

STK1000 Studentene forventes å lese Ch 3.4 og Ch 5 i læreboka (MMC).

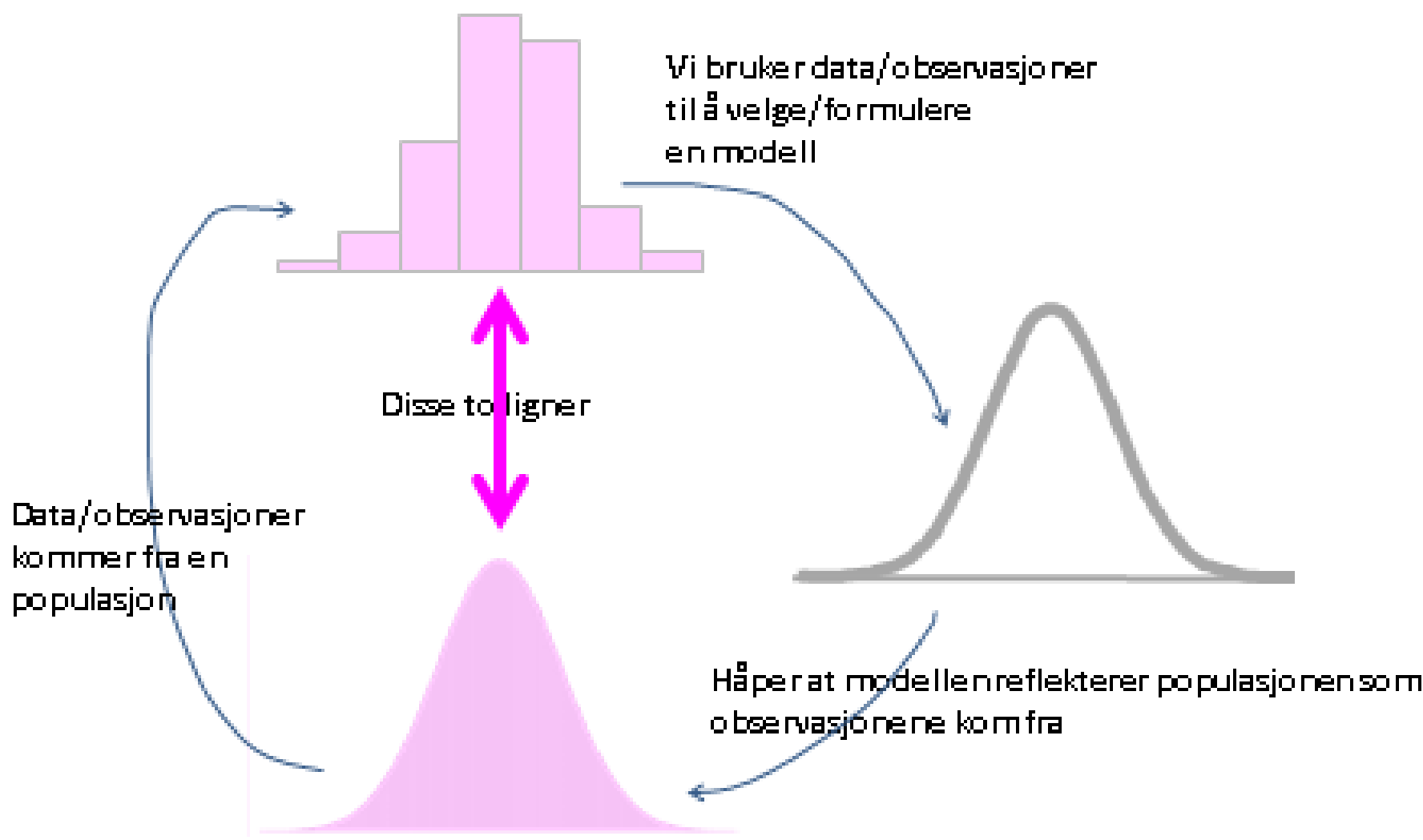
# Inferens: Å trekke slutninger om en populasjon basert på et utvalg

Det er kostbart/vanskelig å undersøke hele populasjonen. I stedet trekkes et representativt utvalg fra populasjonen, og så undersøker vi dette.

Idé: Sånn som et er i utvalget, er det omtrent i populasjonen også.

Utvalg = «Sample»





Vi bruker data/observasjoner til å velge/formulere en modell

Disse to ligner

Data/observasjoner kommer fra en populasjon

Håper at modellen reflekterer populasjonen som observasjonene kom fra

Undersøker og beskriver disse



...for å si noe om alle nordmenn

Undersøker og beskriver disse



...for å si noe om alle spedbarn

Undersøker og beskriver disse



...for å si noe om alle eldre

Vi bruker gjennomsnittet både til å  
**beskrive** dem vi har undersøkt



...og til å si noe generelt om dem vi  
**ikke** har undersøkt; populasjonen

# Inferens: Fra populasjon til utvalg og tilbake igjen

Observasjoner	Teoretisk/tenkt Modell	Populasjon
<b>Histogram</b>	<b>Tetthetskurve, Sannsynlighetsfordeling</b>	<b>Fordelingen i populasjonen</b>
Viser fordelingen av faktiske observerte verdier som søyler i intervaller	Viser den idealiserte fordelingen av verdier som en kontinuerlig kurve	Ukjent, men vi tror at den observerte fordelingen og den teoretiske fordelingen ligner på populasjonsfordelingen
Kan være symmetrisk, skjev, ha en eller flere topper, lette eller tunge haler (outliere/ekstremverdier)	Kan være symmetrisk, skjev, ha en eller flere topper, lette eller tunge haler	Samme her
<b>Statistic = Observator</b>  Tolkbare oppsummeringstall som kan beregnes i utvalget, for eksempel Gjennomsnitt: $\bar{x}$ Median Kvartiler (Empirisk) standardavvik: sd, SD Andeler	Egenskaper ved den teoretiske fordelingen, som tilsvarer de størrelsene vi kan beregne i utvalget: Tyngdepunkt/forventningsverdi Median (tallet som halverer arealet i fordelingen) Standardavvik	<b>Parameter</b>  Oppsummeringstall for den ukjente populasjonen, som vi ønsker å gjette verdien av. Det kalles å estimere. Forventningsverdi $\mu$ Standardavvik $\sigma$ Andel p

## Bias vs Variability

- Bias is the accuracy of a statistic
- Variability is the precision of a statistic

