

# Föreläsning 5: Att generalisera

Pär Nyman

25 januari 2016

Både föreläsning 4 och 5 innehåller en del matematik. På Studentportalen finns därför några sidor med räkneövningar, vilka riktar sig till personer som inte tidigare har läst någon statistik och som även har svårt för eller upplever ett motstånd mot matematik. Vi vill som vanligt understryka att matematiken aldrig är det viktigaste, men att vi erbjuder räkneövningar för att det är där som era förkunskaper skiljer sig mest åt.

## 1 Generaliseringar

I vardagligt tal är det ofta någonting negativt att generalisera. Anklagar vi våra vänner för att generalisera kan det exempelvis handla om att de i sina påståenden om andra utgår från fördomar snarare än fakta. Inom vetenskapen har inte generalisering dessa negativa konnotationer – det är tvärtom en förutsättning för nästan all forskning. Vi använder djurförsök som en indikation på hur samma preparat fungerar på människor och utifrån en opinionsundersökning kan vi – med viss felmarginal – ta reda på den svenska befolkningens åsikter.

Mitt något provokativa bud är att vi huvudsakligen bör intressera oss för generella teorier och stora populationer, snarare än enstaka fall som lyckats fånga vårt intresse. Det är så vi gör samhällsvetenskapen relevant och till ett instrument för att förstå och förändra världen, men det kräver också att vi vågar generalisera. När vi vet vilken teori eller population vi är intresserade av, bör vi därför välja de fall som maximerar våra möjligheter att dra slutsatser om populationen i stort. Eftersom vi sällan kan genomföra en totalundersökning – alltså studera alla analysenheter i den population vi är intresserade av – är generalisering vår enda möjlighet. En totalundersökning är också en logisk omöjlighet om vi vill generalisera till mer abstrakta fenomen eller till andra kontexter.

Det finns två nyckelord i det påståendet – *huvudsakligen* och *bör*. För det första finns det förstås många fall som är intressanta i sig, även om det är svårt att generalisera slutsatserna till andra fall. Kan man ge ett bidrag till litteraturen om hur Hitler kom till makten är det förstås värdefullt, även om det är en väldigt specifik situation. För det andra upplever jag att

många drivs mer av intresset för det enskilda fallet snarare än ambitionen att säga något om den större populationen eller teorin. Möjligheten att generalisera resultaten lyfts då ofta in i uppsatsens slutsatser, i stället för som en motivering till valet av fall. Ett av de viktigaste budskapen med dagens föreläsning är att tankarna om generaliseringsmöjligheter bör komma tidigt i både forskningsprocessen och den färdiga uppsatsen. Vad är det för fenomen eller population du är intresserad av? Varför bör vi i så fall studera just ditt fall? Kommer din studie att erbjuda tillräckliga förutsättningar för att vi ska kunna generalisera slutsatserna?

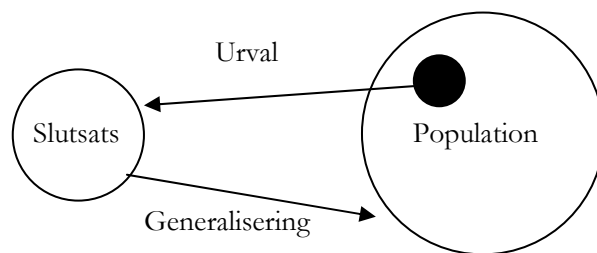
Den vanligaste invändningen mot generaliseringar är att ”varje fall är unikt”. Det är förstås sant i en trivial mening: två situationer är aldrig identiska. Men så länge det också finns saker som förenar så finns det vissa möjligheter till generalisering. Mina slutsatser är för det första att generalisering aldrig är något binärt. Förutsättningarna för att kunna generalisera är aldrig perfekta, men heller aldrig helt frånvarande. För det andra bör vi eftersträva så goda förutsättningar för generalisering som möjligt. För det tredje bör vi utforma och premiera teorier som kan appliceras i många olika situationer.

## 1.1 Olika sorters generalisering

I grunden kan man prata om att vi generaliserar i två olika dimensioner. För det första kan vi generalisera till en högre abstraktionsnivå. Teorell och Svensson kallar detta för *teoretisk* generalisering. För det andra kan vi generalisera till andra situationer, vilket Teorell och Svensson benämner *empirisk* generalisering. I regel ägnar vi oss åt båda typerna av generalisering, även om vi sällan uttrycker oss i termer av teoretisk och empirisk generalisering.

Teoretisk generalisering innebär att vi generaliserar från det vi observerat tillbaka till de teoretiska begreppen. Detta liknas ibland vid att vi ersätter alla egennamn (ett möte med Landskronas kommunfullmäktige, Sverigedemokraterna) med mer abstrakta begrepp (politiska förhandlingar i flerpartisystem, högerextrema partier). När vi funderar på vilka abstrakta kategorier som det vi studerat kan generaliseras till uttrycker vi det ibland med frågan ”vad är det här ett fall av?”. Styrkan i den teoretiska generaliseringen beror på validiteten i operationaliseringen.

Empirisk generalisering innebär att vi uttalar oss om en annan situation än den vi har studerat. Ärketypen av generalisering är när vi drar ett urval ur en större population och använder urvalet för att säga något om populationen. Vanliga exempel är opinionsundersökningar eller när SCB intervjuar några tusen svenskar för att beräkna arbetslösheten i hela landet. En andra typ av empirisk generalisering är när vi drar slutsatser om ett fall utifrån vad vi vet om ett annat fall. Vi vet till exempel att en dörrknackningskampanj ökade antalet röster för Hollande i det franska presidentvalet 2012. Om vi utifrån det drar slutsatser om hur dörrknackning fungerar i Sverige har vi gjort en



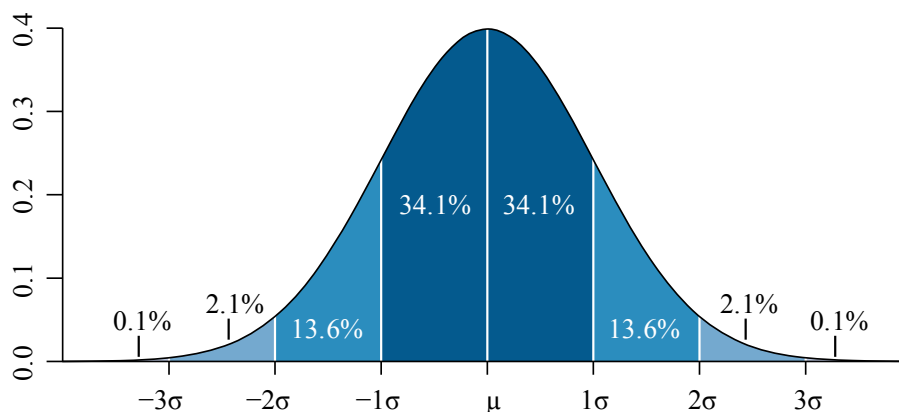
generalisering från ett fall till ett annat. För det tredje kan vi tänka oss att vi går från vad vi vet om "det stora" och uttalar oss om något mindre. Kanske utgår vi från forskning om vilka miljöer som är bäst för barns lärande när vi väljer skola åt våra barn, eller väljer att se en film som har fått höga betyg på imdb.com. Om andra har gillat filmen så borde ju jag också göra det.

Det ofrånkomliga problemet i empirisk generalisering är att det vi vill uttala oss om skiljer sig från det vi har studerat med avseende på variabler (vilka utfall tittat man på och hur är variablerna mätta?), analysenheter (oftast vill vi uttala oss om fler personer eller händelser än de vi studerar) och kontext (vår studie är begränsad i tid, plats och sammanhang). Därför är det alltid viktigt att beskriva i vilka avseenden som en studie skiljer sig från den population eller det fall som man vill dra slutsatser om. Men nöj dig inte med att bara visa på skillnaderna, utan resonera även om hur du tror att dessa skillnader kan påverka resultaten. Utgå från tidigare forskning och egna resonemang för att diskutera hur stor betydelse skillnaderna kan få och i vilken riktning du tror att de påverkar resultaten.

Vi har nu pratat om flera olika typer av generalisering. Resten av föreläsningen kommer ägnas åt en enda form – att generalisera från urval till population – men liknande resonemang kan användas på andra typer av empiriska generaliseringar. Sinnesbilden av hur vi arbetar med sådana generaliseringar illustreras i Figur 1.1. Först fastställer vi den population vi är intresserade av. Därefter väljer vi fall som maximerar möjligheten till generalisering och när vi dragit slutsatser om urvalet försöker vi generalisera dessa till populationen. Även om det inte alltid fungerar på det här viset – ofta har vi inte ens möjlighet att påverka hur urvalet ska se ut – så är det en bra bild att ha med sig.

## 2 Statistisk inferens

När vi genomför en fallstudie är vi i regel mer intresserade av en större population än vad vi är av det enskilda fallet. Samma sak gäller de flesta kvantitativa undersökningar. När SCB intervjuar några tusen svenskar om deras arbetsmarknadsstatus, gör de det för att kunna säga någonting om läget på hela den svenska arbetsmarknaden. På samma sätt genomförs opini-



Figur 1: Normalfördelningen

onsundersökningar för att ta reda på vad den svenska befolkningen tycker, och inte för att dessa mer eller mindre slumpmässigt valda respondenter har så väldigt intressanta åsikter.

Dessa generaliseringar är emellertid alltid behäftade med viss osäkerhet. Hur kan vi veta att det urval vi studerar är en bra beskrivning av hela populationen? Hur stora felmarginaler bör vi räkna med? När vi använder statistiska metoder för att dra slutsatser om en större population än den vi har studerat kallas det för statistisk inferens. Om urvalet är slumpmässigt och alla analysenheter i populationen har samma sannolikhet för att inkluderas i urvalet så kommer våra urval *i genomsnitt* att återspegla populationen. Vår enda felkälla är slumpen. Under vissa antaganden kan vi beräkna hur stora felmarginaler som krävs för att tillgodose den osäkerhet som slumpen ger upphov till. För att förstå dessa antaganden ska vi först bekanta oss med något som kallas för normalfördelningen.

## 2.1 Normalfördelningen

I det här avsnittet ges en mycket kortfattad beskrivning av normalfördelningen. Syftet är att visa vart metoderna i avsnitt 2.2 kommer ifrån. Det är förståeligt om det är svårt att följa resonemangen i det här avsnittet och ni behöver inte heller förstå varje ord. För en djupare förståelse av normalfördelningen och andra statistiska fördelningar rekommenderas att läsa en kurs i statistik.

Figur 1 visar hur en normalfördelning ser ut. Höjden på kurvan visar hur vanligt ett visst värde är. Som ni kan se är det vanligast med värden i mitten av fördelningen, nära medelvärdet ( $\mu$ ). Det som gör normalfördelningen så attraktiv är att vi vet hur stor andel av värdena som befinner sig inom ett visst intervall. Exempelvis befinner sig 68,2 procent av värdena ( $34,1+34,1$ )

inom en standardavvikelse ( $\sigma$ ) från medelvärdet och 95,4 procent inom två standardavvikelser från medelvärdet ( $13,6+34,1+34,1+13,6$ ). Viktigare för vårt syfte är något som inte kan utläsas direkt ur diagrammet: 90 procent av värdena återfinns inom 1,65 standardavvikelser från medelvärdet, 95 procent av värdena hittar vi inom 1,96 standardavvikelser från medelvärdet och 99 procent av värdena ligger inom 2,58 standardavvikelser från medelvärdet. Dessa värden kallas för kritiska värden och vi kommer snart att återkomma till dem. Om fördelningen inte avser hur en variabel är fördelad, utan visar hur sannolika olika urval är, brukar vi använda ordet standardfel i stället för standardavvikelse. Ett stort standardfel innebär att, om vi drog ett oändligt antal urval, så skulle medelvärdena för dessa urval skilja sig mycket från varandra och från medelvärdet i populationen. Ett litet standardfel innebär att de flesta urval ligger nära varandra och populationsvärdet.

Om vi kan anta att en population av möjliga urval följer en normalfördelning, då kan vi även beräkna sannolikheten för att dra ett visst urval ur den populationen, givet att vi känner till populationens medelvärde och standardfel. Detta kan vi använda för att beräkna osäkerhetsintervall kring våra skattningar och för att testa om ett hypotetiskt populationsmedelvärde är rimligt.

Students t-fördelning är en nära släkting till normalfördelningen. Vid stora urval har de båda fördelningarna samma kritiska värden, men vid små urval är värdena något större för t-fördelningen. På den här kursen kommer vi använda normalfördelningen när vi beräknar konfidensintervall runt en proportion och t-fördelningen när vi beräknar konfidensintervall runt ett medelvärde eller runt en regressionskoefficient. Detta kräver egentligen att vi gör vissa antaganden, vilka Teorell och Svensson beskriver mycket kortfattat, men det är i regel ganska oproblematiskt och ingenting ni behöver sätta er in i.

## 2.2 Konfidensintervall

När vi undersöker ett urval ur en population, kan vi enkelt beräkna sådant som medelinkomsten i urvalet eller hur stor andel som är negativt inställda till kärnkraft. Dessa värden kallas för punkttestimat och är ofta vår bästa gissning av vad medelinkomsten (ett medelvärde) eller andelen kärnkraftsmotståndare (en proportion) är i populationen.

Det kan tyckas vanskligt att uttala sig om en hel population när vi bara har studerat en bråkdel av den, men detta görs hela tiden. När vi säger att arbetslösheten i Sverige är 8,7 procent, då bygger den siffran i regel på en intervjuundersökning med några tusen personer. Förmodligen är arbetslösheten i populationen något högre eller något lägre än just 8,7 procent. När vi gör den typen av generaliseringar är det viktigt att vi kan beskriva hur stor osäkerheten är. Är det rimligt att arbetslösheten i själva verket är 5 eller 12 procent, eller kan vi räkna med att den ligger någonstans

mellan 8,5 och 8,9 procent?

En vanlig metod för att illustrera den osäkerheten är att beräkna konfidensintervall, inom vilket vi tror att populationens medelvärde eller proportion återfinns. För att uttrycka hur säkra vi är på denna slutsats använder vi oss av en säkerhetsnivå, vanligen 99, 95 eller 90 procent.

Låt oss anta att vi beräknat ett konfidensintervall för en säkerhetsnivå på 95 procent, vilket är vanligast. Med detta menar vi att om vi drog ett oändligt antal urval från populationen, skulle 95 procent av urvalen täcka in populationens medelvärde. Notera att man inte bör uttrycka detta som att *sannolikheten* för att populationsmedelvärdet ligger inom intervallet är 95 procent. I stället uttrycker vi oss som att vi *vid 95 procents säkerhetsnivå* kan säga att populationens medelvärde ligger inom intervallet. Andra ord för säkerhetsnivå är konfidensnivå eller konfidensgrad.

Helt säkra (100 procent) kan vi nästan aldrig vara, om vi vill ha meningsfulla intervall. Intervjuar vi 1 000 personer och alla är kärnkraftsmotståndare, finns det förstås en möjlighet att alla andra svenskar är positivt inställda till kärnkraft och att andelen motståndare är mindre än 0,1 procent. I stället uttalar vi oss om hur det förmodligen ligger till, alltid med en viss risk att vi har fel.

Ekvationerna för att beräkna ett konfidensintervall bygger på en enkel princip. I mitten av intervallet har vi vårt punkttestimat och runt detta en felmarginal åt varje håll. Felmarginalens storlek beror på två faktorer. För det första vilken kritiskt värde vi har, vilket framför allt kommer av vilken säkerhetsnivå vi valt. För det andra hur stort standardfel som urvalen har, alltså hur stor spridningen skulle vara om vi drog flera urval. Ett lågt standardfel innebär att punkttestimatet i alla urval ligger ganska nära populationens medelvärde eller proportion, medan ett högt standardfel innebär att punkttestimaten i urvalen uppvisar en stor spridning och ofta ligger långt ifrån populationsvärdet.

Givet att vi inte känner till populationens standardavvikelse, utan måste skatta den utifrån standardavvikelsen i urvalet, kan konfidensintervallet för populationens medelvärde skrivas som en funktion av medelvärdet i urvalet ( $\bar{x}$ ), det kritiska värdet för t-fördelningen ( $t_{kv}$ ), urvalets standardavvikelse ( $s$ ) samt antalet observationer i urvalet ( $n$ ). Den sista termen ( $\frac{s}{\sqrt{n}}$ ) är vår uppskattning av standardfelet. Tecknet  $\pm$  innebär att vi erhåller två värden, vilka tillsammans avgränsar konfidensintervallet.

$$\bar{x} \pm t_{kv} \times \frac{s}{\sqrt{n}} \quad (1)$$

När vi beräknas konfidensintervallet för en proportion kan vi i regel anta att proportionens sannolikhetsfördelning följer en normalfördelning (Teorell och Svensson, s. 147). Ekvationen för konfidensintervall för proportioner kan därför skrivas som en funktion av proportionen i urvalet ( $p$ ), det kritiska

värdet ur normalfördelningen ( $z_{kv}$ ) och urvalets storlek ( $n$ ). Precis som tidigare utgör den sista termen ( $\sqrt{\frac{p(1-p)}{n}}$ ) vår uppskattning av standardfelet.

$$p \pm z_{kv} \times \sqrt{\frac{p(1-p)}{n}} \quad (2)$$

Ibland vill vi beräkna ett konfidensintervall för skillnaden mellan två grupper. Anledningen till det kan exempelvis vara att utifrån två urval beräkna ett intervall för hur stor en förändring har varit mellan två tidpunkter eller för hur stor skillnad det finns mellan män och kvinnor. Vi tänker oss att urvalen är dragna ur två olika populationer och att vi vill veta skillnaden mellan de två populationerna. Om noll inte ligger i det konfidensintervall vi beräknar kan vi utesluta möjligheten att de två grupperna inte skiljer sig åt. Det uttrycker vi som att "förändringen är statistiskt säkerställd", "fler män än kvinnor röstar på moderaterna" eller att "skillnaden mellan män och kvinnor är statistiskt signifikant".

Ett så enkelt test kan också användas för att analysera olika typer av experiment. Exempelvis kan vi beräkna ett konfidensintervall för skillnaden i tillfriskningstid mellan den grupp av försökspersoner som fått ett nytt läkemedel och den grupp som endast fått sockerpiller. Om värdet noll befinner sig i konfidensintervallet kan vi inte utesluta att skillnaden mellan läkemedlet och sockerpillret var noll. Det uttrycks ofta som att läkemedlet inte hade någon statistiskt signifikant effekt. Men om värdet noll inte återfinns i intervallet är skillnaden mellan grupperna för stor för att bero på slumpen och vi kan dra slutsatsen att läkemedlet hade effekt.

Nedanstående formel kan användas för att beräkna konfidensintervall för differensen mellan två medelvärden. Informationen som behövs är medelvärde ( $\bar{x}_i$ ), standardavvikelse ( $s_{x_i}^2$ ) och urvalsstorlek ( $n_{x_i}$ ) för de två grupperna samt ett kritiskt värde ( $t_{kv}$ ) vilket beror på vår säkerhetsnivå.

$$\bar{x}_1 - \bar{x}_2 \pm t_{kv} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (3)$$

Och dagens sista formel används för att beräkna ett konfidensintervall för skillnaden mellan två proportioner. Proportionerna i de två urvalen betecknas  $p_1$  och  $p_2$  och urvalens storlek  $n_1$  och  $n_2$ .

$$p_1 - p_2 \pm z_{kv} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \quad (4)$$

### 2.3 Räkneexempel

Låt oss ta ett lite mer konkret exempel. Antag att vi har undersökt den genomsnittliga månadsinkomsten i ett urval av 1 000 yrkesarbetande personer boende i Sverige, men att vi egentligen är intresserade av att beskriva den

genomsnittliga månadsinkomsten för alla yrkesarbetande personer boende i Sverige. Medelinkomsten i urvalet är 23 000 kr och standardavvikelsen ( $s$ ) är 5 700 kr. Vi tycker dock att det känns lite för skakigt att enbart använda vårt punkttestimat (23 000 kr) för att uttala oss om medelinkomsten i populationen. Istället vill vi skatta ett intervall som med en viss säkerhet innesluter populationsvärdet. På basis av den tidigare diskussionen kan vi dela upp beräkningen och tolkningen av detta konfidensintervall i fyra steg:

1. Välj säkerhetsnivå. Den vanligaste säkerhetsnivån i samhällsvetenskapliga sammanhang är 95 procent. Detta innebär att om man skulle göra ett oändligt antal urval från populationen så skulle 95 av 100 konfidensintervall beräknade på detta sätt innesluta det faktiska populationsmedelvärdet. När vi som i detta fall har ett urval på 1 000 individer motsvaras en 95-procentig säkerhetsnivå av ett kritiskt  $t$ -värde på 1,96, det vill säga  $t_{kv} = 1,96$ . Hade vi istället valt 90 procent säkerhetsnivå så hade vi här fått  $t_{kv} = 1,65$  och om vi valt 99 procent säkerhetsnivå  $t_{kv} = 2,58$ .
2. Ta reda på urvalsmedelvärde, urvalsstandardavvikelse och urvalsstorlek. I detta exempel har vi redan tillgång till dessa värden, men vanligtvis måste man beräkna dem på basis av de data man samlat in. Här är  $\bar{x} = 23000$ ,  $s = 5700$  och  $n = 1000$ .
3. Sätt in värdena i formeln för konfidensintervallet. Vi erhåller då ett undre (22 647) och ett övre (23 353) värde för intervallet.

$$\bar{x} \pm t_{kv} \times \frac{s}{\sqrt{n}} = 23000 \pm 1,96 \times \frac{5700}{\sqrt{1000}} = 23000 \pm 353 \quad (5)$$

4. Detta uttrycker vi som att vid 95 procents säkerhetsnivå, ligger den genomsnittliga månadsinkomsten för yrkesarbetande boende i Sverige någonstans mellan 22 647 och 23 353 kr.

Vi tar ett exempel till, men den här gången använder vi en proportion i stället för ett medelvärde. I den senaste Sifo-undersökningen (augusti 2013) uppgav 30,5 procent ( $p = 0,305$ ) av de tillfrågade att de skulle rösta på Socialdemokraterna om det var val idag. Antalet intervjuade var 1 916. För att kunna uttala oss om partiets popularitet i hela landet måste vi beräkna ett konfidensintervall.

Liksom tidigare väljer vi en säkerhetsnivå på 95 procent. Eftersom vi kan anta en normalfördelning i detta fall innebär ett kritiskt värde ( $z_{kv}$ ) på 1,96 oavsett urvalets storlek. Detta är allt vi behöver för att beräkna konfidensintervallet. Vi fyller i ekvationen med uppgifterna  $p = 0,305$ ,  $z_{kv} = 1,96$  och  $n = 1916$ .



$$p \pm z_{kv} \times \sqrt{\frac{p(1-p)}{n}} = 0,305 \pm 1,96 \times \sqrt{\frac{0,305(1-0,305)}{1916}} = 0,305 \pm 0,02 \quad (6)$$

Med andra ord sträcker sig intervallet från 0,285 till 0,325. Vi drar då slutsatsen att vid 95 procents säkerhetsnivå skulle mellan 28,5 procent och 32,5 procent av svenskarna rösta på Socialdemokraterna om det var val idag.

### 3 Strategiska urval

När vi inte kan förlita oss på statistisk inferens måste vi använda andra tekniker och resonemang för att generalisera slutsatser om vårt fall eller urval till en större population. Detta uttrycker vi ofta som olika metoder för att välja fall.

Det finns många strategier för att välja fall. Idag kommer vi att prata om representativa fall och kritiska fall. När vi kommer in på förklaringar kommer vi även att prata om Mills Method of Difference och Method of Agreement (samt Most Similar Systems Design och Most Different Systems Designs). Och efter passen om regressionsanalys kommer vi även att gå igenom hur man kan välja fall utifrån resultaten i en extensiv undersökning (illustrativa fall och avvikande fall).

På kursen pratar vi om två typer av strategiska urval. För det första har vi representativa fall, vilka är typiska för populationen i de avseenden som vi tror är viktiga för våra resultat. Studerar vi de långsiktiga politiska effekterna av revolutioner vill vi inte studera den senaste, den blodigaste eller den mest utdragna revolutionen, utan i stället en revolution som påminner om genomsnittet av revolutioner. Vi ställer då upp vissa egenskaper, som vi motiverar utifrån teori eller konventioner, och väljer det fall som kommer närmast.

För det andra har vi kritiska fall, så kallade most likely- och least likely-fall. Ett sådant val av fall maximerar respektive minimerar sannolikheten för att en viss hypotes eller utsaga ska få stöd, givet vissa teorier eller förningar om vad som påverkar denna sannolikhet. Genom att ge en teori de sämsta tänkbara förutsättningarna för att få rätt, har vi trots att det handlar om ett enstaka fall ett starkt argument till stöd för teorin om den skulle visa sig stämma. På motsvarande vis kan vi förkasta en teori som visar sig vara fel, även när vi ger den de bästa tänkbara förutsättningarna. Därför används least likely-fall för att ge stöd åt hypoteser och most likely-fall för att förkasta hypoteser.

Det här sättet att argumentera används vid både kvalitativa och kvantitativa studier. Kom ihåg att det vi vill uttala oss om ofta skiljer sig från det vi har studerat med avseende på analysenheter, kontext och hur variablerna är mätta. I en kvantitativ studie kan vi dra ett slumpmässigt urval

av analysenheter och med hjälp av statistisk inferens dra slutsatser om alla analysenheter i populationen, men vi kan göra på samma sätt för att uttala oss om alla tänkbara variabler eller kontexter. Därför är statistisk inferens aldrig tillräckligt utan måste kompletteras med andra tekniker.

### 3.1 Att välja strategiska fall - en tillämpning

Låt oss anta att vi vill studera en variant av medianväljarteomet. Närmare bestämt ska vi testa hypotesen att de politiska partierna försöker att lägga politiska förslag som attraherar medianväljaren, snarare än förslag som partiet tror på eller som är populära bland partiets kärnväljare. För att göra det ska vi studera den politiska processen i en svensk kommun.

Ett tillvägagångssätt skulle kunna vara att välja en typisk kommun som i flera avseenden är representativ för svenska kommuner. Vilka faktorer som avgör vad som är en typisk kommun beror förstas på vad det är vi studerar – faktorerna måste vara relevanta. I det här fallet vill vi kanske ha en kommun med ett normalstort antal partier i kommunfullmäktige och där det är maktskifte ungefär lika ofta som i resten av riket, eftersom vi utifrån sunt förnuft och tidigare forskning tror att dessa faktorer påverkar förutsättningarna för medianväljarteomet. På samma sätt är det förmodligen irrelevant huruvida kommunen vi väljer har ett medelbra fotbollslag och en genomsnittlig kaffekonsument. Däremot är det aldrig fel att även inkludera mer generella faktorer, såsom antalet invånare i kommunen. Vi kanske inte har en tydlig idé om varför det är relevant, men motiverar i stället valet med att variabeln ofta inkluderas av andra forskare, att den samvarierar med en mängd andra variabler eller att den har visat sig vara relevant i andra sammanhang.

Ett annat möjligt tillvägagångssätt är att välja ett kritiskt fall som antingen maximerar eller minimerar sannolikheten för att medianväljarteomet ska stämma. För att göra det behöver vi en teori om vad som påverkar sannolikheten för att hypotesen stämmer. Låt oss anta att de politiska förslagen i högre grad anpassar sig till medianväljaren om kommunen brukar ha jämna val och det finns två stora partier som tenderar att alternera om makten.

Ett most likely-fall är ett fall som i alla avseenden har en maximal sannolikhet för att hypotesen ska stämma. Man skulle därför kunna argumentera för att Stockholm är ett most likely-fall, eftersom nästan varje val de senaste decennierna har resulterat i ett maktskifte. Dessutom är Moderaterna starkare i Stockholm än i resten av landet, vilket gör att de tillsammans med Socialdemokraterna har varit mycket dominerande i Stockholmspolitiken. Om vi hittar stöd för medianväljarteomet i Stockholm, är det inte särskilt intressant. Om vi däremot finner att den politiska processen i Stockholm – trots sina gynnsamma förutsättningar – inte alls följer vad medianväljarteomet implicerar, då har vi ett starkt argument för att det är en dålig beskrivning av politiken även i andra kommuner. Most likely-fall används

därför för att förkasta hypoteser.

Motsatsen till most likely-fall kallas least likely-fall och är alltså fall som har en minimal sannolikhet för att en hypotes ska stämma. Om vi fortsätter på medianväljarteoremet kan vi tänka oss en kommun som Munkfors, där Socialdemokraterna ensamla erhö11 över 57 procent av rösterna i det senaste kommunvalet. Det är också ett gammalt industrisamhälle med hög medelålder, vilket förmodligen innebär att väljarna där inte är särskilt rörliga. Behovet av att anpassa politiken efter en strategiskt viktig mittenväljargrupp borde där vara minimalt. Om vi trots detta hittar stöd för hypotesen, har vi goda skäl att tro att medianväljarteoremet kan förklara politiska förslag även i andra kommuner. Least likely-fall används därför för att ge stöd åt hypoteser.

Dagens föreläsningar handlar om beskrivningar, men vi kan också använda kritiska fall i en förklarande ansats. Då motsvarar most likely-fall ett fall som i alla avseenden – förutom den förklarande variabel vi är intresserade av – har en maximal sannolikhet för ett visst utfall. På motsvarande vis är ett least likely-fall en situation där allt talar emot ett visst utfall, förutom just den variabel vars betydelse vi vill argumentera för.

Exempelvis argumenterar Reilly och Phillipot (2002) för att Papua Nya Guinea är ett kritiskt fall som skänker starkt stöd till hypotesen att demokrati fungerar i samhällen som är rika på socialt kapital. Anledningen till att Papua Nya Guinea är ett kritiskt fall är att landet i alla avseenden har dåliga förutsättningar för att vara demokratiskt – utöver att vara ett fattigt land talas där över 800 språk! Trots detta har landet varit demokratiskt sedan 1964, vilket författarna förklarar med ett högt socialt kapital.

## 4 Liten repetition och ordlista

- Generalisering: Att utifrån ett urval uttala oss om den större mängd analysenheter som vi egentligen vill studera.
- Punktestimat: Ett tal som används för att, med utgångspunkt i ett urval, gissa sig till (medel-)värdet i en population.
- Konfidensintervall: Det intervall inom vilket man, med en viss säkerhet, tror att ett populationsvärde ligger.
- Totalundersökning: Att studera alla fall
- Strategiskt urval: Att generalisera genom att välja ett fall som gör på grund av sina specifika karaktäristiska gör det möjligt att dra vissa generella slutsatser
- Representativa fall: Representativa för andra fall

- Most likely case: Ett kritiskt fall med gynnsamma omständigheter (om teorin inte får stöd här, får den sannolikt inte stöd någon annanstans heller. Man gör det lätt för teorin.)
- Least likely case: Ett kritiskt fall med ogynnsamma omständigheter (om teorin får stöd här får den sannolikt stöd även under mindre ogynnsamma omständigheter. Man gör det svårt för teorin. )