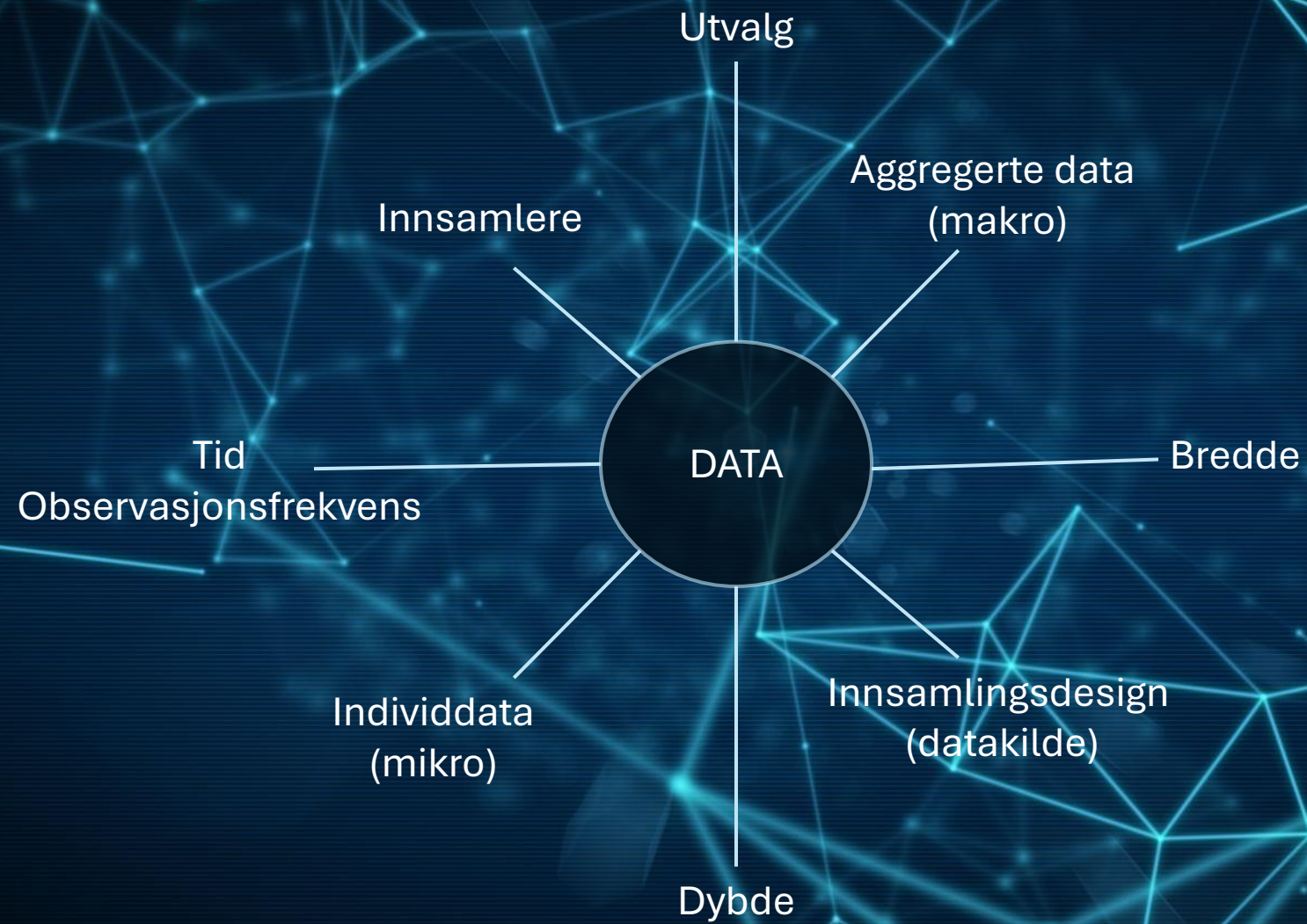




Data

Sok-3024 H25



Innsamlere

– Hvem har samlet inn data og til hvilket formål?

Primærdata

Førstehåndsinformasjon samlet inn av forskergruppen med formål å svare på forskningsspørsmål i deres forskningsprosjekt

Sekundærdata

Andrehåndsinformasjon samlet inn av noen annen enn forskergruppen urelatert til forskningsspørsmålene i forskningsprosjektet

Aggregering

- På hvilket nivå er data samlet inn?

Aggregerte data
(makrodata)

Oppsummerte eller grupperte data, f.eks. nasjonalt eller regionalt nivå.

Eksempel: Gjennomsnittlig alder, prosent kvinner, prosent arbeidsledighet, BNP per innbygger

Individdata
(mikrodata)

Enkeltindivid- eller enhetsnivå data, f.eks. data om enkelte personer, husholdninger, bedrifter.

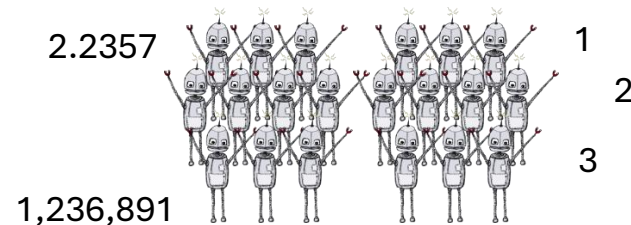
Eksempel: alder, kjønn, arbeidsstatus, årlig inntekt, profitt

Bredde og dybde

Tradisjonell inndeling

Kvantitative data

Tall



Analyse:

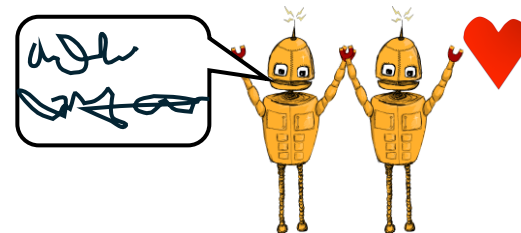
Statistiske forskjeller i gjennomsnitt, median, proporsjoner, ved bruk av statistiske metoder

Datakilde:

Surveydata
Eksperiment
Registerdata (administrative data)
Offisiell statistikk

Kvalitative data

Tekst, lyd, bilde



Analyse:

Tematikker, strukturer, diskurser

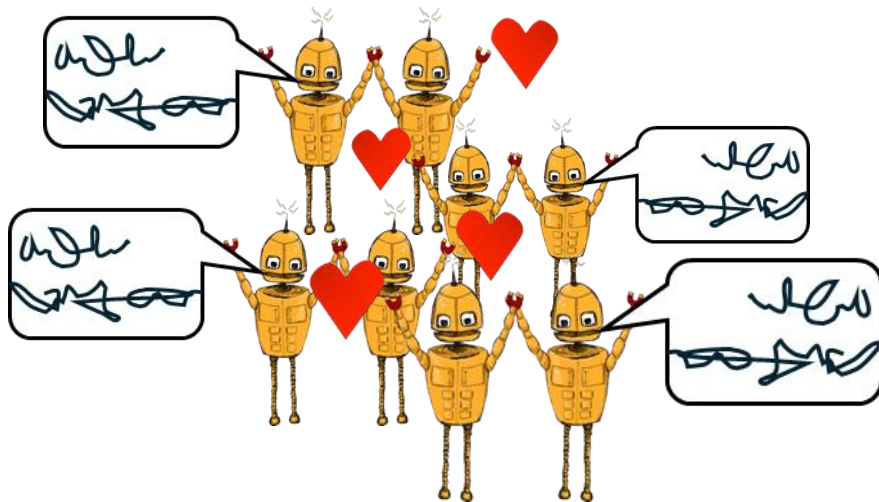
Datakilde:

Intervjuer
Observasjonsstudier
Dokumenter
Spørreundersøkelser med åpne spørsmål

Bredde og dybde – med maskinlæring

Kvalitative data

Tekst, lyd, billede



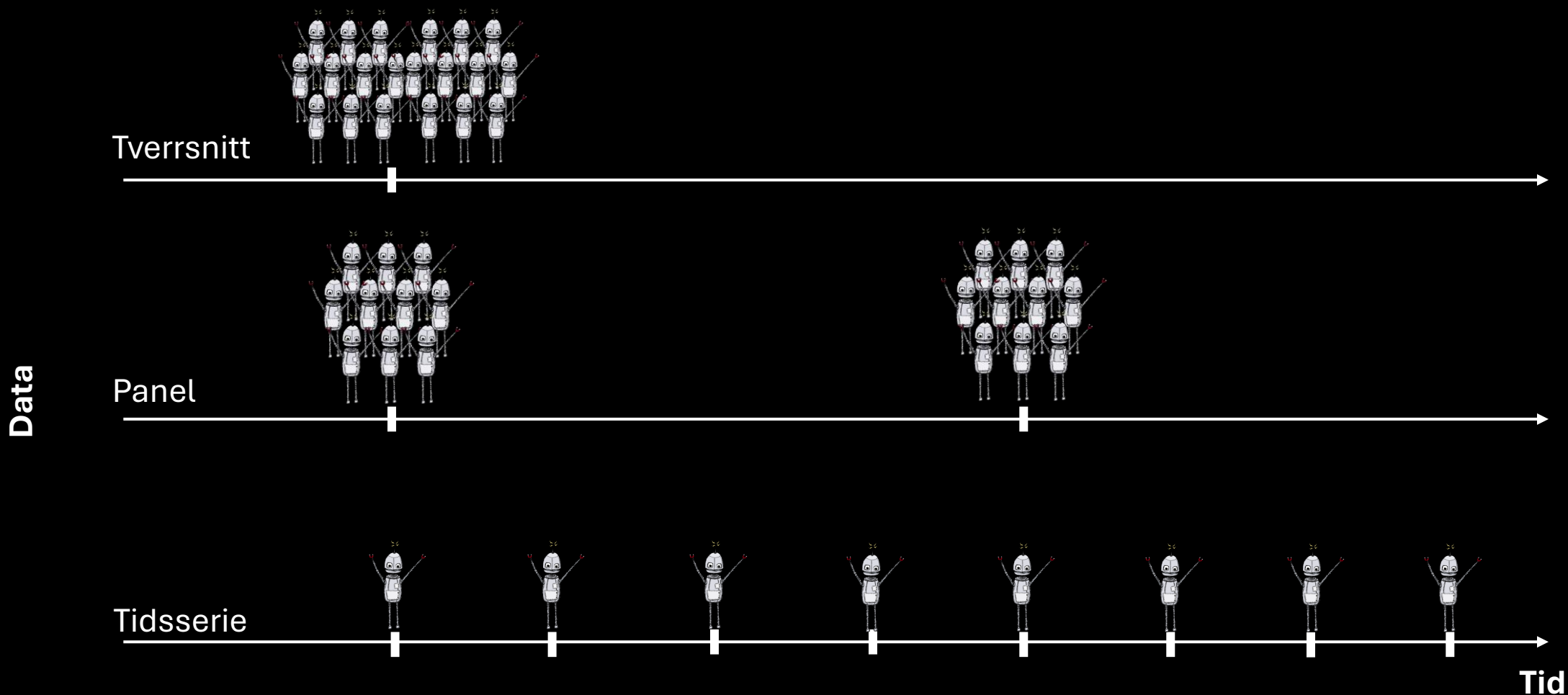
Kvantitativ analyse

Tall

Hassan et al., 2025. Text as Data in Economic Analysis, Journal of Economic Perspectives, 39(3). DOI: 10.1257/jep.20231365

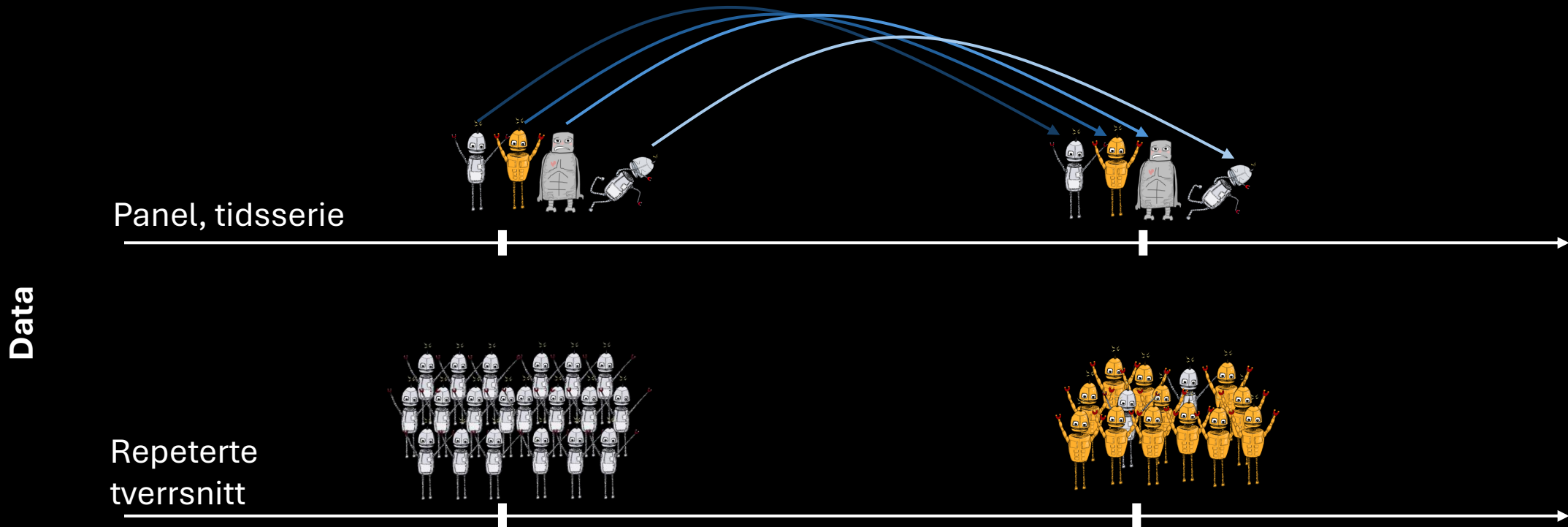
Tid

– ved hvor mange tidspunkter blir enhver enhet observert?



Tid

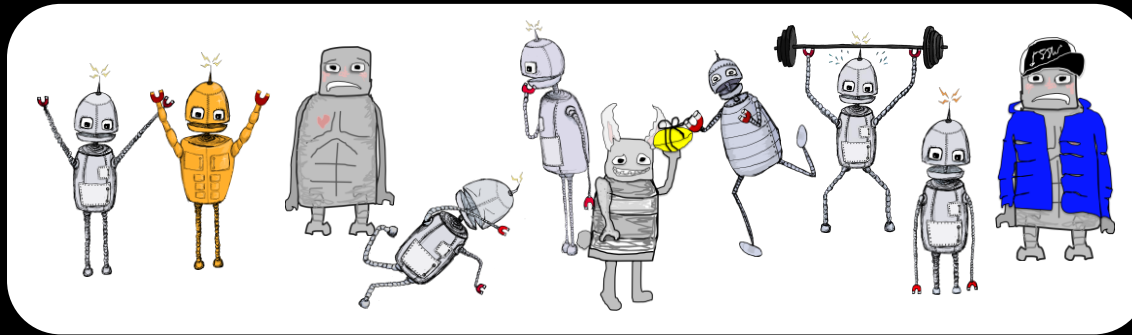
– ved hvor mange tidspunkter blir enhver enhet observert?



Utvalg

- Hvordan har enhetene i data valgts ut?

Populasjon



Sannsynlighetsutvalg

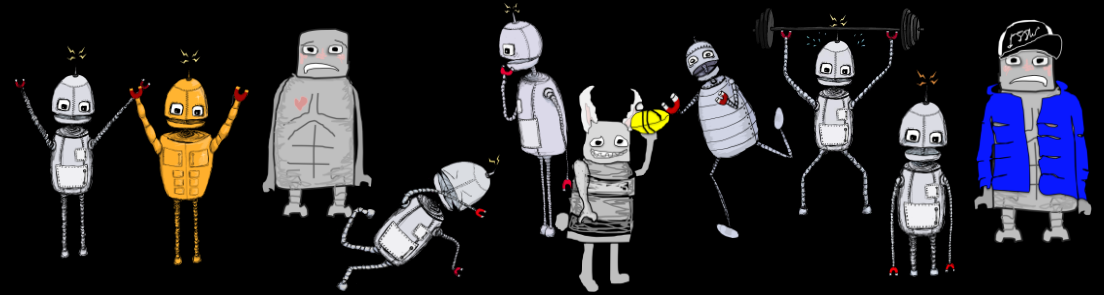
- Enkelt tilfeldigheitsutvalg
- Systematisk utvalg.
- Stratifisert utvalg
- Klyngeutvalg

Ikke-sannsynlighetsutvalg

- Bekvemmelighetsutvalg
- Selvseleksjon
- Kvoteutvalg
- Snøballutvalg

Utvalg

Sannsynlighetsutvalg



Utvalg

Beskrivelse

Fordeler/ulemper

Enkelt tilfeldighetsutvalg

Hvert individ i populasjonen har lik sannsynlighet å bli med i utvalget. F.eks. 0.001%

- + Representativt (gitt bibetingelser)
- Kan gi få observasjoner av noen grupper
- Tid og kostnadsnivå. Krever liste over hele populasjonen

Systematisk utvalg

Hvert n:te individ blir valgt, der n er forbestemt og starttallet er tilfeldig valgt. F.eks. Hver 5. person med start på person 8.

- + Enklere og «billigere» enn tilfeldig utvalg
- Risiko for skjevheter dersom det er mønster i populasjonen.

Stratifisert utvalg

Tilfeldig valg av individer innen gitte strata (grupper). F.eks. 100 tilfeldig trukne menn og kvinner.

- + Sikrer representasjon av viktige undergrupper
- Krever kunnskap om populasjonens struktur for å sikre representativitet.
- Tid og kostnadsnivå.

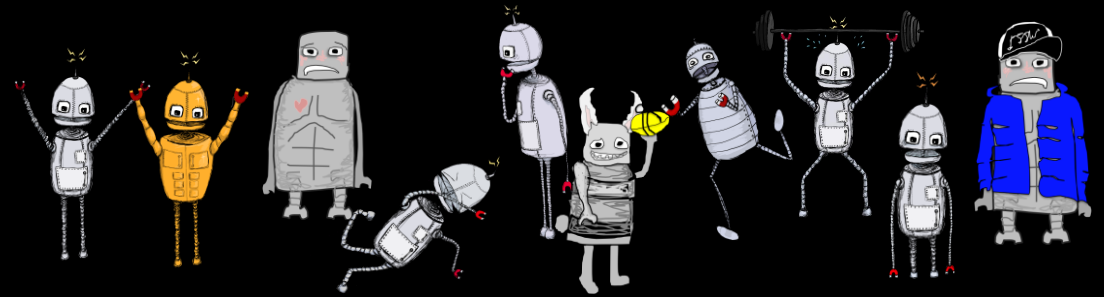
Klyngeutvalg

Tilfeldig valg av klynger og evt. Individer i klyngene. F.eks. skoler

- + kostnadsbesparende når populasjonen er geografisk utspredd
- Risiko for skjevheter

Utvalg

Ikke-sannsynlighetsutvalg



Utvalg	Beskrivelse	Fordeler/ulemper
Bekvemmelighetsutvalg	Utvalg basert på tilgjengelighet. F.eks. studenter ved UiT.	+ Enkelt og billig - Stor risiko for skjevheter (vanskelig å generalisere resultater til populasjonen)
Selvseleksjon	Deltagere velger selve å være med. F.eks. Åpen spørreundersøkelse på nett.	+ Enkelt og billig - Stor risiko for skjevheter og seleksjonsbias
Kvoteutvalg	Forbestemte kvoter med bekvemmelighetsutvalg innen kvotene. F.eks. 50% studenter i Tromsø, 50% studenter i Oslo	+ Sikrer representasjon av viktige grupper + Enkelt og billig - Stor risiko for skjevheter (sammenligne med stratifisert utvalg).
Snøballutvalg	Én deltaker rekrutterer andre deltakere.	+ Kan gi tilgang til populasjoner som forskeren ellers ikke har tilgang til eller kunnskap om) - Stor risiko for systematiske skjevheter

Et (ufullstendig) utvalg av datakilder

Offisiell statistikk (aggregerte data):



<https://www.ssb.no/>



<https://data.europa.eu/en>



<https://ec.europa.eu/eurostat/web/main/data/database>



<https://www.oecd.org/en/data.html>



<https://data.worldbank.org/>

<https://www.imf.org/en/Data>



<https://ilostat.ilo.org/data/>



<https://uis.unesco.org/en/home>



<https://unece.org/data>



<https://data.met.no/>



<https://www.ncei.noaa.gov/cdo-web/>

Et (ufullstendig) utvalg av datakilder

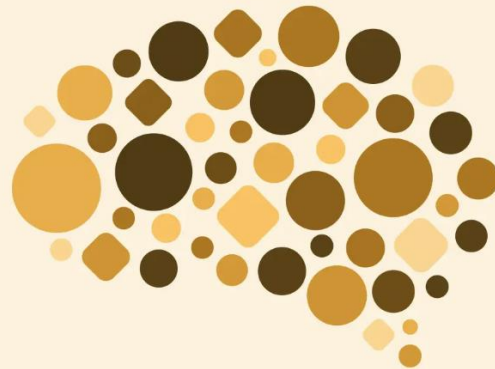
Mikrodata

Surveybanken

Her kan du finne, analysere og laste ned data fra spørreundersøkelser og andre datasamlinger. Surveybanken inneholder 750 000 spørsmål fra nesten 3000 spørreundersøkelser tilbake til 1957 - i tillegg til en rekke kvalitative data.

Se hvordan folks holdninger og meninger har endret seg siden midten av forrige århundre.

<https://sikt.no/surveybanken>



ESS Data Portal

Search, visualise, analyse, build your own datafile, and download data from the ESS, CRONOS and EOSC Future.

<https://www.europeansocialsurvey.org/data-portal>

microdata.no

Søknadsfri tilgang til kobling og analyse av store mengder detaljerte registerdata

<https://www.microdata.no/>

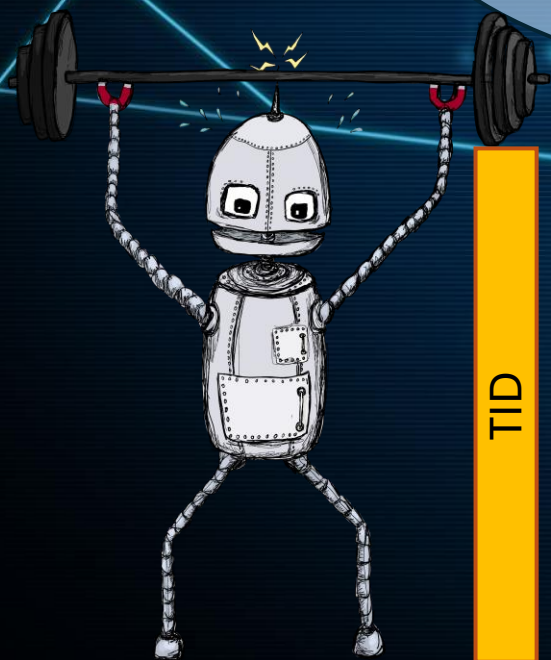
Ting å tenke på ved

Innsamling av egne data



Spørreundersøkelser

Eksperimenter



TID

Design

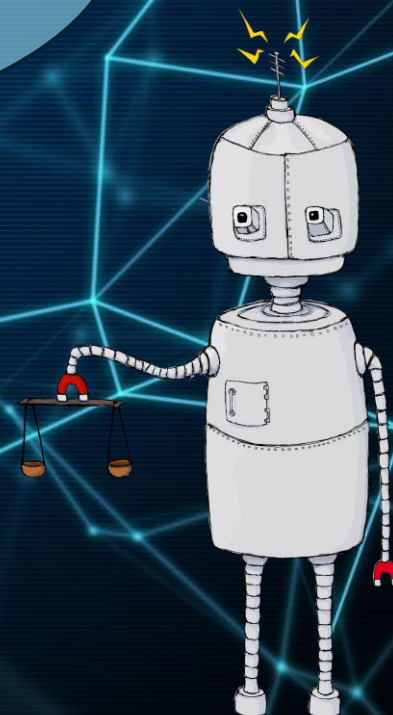
Hvordan sikkerstiller du at du måler det du ønsker å måle?

Utvalg

Hvordan får du tak i (tilstrekkelig mange) deltakere? Hvor viktig er det at utvalget er representativt?

Etikk

Hvordan sikkerstiller du at deltakernes rettigheter blir ivaretatte?



Ting å tenke på ved Innsamling av egne data

Etikk

Hvordan sikkerstiller du at deltakernes rettigheter blir ivarettatte?

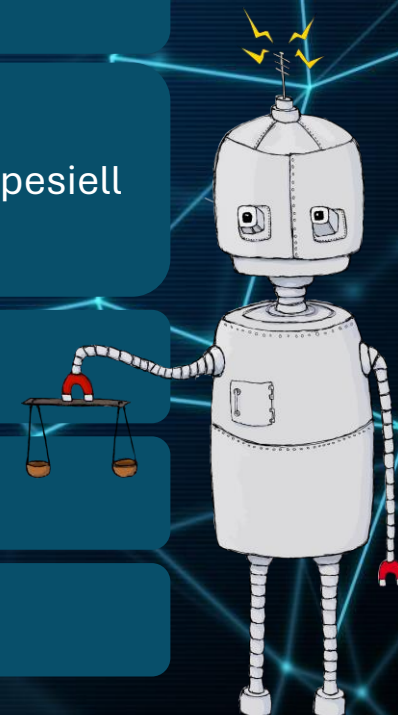
For å respektere menneskeverd og **ivareta personlig integritet, sikkerhet og velferd**, skal forskere informere om forskningen.

Forskere skal som hovedregel få et etisk **samtykke** fra personer som deltar i forskning. Det etiske samtykket skal sikre deltakernes personlige integritet og rett til selv å bestemme om de vil delta. Spesiell hensyn må tas til grupper som har redusert evne til samtykke.

Forskere må sikre at **anonymitet** er ivarettatt hvis det er avtalt, eller hvis andre hensyn tilsier det

Forskningsdata og annet forskningsmateriale skal **lagres og deles forsvarlig**.

Forskere skal vise **respekt for ulike verdier og holdninger**.



Ting å tenke på ved Innsamling av egne data



Etikk

Hvordan sikkerstiller du at deltakernes rettigheter blir ivarettatte?

Informert samtykke: Du må informere deltakerne om hva det innebærer å delta i studien, inklusive hvordan deres data vil bli lagret.

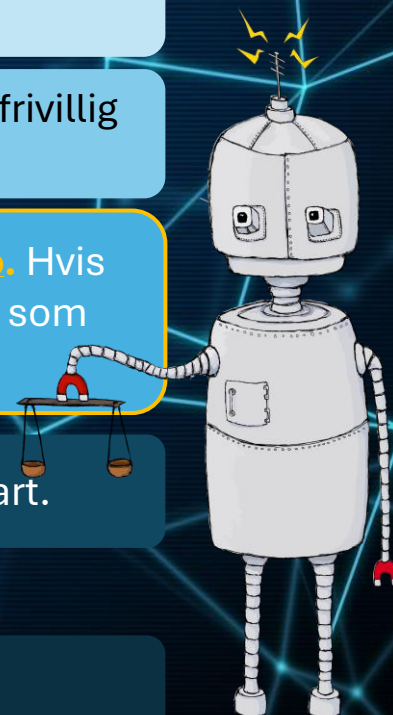
Frivillighet: Det skal alltid være frivillig å delta i forskning. Deltakerne skal informeres om at det er frivillig å delta.

Anonymitet: Hvis data ikke er helt anonyme (direkte eller indirekte) **må søknad sendes til Sikt.no**. Hvis data skal være anonyme må dette sikkerstilles (ikke lagring av IP-adresser, ikke lagring av variabler som sammen kan identifisere et individ).

Lagring: All individdata skal lagres på en måte som sikkerstiller at data ikke kan havne hos tredjepart.

Leverandører av spørreskjema har tilgang til data! Nettskjema.no er godkjent for å bruke til data på individnivå i Norge.

NB: det er ikke vanskelig å sende en søknad til Sikt, men det tar tid! For seint å sende inn i januar dersom du skriver oppgave til våren.



Diskusjonsoppgave

Hvilken type data trengs for å svare på forskningsspørsmålet?

- Aggregerte (makro) eller individdata (mikro)?
- Tverrsnitt, panel, eller tidsserie?

TIPS: Tenk på som er utfallsvariabel og hovedsakelige forklarende variabler

Hvilke datakilder kan du bruke, gitt dine krav på data?

- Finnes det sekundærdata som du kan bruke?
- Kan du samle inn egne data som klarer å besvare forskningsspørsmålet? Hvilken metode vil du bruke for å gjøre dette?