

**Vi må uttrykke oss presist  
om statistiske analyser**

**Vi må undersøke mange  
for å trekke ut essensen**

**Oppsummeringene  
er viktige**

**Igangsetting av tiltak?**

**Grisete datasett?**

**Svære datasett?**

**Data-behandling**

*Notasjon, språk og begreper*

*Hvor mange er «mange»?*

*Hvilke oppsummeringer?*

*Figurer? Oppsummeringstall?*

*Hva ønsker du å oppnå?*

*- Hvilken respons?*

*- Hvilken forklaringsmodell?*

*Datavasking, forberedelser  
til analysen*

*Big data? Forklare? Predikere?*

*Programvare (R)*

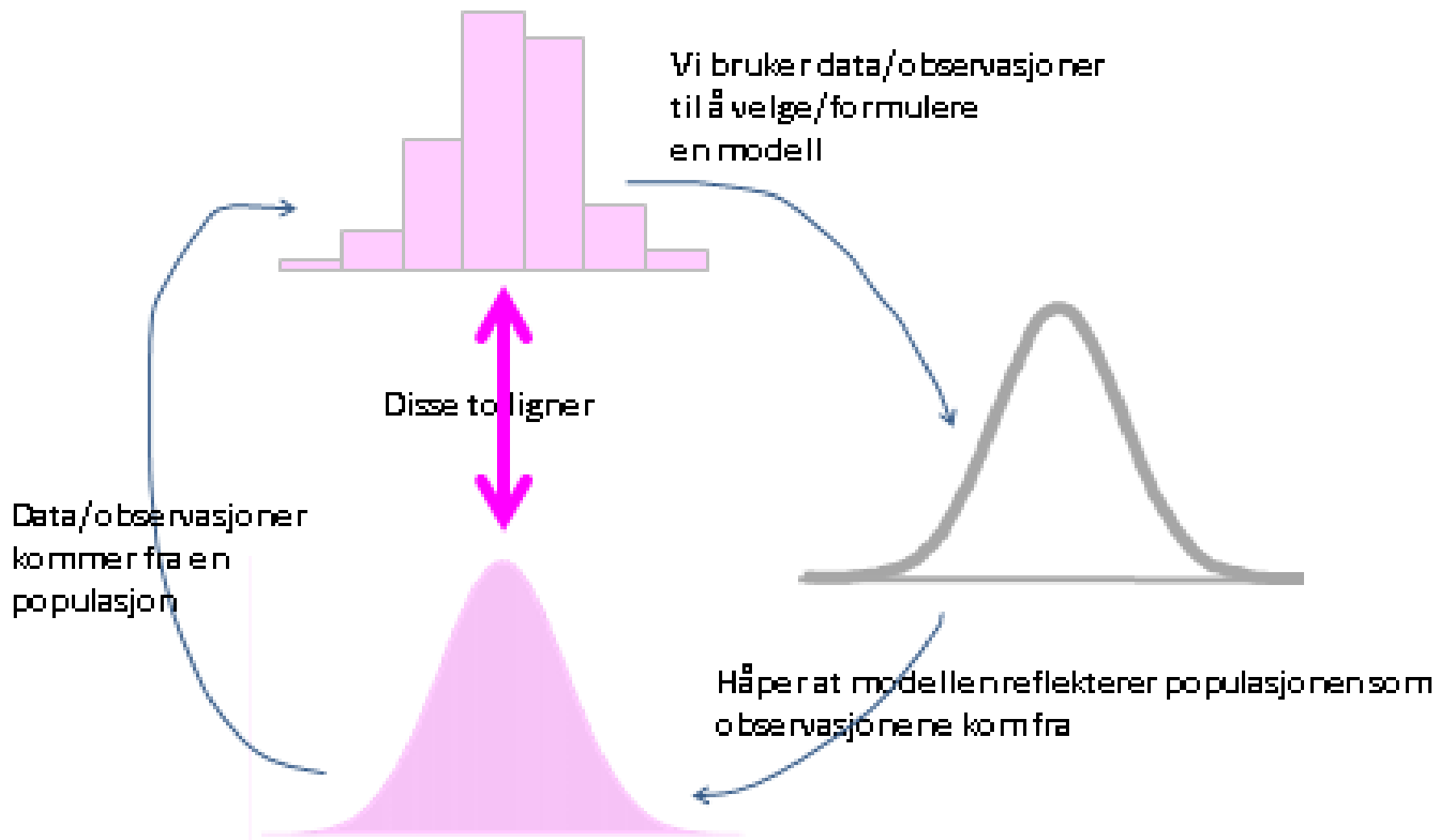
Fra forrige forelesning

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

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

Utvalg = «Sample»





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

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>	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>
Tolkbare oppsummeringstall som kan beregnes i utvalget, for eksempel Gjennomsnitt: $\bar{x}$ Median Kvartiler (Empirisk) standardavvik: sd, SD Andeler		Oppsummeringstall for den ukjente populasjonen, som vi ønsker å gjette verdien av. Det kalles å estimere. Forventningsverdi $\mu$ Standardavvik $\sigma$ Andel p



## Sannsynlighetsbegrepet



## Sannsynlighetsmodeller

### Stokastiske (Tilfeldige) variabler

- ✓ Diskrete
- ✓ Kontinuerlige

### Forventning og varians for stokastiske variabler

## Generelle sannsynlighetsregler

