

# **Kapittel 3**

## **Produsere data**

### **Introduksjon**

#### **3.1 Kilder til data**

#### **3.2 Design av eksperimenter**

#### **3.3 Design av utvalg**

#### **3.4 Etikk**

## 3.1 Kilder til data

- Anekdotiske data
- Tilgjengelige data
- Utvalgsundersøkelser og eksperimenter
- Observasjonstudie vs. eksperiment
- Konfundering

Vær forsiktig med å trekke konklusjoner basert på egen eller andres erfaring:

- **Anekdotiske data** representerer enkelthendelser som lett kan fange vår oppmerksomhet fordi de er slående på en eller annen måte.
- Vi har en tendens til å huske disse nettopp fordi de er uvanlige.
- **Det at de er uvanlige betyr at de antageligvis ikke er representative for enhver stor gruppe.**

**Tilgjengelige data** er data som ble produsert tidligere for et annet formål, men som kan hjelpe til med å besvare et aktuelt spørsmål uten store kostnader. Biblioteket og internett er eksempler på kilder til tilgjengelige data.

Noen spørsmål krever data produsert spesifikt med det formål å besvare disse spørsmålene. Dette leder til å **design**e observasjons- eller eksperiment-studier.

# Bare observere eller gi behandling?

Når målet vårt er å forstå årsakssammenhenger er data fra eksperimenter mest overbevisende.

Skillet mellom en observasjonsstudie og eksperiment er viktig:

En **observasjonsstudie** observerer individer og måler variable av interesse, men forsøker ikke å påvirke individene. Målsetningen er å beskrive en gruppe eller situasjon.

Et **eksperiment** utsetter med hensikt individene for en behandling for deretter å måle responsene. Formålet er å undersøke om behandlingen forårsaker en endring i responsen.

- Eksperimenter nøyer seg ikke med å observere individer, ved å f.eks. måle karakteristikk eller stille spørsmål, slik observasjonsstudier gjør.
- Et eksperiment gir aktivt individene en **behandling** for å måle responsen.

# Utvalgsundersøkelser er en viktig type observasjonsstudie

- I en utvalgsundersøkelse blir et **utvalg** av individer valgt fra en større **populasjon** av individer.
- Utvalgsundersøkelser har vanligvis som mål å utforske egenskaper ved individer i populasjonen. Man studerer en liten del av en populasjon for å få kunnskap om populasjonen som helhet.
- Dette er nyttig fordi det i de fleste tilfeller er for kostbart, eller umulig av andre årsaker, å studere hele populasjonen.
- Konklusjoner som trekkes fra et utvalg er gyldige bare når utvalget er trukket på en veldefinert måte, som vi skal se på senere (delkapittel 3.3).

# Godt designede eksperimenter håndterer underliggende variabler og unngår uønsket konfundering

En **underliggende variabel** er en variabel som ikke er blandt forklarings- eller respons-variablene i en studie, men som kan influere respons-variabelen.

**Konfundering** skjer når sammenhengen mellom to variable er slik at deres effekter på en respons-variabel ikke kan skilles fra hverandre.

Observasjonsstudier av sammenhengen mellom to variabler er sjelden egnet til å fastslå årsak og virkning på grunn av **konfundering** mellom forklaringsvariabelen og en eller flere **underliggende variable**.

## 3.2 Design av eksperimenter

- Eksperimentelle enheter og behandlinger
- Komparative eksperimenter
- Randomisering
- Prinsipper for forsøksplanlegging
- Forsiktighetsregler ved eksperimenter
- Blokk-design og matchede par-design

# Eksperiment: En studie der vi påfører behandling på eksperimentelle enheter for å observere responsen. <sup>8</sup>

En **eksperimentell enhet** er objektet/individet som behandlingen gis til (mennesker, dyr, planter eller objekter).

Forklaringsvariablene i et eksperiment kalles ofte **faktorer**, og verdiene de kan ta kalles for **nivåer**.

En spesifikk eksperimentell tilstand som individene i eksperimentet utsettes for kalles en **behandling**. Hvis eksperimentet har flere forklaringsvariable er en behandling en kombinasjon av spesifikke verdier av disse variablene.

**Utfallet** er variablene som måles for å sammenligne behandlingene.



# To faktorer, med hhv to og tre nivåer, gir totalt 6 ulike behandlinger



Alamy

TV-reklame: Effekt av hyppighet og varighet

		Factor B Repetitions		
		1 time	3 times	5 times
Factor A Length	30 seconds	1	2	3
	90 seconds	4	5	6

**Figure 3.2**

Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e, © 2017 W. H. Freeman and Company

Eksperimenter er den foretrukne metoden for å undersøke en eventuell effekt av en variabel på en annen. **Gode design** er essensielle for vellykkede eksperimenter.

**Placebo-effekten** opptrer når individer har positiv effekt av en behandling uten reell virkning, som en sukkerpille. Det kan være fordi man responderer positivt på enhver behandling, med eller uten virkning. Selve deltagelsen i undersøkelsen blir dermed en underliggende variabel som konfunderer med behandlingen man ønsker å undersøke effekten av.

En studie er **forventningsskjev** hvis den systematisk favoriserer visse utfall.

# Vi kan bare fastslå årsak og effekt ved å <sup>11</sup> **påføre** behandling og samtidig **kontrollere** andre påvirkninger

Eksempel: I et forsøk på å redusere eksamensangst skulle studenter før avsluttende eksamen skrive et essay om følelsene sine rundt den kommende eksamenen.

Studenter → Skrive om følelser → Observer eksamensresultater

Resultatene på denne eksamenen ble sammenlignet med resultatene fra midtveiseksamen. Gjennomsnittskarakteren på avsluttende eksamen var *bedre* enn gjennomsnittskarakteren på midtveiseksamen.

**Er essay-skrivingen effektiv? Hvordan kan man vite det?  
Er du sikker på at økningen kommer av essay-skrivingen?**

For å redusere betydningen av underliggende variable bør man benytte et **komparativt eksperiment** der man sammenligner en kontrollgruppe og en (eller flere) behandlingsgruppe(r)

Eksamensangst-eksperimentet var dårlig designet, vi vet ikke om essay-skrivingen i seg selv hadde en positiv effekt.

De bedre karakterene kan skyldes at den siste eksamenen var enklere enn den første. En annen forklaring kan en mulig **placebo-effekt**.

Sluttresultatet lider av **forventningsskjevheter**.

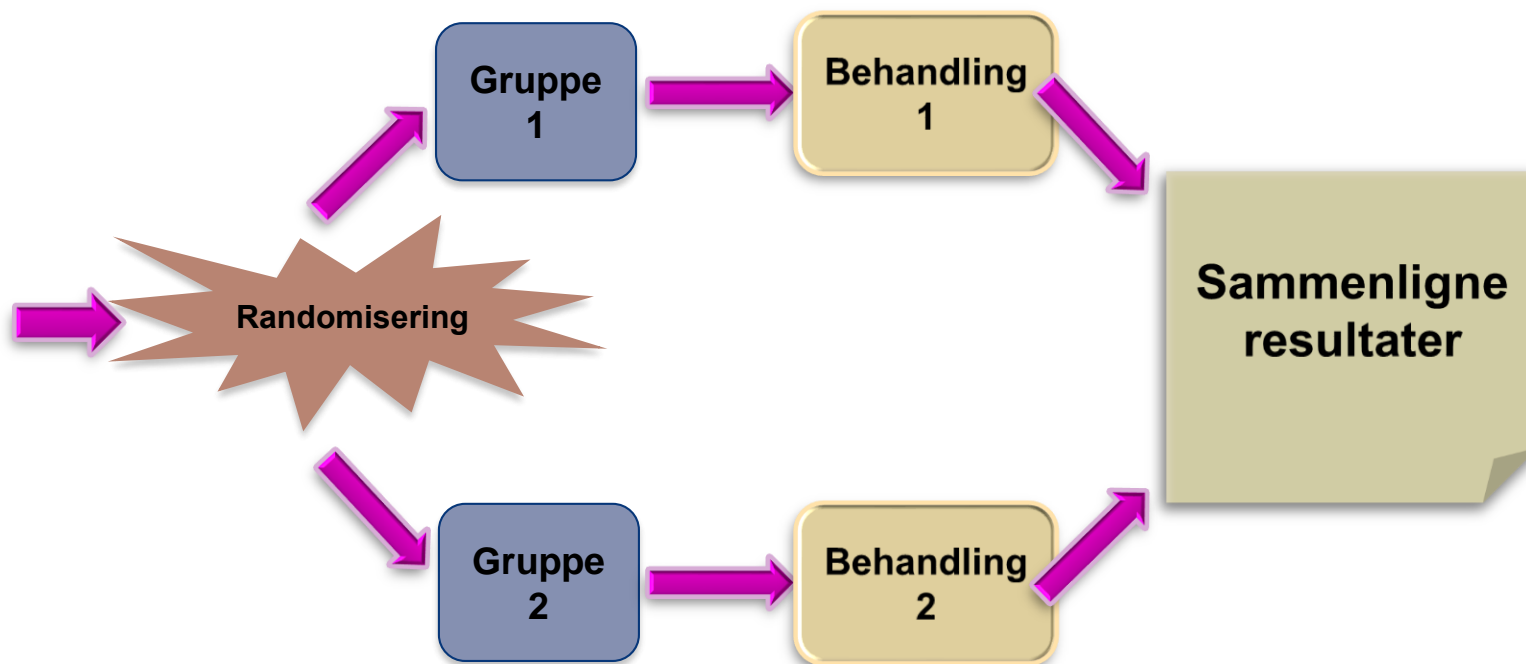
# Randomiserte komparative eksperimenter reduserer forventningsskjevhet

Virkemiddelet mot konfundering er å utføre et **komparativt eksperiment** der noen enheter mottar en behandling og andre, lignende enheter mottar en annen behandling. Deretter sammenligner man utfallene.

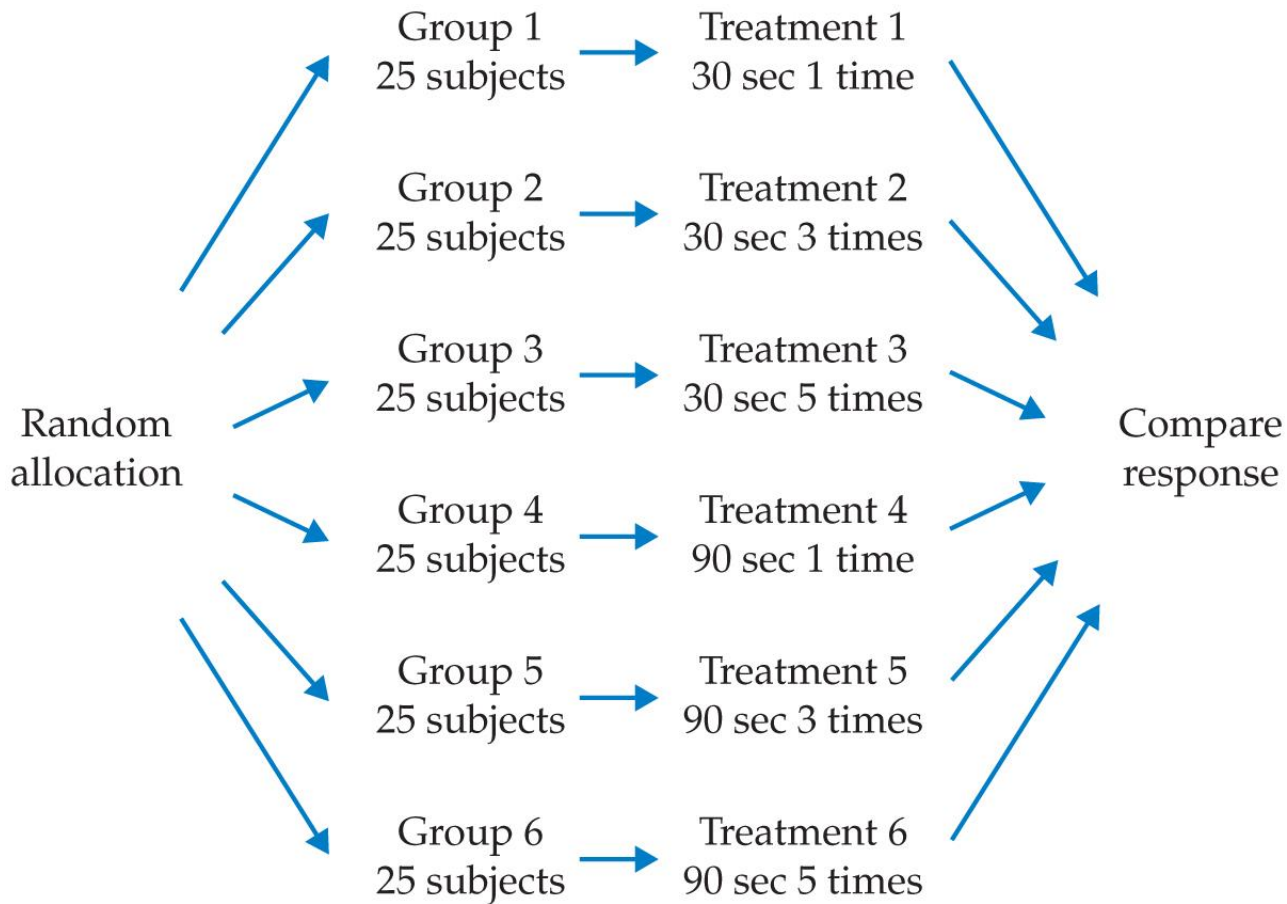
Sammenligning av utfall ved ulike behandlinger alene er ikke nok. Hvis behandlingene gis til grupper som er veldig forskjellige vil man få **forventningsskjevhet**. Løsningen er å **randomisere**.

Det å **randomisere** i et eksperiment betyr at de eksperimentelle enhetene blir tilfeldig plassert i behandlings-grupper. Slik får man grupper med enheter med tilnærmet like egenskaper.

- I et **fullstendig randomisert design** blir behandlingene tilordnet de eksperimentelle enhetene helt tilfeldig.
- Mange eksperimenter (standarden i medisin) har en **kontrollgruppe** som får en behandling uten effekt (eller en eksisterende standardbehandling).



# Eksempel 3.14: Hyppighet og varighet av TV-reklame



**Figure 3.5**

Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e, © 2017 W. H. Freeman and Company

En metode for å **randomisere** et eksperiment er å basere seg på **tilfeldige tall** for å ta nøytrale beslutninger. Vi kan bruke en tabell over tilfeldige tall (slik som tabell B i læreboka) eller en tilfeldig trekkefunksjon i statistisk programvare (som for eksempel R).

## Velge et utvalg på $n$ individer fra en gruppe av størrelse $N$ :

- Nummerer hvert av de  $N$  individene (typisk fra 1 til  $N$ ).
- Se for deg at du skriver alle tallene fra 1 til  $N$  på hver sin identiske papirlapp. Putt så alle papirlappene i en hatt.
- Bland lappene og trekk en lapp tilfeldig.
- Bland de gjenværende  $N - 1$  lappene og trekk tilfeldig en av dem.
- Fortsett på denne måten inntil vi har vårt utvalg på  $n$  lapper.

Du trenger ikke å fysisk gjøre dette, for programvare (for eksempel funksjonen «sample» i R) kan gjøre det for deg!



Randomiserte sammenlignende eksperimenter er designa for å gi gode bevis for at forskjellene vi eventuelt ser i responsene er et resultat av forskjellene i behandlingene.

### Prinsipper for eksperimentell design

- 1. Kontroller** for underliggende variable som kan påvirke responsen, ved å sammenligne to eller flere behandlinger.
- 2. Randomiser:** Bruk tilfeldig plassering av eksperimentelle enheter i behandlingsgrupper.
- 3. Replikasjon:** Bruk nok eksperimentelle enheter i hver gruppe for å redusere variasjon i resultatene som kun skyldes tilfeldigheter.

**I prinsippet kan eksperimenter gi gode bevis for kausalitet.** Vi gjør stegene over for å sikre at den eneste fornuftige forklaringen på de observerte forskjellene skyldes forskjellen på behandlingene som blir gitt.

# Forsiktighetsregler ved eksperimenter

Hvor godt et randomisert komparativt eksperiment avhenger av vår evne til å behandle alle enhetene på samme måte, bortsett fra forskjellen i behandling mellom behandlingsgruppene.

I et **dobbelblindet eksperiment** vet verken enhetene eller de som gir behandlingene og måler responsene hvilken behandling enhetene fikk.

Fullstendig randomiserte design er den enkleste formen for statistisk design av eksperimenter. Men noen ganger er det behov for mer avanserte metoder.

En **blokk** er en gruppe av eksperimentelle enheter som det er kjent før eksperimentet at har en eller annen form for lignende egenskaper som kan påvirke responsen på behandlingene.

I et **blokk-design** blir randomiseringen av eksperimentelle enheter til behandlinger utført separat innen hver blokk.

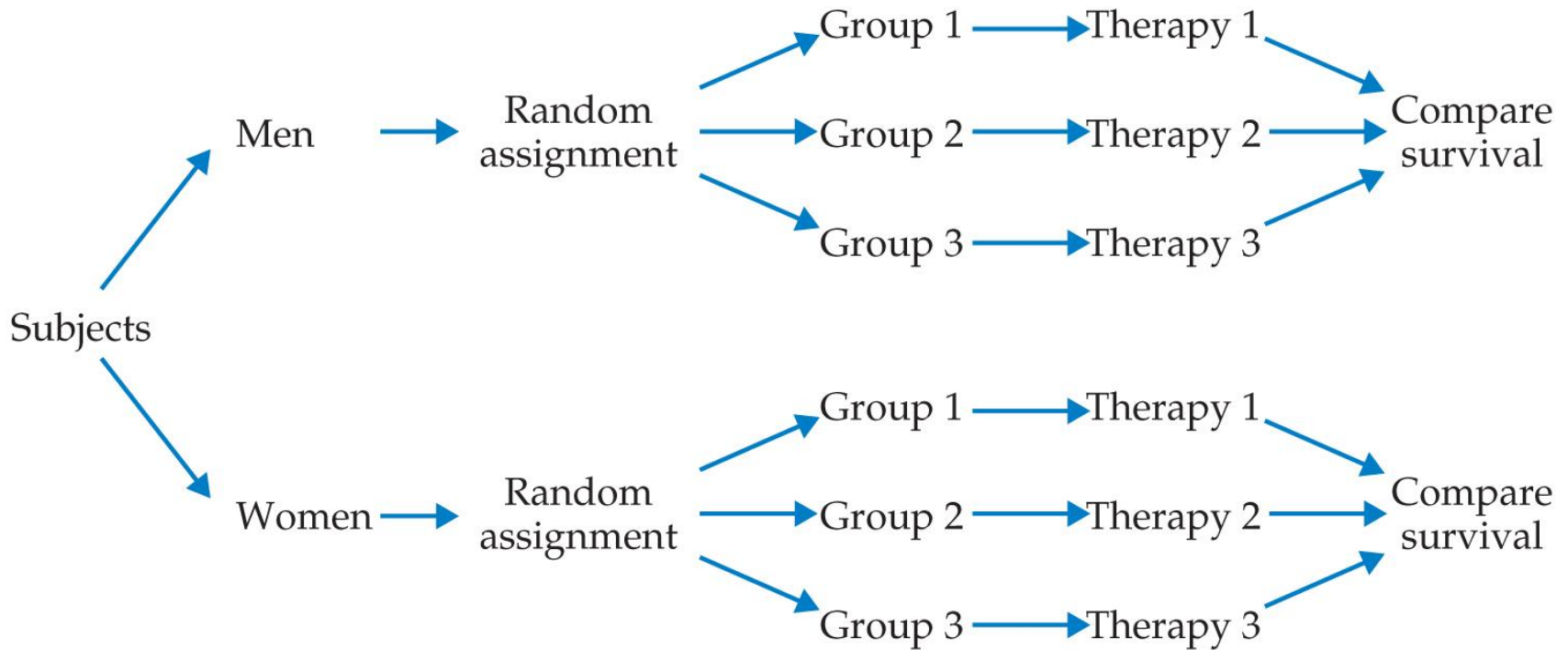
Lag blokker basert på de viktigste uunngåelige kildene til variabilitet (underliggende variable) blant de eksperimentelle enhetene.

Randomisering vil jevne ut effekten av de gjenværende underliggende variablene og muliggjøre en forventningsrett sammenligning av behandlingene.

***Kontroller det du kan, blokk det du ikke kan kontrollere, og randomiser for å lage sammenlignbare grupper.***

# Eksempel 3.18: Blokk-design i sammenligning av kreftbehandlinger

Forløpet for en type kreft er forskjellig for kvinner og menn, derfor behandles kjønn som en blokk-variabel



**Figure 3.6**

Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e, © 2017 W. H. Freeman and Company

En vanlig type randomisert blokk-design for for å sammenligne to behandlinger er et matchede par-design. Ideen er å lage blokker ved å matche par av lignende eksperimentelle enheter.

Et **matchede par-design** er et randomisert blokk-eksperiment der hver blokk består av et matchende par av lignende eksperimentelle enheter.

Enhetene i et par får forskjellig behandling.

Tilfeldighet blir brukt for å bestemme hvilken enhet i hvert par som får hvilken behandling.

Noen ganger kan et “par” i matchede par-design bestå av en enkelt enhet som får begge behandlingene på forskjellige tidspunkter. Fordi rekkefølgen på behandlingene kan influere på responsen blir hvilken behandling som blir gitt først bestemt tilfeldig for hver enhet.

## 3.3 Design av utvalg

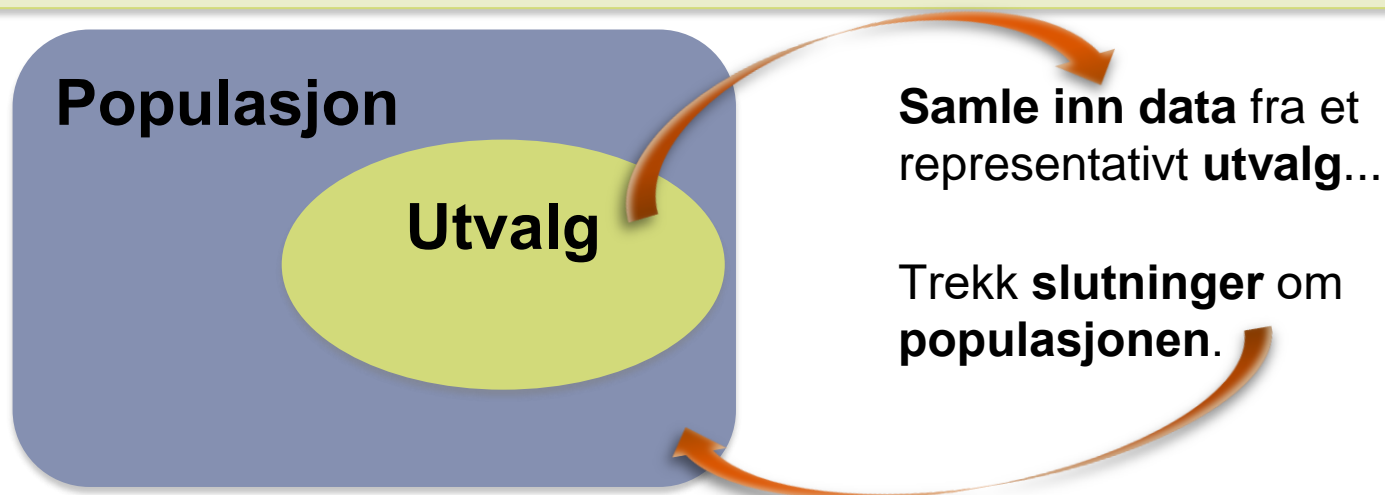
- Populasjon og utvalg
- Dårlige utvalgs-strategier
- Enkle tilfeldige utvalg
- Sannsynlighetsutvalg
- Stratifiserte tilfeldige utvalg
- Forsiktighetsregler ved utvalgsundersøkelser

# Populasjon og utvalg

Forskjellen mellom populasjon og utvalg er sentralt i statistikk. For å sette informasjonen fra et utvalg i sammenheng må man kjenne til hva populasjonen representerer.

**Populasjon** i en statistisk undersøkelse er hele gruppen av individer som vi ønsker å få informasjon om

Et **utvalg** er den delen av populasjonen som vi faktisk henter informasjonen fra. Vi bruker informasjon fra et utvalg til å trekke konklusjoner om hele populasjonen- *en utvalgsundersøkelse* (en viktig form for observasjonsstudie)



Utfordring: Designet til et utvalg er **forventningsskjev** hvis det systematisk favoriserer visse utfall.

Å velge individer bare fordi de er lette å nå resulterer i et **lettevint utvalg**.

Et **frivillig respons-utvalg** består av personer som velger seg selv ved å respondere på en undersøkelse. Frivillig respons-utvalg viser ofte forventningsskjevhet fordi personer med sterke meninger (spesielt negative meninger) kan være mer villige til å svare.



# En ideell løsning: Et enkelt tilfeldig utvalg

**Tilfeldig utvelgelse**, å bruke tilfeldighet til å velge et utvalg, er det sentrale prinsippet i statistisk utvelgelse.

Et **enkelt tilfeldig utvalg (SRS)** av størrelse  $n$  består av  $n$  individer fra populasjonen valgt på en sånn måte at ethvert sett av  $n$  individer har en lik sjanse til å være det utvalget som ble valgt.

I praksis bruker man tilfeldighet generert av en datamaskin (for eksempel funksjonen «sample» i R) for å velge utvalg.

# Sannsynlighetsutvalg og stratifiserte tilfeldige utvalg

**Sannsynlighetsutvalg:** Et utvalg valgt ved tilfeldighet. Den som skal analysere resultatene fra studien må vite

- Hvilke utvalg som er mulige
- Sannsynlighet for de ulike utvalg (ikke nødvendigvis lik for alle)

*Et spesialtilfelle er et enkelt tilfeldig utvalg: Alle utvalg mulige, og sannsynligheten er lik for alle utvalg*

## **Stratifiserte tilfeldige utvalg**

- Del populasjon inn i grupper av lignende individer, kalt **strata**
- Velg et enkelt tilfeldig utvalg innen hver strata
- Kan gi mer eksakt informasjon enn ett enkelt tilfeldig utvalg
- Hvis alle helt like innen hvert strata: Nok med et individ fra hvert strata
- **Strata** for utvalg er tilsvarende til **blokker** i eksperimenter

# Forsiktighetsregler ved utvalgsundersøkelser

Det er viktig med gode metoder for å gjøre utvalg for å redusere feilkilder og forventningsskjevhet. Tilfeldig utvelgelse fra en liste over populasjonen fjerner skjevheter i valget av utvalget. Men: Utvalgsundersøkelser av store populasjoner krever mer enn en god metode for å velge utvalget. Mulige utfordringer:

**Underdekning:** når noen grupper i populasjonen blir utelatt fra utvelgelsen: Viktig med komplett og presis liste over populasjonen!

**Ikke-respons:** når individer i utvalget ikke er tilgjengelige eller ikke ønsker å respondere. F.eks. større manglende respons i urbane strøk enn på landsbygda. Problem: De som ikke svarer kan ha andre egenskaper enn de som svarer, noe som medfører forventningsskjevhet.

Et systematisk mønster av ikke korrekte responser i en utvalgsundersøkelse gir **respons-skjevhet**.

**Formulering av spørsmål** er den viktigste kilden til å skjevhet i svarene/responsen man får fra en utvalgsundersøkelse. Spørsmål kan være ledende og/eller vanskelige å tolke.

# Eksempel på skjevt utvalg: Biologiske egenskaper

- Av interesse: Vekt av lemmen
- Utvalg: Lemmen fanget i feller

**Har lemmen som ikke har gått i fella samme vekt som de som er fanget?**

Innhenting av data reiser ofte etiske spørsmål.

Dette gjelder spesielt når dataene angår personer, men også i andre sammenhenger er det viktig, som i bruk av forsøksdyr.

Derfor er det noen kjøreregler, for eksempel:

- Studier som involverer eksperimenter med dyr og mennesker skal rutinemessig godkjennes av nemnder.
- Deltagerne i undersøkelser skal informeres og samtykke i å delta.
- Dataene skal behandles konfidensielt og ved offentliggjøring skal deltagerne ikke kunne identifiseres. Dette betyr f. eks at tabeller ikke kan være for detaljerte.

# **Kapittel 3**

## **Produsere data- Oppsummering**

### **Introduksjon**

#### **3.1 Kilder til data**

#### **3.2 Design av eksperimenter**

#### **3.3 Design av utvalg**

#### **3.4 Etikk**