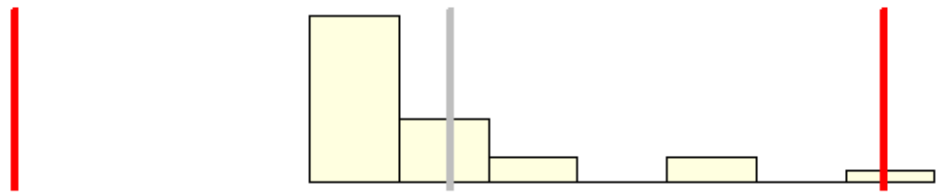
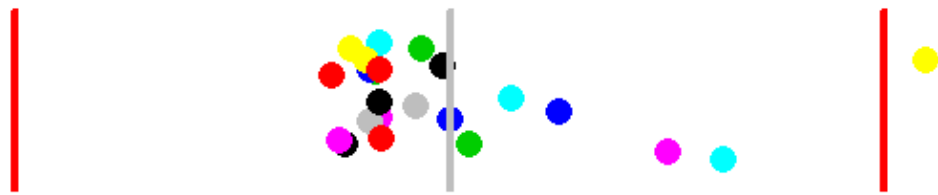
Rosa linjer er gj.sn. +/- 2 standardavvik



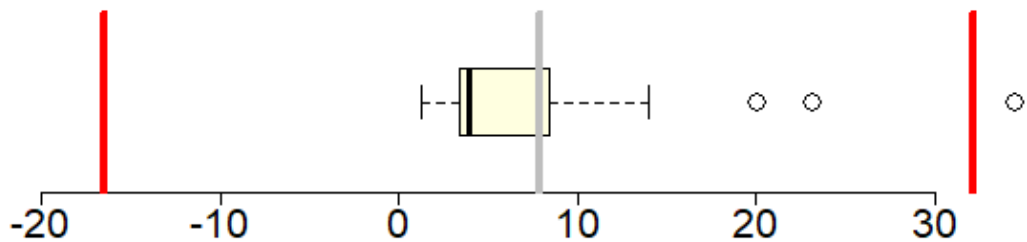
A) Hvor mange standardavvik er det mellom de røde strekene?

B) Hvor stor del av observasjonene havner mellom de røde strekene?

Observasjonene



Røde linjer er gj.sn. +/- 3 standardavvik



C) Er gjennomsnittet og standardavviket meningsfulle oppsummeringstall i dette datasettet?

Hvis ja: På hvilken måte? Hvis nei: Hvorfor ikke?