

# Statistiska undersökningar, urval och felkällor

## Statistiska undersökningar och urval

Information som behandlas i en undersökning kallas **data**.

Den mängd som undersökningen handlar om kallas **population**. Populationen kan utgöras av t.ex. människor eller föremål.

Om man samlar in data från en hel population kallas det **totalundersökning**.

Totalundersökningar är exakta, men tidskrävande om populationen är stor.

Om man väljer ut och undersöker en mindre del av populationen kallas detta för **stickprovsundersökning**. Den är mindre exakt, men är lättare att genomföra.

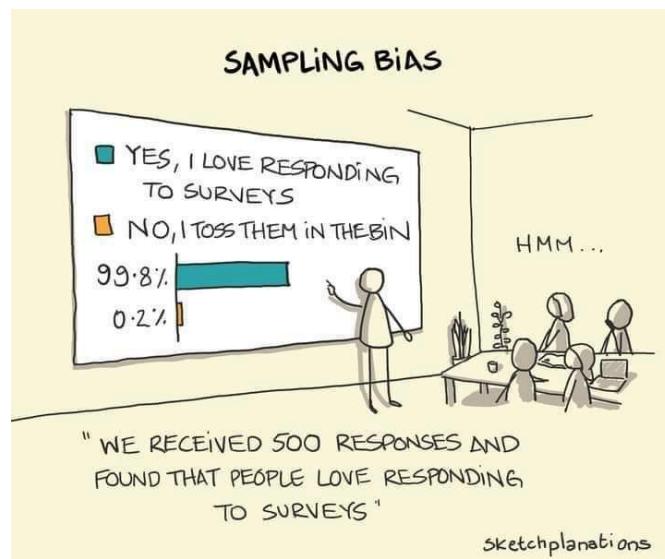
Vid stickprovsundersökningar måste man på något sätt välja vilka i populationen som ska delta i undersökningen. Hur man väljer dessa kallas **urval**.

Det finns tre vanliga urvalsmetoder:

- **Obundet slumpmässigt urval.** Varje individ i populationen tilldelas ett tal. Sedan väljs ett tal med en slumpgenerator. "Dra namnlappar ur en hatt." Urvalet blir helt slumpmässigt.
- **Systematiskt urval.** Man använder ett system för hur man gör sitt urval, t.ex. att välja var femte person från en klasslista. Till skillnad från vad man kan tro är detta sätt inte slumpmässigt (om man inte har slumpat ordningen i klasslistan innan).
- **Stratifierat urval.** Man anpassar valet av individer i stickprovet för att spegla populationen. T.ex. om 75 % av en population är kvinnor, ska även 75 % av stickprovet vara kvinnor, om man tror att kön har inverkan på resultatet.

## Felkällor

Statistiska undersökningar kan ge felaktiga resultat. Orsaker till felaktiga resultat kallas **felkällor**.



Fel som beror på att urvalet inte representerar populationen kallas **urvalsfel**. Ett exempel på det skulle kunna vara att stå utanför en kyrka och ställa frågan "Tror du på Gud?" för att dra en slutsats om gudstron i en stad. Eftersom kyrkobesökare i större utsträckning är religiösa är sannolikheten att de tror på Gud större än för hela populationen i staden. Därför blir urvalet olämpligt.

Fel som gör att de insamlade värdena blir felaktiga kallas **mätfel**. Mätfel kan orsakas av en felgraderad linjal, men även av otydliga eller **ledande frågor** eller svarsalternativ som inte täcker in alla möjliga svar.

- |                                     |                        |
|-------------------------------------|------------------------|
| "Matematik är väl ett roligt ämne?" | – Mycket ledande fråga |
| "Är matematik ett roligt ämne?"     | – Ledande fråga        |
| "Vad tycker du om matematik?"       | – Bra                  |

Om alla som tillfrågas i en undersökning inte svarar får vi ett **svarsbortfall**. Om bortfallet är stort eller om de personer som ingår i bortfallet kan misstänkas svara annorlunda på undersökningen behöver hänsyn till bortfallet tas (detta är fallet i bilden ovan!).

Ex. Du vill undersöka uppfatningen om språk och ställer nedanstående fråga i en enkät:

Vilket språk tycker du är vackrast: spanska eller franska?
<input type="checkbox"/> Katalanska
<input type="checkbox"/> Franska

Finns det några problem med undersökningen?

**Svar:** Ja, svarsalternativen täcker inte in alla möjliga svar. Det är inte säkert att en svarande föredrar något av språken. Det är inte heller säkert att en svarande vet hur t.ex. katalanska låter. Alternativen "Inget" och "Vet ej" saknas alltså.

Ex. En enkät skickades ut till 1 000 personer. Frågan var "Är du positiv eller negativ till utbyggnad av kärnkraft?". Resultatet blev

Positiv: 420 st

Negativ: 380 st

Vet ej: 80 st

- a) Hur stor andel av de som svarade var positiva till utbyggnad av kärnkraft?
- b) Hur stort var bortfallet procentuellt?
- c) Om vi tar hänsyn till bortfallet, mellan vilka värden kan andelen positiva till utbyggnad av kärnkraft ligga?

a) Vi har 420 positiva av  $(420 + 380 + 80) = 880$  svarande. Andelen positiva är då  
 $\frac{420}{880} = 0,477\dots \approx 0,48 = 48\%$

**Svar:** 48 %

b) 1 000 enkäter skickades ut och 880 svarade. Antalet i bortfallet är alltså  $(1\ 000 - 880) = 120$  st.

Det procentuella bortfallet är alltså

$$\frac{120}{1\ 000} = 0,12 = 12\%$$

**Svar:** 12 %

c) Vi får den undre gränsen om ingen i bortfallet är positiv. Antalet positiva är i så fall  
fortfarande 420 och andelen är då

$$\frac{420}{1\ 000} = 0,42 = 42\%$$

Den övre gränsen fås om alla i bortfallet är positiva. Antalet positiva är i så fall  $(420 + 120) = 540$  st, vilket ger andelen

$$\frac{540}{1\ 000} = 0,54 = 54\%$$

**Svar:** Mellan 42 % och 54 %.