



Surveyföreningen medlemsförening till Statistikfrämjandet

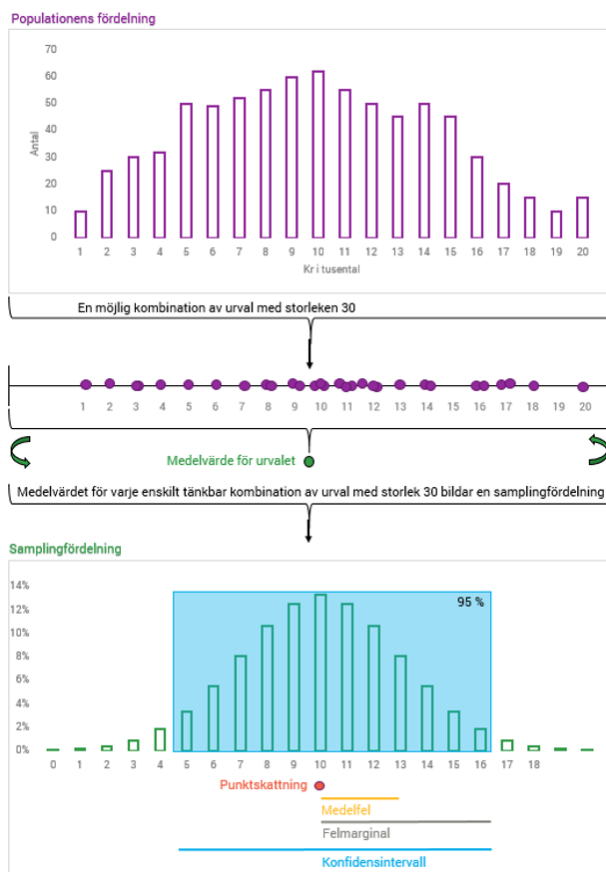


Kort statistisk ordlista

I media refereras ofta till statistiska undersökningar av olika slag. Ibland används statistiska facktermer vilka kan vara svåra att förstå för lekmän och en bred allmänhet. Vissa begrepp och termer är otvetydigt svåra att förstå och kan kräva flera års universitetsstudier för att till fullo förstå. För att ändå försöka bidra till en ökad förståelse för vissa vanliga termer och begrepp har Surveyföreningen tagit fram föreliggande ordlista. Val av ord som ingår är subjektivt och Surveyföreningen är öppen för att lägga till ord om så önskas.

En kort förklaring kring hur ordlistan är uppbyggd:

- Varje ord har *en eller flera* av följande: (i) en definition, (ii) en förklaring och (iii) ett exempel. Dessutom anges den engelska motsvarande termen
- Om en definition finns framgår det genom att direkt efter ordet anges referensen till definitionen, se referenslista sist i dokumentet. Om en referens till definitionen saknas är definitionen Surveyföreningens definition. Notera att den definition som anges är *en* tänkbar definition, andra författare kan ha liknande, men ej identiska, definitioner.
- Vill du ha ett förtydligande kring ett begrepp eller en term, eller vill att vi ska lägga till ytterligare termer, är du välkommen att kontakta Surveyföreningen (epost sekrsurvey@gmail.com)



Nytt ord i version 2022-05-20: Korrelation



Surveyföreningen
medlemsförening till Statistikfrämjandet



Ordlista

Bas	3	Samplingsfördelning	14
Bias	3	Sannolikhetsurval	14
Bortfall	3	Standardvägning	15
Felmarginal	4	Statistisk signifikans	15
Filterfråga	4	Stratifiering	16
Icke-sannolikhetsurval	5	Svarsfrekvens	17
Inferens	6	Urval	17
Konfidensintervall	6	Webbpanelundersökning	18
Korrelation	7	Vägning	19
Kvalitativ undersökning	8	Referenser	21
Kvantitativ undersökning	8		
Longitudinell studie/undersökning	9		
Medelfel	9		
Medelvärde	9		
Median	10		
Mixed-mode	10		
Normalfördelning	11		
Obundet slumpmässigt urval (OSU)	11		
Panelundersökning	11		
P-hacking	12		
Population	12		
Representativ	13		
Respondent	13		



Surveyföreningen medlemsförening till Statistikfrämjandet



Bas

Definition: Den mängd svar/intervjuer som undersökningens resultat bygger på.

Förklaring: En synonym term är svarsmängd, dvs de som har svarat.

Engelsk term: Base.

[Tillbaka till Ordlistan](#)

Bias

Definition: Snedvridning; tillstånd som avviker från vad som väntas. Systematiskt fel [7].

Förklaring: Om det skattningsförfarande som används ger resultat (i långa loppet) som systematiskt avviker från det sanna värdet säger man att skattningsförfarandet är biased (skev).

Man bör dock *inte* använda ett språkbruk som exempelvis att säga att en "undersökning är biased". Om de svarande i en undersökning exempelvis inte har samma köns- och åldersfördelning som i populationen kan man prata om att de svarande är *snedvridna* avseende kön och ålder. Man bör inte säga att de svarande har en bias avseende kön och ålder.

Exempel: Anta att det sanna värdet för en populations genomsnittliga månadslön är 30 000 kr. Vi drar ett sannolikhetsurval för att skatta genomsnittlig månadslön och skattningsförfarandet ger (i långa loppet) ett medelvärde på 27 000. Då är skattningsförfarandet biased. Vanligtvis är det bortfall som kan orsaka att skattningsförfarandet blir biased.

Engelsk term: Bias.

[Tillbaka till Ordlistan](#)

Bortfall

Definition: De individer man har planerat att undersöka, men som man ej fick något svar från. [1].

Förklaring: Bortfall kan uppstå av olika anledningar. Vanligtvis brukar man dela in bortfallet ej anträffade, vägrare samt övrigt bortfall. Ibland används finare indelningar i fler grupper.

Engelsk term: Nonresponse.

[Tillbaka till Ordlistan](#)



Surveyföreningen medlemsförening till Statistikfrämjandet



Felmarginal

Definition: Det tal som läggs till respektive dras från ett punktestimat när man bildar ett konfidensintervall. [7]

Exempel: I ett riksdagsval får X-partiet 8 procent av rösterna. En tid senare vill man veta om det skett en förändring i partisympatierna. Man gör därför en urvalsundersökning och ett procenttal beräknas för X-partiet. Kring detta procenttal beräknas en s.k. felmarginal; andra ord är osäkerhetsmarginal eller osäkerhetstal. Antag att urvalet ger punktestimatet 7 procent (dvs. skattningen av andel som skulle rösta på X-partiet) och felmarginalen 1,6 procentenheter. Konfidensintervallet blir då från 5,4 till 8,6 procent.

Engelsk term: Margin of error.

[Tillbaka till Ordlistan](#)

Filterfråga

Definition: En fråga i ett frågeformulär (eller intervju) som används för att avgöra om personen ska svara på en eller flera efterföljande frågor.

Förklaring: I ett frågeformulär kan det vara så att vissa frågor inte ska besvaras av samtliga. För att styra de svarande till rätt frågor används s.k. filterfrågor. Tillsammans med själva filterfrågan finns en hoppinstruktion som förklarar för respondenten vilka frågor man ska hoppa över vid ett visst svar på filterfrågan.

En närbesläktad term är ordet screeningfråga. Termen screeningfråga avser vanligtvis en *inledande fråga* som används för att avgöra om personen tillhör målgruppen eller inte. I den bemärkelsen fungerar screeningfrågan som ett "filter" för att avgöra om personen ska besvara enkäten.

Exempel: En fråga om resvanor kan lyda "*Har du under de senaste 3 månaderna åkt i kollektivtrafiken?*"

Ja --> Gå till fråga 18

Nej --> Gå till fråga 25

Hoppinstruktionen guidar den svarande till olika frågeblock. De som svarar Ja får olika följdfrågor med start på fråga 18. De som svarar Nej får andra följdfrågor med start på fråga 25.

Ibland behövs inte hoppinstruktionen om filterfrågan består av en (eller få) följdfrågor:

"1a. *Står du som ägare till någon personbil?*"

Ja

Nej

"1b. *Om ja, hur många personbilar står du som ägare till?*" ____ st

"2. *Äger du en båt?*"



Surveyföreningen medlemsförening till Statistikfrämjandet



Filterfrågor kan ibland ge upphov till mätfel i självadministrerade datainsamlingar, t.ex. pappersenkäter. Om en respondent svarar på ett inkonsistent sätt, exempelvis genom att svara *Nej* på filterfrågan om kollektivtrafik och därefter ändå besvara frågorna avsedda för dem som svarar *Ja*, uppstår oklarhet om vad som är det korrekta svaret. Är *Nej*-svaret på filterfrågan det korrekta och svaren på fråga 18 och efterföljande felaktiga? Eller var det så att när respondenten såg fråga 18, och efterföljande, så drog hen sig till minnes att "jag åkte ju kollektivt för en månad sedan" och glömde rätta svaret på filterfrågan?

I datorstödda insamlingsätt, t.ex. telefonintervjuer och webbenkäter, kan man bygga in en funktionalitet som gör att ett sådan inkonsistent svarsmonster inte kan uppstå.

Engelsk term: Filter question

[Tillbaka till Ordlistan](#)

Icke-sannolikhetsurval

Definition: Ett urval där man vid urvalsdragningen inte följt i statistisk teori accepterade regler för slumpmässighet. Ett annat uttryck är icke-slumpmässigt urval. Se vidare sannolikhetsurval. [7]

Förklaring: Vid icke-sannolikhetsurval väljs individerna inte enligt statistiska principer utan enligt mer godtyckliga principer.

I regel vill man dra ett sannolikhetsurval men ibland går inte det. En förklaring till att det inte går kan vara att det saknas register eller listor med kontaktuppgifter för individerna i den population man vill undersöka.

Eftersom urvalet inte är slumpmässigt är det svårt att dra vetenskapliga slutsatser utifrån datamaterialet utan att analysera selektionsprocessen

Om bortfallet är snedvridet är det inte säkert att en undersökning som börjar som sannolikhetsurval faktiskt kan betraktas som det när undersökningen väl är genomförd.

Denna typ av urval tenderar att bli allt vanligare på grund av ökande bortfall och höga kostnader för sannolikhetsurval. Forskning kring icke-sannolikhetsurval pågår i många forskargrupper globalt.

Exempel: Olika exempel på icke-sannolikhetsurval är kvoturval, snöbollsurval, bekvämlighetsurval.

Engelsk term: Non-probability sample.

[Tillbaka till Ordlistan](#)



Surveyföreningen medlemsförening till Statistikfrämjandet



Inferens

Definition: Statistisk inferens betyder slutledning under osäkerhet [10].

Förklaring: Inom statistik används termen inferens när man vill dra slutsatser om en bakomliggande population baserat på ett urval från populationen. I denna situation finns det osäkerhet som beror på urvalet, men även andra osäkerhetskällor så som mätfel och bortfall kan förekomma.

Inferens kan även gälla produktion av registerbaserad statistik, där till exempel saknade värden eller brister i själva registeruppgifterna kan ge osäkerhet.

Ordet inferens används även i allmänt tal och avser då slutsatser om generella förhållanden med utgångspunkt i observerade resultat. Inferens är ett latinskt ord som betyder föra in, sätta in.

Inom statistikutbildning på universiteten kan man läsa hela kurser inom "Inferensteori".

Exempel: De slutsatser man vill dra kan röra deskriptiva parametrar som medelvärden eller procentandelar, eller mer analytiska parametrar som koefficienter i olika modeller, t.ex. regressionsmodeller.

I den svenska arbetskraftsundersökningens (AKU) är görs en skattning av arbetslösheten i Sverige baserat på en urvalsundersökning. Att generalisera resultatet från urvalsundersökningen AKU till den bakomliggande populationen är att göra inferens.

Engelsk term: Inference

[Tillbaka till Ordlistan](#)

Konfidensintervall

Definition: Ett konfidensintervall är slumpmässigt intervall med en given sannolikhet att det innehåller det sanna okända värdet av den statistiska parameter man är intresserad av. [3]

Förklaring: Felmarginalen är det tal som läggs till respektive dras ifrån ett punktestimat. Den övre och den nedre gränsen i detta intervall bildar ett konfidensintervall.

Exempel: Man vill veta något om svenska 10-åringars TV-vanor. Ur populationen alla 10-åringar i Sverige dras ett slumpmässigt urval barn; dessa intervjuas om sitt TV-tittande. Bl.a. får de frågan hur länge de en normal dag tittar på TV. Medelvärdet för urvalet (\bar{x}) visar sig vara lika med 2,6 timmar. Kring detta medelvärde beräknas felmarginalen till 0,2. Det i sin tur ger $2,6 \pm 0,2$ som bildar ett intervall. Intervallet sträcker sig från 2,4 till 2,8 timmar. Talen 2,4 och 2,8 är intervallets ändpunkter; 2,4 är den undre intervallgränsen och 2,8 den övre. Ett annat ord för intervallgräns är konfidensgräns.

Engelsk term: Confidence interval.

[Tillbaka till Ordlistan](#)



Surveyföreningen medlemsförening till Statistikfrämjandet



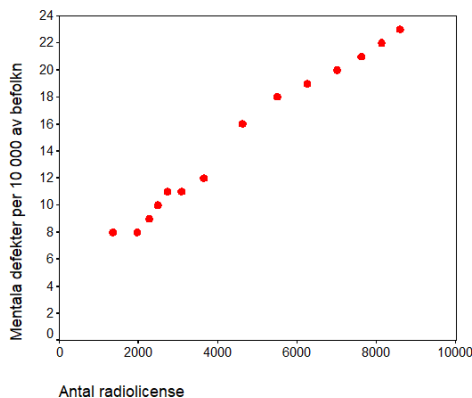
Korrelation

Definition: Inom statistik avser korrelation styrkan av ett linjärt samband mellan variabler [7].

Förklaring: Korrelationen anges ofta med ett tal; korrelationskoefficient som ligger mellan -1 och +1 där 0 innebär avsaknad av *linjärt* samband. Värdet +1 indikerar maximalt *positivt* samband och -1 maximalt *negativt* samband (se exempel nedan). Det är viktigt att understryka att korrelationen *inte* säger något om *orsakssambandet* mellan två variabler. Bara för att två variabler samvarierar (korrelerar) behöver inte betyda att det finns ett orsakssamband mellan dem.

Det finns flera olika sätt att beräkna korrelationskoefficienten. Två vanliga är Pearsons produktmomentkorrelationskoefficient¹ och Spearmans rangkorrelationskoefficient. Om man utan närmare precisering nämner att man beräknat korrelationen är det vanligtvis Pearsons korrelationskoefficient som använts.

I vardagligt tal används ordet korrelation ofta ungefär som synonym för samband. Men man måste då vara observant på förbehållet som nämndes ovan att det inte behöver vara ett *orsakssamband*. Ett klassiskt exempel på så kallad nonsenskorrelation är från perioden 1924–1937 i England (synonymer: skensamband, falsk korrelation). Under den perioden ökade antalet lösta radiolicenser (man behövde licens för radio då) och antalet personer med mentala defekter per 10 000 av befolkningen, se figur nedan. Korrelationen blir 0,99, dvs. nästan ett perfekt linjärt samband. Att nu dra slutsatsen att radiolyssnande kan inverka negativt på den mentala hälsan är naturligtvis rent nonsens. Att båda variablerna samvarierar, där båda ökade under åren 1924–1937, har en naturlig förklaring i en allmän samhällsutveckling.



Exempel: Det finns en *positiv* korrelation mellan längd och vikt; längd och vikt är positivt korrelerade. Samvariationen innebär att ju större längden är desto större är i allmänhet vikten och ju mindre längden är desto mindre är i allmänhet vikten.

Det finns ett *negativt* samband mellan ålder och ögats förmåga att anpassa sig till seende på nära håll. Ju högre åldern i år är desto sämre är i allmänhet denna förmåga och ju lägre åldern är desto bättre är den i allmänhet.

¹ En rolig detalj är att detta ord är bland de längsta ord i svenska språket, 36 bokstäver långt.



Surveyföreningen medlemsförening till Statistikfrämjandet



Engelsk term: Correlation

[Tillbaka till Ordlistan](#)

Kvalitativ undersökning

Definition: Kvalitativ analys innebär en företeelse-, egenskaps- och innebördssökande analys som har som mål att identifiera a) variationen, b) strukturen och/eller c) processen i den identifierade företeelsen, egenskapen eller innebörden. [8]

Förklaring: Den kvalitativa analysen förklaras enklast genom att ställa den i kontrast till den kvantitativa. Den kvalitativa analysens syfte är att identifiera och bestämma icke kända eller otillfredsställande kända företeelser, egenskaper eller innebörder medan man i den kvantitativa analysen undersöker hur på förhand definierade företeelse och dess egenskaper och innebörder fördelar sig i en population, händelser eller situationer. I den kvalitativa analysen ställer man sig frågor som "vad innebär det?" eller "vad handlar det om?" och analysprincipen är abduktiv tillskillnad från den kvantitativa analysen som är deduktiv. I den kvalitativa analysen går man oftast från helhet till del medan man, tvärtom, i den kvantitativa analysen går från del till helhet.

Man kan även förklara den kvalitativa analysen som mer subjektiv och mjuk där den är mer av en konstform än en objektiv vetenskap. Den kvalitativa analysen härstammar från den hermeneutiska skolan (att tolka) och kan jämföras med den kvantitativa analysen som utgår ifrån positivism (kunskap baserad på fakta och sinneserfarenhet).

Kortfattat och något grovt kan man även säga att den kvalitativa analysen använder sig av ord som sitt huvudverktyg medan den kvantitativa analysen använder sig av siffror.

Exempel: Innan man genomför en urvalsundersökning där man känner till väldigt lite om den population man är intresserad av kan man välja ut några få personer och göra mer omfattande intervjuer av kognitiv karaktär. Dessa intervjuer leder till en bredare bild av ev. problem och ger en djupare förståelse för ev. orsaker som kan orsaka mätfel i den kommande urvalsundersökningen.

Engelsk term: Qualitative survey.

[Tillbaka till Ordlistan](#)

Kvantitativ undersökning

Definition: Kvantitativ analys är en företeelse-, egenskaps- och innebördsstyrd analys som har som mål att undersöka hur på förhand definierade företeelser, egenskaper och innebörder fördelar sig mellan olika grupper i en population. Ett annat mål är att undersöka om det föreligger något samband mellan två eller flera företeelser, egenskaper eller innebörder och om dessa eventuellt kan leda till slutsatser om kausala relationer. [8]

Förklaring: Se kvalitativ analys.



Surveyföreningen medlemsförening till Statistikfrämjandet



Exempel: Det flesta urvalsundersökningar är exempel på kvantitativa undersökningar.

Engelsk term: Quantitative survey.

[Tillbaka till Ordlistan](#)

Longitudinell studie/undersökning

Under arbete

[Tillbaka till Ordlistan](#)

Medelfel

Definition: Kvadratroten för variansen för ett skattningsförfarande kallas för medelfelet för skattningsförfarandet. [4]

Förklaring: Ett annat ord för medelfel är standardfel, en direktöversättning av engelskans standard error, ofta förkortat SE. Vardagligt, och förenklat, kan begreppet medelfel tolkas så att det är den avvikelse man i genomsnitt ska räkna med när man med hjälp av en urvalsegenskap uttalar sig om motsvarande populationsegenskap.

Medelfelet beror i huvudsak på variabiliteten i data och på urvalsstorleken. Multiplieras medelfelet med faktorn 1,96 erhålls felmarginalen (med 95 procents konfidens).

Engelsk term: Standard error.

[Tillbaka till Ordlistan](#)

Medelvärde

Definition: Medelvärdet är ett centralmått som används för att beräkna ett genomsnittligt värde för ett urval eller en population.

Förklaring: Det aritmetiska medelvärdet beräknas som summan av observationsvärdena dividerat med antalet observationer.

Medelvärde är ett exempel på centralmått, till skillnad från spridningsmått som exempelvis standardavvikelsen.

Exempel: Medellängd (medelvärdet) för män i Sverige är ca 181 cm och för kvinnor ca 168 cm.

Engelsk term: Mean (arithmetic, geometric and harmonic mean).

[Tillbaka till Ordlistan](#)



Surveyföreningen medlemsförening till Statistikfrämjandet



Median

Definition: Av ett latinskt ord som betyder i mitten befintlig, mellerst. Det mittersta värdet då de observerade mätvärdena ordnats i stigande ordning efter mätvärdets storlek (mittenvärdet). [7]

Förklaring: Median är ett exempel på centralmått, till skillnad från spridningsmått.

Exempel: Om vi har en grupp individer där alla har en inkomst så kommer medianen bli den mittersta individens lön efter att vi sorterat lönen efter storleksordning.

Engelsk term: Median.

[Tillbaka till Ordlistan](#)

Mixed-mode

Definition: Mixed-mode inom kvantitativa undersökningar betyder att mer än en insamlingsmetod används i datainsamlingen.

Förklaring: På svenska används ibland beteckningen "blandad insamling" eller "kombinerad insamling". De vanligaste sätten att samla in data i Sverige är via pappersenkät, webbenkät eller telefon. Om en utvald person kan välja bland två, eller flera, sätt att svara talar man om blandad insamling (mixed-mode).

Även en situation med två *olika* urval med olika insamlingsmetoder kan betecknas som mixed-mode. Exempelvis ett urval från ett befolkningsregister där telefon är insamlingsmetod och ett annat, separat, urval från en webbpanel (webbenkät).

Att använda olika metoder för att *kontakta* de utvalda personerna (t.ex. att vid påminnelser använda vanlig post samt telefon) brukar inte betecknas som mixed-mode.

Det finns flera fördelar med mixed-mode. I litteraturen [11] nämns ökad svarsandel, förbättrad mätning, minskad kostnad, goodwill (tillmötesgå respondenternas önskemål om att kunna välja svarssätt), snabbare datainsamling.

En nackdel med blandad insamling har att göra med mätosäkerhet. Om en fråga ställs på exakt samma i en telefonintervju och i en postal enkät kan man få olika svar beroende på insamlingssätt. Detta är ett känt fenomen vid exempelvis känsliga frågor. Om en enkätfråga avser om personen blivit *utsatt för någon form av brott*, kan man få olika svar i en telefonintervju jämfört med en självadministerad enkät. I en telefonintervju kan det vara så att man är ovillig att lämna ut sådana uppgifter, med följd att man kanske inte svarar helt sanningsenligt. Att ha blandad insamling är alltså utmanande från ett mätosäkerhetsperspektiv.

Exempel: I Energimyndighetens undersökning Energistatistik för småhus är populationen alla småhus för permanentboende. Utvalda fastighetsägare som svarar på undersökningen kan välja mellan att svara på en pappersenkät eller webbenkät.

Engelsk term: Mixed-mode.

[Tillbaka till Ordlistan](#)



Surveyföreningen medlemsförening till Statistikfrämjandet



Normalfördelning

Definition: Normalfördelning är en statistisk sannolikhetsmodell som används för att beskriva variation i mätvärden.

Förklaring: Normalfördelningen är den viktigaste teoretiska sannolikhetsfördelningen i statistik. Vid mätning av fenomen i naturen och i samhället visar det sig ofta att observationsvärdena tenderar att följa ett visst mönster - en normalfördelning. Det kan röra sig om till exempel längden på vuxna människor, vikten på nyfödda barn, mängden nederbörd som fallit under ett dygn.

Engelsk term: Normal distribution.

[Tillbaka till Ordlistan](#)

Obundet slumpmässigt urval (OSU)

Definition: Ett urvalsförfarande som innebär att alla *kombinationer* av n objekt (urvalsstorleken) från en population som består av N objekt har lika stor sannolikhet att bli utvalda. [4]

Förklaring: Definitionen av OSU får som konsekvens att varje objekt har samma sannolikhet att bli utvald och denna inklusionssannolikhet är lika med n/N . Notera att det finns flera olika urvalsmetoder som ger att varje objekt har samma sannolikhet att bli utvald.

Exempel: Om OSU används för att dra ett urval på 1 000 personer bland t.ex. Sveriges befolkning över 18 år kommer samtliga individer att ha samma sannolikhet att komma med i urvalet. Urvalet kommer då ofta att vara en miniatyrkopia av Sverige avseende fördelningen över kön, ålder, födelseland, utbildningsnivå, längd, BMI, partisympati, hårfärg, inställning i fråga X, utgifter på restaurangbesök osv. osv.

Engelsk term: Simple random sample.

[Tillbaka till Ordlistan](#)

Panelundersökning

Definition: En undersökning bestående av samma grupp individer som deltar vid olika mättillfällen över en viss tidsperiod. Det kan vara samma eller olika frågor vid varje mättillfälle.

Förklaring: Ordet panel har olika betydelser i olika sammanhang medan ordet *panelundersökning* är kopplat till statistik. Det finns olika varianter av panelundersökningar, exempelvis rullande eller roterande panel eller webbpanel där alla panelundersökningar



Surveyföreningen medlemsförening till Statistikfrämjandet



uppfyller olika syften. I en panelundersökning är man bl.a. intresserad av att följa individerna över tid för att förklara hur deras åsikter eller beteenden förändras över tid.

Ett av huvudsyftena med en panelundersökning är att kunna generalisera resultat (skattningar) till en större population. Tillfälligt sammansatta paneler som kan ses som kvalitativa undersökningar (expertpanel, smakpanel) är en annan typ av panel med ett annat huvudsyfte än att kunna producera skattningar.

Se även ordet [webbpanelundersökning](#) och [longitudinell studie](#).

Exempel: Den svenska arbetskraftsundersökningens (AKU) är ett exempel på en panelundersökning. Designen innebär i korthet att personer dras slumpmässigt med viss regelbundenhet från registret över totalbefolkningen och de som dras kommer vara med under 2 års tid och svara på undersökningen en gång varje kvartal. Vid varje kvartal fasas en åttondel av urvalet ut ur panelen och en åttondel fasas in (nytt urval).

Engelsk term: Panel survey, panel study.

[Tillbaka till Ordlistan](#)

P-hacking

Definition: Missbruk av den statistiska verktygslådan för att åstadkomma signifikanta resultat.

Förklaring: Det händer att de som tar fram statistiken gärna vill lyckas påvisa statistiskt signifikanta resultat (exempelvis för att driva opinioner) och att man gör det på bekostnad av den vetenskapliga processen. Kan göras både medvetet och omedvetet. Även känt som data dredging, data fishing eller data butchery.

Exempel: Görs exempelvis genom att sätta beslutsregler för signifikanstester post facto eller genom att frångå att visa p-värden och bara rapportera resultaten i text när testvärde inte är lika signifikanta som man hade hoppats på.

Engelsk term: P-hacking.

[Tillbaka till Ordlistan](#)

Population

Definition: Populationen är hela den mängd av objekt som en undersökare valt att undersöka.

Förklaring: Oftast får man nöja sig med de objekt som det finns i en förteckning eller ett register. Registret kan mer eller mindre stämma överens med den population man valt att undersöka.



Surveyföreningen medlemsförening till Statistikfrämjandet



Ett fackord för förteckningen (registret) är ram eller urvalsram. Populationen kan vara stor (t.ex. ett lands befolkning) eller liten (t.ex. populationen av biskopar i Sverige).

Exempel: Ett viktigt arbete vid undersökningar är att *avgränsa* populationen på något lämpligt sätt. Man är intresserad av kvinnor i åldrarna 20-39 bosatta i Halland. En tänkbar ram för en sådan population är folkbokföringsregistret. Ibland görs avgränsningar av populationen så att definitionen av populationen passar en tillgänglig ram.

Engelsk term: Population.

[Tillbaka till Ordlistan](#)

Representativ

Definition: Termen representativ eller representativitet är ett något oklart begrepp, som kan ha olika betydelser. Den som säger att ett stickprov (urval) är representativt menar i regel att stickprovet i någon icke specificerad mening "liknar" eller "kan utgöra bas för slutsatser rörande" den population som man är intresserad av. Att ett stickprov är representativt för en viss population skulle exempelvis kunna betyda något av följande:

1. Stickprovets fördelning med avseende på olika bakgrundsvariabler är densamma som motsvarande fördelning i hela populationen (kön, ålder, utbildning, stad/land etcetera.)
2. Alla intressanta grupper i populationen finns representerade i stickprovet.
3. Varje enskild medlem i stickprovet är en typisk representant för populationen.
4. Urvalet har gjorts så att alla i populationen har haft lika chans att komma med.
5. Urvalet har gjorts så att det går att med lämpligt vägningsförfarande erhålla bra skattningar av de sökta populationsparametrarna. [6]

Förklaring: Eftersom "representativ" är ett mångtydigt begrepp, gör ISO 26362 [3] följande uttalande (avsnitt 4.4.1; i svensk översättning):

Begrepp som "representativ" får användas endast om det tydligt definieras.

Engelsk term: Representative.

[Tillbaka till Ordlistan](#)

Respondent

Definition: Personen som ska besvara en enkät/intervju eller som skall observeras kallas för respondent. [1]

Förklaring: Respondenten behöver inte vara en fysisk person, det kan vara en myndighet eller ett företag. Ofta används termen uppgiftslämnare som synonym till respondent.

Engelsk term: Respondent.



Surveyföreningen medlemsförening till Statistikfrämjandet



[Tillbaka till Ordlistan](#)

Samplingfördelning

Definition: Fördelning för en urvalsegenskap som antas beräknad för *alla möjliga urval* av viss storlek ur en population. [7]

Förklaring: Urvalsegenskapen kan t.ex. vara medelvärde, total, varians eller korrelationskoefficient. Om det mått som används är medelvärde är samplingfördelningen för medelvärdet den teoretiska fördelning som skulle uppstå om samtliga tänkbara urval av en viss storlek (och med en viss urvalsmetod) drogs ur populationen. För vart och ett av dessa teoretiska urval behöver man kunna beräkna medelvärdet baserat på de utvalda individerna. Notera att samplingfördelningen är en *teoretisk* fördelning. Det är med hjälp av samplingfördelningen som t.ex. konfidensintervall kan beräknas. Se figur på första sidan, den understa figuren är samplingfördelningen.

Engelsk term: Sampling distribution.

[Tillbaka till Ordlistan](#)

Sannolikhetsurval

Definition: Sannolikhetsurval innebär (1) att det finns en entydigt definierad population från vilken urval ska göras med hjälp av en viss urvalsram, och (2) att urval görs från denna urvalsram med användande av någon av oss själva kontrollerad slumpmekanism (så kallad "randomisering"), så att varje individ i urvalsramen får en känd sannolikhet, större än noll, att bli utvald. [6]

Förklaring: Ett annat uttryck är slumpmässigt urval. Metodiken för sannolikhetsurval är mycket flexibel och kan anpassas till många praktiska situationer. Hit hör tillämpning av flerstegsurval, klusterurval, stratifierat urval, systematiskt urval och användande av hjälpinformation.

Exempel: Sverige har drygt sju miljoner röstberättigade; de utgör populationen röstberättigade. För en opinionsmätning slumpas, säg, 1 000 av de röstberättigade ut för intervjuer. De utvalda utgör ett slumpmässigt urval.

Engelsk term: Probability sample.

[Tillbaka till Ordlistan](#)



Surveyföreningen medlemsförening till Statistikfrämjandet



Standardvägning

Definition: Statistisk metod som gör att olika mått, ofta medelvärden, kan jämföras mellan grupper även om grupperna har olika fördelning över en viktig bakgrundsvariabel som påverkar måttet.

Förklaring: Standardvägning innebär att exempelvis medelvärden som jämförs mellan grupper standardiseras på ett sådant sätt att jämförelse blir möjlig. Om en bakomliggande variabel, t.ex. åldersfördelning eller yrkeskategori, påverkar medelvärdet kan det vara så att en jämförelse av medelvärden mellan grupper är olämplig eller direkt felaktig. Som att jämföra äpplen och päron. Det finns olika tekniska metoder för standardvägning. En metod kallas standardpopulationsmetoden (eller direkt standardvägning) en annan kallas kapacitetsmetoden (eller indirekt standardvägning).

Exempel: Löneskillnaden mellan kvinnor och män förklaras till stor del av att kvinnor och män i huvudsak arbetar inom olika yrken och att lönerna varierar kraftigt mellan dessa yrken. Att jämföra medellönen för alla män med medellönen för alla kvinnor är därför inte relevant, det speglar endast att män och kvinnor arbetar inom olika yrken. Standardvägning innebär i korthet att medellöner för män respektive kvinnor beräknas med antagande om att antalet kvinnor och antalet män är lika många inom varje yrkesgrupp.

Enligt en LO-rapport [2] (diagram 3.3 på sid 26) medför standardvägning efter klass, sektor och yrke att kvinnors medellön i procent av mäns medellön ökar från 88 procent till 96 procent bland samtliga anställda.

Ett annat aktuellt exempel är jämförelsen av dödstal mellan länder under corona-pandemin. Olikheter i dödstalen beror bland annat på hur inrapportering sker (endast sjukvård eller sjukvård och äldreboenden), definition av dödsorsak, var i fasen landet befinner sig mm. Detta i sig utgör svårigheter vid jämförelse av dödstal mellan länder. Dessutom påverkas dödstalen av befolkningens åldersstruktur. Om land A i jämförelse med land B har en mer ålderstigen befolkning är det naturligt att dödstalet är högre eftersom äldre personer drabbas hårdare. Denna olikhet i åldersstruktur kan man dock justera för genom att utgå från dödstalen i olika åldersklasser för land A och B och sedan anta att länderna har samma åldersfördelning när dödstalet för hela landet ska beräknas, dvs. en standardvägning utifrån ålder. Görs en standardvägning blir dödstalen något mer jämförbara, men andra olikheter (som nämnts) kvarstår och försvårar jämförelsen.

Engelsk term: Direct standardization, indirect standardization.

[Tillbaka till Ordlistan](#)

Statistisk signifikans

Definition: Signifikans är inom statistiken ett begrepp för att ange att ett observerat värde i en undersökning avviker från ett hypotetiskt värde eller annat jämförelsevärde så pass mycket att det sannolikt inte beror på den statistiska osäkerheten (slumpen). [9]



Surveyföreningen medlemsförening till Statistikfrämjandet



Förklaring: Två andra uttryck för sådan prövning är signifikansprövning och signifikanstest. Ordet signifikans med adjektivet signifikant kommer av ett latinskt ord som betyder beteckna, betyda. Adjektivet kan vardagligt översättas med betydelsefull.

Statistiskt signifikant och statistiskt säkerställd (skillnad) kan användas synonymt.

Notera att man *inte* bör använda ett språkbruk och säga att en *undersökning* är statistiskt signifikant eller statistiskt säkerställd. Termerna bör användas i samband med undersökningens resultat och de variabler som man undersökt.

Exempel: I ett riksdagsval får X-partiet 8 procent av rösterna. En tid senare vill man veta om det skett en förändring i partisympatierna. Man gör därför en urvalsundersökning och ett procenttal beräknas för X-partiet. Kring detta procenttal beräknas en s.k. felmarginal; andra ord är slumpmarginal och osäkerhetstal. Antag att urvalet ger punktestimatet 7 procent och felmarginalen 1,6 procentenheter. Konfidensintervallet blir då från 5,4 till 8,6 procent. Det här innebär att det inte finns någon statistiskt säkerställd skillnad eftersom intervallet innehåller resultatet 8 % från riksdagsvalet.

Engelsk term: Statistical significance.

[Tillbaka till Ordlistan](#)

Stratifiering

Definition: Stratifiering av en population innebär att populationen delas in i (urvals-) grupper, benämnda strata (stratum i singularis). Varje objekt ingår endast i ett stratum.

Förklaring: Syftet med att stratifiera populationen innan ett urval dras är i huvudsak

- (i) För att få bättre säkerhet i skattningarna för hela populationen
- (ii) För att ha kontroll över att tillräckligt många objekt från olika redovisningsgrupper kan väljas ut, exempelvis unga kvinnor i Uppsala.

Inom varje stratum dras ett urval, oberoende av urvalen i andra strata. Urvalsmetoden kan vara olika i olika strata, men det vanligaste är att samma urvalsmetod används i samtliga strata (till exempel urvalsmetoden OSU).

Om en population stratifieras behöver den totala urvalsstorleken fördelas (allokeras) på något sätt över samtliga strata. Med exempelvis en planerad total urvalsstorlek på 1 000 objekt och 10 strata behöver de 1 000 allokeras över 10 strata enligt någon princip. Därefter dras ett urval inom varje stratum. Det finns olika allokeringsprinciper: Optimal allokering, proportionell allokering, likformig allokering med flera.

Vid stratifiering är det vanligaste att de utvalda objekten har olika urvalssannolikhet (se även ordet [Vägning](#)). Lika urvalssannolikhet uppstår endast vid proportionell allokering.



Surveyföreningen medlemsförening till Statistikfrämjandet



Kriterium (ii) ovan gör också att man kan ha kontroll över säkerheten i skattningar av små men viktiga delgrupper som kan behöva särstuderas i populationen – t.ex. unga kvinnor i en given storstad.

Notera att en stratifiering av populationen endast innebär en indelning i urvalsgrupper, inte en urvalsmetod. Om man ska beskriva hur ett urval genomförts räcker det alltså inte att säga att "vi gjorde ett stratifierat urval", eftersom detta endast beskriver att populationen delades in i strata. Ofta används dock urvalsmetoden OSU inom varje stratum, även om det inte explicit nämns.

Exempel: Den nationella trygghetsundersökningen, som Brå genomför, stratifierar populationen efter variablerna ålder i 2 grupper (16 - 29 år och 30 - 84 år) och 94 lokalpolisområden. Detta innebär $2 \times 94 = 188$ strata. Den totala urvalsstorleken på ca 200 000 personer allokeras lika per lokalpolisområde och därefter proportionerligt i åldersgrupperna 16–29 år och 30–84 år, med hänsyn tagen till förmodade svarsandelar i dessa grupper.

Engelsk term: Stratification (stratified sampling).

[Tillbaka till Ordlistan](#)

Svarsfrekvens

Definition: Antalet som svarar i relation till antalet utvalda i en undersökning.

Förklaring: En synonym term är svarsandel - vilket egentligen är en bättre term eftersom måttet ofta anges som ett procenttal och termen frekvens i svarsfrekvens associeras till ett tal eller heltal.

I skriften *Standard för bortfallsberäkning*, av Surveyföreningen [5], ges förslag och rekommendationer kring hur svarsandel bör beräknas. Situationen som förespeglas är vid sannolikhetsurval. Bortfallsandel är 1 minus svarsandel.

För beräkning av deltagarandel, som är ett motsvarande mått vid webbpanelundersökningar, hänvisas till [6].

Exempel: I en undersökning där 1 000 personer slumpmässigt valts ut och 600 personer svarar så är svarsandelen 60 %. Detta innebär att bortfallsandelen är 40 %.

Engelsk term: Response rate.

[Tillbaka till Ordlistan](#)

Urval

Definition: I statistik, den del av objekten i en population som valts ut för en undersökning. [7]



Surveyföreningen medlemsförening till Statistikfrämjandet



Förklaring: På fackspråk talar man om ett urval som en delmängd av populationen. Två andra ord är stickprov och sampel. Se figur på första sidan.

Vid opinionsundersökningar drar man ett urval för att kunna generalisera till populationen. Urvalet kan vara draget som ett sannolikhetsurval eller som ett icke sannolikhetsurval.

I vissa sammanhang förekommer termerna *inbjudet urval* (invited sample) och *deltagande urval* (participating sample). Detta är ett språkbruk som Surveyföreningen avråder från. Termen *inbjudet urval* är en synonym till urval, eftersom urvalet avser *samtliga* utvalda, dock behövs inte förstärkningsordet "*inbjudet*". Termen *deltagande urval* är närmast att betraktas som svarandemängden, dvs. de som svarar på undersökningen.

Engelsk term: Sample.

[Tillbaka till Ordlistan](#)

Webbpanelundersökning

Definition: En *webbpanel* är en databas innehållande individuppgifter om personer, som förklarat sig villiga att medverka som respondenter i framtida undersökningar via webben, ifall de blir utvalda. En *webbpanelundersökning* innebär att ett urval från panelen (eller från en delmängd av panelen) dras och de utvalda individerna ombeds delta i en undersökning.

Förklaring: En webbpanel skapas i syfte att under en längre tid tjäna som urvalsram i ett antal olika undersökningar. Prefixet "webb" i detta sammanhang markerar att all datainsamling sker via internet och sättet att svara kan t.ex. vara på en dator, mobiltelefon eller surfplatta. Andra termer, som används i samma betydelse som "webbpanel", är "internetpanel" och "online-panel". Begreppet "accesspanel" förekommer också, men det är ett vidare begrepp, som innefattar även paneler där datainsamling sker på annat sätt än med internet.

Det finns idag i samhället en stor mängd med register med e-postadresser. Vissa av dessa är endast ett register för att kunna komma i kontakt med personer, exempelvis ett medlemsregister i en idrottsförening eller ett kundregister på ett företag. Om en undersökning görs genom att skicka ut en inbjudan till e-postadresser i ett sådant sammanhang är det *inte* att betrakta som en webbpanelundersökning. Gränsen är inte helt entydig när ett register kan kallas en webbpanel. Det finns dock några grundläggande kriterier som bör vara uppfyllda. Deltagare i webbpanelen ska på något sätt ha *rekryterats* och *accepterat att delta i framtida undersökningar*. Dessutom bör deltagarna ha genomfört en s.k. *profilundersökning*. En profilundersökning är ett försteg innan en person får delta i panelen. I den ombeds personen svara på ett antal bakgrundsfrågor t.ex. kön och ålder men även livsstilsfrågor som om man är rökare eller inte, har hund eller inte, vilket bilmärke man äger (om man äger bil). Dessa frågor kan användas som avgränsningskriterier i kommande undersökningar.

Det finns i huvudsak två olika sätt att rekrytera personer till en webbpanel (i) via [sannolikhetsurval](#) av personer och via (ii) självrekrytering. Med självrekrytering avses att en



Surveyföreningen medlemsförening till Statistikfrämjandet



person kan rekrytera sig själv, ofta genom att klicka på en länk på en webbsida. För att locka personer att ställa upp i en självrekryterad panel används ofta argument som "gör din röst hörd" och att man lockar med erbjudanden. Ur statistisk-metodologisk synvinkel är det en stor skillnad mellan dessa två sätt att rekrytera webbpaneler. För sannolikhetsurval finns en etablerad vetenskaplig teori, medan man vid självrekrytering är ute på vetenskapligt mer osäker mark. I praktiken förefaller det dock vara så att det i båda fallen görs undersökningar av skiftande kvalitet.

Många traditionella undersökningsföretag har webbpaneler, men även andra typer av företag t.ex. mediebolag kan ha egna webbpaneler.

Antalet webbpanelundersökningar har ökat mycket under de senaste åren.

Surveyföreningen gav år 2014 ut en skrift, se [6], med titeln *Kvalitet i webbpanelundersökningar – metoder och mått*. I skriften föreslås ett antal mått som beskriver olika dimensioner av kvalitet i en webbpanelundersökning.

Exempel: Många opinionsundersökningar idag görs via webbpaneler.

Engelsk term: Web panel. Online panel.

[Tillbaka till Ordlistan](#)

Vägning

Definition: Om urvalet är draget som ett sannolikhetsurval har alla utvalda individer en given urvalssannolikhet (inklusionssannolikhet). 1 dividerat med urvalssannolikheten ger en vikt (designvikt) som används för att "räkna upp" eller "väga" urvalet till populationen.

Förklaring: Termen uppräkningsvikt eller basvikt används ofta synonymt med designvikt. Termen designvikt är tänkt att associeras med att vikten beror på vilken urvalsdesign som använts.

Om bortfall inträffar kan man justera designvikten för att kompensera för eventuella snedvridningar som bortfallet kan ha orsakat. Detta kräver tillgång till (stark) hjälpinformation, ofta i form av variabler som kön, ålder, utbildning, inkomst m.fl. Om de svarande uppvisar skev fördelning över sådana hjälpvariabler i förhållande till populationens fördelning kan designvikten justeras för att kompensera för denna snedvridning. Efter en sådan justering uppvisar de svarande, vid vägd sammanställning, *samma* fördelning över hjälpvariablerna som i populationen. Syftet med detta är reducera eventuell bias i skattningsförfarandet. Poststratifiering eller kalibrering är vanliga metoder för att åstadkomma detta tekniskt.

Inom opinionsmätningar talar man ibland om vägning av resultat. Syftet är även här att reducera bias i skattningarna, t.ex. på grund av bortfall. Inom opinionsundersökningar används ibland hjälpvariabeln "hur man röstade i föregående val" för att justera uppräkningsvikten.



Surveyföreningen *medlemsförening till Statistikfrämjandet*



Om urvalet *inte* är draget som ett sannolikhetsurval vet man inte urvalssannolikheten för individerna. Det existerar då ingen designvikt och inte heller någon (designbaserad) statistisk teoretisk grund att stå på för att göra uppräknings till populationen.

Exempel: Om vi har ett urval med 1000 personer där 500 är män och 500 är kvinnor. Anta att 250 män svarade och 450 kvinnor svarade. Då är kvinnorna relativt sett överrepresenterande. Vid ett vägningsförfarande kan man "väga upp" männens svar och "väga ner" kvinnornas svar.

Engelsk term: Weighting.

[Tillbaka till Ordlistan](#)



Surveyföreningen
medlemsförening till Statistikfrämjandet



Referenser

- [1] Dahmström, K. (1996). *Från datainsamling till rapport – att göra en statistisk undersökning*. Studentlitteratur.
- [2] Larsson, M. (2019). Lönerapport 2019 - Löner och löneutveckling år 1913–2018 efter klass och kön. LO, enheten för avtalsfrågor.
- [3] SS-ISO 26362:2009 Accesspaneler inom marknads-, opinions- och samhällsundersökningar - Vokabulär och servicekrav (ISO 26362:2009, IDT)
- [4] Särndal, C-E., Swensson, B. och Wretman, J. (1992). *Model assisted survey sampling*. Springer-Verlag.
- [5] Surveyföreningen (2005). Standard för bortfallsberäkning. Finns på webben på <https://statistikframjandet.se/survey/arkiv/en-standard-for-berakning-av-bortfall/>
- [6] Surveyföreningen (2014). Kvalitet i webbpanelundersökningar – metoder och mått. Finns på webben på <https://statistikframjandet.se/survey/arkiv/kvalitet-i-webbpanelundersokningar-2/>
- [7] Vejde, O. och Leander, E. (2000). *Ordbok i statistik*. Olle Vejde förlag. Finns i förkortad variant på webben på <http://www.ollevejde.se/statistikord/>
- [8] Starrin & Svensson (1994). *Kvalitativ metod och vetenskapsteori*. Studentlitteratur.
- [9] Wikipedia. <https://sv.wikipedia.org/>. Sökord: Signifikans. Sidan uppdaterades senaste 21 april 2021 kl. 17.21.
- [10] SCB. Vad är statistik? <https://www.scb.se/hitta-statistik/artiklar/2016/Vad-ar-statistik/>. Sidan uppdaterades senaste 3 november 2016.
- [11] SCB (2016). *Frågor och svar - om frågekonstruktion i enkät- och intervjuundersökningar*. SCB-Tryck, Örebro. ISBN 978-91-618-1653-8