SOM-INSTITUTET
SAMHÄLLE OPINION MEDIER

GÖTEBORGS UNIVERSITET

Datakvalitet i SOM-undersökningen

1. Slumpens förträfflighet

Den nationella SOM-undersökningen 2015 har genomförts på ett slumpmässigt befolkningsurval bestående av 17 000 personer.

Ett urval måste vara slumpmässigt draget ur den befolkning som det representerar för att vara generaliserbart. Genom att slumpa fram personerna i undersökningen kan vi nämligen anta att både sådana personegenskaper som går att mäta och egenskaper vars spridning varken går att kontrollera eller säkerställa, får samma spridning i urvalet som i befolkningen. Ju fler personer vi slumpar fram till att delta i undersökningen, desto mer likt befolkningen kan vi anta att urvalet är. Eftersom vi omöjligt kan veta allting om de personer som ingår i befolkningen kan fullständig representativitet inte konstrueras av forskaren.

En annan förutsättning för att ett urval ska vara representativt är att urvalsramen, alltså det register ur vilket urvalet dras, täcker in hela befolkningen. SOM-undersökningens urval är draget ur folkbokföringsregistret vilket innebär att täckningsfelet är minimalt.

Den befolkning som vi uttalar oss om i undersökningen är samtliga personer i åldern 16 till 85 år som är folkbokförda i Sverige. Både svenska och utländska medborgare ingår alltså i urvalet. Eftersom undersökningen genomförs på svenska är ”SOM-befolkningen” dock begränsad till personer som förstår svenska.

Tabell 1. SOM-undersökningens urval

Undersökningens population:	Sveriges vuxna befolkning (16-85 år)
Typ av urval:	Obundet slumpmässigt befolkningsurval
Register som urvalet är hämtat från:	Skatteverkets folkbokföringsregister
Urvalets storlek:	17 000 personer
Sannolikhet att komma med i urvalet:	0,22% eller 1 på 460 personer
Naturligt bortfall:	907 personer
Antal svarande:	8 261 personer
Svarsfrekvens:	51 %

Kommentar: Med naturligt bortfall avses adress okänd; avflyttad; bortrest under fältperioden; bosatt/studerar/arbetar utomlands; ej svensktalande; fysiskt eller mentalt oförmögen att svara; avliden. Dessa personer stryks från urvalet när svarsfrekvensen räknas ut.

► Det är slumpen och urvalets storlek som tillsammans med kvaliteten på urvalsramen gör att svaren från SOM-undersökningen 2015 går att generalisera till hela befolkningen.

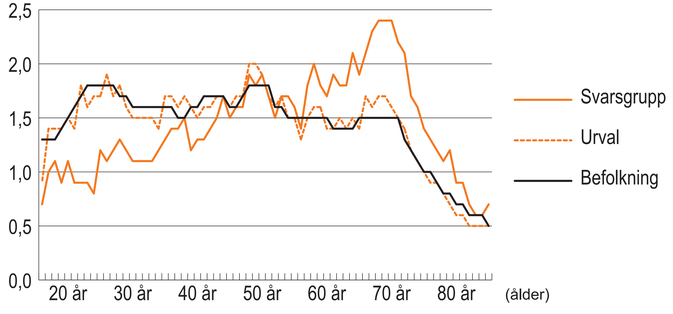
2. Representativitet

Nästa steg i värderingen av datakvaliteten hos en frågeundersökning handlar om vilka som svarar på enkäten – och vilka som inte gör det. En god representativitet i ett urval är inte mycket värd om det finns stora skillnader i sammansättningen hos dem som skickar in en ifylld enkät och dem som låter bli.

Hade alla personer som ingår i urvalet besvarat enkäten hade urvalet och svarsgruppen sett likadana ut. Då hade vi med god säkerhet kunnat säga att också svarsgruppen varit helt representativ för befolkningen. Men så är tyvärr inte fallet i SOM-undersökningarna eller någon annan befolkningsundersökning.

I takt med att svarsfrekvenserna sjunker blir det allt viktigare att följa och analysera konsekvenserna av svarsbortfallet för representativiteten. De mätbara personegenskaper som går att kontrollera mot befolkningsstatistik är få. Det är ålder, kön, civilstånd och medborgarskap. När det gäller kön, civilstånd och medborgarskap är representativiteten god. Det är åldersfördelningen som avviker mest. Unga svarar på enkäten i lägre utsträckning än äldre (se figur 1).

Figur 1. Åldersrepresentativitet i SOM-undersökningen 2015
(procent)



Kommentar: Urvalets åldersfördelning speglar befolkningen perfekt medan svarsgruppen utgörs av färre unga och fler äldre än befolkningen som helhet.

► Representativitet är ett mått på datakvalitet. Hög representativitet ger mer tillförlitliga resultat. SOM-undersökningen 2015 har hög representativitet i alla mätbara egenskaper utom ålder.

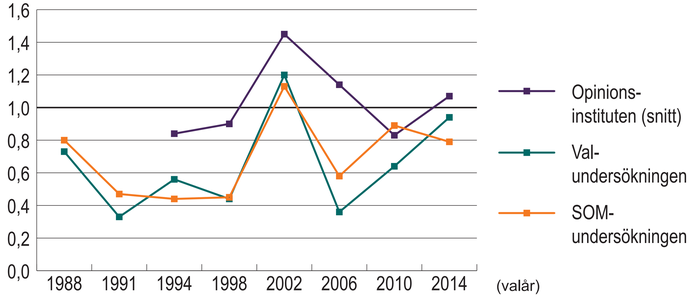
3. Svarskvalitet

Ett sätt att kontrollera vilka konsekvenser den skeva representativiteten får för svarskvaliteten är att studera undersökningens träffsäkerhet. Det grundläggande bekymret i befolkningsundersökningar är att det nästan aldrig finns något ”facit” att bedöma träffsäkerheten mot.

Utfallet i de allmänna valen är dock ett sådant facit. Valåren kan vi nämligen jämföra SOM-undersökningarnas data med de officiella valresultaten. Utfallet från de senaste åtta riksdagsvalen presenteras i figur 2. SOM-undersökningarna har alltid haft bra träffsäkerhet när det gäller utfallet i riksdagsvalen – trots att svarsfrekvensen sjunkit under perioden. Det beror på att SOM-undersökningen använder ett slumpmässigt befolkningsurval – och att vi förlitar oss på slumpen.

Trots att SOM-undersökningen är träffsäker ska nivåskattningar i enkätundersökningar tolkas med försiktighet. Istället för enstaka nivåskattningar är det förändringen över tid som är värd att ta fasta på. Vi kan också anta att de underliggande orsakssambanden, med andra ord hur svarsmönstren i en viss enkätfråga hänger samman med svaret på en annan, gäller även om representativiteten inte är perfekt.

Figur 2. Träffsäkerhet i opinionsmätningar i förhållande till valresultat
(genomsnittlig procentenhetsavvikelse per parti)



Kommentar: Figuren visar den genomsnittliga procentenhetsavvikelsen per parti för olika undersökningar. Opinionsinstituten var 2014 Sifo, Demoskop, Novus, Ipsos, Yougov, Skop och United Minds.

► Ett stort och slumpmässigt urval ger ett robust svarsunderlag även om representativiteten inte är perfekt.