**Hur många måste man fråga?**

# C:\Documents and Settings\aehmge\Local Settings\Temporary Internet Files\Content.IE5\EZK7MCCX\MC900390990[1].wmf

Har jag chans på dig?

©Martin Gellerstedt

2011

Innehåll

[1. Inledning 2](#_Toc285139330)

[2. En formell stickprovsberäkning 3](#_Toc285139331)

[3. Hur många enkäter? 6](#_Toc285139332)

[3.1 Stickprov baserat på: Medelvärde 6](#_Toc285139333)

[3.2 Stickprov baserat på: Procentsats % 7](#_Toc285139334)

[4. Något om kategorier och bortfall 10](#_Toc285139335)

[4.1 Att dela upp resultatet i kategorier 10](#_Toc285139336)

[4.2 Bortfall 10](#_Toc285139337)

[5. Signifikans inte alltid tillräckligt och kort om etik 11](#_Toc285139338)

[5.1 Signifikans inte alltid övertygande? 11](#_Toc285139339)

[5.2 Kort om etik 11](#_Toc285139340)

[6. Intervjuer 12](#_Toc285139341)

[7. Summering 15](#_Toc285139342)

# Inledning

En av de vanligaste frågor man får som statistiker handlar om en undersöknings storlek. Ur många försöksobjekt ska studeras? Hur många ska man intervjua? Hur många enkäter ska sändas ut?

Dessvärre finns det inget snabbt svar på denna fråga. För att utföra en detaljerad beräkning av lämplig stickprovsstorlek krävs relativt god kunskap i statistik. Vidare kräver en detaljerad stickprovsberäkning för det mesta att man även har tillgång till tidigare data kring det fenomen som ska studeras. Denna situation är relativt vanlig i medicinska studier och en formell stickprovsberäkning brukar vara standard att bifoga ansökan till forskningsetisk prövningsnämnd.

Har man ingen tidigare empiri är situationen betydligt mer komplicerad och gissningsartad. Det finns de som rekommenderar att man alltid bör utföra en undersökning i liten skala först, en så kallad pilotstudie, för att på så sätt få ett hum om hur det fenomen man vill studera ser ut och möjliggöra bra uppskattningar av hur stor den efterföljande fullskaliga undersökningen bör göras (man får även bra information om hur frågor och skalor fungerar – så rådet om pilotstudie är inte dumt…).

Men, dessvärre är situationen ofta den att tidigare data saknas och att det varken finns tid eller resurser för att utföra en pilotstudie. I en sådan situation brukar man ofta nöja sig med att göra någon slags rimlighetsbedömning över stickprovsstorlek. Vad som menas med ”rimlighet” är dock ovisst, men bör inkludera faktorer som tid, ekonomi och rimlighet. Ofta används samma storlek som i en tidigare studie.

Nåväl, i detta lilla kompendium ska jag försöka beskriva vad som krävs för en stickprovsberäkning av det mer detaljerade formella slaget. Jag ska även försöka ge lite råd om hur man kan bedöma lämplig stickprovsstorlek även i mer vaga situationer, exempelvis när en enkät ska användas för första gången, eller i en undersökning baserad på intervjuer.

Teorin om beräkning av stickprovsstorlek är mycket omfattande och det finns rikhaltiga böcker i ämnet. Detta lilla kompendium ger en förenklad bild och innehåller inte sofistikerade analysmodeller för ändamålet. Men, min förhoppning är att teorin ska vara tillräcklig för exempelvis examensarbeten på högskolenivå och enklare utredningsarbeten. Anvisningarna är mer detaljerade än de man normalt finner i litteratur om vetenskaplig metodik och grundläggande böcker i statistik. Vidare ges ett förslag på hur stickprovsstorlek kan beräknas i kvalitativa undersökningar baserade på intervjuer. Något som jag inte sett i någon metodbok eller statistisk test tidigare. Jag ger även förslag på hur stickprovsberäkningen kan beskrivas i metodavsnitt.

*Liten varning!*

*Det första avsnittet är komplicerat och förutsätter grundkunskap i statistisk hypotesprövning. Är man på jakt efter tips på hur man gör i en enkätundersökning (avsnitt 3.1-3.2 mest relevanta) eller kanske i en undersökning baserad på intervjuer (kapitel 5) kan man hoppas över detta avsnitt.*

# En formell stickprovsberäkning

Till att börja med måste man utse undersökningens viktigaste variabel, den variabel som fångar in syftet på bästa sätt. Denna variabel brukar kallas för primärvariabel. Har man flera variabler som är ”lika primära” kan man dimensionera undersökningen efter den variabel som kräver störst stickprovsstorlek (i vissa fall bör man även beakta dilemmat med massignifikans och justering av p-värden).

När en primärvariabel är utsedd måste man ge ett så gott svar som möjligt på följande frågor:

* Vilken signifikansnivå ska användas?
	+ *Vanligen används 5%, men ibland kan det finnas anledning att använda en högre eller lägre nivå, exempelvis om man har flera primärvariabler*
* Vilken styrka (power) ska undersökningen ha?
	+ *Styrka är sannolikheten att förkasta den statistiska nollhypotesen förutsatt att den är falsk. Lite populistiskt kan man uttrycka det som ”chansen att vår undersökning bekräftar det vi tror, givet att det vi tror är sant”. Vanligen brukar man välja att ha minst 80% styrka. Ibland kan det finnas anledning att gå ifrån denna standard. Om man exempelvis vet att studien är mycket svår att upprepa, på grund av tillgång på objekt eller av etiska skäl, kan det vara rimligt att öka styrkan, exempelvis till 90%, vilket innebär större chans att finna mönster i ” the-one-and-only-study”.*
* Vilket test ska användas för att analysera primärvariabeln? (Alternativt vilken typ av konfidensintervall ska beräknas? )
	+ *Man måste veta hur primärvariabeln ska analyseras, antingen vilket test som ska användas (exempelvis t-test, chi-två-test,etc). Alternativt kan man välja att analysera primärvariabeln med ett konfidensintervall, man måste då ha klart för sig hur detta intervall ska beräknas. Man måste också beakta om testet ska vara tvåsidigt (mer eller mindre standard) eller ensidigt (motsvarande för konfidensintervall).*
* Hur stor variation har primärvariabeln?
	+ *Ju större variation en variabel har desto större stickprov krävs. Normalt utgår man från tidigare studier och söker där efter mått på spridning, exempelvis standardavvikelse. Finns det flera tidiga studier är det förmodligen så att variationen skiljer sig mellan de olika studierna. Det kan vara värt att studera och reflektera över varför vissa studier har lägre variation. Det är en klar fördel om man kan lära av detta för att på så sätt minimera variationen även i den egna studien. Finns inga tidigare studier får man antingen genomföra en pilotundersökning eller gissa. I vissa fall kan man även utgå från ”värsta fallet”. Exempelvis när man vill uppskatta en proportion, exempelvis: andelen lyckade operationer, kan man utgå från den proportion som är svårast att undersöka statistiskt, nämligen proportionen 50%. (Det är lättare att studera %-siffror närmre 0 eller 100% än när %-siffran ligger kring 50%.)*
* Hur starkt är det mönster man vill upptäcka?
	+ *Antag exempelvis att vi studerar skillnaden mellan en ny behandling och en gammal behandling. Vi tror (och hoppas) att den nya behandlingen är bättre. Frågan är hur mycket bättre? Är skillnaden liten krävs stort stickprov. Är skillnaden stor krävs ett mindre stickprov. Här är det viktigt att göra en realistisk bedömning. Det är inte helt ovanligt att en forskargrupp är överoptimistiska och tror att skillnaden är större än vad den är. Detta leder till att den beräknade stickprovsstorleken blir för liten. Rådet är därmed att vara lite smått pessimistisk. Utöver att den uppskattade skillnaden måste vara realistisk måste man förstås bedöma att den är av intresse. Det är knappast någon idé att utföra en jättestor studie för att finna en skillnad som är så pass liten att den är ointressant. Det är inte heller alla studier som handlar om skillnad, men oavsett om det handlar om en skillnad, en kvot, ett linjärt samband, eller vad man nu hoppas finna, måste man göra en bedömning över hur starkt man tror mönstret är.*

Med all denna information i bagaget kan man utföra beräkning av stickprovsstorlek. Beräkningarna är avancerade och bör inte utföras manuellt. Mitt råd är att använda professionella validerade programvaror för stickprovsberäkning, alternativt kontakta statistiker och förse denne med informationen beskriven ovan. Personligen använder jag oftast en programvara som heter StudySize (http://www.studysize.com/), men det finns gott om liknande programvaror, exempelvis har SPSS en tilläggsmodul för stickprovsberäkning.

Fördelen med programvara för beräkningen, utöver att man slipper krångliga beräkningar, är att det är väldigt smidigt att utföra beräkningen flera gånger under olika förutsättningar. De nämnda programvarorna ovan har även denna möjlighet. Det är en stor fördel att kunna ge flera olika förslag på hur stor spridningen är samt hur starkt mönster man vill upptäcka är. På detta sätt kan man få ett scenario över erforderlig storlek under såväl pessimistiska som optimistiska premisser.

**Exempel**. Antag att vi vill undersöka om personer som äter mycket fisk är intelligentare än gemene man. Som primärvariabel använder vi ett standardiserat IQ-test. I populationen ”gemene man” vet vi att genomsnittligt IQ är 100 och att standardavvikelsen är 15. Vi misstänker att populationen: ”Människor som äter fisk minst en gång i veckan” har ett högre genomsnittligt IQ. Vi vill utföra ett test med 5% signifikansnivå och 80% styrka. Vi har tänkt använda ett tvåsidigt t-test baserat på normalfördelning. När det gäller frågan om information utgår vi från att standardavvikelsen är 15 precis som i populationen ”gemene man”, men vi misstänker att standardavvikelsen är något lägre, eftersom detta är en specifik subpopulation som troligen är mer homogen än i populationen ”gemene man” som är en ”mix av alla människor såväl fiskätare som fiskratare”. Låt säga att en tidigare studie indikerar en standardavvikelse kring 13. Slutligen måste vi ha en idé om hur stor skillnaden kan vara i IQ? Var tror vi att denna speciella population ligger i snitt? I snitt på 103? Eller 105? Eller…? Tja, låt oss testa tanken på att denna speciella population har ett snitt på 103. Detta ger följande förslag, enligt programvaran studysize:

 

Som synes, har jag matat in 5% signifikansnivå och 80% styrka. Den statistiska nollhypotesen är att det saknas skillnad, dvs att även denna subpopulation har IQ-snitt 100. När det gäller mönster att upptäcka, i detta fall: skillnad att upptäcka, har jag matat in 103. Förslaget är att det behövs drygt 196 personer i undersökningen (standardavvikelsen satt till 15).

Nu kan ju detta tyckas vara lite av en vild gissningslek förklädd i matematik, speciellt gällande gissningen på skillnad att upptäcka som är väldigt oviss (åtminstone om vi inte har någon tidigare studie som indikerar detta). Vidare kanske standardavvikelsen i denna subpopulation är mindre (vilket kräver mindre stickprov). Jag använde samma programvara för att få följande scenario:



Här kan vi se att stickprovsstorleken varierar kraftig, från ett förslag på 53, vilket baseras på den optimistiska gissningen med stor skillnad (IQ snitt på 105) samt lägre standardavvikelse (13). Den mest pessimistiska kalkylen ger ett förslag på 1767 personer, men frågan är om det är intressant att genomföra en så stor studie för att upptäcka en så pass liten skillnad? Naturligtvis blir beräkningen säkrare om man har flera tidigare och samstämmig studier att förlita sig på. När det gäller IQ finns det gott om studier där man selekterat en subpopulation, exempelvis rökare, religiösa, osv. Intressant att jämföra med dessa studier, se hur stor skillnad man upptäckte och hur stor standardavvikelsen var. Därefter får man reflektera över om man tror att effekten av att äta fisk är i paritet med religiositet eller om den är lägre/högre, samt studera hur ofta standardavvikelsen är lägre i selekterade populationer. Med detta i bagaget kan man tämja den vilda gissningsleken något.

Programvaror för stickprovsberäkning brukar innehålla de vanligaste testerna, t-test, rangsummetest, test av proportioner, Anova, McNemar, Fisher, etc.

# Hur många enkäter?

Mitt råd är att börja med att fundera på vilken av enkätens frågor som är den allra viktigaste - vilken fråga som är den primära. Är det flera frågor som är lika viktiga får man beräkna erforderlig stickprovsstorlek för alla frågorna och därefter dimensionera undersökningen efter den fråga som kräver störst antal enkäter. Men, försök att hålla antalet primära frågor så få som möjligt – helst bara en. Det är alltid en god idé att försöka diskutera vilka frågor som fångar syftet bäst och vilka frågor som är sekundära.

Ibland kan det vara lämpligt att skapa ett index av flera primära frågor. Exempel: Antag att enkäten innehåller fem viktiga Ja/nej-frågor. Kanske kan man slå samman dessa i ett index (även kallat score), exempelvis: antalet ja-svar bland de fem frågorna, eller 0=minst 3 nej svar 1=minst 3 ja svar. På detta sätt görs de fem frågorna om till en enda variabel. Kanske är denna variabel mer valid för syftet?

Vi ska nu utgå från att vi har en fråga och utifrån denna fråga ska vi bedöma stickprovsstorleken. Den optimala lösningen är att man har tillgång till tidigare undersökningar där enkäten använts, eller en pilotstudie, samt kunskaper i statistisk hypotesprövning, då finns möjligheten att utföra en formell stickprovsberäkning beskriven i föregående avsnitt.

Men, låt oss nu studera några olika situationer.

## 3.1 Stickprov baserat på: Medelvärde

Låt oss börja med situationen där den viktigaste variabeln är kvantitativ och att vi i undersökningen kommer att beräkna medelvärdet. Om man har information om variabelns standardavvikelse går det att beräkna urvalsfelets storlek (baserat på ett 95% så kallat konfidensintervall, vilket är det vanligaste valet) genom att använda följande formel (n=stickprovsstorleken, s=standardavvikelsen):

 *Urvalsfel* = $1,96∙\frac{s}{\sqrt{n}}$

Givet ett visst krav på urvalsfelets storlek kan man nu beräkna n (stickprovsstorleken) genom följande formel:

 $n\geq \frac{1,96^{2}∙s^{2}}{urvalsfel^{2}}$

Förutsättningen för denna kalkyl är att den variabel man studerar är normalfördelad eller att stickprovet är stort. Vidare förutsätts att stickprovet är slumpmässigt valt (OSU) från en stor population (urvalet är mindre än 10% av populationen).

**Exempel:** Antag att vi vill studera kroppslängden i populationen: vuxna kvinnor som under uppväxten ätit fisk minst en dag/vecka. Vi gissar att standardavvikelsen bland dessa kvinnor är lika stor som i en större undersökning där standardavvikelsen för kroppslängd visade sig vara 6 cm. Baserat på tidigare forskning bedömer vi att antagandet om normalfördelning är rimligt (inget avgörande antagande i stora studier). Vi vill genomföra en undersökning som kan uppskatta medellängden i denna population med ett urvalsfel på max 1 cm. (Detta innebär skillnaden mellan stickprovets medelvärde och det medelvärde vi fått om vi mätt alla kvinnor i populationen, dvs urvalsfelet, får vara max 1 cm, detta med 95% sannolikhet). Sätter vi in våra värden får vi:

 $n\geq \frac{1,96^{2}∙s^{2}}{urvalsfel^{2}}=\frac{1,96^{2}∙6^{2}}{1^{2}}=138,3$

Tar man en stickprovsstorlek på 138 personer ur denna population kommer det medelvärde man får i undersökningen som mest avvika med 1 cm från medelvärdet i hela populationen (alltså det fel vi får på grund av att vi inte mäter alla kvinnor i populationen utan endast ett stickprov), detta med 95% sannolikhet. Det finns alltså 5% risk att undersökningens medelvärde avviker med mer än 1 cm från populationens snitt. Vill man minska denna risk till 1% kan man ersätta värdet 1,96 ovan med värdet 2,58 i ekvationen ovan (n blir då 240). Kan man acceptera 10% risk ersätter man 1,96 med 1,64 i ekvationen ovan (n blir då 97).

I undersökningen skulle man i metodavsnittet kunna skriva något i stil med:

*”För att kunna uppskatta medellängden i denna population med ett urvalsfel på max 1 cm beräknades stickprovsstorleken till 138 personer. Kalkylen baseras på att standardavvikelsen är 6 cm och att risken för att urvalsfelet överstiger 1 cm är 5%.”*

## 3.2 Stickprov baserat på: Procentsats %

Om den viktigaste frågan kan sammanställas i form av en proportion (procentsats) är beräkningen något enklare.

Exempel på en fråga som kan sammanställas i form av en proportion är: Äter du vanligen fisk minst en gång i veckan? Ja/Nej?

Denna fråga kan med exempelvis presenteras som % med Ja svar, ex: ”40% svarade Ja”.

Även andra typer av svarsalternativ kan sammanställas om en proportion. Exempel på fråga, eller snarare ett påstående med tillhörande skala är:

*”Man bör äta fisk minst en gång per vecka”*

Svarsalternativ:

*Tar helt avstånd – tar delvis avstånd – Varken eller – Instämmer delvis – Instämmer helt*

Här kan man tänka sig många olika redovisningar i form av procentsatser, exempelvis:

*”5% tog helt avstånd, 15% tog delvis avstånd, 50% varken eller, 20% instämde delvis och 10% instämde helt”*

Eller lite summerande: *”30% instämde delvis eller helt”*

Ja, att resultat sammanfattas i en procentsats är oerhört vanligt i enkätsammanhang. Men hur stor är precisionen för en beräknad proportion? Hur stort kan urvalsfelet bli? Om vi exempelvis vill veta hur stor andel av befolkningen som äter fisk minst en gång i veckan skulle vi teoretiskt sett kunna ta reda på det genom att fråga alla (men det hade varit svårt att genomföra en sådan undersökning med god kvalitet). Det är förstås snabbare, billigare och förmodligen mer träffsäkert att genomföra en stickprovsundersökning. Frågan är hur stort urvalsfelet riskerar att bli, dvs skillnaden mellan den proportion vi får i vårt stickprov jämfört med den sanna proportionen i hela populationen. Om det exempelvis är 40% i hela populationen som äter fisk minst en gång i veckan, kan vi då i vår stickprovsundersökning råka få 35%, 45%? Ja, naturligtvis, det beror på stickprovsstorleken. Frågar vi bara en enda person blir proportionen antingen 0% eller 100%! Ju fler vi frågar desto närmre proportionen i populationen kommer vi. Jämför med när vi singlar ett mynt, ju fler vi singlar ju säkrare kan vi vara på att andelen klave respektive krona hamnar nära 50% (förutsatt att myntet är symmetriskt). Att en uppskattad proportion är säkrare ju större stickprov vi tar är intuitivt. Hur stort urvalsfelet maximalt blir (med 95% sannolikhet) kan beräknas med följande formel (n=stickprovsstorleken och p=den proportion vi fått i stickprovet):

 $Urvalsfel= 1,96∙\sqrt{\frac{p(1-p)}{n}}$

Om man sätter ett krav på hur stort urvalsfelet får vara kan man beräkna stickprovsstorleken genom att använda denna formel:

 $n\geq \frac{1,96^{2}p(1-p)}{urvalsfel^{2}}$

Denna formel förutsätter att stickprovet är slumpmässigt. Vidare finns tumregler som säger att urvalsfraktionen (stickprovets storlek/populationens storlek) får vara max 10% (annars kan man multiplicera uttrycket under roten ur med faktorn (1-urvalsfraktionen). Vidare ska np(1-p) >5.

Nu undrar säkert den uppmärksamma läsaren hur man ska kunna beräkna urvalsfelet för en proportion p redan innan man vet värdet på p. Ja, antingen så måste man göra en bra gissning baserat på tidigare studier eller baserat på en pilotundersökning (testa undersökningen exempelvis på några tiotals personer) eller så får man räkna med värsta fallet, vilket inträffar då p är 0,5, dvs 50%.

**Exempel**. Vi vill studera andelen som är för EMU. I några tidigare undersökningar har andelen för EMU minskat från 48% till 46% och 43% i ungefär samma takt som svenska kronan stärkts gentemot Euron. Eftersom kronan stärkts ytterligare gissar vi nu att andelen för EMU sjunkit ytterligare ned till ca 40%, men vill genomföra en undersökning för att bekräfta detta. Vi vill göra undersökningen så stor att urvalsfelet blir max 0,02 (dvs 2%), med 95% sannolikhet. Då behövs följande stickprovsstorlek:

 $n\geq \frac{1,96^{2}0,4\left(1-0,4\right)}{0,02^{2}}=2305$

Vi kan i metoddelen skriva. *”För att kunna uppskatta proportionen för EMU med ett urvalsfel på max 0,01 beräknades stickprovsstorleken till 2305. Kalkylen baseras på att andelen för EMU är ca 40% och att risken för att urvalsfelet överstiger 0.01 är 5%.”*

**Exempel**: Vi vill studera andelen för EMU, men har inga tidigare studier eller någon susning överhuvudtaget om vad p kan tänkas vara. Vi räknar därför med ”värsta fallet”, dvs med p=0,5. Vi vill fortfarande ha urvalsfelet till max 0,02. Detta ger:

 $n\geq \frac{1,96^{2}0,5\left(1-0,5\right)}{0,02^{2}}=2401$

*”För att kunna uppskatta proportionen för EMU med ett urvalsfel på max 0,01 beräknades stickprovsstorleken till 2401. Kalkylen baseras på att andelen för EMU är ca 50% och att risken för att urvalsfelet överstiger 0.01 är 5%.”*

Nedan följer en tabell över erforderlig stickprovsstorlek för olika urvalsfel, alla beräkningar är gjorda utifrån värsta fallet, dvs att p=0,5.

**Tabell med största nödvändiga stickprovsstorlek (värsta fallet p=0,5) för olika urvalsfel:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Urvalsfel | 0,005 | 0,01 | 0,015 | 0,02 | 0,025 | 0,03 | 0,035 | 0,04 | 0,045 | 0,05 | 0,055 | 0,06 | 0,065 | 0,07 | 0,075 | 0,08 | 0,085 | 0,09 | 0,095 | 0,1 |
| n | 38416 | 9604 | 4268 | 2401 | 1537 | 1067 | 784 | 600 | 474 | 384 | 317 | 267 | 227 | 196 | 171 | 150 | 133 | 119 | 106 | 96 |

Eftersom det kan skilja en del i erforderlig stickprovsstorlek mellan värsta fallet (p=0,5) och proportioner närmre 0 eller 1 bör man använda den första formeln så ofta det är möjligt.

När det gäller ”Ja/Nej-frågor” har jag i exemplen ovan räknat urvalsfelet för proportionen ”Ja”, men urvalsfelet för proportionen ”Nej” blir precis lika stort (vilket beror på att p(1-p) alltid är lika stort för ”Ja” som för ”Nej”).

Men, om man har en fråga eller påståenden med flera svarsalternativ, exempelvis den femgradiga skalan i det inledande exemplet (”instämmerskalan”) får man räkna ut urvalsfelet för varje skalsteg för sig. Man bör då i förväg ha ett hum om hur svaren kommer att fördela sig på alla fem alternativen. När man har många svarsalternativ är risken förstås att det endast blir ett fåtal observationer i vissa svarskategorier och urvalsfelet blir stort (trots att p ligger långt från 0,5). Men i en analys kan man slå samman kategorier, exempelvis kan man tänka sig att slå samman ”tar helt avstånd & tar delvis avstånd” samt ”instämmer delvis & instämmer helt) för att öka antalet observationer i kategorierna. Undersökningens trovärdighet är större om man i en studieplan i förväg diskuterat sådana eventuella sammanslagningar och kriterium för när sammanslagning av kategorier ska ske. Att trovärdigheten för undersökningen är mindre om man gör det i efterhand beror på att man då kan frestas till att välja sammanslagning eller ej, beroende på vilket som ger en resultat som stödjer den egna teorin.

# Något om kategorier och bortfall

## 4.1 Att dela upp resultatet i kategorier

I tidigare avsnitt har vi diskuterat precision och hur man kan beräkna urvalsfelet. I beräkningarna har vi utgått från att ett medelvärde eller en proportion ska beräknas baserat på alla respondenterna. Men, den vanliga situationen i en undersökning är att man också vill göra uppdelningar och jämföra olika grupper. Man kanske vill jämföra proportionen ”Ja” bland kvinnor jämfört med män, eller se hur andelen ”Ja” ser ut i fem olika ålderkategorier. Eller varför inte uppdelat i fem ålderskategorier som i sin tur är uppdelade i man/kvinna, dvs sammanlagt 10 olika kategorier. Då måste man i förväg reflektera över hur respondenterna faller in i olika kategorier och bedöma att man får ett rimligt antal i varje kategori.

Jag minns en gång när jag läste en marknadsundersökning som angav att ”100% av ensamstående män, boende i lägenhet i förort till Stockholm svarat ”Ja” på frågan”. Eftersom det var en väldigt selekterad grupp och eftersom siffran 100% gjorde mig skeptisk, granskade jag hur många personer som denna siffra baserades på. Det visade sig vara 1 person! Om han svarat ”Nej” hade andelen ”Ja” varit 0%. Kort sagt, man ska se till att man får ett rimligt antal observationer i alla kategorier man anser är viktiga att dela upp materialet i. (Se föregående avsnitt för kort diskussion om sammanslagning av kategorier). Man kan förstås använda formeln för urvalsfel enligt tidigare avsnitt. Men en grov tumregel är att man bör ha åtminstone några tiotals observationer i varje kategori.

## 4.2 Bortfall

Observera att formlerna som visats i tidigare avsnitt baseras på antal svar (n). Man bör räkna med ett visst bortfall. För att bedöma bortfallet kan man kika på liknande studier och utgå från svarsfrekvens i dessa. Man ska inte bli förvånad om endast 20-30% av enkäterna blir besvarade efter första utskicket. Efter ett par utskick kan man med lite tur och mycket skicklighet kanske komma upp i 50-60%, när det gäller marknadsundersökningar. I medicinska studier har jag sett svarsfrekvenser över 90% men då efter mycket arbete med påminnelser och personliga samtal som påminnelse. Dessutom kan det bland patienter finnas starka incitament för att hjälpa till med forskning och utveckling.

Hur man minimerar bortfall och utför bortfallsanalyser är viktiga frågor för att en undersökning ska få hög kvalitet, men dessvärre är inte detta lilla kompendium forum för detta.

Men, när det gäller bortfall är det klart att man måste ta hänsyn till detta när man dimensionerar undersökningen. Om man behöver 500 enkätsvar och bedömer att bortfallet är 40% (60% svarar) måste man sända ut 500/0,6=833 enkäter (833\*0,6 är ungefär 500).

# Signifikans inte alltid tillräckligt och kort om etik

## 5.1 Signifikans inte alltid övertygande?

Vid en stickprovsstorlek på omkring 100 enkäter blir urvalsfelet max 10%, se tabellen ovan. Frågan är om man kan acceptera urvalsfel som är så pass stora? Ja, om exempelvis vi vill studera en Ja/Nej fråga i syfte att se om det är majoritet Ja eller Nej kan det räcka med 100 enkäter, eller än färre. Om exempelvis proportionen Ja ligger kring 80% spelar det knappast någon roll att urvalsfelet kan vara 10%. Proportionen på 80% är så pass skild från 50% att en felmarginal på 10% saknar betydelse, det är klar majoritet för ”Ja” även om vi drar bort 0,1 från 0,8 (då får vi ju 70%). Faktum är att vi även kunnat acceptera större felmarginaler utan problem. Om vi tar 35 personer blir urvalsfelet max:

 $Urvalsfel= 1,96∙\sqrt{\frac{0.8(1-0,2)}{35}}=0,13$

Men även ett urvalsfel på 0,13 är acceptabelt. Vi kan tryggt påstå att en majoritet (över 50%) i populationen svarar ”Ja” på frågan. Detta baserat på vårt lilla stickprov om 35 observationer.

Men, nu till dilemmat. Även om en procentsiffra på 80% baserat på ett stickprov med 35 svar med statistisk säkerhet kan ge oss slutsatsen att proportionen i populationen är över 50%, dvs att en majoritet svarar ”Ja”, är det inte säkert att denna undersökning vinner acceptans. Min erfarenhet är att människor i allmänhet och beslutsfattare i synnerhet har svårt att tro på mindre undersökningar, även om resultatet är mycket tydligt och statistisk säkerställd. Jag har flera gånger mött kommentaren: ”Men den undersökningen kan man väl inte lita på, det var ju bara ett tiotal enkäter, det brukar ju vara tusen enkäter minst i undersökningar…”.

Med andra ord finns det ibland andra dimensioner än rent statistiska att beakta. Hur stora undersökningar man brukar göra i den aktuella frågeställningen är ur ovanstående aspekt klart relevant att hålla koll på.

## 5.2 Kort om etik

En annan dimension som bör beaktas är den etiska. I varje undersökning bör man fundera på hur undersökningen kan påverka respondenterna (de som svarar). Det finns forskningsetiska principer som bygger på att man alltid måste väga eventuell nytta med undersökningen i förhållande till den eventuella skada som respondenterna kan utsättas för. En forskningsetisk prövningsnämnd bedömer om undersökningar är försvarbara att genomföra eller ej. Alla medicinska studier granskas av en dylik nämnd.

När det gäller stickprovsstorlek vill man ur etiskt perspektiv inte göra undersökningen onödigt stor, dvs påverka onödigt många personer. Men, det kan även vara felaktigt ur ett etiskt perspektiv att göra undersökningen för liten. Är undersökningen för liten riskerar man att missa relevanta mönster och då går den tänkta vinsten med studien förlorad och man har ”laborerat” med de (för få) respondenterna i onödan. Med andra ord talar etiken för att man ska undvika såväl under- som överdimensionering av en undersökning.

# Intervjuer

I en situation där intervjufrågorna är av ungefär samma karaktär som i en enkät, exempelvis frågor som kan sammanställas med en procentsats, kan samma stickprovsberäkningar utföras på samma sätt som för en enkät. Men, hur gör man i situationer där intervjufrågorna inte är av så pass enkel karaktär? Ja, för att vara helt uppriktig är min kunskap om detta ringa. Jag har endast begränsad erfarenhet av dimensioneringen av kvalitativa studier. Men, jag ska trots detta ge en statistisk syn på denna fråga. I litteraturen som beskriver kvalitativ metodik, stöter man ofta på förslaget:

”Interview until saturation”

Detta innebär att man ska fortsätta intervjua fler respondenter så länge inte informationen blir repetitiv tills man når ”mättnad”. Man ska alltså sluta när man märker att intervjuerna känns bekanta och att samma aspekter ältas om igen. När intervjuerna känns som en repris kanske man skulle kunna säga.

Nu till den, för en statistiker, intressanta frågan: går det att räkna på detta?

Till att börja med antar jag att undersökningen innehåller öppna frågor där respondenten förväntas beskriva och ge sin syn på frågeställningen/fenomenet i fråga. Jag tänker mig intervjufrågor av beskrivande karaktär, exempelvis:

”Kan du ge en beskrivning av vad som enligt dig, karaktäriserar en god pedagog?”

”Vilka tankar ställdes du inför när du fick beskedet om sjukdomen?”

”Kan du beskriva hur du uppfattar relationen med…” osv.

I ovanstående frågor kommer förmodligen respondenten nämna flera olika aspekter (infallsvinklar, tema) som svar på frågeställningen. Jag tänker mig att sammanställningen av den insamlade informationen (bandade/nedskrivna intervjuanteckningar) i princip handlar om att kartlägga och identifiera vilka aspekter som tagits upp och hur frekventa dessa är och eventuella relationer mellan aspekter och andra faktorer, exempelvis respondentens bakgrund. Man kan även tänka sig att sammanställningen bygger på att sammanställa vilka begrepp som används och hur vanliga dessa är. Kort sagt, sammanställningen handlar om att kartlägga förekomst och frekvens av gemensamma nämnare i intervjumaterialet, detta kan handla om exempelvis aspekter, tema, infallsvinklar eller begrepp. I fortsatta texten använder jag mig ”aspekt” som gemensam nämnare.

Låt oss alltså anta att vi i den kvalitativa studien syftar till att belysa de vanligaste aspekterna kring en viss frågeställning och att vi vill intervjua respondenter ända tills dessa aspekter återkommer – tills vi fått ”redundans”.

Fortfarande är detta lite vagt i min smak. Låt oss börja med att utkristallisera vad som menas med ”vanligaste aspekterna”. Vad menas med att en aspekt är vanlig? Ja, ett sätt att definiera detta på är att ange hur stor sannolikheten är för att et respondent tar upp denna aspekt. Om sannolikheten är 0,25 innebär det att denna aspekt i genomsnitt tas upp av var 4:e respondent.

Inför en kvalitativ ansats av detta slag måste man först göra en avgränsning och ange hur ovanlig aspekt man vill kunna fånga upp. Om man vill fånga upp en aspekt som har sannolikheten 0,01 förväntas man få intervjua 100 respondenter för att höra aspekten en gång. Ska man intervjua tills den återkommer en eller flera gånger blir det många intervjuer…

När man satt en gräns för hur vanliga, elelr snarare ovanliga, aspekter man vill fånga upp, dvs satt den nedre gränsen för vad som ska ingå i ”vanligaste aspekterna” återstår att definiera när man anser att redundans uppstår. Är det när alla ”vanliga aspekter”, även ”den ovanligaste bland de vanliga”, nämnts åtminstone en gång, två gånger, tre gånger eller kanske fyra? Givetvis kommer redundans uppstå för de vanligaste aspekterna snabbare än för de mindre vanliga, det kan ju finnas en aspekt som alla respondenter tar upp. Även här känns det rimligt att sätta gränsen efter den ovanligaste aspekten man vill fånga upp.

Sammanfattningsvis behöver vi:

* Ange sannolikheten för den ovanligaste aspekten bland de aspekter vi vill fånga i undersökningen
* Ange hur vi definierar redundans. Ska vi nöja oss med att den ovanligaste aspekten nämns endast en gång (och anta att intervjuerna då känns redundanta baserat på att vanligare aspekter nämnts flera gånger). Eller ska vi kräva att även den ovanligaste aspekten börjar kännas redundant, dvs när den nämns andra gången, eller tredje, eller än fler…?

När vi väl beaktat punkterna ovan måste vi bestämma vilken sannolikhet vi vill ha för att den ovanligaste aspekten når redundans med vald stickprovsstorlek. Antag exempelvis att vi vill fånga upp en aspekt som diskuteras av i snitt var femte respondent (dvs med sannolikheten 0,2). Antag vidare att vi anser att redundans uppstår när vi hört aspekten för andra gången. Antag nu att vi väljer 10 respondenter, då *förväntas* vi få 2 respondenter som nämner aspekten, men det är slumpen som avgör hur många det verkligen blir. Sannolikheten för att få höra denna aspekt för andra gången (eller fler ggr) inom 10 respondenter är ungefär 62%. Det finns alltså 38% sannolikhet för att aspekten dyker upp färre än två tillfällen även efter 10 intervjuer. Om vi vill ha 80% chans att nå redundans (aspekten nämns för andra gången) måste vi enligt sannolikhetskalkyler genomföra 14 intervjuer.

Att välja 80% sannolikhet för att ovanligaste aspekten nämns erforderligt antal gånger för redundans kan ses som en parallell till att i en undersökning ha 80% styrka.

Nedan följer en tabell som anger antal respondenter man bör intervjua för att den ovanligaste aspekten (bland de aspekter vi önskar fång upp med undersökningen) med 80% sannolikhet ska ge redundans. Tabellen innehåller olika sannolikheter för ”ovanligaste aspekten” samt olika definitioner på redundans.

Exempel på hur tabellen ska användas: Antag att den aspekt som är den ovanligaste aspekten bland de vi vill kartlägga har en sannolikhet på 5% att tas upp av en respondent (dvs tas upp av i snitt var 20:e person). Antag vidare att vi anser att intervjuandet når redundans när aspekten nämns för andra gången. Då säger tabellen att vi ska intervjua 59 personer. Gör vi det har vi 80% chans att aspekten nämns av minst två respondenter. Nöjer vi oss med att definiera redundans som att aspekten (den ovanliga) nämns endast en gång, kan vi nöja oss med 32 personer. Med 32 personer är det 80% sannolikhet att denna aspekt nämns av minst en respondent.

|  |  |  |  |
| --- | --- | --- | --- |
| Hur vanlig är aspekten? Nämns i snitt av var: (Sannolikheten inom parentes) | Redundans innebär att aspekten nämns minst en gång | Redundans innebär att aspekten nämns minst två gånger | Redundans innebär att aspekten nämns minst tre gånger |
| 50:e person (0,02) | 80 | 149 | 213 |
| 25:e person (0,04) | 40 | 74 | 106 |
| 20:e person (0,05) | 32 | 59 | 85 |
| 10:e person (0,1) | 16 | 29 | 42 |
| 5:e person (0,2) | 8 | 14 | 21 |

I metodavsnittet kan man skriva:

*”För att kunna fånga upp en aspekt som tas upp av i snitt var 20:e respondent beräknades stickprovsstorleken till 32. Med 32 respondenter är chansen 80% att en så pass ovanlig aspekt tas upp av minst en respondent i den aktuella undersökningen.”*

# Summering

* Formella kalkyler av stickprovsstorlek kräver information om:
	+ Signifikansnivå, styrka och vilket test (konfidensintervall) som ska användas
	+ Information om primärvariabelns variation
	+ Hur starkt det mönster är som man vill bekräfta med studien.
	+ Med denna information kan enkla **programvaror** användas för att få stickprovsberäkningar och scenarium.
* I en enkät kan man utgå från:
	+ Medelvärde och sätta en gräns för **urvalsfel** – detta ger ett förslag på stickprovsstorlek enligt **formel**.
	+ Procentsats och sätta en gräns för **urvalsfel** – detta ger ett förslag på stickprovsstorlek enligt **formel**.
* Man ska även beakta vilka analyser och **kategoriseringar man önskar göra**. Se till att antalet observationer är åtminstone några tiotal i varje kategori.
* Ta hänsyn till **bortfall**! Tänk på att formlerna basseras på antal svar – inte antal utskickade enkäter!
* Ta hänsyn till **etiska aspekter**
* Även kvalitativa undersökningar baserade på **intervjuer kan dimensioneras** i förväg, se utvecklad tabell.